A Method for the Measurement of the Latency Tolerance Range of Western Musicians

Jorge Medina Victoria

Department of Electrical and Electronic Engineering, Cork Institute of Technology, info@jormedinavic.de

Follow this and additional works at: https://sword.cit.ie/engdiss

Part of the Electrical and Electronics Commons, Musicology Commons, and the Other Music Commons

Recommended Citation

This Doctoral Thesis is brought to you for free and open access by the Engineering at SWORD - South West Open Research Deposit. It has been accepted for inclusion in PhDs by an authorized administrator of SWORD - South West Open Research Deposit. For more information, please contact sword@cit.ie.
A Method for the Measurement of the Latency Tolerance Range of Western Musicians

Jorge Enrique Medina Victoria

Volume I of II: Thesis

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.D.).

Department:
Electrical & Electronic Engineering

Supervisors:

Dr. Joe Connell, Cork Institute of Technology
Dr. Susan O’Regan, Cork Institute of Technology
Prof. Dr. Kyrill Fischer, Darmstadt University of Applied Sciences

Submitted to Cork Institute of Technology, May 2019.
Declaration

This thesis is entirely my own work except where otherwise accredited. The thesis has not been submitted for an award at any other institution.

Jorge Enrique Medina Victoria, May 2019

Dr. Joe Connell, Cork Institute of Technology

Dr. Susan O'Regan, Cork Institute of Technology

Prof. Dr. Kyrill Fischer, Darmstadt University of Applied Sciences
Abstract

This thesis presents a new systematic method to measure the ability of western musicians to cope with latency. The core of the method is a listening test and the development of a measure. The viability of the method is statistically tested with an empirical observation of 31 test subjects performing on 17 different musical instruments.

The primary goal of the investigation is the development of a systematic, reliable and replicable method that can be applied to different western music instruments, in order to provide data for analysis on latency issues while performing music in non-collaborative performances. In addition, a measure of the latency range tolerance for different musical instruments groups is defined and developed on the basis of the data gathered.

The experimental application of the method developed provides empirical results showing that different musical instruments produce different results with regard to latency. This indicates that, in terms of latency, the type of musical instrument plays a decisive role with respect to the ability to perform music. Furthermore, evidence of the dissimilarities in the ability to cope with latency could be observed and classified according to musical tempo and the four musical instrument groups of aerophones, chordophones, idiophones and membranophones.

This investigation is a further contribution to the understanding of the relationship between musician, musical instrument and musical performance.

Keywords: Latency, Tempo, Latency Tolerance Range, Western Musical Instruments, Musical Instrument Groups.
Acknowledgements

Firstly I’d like to thank my advisors Dr. Joe Connell, Dr. Susan O’Regan and Prof.Dr. Kyrill Fischer for all the support, inspiration and ideas in topics such as musicology, physics and technology. This work wouldn’t be possible without the help of Prof.Dr. Arnd Steinmetz who facilitated my studies at the Darmstadt University of Applied Sciences and the Cork Institute of Technology. I also would like to thank Hugh McCarthy for his support from the beginning of my Ph.D. and for his valuable help while I was in the CIT Cork School of Music.

Special thanks to Dr. Janina Fangel for all the questions answered, advice and support in those difficult moments. Thanks also to the people I met through the Graduate School, especially Dr. Fred Joop who aroused my interest in statistics, Dr. Michael Dyrenfurth who patiently explained how to write a Ph.D. and Florian Junge of the CCSOR who gave of his time for the statistical review.

Many thanks to Dario Quiñones, a musician and friend, who was there to help on any artistic and technical inquiry, to Martin Streit who always read my papers and patiently explained many things, to Ingo Weismantel who read and corrected the final draft and Carolina Ramirez (Cabbage) for helping me with the pictures but especially for her supporting over the whole time.

Thank you to all the musicians and music students who took part in the listening test. I’d particularly like to thank Filipe Abbreu and all the professional musicians of the Staatstheater Darmstadt, as well as Ben Brigmann and his former teachers, Gernot Hosang and Andreas Pfaul, who facilitated many listening tests with their big band and orchestra students.
Finally, thank you to my parents, Cecilia Victoria and Jorge Medina, and my brother Javier on the other side of the world for their unconditional love, for believing in me and for their support in difficult times. Thanks to my mother-in-law, Charlotte Nolte-Uhl, my sister-in-law, Anke Uhl and “Tante” Karen Schmidt for helping me with difficult translations and spending much time taking care of my little Ferdi. Friederike my beloved wife, thank you for being there; none of this would be possible without your help and the patience, commitment and time you have given me while taking care of our Friedrich Ferdinand.
Acronyms

AAF  Altered Auditory Feedback
BPM  Beats per Minute
DAF  Delayed Auditory Feedback
DIP  Distributed Immersive Performance
DVF  Delayed Visual Feedback
EDA  Exploratory Data Analysis
EPT  Ensemble Performance Threshold
IEM  In-Ear Monitoring
IOI  Interonset Interval
IPD  Inter Pulse Delay
KDE  Kernel Density Estimation
LAT  Latency Adaptive Tempo
LTR  Latency Tolerance Range
MIDI  Musical Instrument Digital Interface
NMP  Networked Music Performance
OWD  One-Way Delay
PAT  Perceptual Attack Time
RTT  Round-Trip Time
SPL  Sound Pressure Level
TIP  Temporal Information Processing
TO  Temporal Order
TOT  Temporal Order Threshold
List of Publications Derived From this Research

1. Medina Victoria, Jorge., An Exploratory Data Analysis of the Ability of Western Musicians to Cope with Latency, to be presented at the CERC 2019. Collaborative European Research Conference, March 29, Darmstadt, Germany, 2019

2. Medina Victoria, Jorge., Latency Tolerance Range Measurements in Western Musical Instruments, AES 146th Convention, March 20-23, Dublin, Ireland, 2019


5. Medina Victoria, Jorge., An Approach to Quantifying the Latency Tolerance Range in Non-collaborative Musical Performances, AES 136th Convention, April 26-29, Berlin, Germany, 2014


1 Peer reviewed conference papers
## Contents

### Chapter 1 Introduction
1.1 Background ................................................. 1
1.2 Description of the Problem ............................... 4
1.3 Research Motivation ...................................... 6
1.4 Research Question ......................................... 8
1.5 Thesis Structure ........................................... 10

### Chapter 2 Literature Review
2.1 Western Orchestral Musical Instruments ................. 12
  2.1.1 A Model of Western Musical Instruments ............. 12
  2.1.2 Musical Instruments Groups .......................... 13
  2.1.3 Idiophones ........................................... 15
  2.1.4 Membranophones ..................................... 17
  2.1.5 Chordophones ........................................ 18
  2.1.6 Aerophones .......................................... 21
2.2 Music and Network Technology ............................ 27
  2.2.1 Technical Fundamentals of Networks ................. 30
  2.2.2 Networked Music Signal Path ......................... 30
  2.2.3 Taxonomy of Musical Networks ....................... 37
2.3 Sound and Latency Perception ............................. 40
  2.3.1 Sound Perception in Time ............................. 40
  2.3.2 Musical Events and Time and Body Movements ........ 43
2.4 Latency Measurement Approaches .......................... 46
  2.4.1 Latency and the Performance of Music ............... 47
  2.4.2 Effects of Latency on Musicians ...................... 54
  2.4.3 Delayed Auditory Feedback (DAF) ..................... 68
Appendix F  Measurements of Sound Pressure Level on Instruments and Rooms          F1

Appendix G  General Information Sheet for Participants                        G1

Appendix H  Listening Test Consent Form                                        H1

Appendix I  Questionnaire for the Listening Test                               I1

Appendix J  Friedman Test Data and Results                                     J1

Appendix K  Data tables
            K.1 Complete Latency Values for all musical instruments          K1
            K.2 Categorical Information                                      K4
            K.3 Recording Information for all musical instruments           K6

Appendix L  R code for the statistical analysis                                L1

Appendix M  Listening test results and questionnaire results                  M1

Appendix N  CIT Research Ethics Application Form                               N1
List of Tables

2.1 Taxonomy of the musical instruments of the western modern orchestra . . 14
2.2 Body parts and playing tools involved in sound generation . . . . . . . . 26
2.3 NMP research categories [104] . . . . . . . . . . . . . . . . . . . . . . . . . 29
2.4 Taxonomy of a networked music performance [104] . . . . . . . . . . . . . 38
2.5 Network musical approaches [228] . . . . . . . . . . . . . . . . . . . . . . 39
2.6 Timescale of musical events [17] . . . . . . . . . . . . . . . . . . . . . . . . 42
2.7 Strategies to cope with latency in collaborative performances adapted from [42] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
2.8 Types of errors produced with DAF in music performance [168] . . . . 72
3.1 Questionnaire for the pilot test . . . . . . . . . . . . . . . . . . . . . . . . 114
3.2 Brief taxonomic description of the listening test . . . . . . . . . . . . . . 120
4.1 Statistical summary for measurements related to the listening test . . . 141
4.2 Sound pressure level range for western musical instruments . . . . . . . 142
4.3 Pilot test musical instruments . . . . . . . . . . . . . . . . . . . . . . . . . 149
4.4 Latency values for the pilot test . . . . . . . . . . . . . . . . . . . . . . . . 150
4.5 List of numerical variables . . . . . . . . . . . . . . . . . . . . . . . . . . . 156
4.6 List of categorical variables . . . . . . . . . . . . . . . . . . . . . . . . . . 157
4.7 Summary of the latency value results ($L_d$) . . . . . . . . . . . . . . . . 158
5.1 Descriptive statistics for the tempo in BPM . . . . . . . . . . . . . . . . . 160
5.2 Shapiro test results for normality . . . . . . . . . . . . . . . . . . . . . . . 173
5.3 Descriptive statistics for tempo in BPM (average of all 3 metronomes) . . 184
5.4 Descriptive statistics for the aerophones . . . . . . . . . . . . . . . . . . . 191
5.5 Descriptive statistics for the chordophones . . . . . . . . . . . . . . . . . . 192
5.6 Descriptive statistics for the membranophones . . . . . . . . . . . . . . . 194
5.7 Descriptive statistics for the idiophones . . . . . . . . . . . . . . . . . . . 195
5.8 Descriptive statistics for air reed sound generation .................................. 201
5.9 Descriptive statistics for lip reed sound generation .................................. 202
5.10 Descriptive statistics for mechanical reed sound generation ............... 202
5.11 Descriptive statistics for bowed sound generation ............................... 203
5.12 Descriptive statistics for plucked sound generation ............................... 203
5.13 Descriptive statistics for struck sound generation (chordophones) ......... 204
5.14 Summary of the characteristics of the data ............................................ 212
5.15 Friedman test results for the three different metronomes at each tempi .... 214
5.16 Post hoc Friedman test for metronomes at 150 BPM ............................. 215
5.17 Friedman test results supporting the alternative hypothesis $H_1$ .......... 216
5.18 Post-hoc test significant results for the pairwise comparisons ................ 217
5.19 List of the musical instruments of the test ............................................ 219
5.20 Experimental data for all tempi .......................................................... 222
5.21 Friedman test results for the influence of the different tempi ................. 222
5.22 Post-hoc test for all pairwise comparisons for tempi 90 BPM to 210 BPM 223

B.1 Measurement of the metronome level (headphone level = -20dB) ......... B2
B.2 Measurement of the metronome level (headphone level = -15dB) ............ B3
B.3 Measurement of the metronome level (headphone level = -10dB) ............ B3
B.4 Measurement of the metronome level (headphone level = 0dB) ............... B4

C.1 Headphones attenuation ................................................................. C2

D.1 Audio interface output level measurements ........................................... D2
D.2 Linear regression coefficients .......................................................... D4

E.1 Fireface 400 latency measurement values ........................................... E2
E.2 Fireface UCX latency measurement values ......................................... E4

F.1 Musical instruments recording reference values .................................... F2
F.2 Room measurements ...................................................................... F3

I.1 Questionnaire for the listening test ....................................................... I1

J.1 Latency values for the three different metronomes at 90 BPM ............. J1
J.2 Latency values for the three different metronomes at 120 BPM ............ J2
J.3 Latency values for the three different metronomes at 150 BPM ............ J2
J.4 Latency values for the three different metronomes at 180 BPM ............ J3
J.5  Latency values for the three different metronomes at 210 BPM . . . . . . . . . J3
J.6  Matrix for Hypothesis Testing (instruments 1 to 15) . . . . . . . . . . . . . J4
J.7  Matrix for Hypothesis Testing (instruments 16 to 31) . . . . . . . . . . . . . J5
J.8  Pairwise comparisons, post-hoc test main hypothesis, list 1/8 . . . . . . . . J6
J.9  Pairwise comparisons, post-hoc test main hypothesis, list 2/8 . . . . . . . . J7
J.10 Pairwise comparisons, post-hoc test main hypothesis, list 3/8 . . . . . . . . J8
J.12 Pairwise comparisons, post-hoc test main hypothesis, list 5/8 . . . . . . . . J10
J.13 Pairwise comparisons, post-hoc test main hypothesis, list 6/8 . . . . . . . . J11
J.14 Pairwise comparisons, post-hoc test main hypothesis, list 7/8 . . . . . . . . J12
J.15 Pairwise comparisons, post-hoc test main hypothesis, list 8/8 . . . . . . . . J13
List of Figures

2.1 Simplified acoustic instrument model adapted from [221] . . . . . . . . . . . . . 13
2.2 Total latency signal path, adapted from [42] . . . . . . . . . . . . . . . . . . . . 31
2.3 Hypothetical physical and perceptual onsets and perceptual attack time [231] . 47
2.4 Individual latency tolerance [18] . . . . . . . . . . . . . . . . . . . . . . . . . . 75

3.1 Simplified cause-effect diagram related to latency and based on the “model of a solo music performance situation”, adapted from [203]. . . . . . . . . . . . . 84
3.2 Methodology stages of the research design . . . . . . . . . . . . . . . . . . . . . 88
3.3 Variables and their effects in an experiment adapted from [140] . . . . . . . . 109

4.1 Visual metronome . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 126
4.2 Score for the listening test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
4.3 Bass clef score . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
4.4 Alto clef score . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
4.5 Neutral clef score for instruments with no precise pitch . . . . . . . . . . . . . 128
4.6 DAW template . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 132
4.7 DAW transport panel . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 133
4.8 DAW delay plug-in . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 134
4.9 Metronome set-up . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 135
4.10 Metronome headphones levels . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 136
4.11 Normal hearing measurement . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 137
4.12 Equipment set-up for the final listening experiment . . . . . . . . . . . . . . 138
4.13 Recording set-up (a) piano (b) snare drum (c) double bass (d) harp . . . . 140
4.14 Listening experiment set-up . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 146
4.15 Latency ($L_d$) vs. tempo in BPM . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 151
4.16 Latency ($L_d$) vs. instrument groups . . . . . . . . . . . . . . . . . . . . . . . . . . 151
4.17 Latency ($L_d$) vs. musical tempo (aerophones and chordophones) . . . . . 152
5.1 Age and gender distribution .................................................. 162
5.2 Distribution of test subjects according to instrument and sound generation group .................................................. 163
5.3 Additional instrument performed ............................................. 164
5.4 Scatter plot matrix for all numerical variables as indexed in Appendix K.1 .................................................. 166
5.5 Scatter plot matrix of all numerical variables averaging the metronome data indexed in Table 4.7 .................................................. 168
5.6 EDA for a random distribution .................................................. 170
5.7 EDA for the data distribution of the research .................................. 170
5.8 EDA for the square root transformation of the data .......................... 172
5.9 EDA for the logarithm transformation of the data .......................... 172
5.10 Gender distribution for latency values ........................................ 174
5.11 Metronome distribution (aural, visual and both) .......................... 175
5.12 Correlation matrix summary for all variables ................................ 177
5.13 Correlation matrix plot for latency values .................................... 178
5.14 Correlation matrix plot for latency values including linear regression .................................................. 179
5.15 Density plot for all instruments according to the different tempi (average of all 3 metronomes) .................................................. 181
5.16 Latency (Ld) vs. tempo for all instruments .................................... 183
5.17 Latency (Ld) vs. tempo for every musical instrument ......................... 185
5.18 Box plots of the latency values (Ld) for each musical instrument ............... 186
5.19 Latency (Ld) vs. tempo for the four different instrument groups ............... 187
5.20 Latency (Ld) vs. instrument group (average of all 3 metronomes) ............... 188
5.21 LTR according to musical instrument group .................................. 189
5.22 LTR with respect to musical instruments group and tempi ..................... 190
5.23 Density plot for the aerophones at different tempi (average of all 3 metronomes) .................................................. 192
5.24 Density plot for the chordophones at different tempi (average of all 3 metronomes) .................................................. 193
5.25 Density plot for the membranophones at different tempi (average of all 3 metronomes) .................................................. 195
5.26 Density plot for the idiophones at the different tempi (average of all 3 metronomes) .................................................. 196
5.27 Latency (Ld) vs. tempo for the different sound generation techniques ........... 198
5.28 Latency (Ld) vs. sound generation technique (average of all 3 metronomes) .................................................. 199
5.29 LTR according to the sound generation technique ................................ 200
List of Equations

2.1 Blocking delay ................................................. 33
2.2 Total delay experienced by the user of an NMP System .............. 37
2.3 Minimal self-delay ............................................. 52
2.4 Actual offset .................................................. 53
2.5 Tempo model ................................................. 58
2.6 Human player behaviour ....................................... 58
3.1 Total latency ................................................. 92
3.2 Latency Tolerance Range .................................... 94
Chapter 1

Introduction

1.1 Background

Western music has always been in a continual state of change. The development of music towards the musical ensemble known as western orchestra started with the culmination of modal vocal polyphony in the 16th century. Music printing, around 1501, constituted a technological development enabling wider music diffusion and setting the path for new creative possibilities [35]. The refinement of compositional techniques of vocal polyphony, such as the madrigal, the French chanson and the German Lied influenced instrumental ensembles from the Renaissance onwards [35]. Renaissance consorts were the prototypes for later chamber ensembles.

The Renaissance brought new musical instruments, new genres and new styles. Instrumental ensembles of a handful of musicians with a homogeneous sound, as well as mixed ensembles were common [35]. The majority of the musical instruments, such as wind, percussion and strings, were already being used in the Middle Ages. In the Renaissance period, instrumental music grew in significance. This importance is reflected in a vast number of instrumental works [35].

The western orchestra, as it is known today, began an evolution process from around 1600. At this time the bass line was performed by supporting instruments in a mainly improvised way. By the end of the 18th century, large ensembles were grouped into wind and string instruments [211]. The first large ensemble of the violin family and precursor of the modern orchestra emerged in France around 1670 [211]. The modern orchestra had its
roots in the court orchestra of Louis XIV at Versailles, under the direction of Jean-Baptiste Lully [35]. By the end of the 17th century, the name orchestra denoted a specific group of instruments with no more than 25 musicians. In the mid-18th century, the strings performed the main musical material, whereas wind instruments doubled, reinforced and filled the harmonies [35].

From the beginning of western orchestra, the environmental and technical possibilities, particularly those with regard to the physical possibilities of the instrument, have shaped the way music has been performed. Musical styles and the acoustics of the rooms where music is performed have continuously interacted over the last centuries. In the 17th and 18th centuries, audiences observed an increasing relationship between acoustics, compositional style and performance practice. However, acoustic determinism never existed [211].

As public concerts developed in the late 18th century, music was played in new spaces. The performance of music was no longer only an event for the nobility in the courts or for specific religious purposes in the church. Larger indoor spaces with a greater distance from the audience led to the growth of small ensembles and there were also further modifications in the construction of instruments [211].

Large orchestras doubled or tripled the winds and added brass. Violins were divided into first and second and left and right, respectively. The number of musicians in an orchestra increased from about 40 at the beginning of the 19th century up to 90 by the end of that century [35]. The placement and seating in orchestras were influenced not only by social considerations, as in the 16th and 17th centuries, but also primarily by logistical and practical aspects relating to the physics of sound [211].

The role of the conductor as the leader of the orchestra was shaped mostly in the 19th century. The orchestra is a “body of sound”, and not only was it necessary to keep all musicians together by indicating the tempo, but the music had to be interpreted in a special way. Therefore, the conductor was responsible for coordinating the performers as well as shaping and, phrasing the musical character [35].
Today, network technologies, including the Internet allow the expansion of musical artistic expression and offer new ways for experimentation in music performances. A network is defined by the Audio Engineering Society (AES) as:

“...a collection of more than two devices, dispersed in space, and connected together via hardware and software, such that any device can communicate with any other connected device.” [28, p. 730]

It is important not to overestimate the role of new technologies in the origin of new musical forms. While network technologies offer new performance conditions that encourage modern musical practices, they do not determine them [211].

Networks such as the Internet allow “collaborative performances”, performances where musicians are geographically and physically separated [18]. In addition, many new multi-channel digital media networking technologies are emerging. The flexibility of these media technologies using networks is greater and the administration is easier for audio requirements at live events and in the studio in comparison to some of the current analogue solutions [15, 4].

The history of networked media [68] began in the 1960’s. In 1966, Max Neuhaus experimented with music and telephone networks creating an artistic work under the name “Pieces for Public Telephone Network” [28]. For this occasion, mixed calls arrived on ten telephones at the WBAI radio station in New York and the resulting mix of sounds, noises and calls were broadcast [173]. Today, every imaginable music performance set-up over networks, from jazz to classical music, is feasible. The new architectonic spaces are virtual and include PCs, networks, and the Internet, to name a few. Music is performed in a whole new way thanks to these new mediums.

The first decade of the 21st century, has seen some relevant network music performances. The “Pacific Rim of Wire” at Stanford on April 29, 2008, was a successful example combining network technology and music. A concert performance of Terry Riley’s “In C” took place between Stanford and Peking across a distance of 6,000 miles. This concert also demonstrated the possibilities of experimenting with non-traditional instrumental combinations [39]. A remarkable project is the Distributed Immersive Performance (DIP), one of the first experiments evaluating synchronous collaboration of remotely located musicians which was conducted in 2002 between two buildings at the Viterbi School of Engineering [63]. The LOLA project (“LOw LAtency”) conceived in 2005 and developed between 2008
Introduction

and 2010 is one of the first projects using not only audio, but also video, across large distances of thousands of kilometres [84]. Some results of experiments on the effects of musical interactions over the Internet for realistic performance settings will be further discussed in this work.

Many questions arise with these new possibilities and research has been carried out since the beginning of this century, primarily in order to understand how these collaborative performances could work. Nevertheless, quantitative studies based on additional parameters, for e.g. the characteristics of the instruments, were not very common until the first decade of the 21st century [199]. This research problem is still unexplored and should also be approached from an unorthodox point of view with respect to musical practice [198].

This is essentially a quantitative research in music technology. Different approaches and topics, such as musicology, psychology and several technical issues must be addressed. The limits and approaches of each subject are clearly defined. This work also addresses how the topics complement each other. The main scope of this work is limited to western musical instruments present in the classical western orchestra. Nevertheless, popular instruments (e.g. the guitar) which are not part of the standard orchestra configuration are also included.

1.2 Description of the Problem

A network like the Internet creates a virtual space where many musicians from all around the world can perform together at the same time. The network defines a space with specific characteristics and, as such, is also an acoustical environment. A network may even be used as a musical instrument [28].

An intrinsic issue when using networks is latency. Latency may be even used as an artistic tool. In the 16th century, the use of long reverberation times and large delays were exploited in the Basilica di San Marco in Venice [21] by the cori spezzati, the Venetian polychoral music style of separated choirs [18]. The director was in the middle of the church to synchronize the performance, and the singers were distributed around the cathedral balconies. Due to the size of the church, singers had to cope with delays of more than 200ms (60m distance) [145].
Current research has focused on defining timing conditions in which a collaborative performance is feasible [205, 55, 46, 63, 113, 54]. The primary question is how much latency\(^1\) a test subject can cope with. In almost all current research, delays have been electronically introduced. Performers listen to their own and other musician’s signals (i.e. music) through headphones. Test subjects perform or play until the performance breaks down because they cannot cope with any more latency, (i.e. the musicians are not able to play together). In the last decade there has been increasingly more literature and research dealing with composition techniques, aesthetics or software. Researchers are also interested in the practicability of performances with musicians located remotely in different physical spaces [104].

During the 20th century, the role of recording technologies was decisive and important for the development of both music performance and music perception. In his book, Katz [134] describes how the use of vibrato in violins was considered “tacky” and “kitschy” before the advent of music recording. Not all new approaches are so evident, but are rather highly transcendental. Because of new technologies, musical tempo has become very precise and performers are more aware of their internal metronome [36].

Audio networks are the new technological standard used in music production, the concert venue and for the exploration of new musical forms. Previous research on latency [158, 46] has shown that some factors such as musical style and tempo influence a musical performance. Latency makes musical performances annoying or even impossible. However, the role of the musical instrument regarding the latency issue is not yet clarified. A systematic approach is necessary. The problem addressed in this work is the development and testing of a method to estimate to what extent the ability to cope with latency when playing music is related to the musical instrument being performed. Furthermore, it is necessary to develop a measure, in order to establish similarities or differences among musical instruments. As current knowledge is limited to specific ensemble configurations a new approach is necessary to help improving the work of musicians, music producers and engineers. An approach that is focusing on the differences between musical instruments.

The performance of music is supported by new structures and roles [18]. Network music is a new phenomenon whose concept changes constantly depending on the parameters involved, the medium and the art of performance. A consistent definition is found in the

\(^1\) Delay and latency are concepts used unambiguously through this work.
work of Barbosa. He states that “Music is Networked” [18, p. 14]. The orchestra might be a client-server where the clients are the musicians all interacting together. There may or may not be a conductor. This concept has been applied in connection with collaborative performances or ensemble performances (two or more performers). However, a musical network also allows what can be defined as “non-collaborative performances”. In contrast with collaborative performances, non-collaborative performances are those in which just one musician performs, a solo performance.

At this point it is necessary to introduce a definition of latency to describe the meaning of this concept in this work:

"Latency is a measure of the time it takes for sound (or image) to be captured, transmitted, and reproduced at a remote destination." [209, p. 16]

As this is a general latency concept, it is also important to clarify which specific components belong to this concept. The "remote destination" is the remote listener or music performer. When musicians perform, latency is the delay due to the time a sound wave takes to travel from one place to another. If there is any digital equipment involved in the performance (i.e. sound recording or sound monitoring equipment), the processing time for digital conversion has to be taken into account [135]. In this work, an additional new element is also part of the equation: the delay produced due to transmission on a network.

Latency is also an issue when using network technologies at live sound venues. Monitoring signals for the musicians, (mainly in-ear monitoring), can be annoying even when the latency figure is under 3ms [207]. In summary, the scope of research for this work is the development of a method in order to evaluate the disruptive effect of latency in “non-collaborative performances” and its relationship with musical instruments.

1.3 Research Motivation

Research on latency for non-collaborative performances has been carried out since the 1950s, mainly in language research. In other words, feedback from oneself when speaking [217]. Questions about the influence of the musical instruments used in the research have been outlined but not fully researched in a variety of musical instruments.
Latency is, and continues to be, an issue for many artistic and educational projects where interaction between practice rooms, classrooms or even stages are required [24, 28]. In 2002, teaching applications were tested with the well-known violin player, Pinchas Zueckerman, using the ultra-video conferencing system with multi-channel audio and uncompressed video [162].

Physical conditions such as the transmission velocity over the Internet still cause difficulties with simultaneous performances of music over a network between musicians. Nevertheless, new audio technologies for recording studios, live sound and broadcast [4, 15] rely increasingly on networks. Nowadays, live venues are very complex and demanding tasks, whereas new technical solutions offer flexibility and are also financially attractive. These new technical approaches have a latency side effect [28], and even minimal amounts of latency could be a critical problem for the development of live music events [28]. Historically, the delay has always been an issue when performing music. Musicians such as singers, orchestra musicians and church organists have always performed under auditory delay [62]. To overcome this problem, acoustical solutions like balcony placement of orchestras in churches and a better arrangement of the musicians were pragmatic answers to solving the problem [211]. Being nearer to the ceiling and walls, the sound was immediately reflected and the delay between direct and reflected sound was reduced to a minimum [211].

It is relevant to understand how latency interacts with instruments and performers [75], and this understanding is necessary for finding possible solutions and for estimating the impact of latency for different instruments set-ups. Moreover, the understanding of perceptual issues regarding networked performances requires more experiments that should be developed systematically [198].

Nowadays network solutions are the state of the art in audio engineering, not only in the recording studio but also in the concert venue. Flexibility, cost and configuration are the arguments for the use of this technology trend. The drawback of latency is very well known, but how it affects the work in the studio with musicians is still not completely answered. A further understanding of the relationship between musical instruments and the latency issue is highly relevant for musicians, music producers, recording engineers and developers of network technologies in order to improve results directly related to the musical performance.
Research on latency is not only interesting when it comes to music and recording technology. In the fields of virtual reality design, telepresence, virtual acoustics, haptics, immersion [138, 128] and even “robotic musicianship” [198], research on latency is an important development.

A review on research in networked music performances (NMP) published in 2016 [198] recognizes the inadequate amount of research done on the topic of individual latency tolerance applied to the auditory feedback from the musician’s own instrument. However, it is clear how subjective the task may be with respect to the related factors such as instrument type and musical dexterity. Research on latency conditions in which a performance is acceptable is an issue that should be further studied [74].

1.4 Research Question

This study presents the development of a new research methodology to obtain quantitative results regarding the musician’s ability to cope with latency in non-collaborative performances. Accordingly, the research question is as follows:

Can the relationship between musical instruments and the ability to cope with latency be determined?

This research question can be broken down into the following sub-questions.

1. How can a measurement for different groups of musical instruments (idiophones, chordophones, membranophones and aerophones) be defined?

2. How can an appropriate and objective measurement method be designed and conducted?

3. How can the ability to cope with latency depending on the musical instrument be determined?

The aim of this research is the development of a method for the analysis of musicians’ ability to cope with latency in connection with western musical instruments. The test must be repeatable, accurate and involve musicians or music students as test subjects.
Introduction

Characteristics of the study

It is important to emphasise that this is not a research into the musician’s timing or talent. Furthermore, some assumptions prior to the development of the testing and the analysis of the results have been taken into account. It is assumed that music students and professional musicians behave similarly regarding the performance of music and the ability to read a music score, in order to allow for comparable results.

It is also not within the scope of the study to research the effects of latency or the strategies developed by musicians to cope with it, although literature on these topics is part of the review on literature and include the works of Chafe et al., Carôt, Schuett and Olmos et al. [49, 46, 205, 176].

This work presents a systematical and replicable method to evaluate the ability of musicians to cope with latency. The empirical quantitative results generated are statistically evaluated. For the evaluation, most of the representative western musical instruments of every instrument group were tested for this work. Based on the quantitative results, a hypothesis could be tested:

1. Null hypothesis $H_0$: The role of the musical instrument is not relevant with respect to the ability to cope with latency.

2. Alternative hypothesis $H_1$: The ability to cope with latency when performing music is directly related to the musical instrument played.

In the case where the null hypothesis is rejected, the alternative hypothesis is supported. Different graphics displaying latency vs. instrument group for every group might be expected. On the basis of previous research [18, 141, 152], it has been shown that performers of different instruments can cope with latency depending on the instrument played.

This work looks even further forward in order to find empirical evidence on the connection between the ability to cope with latency and the different instruments played. Moreover, this study may find a possible relationship between instruments groups (idiophones, chordophones, membranophones and aerophones) and latency. The role of the musician with respect to the ability to cope with latency is obvious. It is clear that every musician has
their own abilities and dexterities which are evidenced by the performance of their musical instrument.

One of the research goals is to determine a measure of the relationship between musical instruments and the ability to cope with latency. This measure could play a role in the further development and design of virtual and digital instruments, sound applications and professional audio products involving musical instruments. This work aims to provide a method based on a listening test. Of additional importance is the expansion of knowledge in the field of human perception and music performance.

There are several factors involved in the perception of latency, some of which may be measured and some not, even using current technology. The current work provides a further step in understanding the direct relationship between music performance, musical instruments and the musicians’ ability to cope with latency.

1.5 Thesis Structure

After this introductory section, the following chapters systematically illustrate the research process. Chapter 2 includes the literature review and established methods used previously to gather quantitative and qualitative information. Relevant topics are musical instruments and network taxonomy. A theoretical framework, in order to develop a general understanding of the musical instrument groups, is presented. In addition, information about the physical characteristics of musical instruments and body parts involved in the performance process is summarized. Furthermore, the role of networks in music and its development are described. The theoretical background and current research on latency are outlined and the main definitions and results regarding human perception and latency are introduced. The relationships between delayed auditory feedback (DAF), perceptual attack time (PAT), and latency adaptive tempo (LAT) are discussed.

In Chapter 3 the author develops the experimental methodology and designs a listening test based on the groundwork and its outcomes. The experimental set-up of the methodology and the listening experiment are presented in this chapter. A detailed description of the measurement procedures that support or reject the hypothesis and the different questions and steps involved in the listening test are interpreted and explained.
Chapter 4 is an overview of the first results obtained using the pilot test. Furthermore, the concept of latency tolerance range (LTR) is introduced. In this chapter, the development and difficulties related to the pilot listening test are outlined, discussed and evaluated. Further improvements for the main test as well as its development are reviewed and developed. The final experiment and data results are presented at the end of this chapter.

Chapter 5 presents a statistical analysis of the results. The first approach is exploratory and descriptive. In the second part, inferential statistics are used to present the results. The analysis of trends is based on graphical analysis and the information gathered.

Chapter 6 concludes the study. A summary of the conclusions, limitations and suggestions for future work are discussed.
Chapter 2

Literature Review

This chapter summarizes the most relevant research results of the last decades with respect to the issue of latency in music performance from a technological and music psychological approach. Although some of the findings and research studies are contradictory, it is considered necessary to present the different approaches here for a better understanding of the phenomenon.

2.1 Western Orchestral Musical Instruments

Musical instruments are systems that can approached through a model. When playing a musical instrument, different parts of the body as well as external tools are involved in producing sound.

2.1.1 A Model of Western Musical Instruments

Figure 2.1 is a diagram based on the work of Trueman [221] and his acoustic violin model. The simplified acoustic instrument model is an attempt to display and summarize a very complicated relationship between performer, instrument and sound production. This model may help in understanding the ability to cope with latency based on the instrument used.
The performer receives information not only from the direct sound and the room (acoustic feedback) but also from the haptic and visual feedback (physical interaction). This information affects the ability to cope with latency. In addition, the instrument body and its physical structure influence the performance [18].

### 2.1.2 Musical Instruments Groups

The musical instruments in the western orchestra are normally divided into four families, i.e. strings, woodwind, brass and percussion [41], with respect to both the sound quality of the instruments and also with the physical positioning within the orchestra. Acoustic principles are not taken into account in this classification. A widely accepted scientific classification standard is the one proposed by Curt Sachs and Erich M. von Hornbostel [223], revised by Sachs some years later in 1940 and re-edited under the name: “The history of musical instruments” [200]. The taxonomical system proposed by Sachs and von Hornbostel divides instruments into five main groups, idiophones, membranophones, chordophones, aerophones and electrophones.

In the current work, four of the five groups are relevant: idiophones, membranophones, chordophones and aerophones. The music instruments belonging to these groups are found in any conventional western orchestral set-up. Table 2.1 groups the different instruments
Literature Review

of the western orchestra and describes the excitation method to produce sound in the instrument family. These family grouping are the most commonly used taxonomic descriptions for the musical instruments.

For this study, the traditional classification ontology based on the material characteristics of the musical instruments was chosen. In order to enable an empirical approach, the classical definition of a musical instrument as a purpose-built technology [219] was adopted.

<table>
<thead>
<tr>
<th>Group</th>
<th>Excitation method</th>
<th>Instrument family</th>
<th>Musical instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiophones</td>
<td>Striking</td>
<td>Percussion</td>
<td>Marimba, Xylophone, Vibraphone, Gong, Triangle, Tubular bells, Glockenspiel, Tambourine (indirectly struck), Wood block, Cymbals, Tam-tam, Celesta</td>
</tr>
<tr>
<td>Membranophones</td>
<td>Striking</td>
<td>Percussion</td>
<td>Timpani, Snare drum, Tenor drum, Bass drum, Tambourine (including struck)</td>
</tr>
<tr>
<td>Chordophones</td>
<td>Striking</td>
<td>Keyboards</td>
<td>Piano, Harp, Guitar</td>
</tr>
<tr>
<td></td>
<td>Plucking</td>
<td>Strings</td>
<td>Violins, Violas, Cellos, Double basses</td>
</tr>
<tr>
<td></td>
<td>Bowing</td>
<td>Strings</td>
<td></td>
</tr>
<tr>
<td>Aerophones</td>
<td>Mechanical reed</td>
<td>Woodwinds</td>
<td>Clarinet, Saxophone, Oboe, Bassoons, Flutes</td>
</tr>
<tr>
<td></td>
<td>Air reed</td>
<td>Woodwinds</td>
<td>French Horn in F, Trumpet in Bb, Trombones, Tubas</td>
</tr>
<tr>
<td></td>
<td>Lip reed</td>
<td>Brass</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Taxonomy of the musical instruments of the western modern orchestra

The Sachs-Hornbostel classification is based on sound production mechanisms, physical characteristics [219] and the global view of instruments as intermediaries between work performed and performer [6]. However, cultural origins and social interactions are not taken into account for this classification [137]. Nowadays, the classification of musical instruments aims to include not only the physical aspects of sound production but also to introduce a more global concept [6]. Instruments can be seen as extensions of the musician’s physical body, i.e. “embodied entities” [6]. This new approach has been accepted since the adoption of electrical and virtual instruments [137, 180].

The classical approach makes it possible to consider almost all musical instruments as a system with two vibrating devices: a generator and an amplifier coupled to the generator.
More generally, the acoustics of musical instruments may be described by means of a sound source and sound modifiers. The sound source is the input and the sound modifiers are the “system”, mainly the instrument body and the resonator. The output of the instrument (not necessarily what the listener hears) is the actual output of the instrument. In addition, acoustic effects of the environment (room acoustics) should be taken into account [121].

The performer initiates the production of the sound by blowing, bowing, striking or by any other excitation method. The amount of energy of this initial vibration is not enough for the creation of sound itself. The resonator which, in most cases, is the musical instrument body or the air column inside it and is coupled to the basic device, amplifies the different harmonics generated to a greater or lesser degree [218], enabling an adequate sound pressure level (SPL) and therefore sound production.

2.1.3 Idiophones

Idiophones, also known as self-sounders, are made of naturally sonorous material [200]. To set idiophones in vibration the “striking” method is normally used. Idiophones which require striking include xylophones, marimbas, gongs, triangles and celestas. These instruments consist of one or more pieces made of a sonorous material, e.g. wood, bamboo, stone, glass or metal. The instrument is struck with a stick or a similar device with a rotary motion of the arm [200]. For striking keyboard percussion instruments such as the xylophone or marimba, mallets are used, while for all other instruments, for e.g. the gong, beaters are chosen [3]. Other idiophones that are struck, e.g. the concussion instruments, consist of a pair of resonance elements which are struck together using either two hands or one hand and include castanets, cymbals or clappers [41]. Current materials are metal, wood, bone or ivory [200]. Idiophones use either wood bars, for e.g. in the glockenspiel, celesta, triangle or xylophone, or plates, for e.g. in the cymbals [121].

Because of their physical nature, idiophones are divided into instruments with either a definite pitch or an indefinite pitch. The following section describes those subgroups.
Idiophones with a definite pitch

Tuned idiophones, such as the xylophone, marimba, glockenspiel, chimes, celesta, gong and vibraphone produce tones with a definite pitch [98]. These instruments can also be excited with other elements such as as a bow, generating sustained tones in the case of the gong [98].

The orchestra bells or glockenspiel is normally played with wood or hard plastic mallets. Usually the instrument is performed with two mallets [3]. The tubular resonators of the instrument create a loudness increment by producing shorter decay times of the sound.

As with the marimba, the xylophone bars are made of wood or synthetic material and also have tubular resonators that achieve the same function. Mallets, depending on their construction material, e.g. metal or wood, and also the material covering the striking end [121], influence the timbre of the instrument played (e.g. the marimba or xylophone). The majority of marimbas and modern xylophones that are part of the western orchestra set-up have tubular resonators to increase tone loudness. The bars are mainly of wood or synthetic material [98]. The bars of the xylophone and the marimba have different lengths and are arranged in the same way as in the piano. For playing the xylophone, two mallets are normally used (one per hand). Marimba players use two mallets per hand sometimes even three [3].

The celesta is probably the most used keyboard instrument in the orchestra. Although it has a keyboard, the sound produced is very similar to that of a glockenspiel. The sound is produced by hammers striking steel bars. These bars lie across a wooden resonator box. Pitches are sustained and it is not possible to play staccato [3].

Idiophones with an indefinite pitch

Cymbals are instruments without a resonator [218]. In jazz and popular music, the instrument uses different names and shapes like swish, splash, ping, pang, ride and crash. In the western orchestra, cymbals can also be found under different names depending on the way they are played and their forms (e.g. crash cymbals, suspended cymbals, hi-hat cymbals, sizzle cymbals, China-type cymbals and finger cymbals) [3]. The most commonly used material is bronze. Cymbals are excited by striking at various points on the instrument
with either a soft beater, a wooden stick or even a cymbal crash (i.e. two cymbals struck together) [121]. The physical onset of the instrument’s sound depends on the form of excitation used.

Triangles are one of the oldest percussion instruments of the orchestra [3]. A steel rod bent into a triangle shape with a gap separating the two ends gives the name to the instrument [41]. The sound of the triangle is both dependent on the strike point and the energy delivered.

2.1.4 Membranophones

Membranophones are normally referred to as drums. Their sound is produced by a vibrating membrane stretched over an opening. For this research, the main focus lies in the membranophones excited for sound production by “striking”, e.g. the snare, bass drum or tom-tom [41].

Percussion instruments are excited and receive energy in short bursts. A common method is to strike the membrane of the drum with a single stroke, thereby enabling the vibration of the significant parts of the instrument responsible for sound production at their natural frequencies. Due to the strike, many frequencies with no relationship to each other are generated, producing an “unpitched” sound. The vibrations normally last until the instrument is struck again [218]. Strokes for drums are produced primarily with sticks [3].

After the human voice, drums are the oldest musical instruments [98]. Modern drums can be divided into two groups: those producing a sense of pitch and those with no identifiable pitch. The first group includes the kettledrums. Instruments such as the snare drum, tenor drum and bass drum belong to the second group [98].

Kettledrums have had a place in the modern orchestra since the invention of screw tension devices in the 17th century. Normally this instrument is played by a single person using two to five kettledrums known as timpani [121]. The instrument is constructed in various sizes enabling the production of different musical intervals, normally a separation range of approximately an interval of a fifth.
The bass drum radiates the greatest degree of power of all the orchestra instruments [98]. The snare drum, which is found in the western orchestra, consists of two heads, the top (playing head) and the bottom head (the batter). By striking the instrument the snare head vibrates against the snares. Tom-toms are available as a one-headed or two-headed tom. Normally tom-toms have no identifiable pitch, especially the one-headed type [98]. However, the instrument may be tuned to obtain certain pitches [3].

Sound decay time depends on many factors including the tension of the membrane, the kettle weight and material, the type of drum-head and, in particular the way the drum is supported [98]. Striking the membrane of the instrument produces a vibration in all relevant parts of the instrument. These drums are mounted on stands, one tom-tom pair per stand, and are ordered from high to low [218].

2.1.5 Chordophones

Instruments where sound generation is produced by a vibrating string belong to the chordophone group. For amplification, the strings are normally attached [41] to some sort of resonator or soundboard which, in most instruments, is made of either wood or skin [200]. Three different acoustical excitation methods for sound production are relevant. The sound is produced by “striking”, where a hammer strikes the string which normally remains in vibration after immediately rebounding off the hammer, e.g. the piano. Damping of the string can be achieved by contact with felt or other materials [218]. “Bowing” is where the energy is supplied from a moving bow. For an enduring note at the same pitch the vibration should be maintained, e.g. the viola, violin, cello and double bass [218], and “plucking” causes the string to vibrate by means of a plectrum or other device that plucks the string, e.g. the guitar or harp.

Bowed string instruments

The violin is in contact with the musician’s body. When played, the left shoulder and the left arm and hand hold the instrument which is also supported by the left side of the chin [3]. The viola is held in the same way as the violin, but the left-hand tension for playing is greater than for the violin [3]. The cello, also called the violoncello, is more or less double
the size of the violin [98]. Due to its size, the cello is held between the knees and the player is seated. An adjustable peg which reaches the floor provides additional support [3]. A double, bass or contrabass, retains most of the characteristics of the violin family. It has four strings like the other bowed instruments of the orchestra, but sometimes five strings are also used [98]. The instrument is also supported by a peg. The player can stand or sit on a special stool. The instrument is held with the body and the left knee by the musician [3]. The influence of the bow in the production of sound in violins, violas, cellos and double basses is recognized but is not well researched. Some of the static properties of this element are the shape, mass, camber and stiffness [98]. Bow hand positions are very important for sound production [3].

Bowed instruments have a slower attack when compared to wind instruments. The attack influences the sound of the instrument and delivers important information for timbre perception [153]. The perception of the sound may be influenced by this fact, even when wind and bowed instruments have the same physical onset [111]. Bowing pressure and acceleration primarily influence the string motion for a fundamental frequency and the distance between bow to bridge during the attack phase. Bow pressure and speed influence the string motion after the attack when full amplitude has been achieved [13]. With high pitched instruments, bowing variation in order to obtain noise-free attacks with a specific string is not as restrictive as for low pitched instruments [13].

With respect to bowing, it has been observed that for some players it is better to start the stroke from the air, which produces an “airy” or “slippy” onset of sound. For a controlled and consistent bow-arm movement, an upper body posture allowing flexibility is necessary. Double bass players should avoid having the instrument’s body locked to the chest [111]. In order not to produce “raucous” or “creaky” sounds, the string player should avoid applying a very high bow pressure or should play at a very low speed. Bowing position changes towards the bridge may have the effect of producing increased brilliance and therefore the perception of louder sounds, even when the performer compensates with a lower bow speed [111]. Releasing the bow pressure while starting a new stroke in the opposite direction reduces brilliance which is easier to perceive in the double bass than the violin. The timbre of the violin is mainly influenced by changes in bowing pressure and bowing point. Changes in bowing velocity produce variations in loudness [17].
In psychological research it is assumed that string players constitute a homogeneous group. However, there are likely to be differences between musicians. In contrast to piano players who always know which key they should press, string players have to rely on other cues. It is necessary for string players to create aural schemata or “mental representations” [111], in order to assess a musical performance.

**Plucked-string instruments**

The modern harp is a chromatic instrument developed in the late 18th and early 19th centuries [3]. A combination of pedals enables the chromatic set-up [98, 41].

The guitar is mainly found in so-called “popular music” (e.g. jazz, folk or rock) and is also part of the normal orchestra repertoire [3]. Historically, this instrument evolved in different ways in France, Spain, Italy and Britain [222]. The modern guitar, as we know it, is the product of a constant development initiated around the year 1785. From its early days, some composers expected the guitar to achieve a place in the orchestra and be accepted as a “serious instrument”. This was initially achieved mainly for works involving the guitar as a solo instrument or when performing with other chordophones, such as violins, bass and the piano [222].

Classical guitars are played with nylon strings. Steel strings are used in flat top guitars. Both types of strings can be played with the fingers or with a plectrum. The main difference between classical and folk guitars are the dimensions. Folk guitars, also known as country or western guitars, have the same shape. They are larger but flatter at the top and their body is heavier to support the higher stress caused by the steel strings. The transient plucked sound is very different from the bowed sound of other string instruments: it rises and falls very quickly. The use of fingers or a plectrum changes the harmonics that are accentuated and therefore the sound produced [41]. As in most musical sounds, the manipulation and elimination of the initial transient form produces an uncertainty in timbre recognition. For plucked instruments, this information lies in the first 50ms. In contrast to bowed instruments, the plucking hand has no influence on the sound after tone initiation [17].

For this research, with regard to the way the instrument is held, it is assumed that musicians adopt the universally accepted sitting position favoured for serious playing [222],
with or without resting the right foot on a small object.

**Struck string instruments - the piano**

The piano belongs to the chordophone group. However, it can also be used as a percussion instrument [3]. The percussive characteristics of the piano sound are intrinsic to the onset of sound by this instrument. Two main instrument configurations are available: the concert and the upright piano [98]. Fingers and arms play an important role in playing the piano. Optimal finger positioning depends on many factors, e.g. anatomic, motor, physical and cognitive faculties, in addition to interpretation, making the playing of the piano, as for every other instrument, a very subjective task [178].

In most cases, finger-key-noise is masked by the tonal part of the sound. Other relevant factors are dynamics, room acoustics and the listener’s position. Due to sound characteristics in the higher register, the finger-key-noise is not as well masked [178]. The pianist can influence the speed with which the string is hit by the hammer. However, contrary to popular belief, the timbre of the instrument cannot be influenced by the way the musicians depress piano keys, only by the velocity [178]. Pianist movements affect control, which may indirectly affect the timbre. The use of the pedal may also influence timbre perception.

From research on piano sound production dynamics by Askenfelt et al. [14], the time measured for key depressing from rest (surface) to bottom (key bed) was 160ms for *piano*, 80ms for *mezzo-forte* and 25ms for *forte*.

### 2.1.6 Aerophones

In the aerophones, the sound is produced by a vibrating column of air originating in the mouthpiece of the instrument [218] and enclosed by a tube, and a device which could be either the compressed lips of the player, as in the trumpet, or the to-and-fro movement of a reed, as in the clarinet or the oboe [41]. Both tube and device set the air into vibration by changing the constant breath of the player into pulsations [200]. In the western orchestra set-up, the normal methods of excitation include a mechanical reed which controls the air flow from the musician’s lung to the instrument [121], e.g. the
woodwind clarinet or saxophone, an air reed, e.g. the woodwind transverse flute, and a lip reed where the vibrating motion is created by the musician’s lips on the mouthpiece and is independent of the instrument [218], e.g. the brass, trombone, trumpet or horn [41].

For the production of sound the energy comes from the performer’s respiratory system. The reed is the element which controls the air stream. There are different kinds of reed [103]:

- Cane piece (e.g. bamboo)
- The lips
- A metallic tongue
- Air jet (e.g. flutes)

Controlling loudness, attack, intonation and timbre is possible due to the embouchure setting, airflow, blowing pressure and the length of the air column. The embouchure is the instrument coupling with the mouth and it takes into account the forces, skills and muscle regions (lips, mouth, jaw, tongue, face and control) involved in sound production in order to generate the characteristic sound of reed and brass wind instruments [6, 103, 180].

The movements of the respiratory system also allow different pressure ranges and coordinated oscillations. These oscillations play a significant role in vibrato production. The air from the lung may also play a role in intonation [103].

Sound and musical expression are directly related to the air which is delivered by the performer by means of the embouchure. In addition, fingers, hands, arms, control levers, valves and holes influence the production of the sound in the different instruments. Respiratory techniques are very important when it comes to playing wind instruments. For an effective wind instrument performance, sensory functions in the abdominal, thoracic and lunge regions are fundamental [103].

Fine control of the fingers, hands and arms, in addition to body posture, tongue articulation and even the capacity to deal with stressful situations are all necessary for good performance with the aerophones. Tone quality depends on the embouchure control and blowing control [103]. No single correct way for producing expiratory movements in order
to play wind instruments has been found. However, a constant, systematic and automatic expiratory system is recommended while the player focuses on other aspects like embouchure control and blowing pressure [103].

In wind instruments, the system is considered a self-excited system. Air produced by the performer and introduced to the instrument with the reed, i.e. air, lips or other material, goes through the length of the air column and returns to the reed. When the reflected wave pressure has enough energy, the reed will reverse its direction [23]. By this means, a new cycle can be initiated (i.e. self-sustained oscillations) [103].

As stated before, the three common sound generation methods are:

1. Mechanical reed
2. Air reed
3. Lip reed

**Reed aerophones (woodwinds)**

Reed woodwinds are characterized by a reed fixed at one end and free at the other end to enable vibration. Reed woodwinds are either single reed or double reed instruments [41]. As the strength of the mechanical reed increases, it is necessary for the performer to increase the blowing pressure and the lip tension [103]. The reed in every woodwind instrument is held between the player’s lips. The performer can adjust the resonance frequency and the opening of the reed, thus having more control which is essential for producing notes [98]. The air flow for both reed aerophones and brass instruments is also regulated through the action of the lips, the tongue and the moving structures of the respiratory system. All of these work as additional valves [103].

The fingering system is the same for all clarinetists [3] and its dynamic range is very wide [165]. Due to developments in harmony in the clarinet, a performer is able to produce a great psychophysical impression of change in loudness [98] with a dynamic range of approximately 50dB [165]. The oboe family is not as extensive as the clarinet family. In contrast to the clarinet, the oboe has a narrower dynamic range of 15 to 20dB [165] it has a lower loudness when compared to the clarinet [41]. When playing the oboe, rest periods
are necessary due to breath and embouchure control [3].

The bassoon has a dynamic range of about 20dB. Soft and loud playing are not significantly different and the pitch increment and dynamic level translates into a higher blowing pressure [98].

The saxophone is a relatively new instrument. Its tone fingering and playing techniques are similar to the clarinet [3].

**Nonreed (air reed) aerophones (woodwinds)**

In the flute, the airflow is produced and injected through the player’s lips into a mouth hole. An air column generates airflow fluctuations inside the tube [103]. When playing the instrument, breath volume is more important in comparison with other aerophones [3]. For loudness control, the main parameter is volume flow and not blowing pressure. Different players may produce different intonations, even when playing the same instrument [98].

The orchestral flute is the instrument known as the transverse flute due to it being blown from the side. It is also called a Boehm flute. In contrast to other reed instruments, the flute is controlled by airflow. In the flute, the air flows in an alternating form into the instrument and out to the environment [23]. The sound waves produced in the reed go through the tube and mouth cavity. A large amount of energy absorbed in the respiratory system is therefore lost [23].

The embouchure in playing the flute is crucial. The performer has a great deal of control over note production in terms of pitch and tone colour [97]. Nowadays, a more relaxed embouchure is preferred by most musicians and muscular tension is kept to a minimum [103]. However, air production by the performer remains one of the main characteristics with respect to sound production. The musicians can influence parameters like blowing pressure, jet length and lip opening when performing the instrument [98]. For the playing characteristics of the instrument, the coupling between mouth and head joint (the mouthpiece of the flute), is maybe as important as the instrument’s body, together with the embouchure [103].


**Lip reed instruments brass**

The column of air of the instruments of this group is excited by vibrations of the performer’s lips [41]. The timbre of the individual brass instruments horn, trumpet, trombone and tuba depends on the shape and size of the mouthpiece specific to each instrument. Note production is achieved through the embouchure. Loose lips produce low notes while for higher notes the lips have to be pressed tightly against the mouthpiece [103]. The attack in the lower register is complicated due to the looser lips in the embouchure. On the other hand, for the opposite reasons soft attacks are more difficult to achieve in the upper registers [3].

More wind is required for lip reed instruments than for woodwinds. Rests are even more necessary due to the physical stress when playing these instruments. Phrasing is almost the same as in the woodwinds. *Legato* phrasing is normally played in one breath [3].

Of the lip reed instruments, the trumpet is considered the soprano of the brass family. It is a versatile instrument in which fast and slow passages can be played. Previous models were valveless but nowadays three valves dominate the trumpet in Bb and C [3]. Louder passages are easier to play than softer ones. *Pianissimo* in the lower register is difficult to control. In contrast, the Bb trumpet makes playing in the higher register easier [41]. The instrument works better with louder notes than with softer ones. *Legato* notes can be played in one breath [3] On the other hand, the trombone works well both as a solo or as a harmony instrument. It can be thought of as a large trumpet. To achieve the different pitches, a system called a slide which makes the outer tube longer or shorter enables fine-tuning. *Legato* is only possible between two notes of the same harmonic series [3].

There are two types of horn: the natural horn and the valve horn. The first has a C pitch and its performance requires the correct embouchure and right-hand manipulation. Nowadays, the standard in the orchestra is the double horn in Bb or F. However, different horns can be used depending which register is to be played [41]. Notes in the lower register are more difficult to produce. It can also be strenuous for the horn player to perform continuously at higher pitches [3]. The horn is a demanding instrument and horn players must “mentally” hear the notes to be played. Subsequently, the note has to be played with the embouchure. Scores having fast passages or wide changing intervals are difficult to perform [3].
The bass instrument of the brass section is the tuba. The working mechanism is a piston or a rotary valve system. Slurred notes can be played with one breath and the instrument is suited to play attacks [3].

Table 2.2 includes a summary of the playing tools necessary to generate sound with the instrument and the body parts involved in direct (contact with the body) or indirect (using a playing tool) sound production.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Human body input</th>
<th>Playing tools</th>
<th>Technique</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tubular bells</td>
<td>Hand</td>
<td>Wooden or hard plastic mallets</td>
<td>Striking</td>
<td>Idiophones</td>
</tr>
<tr>
<td>Marimba, xylophone</td>
<td>Hand, foot</td>
<td>Soft mallets</td>
<td>Striking</td>
<td>Idiophones</td>
</tr>
<tr>
<td>Celesta</td>
<td>Hand, fingers</td>
<td>None (direct contact)</td>
<td>Striking</td>
<td>Idiophones</td>
</tr>
<tr>
<td>Cymbals</td>
<td>Hand</td>
<td>Soft beater, wooden stick, another cymbal</td>
<td>Striking</td>
<td>Idiophones</td>
</tr>
<tr>
<td>Tam Tam</td>
<td>Hand</td>
<td>Wooden, plastic or metal mallets</td>
<td>Striking</td>
<td>Idiophones</td>
</tr>
<tr>
<td>Triangle</td>
<td>Hand</td>
<td>Beater</td>
<td>Striking</td>
<td>Idiophones</td>
</tr>
<tr>
<td>Drums</td>
<td>Hand, foot</td>
<td>Timpani mallet</td>
<td>Striking</td>
<td>Membranophones</td>
</tr>
<tr>
<td>Violin, viola, cello and</td>
<td>Hand, fingers</td>
<td>Bow</td>
<td>Bowing</td>
<td>Chordophones</td>
</tr>
<tr>
<td>double bass</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piano</td>
<td>Hand, fingers</td>
<td>None (direct contact)</td>
<td>Striking</td>
<td>Chordophones</td>
</tr>
<tr>
<td>Guitar</td>
<td>Hand, fingers</td>
<td>Plectrum or direct contact</td>
<td>Plucking</td>
<td>Chordophones</td>
</tr>
<tr>
<td>Harp</td>
<td>Hand</td>
<td>None (direct contact)</td>
<td>Plucking</td>
<td>Chordophones</td>
</tr>
<tr>
<td>Clarinet, bassoon, oboe,</td>
<td>Fingers, embouchure (mouth, lips and tongue)</td>
<td>None (direct contact)</td>
<td>Mechanical reed</td>
<td>Aerophones</td>
</tr>
<tr>
<td>saxophone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flute</td>
<td>Fingers, embouchure (mouth, lips and tongue)</td>
<td>None (direct contact)</td>
<td>Air reed</td>
<td>Aerophones</td>
</tr>
<tr>
<td>Horn, trumpet, trombone</td>
<td>Fingers, embouchure (mouth, lips and tongue)</td>
<td>None (direct contact)</td>
<td>Lip reed</td>
<td>Aerophones</td>
</tr>
</tbody>
</table>

Table 2.2: Body parts and playing tools involved in sound generation

The list presented in Table 2.2 includes all the musical instruments relevant to the western orchestra. The information provided in this list is necessary in order to develop the methodology and it also enables the classification of the results from this research.

Sound generation is produced either through direct contact or by means of playing tools. Listing the different methods used to generate sound may enable a better hindsight when
it comes to understanding differences in latency perception with respect to performance. Haptic information gathered by the musician while performing musical instruments using direct contact, as in the case of the harp or piano, produces a different perception when compared to performing an instrument with a tool (e.g. mallets or bow) where the contact with the musical instrument is not direct.

Having described the different musical instruments involved in the research together with their main characteristics, it is now necessary to introduce the network technologies that play a relevant role in musical performances.

### 2.2 Music and Network Technology

This section presents the concepts and terminology of sound perception, latency and time. The current research and its outcomes which are related to this work are explained and discussed.

Telecommunication networks are not a new technological development. We have been using telephones and wired communication since the previous century. Networks enable information distribution and human communication [67]. Performances involving networked music are state-of-the-art in contemporary music composition and are becoming more and more a part of the artistic repertoire. Beyond this, network technologies improve communication between musicians, facilitating rehearsals, conducting, improvisation and even education [104]. Network music performances (NMP) have become a very interesting research field for musicians, especially those interested in new technologies, where an interdisciplinary approach in areas such as communication, psychoacoustic research and musical aesthetics is indispensable [104].

The introduction of personal computers or PCs such as the 1976 Commodore KM1 was a decisive event in establishing controllable independent networks and was a first step in the development of network music [228].

Interconnected musical networks enable musicians to influence, shape and share music in real time. These networks are assumed to be independent, dynamic and with a tendency to facilitate social interaction. Networks can be classified into two major categories:
“process-centered networks” and “structure-centered networks”. These two different categories may have resulted from the musical processes in the 20th century when “avant-garde” and “post serialist” European composers such as Karlheinz Stockhausen and Pierre Boulez advocated for composer control and musical structure. On the other hand, experimental American composers such as Steve Reich and John Cage preferred an escape from structure towards “process music”. For Reich and Cage, musical art was defined as a process with many different experiences (social, creative, exploratory) and participants (composers, performers, listeners) who collaborated to develop the musical composition. In the “process music” approach, the composer does not actively control all aspects of music and the construction of new musical structures far beyond the original work is also enabled [228]. Nowadays, the CCRMA group of Stanford University is one of the successors of the process-centered category. The network should be explored as a new medium for interaction and latency is not regarded as an impediment, but rather a feature that also shapes music and sound [104, 50] in an improvisatory scenario which relies on human performers to create a unique performance [62].

The role of NMP in education is extremely important for those living in remote locations. With the possibility of teaching musical nuances such as gestures, touch and different timbre issues over long distances in real time, the learning perspectives of subjects living abroad are significantly enhanced [84]. The number of distance learning programs has increased over the last decade and the most renowned universities offer different courses which may also include practical content [124]. In addition, universities offer regular live concerts between institutions as well as remote recording. Tuition in musical instruments on a one-to-one basis benefits from videoconferencing software such as Skype [166]. Dedicated software like Jack Trip and JMess [37] are also available for education and research purposes on the Internet.
Established research groups include the Soundwire Group at Stanford University\(^1\) led by Chris Chafe, the Network Musical Performance Project\(^2\) at the University of California, Berkeley with John Lazzaro and John Wawrzynek, the Real-Time Networked Media\(^3\) group at McGill University with Jeremy Cooperstock and the Princeton Sound Lab\(^4\) with Perry Cook [62].

Research done in networked music performance has been increasing since the first decade of the 21st century. Artistic, technical and even philosophical approaches take advantage of the current interdisciplinary discussion with respect to NMP. The most researched categories are listed in Table 2.3. The outcomes of the present work may bring new elements to further advance the understanding of psycho-acoustical aspects of NMP.

<table>
<thead>
<tr>
<th>Category</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td>Technical improvements in software and hardware</td>
</tr>
<tr>
<td>Sim</td>
<td>Use of network for music composition, performance, installations and choreography</td>
</tr>
<tr>
<td>Psychoacoustics</td>
<td>Relevant findings in the field with direct relation and application to NMP</td>
</tr>
<tr>
<td>Reports</td>
<td>Reports, survey, projects and installations</td>
</tr>
<tr>
<td>Philosophical</td>
<td>Conceptual framework and aesthetics with regard to NMP</td>
</tr>
<tr>
<td>Other</td>
<td>Development of technologies towards the use of NMP as an educational tool</td>
</tr>
</tbody>
</table>

Table 2.3: NMP research categories [104]

---

\(^1\) www.ccrma.stanford.edu/groups/soundwire.
\(^2\) www.cs.berkeley.edu/~lazzaro/nmp.
\(^3\) www.cim.mcgill.ca/sre/projects.rtnm.
The relevance of this work lies in the psycho-acoustical approach and the search for quantitative parameters which lead towards a better understanding of the relationship between the musician and the instrument.

2.2.1 Technical Fundamentals of Networks

For real time audio data to be sent over networks, a minimum level of technical conditions has to be provided. Three different networking area technologies are relevant for networked music performances. Local area networks (LAN) are the best option in terms of latency as the short distances make dealing with latency easier. In metropolitan area networks (MAN), latency is more relevant as distances are bigger, up to tens of kilometres. Using wide area networks (WAN), a connection between countries and even continents is established, meaning latency is a bigger issue. On the other hand, a WAN connection is interesting for the purposes of performing worldwide collaborative music [104].

The primary purpose of networks is the distribution of information. This information is sent in the form of units of data known as packets [42]. Although an unobstructed transmission is always envisaged, packet loss and delay variations within the transmission may be expected. Such issues may lead to dropouts resulting in signal errors and audible clicks or cracks, which are annoying and disturbing when performing music [193, 209].

2.2.2 Networked Music Signal Path

For a networked music performance (i.e. over the Internet), four main paths contribute to the total latency that has to be taken into account when playing music over networks [42].

Figure 2.2 presents an abridged version of the total latency paths described by Carôt [42]. The figure summarises the latency sources and their influence within the different paths.
The natural and electronic paths

The natural path describes the air transmission path which is a physical requirement in audio production and perception. The electronic path groups the main elements involved in the conversion of acoustic signals into electrical signals and again into air pressure waves (audible sound signals) as in the case of the microphone, amplifier and loudspeaker-headphones signal chain [42].

Musical instruments generate acoustical sound vibrations. To transmit those signals over long distances in a network, the sound pressure waves generated are converted into electrical analogue signals and then converted again into digital audio signals [190]. Analogue to digital and digital to analogue converters may each add 1ms of latency in the path. This latency will increase with the use of processing equipment and network transport. There is also more gear and processing in the path such as mixers, cables and electronic equipment. Once latency has been added to a system by any of these components, it can never be removed [28].
The digital path

For an analogue audio signal (electrical signal) to be converted into a digital audio signal (i.e. in a computer), the digital signal process introduces new stages. The processes of sampling, filtering and any digital signal processing, in addition to the effects of audio codecs by transmitting raw audio data, introduce noticeable delays [141].

In the worst case, comb filtering, phase cancellation [135] or echo are produced [152]. Latency can produce comb filtering, especially with in-ear monitoring (IEM). Sound conducted through the head is mixed with delayed IEM sound [135]. Because there is no linear relationship between latency and comb filtering, decreased latency does not always traduce in a decreased comb filter effect. Reversing the signal polarity in IEM systems can sometimes achieve better results than decreasing the latency of the system [135].

Besides the previous steps, it is sometimes necessary to encode and decode the signal to reduce the data stream. This reduction enables the transmission of data when the bandwidth is insufficient for the amount of digital information [42]. The process of decoding and encoding audio signals adds significant latency (tens or hundreds of milliseconds) and therefore should not be an option for a real time music performance over networks [104].

In order to provide a stable audio stream, digital data should be buffered and organized in blocks. Sound cards generate audio in blocks with a constant number of samples [42]. The device input blocking together with the driver buffering is one of the stages that adds a large amount of latency to the whole configuration. Two delays that contribute to the total amount of latency in the system are introduced when using buffers. Firstly, the record delay, where the processing of samples is preceded by the filling of a buffer. Secondly, the synchronization delay where playback is only possible when the next buffer is ready [130].

Audio processing in a personal computer is not executed immediately and time values for processing vary between systems. In general, when processing audio, a number of samples must be collected in order to be processed [42]. These samples are referred to as the “sample block”, “sample frame” or “audio buffer”. The audio buffer size depends on the system’s performance and it is limited to a minimum value of 32 samples. The size of a sample is two bytes [42]. A simple audio link with one direction, which implies acquiring and transmitting on one side and receiving and playing back at the other side,
may have a minimum amount of latency due to a buffer time at each end. Several audio slots fill a buffer. These audio slots are called periods. The values for these periods in modern computer architecture in 2016 are 128 to 512 samples at 44.1 to 96kHz or, in time units, 1.3 to 11.6ms. This figure is still high in comparison with modern digital mixer input-output delays due to the signal, which are around 0.8ms and up to 2.35ms [207, 104].

The delay increases with larger block sizes and decreases with higher sample rates. Higher sample rates produce a lower amount of latency but increase the bandwidth and the packet rate [29]. The following equation describes the trade-off between block size and sample rate when adjusting latency in this stage [42].

\[
\text{blocking delay} = \frac{\text{block size}}{\text{samplerate}}
\]

Equation 2.1: Blocking delay

It is not only the blocking delay that contributes to total latency in the digital path. Drivers for the audio interface and those of the personal computer, like “Core Audio” by Macintosh with the operating system OSX, also contribute to latency [42, 123]. The introduced delay which lies in the order of samples is too short and may not have been taken into account. For digital signals to be in synch, a word clock is mandatory. This word clock may also be a source of latency through the introduction of jitter in the system [42]. Word clock jitter is the deviation or variation between the theoretical and the real trigger event for synchronisation in digital audio systems [190].

The Internet path

The delay produced by the characteristics of the network connection or network delay is the main latency factor from the Internet path. The network delay is affected by two components. The first is the physical component due to the propagation of the signal in the medium. The second is the packet switching and routing within the network [104]. Other authors, such as Rottondi et al. [198], describe the transmission delay produced by the bandwidth and the propagation time as separate factors which influence latency.
An electrical signal flows with a velocity of approximately the speed of light at approximately $c=300,000\text{km/s}$ (i.e. electronic path). Digital information is transmitted with a damping factor of approximately $0.67c$ or $200,000\text{km/s}$ when using fibre optic [229, 57, 47]. Each $1,000\text{km}$ of transmission line adds $5\text{ms}$ latency even in the event of a direct path between A and B [43]. A round-trip time (RTT) across the USA (there and back) is approximately $40\text{ms}$ [57]. This latency value is not an issue [42]. However, routing between distant locations does not follow a straight line path between two points. In most cases, the distance is longer than the direct path [104]. As a matter of fact, even with centralized server topologies, acceptable performance distances are about $800\text{km}$ or less [44]. A direct fibre optic link may not exceed $5,250\text{km}$ assuming $25\text{ms}$ of delay. In practice, this figure is usually below $1,000\text{km}$ [47].

An audio packet arriving too late, or even not arriving at all, may produce a dropout. An additional source of dropouts is the loss of audio clock synchronisation either from transmitter or receiver [104]. When a packet arrives after its playback time, there is a gap before the packet is immediately played. The gap is perceived as a dropout [42]. Even under the best conditions, dropouts might be present. Zero dropout tolerance in practice is difficult to achieve and this is one of the challenges for musicians. They have to work with network conditions as they are and be able to interact in real time with other performers while monitoring themselves [167].

In addition to the delay produced by the IP (Internet Protocol) routers, where information packets are analysed and then delivered through a specific path [65], the Internet delivers a variable data traffic delay thereby introducing a large packet delay variation known as network jitter (different to word clock jitter). This jitter is also unpredictable and not constant. The effects of this jitter can be compensated for by allowing sufficient buffering. In this way, delayed packet arrivals are possible. Ultimately it is always a compromise: a larger buffer means a safer playback but also a greater latency. When performing music over networks a “latency budget” may be assigned to provide a feasible audio transmission [51]. The transmitted information over the Internet is not just the information itself (i.e. audio), but also additional data such as the Ethernet header, IP header and UDP header, whose functions are also to guarantee the delivery of the information to its final destination. This extra information causes additional latency [65].
In networks such as the Internet, packages of information are sent through the fastest route, depending on the availability of resources. A narrow bandwidth produces congestion, which may cause irremovable network jitter, delaying certain packages more than others. This can change the packet order, degrading the perceptual quality of sound [43, 112]. Every data packet delayed in audio networking will produce a perceivable error [57].

As a rule of thumb, for normal real audio flow in a high quality service (i.e. a research institute or university network with at least 1Mbit/s), the network jitter should be less than 0.4%. If this value exceeds 0.4%, dropouts may be audible as a result [52]. Low latency can also be achieved by routing signals manually and reducing the number of switches in the path. New technologies enable 100Mbps over Cat5 cable or even 1Gbps for 1000Base-T Gigabit Ethernet as state-of-the-art in the year 2016 [104].

The Internet, being an asynchronous\(^5\) medium, is not the best choice for transmitting real time audio data [47]. However, the Internet is, nowadays, the most popular and accessible way to send and receive information. The problem related to the asynchrony of transmission is partially solved with an increase in bandwidth, which also increases the propagation speed. The network jitter problem is not as significant as it was some years ago [104, 42]. Research groups are working on the development of global metronomes using algorithms such as the Network Time Protocol (NTP) to synchronize time on the Internet. Nevertheless, the influence of many factors, including the network route asymmetry and accuracy of time servers, complicates the implementation of standardised metronomes for music applications over the Internet network [175].

As stated before, some factors influencing the latency budget are the audio quality and the sampling rate. The music style is also a decisive factor. Classic, popular and jazz music require a real interaction and faster audio feedback than other music styles such as avant-garde or classical contemporary music, which rely on improvisation and therefore are more latency-tolerant [104].

Another issue which makes audio over the network more complicated is the so-called “network porridge” [133], perceptual audio artefact (crackly and poppy audio) produced by

\(^5\) Asynchronous, synchronous and isochronous are the three different network information transfer modes. In an asynchronous transfer mode, network signal transmission happens randomly [47].
some network transmission issues. The Transmission Control Protocol (TCP) retransmits lost audio packets. A confirmation of data arrival is always necessary for this protocol, therefore an increase in the latency time and errors in the transmission are to be expected [42].

In live audio networking systems (e.g. music venues, concerts, etc.), the TCP produces dropped or delayed audio packets. Humans perceive and notice any absence of audio data transmission in a continuous playback. Glitches in data voice recordings or transmissions are distortions in the audio signal reproduction [199] and these are easily detected by listeners [213].

Other transmission protocols, such as the User Datagram Protocol (UDP) are not so reliable in terms of message transportation. Having a smaller packet header in comparison with the TCP, UDP is better for real time applications and is more efficient in terms of data delivery velocity. However, there is no confirmation of data arrival in the UDP as in the case with TCP [198]. UDP has become the protocol of choice for network music performances [104]. Indeed, the “SourceNode” project has shown that within a local area network (LAN) it is possible to connect computers with their respective digital audio workstations. These remain synchronised due to UDP and MIDI clock [195].

Latency over the network will always be present even after disregarding signal processing. Long distances worldwide between source and destination and the laws of physics introduce latency values [46] which make real time musical collaboration difficult [18]. Nevertheless, audio transmission in real time over the network is feasible [99].

There are two possible scenarios when referring to latency over the network. The first scenario is “one way at a time”. An example is a musician conducting a master class in a location while the students listen to the musician in another location. Delay should not be an important issue in this approach. The second scenario is the “two-way real time interactive network performance”, where real time interaction between musicians is possible, but latency becomes an issue. This is the case with collaborative performances over the network (two musicians performing) and non-collaborative performances (a teacher holding a lesson for a student overseas) [21].
The following equation is an estimation by Rottondi et al. [198] of the different delays contributing to latency as a whole in a music networked performance (NMP). In their “Overview on NMP Technologies”, the authors describe the total delay with the following equation:

\[ D_t = 2(D_a + D_c + D_s) + D_b + D_p + D_t + D_q \]

Equation 2.2: Total delay experienced by the user of an NMP System

where:

- \( D_a \) = Analogue to digital conversion delay.
- \( D_c \) = Delay due to audio signal compression.
- \( D_s \) = Blocking delay.

In the equation these factors are multiplied by two, simulating the process by the transmitter and the receiver [198]. The air propagation delay is not taken into account and defined as irrelevant [198].

- \( D_b \) = Application buffer delay.
- \( D_p \) = Propagation delay.
- \( D_t \) = Transmission delay.
- \( D_q \) = Processing delay.

### 2.2.3 Taxonomy of Musical Networks

At this point, the relevant technical issues have been defined. It is pertinent to define the taxonomy of a networked music performance. Table 2.4, extracted from the work of Gabrielli et al. [104], specifies the more general aspects.
An explanation on the use of networks for musical performances and different approaches is described in the work “Interconnected Musical Networks: Toward a Theoretical Framework” by Weinberg [228]. Table 2.5 summarizes this information and enables an easier understanding and comparison of the approaches. For this research, the main interest lies in the “bridge approach”. It is a purist approach and uses high-bandwidth technology conditions for human performers. Nowadays, this approach is used primarily for research purposes [104].
<table>
<thead>
<tr>
<th>Approach</th>
<th>Definition</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Server</td>
<td>The network is used primarily to send musical data. There is no participant interconnectivity or communication among players. The process is passive, players are allowed to trigger and listen to sounds when interacting with a specific Internet link.</td>
<td>The Sound Pool Web application contained in the interactive piece “Cathedral” by William Duckworth [86, 78].</td>
</tr>
<tr>
<td>The Bridge</td>
<td>Musicians connected over large distances as if they were performing in the same space. Performers can listen and play music with other participants. The network provides a platform to perform in which musical collaboration can be enriched. The main challenge is to find the best way to cope with latency and therefore to enable a performance.</td>
<td>The bi-directional jazz/party event between McGill University in Montreal and the CCRMA in Stanford under the name 2002 Distributed Jazz Jam + party [53].</td>
</tr>
<tr>
<td>The Shaper</td>
<td>In this approach, the network plays a central role by generating musical material by means of algorithms. Participants are allowed to collaborate due to modification and shaping of the musical structures. Direct algorithmic interdependences among players are not supported.</td>
<td>Distributed interactive music [179].</td>
</tr>
<tr>
<td>The Construction Kit</td>
<td>Musicians play and contribute to musical creations beyond performing. The interconnectivity is the main characteristic of this approach. Musical pieces might be downloaded, updated with new musical information and then uploaded creating a new whole musical piece.</td>
<td>The WebDrum application. Users are allowed to turn notes on or off on a grid from a traditional drum-pattern [34].</td>
</tr>
</tbody>
</table>

Table 2.5: Network musical approaches [228]

These different approaches concerning the use of the networks in music performance are part of the taxonomy and theoretical work necessary for the scope of this work.
2.3 Sound and Latency Perception

The following section describes and summarises recent work on the fundamentals of sound perception. Auditory and visual perception can differ from person to person [202] and this makes it difficult to develop methods or tests when attempting to generate quantitative results in order to confirm or reject theories concerning the perception of audio. However, it is possible to measure perception effects in an indirect manner. For this research, correlated perceptual events inherent to human actions are the main focus. These events are also referred to as perceptual feedback [147].

2.3.1 Sound Perception in Time

It has been estimated in experimental set-ups that auditory stimulation is perceived to last longer than its visual counterpart [101]. This assumption is valid for experiments with a duration of more than one second and with sound and light stimuli presented simultaneously.

To get an idea of the relevant time-scale for this research, it is important to outline previous work connected to human perception. The psychological present is a short-term auditory memory with an average value of approximately 2 to 3s, and in very rare cases it exceeds 5s. Within these time limits, it is possible to speak of the perception of duration [101].

Human sensitivity has been the subject of methodological studies since Wilhelm Wundt in 1870 [153, 109]. Topics studied include simultaneity and the ability to detect differences in the order of milliseconds for onset times. For humans to be able to hear a click sound without special pitch or timbre, it is necessary for the sound event to have at least a duration time of 2-5ms, so-called lower “click” threshold. Pitch certainty increases with an increment of this value up to 50ms, which corresponds to a frequency of approximately 20Hz [17]. Below 20Hz, impulses are perceived as clicks. The 20Hz threshold is considered the point of event discrimination. Some physical properties like ear geometry, ear channel or even the synchronisation of nerve fibre oscillations are involved in this detection threshold of the ear [17].
Age is an important factor regarding the ability to process information temporally and to identify stimuli separated by a gap. Age-related deterioration also depends on the physical properties of the stimuli [215]. Humans can distinguish between two different auditory stimuli when the temporal order (TO) corresponds to a gap of approximately 30ms. This value is also reported in experimental data supporting the idea of temporal information processing (TIP) where a time window within discrete information is processed [215].

Research has shown that the average threshold, the temporal order threshold (TOT), lies between 20 and 60ms and that the TOT declines considerably with age, which could be a consequence of a slower processing speed in the brain (a slower TIP) or even due to the deterioration of a hypothetical internal clock [215, 101]. It is important to stress that not only age, but also the sound stimuli used, can have an effect on the TOT. A crackling sound between very brief stimuli is perceived and identified at a separation of approximately 1ms [101].

Human beings have a very high adaptability. A vocal conversation is possible even with one-way delays (OWD) up to 500ms [119]. Researchers have found that quasi-repetition\textsuperscript{6} of two sound events should exceed 50µs between sounds for them to be perceived as separate events. Sound repetitions mostly happens in the so-called “psychological present\textsuperscript{7}” within 100ms to 2s. Periodic events between this range could be perceived as a rhythm or meter [17]. Meter is actually a rhythm component in addition to rhythm patterns, tempo and timing [120]. The threshold for pitch and rhythm perception as mentioned before is between 50 to 100ms. Some authors [17] use 100 to 130ms as the time value for rhythm perception.

Perception of synchronicity can exist even when two sounds are not physically synchronous [231]. For two musical notes to be perceived as separate time events, 2ms are necessary [154]. Further sound characteristics like timbre, pitch or loudness, the context of the musical style performed and visual and physical stimuli have a strong influence relative to the perception of two different sounds performed at the same time [18]. For example, two legato string notes are harder to separate than two glockenspiel strikes. Asynchrony of the sounds, but not the order, can be identified within the range 3 to 20ms. The brain

\textsuperscript{6} A sound event never repeats exactly because of noise floor and added noise even in a perfect recording [119].

\textsuperscript{7} Time values relating to the psychological present may appear contradictory. However, there is no single method that allows a precise measurement of the duration of the psychological present [129].
uses the time difference to extract source direction cues [209]. Experiments made with clarinet and violin players by Bartlette et al. [21], found that tone onsets are identified at around 20ms separation for both musical instruments.

Temporal structures or spatial tasks are better perceived through the auditory modality in comparison to the visual modality, while the opposite can be said about spatial structures. The spatial resolution is higher as in the ventriloquism effect [12, 206]. Sounds are localised where the visual information is placed, even if the physical source is elsewhere [128].

Table 2.6 from Bader [17] summarizes the different time intervals with regard to the perception of sound, rhythm and patterns.

<table>
<thead>
<tr>
<th>Time interval</th>
<th>Frequency</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sound</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62.5 - 50µs</td>
<td>16-20 kHz</td>
<td>Upper hearing threshold</td>
</tr>
<tr>
<td>~ 250µs</td>
<td>~ 4kHz</td>
<td>Upper pitch threshold</td>
</tr>
<tr>
<td>2 - 5ms</td>
<td>200 - 500Hz</td>
<td>Lower timbre threshold above simple “click” perception</td>
</tr>
<tr>
<td>62.5 - 50 -ms</td>
<td>16 - 20 Hz</td>
<td>Lower pitch threshold, sound event separation threshold</td>
</tr>
<tr>
<td><strong>Rhythm</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 - 130ms</td>
<td>10 - 8 Hz</td>
<td>Upper rhythm production speed threshold</td>
</tr>
<tr>
<td>~ 200ms</td>
<td>~ 5Hz</td>
<td>Upper timbre perception threshold</td>
</tr>
<tr>
<td>500ms</td>
<td>2 Hz</td>
<td>Walking/movement speed peak</td>
</tr>
<tr>
<td>1.8 - 2 s</td>
<td>~ 0.5 Hz</td>
<td>Fastest beat performance tempo</td>
</tr>
<tr>
<td><strong>Pattern</strong></td>
<td>2 - 5 s</td>
<td>Groove pattern, melodic unit time (short term memory)</td>
</tr>
</tbody>
</table>

Table 2.6: Timescale of musical events [17]

Some of the data has been re-evaluated over the years and it is to be expected that further research may arise suggesting different time ranges. Different time intervals are strongly related to different frequencies. The more complex the sound event, the more time the brain requires to perceive it [17].
2.3.2 Musical Events and Time and Body Movements

Music itself exists and can be perceived only in connection with time. When performing collaboratively, hearing one’s own performance and the performance of the other musicians at the right time is what assures the ordering of musical events [21]. However, it is crucial not to adopt the idea of music being something that consists of discrete events [231]. A sense of pulse, the feeling of temporality which is shared by individuals, belongs to performing as an ensemble [209]. For solo performers, temporality is also inherent to playing music. It is important to underline that musical perception is also susceptible to elements not inherent to itself. Perception is affected even when listening to a piece of music alone or with others [154].

Levels of tolerance relating to timing variances while listening to solo or ensemble musical performances and in connection with acoustic environments and musical content are to be expected. It is likely that different musical instruments may have an effect on the musician’s adaptability to the room’s acoustic environment [203].

There are some explanations for the question of how musicians separated by more than few meters can perform together, regardless of internal timing variances. Ear sensitivity is not perfect. Moreover, a perfect synchronization is not needed. Expectation of what note comes next may also play a role. A standard performance involves isochronous timing. In other words, timing with equivalent durations [218]. No musician can perform with perfect and accurate timing. Furthermore, very good musicians violate the expected timing. “Temporal contrast” is the name of this deviation of the observed from expected timing which provokes a “surprise” in the listeners [129].

Several piano research studies have established a variation of the “interonset interval” (IOI) [154]. The IOI is the time between the attack points of successive events or notes, without including the duration of the events [154]. The time between two successive key presses varies constantly from event to event [182]. Asynchrony by a very few milliseconds and not mechanical exactitude is what makes a performance worthy [21]. Deviations from the isochronous beat, due to accelerando or ritardando produce “expressive timing”. Even less obvious variations from the strict metronomic time measurement are enough to enable musical expressivity [218]. Minimizing such variations and achieving regular timing is part of the training of a good musician. Eliminate the very small variations required
to achieve perfectly isochronous timing is an impossible task. The physical properties of the human body are unable to achieve mechanical timing [218].

It has been observed in research with master musicians that produced meter does not always reflect scored meter [129]. Some musicians may introduce slight variations. In his dissertation Carot defines [42] the inter pulse delay (IPD), which is the time shift between a local and external pulse, as highly dependant on the musician’s style and personality.

Rhythm is approached without conscious effort and rhythm surfaces can be easily reconstructed, particularly by musicians. Furthermore, our capacity for rhythmical response is present from the very beginning of our lives. These capacities can be modified by the environment through learning. They evolve throughout our lifetime and the rhythmical response emerges through different activities for non-musicians and in the production of sounds for musicians [154].

Since research suggest that our auditory experience is mainly influenced by learning and through exposure to auditory events, it is unlikely that the meaning of sounds remains invariable through time [122]. Physical and phenomenological factors intervene in every interaction which involves audio and visual information [209]. The learning process is driven between intent and perception, so the musician adapts himself in a conscious or unconscious way [231] and therefore motor control over a musical instrument can be achieved.

While performing music, different kinds of sensory feedback are available: visual feedback from the movements of every person involved in the performance, including the musician himself; tactile feedback due to physical contact with the instrument; kinesthetic feedback from the proprioception in joints and muscles and obviously also auditory feedback [74]. Certain body movements are not related to or even needed for sound generation, although these conscious or unconscious movements have been observed during performances. These movements not only enhance communication between performers, but are to some extent individual interpretations and therefore also related to experience and behaviour. Movement is an additional form of communication with the audience [77]. Studies on pianists indicate that musicians rely more on alternative sensory information when the original sensory information is removed [177].
Musical expression cues, particularly those related to articulation, sound level, tempo, tone attack and timbre in music are essential to musical performance. Examples include slowing down or speeding up over time, the use of agogics, which is the musical artistic shape that influences the duration of a note [3, 196], or changes to the types of attack on notes which delays or accelerates attacks and varies the dynamic. These expression cues are a very important response of musicians to score directions and tempo markings [21]. Making a performance “interesting” demands the use of intentional deviations from the music notation in order to communicate the musician’s interpretation of the musical structure. This communication is achieved apart from the music, mainly due to physical motion, and is helpful to the listener [217]. In research on expression in drumming performances, some interesting results were found. Sad performances are usually slower in tempo at 62 beats per minute (BPM) in comparison to happy performances at 192 BPM. Sad performances often utilise legato articulation, whereas happy performances have a fast mean tempo, higher sound levels and the articulation is primarily staccato [149]. Compositions with higher and more complex structural relationships are performed slowly in comparison to those with simple structural relationships. A decrease in the measured tempo was noted for both solo performers and duet performers [177].

Although it is a physical impossibility for musicians to keep constant tempo without the use of a metronome, professional musicians can achieve better results in timing when using one [205]. However, no musician, even with metronome support, can achieve a perfect time striking every note at its onset [218].

Some of the reasons why humans perceive music as synchronous, even though musical instruments are not performed at the exact same time are described by Rasch [194]. The most important reasons are:

- Pitch and timbre are variable for different musical instruments. Therefore temporal structures are not the same and the attributes of tones differ.

- Onset differences follow a near random distribution.

- Listeners are normally not interested in the different instrument onsets (vertical distribution on the score) but in the melody (horizontal distribution).

- Sometimes even musicians are unable to notice the amount of de-synchronisation, except by “temporal mistakes”. These mistakes are found frequently in the challenging parts of a performance.
2.4 Latency Measurement Approaches

The next section outlines a summary of the research and methods related to the measurement of latency in performances. This review includes approaches and definitions which are relatively new, while others that have been part of research in fields such as psychology and physiology have been around for decades.

An “auditory inconsistency” is produced by delayed signals when musicians are exposed to latency. This exposure to delayed signals exists by performing solo in a non-collaborative way or when listening to each other with added latency while playing in a collaborative performance [27]. Bartlette et al. [21] have described the delay as being due to physical factors (travel distance from the instrument to the performer), i.e. “external latencies”. For small chambers, this figure lies between 5 and 10ms. In a western music orchestra, long distances between musicians as experienced by double bass players [176] would produce delays due to sound transmission through the air of around 80 to 100ms [21]. Delays may vary depending on the orchestral seating arrangement. In the European (German) seating arrangement the double basses are to the left. In the American arrangement the double basses play from the right [165].

Synchronous rhythmic interaction in music demands simultaneity in hearing, sharing and feeling the beat, in contrast to speech and conversation where latencies up to the order of half a second are tolerable [54]. Between sound production in a musical instrument and the perception of this sound, latency is always present. In “meaningful auditory sequences”, auditory events such as those produced by speaking or playing a melody on the keyboard, latency is also present [186].

The degree of external latency varies depending on the kind of performance. For a trio, quartet or quintet, where the physical separation of the musicians lies in the range of 2 to 3m, the one-way delay (OWD) due to separation is around 6 to 9ms [54]. Assuming standard temperature and pressure conditions, each 1m of distance adds 3ms of delay [104]. For an orchestra, the distribution of latency times is slightly different. The distances between the instruments and the performer’s compensation for acoustic latencies produce a different perception of music relating to time at every point in the hall. Depending on the position of the musician in the orchestra, musical notes could be delayed due to distance and sound travelling velocity up to a semiquaver or a quaver apart [209]. In an orchestra,
delays in the range of 46 to 60ms are possible due to the separation between musicians [104, 67]. The role of the conductor mitigates this problem. In his role, expression and coordination are integrated. One important task is to keep the performance synchronized [194]. For artistic reasons the conductor can alter the tempo and accommodate both coordination and expression, but in most cases the delays should be compensated for to reach the desired expression. The conductor’s expertise is based on knowledge of attack times (important in determining the instrument’s entrance) and relative sound pressure level of the musical instruments [176, 21]. In other words, when a conductor is present, latency is normally not an issue because the conductor compensates for the delay by thinking and acting ahead of time [45].

Music played by performers distributed spatially over networks such as the Internet is prone to lose synchronicity due to even larger amounts of latency (digital path, Internet path) in comparison with the delay produced by the acoustic characteristics of the room (natural path).

2.4.1 Latency and the Performance of Music

The perceptual onset of a note is the moment when the stimulus is perceived by the listener [225, 231]. A physical onset begins when the sound level amplitude of the instrument is different to 0dB. At this moment, the generation of the stimulus has begun [225]. A graphical description of this moment is displayed in Figure 2.3 showing the amplitude envelope of a hypothetical sound [231].

![Figure 2.3: Hypothetical physical and perceptual onsets and perceptual attack time](image)

Figure 2.3: Hypothetical physical and perceptual onsets and perceptual attack time [231]
Time differences between perceptual and physical onsets depend on the musical instrument being performed. The differences increase for instruments with low tone intensity like bowed string instruments with relatively long rise times [231]. In this case, the perceptual onset depends mainly on the relative threshold. For instruments with short rise times (piano, drums, harpsichord and percussion instruments), perceptual and physical onset are not so different [194] or could even be the same [231]. The perceptual onset depends primarily on the human hearing ability to adapt to specific relative stimulus levels. Both physical and perceptual onset exist only as discrete musical events [231].

Musical instruments, together with the acoustic characteristics of the room (i.e. the stage and audience) and the performed musical structure, play an important role with respect to latency perception [201]. It is well known that the tempo in which a musical passage can be performed depends on the acoustic characteristics of the location [211]. For instance, a musical piece written for allegro is faster and more easily performed in places where echo and outdoor influences are not an issue. Performing the same piece in a concert hall could be just possible at a slower tempo [11].

The question of the influence of musical instruments on latency has been postulated for a long time [116]. Research from 2006 by Bartlette et al. [21] continues with this inquiry in a subjective form by asking the musicians. In 2015 Rottondi et al. [199] used the timbre and spectra of seven different instruments (acoustical and electrical) and characterize them the statistical moments of their spectrum, such as spectral kurtosis, spectral skewness, spectral spread and centre of gravity. Guitars and drums, for instance, produce a larger tempo slowdown when the latency lies above 35ms. However, in general, rhythmic complexity results in a tendency to decelerate [199].

Some humans are able to perceive latencies below 10ms. For musical performances latencies can increase to 50ms while performance is still possible [141]. Musicians can cope with latencies of several hundredths of milliseconds, a common situation in church organ performances. Organ players compensate for the delay naturally between the key pressure onset, the sound emission at the pipes and the time it takes the sound waves to travel back from the pipes to the performer [199]. A similar situation is also true for piano players. For the piano, delays vary between 30 to 100ms depending on articulation (e.g. legato,
accent, etc.) and loudness of the instrument [14]. Continuous practice under the same circumstances, in particular for the organ, may be the key to coping with those latencies [156].

Having researched the effects of latency on live sound monitoring, Boley and Lester [152] came to the conclusion that drummers and keyboardists perceive latency in similar ways through different performance situations. This sensitivity to latency may depend more strongly on the musical instrument rather than on the individual. However, as in many other studies, Boley et al. [152] also came to the conclusion that having some latency while performing is preferable to no latency at all. In order to cope with latency, musicians develop various strategies, one of the most important is to anticipate the actual attack time.

A possible ranking for latency sensitivity based on experiments in live monitoring for musicians on stage was introduced by Lester [152]. According to the authors, the musical instruments in the study could be classified from “not confident” to “highly confident” with respect to latency in the following order: saxophone, vocals, electric bass, electric guitar, drums and keyboard. This means, for example, that vocalists are more sensitive to latency than keyboardists. Live sound mixers claim that even 3ms latency by in-ear monitoring can make a performance annoying for a musician on stage. Besides this, comb filter effects due to latencies in audio signals lead to confusion for the majority of musicians [207]. Nevertheless, further research is required to confirm this result. It is expected that for musicians playing fast attack instruments like the drums, latency is more of a problem than for those playing instruments with less attack [218]. Other experiments will confirm these results, at least for some instruments. Research on the keyboard by Kleimola [141] showed that in collaborative performances musicians tolerated latencies up to 69ms without any inconvenience, except for the bass and the drums. It is important to note that funk music was being performed in this study. Playing such music requires a tight time interaction between drummer and bass player.

Even for the same musical piece, performers are able to cope with less or more latency, depending on the sound of the performed instrument, as found by Sawchuk [201]. A Piazzolla piece was played on a keyboard having an accordion sound and a synthesized piano sound. The respective tolerated latencies were 25ms and 100ms. Possible explanations may be the acoustic properties of the sound. In their experiment, Boley and Lester [152]
had the same musicians perform different instruments. The ratings with respect to latency sensitivity were consistent within the instrument type. The same musician was also able to cope with different latencies on different instruments.

Musical style as mentioned before also influences the ability of musicians to cope with latency. Slow rhythm music allows very long delays. On the other hand, faster rhythms and highly complex musical patterns require very small delays [145].

Another aspect related to the way in which musicians cope with latency is the haptic information (touching) and the influence of body movements. With regard to sound production, body movements can be functional or non-functional [168]. Functional movements are those necessary to perform the instrument. Non-functional movements do not directly influence the production of sound (e.g. in a piano performance the neck, hip or shoulders) and these movements are not directly related to the action of playing the instrument [168]. As stated before, expressivity is strongly related to body movements [76]. Musicians who play instruments that have a physical coupling with the body (vocals or saxophone) are prone to easily identify the artefacts of delayed and direct sound. In-ear monitoring (IEM) may produce additional psychological effects such as annoyance because of incongruous signals between the vibrations of the instrument and the delayed sound [152]. A relevant issue is the occlusion effect, which occurs when an object fills the outer portion of the ear canal. The bone-conducted sound vibrations are trapped, producing a “boomy” or hollow sound due to amplification of lower frequencies (below 500Hz) by 20 to 30dB (e.g. hearing aids) [135, 136].

The physical sound characteristics produced by a musical instrument might influence the ability of the musician to cope with latency. For instance, the physical onset of percussion instruments is well defined. The duration of the excitation is short and control over an initiated tone by the percussionist is very restricted in contrast to the wind or string instruments. Moreover, percussionists depend more on timing and dynamic level [74]. Critical listening skills in addition to the musical instrument are, for Boley and Lester [152], some of the factors which might influence the perceived quality of a given amount of latency. Depending on the instrument, some musicians would require more immediate feedback than others, as is the case for the saxophone but not for the keyboards. Percussive instruments, due to the absence of tonality (at least in comparison with other instruments), are not so immune to perceived comb filtering.
Under the name of Distributed Immersive Performance (DIP), Elaine Chew, working together with other researchers [62, 63, 233] analysed realistic music interactions and the effects of latency on the aural response of performers. The experiments relied mainly on information obtained for musical interaction using a musical instrument digital interface (MIDI) keyboard. Experiments on DIP began in December 2002. In May 2004 a new experiment was conducted [63]. This time two players performed a movement of Poulenc’s “Sonata for Piano Four-Hands”. The audio delay varied between 0 to 150ms. In the second part of the experiment, players switched parts. Three different movements were played: “Prelude”, “Rustique” and “Final” with respective tempos of 132 BPM, 46 BPM and 160 BPM. In a scale rating with values from 1 to 7, the musicians described how difficult it was playing, adapting and creating a musical interpretation. For both players it was possible to perform as ensemble up to a delay of 50ms for the “Prelude”. For the slow movement, “Rustique”, musicians were even able to cope with a latency of 75ms. None of them could cope with latencies around 150ms. The ability to cope with latency is strongly affected by the tempo [63].

With a similar methodology, the same author confirmed the results [62]. In addition, the latency threshold of 50ms could be increased up to 65ms when introducing a delay to their own performance. In other words, it was not just the other performer who was delayed, but also the individual performance. The auditory perspective becomes increasingly similar after adding this delay.

A very interesting approach based on the musical performance of both bass and drummers is presented in the work of Carôt “Towards a comprehensive cognitive analysis of delay-influenced rhythmical interaction” [5]. In this research, five professional drummers and a bass player performed drum patterns at different tempos (60, 100, 120 and 160 BPM) and increasing delays in 5ms steps until the musicians could not cope with any further latency.

Results showed that there was no such thing as a “common latency acceptance value”, and it was not possible to define a delay threshold for ensembles. Moreover, the relationship between speed and note resolution of the pattern performed with regard to the delay acceptance threshold is directly proportional. Higher BPMs and higher note resolution lead to lower delay thresholds [5], a relationship that is found in different research.
The term EDAL, “ensemble delay acceptance limit”, was introduced by Carôt and Werner [5]. This figure is set-up dependent. For increasing note resolution and BPM values, this value decreases. Delays beyond the EDAL make performing music difficult.

For musical interaction or collaborative performances, the influence of the corresponding playing styles and the “personal beat shift range” (PBSR) have to be taken into account. The PBSR is a musician’s temporal range of acceptance. It defines a deviation from the musician’s perceived pulse and the theoretical beat reference [42].

The PBSR is highly dependent on the musician’s performing abilities. It evaluates personal tolerance for deviations from perfect onset synchronisation when performing collaboratively [199]. Having different musicians with different PBSR values, Carôt concludes that is not possible to define a common valid latency threshold where collaborative performances are possible [42]. The experiments considered two possibilities. Firstly, delaying one musician’s signal with the round-trip delay (RTT). A player that is able to cope with the actual self-delay is in perfect synchronization with their counterparts when the resulting delay gap is set to zero. This principle is called “single delayed feedback” (SDF). In contrast, a second possibility is “dual delayed feedback” (DDF), and involves symmetrical or asymmetrical delayed feedback between the two performers in order to achieve the same playing conditions.

The following mathematical expressions describe the minimal required total self-delay in a collaborative performance.

\[
dF_{\text{min}} = (OWD - EDAL) + \varepsilon
\]

Equation 2.3: Minimal self-delay

where:

- \( OW D \) = One-way delay
- \( EDAL \) = Ensemble delay acceptance limit
- \( \varepsilon \) = Delay due to non-determinate human error
The value of $\varepsilon$ increases with the amount of self-delay. In some cases, the increase leads to a collaborative performance breakdown. This delay reduction after the minimal self delay produces a delay offset between the players, described by subtracting $dF_{\min}$ from the RTT to calculate the actual offset.

\[ \text{offset}_{\max} = \text{RTT} - dF_{\min} \]

Equation 2.4: Actual offset

where:

RTT: Round-trip delay

These results are the product of a specific musical performance experiment between a reduced number of musicians and musical instruments. The authors state that there are slight variations between player to players but, in general slow pieces (slow BPM) allow higher latencies.

After analysing three different musical pieces performed by musicians under simulated adverse network conditions, Rottondi et al. [199] concluded that the choice of musical instruments and their combination is crucial with regard to the ability to keep a steady tempo. Complex musical parts also have a negative impact on the ability to perform music and have a tendency to decelerate the performance when the network latency is higher. In addition, it has also been observed that timbral features of the musical instruments have a relevant impact on the subjectively perceived delay. Guitars and drums are some of the instruments with a higher spectral entropy and spectral flatness. Those musical instruments have a tendency to slow down in the presence of higher network delays. For Rottondi, the role of the selection of musical instruments and their combination within a musical performance may be as relevant as the network delay itself. While latencies above 75ms may not be an issue for collaborative performances, the choice of musical instruments and the kind of music are relevant.
2.4.2 Effects of Latency on Musicians

Network latency, also known as net-delay, has a highly disrupting effect on music with traditional melody and rhythm elements where tight synchronisation is necessary [18]. From research on keyboard performances [94], it is known that auditory delay causes significant errors in piano performances. One of the sources of these errors is the dissimilarity between sound and tactile feedback. For collaborative performances, network latency is a critical issue. Musicians are affected by their own acoustic feedback, which influences the overall collective music response [20]. Small latencies of a few milliseconds are not perceived as a delay in the signal [152]. However, these latencies can produce a comb-filtering effect, which could eventually affect a musical performance making it difficult to estimate when to begin with sound production in the instrument. It is also known that shorter network latencies may cause a higher pitch. A very long lag (latency > 50ms) could induce very low pitches and echoes [50].

Other effects are related to the difficulty in identifying the direction of audio sources because of latencies in the left or right channel. Due to comb filtering, the delayed sound may be very uncomfortable, especially for drummers and percussionists. In this regard, delays adding more than 25 to 50ms to the original signal may force the brain to understand the original and the delayed signal as two distinct sounds, which could be very annoying when performing a musical instrument [100]. However, it is known from Chew [63] that with practice, adaptation to latency for delays below 50ms is achievable.

The annoying effects of latency on musicians in connection with collaborative music performances can be summarized as follows:

- Difficulty to maintain synchrony. In collaborative performances, musicians try to align their beat with the other musicians [205]. In other words, rhythmic precision decreases as latency increases. This may continue to a point where playing music together becomes impossible. The performance is disrupted [56], ending with a the loss of musical coordination [84].

- Due to delays, musicians are unable to play on time and thus performers do not play together. As a result, during the performance, the musicians slow down the tempo. In other words, they play rallentando. When the delay is perceptible, musicians tend to swing the long beat [205].
• Through latency, syncopation turns into unison, synchrony into asynchrony and anticipations are perceived as arrivals [209].

Musicians are able to cope with latency to some extent. Large asynchronicities may not cause performance problems [156]. Higher latencies, even up to 20 to 30ms, are not an issue for most multimedia and music applications [172]. A value of 150ms is acceptable for quality telephony according to the International Telecommunication Union [29]. Even higher latencies are not an issue when it comes to localizing sounds. Front-back confusions were minimal in listening experiments with latencies as high as 500ms [230]. In hand-clapping experiments done by Farner et al. [90], it was observed that musicians were prone to slow down the tempo more than non-musicians. A possible explanation is the awareness of other performers while playing. Slowing down ensured a better ensemble playing.

Different authors have researched the values of latency which make performing difficult or which could even disrupt a musical performance. Bartlette et al. [21] found that at the 100ms threshold, musicians experienced the performance as neither musical nor interactive. Performers began to play in isolation from each other, listening only to their own instruments. From personal observations carried out by Keltz [135], timing is very difficult to maintain around 15 to 20ms, and larger delays slow the performance down. On the other hand, he states that performers, audiences and even audio engineers can tolerate long latencies. Latencies below 10 to 15ms are an issue under frequency response changes. Sounds that are more than 35ms apart are perceived as echoes or separate sounds.

Other authors like Chafe [56] believe that musical ensembles are able to perform over the network even with one-way delays (OWD) in the range from 20 to 60ms. In other research [54], the same author found that at delay values of between 55 to 66ms “the edge of playability” is reached. At this point, lag and asymmetry are so disturbing that even strategies to cope with latency are not a solution. It is important to note that these values mainly resulted from experiments with random people clapping their hands under musical parameters and that it was not a musical performance involving instruments. Listening experiments with real musicians in real performances develops a further and better understanding with regard to latency and music [42].

The Ensemble Performance Threshold (EPT) is defined by Schuett as a threshold between 10 to 40ms in his work “The effects of latency on ensemble performance” [205]. According
to Schuett, a performance is still possible with latency values of 11.5ms or more. However, beyond 40ms, synchrony and tempo are compromised and all ensemble features are lost. It is important to note that the EPT was also found for experiments involving hand-clapping.

Similar research done on latency came to different results. Mäki-Patola [157] states that a threshold between 20 to 30ms was found to be just noticeable when compared to 0ms latency. Test subjects noticed the effect after 60ms. After the 30ms threshold, the probability of detecting latency gradually rises. As stated before, the threshold of perception for temporal segregation is approximately 30ms [84, 215]. Playing with vibrato and in a slow tempo resulted in no detection of 100ms of latency. The authors describe that younger test subjects were able to detect latency better than older ones. It was also found that test subjects made use of kinesthetic memory while playing music. They learned the position where the hand should be placed for playing the theremin (the instrument used). In the listening test, latency values between 120 to 240ms were achieved. Beyond 240ms test subjects were unable to cope with latency and were unable to compensate with tactile memory. On the contrary, as stated before, other researchers, such as Carôt [5], support the idea that is not possible to define a common valid latency threshold based on listening experiments with drums, bass and saxophone.

In research on digital instruments a “changing responsiveness” of the instrument was detected after introducing latency to a digital instrument, as found by Jack et al [127]. In other words, musicians felt that their strikes on the digital instrument were not enough, therefore the dynamic level was increased. To cope with latency musicians “put more effort” into performing the instrument. The behaviour of the musician towards the performance of the digital instrument varies depending on latency levels.

Based on telematic performance experiments simulating round-trip latency conditions (RTT), Smith [209] categorized five ranges where different musical actions are framed. The listening experiment involved a wide range of music instruments (e.g. alto saxophone, violin, keyboard, voice) in collaborative performances. These five ranges are:

1. < 20ms: Imperceptible delay. Musical performance is suitable.
2. 20ms < 120ms: Audibly asynchronous range. Musicians slow the tempo down. Music performance is challenging and difficult. Music which is not subject to tight rhythm patterns could be performed.
3. 120ms < 400ms: The delay is an eighth or sixteenth note and is heard as part of the beat. Music with this musical note pattern cannot be played.

4. 400ms < 600ms: At this range, delay may become part of the rhythm and is easy to perceive. Music that has a slow tempo may become out of sync.

5. > 600ms: Synchronization becomes loose very easily. Rhythmic music cannot be performed unless a metronome or any other (mechanical or electronic) assistance is included in the set-up.

A very important conclusion of this work is that the mental conception formed by musicians in connection with the degree of latency at the beginning of a performance is difficult to change [209]. Similarly, Brochier [33] states that initial tempo and auditory delay strongly affects tempo changes throughout a performance. In normal network set-ups, 30ms latency is acceptable for the performance of music. This is about the same as having a physical separation of 30ft (approx. 9.14m). A broader approach estimates that musicians are able to tolerate a latency up to 75ms, which is a very high figure for the latency budget [51]. This corresponds approximately to the maximum distance between the conductor and the last musician of a very large orchestra [145]. Rhythmic musical interaction becomes very difficult without the conductor when the distance between musicians is more than 8.5m [47].

Another conclusion of the research on musicians formulated by Lester and Boley [152] suggests that no magnitude is considered by everyone as a threshold. Different test subject’s responses for varying degrees of latency follow a Gaussian distribution. This Gaussian distribution remains similar for a given musical instrument regardless of the test subject.

The inverse linearity function delay in milliseconds vs. tempo in beats per minute for hand-clapping experiments is described in the work of Farner [90] and was confirmed for musical instruments by Barbosa [18]. With delays up to 68ms, the tempo decreased as the delay increased and the so-called “Chafe effect” was observed with delays of less than 15 to 23ms. This effect describes how short delays, < 11.5ms produced a modest but surprising acceleration, while longer delays produced tempo deceleration [56].

Further analysis from the work of Chafe et al. [56] presented in [113] introduces a mathematical model describing the relationship between tempo and latency. The model assumes...
that the tempo $M$ decreases according to the following equation.

\[
M(n) = \frac{60}{T_0 + nd}
\]

Equation 2.5: Tempo model

where:

- $T_0 = 60/M_0$ is the starting period in seconds.
- $n$ is the quarter-note number
- $d$ is the delay time

A delay $d > 0$ produces deceleration. In this model, where $d=0$, a normal steady tempo is achieved [113].

Another mathematical approach modelling people playing in delayed performance scenarios is presented by Driessen et al. [83]. Fitting data and using coupled oscillators theories, the model approximated human player behaviour with a mathematical expression.

\[
M = M_0 - kM_0d
\]

Equation 2.6: Human player behaviour

Data fitting gave the constant $k$ the approximate value of $k \sim 58$. Driessen et al. [83] simulated NMP set-ups. The results obtained may be comparable to performance situations in acoustic environments where the distance between performers is enough to produce delays comparable to those achieved in real networked music performances [104].

For better performances regarding tempo stability, latency between 10 and 20ms is preferable to no latency. This assumption stated by Schuett [205], was also confirmed by Chafe [56]. In the range 10 to 20ms, the performances were stable and tempo changes such as slowing down or accelerating were measured. Musicians do not play in precise synchrony.
There is always a rhythmical inaccuracy which Carôt and Werner [5] called inter pulse delay (IPD). Some musicians accept higher IPDs than others. However, this ability has a direct relationship with note resolution.

Some latency may be preferable to none when playing with a non-delayed metronome or with another non-delayed musician [152]. Especially, for in-ear monitoring (IEM), a latency similar to the free air distance from the instrument to the ear may be preferable. This result also fits the approach formulated by Schuett [205] which suggested that a 4 to 20ms latency would translate to a distance of 4 to 20ft. This distance is a normal separation between musicians during a performance. Acoustically isolated performers playing with headphones would benefit, and therefore perform better, with an added delay of between 5 to 20ms.

The ability to cope with latency in collaborative performances may vary depending on the music and its characteristics, e.g. style, tempo, attack of the instruments, etc. [205]. In addition to the work of Schuett, similar conclusions were also reached by Kleimola [141]. Latency tolerance varies depending on instrument choice. Musical styles requiring strict time synchronicity may be more tolerant to latency compared to styles where rubato is allowed. Moreover, the playing style influences the ability to cope with latency. From research done on a theremin [158], latency was found to be more difficult to notice when playing vibrato.

It is important to stress that some of the above conclusions were obtained through listening experiments where no music or musical instruments were performed, only rhythm patterns played by clapping hands. Smith [209] notes that some of the conclusions from the works of Schuett [205] and Chafe [56] could lead to incorrect assumptions. As one of the systematic problems, Smith [209] found that test subjects were “anyone who could clap”, and it is to be expected that the majority of the population has no capacity or training to maintain a steady rhythm. There is also no plausible explanation in these works [205, 56] for the assumption that under delays of less than 10ms musicians tend to speed up. Furthermore, some results vary between the different studies. One example is the delay threshold, where the tempo at 8ms is stable for Chafe [54], while for Farmer [90] this value is 15ms and for Brochier [33] it is 6.5ms. However, Chafe is aware that there is a mismatch between his “edge of playability” and the higher delay times that are considered
normal for musicians to cope with in musical performances. He pointed to a difference in tasks as the reason for this difference.

In addition to latency numbers, research by Mäki-Patola [158] suggests that the ability to cope with latency depends on three main aspects:

1. The music style performed.
2. The musical instrument used for the performance.
3. The tactile feedback, which is either existent or not existent (e.g. virtual instruments).

For instance, pianists could perform without having aural feedback from their instrument [156]. With added latency pianists cannot rely on their “inner prediction model”. The instrument has to be relearned during the performance. Research carried out on pianists performing three movements of Poulenc’s “Sonata for Piano Four-Hands” [64] states that musicians should share a common point of view in collaborative performances. These common goals are even more important than having immediate feedback from the player’s own actions. The auditory latency threshold was found to be between 50 and 75ms. The figure of 50ms was the upper limit for the acceptable delay on fast movements and 75ms for slow movements. When both signals were delayed, the own performance and the other performer’s signal, musicians were able to tolerate delays of up to 65ms for the fast movements. In addition, as a general rule, it has been observed by many researchers that the greater the musician’s skill, the more stable their musical performance and coping with latency may not be such an issue [42].

Another interesting result from the research involving clapping experiments is the important role of reverberation. The findings of Schuett [205] showed that the ability to cope with a delay in ensemble performances is better in the presence of a natural reverb rather than a digitally manipulated reverb. Drummers could play with stable tempo up to distances of 100ft or 100ms delay. However, a synthesized reverb did not provide any improvement. In another study, artificial reverberation as reported by [5] does not result in improvement by musicians. Players even prefer to perform with an unmodified signal. Farner [90] also found that results in virtual anechoic conditions were not as precise with respect to timing as in reverberant conditions. Farner assumes in his work that reverberation enables ensemble playing. This could be explained [56] by the masking of sharp-edged
signals due to reverberation. Auditory masking occurs when the perception of one sound is affected by the presence of another [169].

The negative effects of latency for vocalists can be ameliorated through room information [159], especially for IEM. In accordance with this assumption, Lester and Boley [152] encountered similar results (even for vocalists) in their research.

**Anticipation**

Opera singers behind the stage are not seen by the conductor. Therefore, singers must anticipate their entrance onto the stage in order to synchronize with the other singers performing on stage. It was the same situation during the Renaissance with the *cori spezzati* of the Basilica di San Marco [145]. In order to achieve synchronization and anticipation, the music being performed plays a very important role. To begin a musical action slightly ahead of the sounding time is a very common task for musicians such as drummers or conductors.

It is possible to learn to play while compensating for latency [67]. Delay compensation up to more than one second is also known in organ playing. This ability may be linked to a strong musical background and also to performer dexterity [198].

Synchronization deviates from the stimulus onset. The results of tapping research by London [154] show that humans tend to tap 20 to 60ms ahead of a metronome click in experiments in which people had to tap in synchrony with the metronome. In research by Aschersleben [12], this figure is around 20 to 80ms for the tap preceding the click. Humans are very sensitive to the synchronization of audio and tactile experiences and the accuracy is around 40ms [8]. Further studies on the implication of these figures for the design of haptic interfaces suggest values between 10 to 30ms [2].

Research involving not only tactile but also visual information done in the 1970s by Posner et al. [191] came to the conclusion that the auditory system is faster than the visual. The term used in Altison’s research to describe the visual system was “sluggish” [8].

It is relevant for the design of multimedia applications to prevent desynchronization between the different senses involved, and thus avoid compromising realism [9]. Moreover,
analysis of brain function shows that interaction and integration between the different sensory modalities is more common than expected [206]. Synesthesia may be the best-known example to describe this modularity in contra-position to a sensory uni-modal approach.

The effect of anticipation has also been also documented by Chafe [54] in the hand-clapping experiments. An intrinsic tendency to anticipate when the delay is present was observed. This tendency may be part of the human rhythmic production. Due to the physical impossibility to react instantaneously, the motor commands should be anticipated in order to perform on time. Humans are able to calculate the time an action needs, even in changing situations [156]. Research in bowing and fingering done by Baader et al. [16] concluded that tone sequences include serial concatenations of finger movements and as well as anticipatory behaviours. A variation in the range of 30 to 60ms was observed from perfect synchronization between bow reversal and lifting of a finger to initiate a tone. The brain generates motor commands with enough time to anticipate an action due to physical activity, e.g. playing an instrument. These relationships between anticipation and action are fine-tuned [231]. Playing from the score is not immediate and there is an anticipatory reading where the performer sees the notes about to be played next [218]. The latency gap between the onset of a triggered note and its audible feedback is the delayed onset present in many instruments [172]. For musicians, a normal part of their musical performance is to consider this gap and unconsciously play notes in advance [42].

Scientific research in consciousness has shown through experimentation that there is a lag time of a third of a second between the input of the real world and its conscious perception of the human brain [79]. However, there are compensatory mechanisms for this perception delay. One of these is relying on an unconscious “auto-pilot” which may be a possible explanation of the lower tempo variations when performing collaboratively with latencies above 100ms [198]. The other mechanism is anticipation. Humans look ahead to events in the outside world and previous learning plays a decisive role in this capacity [79].

The simultaneity of perception of audio and video stimuli

Literature from the 20th century describes a visual dominance when it comes to perception. In his work of 1933, “Adaption after-effect and contrast in the perception of curved lines”, Gibson outlined how the visual input dominates perception. Further research in 1976 by Posner et al. [191] showed a relevant role of the visual sense. However, it noted
the difficulty switching between modalities (visual, audio or kinesthetic) when the attention is focused on one of those modalities.

Nowadays, scientific research advocates a more holistic view [148]. The term used is “multisensory integration”. Two rival theories are open to discussion in the scientific world, either information in one sensory modality affects perception in another sensory modality or there is just a simple interaction of sensory inputs [204]. However, experiments have shown that spatial processing is mainly dominated by vision and temporal processing relies mainly on audio [114]. The key to producing a congruent unitary perception is an interaction between the different senses.

The sensitivity of humans to the tolerance of delays is higher for delayed audio signals than for delayed video signals (video leading) [8] assuming an equal duration of the stimuli [146]. In other words, two short stimuli (short signals under 100ms) with the same duration are perceived in subjective synchrony, even when the physical onset of the video is presented ahead of the audio, i.e. a higher delay tolerance for audio signals [9].

The perceived subjective maximum is known as the point of subjective simultaneity (PSS) [9]. Furthermore, the increase in the duration of one of the signals, either audio or video, produces a shift related to the relative apparent onset [146]. In summary, the perceived synchrony shift is either in the positive direction for longer video durations and in the expected negative direction for longer audio durations.

Audio-visual delays are far easier to detect when the audio is in advance and the video slightly delayed. There is a synchrony offset with a range of 30 to 50ms. In this range, both audio and video are perceived as simultaneous. A possible reason for this asynchrony may be related to the cell’s response to audio and visual stimuli in the brain [143].

In short, the different sensory modalities such as auditory, visual and tactile may have a strong interaction. Evidence for this is available in the research of Kuling [146] on auditory-visual asynchronies and Altinsoy [9] on auditory-tactile information. Shimojo and Shams [206] advocate for an interaction between the different sensory modalities in contrast to a “naive modularity” point of view. Auditory and visual perceptions are not to be seen as independent processes. Moreover, stimuli affecting one sensory modality can impair the reaction in another sensory modality [143].
Musical events onset and disruption

The beginning of a musical event is defined by Wright [231] as its onset. As stated before, there are both physical and perceptual onsets [225]. These onsets vary in time depending on the characteristics of the sounds produced by music instruments [194] and are relevant for the perceptual characteristics of a sound [146]. The physical onset is the acoustic beginning of an event. The perceptual onset is when the listener can hear that an auditory event has begun. This classification is meaningful only for events perceived as discrete in time [231].

The disruption or breakdown time for musicians playing in non-collaborative performances is an essential topic for this research. With increasing delays, musicians have to cope with latency which translates into a temporal separation between actions produced and the feedback. A performance disruption is more likely to happen. Early research done by Gates [107] on an electronic organ found that the maximal disruption time is around 270ms, which is higher than most of the values found in speech research. Research on tactile feedback for digital percussion instruments [73] shows that breaking points could be found in the range 40 to 55ms.

A relevant research approach by Kobayasi [232] presented three separate time lags or latency ranges. Within the first range of 0 to 50ms of “time lag” between two performers, the statistical average and the standard deviation of the phase difference have low values. There is almost no difference between 0 and 50ms with regard to the effects on the musical performance. In other words, in this range, a performance between two musicians is feasible. The onset difference between performer one and two for the same musical event or their phase difference is small and therefore they play the same note at almost the same time. When perceiving latencies < 50ms, musicians feel like they are on the same stage [162]. Similar results were obtained in [62] where 50ms was defined as the commonly acceptable threshold where a performance is possible. When the range is > 60ms, the statistical average increases significantly, the standard deviation remains unchanged. In this range, the performer may adopt a leader-follower relationship [205]. One musician is the leader and sets the pace and the other follows. In the range of 80 to 90ms the standard deviation increases significantly. The musical performance of the ensemble is prone to breakdown [232].
There is a breaking point where the musical performance is disrupted. This performance breakdown is influenced by metronomic beat, musical styles and musical instruments [152]. Research results by Pfordresher [182] suggest that the disruption of a musical performance could be produced by altered feedback content, but that it depends on the structural similarity between the planned sequence and the sequence produced by altered auditory feedback events. Under optimal auditory conditions of performance, there is no disruption to a musical performance. Optimal conditions exist when planned and performed sequences match each other and the order of sequences also coincide [182]. In short, the musician listens to the music performed without modifications. In contrast, when planned and perceived events are temporally shifted relative to each other, performance disruption may occur. The absence of feedback normally has no effect on disruption. Previous results were obtained for experiments exclusively involving keyboards. Adjustable delays may easily result in a performance disruption in keyboard experiments [183]. A performance disruption is more probable when the auditory feedback matches the next event. In other words, the delay is the same as the interonset interval (IOI) [74].

**Strategies to cope with latency**

Keeping a stable tempo in a performance is achieved by many factors such as eye contact, trust, the singer’s breathing and the conductor’s instructions [176]. However, in performances over a network or with added latency, a new set of strategies are necessary to overcome the effects of latency and be able to perform with the ensemble.

It is important to remember that latency could even be used as an artistic approach and may be an intrinsic part of the musical work [49] if it was conceived from the beginning as “network music”. This kind of music would not work in a different setting [50].

It is well known that musicians use a coping strategy to maintain a solid tempo for latencies between 50 and 70ms [205]. Schuett describes it as a “leader-follower relationship”. However, this strategy produces a decrease in synchrony on the leader’s side of the musical performance and from a musical point of view is not a true ensemble performance. The same musical strategy is described in the work of Carót and Werner [46], as the master-slave approach (MSA). Another strategy described by Schuett [205] is the compensation of the delay by synchronizing the performer’s own pattern with the sounding result of the other performer’s pattern. The result is a slowdown of the overall tempo with each
iteration of the pattern. In the presence of normal auditory feedback, the leader always adapts to the changes of the follower. This timing adaptation is not just a musical one but an essential temporal coordination inherent to human beings [177].

To solve the problem of latency in collaborative performances, Cáceres [38] describes an approach where the delay is added to one of the performers. From the performer’s point of view, it is like performing on top of a recording. This approach could be annoying and difficult to achieve for the musician. Voice leading is also a method [217] which has been observed in piano performance research. One note is played slightly before the others. In the research done by Bartlette et al. [21], the clarinetists played ahead of the other musicians’ sound while using headphones and maintaining a constant tempo. The strategy of the string players was quite different. The violist played in time with the violinist’s headphone sound.

Strategies to cope with latency have been observed and researched for a long time [107]. The ability to cope with latency may be individual and varies from subject to subject. In research on violinists and percussionists the results differed. Violinists were able to admit more latency than percussionists. A possible explanation is musical training. The percussionist in this experiment had learned to keep tempo in small ensembles. He was more affected by delays. In comparison to the percussionist, the violinist had more freedom [73].

In addition to some of the previously mentioned strategies, there are also more radical procedures to cope with delay, such as ignoring the delay or confusing sounds and concentrating primarily on the internal tempo representation [73]. It has been observed that increases in loudness and the addition of heavy accents help to overcome the effects of latency [127]. A possible explanation may be found in Goldstone [110]. Psycho-acoustical research found that for tones shorter than 200ms, sound pressure levels had to be increased to match the loudness of a longer tone [69]. More intense sounds are perceived to be longer than less intense auditory sounds. This assumption also applies to visual stimuli. Time duration is a human construct [101]. On the other hand, the louder the instrument’s direct noise due to higher latency, the more difficult it is to perform the instrument and to cope with latency [5].
A categorization of musical interaction styles and a solution to overcome latency problems in network music performances, especially for collaborative performances, is presented in the Table 2.7 which summarizes the work of Carôt and Werner [46, 47, 42].

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Latency range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Realistic Interaction Approach (RIA)</td>
<td>delay ≤ EDAL</td>
<td>Comparable to musicians playing in the same room with up to 8m distance between them. A stable one-way latency below the EDAL is assumed [42].</td>
</tr>
<tr>
<td>B1</td>
<td>Master/Slave Approach (MSA)</td>
<td>delay &gt; EDAL</td>
<td>One musician is the master and keeps the rhythm (normally drums or bass) without listening to the delay. The others follow.</td>
</tr>
<tr>
<td>B2</td>
<td>Laid Back Approach (LBA)</td>
<td>max 80ms &gt; delay &gt; EDAL</td>
<td>Based on the “laid back” musical approach. It means playing slightly behind the groove. Similar to (MSA).</td>
</tr>
<tr>
<td>B3</td>
<td>Delayed Feedback Approach (DFA)</td>
<td>delay &gt; EDAL (SDF) or delay &gt;&gt; EDAL (DDF)</td>
<td>Musician’s own signal is delayed artificially (SDF). The best synchronization is achieved when the self-delay is equal to the latency round-trip-time (RTT). A variation is foreseen when both performers are self delayed (DDF). However, the larger the delay, the worse the performing conditions.</td>
</tr>
<tr>
<td>B4</td>
<td>Fake Time Approach (FTA)</td>
<td>variable</td>
<td>Musicians play asynchronously to other musicians delayed exactly one measure.</td>
</tr>
<tr>
<td>C1</td>
<td>Latency Accepting Approach (LAA)</td>
<td>delay &gt;&gt; EDAL</td>
<td>Not for conventional musical styles, the delay is used as an artistic means of expression.</td>
</tr>
<tr>
<td>C2</td>
<td>Remote Recording Approach (RRA)</td>
<td></td>
<td>Remote musicians participate in recording sessions. The Internet is used as a connection, replacing traditional data transfer mechanisms such as CD or tapes. It is comparable to DFA where the musician plays the slave role and the playback/recording device is the master.</td>
</tr>
</tbody>
</table>

Table 2.7: Strategies to cope with latency in collaborative performances adapted from [42]
Literature Review

In previous versions of his work Carôt assumed a minimal delay threshold of 25ms. This assumption has been replaced in this table with the concept of EDAL, which is not a known number and has to be estimated by means of a test [47]. Under 25ms it is not necessary to utilise any strategy. This latency figure is not perceivable for most musicians [46] and is also not disturbing [203].

The feasibility of a network music performance is determined by the degree of latency, particularly in the real interaction approach (RIA), which is the most technically challenging of all the strategies presented [104]. It requires the lowest latency.

2.4.3 Delayed Auditory Feedback (DAF)

Auditory feedback is defined by Pfordresher [182] as the sound created through actions involved in sequential behaviour such as singing, playing an instrument or speaking. In the last decades, a lot of research on feedback alteration has been done under the concept of altered auditory feedback (AAF) [183]. Perceptual feedback includes audio, visual or even haptic information [218]. The role of auditory feedback is far more relevant for performance disruption and performance errors than the role of the visual feedback. However, recent research has shown that visual feedback also has an important effect on timing variability [147].

The introduction of a delay $\Delta t$ to the musician between the onset of a played note and the auditory feedback is defined by Dahl [74] as DAF. On the basis of this definition, the implications of DAF for our research will be explored. During a musical performance, the effect of DAF can be compared to the performance of music in a large echoic room like a cathedral [218].

Musicians are particullary sensitive to temporal relationships between auditory feedback and actions towards the production of a melody [192]. The majority of publications are related to the effects of alteration in timing (synchrony) of perception and action, which is defined as serially shifted delayed auditory feedback (DAF) and is a specific type of altered auditory feedback [186]. On the other hand, there is also research in DAF on feedback pitch alteration [188] which involves other factors and is not relevant to the present work.

DAF research has existed since approximately 1950 and in the beginning was dedicated exclusively to speech. In his work, “Effects of delayed speech feedback”, Bernard Lee delayed
oral speech signals by up to 300ms and documented the effects produced in individuals. Since then some of the results are known under the name of “Lee-effect”. Above 200ms delay, the effect is disruptive and speech is no longer possible. The delay of 200ms may coincide with the duration of a syllable [216]. There are some common features between music production and spoken language. These are the production of ordered sequences with hierarchical structures, the requirement for precise sequence timing and the important role of the auditory modality. Moreover, there are also features similar to those found in motor learning, such as the movement of hands and fingers [95]. Therefore, the effects of DAF are not just related to speech but also affect rhythmic production and musical performance.

Recent research has confirmed another important effect of DAF on speech, namely that stuttering decreases and fluency increases. Similar experiments on pianists with musician’s dystonia\(^8\) have shown that the presence of AAF disrupted the performance of healthy and unhealthy pianists, i.e. those suffering from “musician’s dystonia”. There was no improvement observed with DAF for pianists [61]. In DAF speech experiments, test subjects need more time to speak and raise their voice level. These effects show the importance of auditory feedback in speech production [171].

The presence of auditory feedback is not necessary for an effective musical performance. Research done by Finney [94] has shown that musical performances on a keyboard were not compromised when auditory feedback was absent. Measurements done in the keyboard experiment were not different between normal feedback and no-feedback conditions. However, it was noted that when the auditory feedback is present, it should match the sequence of executed actions in the performance.

When performing the piano, dynamics and articulation can be imagined by the musician while playing the music. Musicians may have inner representations of the sound they want to produce and they are also able to adapt themselves to environmental conditions [203]. The absence of auditory feedback can be also compensated for. This effect seems to be stronger in musicians with higher expertise as found in musical imagery experiments [25].

\(^8\) A degradation of voluntary control of highly skilled movement patterns [61].
Finney [96] published his dissertation titled “Disruptive Effect of Delayed Auditory Feedback on Motor Sequencing”. He found that there is not such a maximal delay value of impairment. Experiments with tapping fingers showed that above the value of 200ms the effect of DAF still increases. Further research by Pfardresher [187] confirms the non-existence of a specific value. Alterations of DAF in the order of about 200ms, could induce errors such as slowing the timing or even disrupting the performance [218]. Some studies point to a dominance of the auditory modality over the visual modality relative to the encoding of temporal information, which is used to time actions [147]. When performers learn in the presence of auditory feedback, musicians’ memory for music seems to improve [95].

Solo performers rely on the coupled auditory and motor information from the musician’s feedback. Both the motor information and particularly auditory information affects timing. An ensemble performance yields auditory information in the presence or absence of the musician’s own motor movements. This could be the reason why professional musicians who normally perform in ensembles may or may not be annoyed by motor or auditory information intervening with their performance [177].

Disruption in DAF exists due to a mismatch between perception-action coordination at a specific timescale [183]. Sequencing and timing are not a unit. In addition to the alteration of timing, some research has focused on pitch alteration. Finney [94] showed that pitch alterations could be ignored to some extent, in contrast to timing manipulation, which is in some way specific to DAF and leads to a disrupted performance. In contrast, research done by Pfardresher [181], which has been reconfirmed [182], has shown that disruption is also possible when pitches of previously planned events are repeated as in serial shifts but not by absent pitches or those randomly presented. In serial shifts, pitch events that should match past or future sequences are presented in synchrony with the key pressed in the actual moment [182]. These shifts are similar to the serial ordering errors that pianists make in normal performance conditions [185].

Pfordresher [183] states that disruption suggests not only a passive response of the system for the duality between planning and perception of events, but could be a more active response to counteract the influence of auditory feedback. Auditory feedback for speech [171] and music performance is used as an additional source of information to establish the accuracy of sequence production [185].
Effects of DAF

Research results have shown that human performers can cope well with delays up to some milliseconds in the order of 35ms [113]. Research has also found [116, 106, 94, 168, 218] that delays, particularly longer delays, produce the following effects during musical performances:

- Distraction.
- An increase in dynamics. Playing with heavy accents.
- Slower tempo.
- Asynchrony between left and right hand (for keyboard instruments).
- The number of musical errors in the performance increases significantly.
- A conflict between expectancies and auditory imagery.

These effects are very similar to those found on research done exclusively in speech. Rhythmic tapping and speech interference has been observed from early delayed feedback experiments. To overcome these adverse effects, test subjects raised the intensity of the speech or repeated syllables. Nevertheless, the sensitivity to delay depended primarily on the test subject [150]. Havlicek [116] continued the research initiated by Lee and concentrated on DAF and its disruptive properties on musical performances. In his work, he also describes that test subjects tried to ignore the DAF and concentrate on rhythm, keeping a constant tempo. In these early experiments, the use of strategies to cope with latency was also documented. Musicians slowed down or even accelerated the tempo and the number of articulation errors also increased. Some of the test subjects relied on tactile and proprioceptive feedback, which is always present regardless of DAF. This kind of feedback is enabled by proprioceptors, which are nerve endings in muscles and other organs. These nerve endings respond to stimuli of position and movement of the body with respect to gravity [74].

In recent research on DAF with musical instruments, mainly on MIDI keyboards, the group of Louhivouori et al. [168] encountered four basic types of errors. It is important to note that a large amount of research has been carried out with keyboard instruments.
Electronic keyboards allow the analysis of MIDI data. Personal computers and MIDI enable better empirical research especially in rhythm through the use of a wide number of methods [120]. Table 2.8 summarizes these findings.

<table>
<thead>
<tr>
<th>Number</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Order errors</td>
<td>Two notes played in the wrong order</td>
</tr>
<tr>
<td>2</td>
<td>Deletions</td>
<td>Omission of one or more notes</td>
</tr>
<tr>
<td>3</td>
<td>Wrong notes</td>
<td>The replacement of the notated tone by another</td>
</tr>
<tr>
<td>4</td>
<td>Insertions</td>
<td>Playing notes not previously notated</td>
</tr>
</tbody>
</table>

Table 2.8: Types of errors produced with DAF in music performance [168]

The musical instrument regulates the combination of perception and action. This process is defined by Louhivuori et al. [168] as a circular action-reaction process that can be conceived from an “embodied perspective”. With DAF, the adaptation of the action performed, the judgement of the player’s action and the generated sound are distorted.

The effect of DAF could be also explained through “mirror neurons” which may play an important role in the mental simulation of action. This unorthodox approach states that presented sequences differing only in time, displaced due to auditory feedback, may activate mirror neurons at the wrong time, producing undesired effects [218]. Research has also established that in primates and some birds, mirror neurons work (the term in neuroscience is “fire”) when performing an action or also when seeing someone performing that action. It is presumed that the function of mirror neurons is the preparation and training of movements that have never been executed by the organism [153].

Research on drums showed that DAF in the order of 50ms is not an issue when playing this musical instrument according to Dahl [74]. Musicians are able to play and perform

---

9 Named after the work “The mirror-neuron system” of Rizzolatti and Craighero in 2004. They researched the neurons that regulate the motor behaviour in monkeys.
music together with DAF. Larger delays produced errors and difficulties for the musical performance. Above 120ms, players relied on tactile feedback of the stroke. These effects were also observed in previous research by Havlicek [116]. Musicians increased the striking force, which is the same as increasing the dynamic. For some experiments a metronome was used. The use of a metronome delivered better results in comparison with the decrease in tempo observed in musicians without a metronome. In other experiments with the metronome [73], test subjects gradually matched the delayed auditory feedback of the metronome sound. By slowly introducing delays, musicians adjusted their performance without noticing the delay changes until the delay was annoying and noticeable.

Another relevant conclusion from DAF research is that, at least in speech research, age is a variable that may have an influence on the level of disruption [171]. Gender is another variable. However, Gates [106] found no statistical differences between the performances of males and females in his research on DAF. Lateralized DAF conditions, i.e. hearing a delayed auditory feedback right or left, have no measurable significance on the outcomes.

DAF, as mentioned before, is not the only type of delayed feedback that could alter perception. Another approach used in research is the delayed visual feedback (DVF). Based on the auditory dominance hypothesis [147], the effect of the visual delay is not as dominant in performances when auditory feedback is present. The effects of DAF and DVF are additive [147].

In tasks where no musical knowledge was expected, there were no differences in the response to alterations in feedback content between pianists and non-pianists as confirmed by Pfordresher [183]. This result points towards a general tendency: in humans, perceptual information and action plans are related. Nevertheless, musical knowledge [182] provides a better dexterity with regard to timing variability, reflecting a better temporal precision in motor control. Adaptation and the increase in skills to cope with DAF due to practice are possible and could explain why organists can perform well, regardless of a high DAF [183]. The DAF effect on organ performance is the product of the instrument mechanics and acoustics, normally in churches [168], in contrast to the majority of musical instruments that are performed having a more direct auditory response [74].
2.4.4 Perceptual Attack Time (PAT)

Wright has developed a theory where the perceptual attack time (PAT) of sounds are not conceived as a single instant but as a continuous probability density function.

Wright defines PAT as:

“A musical event’s Perceptual Attack Time ("PAT") is its perceived moment of rhythmic placement. Note that this is a subjective, perceptual parameter.” [231, p. iv]

It is important to emphasise that PAT is subjective and therefore inherently relative and depends on the perception of the listener. Different sounds, e.g. percussive, bowed, etc., produce different perceptions, regardless of the measurement method. Results will always vary slightly [231].

There is a lag between the onset of a musical note and its PAT and the brain learns to compensate for this time. This could be another reason why organists can learn to play in the presence of delays of hundreds or thousands of milliseconds. The musical instrument plays an important role. Wright [231] even thinks that skilled musicians may be able to control the shape of the probability density function (PDF) of PAT. Controlling the PDF-PAT shape can be very difficult for percussive instruments because their PAT and physical onset are very close to each other, but it may be not as difficult for voice and continuously bowed or blown instruments.

The influence of PAT in network music performances is also described in the work of Barbosa et al. [20]. The human auditory system focuses on those instruments with an impulsive attack time. On the other hand, instruments with slow attack times are not immediately perceived. Using short attack instruments in a performance may lead to better synchronism when latency is present. The musical instrument choice may become an important one regarding latency issues [199].

2.4.5 Latency Adaptive Tempo (LAT)

The concept of latency adaptive tempo (LAT), which defines the direct relationship between musical tempo and latency tolerance, was introduced by Barbosa [18]. Furthermore,
it defines the disrupting effect of latency. In one listening test, as a by-product of his dissertation [18], Barbosa simulated network latency conditions for the performances of four musicians playing jazz standards with bass, percussion, piano and guitar.

Figure 2.4 shows the results of the maximum individual latency tolerance applied to the auditory feedback from the player’s own instrument as performed by each musician [18]. Regardless of methodological issues, the results obtained shows the LAT of four different instruments. Latency in the work of Barbosa is simulated network latency.

As already noted, the LAT was a by-product of the dissertation and contains some methodological issues. Performances were synchronised by means of a metronome with tempos between 80 to 220 BPM, without specifying the technical characteristics of the metronome and the measurements. Variable delays, probably digital, between 50 to 250ms were introduced via headphones for each musician. From the graphic, an inverse relationship for latency tolerance vs. musical tempo can be established [18].

In other works, it has also been observed that the ability to cope with delay is directly related to speed and note resolution of the performed pattern [42, 113, 152]. A faster
Literature Review

tempo in BPM combined with higher note resolutions diminish the ability of musicians to cope with latency when playing in an ensemble [5]. Although the direct relationship between delay acceptance and musical tempo is confirmed by other researchers, it has also been shown that below a musical tempo of 100 BPM the relationship may not be as significant [5]. Konstantas et al. [145] estimates in his research that a round-trip delay (RTT) should not exceed 85% of the duration of a sixteenth note. In other words, having a delay of 90ms, the musical tempo should not exceed 142ms per quarter notes. Classical music is more appropriate for such requirements than contemporary music because of its stable rhythm structures.

2.4.6 Measuring the Ability to Cope with Latency

The issue of latency while performing has been approached in the literature qualitatively and quantitatively. Researchers including Rottondi [198], Gurevich [113] but especially Carôt [5, 42] used the expression “to cope with latency” to indicate that musicians were able to perform collaboratively up to a certain moment in time while listening to delayed audio signals, beyond this moment musical interaction was impossible. Different tempi and delay values were introduced during the performance until one of the musicians felt uncomfortable or musicians tried to slow down the performance. A similar subjective approach was presented by Boley et al. [152]. Here, the issue of latency and the performance of music, in particular the question of sensitivity to latency was answered with a subjective listening test: Subjects rated different signals and wrote down the result in a scale ranging from excellent to horrible or a numerical scale from 100 to 0. In this case, horrible means that the musician found it very difficult to work under these conditions. Absolute thresholds are not defined.

The research of Bartlette et al. [21] investigates the effects of different latency levels with regard to coordination, pace and timing regularity. The latency levels were predefined and were not changed during the performance. Inter-onset intervals (IOI) were measured and compared establishing differences for pairs of musicians (clarinetists and string players). This approach intended to find a limit for the level of external latency that allowed a musical performance. For Barbosa [18] the issue was to determine the individual latency tolerance. Four different musicians (bass, percussion, piano and guitar) performed together, latency was introduced until a synchronous performance as an ensemble was not
possible at all.

For Drioli [84] the main factors that force musicians to stop performing are:

- Loss of musical coordination.
- Slowing down the tempo.

The metric used in previous approaches is based on predefined latency values and subjective answers. No research has expected to find an exact value, only thresholds where musical interplay is possible [42, 113]. According to Barbosa [18], the musical tempo is a very important variable regarding the latency issue. In addition, other factors such as the musical background and dexterity [198], the music style performed and the musical instrument as suggested by Mäki-Patola [158] may contribute to the musician’s ability to cope with latency.

Based on the results of previous research it may be possible to expand the concept of “cope with latency” used for qualitative approaches and define a latency tolerance range concept which can be related to a quantitative measurement. To measure the latency tolerance range, it is necessary to find a more accurate definition of the concept of “cope with latency”. In order to enable comparisons between different musical instruments, the definition has to be the same for every musical instrument.

2.5 Summary

In this chapter, an overview of the current literature was presented including the methodology and sources used to gather the relevant information. The principal features of the western orchestra musical instruments were summarized. Their classification regarding sound generation methods such as striking, plucking, bowing, etc. was introduced. An understanding of this is relevant to the statistical analysis in later chapters.

The different effects produced by latency when performing a musical instrument have been the subject of research over a long period of time. In the 1960s the question of a possible relationship between the different musical instruments and the delayed auditory feedback was asked for the first time [116]. Nevertheless, there are few publications where
systematic and replicable research has been done towards establishing a relationship between latency and the performance of musical instruments [18, 156, 74].

Technical concepts and explanations involving latency in networks are presented and explained in relation to the performance of music, as well as additional information on theories of human perception of latency. Current research on delayed auditory feedback (DAF), perceptual attack time (PAT) and latency adaptive tempo (LAT) and their relevance to the relationships between musical instruments and the latency tolerated by musicians are discussed and their approaches explained. In particular the concept of “cope with latency” has been discussed regarding its prior use in the literature. Metrics defining the limits of playability in music ensembles have also been defined. The Ensemble Threshold Performance (EPT) from the work of Schuett [205] and ensemble delay accepted limit (EDAL) by Carôt [5] are the most substantial quantitative approaches towards a better understanding of the musician’s ability to cope with latency when performing collaboratively.

Possible explanations for the variability of results and the difficulties relating to the issues of latency and perception in humans [224] were addressed. Current research has focused on the question of DAF mainly for speech, although some authors have also been working with music. However, very few musical instruments are used in research, with the piano being the most analysed musical instrument, mainly in the form of MIDI keyboards, due to the relative ease of data collection.

The present work aims to clarify whether there is any relationship between groups of musical instruments and the ability to cope with latency. This information may be helpful for musicians, producers, software engineers and researchers. Every digital process implies latency and in the future it is likely that technology will only be digital, in particular in the field of music technology. Nowadays, the majority of concerts, studio productions, musical experiments and so-called emerging technologies such as virtual reality (VR), augmented reality (AR) and immersive sound, rely on digital technology.

To some extent qualitative research has shown that there might be a relationship between latency and the musical instrument performed, in particular for collaborative performances. However, it is not known if this influence has its origins in the physical characteristics of the instrument or if it is only the performer who plays a role in the ability
to cope with latency. Obviously, it is extremely difficult to separate the performer from
the instrument. A quantitative approach based on measurements may enable to explore
the relationship latency and musical instrument. This study develops a new quantitative
approach based on the research methodology and the experimental design outlined in the
following chapters.
Chapter 3

Research Methodology

To answer the research question it is necessary to chose a research method which can be used to explore the relationship between musical instruments and the ability to cope with latency.

Before selecting a strategy of inquiry, it is important to define the personal worldview. Humans are prone to develop worldviews based on mentors, research experience and discipline orientations [70], just to name a few. Due to my academic background and my beliefs, my philosophical worldview can be considered as postpositivist. Being postpositivist relates to a deterministic philosophy, where cause and effect are strongly connected [70].

A reason for adopting this worldview -in addition to the influence of mentors and research experience- is the belief in an objective reality that can be observed and measured.

Research can basically be done by using three different approaches: quantitative research, qualitative research and mixed-methods research [151]. Quantitative research has been developed since the late 19th century and consolidated throughout the 20th century [70]. Central to this methodology is the experiment. However, it is not the only quantitative research approach, non-experimental designs are also allowed [70]. Due to an elaborated experimental design it is possible to establish numerical relationships among variables, these relationships can be described and analysed statistically.
The most salient characteristics of this approach are [70]:

- Instrument based questions.
- Statistical analysis.
- Statistical interpretation.

Results can be validated numerically through an statistical analysis. Furthermore, the reductionistic approach and the analysis of data based on observations and measurements enable the formulation of hypotheses [70]. The information gathered is reliable and the findings can be generalized. The biggest challenge the quantitative approach has to deal with is having humans as test subjects, inasmuch as the environment differs from normal life situations [156] and is not natural. The control of the environment [151] is a relevant issue to take into account when using this methodology.

Qualitative research has been further developed during the last decades of the 20th century and also in recent years. This research is used broadly in the humanities [70]. Phenomena are described in a narrative fashion, major schemes are identified and the researcher intents, very often, not to simplify or reduce what is observed [151].

The most relevant characteristics of the qualitative approach are [70]:

- Open-ended questions.
- Analysis of text and images.
- Themes interpretation.

The qualitative approach is based on the narrative of the subjects [70] and is helpful to explore research questions and gain initial insights [151]. However, there is a high influence of personal point of views and beliefs, not only from the subjects but also from the researcher who mainly describes phenomena and experiences.

The mixed-methods approach combines characteristics of the quantitative as well as the qualitative approach [70]. Based on the integration of both methods, it is expected to gain better insight than using either a quantitative or a qualitative method [151].

Some of the most important characteristics of the mixed-methods approach are [70]:

81
• Open- and closed-ended questions.

• Different forms of data gathering.

• Text and statistical analysis.

Mixed-methods are a compromise and require more time and energy than working with either quantitative or qualitative methods[151]. The main drawback of this methodology besides the time issue, is the complexity and the resolutions of discrepancies of the results.

For my thesis the quantitative approach is the most suitable to work with. An empirical research, based on observations, delivers numerical results from the data collected. This data can be statistically analysed to support or reject the hypothesis [151]. The empirical approach, which is based on experiments and observations, intends to explain phenomena from the particular to the general [22]. The experimental approach is the method of choice for establishing physical relationships [30].

Experimental designs define test explanatory variables and measure the relationships between these variables through the analysis of data [140]. In contrast, a theoretical deductive approach might deliver assumptions difficult to verify. Although, it may be difficult to compare different results between musical instruments, the statistical analysis of the numerical results confers validity and is also common practice in cognitive sciences [42]. Basic principles of the experimental research are replication, randomization and planned grouping [117]. For this research, mainly replication and randomization play an important role.

The qualitative and the mixed-methods approaches are not considered for this research. Previous research done by Lester and Chew [152, 63] explored the relationship between latency and ensemble performance using a qualitative approach. Test subjects performed music while listening to latency and rated the experience or answered a questionnaire. The qualitative approach is not a good choice for this study as results and observations can be influenced by personal attitudes and opinions. Making generalisations may be difficult, comparisons between different test subjects are not possible. For this study, narrative is not a reliable measure.

Mixed-methods approaches are also used to gain insight with regard to the issue of latency as done by Jack et al. [127], in particular when the research question is related to sub-
jective perception and combined with the analysis of conditions. To answer the research question of the Chapter 1, a mixed-method approach would require a bigger logistical effort and should include text analysis which may not improve the outcome.

For the research presented in this work, the forms of data collection, analysis and interpretation used by qualitative and mixed-methods approaches are not suitable in order to answer the research question. Procedures such as interviews, open questions or the use of narrative or participant’s opinions to clarify phenomena may not be comparable across different musicians and therefore across different musical instruments. Furthermore, in this research not only the observation but also the measurement of phenomena is expected. This measurement could differ from personal opinions.

While empiricism has often been criticized because of its focus on confirmation of expected results [22], a quantitative approach -based on previous latency research studies- with a numerical description of variables has been found to deliver a very reliable method [70]. Research on latency has also used quantitative methods [55, 182] in a very rigorous manner using the same musical instruments (e.g. hand-clapping or keyboards). However, the methodology used in these studies, complicates the comparison between different musical instruments. In current research the numerical data enables the establishment of differences in the measurement of variables [48]. A ratio scale, as the standard scale for time measurements [88], is used in order to compare the ability to cope with higher or lower latency levels. Such a scale has the same properties as an interval scale and, in addition, also has an absolute 0 point which allows quantifications and comparisons [82]. For this study, the 0 point is 0ms or no latency at all. As stated in Chapter 2, every digital system has latency. The latency measured by the audio interfaces used in this research is presented in Appendix E. Before any delay is introduced to the musicians in the listening test there is an initial latency difference and its value is above 0ms.

In order to answer the research question, a listening test, as the core of the methodology, is defined in this chapter. The aim of the test is to verify the hypothesis, i.e. The role of the musical instrument is not relevant with respect to the ability to cope with latency. It is expected that this ability may depend to a great extent on the musical instrument on which the performance is executed. Instruments belonging to a specific group (e.g. chordophones, membranophones, aerophones and idiophones) may have similar intrinsic characteristics with respect to their ability to cope with latency.
A cause-effect diagram can facilitate the research design and establish the different interactions of the system components [31]. Figure 3.1 is a simplified cause-effect diagram presenting the different elements which should be taken into account in a solo performance where latency is an issue.

Figure 3.1: Simplified cause-effect diagram related to latency and based on the “model of a solo music performance situation”, adapted from [203].

Some elements are assigned specifically and grouped into categories such as perception feedback and environments. The role of the different elements relating to inflows (increments) and outflows (decrements) [31] is not specified in the diagram, only the possible relationships. The novelty of this approach as presented in Figure 3.1 is the consideration of the interaction between the different factors. In order to develop a new research design, this approach focuses on the interactions between the different factors.

The quantitative research method adopted in this study differs from previous works in the following aspects:

- Definition of variables.
- Control of the experiment with an holistic approach.
- Non-collaborative performances
3.1 Research Design

To test the hypothesis, quantitative research supported by an experiment in the form of a listening test is defined, developed and conducted. The examination of the evaluated data-set and variable relations, basically those between instrument groups and latency time, are part of the exploratory research. The empirical research is primarily confirmatory and is based on an a priori hypothesis. Through the control of variables that can be manipulated, it is expected that cues, such as differences or similarities between different instrument groups, will be found. These cues may support a possible relationship between latency and the musical instrument performed.

The experimental design is characterized by its temporal limitation, careful arrangement and adequate standardization [30]. In other words, the experimental methodology is suitable for researching the proposed relationship. In contrast to observatory studies, an experiment allows control of the different conditions [117]. With an experimental design it is possible to use statistical principles that allow the comparison of the experimental responses obtained from homogeneous experimental units or measurements [161]. Three characteristics summarize the advantages of experiments [140] and these influenced the decision to use this research design.

1. Disturbing variables can be eliminated through randomization.

2. Due to control over the introduction and variation of the predictor variables, relationships and possible causalities can be easily observed.

3. Experiments are flexible, efficient and enable a statistical manipulation.

In short, considering its feasibility, the experiment is the scientific method of choice [140]. Compared with other designs, such as quasi-experimental or ex-post facto research designs, a true experiment enables better control and also a better internal validity, thus increasing the chance of producing valid and consistent results [132].

Through randomization potential errors or experimental drawbacks may not disappear but these can be randomly distributed. In short, the bias effect is averaged over all levels of the variables of the experiment [161].
From the different four “true experimental designs” known, the “within-subjects design” is the one that best fits the requirements and nature of the proposed experiment [151]. In this design, there is one study group and no control group. Furthermore, the within-subjects design allows the repeated measurement of the study group. All participants receive the same treatment [132]. A specific attribute or condition is measured more than once at a different time for the same test subject [117]. This design is also known under the name “repeated measures design” [92]. The reasons for using this design are:

- All participants have to undergo every experimental condition. There is no control group [151, 92].

- There is economy of time in running the experiment. The participant is exposed to the treatment several times without long periods of time between every experiment [92].

- Random noise is reduced due to sensitivity. Since the test subjects are human, they differ from each other in many ways. However, the within-subjects design allows a better control. Each participant has to be evaluated with the same listening test. It is to be expected that differences between outcomes from the experiment are produced by the manipulation of the variables of the experiment and not because of differences in the subjects [92].

Every research methodology has its drawbacks [92]. Repeated measures designs are not an exception to the rule. The more salient threat is the “general practice effects” where participants become more proficient in their reactions and answers when exposed to the treatment [132]. Some of the disadvantages of using this method and the possible solutions for this specific case are:

- The participants are human therefore do not always behave in the same way. Even under the same conditions, variations in the answers of the participant are to be expected. On the other hand, the so-called “carry-over effect” may appear due to systematic variations, also known as “confounding” effects, i.e. bias due to practice or fatigue. This problem may be avoided with randomization of the order of presentation of the different conditions (for e.g. in the designed listening test or the BPM tempi values) or counterbalancing the order of the different conditions [92]. It is important to be aware that counterbalancing does not completely eliminate order effects, but helps to evaluate possible order effects where present [132].
Research Methodology

- Irreversibility of conditions. The repeated measures design can be used if there are no irreversible effects being in a specific condition. In other words, if a participant is tested in one condition this condition will not affect further tests [92]. This is not the case in this research. The effect of the listening test is not permanent.

The most relevant information relating to the research design method used in this study are as follows:

- A quantitative empirical research approach.

- The within-subjects measures method is the chosen experimental design.

Differentiation from other design methodologies

The question of dismissing other methodologies in order to choose a method to confirm the hypothesis arose at the beginning of the research design. For this research the answer is easy because of the characteristics of the proposed experiment. A phenomenological study, where the evaluation of perception is more subjective and qualitative oriented [151], may not be the best method to test the hypothesis.

Being a quantitative research and taking into account that variables may be manipulated, an experimental design is an approach that better fits these requirements. The different experimental designs can be summarized into three groups: experimental design, quasi-experimental design and ex-post facto design [151].

As stated before, an experimental design was chosen. A quasi-experimental design was not an alternative due to the absence of randomness both for the selection of group members and for the presentation of treatments to the groups [151]. In addition, the confounding variables are difficult to control or may not be controlled at all, meaning many interpretations are possible. An ex-post facto design was also discarded. In such a design, the manipulation of variables would be a danger for the participants [151]. In this research, the manipulation of the variables does not translate into any risk for the participants.
3.1.1 Methodology stages of the research design

Figure 3.2 summarizes the methodology milestones that need to be reached when accepting or rejecting the hypothesis. Intermediate steps are described in the next sections.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description</th>
<th>Expected outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts Selection</td>
<td>Definition of relevant musical terminology</td>
<td>Correct use of musical terms toward the design of control mechanisms for the experiment</td>
</tr>
<tr>
<td>Measurement Development</td>
<td>Definition of a measurement regarding musical instruments and latency</td>
<td>Choice of the experimental design</td>
</tr>
<tr>
<td>Experiment Development</td>
<td>Design of the listening experiment</td>
<td>Internal and external validity, test subjects, variables and reliability</td>
</tr>
<tr>
<td>Pilot Test</td>
<td>First approach of the experimental setup</td>
<td>Observations on the validity of the listening test and first results (tendencies)</td>
</tr>
<tr>
<td>Experiment Adjustments</td>
<td>Descriptive statistics, detection of tendencies</td>
<td>Validity of the listening test set-up and changes adopted for the final experiment</td>
</tr>
<tr>
<td>Experiment</td>
<td>Conduction of the final listening test</td>
<td>Numerical data gathering</td>
</tr>
<tr>
<td>Analysis</td>
<td>Descriptive and inferential statistics</td>
<td>Relationship between musical instruments and latency</td>
</tr>
<tr>
<td>Conclusions</td>
<td>Interpretation of the results</td>
<td>Confirmation/rejection of the hypothesis and further discussion</td>
</tr>
</tbody>
</table>

Figure 3.2: Methodology stages of the research design
The research design includes a pilot test, implementation of the experiment and the analysis of the results. The pilot test is necessary in order to assure and test the reliability of the listening experiment. This test may be conducted on a few test subjects and the results analysed. For pilot studies, samples sizes of 10 are sufficient and are enough to reveal effects [125]. However, statistical significance is difficult to achieve with this sample size.

The pilot test allows a better understanding of the different parameters involved in the research and may highlight possible flaws in the methodology. In addition, the first results or tendencies prior to the main experiment can be monitored and further corrections to the methodology can be made.

### 3.1.2 Concepts Selection

Musical terminology is relevant for the development of the experiment. It is necessary to avoid any ambiguity in meaning and wording. Some cues presented in musical communication can be quantified and described by using physical units such as time for the sound duration and amplitude for the sound pressure level. However, there are also musical terms that cannot be described with physical units.

**Tone duration**

Tone duration defines the time interval between the physical onset or start of the tone and its offset or end [102]. The physical unit used to measure tone duration is time in seconds or milliseconds.

**Rhythm**

Rhythm is based on relative and not absolute (physical) time. Rhythm remains constant regardless of changes in tempo [218] and relates to a perception of sound patterns in time as a phenomenon. It refers to the structure of the temporal stimulus [154]. A series of sound with a duration can be called “rhythm” [120].
“...it is the time pattern created by notes as music unfolds over time. More specifically, rhythm is a set of time-spans that elapse between note onsets.”[218, p. 96]

**Tempo**

Tempo is defined as the overall pace of a musical piece [129, 153]. It is commonly known as the speed of the composition [120] and it is measured in beats per minute or BPM.

“...tempo refers to absolute time. More specifically, tempo concerns the speed at which rhythmic patterns unfold. Tempo is typically considered to be the rate of the “beat” - a time-span associated with the rate at which a listener will tap his or her foot.” [218, p. 99]

Tempo, being time-dependent, is most affected by delay [205]. In a musical performance, there are always accelerations and decelerations, regardless of the fixed tempo [129]. The performer influences it by imprinting his own motional and emotional character [102]. A metronome is mandatory to avoid huge tempo variations. Music written in former centuries had no BPM indication but tempo suggestions like *andante, allegro* or *presto*.

**Beat**

A fast tempo has short beats while slower tempos have longer beats [205]. The beat is an isochronous time-span which is perceptually outstanding in the musical structure context [102]. Beat be perceived physically through time. The beat has a cyclical nature. A more formal definition is:

“A recurring moment when tone-onsets are more expected. In contrast to tactus, beats are differentiated from strong to weak and occur within a repeating pattern of beats, called a meter”. [122, p. 410]

**Tactus**

Tactus is recognized by musicians as the basic beat, which forms the most notable periodic pulse in a musical passage. It often coincides with the beat rate. The main difference lies in the undifferentiated pulse. In other words, the tactus does not have the alternating structure of strong and weak [122].
Research Methodology

Meter

In western music, the meter, which is a periodic pattern of beats, conform to the duration of a bar or measure in a musical score [122]. Cycles that are multiples of twos (e.g. marches) or threes (e.g. waltzes) are preferred by western musicians [218].

The meter is inherent to attentional and motor behaviour and relates to the perception and cognition of the temporal stimulus [154]. Therefore, geometric time is the physical unit associated with meter. Meter also involves the anticipation of rhythm patterns [154]. The meter is the way beats are grouped together [153].

Accents

Research findings from Friberg [102] have shown that it is possible to identify two kinds of accents: “immanent accent” and “performed accent”. The first is an accent perceived due to the structure of the score when notes are placed in metrically strong positions. The second is the musician’s contribution. The performed accent is used to reinforce immanent accents, normally due to an increase in the loudness of the performance. The most effective way to perceive accent is by increasing loudness and perception of loudness is related to sound pressure level (SPL).

3.1.3 Measurement Development

One of the test’s objectives is to find a measurable relationship between musical instruments and the ability to cope with latency. A definition of a measurement referring to both elements is necessary for establishing the relevant parameters. It enables clarity in the development of the experiment and its outcomes.

Measurement of Latency

The experiment is designed to find a relationship between the ability of musicians to cope with increased latency and the performance of a specific musical instrument. For this research, the total latency is the latency a musician perceives when the audio signal of the musical instrument is delayed and returned by any means (e.g. headphones) to the musician’s ears.
The human voice can be also considered an instrument and is included in the aerophones group. However, research and analysis of a singing performance presents additional difficulties such as the isolation of non-delayed and continuous auditory feedback. Bone conduction is, to this day, impossible to isolate. Results may be biased due to immediate auditory feedback. Continuous auditory feedback is present in normal speech [106]. Thus, the human voice is not part of this research.

The following expression describes the contribution of the different latencies present when a musical instrument is performed over networks. The total amount of latency is named the total latency \( L_t \). The mathematical expression is:

\[
L_t = L_a + L_c + L_d
\]

Equation 3.1: Total latency

where:

The total latency \( L_t \) is the sum of the latency due to the sound transmission in the air \( L_a \), the latency produced from the analogue to digital to analogue conversion \( L_c \) and the network delay \( L_d \).

\( L_t = \) Total latency in milliseconds.

\( L_a = \) Delay (latency) due to sound transmission through the air assuming a speed of sound of approximately 343.21 m/s at 20° Celsius and an air density \( \rho (\text{kg/m}^3) = 1.2041 \) [139].

\( L_c = \) Delay (latency) due to conversion, including the analogue to digital and digital to analogue conversion. The value \( L_c \) refers to the electronic path (virtually no delay) described in Chapter 2 in combination with the digital path, which makes a significant contribution to delay. For this research, no data compression is used.
$L_d =$ Variable delay (latency) of the network delay (e.g. the Internet) which includes buffering, propagation, transmission and processing delays.

The physical unit of latency is time, normally in the range of milliseconds (ms). Under normal conditions all three latencies ($L_a$, $L_c$ and $L_d$) are variable. However, for the listening experiment in this research, $L_a$ and $L_c$ are considered constant values. In the listening experiment, the same devices (audio sound card, cables, microphone, etc.) are used. $L_d$ is variable and is introduced by electronically simulating the network delay and its components as described in Chapter 2.

The total latency equation is a simplification of Equation 2.2. The total amount of latency in this research is grouped into three main factors. However, the total amount of delay (latency) in both equations is the same.

**Defining a Measurement for Latency Tolerance Range**

In previous research [18], the relationship between latency in milliseconds and tempo in BPM is well documented. For this work, the definition of the latency tolerance range (LTR) is based on the following characteristics and assumptions:

- The latency tolerance range of an instrument group is the latency time in milliseconds that musicians of a specific instrument group are able to cope with before disrupting a performance.

- Different groups of instruments (chordophones, membranophones, aerophones and idiophones) may have different LTR values.

- The LTR should prevent the effect of extreme values from the data gathered which may significantly modify the outcome.

To measure dispersion in statistics, the lowest value is subtracted from the highest [93]. Extreme values are always a problem. Excluding extreme values for higher and lower scores up to 25% and calculating the range of the middle 50% is the most common approach. This result is known as the interquartile range [93, 117]. In other words, the difference between the 3rd and 1st quartile represents the latency tolerance range (LTR). The mathematical expression is:
\[ LTR = Q_3 L_d - Q_1 L_d \]

Equation 3.2: Latency Tolerance Range

where:

\( LTR \) = Latency tolerance range measured in milliseconds.

\( Q_3 L_d \) = Third quartile of the \( L_d \) latency values in milliseconds.

\( Q_1 L_d \) = First quartile of the \( L_d \) latency values in milliseconds.

A performance disruption occurs when the test subject, i.e. the musician, is not able to continue with the performance and the musical performance has to be disrupted. At this point the latency can be so disturbing that even strategies to cope with latency are meaningless. The performers make a decision and stop the performance. The measurement of the latency tolerance range is not expected to be an exact onset but a range established from various measurements made on instruments of musical groups such as chordophones, membranophones, aerophones and idiophones. The listening experiment may help to establish the time range at which musicians are unable to play a musical instrument further due to latency. For the latency tolerance range (LTR), the measured latency is \( L_d \).

From previous research, it is well known that some musical performance errors are difficult to detect even for experts, such as music conductors, regardless of their ability to read the score and listen to the performance [145]. Performing errors and strategies to cope with latency as observed by musicians in experiments with delayed audio feedback are also to be expected for this research.

The LTR is defined as a range and not as a threshold (i.e. the interquartile range). From previous work on psychoacoustics by Fastl and Zwicker and based on the Compendium of the Implementation of Listening Tests in Science and Industrial Practice (Kompendium zur Durchführung von Hörversuchen in Wissenschaft und industrieller Praxis) published by the German Acoustical Association (Deutsche Gesellschaft für Akustik e.V. DAGA)
Research Methodology

[88], it is known that the perception of transitions or changes is gradual and contained in a range. A test subject will not always perceive a stimulus even if the stimulus is presented repeatedly in the same way.

3.2 Experimental Approach

For an experiment to be successful, different variables, such as independent, dependent and control variables have to be defined before gathering the data [132]. Furthermore, the control of confounding or extraneous variables defines the extent to which conclusions are near to the reality of the studied phenomenon. This control yields a better internal validity of the experiment [151]. The next sections present the control mechanisms necessary for the experiment.

The experimental set-up foresees:

1. Definition of a statistical sample from a population: the test subjects.
2. A set of questions prior to the listening experiment, in order to gather categorical and numerical data: the questionnaire.
3. A tool which enables measurement accuracy and equality of conditions during the experiment for different test subjects: the metronome.
4. An input to evaluate which should be similar for all test subjects: the musical score.
5. A methodical approach to evaluate the variable: the listening test based on “true experiment” research design (repeated measures design).

3.2.1 Validity of the Experiment

In order to search for explanations and forecasts that can be generalized for a vast number of persons, a quantitative research design was chosen. The validity of the results and the reliability of the instruments used in the experiment are elements to be considered before the experimental phase [151]. The validity of an experiment is proportional to its replicability. Moreover, good experimental design yields higher validity [92].
The data analysis process chosen for this research relies mainly on inductive reasoning. In experimental research, it is necessary to ensure the validity of the experiment and therefore support the findings of the research. For quantitative designs, two validity concepts are defined: internal and external validity [40, 151, 92].

**Internal validity**

The extent accurate conclusions can be generated based on the research design and the selected methodology is directly related to the internal validity of the research [151]. A well-designed experiment is a requirement for internal validity [92]. Furthermore, the more controlled an experiment is, the easier it is to observe any potential threats to the experiment [88].

In establishing a cause-and-effect relationship, special care has to be taken to eliminate alternative explanations regarding the outcomes or results obtained from the experiment [151]. To increase the likelihood that the results observed yield the correct explanation for the observation, planning ahead to eliminate alternative explanations is recommended. A common strategy to achieve a better internal validity is used in this research:

- Controlled laboratory study.

To avoid any threat to the internal validity of the experiment, special care has to be taken with the equipment used and its calibration. The listening test presented in this chapter was conceived to eliminate and, in some cases to ameliorate, the risks to the internal validity of the experiment. The following list, summarized from the works of Campbell and Stanley [40] and Leedy and Ormrod [151] presents the main threats to the internal validity and how these are avoided in this research:

- **History**: The likelihood of uncontrolled events taking place during the experiment is very low to non-existent. The listening test is done over a relatively short interval of time.

- **Maturation**: Such an effect might be present. Some of the participants will be able to gain some dexterity while doing the task. However, the randomization of the stimuli presented (i.e BPM tempi) may diminish the maturation effect.
• **Testing:** The test is only taken once, therefore subsequent tests are not part of the research design.

• **Instrumentation:** The experimenter will be the same for all experiments. There are no other observers. In other words, other people do not have to be trained to conduct the experiment. In addition, the equipment will always be the same and its calibration is controlled prior to the experiment (see Appendix B, C, and D). A questionnaire is also handed out to the participants.

• **Statistical regression:** As stated previously, the participants will only take the test once. However, the participants undergo the same test, measuring latency with three different metronomes (see Chapter 4). It is possible to obtain different measurements each time. The simultaneity of perception regarding aural and visual stimuli, as presented in Chapter 2, may play a role.

• **Selection:** Not relevant for this research. Comparison groups are not part of the experiment.

• **Attrition:** Formerly described by Campbell and Stanley [40] as “mortality”, attrition is also not relevant in this research. Every participant has the possibility to withdraw from the experiment. In doing so the sample is reduced but this is not considered a threat for the experiment as a whole.

**External validity**

To generalize the results of the experiment for the population, external validity must be achieved [151]. Defining the sample for the whole population is one of the keys to external validity. The problem arises when dealing with humans due to diversity, personality and character. Depending on the restrictions used to define the sample, it may be possible to generalize the conclusions. However, a representative sample is a compromise between the many factors involved in the experiment. The representativeness of participants is relatively moderate, taking into account the fact that the majority of test subjects are music students.

Experimental research designs conducted as “field experiments” have a higher external validity when compared to classical experiments with a lower degree of generalization for application in the real world, as is the case in this study. Nevertheless, control of the different variables in the experiment is unavoidable and yields a better internal validity.
Artificially controlled environments may lower the external validity of the experiment. A controlled experiment is not the same as a real world situation. Moreover, the duration of the study (up to 30min) may not be realistic [48]. However, hearing perception remains the same in both listening experiments and in real life [88], but what really is different is the way the musical instrument is performed, the room acoustics and the influence of the equipment used on the musicians and therefore on the performance.

3.2.2 Definition of the Variables of the Experiment

The variables of the experiment have to be carefully described, assigned and controlled to ensure that the conclusions of the experiment relate to these variables and not to random uncontrolled variables that may be confused with the primary variables [22].

A laboratory environment is mandatory if the variables have to be controlled. The less number of unkown variables in an experiment, the higher its accuracy. Field experiments are not suitable for this task [22]. In his work “Statistical Design for Research”, Kish [140] presents four classes of variables:

1. Explanatory variables (class E).
2. Controlled variables (class C).
3. Disturbing variables (class D).
4. Randomised variables (class R).

Explanatory variables

Explanatory variables, also known as “experimental variables” are the variables which better describe the purpose of the experiment [140]. With the definition of these variables, the experimenter intends to measure the specified relationship. Explanatory variables are divided into independent variables, known as “predictor (X)”, and dependent variables, known as “predictand (Y)”. Scientific theories, knowledge and insight into the study field are the sources for the design of the explanatory variables [140]. On average, in listening tests, independent variables are physically measurable parameters and dependent variables are psychological parameters [88].
Independent variables

The variables that may be manipulated by the experimenter during the listening test are:

- Latency.
- Musical tempo in BPM.
- Musicians/musical instrument: this is a controllable variable or factor with one specific value, the musical instrument itself.

The degree of latency of the monitored musician’s audio signal, i.e. their own performance sent through headphones, is manipulated. On the other hand, the tempo of the performance during the trial remains the same, while the amount of latency increases. However, there are several trials with different tempi in BPM. For this reason, the tempo in BPM can also be considered an independent variable. The influencing values are called factors and their values are called levels [117]. Both variables, latency and tempo, are the factors in the study. The levels of the factor latency are milliseconds (ms) and for tempo they are beats per minute (BPM). The different levels of the factor tempo are randomized. Due to randomization, data with a statistical normal distribution is expected to be obtained and both the estimation of the effect to be examined and the estimated error are unbiased [117].

Another variable that is controlled by the experimenter and is therefore independent is the test subject [22]. Musicians performing a musical instrument are the test subjects, the instrument can also be considered an independent variable strongly related to the musician. Therefore, both musician and musical instrument are seen as one independent variable. The analysis of musician and musical instrument is difficult to separate. However, control instruments assure measurable conditions where the influence of the musical instrument can be better observed.

It has to be assumed that results may even differ by repetition under the same conditions with the same musician and the same instrument. There are a lot of dependent variables relating to personality and cognition which are very subjective and very difficult, or even impossible to measure using state-of-the-art methods. The performance of music is strongly related to its environment [203]. An experiment is an artificial environment, which may influence the development of the musical performance. Furthermore, it is impossible to standardize playing style, agogical accent and internal timing. Their influence on the results is clear but difficult to estimate [158].
Dependent variables

Dependent variables may be influenced by independent variables [151]. In short, the dependent variable is the subject’s answer [22]. The dependent variable is physically measurable but represents a psychological parameter. For this study, the variable to be observed is:

- The ability to cope with latency (measured in milliseconds).

To estimate the ability of a musician to cope with latency, the experiment measures the time when a performance breakdown occurs. A performance disruption or performance breakdown is when the musician stops the performance, he/she plays no more the musical instrument. Stability is an important criterion regarding the dependent variable. When repeating the experiment, the same scores have to be obtained under the same conditions [132]. However, having the same conditions is not a guarantee for obtaining the same values with human participants. Instability between scores or latency values is to be expected even when measuring the same musician under the same listening test conditions.

The relevant value in this study is not the point or threshold where the sensation is achieved (i.e. hearing the latency), but where the musician is unable to perform further. A radical difference in comparison to the classical psychophysics methods lies in the answer possibilities relating to the test subjects. For this study, the performer plays the instrument until it is impossible to continue performing. In short, the performer does not give an answer as in classical psychophysics methods, but rather a behaviour (i.e. the stop or the breakdown of the performance). However, some of the common biases of the former classical methods may still be present. Possible biases of this method are [109, 88]:

- “Error of habituation”.
- “Error of expectation”.

The musicians can adapt themself gradually to latency and deliver a performance without breakdown. With the second kind of error, the musician may breakdown the performance before being unable to perform anymore. These errors do not cancel each other statistically.

It is well known that humans learn via repetition [122]. It is an impossible task to eliminate any learning effect due to habituation in the listening experiment. The longer the
test, the easier it is to induce a reduction of the response magnitude on test subjects. The mental process and the answer to a stimulus is faster when the exposure time to the stimulus is increased [122]. Humans habituate easily to some situations and are very capable of adapting themselves to produce the same response even when the stimuli are slightly different. Increasing or decreasing stimuli may produce the same result [164]. Nevertheless, it is possible to design some strategies in order to avoid any “learning effect” through repetition. Those strategies can be achieved with thoughtful musical score design and in the way the listening test is approached.

As stated before, the learning effect is constant in experiments with human beings. From research done by Mäki-Patola [157], it was observed that even after randomizing the order of the latencies, a lot of “noise” caused by learning was still observed in the data. In addition, some factors are difficult to eliminate even with the strategies mentioned above. Musicians are able to play with a subjective time interpretation regardless of what is written in the score. They will never perform the exact time duration of notes [91]. Musicians always imprint their own style [129]. It is well documented that performers can change the emotional character of a piece by changing its performance [102].

It is important to define in advance exactly what is to be measured and to avoid any possible confusion relative to the terminology used. It is reported that values in psychoacoustic results are mistaken in different ways (e.g. loudness and pitch) [17]. Therefore, it is mandatory to explain to test subjects any possible variable to be analysed in the experiment. Moreover, prior to the beginning of the experiment, the test subject has to understand that there is no right or wrong way of performance delivery. This understanding may help test subjects to feel that they are not under test conditions [112].

**Control variables**

Experiments should be designed in such a way to prevent the variation in subjects masking the effects of the experimental variables (factors) [161]. In the experiment design, control variables play a crucial role.

The controlled variables are independent and mostly maintained constant during the experiment [132]. The experimenter knows the control variables [22] and controls them either by including them in the experimental design or by randomization. Control occurs
either through design, as is the case for this research, or by estimation techniques in the statistical analysis. A combination of both is also possible. The main goal of this control is a reduction in random errors or a decrease in the biasing effects of disturbing variables. Obviously, both cases can be also present in the experimental design [140]. Influencing values, which are not set as factors, i.e. independent variables, have to remain constant to prevent effects that can influence the outcome of the experiment [117]. Even though the statistical sample from the whole population depicts homogeneity, i.e. music students or music professionals, humans are very different with respect to intellectual capability or personality. These characteristics produce different reactions when interacting in a listening test which may produce bias [88]. Therefore, the exact experimental conditions have to be defined prior to data collection [161].

In this research, the purpose of the controlled variables is mainly to avoid any biasing effect. Two control variables are essential for the listening test:

- The metronome in its different versions (aural, visual and aural-visual).
- The musical score performed by the participants.

Other control variables that are also independent but not especially developed for the experiment are:

- Recording and reproduction system with all its components such as a screen, microphones, headphones and cables.
- The listening room and its acoustic and physical properties such as temperature and noise levels. However, the effect of the room is less relevant due to the use of headphones.
- The sound pressure level of the audio signal on the headphones.
- The gain of the microphone when recording the signal.
- The measurement to determine normal hearing ability is a control mechanism to prevent subjects with hearing disease from participating. Musicians without healthy hearing may influence the outcome of the experiment and produce bias in the measurements.
- The measurement of the SPL levels of the musical instrument’s output may control the range (volume) in which these instruments have to be performed.
Research Methodology

To avoid the influence of the musician’s interpretation in connection with tempo due to motions and emotions [102], the use of a metronome is mandatory. According to Fraisse [101], it is well known that perceived and physical simultaneity may differ from each other. Musicians cannot synchronize perfectly with a metronome and, furthermore, every musician has an inherent phrasing and timing when performing a musical instrument [66, 205]. However, researchers have used a metronome in their experiments to guarantee equality of conditions among the musicians [5]. The use of a metronome should be seen as an additional guide and help for the performing musician.

From qualitative research on latency perception [226], it is known that the role of a metronome and its success as a control mechanism depends on previous experience of the musicians. Moreover, the synchrony on the perception of audio and video signals as explained by Kuling et al. [146] and Altinsoy et al. [9] may play a very important role for the results of the different metronomes used in the listening test. Researchers have found that aural information is preferred over visual. Visualization supports aural information [142].

The second important mechanism to control the experiment is the score. This should be relatively simple to play. The physical characteristics of the musical instruments have to be taken into account for the design of the musical score. It is one of the most important elements of the research. It provides a specific input with which to compare the results obtained between the different musical instruments. The aim of the experiment is not to test the musical dexterity of the musicians but to enable equal conditions regardless of the physical advantages and disadvantages related to the performance of any musical instrument.

The proposed score should satisfy requirements such as:

- Any musician should be able to play the melody regardless of the musical instrument performed.
- Any ornamentation must be avoided.
- The musical key should not be altered during the performance.
- Familiarity or reference to known musical pieces should be avoided.
• Memorizing the score should be difficult.

• Estimating a melodic line should be difficult.

• A common meter should be used.

These requirements are necessary in order to achieve a better evaluation of the desired latency parameter. Some of the requirements are easier to achieve than others. Due to the different groups of musical instruments to be evaluated, it is necessary to define a common score which can be performed in the same way by different musicians.

Having a score which is difficult to memorize without musical familiarity [80, 168] and with a relatively difficult to estimate melodic line forces musicians to concentrate on the metronome and on the score. From previous research, it is known that unconscious expectations in the melody have to be taken into account regarding western musical structures [25]. Otherwise, the introduced latency may not be a big obstacle or may not significantly interrupt the performance.

Studies on percussionists have shown large differences among players when performing the same task. For percussionists, motion trajectories, played patterns and preparation for accents and strokes (upstroke and downstroke) differ from one musician to another [210]. However, it is possible for every musician to play every note and rest with similar time ranges [42]. Rhythm is expected to be the same under different tempo velocities. It may create an undistinguishable sound pattern. From previous research, the direct relation between the delay acceptance threshold and speed and note resolution is well documented [5].

Three basic musical gesture categories are necessary for music performance. These include effective gestures which are necessary for sound production, figurative gestures which are sonic gestures not directly related to the music production and accompanist gestures which are visible ancillary gestures (e.g. head movements) with no connection at all to the production of sound [80]. Musical gestures are coupled with velocity patterns because body movements are involved [129]. Body movements in a musical performance not only influence the production of sound but also the achievement of expression [227].

Studies on pianists have established a close relationship between body movement degree and musical expression intensity. These results may be applicable to a wider range of instruments [102]. Larger movements are related to more expressive musical intentions [77].
Moreover, there is evidence of a direct link between rhythm perception, physiology and body metrics. Musical elements like timbre, dynamics, intonation and timing are strongly affected by the musician’s interpretation of the performance [77]. Perception of rhythm may be influenced by the human vestibular system, body size and even the length of the performer’s leg [220]. In brief, body movements also affect the outcome of a performance [129, 76, 168, 177].

The majority of studies provided the music scores or rhythm patterns before the listening test [198]. For this experiment there is no rehearsal prior to the test to minimize the effect of memory recall [95, 176]. Therefore, the effect of delayed feedback due to having learned the score previously provides no bias.

The listening experiment should not exceed 30 minutes since a possible learning effect may be taken into account. Exposure times of approximately 21 minutes are required to learn passively [122]. Randomizing the BPM tempo may minimize the learning effect. The more a musician performs a score the easier it is for the musician to cope with latency.

Tactile feedback will be always present. It is impossible to avoid tactile feedback in the listening experiment. In listening experiments the only feedback cues that can be eliminated are aural and visual [147].

**Disturbing variables**

In order to enhance internal validity, towards the identification, if present, of a cause and effect relationship, it is relevant to control extraneous variables. These uncontrolled extraneous variables may otherwise threaten the internal validity of the results. When the extraneous variables relate to both explanatory variables (independent and dependent), the extraneous variables may deliver an alternative explanation for the outcome and become a “confounding” variable [140].

There are a lot of possible and even unavoidable effects during the listening test due to the non-mechanical character of the test subjects. Inaccurate measurements could be the outcome of this. It is very important to know which disturbance variables are present [58] and how they can be controlled, if not eliminated, so that the effect of introduced noise can be reduced in the output data. The role of perception and the “human factor” has
been observed and questioned in other works [49]. Research carried out on distributed
music ensembles and in laboratory conditions eventually provides inconsistent results.

The role of uncontrolled extraneous variables should be minimized to avoid them as a pos-
sible explanation for any effect observed [151]. However, avoiding all extraneous variables
in an experiment is not possible. Randomization and control strategies may diminish the
role of these variables [22]. An experimental design is better suited for the methodology
of this research. For a non-experimental design, it is very difficult or even impossible, to
translate a class D variable into class C (controlled) or class R (randomized) variable [140].

Commonly found extraneous variables are:

- Individual participant differences such as subject’s fatigue, expertise, mood, gender
  and age.
- Instrumentation.
- Environment.

Environment and instrumentation are extraneous variables which are controlled in this
experiment. The first is controlled through headphones, minimizing the effect of the room
environment. The latter is controlled through the reliability of the equipment used (see
Appendix B and C).

Participant characteristics such as fatigue or mood are very difficult to control. On the
other hand, the influence of gender, expertise and age of the test subjects may not be
controlled at all. In classical psychophysics, the context effect describes the effect of per-
ception with regard to the ability to make any judgement. In other words, judgements
are relative to the context [88].

The role of the effect of other stimuli present in the test section is known in psychophysics
in the concept of “contextual effects” [109]. The context may strongly influence the par-
ticipant’s judgement. The problem arises specifically when test subjects have to give an
answer on the intensity of stimuli. Even the presentation order of the stimuli may bias the
responses of the participants [109]. With the contextual effect, two categories are found:
interindividual and intraindividual variables [88]. The first category includes variables
such as gender, age and expertise. The second category includes mood, psychological characteristics or even the hour of the day.

Experience, expertise, mood and diverse individual variables relating to the personality of the test subject may have an effect on the response of the participant. The problem relating to the relativity of the answers can be minimized by giving a uniform and constant frame of reference to all test subjects [88]. By using a questionnaire (see Appendix I), it is possible, to some extent, to document all the different individual characteristics of the participants in a standardised way. In addition, two powerful strategies are used to control the extraneous variables in this research [151]:

- Similarity of conditions.
- Exposure of the participants to all experimental treatments.

The use of the within-subject design enables repeated measures with the subjects (repeated-measures design). All participants are exposed to the same treatment (listening experiment) [151]. Furthermore, the conditions of the experiment have to be standardized in such a way that every participant undergoes the same treatment. In other words, any variation observed in the data might be a product of the participant and not from the internal differences between listening tests. To avoid possible sequential effects the within-subject design provides a better design tool compared with other research designs [88].

Measuring the same test subject with the same methodology and equipment will produce different results from trial to trial even during the same testing session [109]. As stated before, these sequential effects can be diminished through the research design method. In addition, the listening test involves a repeat of the measurement within the different stimuli (three different metronomes producing three different measurements for every tempo in BPM).

To ensure the similarity of conditions the music style, or score, for all instruments is the same in all listening tests. From previous research [158], it is known that the music style performed has an effect on the outcome and results on the perception of latency.

The suggested control mechanisms are:

- The musical score.
• The three metronomes (aural, visual and aural-visual).

• The equipment used: gear such as headphones, microphones, cables, DAW and PC are the same for all participants.

• Participants are musicians either professional musicians or music students.

The extraneous variables, which are randomly distributed, may increase the difficulty of achieving the researched effects. In other words, the sensitivity of the experiment for detecting certain effects is reduced. On the other hand, the confounding variables vary systematically with the dependent variable and without control these variables may produce bias in the measurements [48].

In order to minimize the effect of the confounding variables that may affect the dependent variable, some extra additional control mechanisms are adopted:

• SPL measurements on instruments.

• Use of isolation headphones.

• Measurements to determine a normal hearing ability.

High internal validity is achieved through the control of variables. Therefore, changes observed in the dependent variable are the product of changes in the independent variable [48].

**Randomized variables**

Randomized variables are also uncontrolled extraneous variables. However, they are treated as random errors [140]. In experiments it is possible to randomize those uncontrolled extraneous variables in an operational way. The task of the experimental design is to convert any possible extraneous variable into controlled variables and therefore eliminate any possible alternative explanation for the relationship between explanatory variables. The effect of randomized variables may cancel itself. Furthermore, its effect on the explanatory variables may not be sufficiently strong and it will, therefore, not introduce a relevant bias [140].
The goal of an ideal experiment is to eliminate all the class D variables, either by controlling them, i.e. switching to class C, or by placing them in the randomized class R. However, the control of all variables is not realistic and a lot of them remain uncontrolled.

Figure 3.3 is based on the “Effects of three classes (C, R, D) of extraneous variables on the explanatory (E) variables (X→Y)” from Kisch [140]. This figure summarizes the role of the different variables of the experiment. In a true experiment, any disturbing variables should be eliminated, either by controlling or randomizing the variables so that their effect is not relevant.

The thickness of the arrows represents the influence of the variables. Through control due to the experiment design, represented by a cube, some variables are not considered a threat to the experiment. The effect of randomized variables in the experiment is also not relevant (almost a straight line), and they cancel each other out [140].

In contrast to a sample survey or an observational study, running an experiment “theoretically” eliminates the effect of all disturbing variables at the cost of a real representation of reality. The outcome of the experiment is far removed from the target parameters. In short, the experiment delivers estimations and does not represent broader populations as in the case of a sample survey [140].
Sample population

The advantage of experiments compared with other research methodologies lies in the control of explanatory variables. The drawback of experiments is mostly the representation of defined target populations. In addition, measurements are often not very realistic [140].

In order to define a representative sample from the vast population corresponding to the broad number of western musicians, it is necessary to define and specify the characteristics of the test subjects [161]. The sample of the population allows the experiment to be conducted in a practical way. It enables assumptions to be made about the whole population. It is important to recognize that statistical inferences gained from the experimental research design are limited to the selected population [140].

The relationship between dependent and independent variables depends on the population subjected to the research [140]. The sample size depends mainly on the population and the expected accuracy of the results. Other factors that should be taken into account are time, availability of test subjects and financial issues. Samples sizes between 10 and 30 subjects may be allowed when larger samples are economically unfeasible [125]. Although bigger samples are usually better than smaller ones [197], bigger sample sizes do not always lead to an increased accuracy. In contrast, they could even also reinforce statistical errors [82].

An accurate answer to the question on the number of subjects is difficult to answer. There is no accepted method for determining the necessary sample size [118]. Some authors such as Gay and Diehl [108] mention that the sample should be large enough. However, it may strongly depend upon the type of research. For experimental research 30 subjects per group are often quoted as the minimum [108]. This assumption of 30 samples or more is also share by Roscoe [197] who describes some rules of thumb when it comes to the number of subjects for experimental research. He states that for experimental research having tight controls, even smaller samples sizes between 10 and 20 may work. A similar assumption is made by Chassan [60], who estimates that 20 to 25 subjects per group may be a reasonable minimum. Nevertheless, statistical analysis is not possible if the sample number is less than 10 [197].

Regarding the number of test subjects, it is possible to make a general assumption. Having chosen the “within-subject-design” as the research methodology, it is expected that
the variance between test subjects does not play a dominant role compared with other research methods such as the “between-subject-design”. In short, for a “within-subject-design”, fewer test subjects are needed. There is only one group which undergoes the different treatments including the control condition, if available [151]. Differences between participants are easier to check during the test [88].

Participant and subject have the same meaning in this research. Both indicate human beings who are able to perform an instrument and read a score. However, a subject is defined in the literature as something broader describing different populations such as humans or animals [151].

Some restrictions have to be made regarding the sample population [22]. The test subjects for the listening test are exclusively professional musicians or music students. Assumptions have to be made regarding the two main categories of test subjects for the experiment. Professional musicians are those who earn a living through music. It is expected that their availability for the listening test may be restricted because of their tight agenda and variable working issues. On the other hand, music students are those who perform an instrument over many years not just in an academy context of a music university but also in orchestra communities, at high school and obviously during the pursuit of an academic title. Both groups have to be proficient with regard to reading notes in a score.

Previous research [158] has come to the conclusion that there are no measurable differences between musically skilled and less skilled subjects. However, research done by Farner et al. [90] and altered auditory feedback (AAF) research done by Pfordresher [184] showed an important difference between measurements on musicians and non-musicians. Sensitivity to delayed auditory feedback (DAF) can be better mastered by musicians. They can ignore interfering auditory events such as the sound produced by other musicians when performing together [184]. Furthermore, in experiments on sensorimotor synchronisation, musically trained subjects respond, on average, differently to subjects without musical training [12].

In research and listening tests, such as Chafe’’s ensemble hand-clapping experiments by pairs of subjects [54], performers with musical experience slowed down the tempo more than non-musicians. Furthermore, the ensemble imprecision by musicians was lower [90]. Researchers, such as Fraisse [101], were aware of the difficulty of comparing results be-
Research Methodology

tween trained and untrained subjects in experimental situations.

For this study, the ability to play a musical score is mandatory. In fact, a standardization of the skills of test subjects is guaranteed by allowing only trained musicians or music students to take part in the study.

To represent the population of musicians directly related to network music performances (NMP), some assumptions have to be outlined:

- Network music performance is a relatively new research topic whose performers are found mainly in the academic environment of musical universities.

- Young people are more receptive to new technologies such as the Internet and the use of new network technologies.

The population contains mainly European musicians, especially music students, which constitutes a restriction or control of some of the conditions, meaning the external validity of the experiment may be compromised [151]. On the other hand, the limitation of the population means that making accurate assumptions based on the results obtained from the sample population is possible. In addition, when control in the experiment is strong and the variables are small in number or totally unpredictable, a number of 30 samples is also feasible [125]. However, the more data, the better the experiment, and the statistical power may be increased. Time, budget and even space [7, 197] are relevant parameters when it comes to the collection of a large dataset and may justify small sample sizes [125, 160].

Having chosen a research design, it is important to fit the resources to the situation. The target population and the selection of sampling units are a compromise regarding the “representativeness” of the sampling units [140].

The characteristics of the sample population are very specific. Purposive sampling is the most appropriate sampling design for this test study. Purposive sampling is a non-probability sampling design [151] and is mainly used when it is not possible to enumerate all the population elements or when preliminary studies are developed. People are chosen with regard to certain characteristics. The main attributes are the ability to play a score and perform a western musical instrument. Using non-probability sample designs allows
statistical inferential analysis of data from the sample. However, results may not be generalisable with regard to the whole population [26]. Some drawbacks of non-probability sampling are the selection bias. In other words, some subjects may be excluded, i.e. biased samples [26]. In addition, the sampling error cannot be calculated.

Drawing random samples implies that each test subject of the whole population is included with only some of them selected randomly. This assumption is very often not practicable [82]. Population samples in the social sciences, medicine and biology are frequently non-probability samples. The available subjects undergo the research. These samples represent a virtual population. For this population, the purposive sample may constitute a random sample [26].

**Questionnaire**

A questionnaire is not a control mechanism but can provide additional information on the test subjects before the listening test [70]. Categorical and numerical data is collected in a personal interview in the same session before the listening test. Sessions are indexed, thus preserving the anonymity of the test subjects. Participants are debriefed on the scope of the research and the use of the results. Musicians agree to participate voluntarily and are allowed to withdraw from the experiment for any reason whatsoever, whenever they wish.

The purpose of the different questions presented is to recognize patterns that might describe certain characteristics presented in particular musical instrument groups. These patterns can be further analysed through statistics. A questionnaire enables the classification of the relevant information before numerical data is collected.

As explained in the previous chapter, the questionnaire is not a control mechanism but provides additional information. Table 3.1 shows the questionnaire. No questions were formulated in any suggestive way to obtain specific answers, the approach is strictly quantitative. On the contrary, the questions are mainly informative and the answers are only categorical or numerical. The information gathered in the questionnaire provides additional categorical and numerical clarification for the results obtained from the listening test. In some cases, it may be confirmatory and could explain some outliers present in the descriptive statistics.
### Research Methodology

<table>
<thead>
<tr>
<th>Item</th>
<th>Question</th>
<th>Possible Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Instrument</td>
<td>Name and group of the musical instrument</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td>Age of the test subject</td>
</tr>
<tr>
<td>3</td>
<td>Gender</td>
<td>Male or Female</td>
</tr>
<tr>
<td>4</td>
<td>Metronome Preference</td>
<td>1. Visual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Aural</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Both</td>
</tr>
<tr>
<td>5</td>
<td>Expertise</td>
<td>1. Professional Musician</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Music Student</td>
</tr>
<tr>
<td>6</td>
<td>Level of Expertise</td>
<td>1. Amateur</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2, 3, 4 In between</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Professional</td>
</tr>
<tr>
<td>7</td>
<td>Years of Experience</td>
<td>Number of years</td>
</tr>
<tr>
<td>8</td>
<td>Playing technique (chordophones)</td>
<td>1. Plucked</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Bowed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Struck</td>
</tr>
<tr>
<td>9</td>
<td>Current hours of practice per week</td>
<td>Average number of hours</td>
</tr>
</tbody>
</table>

Table 3.1: Questionnaire for the pilot test

### Listening test

The most salient characteristics of every listening test are [88]:

- Methodical and well planned.
- Repeatable.
- Controllable.

The listening test is based on systematic observations [88]. The ultimate goal of the test is to collect data. Normally, data can be acquired through verbal statements or through
Research Methodology

body movements such as key pressing [88]. A well-defined and simple set-up enables reliable data acquisition. For this study, a direct response, either verbal or due to body movements, is not expected.

The most important elements of the listening test are:

1. The input signal of the instruments for further evaluation and comparison: recording the musical instruments.

2. A measurement which enables the comparison of the initial conditions between different musical instruments: sound pressure level (SPL) measurements.

3. A piece of equipment to ameliorate the effect of non-controllable variables: headphones.

4. A method to evaluate the ability of test subjects participating in the listening test: measurement to determine normal hearing ability.

The methodology approach for the listening test is similar to that used in classical psychophysics, specifically the method of limits [109].

Reliability of the measurement instruments

The measurements obtained from an instrument have to be consistent [151]. The experiment has to be replicable by any other person [92]. Reliability is necessary for validity and therefore it is necessary that the equipment is reliable and consistent. The equipment used during the experiment is:

- Digital audio workstation (DAW): a sequencer software for recording and reproducing the audio material.
- Personal computer (PC) to run the DAW software.
- Audio interface which is connected to the personal computer to enable the input and output of analogue and digital audio signals with the DAW (see Chapter 4).
- Microphone to record the audio signals.
- Headphones to listen to the delayed audio signals and the aural metronome.
• HD LCD monitor to display the musical score and the visual metronome.

• Ancillaries such as cables to connect the different pieces of equipment and microphone stands for the microphone.

The reliability of instruments when measuring physical (substantial) phenomena is higher than those measuring social insubstantial phenomena. However, it is relevant to know how reliable the measurement instruments are. Four parameters help to determine the reliability of measurement instruments [151]:

1. Interrater reliability: identical judgements from different individuals are expected.

2. Test-retest reliability: the equipment used should deliver the same results from the same participant on two different occasions.

3. Equivalent forms reliability: different versions of the same measurement instrument yield similar results.

4. Internal consistency reliability: the magnitude to which all items within a single instrument deliver the same results.

In the listening tests, parameters 1, 3 and 4 are simple to provide. Parameter 2 may not be satisfied, not because of the unreliability of instruments, but due to characteristics inherent to humans and music performers. Even with calibrated measurement instruments, it is difficult to obtain the same measurement when repeating the experiment. On the other hand, the instruments used enable different participants to deliver similar measurements with the dependent variable yielding a specific result measured in the DAW. Using other DAWs, the same measurements should be obtained. Other DAWs were discarded because of the absence of some relevant features (see Experimental Procedure chapter). The fourth parameter, relating to reliability is also achieved through the selection of good instruments and their calibration.

The measurements in this research are direct. Therefore, biasing factors regarding prejudices, cultural interpretations and other factors not relevant to the performance of music can be discarded. Nevertheless, measuring substantial phenomena may potentially deliver slightly different values from time to time due to a wide variety of reasons. The imperfect reliability of the measurement determines how different the measurements obtained are.
A consistent, accurate measurement is the best mechanism for avoiding reliability errors [151]. The production of measurement errors is unavoidable, however, the experimental design presented in this chapter is robust enough to minimize such errors.

Several measurements on the equipment used in the experiment are described in Appendices B, C and D. Information from the manufacturers was compared with measurements on the equipment in order to provide and estimate the reliability of the equipment used in the listening test. In addition to the procedures described in Appendices B, C and D, the experimenter has the following options to enhance the reliability of a measurement instrument [151]:

- Standardization.
- Pilot Studies.

Standardization relates to the manner in which measurement instruments are employed in the experiment. In this research, where the author is responsible for all the measurements of the participants, it can be expected that the experiment routine does not vary between participants. Furthermore, the whole experimental process is planned ahead and the procedure is always the same.

Based on previous literature the experiment was designed according to accepted methods and experiences from a variety of authors. Nevertheless, an integral part of this research is the pilot study. This study enables a better planning of the final listening test experiment and highlights the potential flaws that can be corrected before the main experiment takes place.

### 3.2.3 Systematic and Random Errors

Results and the process of gathering data may be affected by errors. Measurements contain measurement errors. Repeating measurements minimizes the measurement errors enabling a better estimation of the real value [88].

Actions to prevent systematic errors have been described in the previous sections. These biases can be minimized by the following actions in the listening test:
• Calibration of the equipment.

• Instrument resolution (this should be higher than the measurement range).

• Double-check the data obtained (re-listening to the audio recordings).

A common systematic error is the offset error. For this research, the measurements begin with 0ms latency. However, it is important to note that a 0ms latency is not achievable using digital equipment. As explained in Chapter 2, there is also latency during sound transmission through air and also through analogue to digital to analogue audio conversions. The next chapter clarifies the “0 point” for the latency measurements.

Random errors are related to measurement accuracy and are not predictable. Strategies to minimize these errors are:

• Averaging measurements from a set of measures.

• Sample size increase.

The first strategy is adopted in the methodology. Every musical instrument is measured three times, each time with a different metronome (aural, visual and aural-visual).

In addition to the errors in the experimental set-up, the test subjects themselves may also affect the accuracy of the results [88]. Measurement inaccuracies in the test subject’s answers may be created through attention fluctuations or attention deviations. These can be considered random errors.

Systematic errors in test subjects are related to memory effects or practice effects [88], known as errors of habituation. The choice of test subjects may also constitute a source of error, although this error can be minimized through random sampling [88].

**Recording the musical instruments**

Special care should be taken with respect to the placement of the microphone to record the performance of the instrument. The recorded sound will return as a delayed signal (latency) to the test subjects.

The microphone distance is not the same and may depend on the sound pressure level of the instrument. Issues to be considered regarding the miking of the instruments are:
• Directional sound radiation patterns of the instruments and sound radiation directional effects are not equal in all directions [165].

• Close-miking is a reliable method to avoid the recording of room information and the influence of room acoustics [145]. In addition, extra room information makes it even more difficult for musicians to cope with latency while playing [104].

• The microphone polar pattern was cardioid. Due to close-miking, directional characteristics of the microphone are not so predominant. However, the proximity effect may be unavoidable. Low frequencies can be boosted.

For every listening test, sound pressure level (SPL) measurements with an SPL meter allow for a better comparison between different musical instruments and their SPL outputs. This control enables equality of conditions with respect to the recording of musical instruments.

Headphones provide a better control over variables such as reverberant room conditions, background noise or sound transmission through the air. To ensure the reliability of the hearing capability of the test subjects, every musician has to undergo a measurement to determine a normal hearing ability prior to each listening test. A healthy and accurate hearing mechanism of the test subject is mandatory for taking part in the listening test. Reliable participants are vital for the collection of experimental data [22]. Two different test procedures are admitted. The first consists of asking the test subjects about their hearing abilities in a questionnaire and a reliable answer is expected. A second procedure is an audiogram [88]. For this experiment, an audiometer is not foreseen as a tool to determine the hearing ability. Further information about conducting this test can be read in the Experiment Procedure chapter.

Using headphones is a compromise where the most relevant characteristics of a room such as the size, reverberation and resonances are eliminated. Musicians adapt themselves individually to these relevant characteristics when they perform music. They manipulate their tempo, dynamic and articulation [203]. On the other hand, controlling the room with headphones produces an artificial environment which has an influence on the external validity. Headphones may also be perceived as disturbing [203].

Although reverberant room conditions enable collaborative performances with a slower tempo, the use of artificial reverb may reduce clarity of sound and the ability to hear
the note onset [5]. In previous research pure signals were preferred over those with artificial reverberation. The delayed returned signal was sent back without any reverberation.

### 3.2.4 Pilot Test

The main purpose of the pilot test is to test the measurement method proposed and to observe the different measurement issues that could arise during the experiment under real conditions. In addition, the pilot test or pilot experiment is done to check the experimental set-up, calibration issues and feasibility of the experiment on a small scale [22].

Other important issues to consider are the possible flaws encountered during the execution of the experiment. Procedure errors that should be avoided for the final listening experiment should be detected in the pilot test. The interaction between test subjects and the measurement system should be observed and, if necessary further improved.

Based on the taxonomy of a networked performance, an outline of the relevant elements for this research is presented in Table 3.2.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human role</td>
<td>Instrument performer</td>
</tr>
<tr>
<td>Network topology</td>
<td>Point to point</td>
</tr>
<tr>
<td>Transmitted signals</td>
<td>Audio</td>
</tr>
<tr>
<td>Distance of performers</td>
<td>Remote, simulating distances in the order of kilometres</td>
</tr>
<tr>
<td>Latency and synchronization</td>
<td>Latencies from 0 to 300ms with increments of 10ms. Sync via aural and visual metronome</td>
</tr>
</tbody>
</table>

Table 3.2: Brief taxonomic description of the listening test
Research Methodology

The information in Table 3.2 briefly describes the relevant aspects of the pilot test. The test simulates a simple point to point connection of several kilometres, transmitting audio signals generated through acoustical instruments.

Procedure

Test subjects will undergo a listening test. This procedure includes answering relevant questions about performing the instrument and experience in handling the musical instrument. Afterwards, the listening test begins with the measurement of the relevant variable, i.e. the ability to cope with latency estimated through the measured latency time. The procedure is the same for all instrumentalists. The measurement equipment remains the same during the test. The listening test is carried out in a systematic way to guarantee equality of conditions. The systematic procedure enables the reproduction of the experiment in any environment and also allows general assumptions about the outcomes and results obtained.

In other studies, delays have also been increased up to 150ms [63]. In the first Distributed Immersive Performance experiment (DIP), two musicians playing viola and piano performed Piazolla’s “Le Grand Tango” and excerpts from Hindemith’s “Sonata Op. 11 No. 4” with similar audio delays artificially introduced. The values were in the range of 20 to 300ms [63]. Further reasons for selecting a value of 300ms were the research done by Gates [107] with maximal disruption values around 270ms as well as the categorization work by Smith [209] presenting different latency ranges. The work on delayed speech of Bernard Lee in the 1950s also used 300ms as the delay limit.

The latency increment of 10ms has also been used in previous research and seems to be the best compromise with regard to internal audio interface resolution [232]. Research done by Jack et al. [127] concludes that there is no difference between zero latency and 10ms. Moreover, according to the experience of sound engineers and musical network systems developers such as Yamaha and Audinet, latency values between 20 to 30ms are not an issue for most of the musicians working with musical network systems\(^1\).

Different experiments use different latency levels. From other experiments, it is known that confident latency values might vary from 0 to 200ms, which is the maximum disrup-

\(^1\) Arthur Koll telephone conversation on March 03, 2015.
tive level of delay for people when speaking [158, 54, 55, 21]. Setting latency values and thresholds is a compromise. For this research, experience together with former research were the guidelines used in establishing the latency values.

**Adjustments**

After running the pilot test with a few test subjects, adjustments might need to be made. These adjustments are expected to be minor changes with respect to the execution of the listening test. Based on the first results, it might be possible to estimate if the measurement procedure constitutes a reliable method to measure the expected variable.

**3.2.5 Experiment**

The final test is the combination of the pilot test and any adjustments made. The listening test provides the quantitative data and further information is collected in the questionnaire.

Data collection and data editing (cleaning) are described in the last section of the next chapter. The data gathered is presented in a table. Data cleaning is not expected. Data is obtained through direct measurement in the experiment. However, it is necessary to double-check the data obtained by re-listening to the audio recordings.

**Results analysis**

An important process within data cleaning is to establish the distribution of data [117]. To discover special structures or peculiarities within the data, an exploratory data analysis (EDA) is first recommended. A previous model is not required [117].

The EDA enables the detection of structures, the presentation of the data and also the recognition of important characteristics, especially by studies where the population is not completely defined and where there is not an established model [117]. For our study, the population is already defined. However, its variability may be high.

A second approach after the EDA is an inferential statistical analysis which enables conclusions to be made about the data gathered. For this research, the analysis of variance
Research Methodology

(ANOVA) is used only when results are normally distributed. Otherwise, other variance analysis methods such as the H-Test of Kruskal and Wallis or the Friedman test can be an alternative to establish significances among latency results and specific musical instruments and also between instrument groups. The standard deviation of the mean is another measurement that has to be calculated in order to corroborate or discard previous results. Box plots presenting median and quartiles are the choice for the graphical representation of trends.

Conclusions

Based on the results and the statistical analysis, a relationship between musical instruments and latency may be established. The final outcome should answer the questions proposed in the introduction of this work.

3.2.6 Ethical Considerations

Ethical issues are indispensable for the planning and performing of the experiment. The mental and physical “well-being” of all participants has to be ensured [48].

According to the recommendations established in the Regulations for Postgraduate Research Study of the Cork Institute of Technology, the Code of Good Practice and the codes of conduct accepted in psychology have to be followed.

The following aspects should be taken into account when conducting an experiment [48]:

- Voluntary participation.
- Right to withdraw.
- Full debriefing after the experiment.
- Confidentiality of data gathered and anonymity of participants.

Relevant issues to be taken into account which, according to psychological research [22, 88], are to be strictly avoided include:
• Unfair discrimination.
• Harm (psychological or physical).
• Interest conflict.
• Exploitative relationships.
• Uninformed consent.

A general information sheet for participants and a listening test consent form will be handed out in English or German (see Appendix H) before the test starts. To avoid misunderstandings it is mandatory to present the listening test instructions in written form. Thus, any possible influence of the person in charge of the experiment may be avoided [88].

Ethical clearance for this research was obtained by way of an application to the CIT Ethics Committee. A copy of the CIT Research Ethics Application Form is attached in Appendix N.

3.3 Summary

After discussing alternative approaches in detail, the research question is explored using a quantitative approach. An experimental design is defined, in particular the issues concerning the validity of the experiment, the selection of the variables and the control mechanism with an holistic approach. The result is an innovative listening test which is presented as the core of the experimental design. Chapter 4 presents the next steps describing the experimental procedure and the trends of the pilot test.
Chapter 4

Experimental Procedure

In this chapter, the listening test and its procedure is presented.

4.1 Development of the Experiment

This section describes the different steps relevant to the conception and execution of the experiment. The main approach is based on the work “An approach to quantifying the latency tolerance range in non-collaborative musical performances.” [163] developed and presented by the author in 2014 at the AES 136th Convention in Berlin. Several technical and conceptional improvements developed for the listening test since the first version will be presented in the next sections.

4.1.1 Participants

From the vast number of musicians who may be considered as test subjects for this research and from all musicians who perform western musical instruments in the world, the selection of a representative number can be achieved by taking into account both the previously defined parameters and the logistical aspects. Students from the European music universities in the area of Frankfurt am Main and Darmstadt in Germany and Cork in Ireland represent the population. At the same time, this population defines a limitation with regard to the results obtained. For the final experiment, test subjects are sampled randomly from this population.
4.1.2 Stimuli

The following stimuli are presented to the test subjects during the listening experiment.

**Metronome**

To guarantee uniformity of conditions both between the test subjects and for data comparison, both a visual and aural metronome are used during the experiment. The aural metronome is a regular DAW-audio software-based audio click metronome with sounds for bars and beats. The sound level of the bars and beats is the same. Bars and beats are distinguished by means of a different pitch. The sound is heard as beeps.

The visual metronome was designed with the MAX/MSP software [72] as described in Appendix A. This application is controlled through MIDI data from the DAW. The visual information is displayed on a small HD LCD 7-inch monitor placed in front of the musicians.

![Visual metronome](image)

*Figure 4.1: Visual metronome*

The metronome as shown in Figure 4.1, consists of two circles, with different radiiues, the bigger circle (red colour) represents the bar and the smaller one (green colour) the beats. The blink time is 150ms, which is far beyond the threshold of perception for a visual image. This threshold varies from 40 to 60ms [79]. To increase the contrast the background is black. The colour selection was chosen after performing some trials. Test subjects described which colours were better suited for following the metronome. The use
Experimental Procedure

of a pendulum in the metronome as a tool to indicate tempo was intentionally avoided. Some musicians described their ability to stay in tempo not by the metronome itself, but by anticipating the movement of the pendulum.

As stated before, MIDI data information sent from the DAW enables the operation of the application at any BPM value. For the pilot listening test, different BPM values were applied. Those values ranged from 60 BPM, corresponding to a very slow or adagio tempo, up to 240 BPM, which is very fast or, in musical terms, prestissimo. The interval between the different tempi was 30 BPM, thus producing seven different tempi for the listening test. Small BPM variations, below 5 BPM, are not perceived by professional musicians [153]. In addition, a shorter interval than 30 BPM would extend the duration time of the listening test, making it difficult to execute.

Score

Based on the considerations presented in Chapter 3, the proposed score is shown in Figure 4.2:

![Figure 4.2: Score for the listening test](image)

Ornamentation is another way of influencing the latency measurement. The elimination of any ornamentation in the score may reduce the performance differences between instruments, e.g. a vibrato is relatively easier to play on a cordophone than on an idéophone, which influences the latency measurement and therefore the validity of comparison between measurements. Accents were avoided to restrict, for as long as possible, the intermission of musical expression and therefore highly different interpretations of the score.

Different musical keys enable musicians to perform in different ways [153], making a comparison impossible. The use of a common meter such as 4/4 allows simple ratios on the metronome. Other meters are also possible in accordance with the metronome values of the experiment (60, 90, 120, 150, 180, 210, 240 BPM). Difficulty increases proportionally
to the musical tempo value.

Additional scores in different keys were generated from the original score presented in Figure 4.2. Figures 4.3, 4.4 and 4.5 show, respectively, the bass clef, the alto clef and a neutral clef without pitch especially for drummers.

The notes used in the score are crotchets (quarter notes), quavers (eighth notes) and crotchet rests. The score is two bars long and should be repeated or played in a loop after the second bar. The different tempi in the listening tests (60 BMP to 240 BMP) define the note duration. It is impossible to expect the exact same note duration for different instruments or performers. Furthermore, it is a fact that very good musicians often violate ratio-timed norms when performing a score [129]. As stated before, this deviation may be one of the reasons why these performers are considered very successful. Differences in timing imply specific differences in the performer’s finger movements [177].

The score has four different versions in the clefs of G, F and C and in the neutral clef for percussion instruments with no precise pitch. Other characteristics are the short two
measures. If necessary, a pitch transposition is allowed depending on the musical instrument only in the case where the performer has a disadvantage (hand or arm position, embouchure difficulties, etc.) when performing the score compared to other musicians.

The chosen meter was 4/4 and the selected key was C major. The most common meters in western music are 4/4, 3/4 and 2/4, being even numbered meters [153]. The 4/4 meter melody allows the crotchet to be the beat unity on each bar. Other meters could be very complicated to perform in higher BPM [168]. Normally, the first beat of a metric cycle, the downbeat, is a “strong” beat and therefore accentuated in western European music [231].

The main characteristics and properties of the score can be summarized as follows:

- The score is two measures long and has to be played in a loop until the performance breaks down.
- Duration of musical tones are: crotchet (quarter notes), quaver (eighth notes) and crotchet rests (quarter rests).
- The score contains no ornamentation.
- The time signature is 4/4 and tone onsets are aligned with metric accents. The accentuation is part of an immanent accent. However, the first note of each measure is an eighth note, to restrict the time of accentuation, especially for faster BPMs.
- Any syncopation\(^1\) (offbeat) is unwanted.
- The melodic contour or shape of the melody line [218] contains rises and falls in an alternating manner.
- Rests are present to enable breathing for the performers of aerophones, especially at higher BPMs.
- The tempo varies randomly from 60 to 240 BPM, with seven different tempi separated by 30 BPM intervals.
- A total of 11 notes, including rests, are part of the score. Normally five or six notes are the “eye-hand span” for sight reading [208].

\(^1\) Tone onset aligning with weak metrical accents [218].
- The musical pattern (melody) was created to avoid the exposure effect and therefore avoid any reminiscence of known melodies and familiarity with other patterns. Test subjects may prefer familiarity [122].

- The proposed melody avoids any reminiscence of widely known musical patterns. Nevertheless, associations could arise depending on the performed BPM [131].

- Melodic tension is avoided due to close relationship. Modulations in distant (according to the circle of fifth) notes increase tension [102] between notes.

- All notes should be played non-`legato`\(^2\).

- The melody has no repeating pitch patterns to prevent easily acquired motor movements. However, loop repetition for the whole musical score is unavoidable.

At this point, it is important to clarify that there are a lot of musical considerations, preferences and expectations that are impossible to specify through a score [142]. In addition to the previously described characteristics, it is necessary to consider the following issues with regard to the performance of the score:

- Short breaks are allowed when starting a new trial with a new tempo.

- The duration of the experiment should not exceed 30 minutes. Previous experiments have a similar duration [56].

- The score is not rehearsed but performed “a prima vista”.

The score should be as simple as possible. Different musicians should be able to perform the basic structure.

**Digital audio workstation (DAW) template**

The main stimuli are controlled using a DAW. Requirements of the digital audio workstation (DAW) in order to be part of the equipment of the listening test are:

\(^2\) According to research in piano performances, playing staccato may correspond with up to a 40% of the total IOI of the note [32].
Experimental Procedure

- Constant latency for analogue to digital to analogue conversion.
- A monophonic delay plug-in.
- Automation features for plug-ins.
- MIDI in/out (to control the visual metronome).

The DAW software of choice based on flexibility characteristics and also experience from other research [192] was Cubase Artist [212]. Its delay plug-in proved to be very reliable. In addition, the MIDI set-up and configuration in connection with the automation and the interface control was easier compared to other available DAW systems. The recorded signal from the performance was returned to the musicians through closed headphones. Latency was added to the recorded signal and delivered to the musicians.

A metronome was available as a visual signal displayed in a 7-inch video monitor connected to the laptop through a high-definition multimedia interface (HDMI) cable. The second metronome was an audio signal sent directly to the headphones. The visual metronome signal was generated from the MIDI signal produced by the DAW template and transmitted via HDMI to the monitor. For the aural metronome, the internal metronome was used and is also controlled by the MIDI protocol.

For the recording, a sample rate of 48kHz with a resolution of 24bit was used. The audio interface internal latency was measured to be approximately 14ms for Mac and 12ms for PC using the software SATLive [174] as presented in Appendix E. This is the value $L_c$ introduced in the total latency (Equation 3.1), which is not noticeable during performance by the musicians. The last assumption was based on technical documents [144] and personal communication with Arthur Koll\(^3\).

The DAW template remained the same during all listening tests. The audio devices, microphones, and even cables, were always the same to guarantee similar test conditions. Figure 4.6 is a screenshot of the DAW listening experiment template. The Cubase Artist sequencer and the latency plug are shown during the recording.

\(^3\) Telephone conversation on March 03, 2015. Arthur Koll works as broadcast and pro audio engineer for Yamaha Music Central Europe.
The relevant elements of the template used for the listening test are:

1. Hearing measurement track: in this track the different test frequencies are played at the beginning of the test to test normal hearing.

2. Instrument track: the musical instrument is recorded in this track.

3. Latency track: this is the track were the delay plug-in is inserted and automated.

4. Transport panel: this panel controls the metronome, BPM and position of the cursor in the DAW. Every available DAW has a transport panel.

5. Delay plug-in: this plug-in introduces delays to the recorded signal of the musical instrument that is sent back to the musician to simulate latency.

In Figure 4.6, two control items are essential for running the listening test: the transport panel and the delay plug-in. The transport panel is necessary to introduce the different BPM values, to locate the cursor for the beginning of a new recording and to initiate the recording process. The delay plug-in simulates the latency and displays the latency value. Both control items are presented in Figures 4.7 and 4.8.
Every DAW has a transport panel with essential functions included. For the listening test, the following function were used:

1. Tempo signature: in this slot the different tempi, e.g. 90 BPM, 120 BPM, etc., can be manually introduced.

2. Time signature: for the listening test, the time signature was 4/4.

3. Metronome button: by clicking the button the metronome is activated or deactivated.

4. Main transport locators: the cursor value to begin a new listening test with a new tempo value can be introduced manually in this panel.

5. Main transport record: the record button enables the recording of the listening test. With the play button, audio signals on the track can be listened to.

6. MIDI activity: the bars allow control of the signal of the MIDI metronome (bars and beats).

7. Audio activity: audio input and output levels are monitored in this section.
Figure 4.8 shows the relevant controls of the monophonic delay plug-in used to simulate the network latency.

![Figure 4.8: DAW delay plug-in](image)

Three items of the delay plug in are relevant for the listening test:

1. **Delay**: this is the latency value read by a performance breakdown. The plug-in is automated. During the performance the latency value increases automatically up to value of 300.

2. **Mix**: with this button, the level of latency that the musicians receive in their headphones can be regulated. The mix level is always 100. This means that musicians only listen to the delayed signal from their own musical instrument.

3. **Read & Write**: when R and W are on, as in this case, the automation modus is enabled.

**Metronome set-up for the listening test**

The set-up parameters of the metronome can be introduced in the Cubase DAW software. For the visual metronome, click outputs should be active and send the signal to the MIDI Port (loop MIDI Port) as shown in Figure 4.9. The aural metronome is activated with the control button “Active Audio Click”. The sound signals are beeps. Figure 4.9 shows the configuration for the PC. The same set-up with a slightly different layout is available for Macintosh.
Experimental Procedure

The two sections relevant for the metronome configuration are:

1. Activation of the click and MIDI port. In this section the visual metronome is activated. Bars and beats are different according to note and velocity.

2. Activation of the audio click. In this section the audio clicks (beeps) are activated and the level settings defined.

The visual metronome was never deactivated during the listening test. For the test with only aural metronome, the window displaying the visual metronome was closed in the HD LCD 7-inch monitor.

The slides for “Level” (lower right corner) controls the volume of the audio clicks (beeps). The upper slide control is the bar volume and the lower slide control is the beat volume. During the experiment, bars and beats have the same output level but a different pitch (see Hi and Lo slides). The dB range between the 1/4 slide position (Figure 4.10 far right) translates into a 28.5dB +/- 0.5dB difference as described in Appendix B. The measurements include the beep signals of the metronome. The level differences produced at the headphone by moving the slides are linear.
For the listening test, the metronome headphone level was set to 1/4 as seen in Figure 4.10. Any additional change to the level was done at the audio interface (RME Fireface 400 or RME UCX). The settings in Figure 4.10 are:

1. Activate audio click: this item was deactivated when only the visual metronome was used.

2. Beeps or sounds: beeps were the acoustical signal selected for the aural metronome.

3. Pitch: Hi and Lo pitch for bars and beats respectively.

4. Level: the level output for bars and beats was the same during the listening test.

After measuring different values and averaging the results (see Appendix B), differences of approximately +/- 2dB are present. In other words, it is possible for the experimenter to regulate the audio metronome output of the headphones by 27.26dB +/-2dB. In actual fact, in the experiment, the range used was in between the slide positions 1/2 and 1/4, which corresponds to the value 27.26dB -14.16dB. The result is 13.1dB. Results are available in Appendix B.

**Measurement to determine normal hearing ability**

The test to determine normal hearing ability consisted of a series of sinus tones of different frequencies at 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz with a duration of one second. The tones were played through the headphones in ascending order from low to high frequency with a silence gap of 3s between each tone. The sound level was the same for every test subject and was adjusted to match the threshold at which a healthy person should be able to hear a tone, conforming with the “Normal equal-loudness level contours”
Experimental Procedure

[126]. Figure 4.11 shows the template used for testing normal hearing with six sinus tones at different frequencies and the same output level.

![Figure 4.11: Normal hearing measurement](image)

The task for each musician was to play any note on his/her corresponding instrument as soon as a sinus tone was heard. This allows for testing of hearing ability as well as reaction time. Participants who were not able to hear any of the six different frequencies were discarded as test subjects. All six frequencies were played at 30dBSPL which is audible for any human with healthy ears in accordance to the ISO Standard [126].

4.1.3 Physical Set-Up

The system to collect the data is a digital audio workstation (DAW), a microphone, headphones and a small 7-inch HD LCD monitor. The DAW records the audio signal of the musician’s performance. In addition to recording the signal, the DAW returns the performed signal back to the musician. Musicians listen to the returned signal through headphones. The returned signal is delayed and the delay incremented in 10ms steps from 0 to 300ms. It simulates the latency the musicians have to cope with. The DAW generates the time control system, in this case, the metronome.

The experimental set-up used in the pilot test consisted of a small condenser diaphragm microphone (Sennheiser MKH40) with a cardioid pattern. This microphone has been rated as a good choice to capture the instrument’s sound nature [84] in other research projects. The audio interface, an RME Fireface 400 connected to a Mac Book Pro, was used for the pilot test. For the final test, an RME UCX connected to a Think Pad PC, provided the recording set-up.
Both audio engine systems (MAC and PC) delivered a good performance during the listening pilot and final test. Since 2001 the Apple MacOS system (Core Audio API) has been very efficient with regard to low latency performance [155]. Similar to the pilot test set-up, the PC system used for the final listening test also performed well.

No extra internal latency or only very low latency added to the listening test is expected via Firewire 400 or USB2 when using both systems (see Appendix E). Audio hardware and microphone characteristics were similar to those used in previous research [84]. An additional microphone, a no-brand cardioid pattern microphone, was used as a talk-back to communicate directly with the musicians. This microphone has no added latency and its function was to facilitate the direct answering of any possible questions or to indicate when a trial was going to begin.

The final equipment set-up is presented in Figure 4.12.

![Figure 4.12: Equipment set-up for the final listening experiment](image)

The equipment for the final listening test included:

1. Computer with a DAW.
Experimental Procedure

2. Audio interface.

3. Condenser microphone.

4. Talk-back microphone.

5. Passive noise reduction headphones

6. HD-LCD 7-inch monitor.

7. Stand for the 7-inch monitor.

8. Microphone stand for the condenser microphone.


10. Sound pressure level meter\(^4\).

Recording the musical instruments

From the mathematical expression defining total latency \( L_t \) (Equation 3.1), \( L_a \) is the latency added due to sound transmission through the air. Normally, in recording situations a distance of 0.30 to 0.40m is standard between microphone and instrument. Adequate placement of the transducers lowers the in-air sound propagation delay [198]. The recording distances between microphone and instrument produced a delay or latency of approximately 0.88 to 1.17ms, a latency value which is hardly perceived by humans [17]. For this experiment, the distance from the microphone to the instrument can be considered a constant. Variations of some centimetres between microphone and instrument should not change the value of \( L_a \) significantly, if at all, and only in the order of microseconds. Therefore, \( L_a \) can be assumed to be constant. To ensure equal condition for the participants, the sound pressure level (SPL) of every musical instrument and the distance to the microphone were measured before the experiment began as shown in Appendix F.

Some basic aspects were considered when miking the musical instrument. Figure 4.13 shows the recording set-up of four specific miking situations for the instruments piano, snare drum, double bass and harp.

\(^4\) Phonic PAA3 [189].
Figure 4.13: Recording set-up (a) piano (b) snare drum (c) double bass (d) harp

Data of microphone distance, musical instrument sound pressure level, microphone and headphones obtained for every recorded instrument is presented in Appendix F. Table 4.1 is a statistical summary of the ranges, differences, median, mean and standard deviation for the different values measured in Appendix F.
**Experimental Procedure**

<table>
<thead>
<tr>
<th>Microphone distance (m)</th>
<th>Instrument dB SPL (A)</th>
<th>Microphone gain (dB)</th>
<th>Headphones gain (dB)</th>
<th>RT60 (sec)</th>
<th>Room dB SPL (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
</tbody>
</table>

| nbr.val     | 31 | 31 | 31 | 31 | 31 | 31 |
| min         | 0.23 | 75.10 | 0.00 | -10.00 | 0.30 | 31.20 |
| max         | 0.71 | 96.40 | 12.00 | 10.00 | 1.64 | 38.70 |
| range       | 0.48 | 21.30 | 12.00 | 20.00 | 1.34 | 7.50 |
| median      | 0.39 | 83.40 | 10.00 | -5.00 | 0.50 | 33.50 |
| mean        | 0.41 | 85.40 | 6.52 | -3.23 | 0.51 | 33.62 |
| var         | 0.01 | 29.48 | 24.26 | 57.58 | 0.06 | 5.82 |
| std.dev     | 0.11 | 5.43 | 4.93 | 7.59 | 0.25 | 2.41 |

Table 4.1: Statistical summary for measurements related to the listening test

Mean and median are very similar. The low standard variation, which is a measure that indicates the average deviation from the mean [93], shows the similarity of room and instruments in dB SPL(A) values. The low value facilitates and enables a comparison of the results between instruments. The microphone distance to the different instruments varies only in the range of 48 cm. Measurements shown in Appendix F and their statistical description in Table 4.1 are the numerical evidence that allow result comparisons for the final tests.

**Measurements in instruments**

The sound pressure level (SPL) is a reliable way of measuring the physical characteristics of a musical instrument. The dB(A) weighting scale, which is used for lower pressure levels, was chosen. This scale approximates the 30 phon equal-loudness [22] and supports the most salient features of the characteristics of the ear [81].

Table 4.2 is a summary of the sound pressure level for the musical instruments of the listening test. The SPL range value is only an approximation and depends strongly on the measured tones and the velocity of the musical performance [165].
Experimental Procedure

<table>
<thead>
<tr>
<th>Instrument</th>
<th>SPL range (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violin</td>
<td>58 - 94</td>
</tr>
<tr>
<td>Cello</td>
<td>63 - 98</td>
</tr>
<tr>
<td>Double bass</td>
<td>74 - 91</td>
</tr>
<tr>
<td>Classical guitar*</td>
<td>37 - 88</td>
</tr>
<tr>
<td>Piano</td>
<td>88 - 104</td>
</tr>
<tr>
<td>Piano (upright)*</td>
<td>86 - 94</td>
</tr>
<tr>
<td>Harp</td>
<td>88 - 100</td>
</tr>
<tr>
<td>Transverse flute</td>
<td>67 - 86</td>
</tr>
<tr>
<td>Bassoon</td>
<td>72 - 102</td>
</tr>
<tr>
<td>French horn</td>
<td>86 - 107</td>
</tr>
<tr>
<td>Trumpet in B</td>
<td>89 - 104</td>
</tr>
<tr>
<td>Trombone</td>
<td>89 - 105</td>
</tr>
<tr>
<td>Alto saxophone*</td>
<td>70 - 90</td>
</tr>
<tr>
<td>Tenor saxophone*</td>
<td>70 - 92</td>
</tr>
<tr>
<td>Snare drum</td>
<td>74 - 100</td>
</tr>
<tr>
<td>Timpani</td>
<td>67 - 115</td>
</tr>
<tr>
<td>Marimba*</td>
<td>75 - 95</td>
</tr>
</tbody>
</table>

Table 4.2: Sound pressure level range for western musical instruments

The values obtained by Jürgen Meyer [165] for the instruments listed are only an approach under specific conditions. SPL measurements are very variable and dependent on the velocity of the tone and the acoustics of the performing room (see Appendix F). The data in Table 4.2 was measured while playing fast scales across two octaves. The dynamic level corresponded to the softest *pp* (*pianissimo*) and the loudest *ff* (*fortissimo*). Instruments with (*) were measured by the author with an SPL meter at 1m distance using the A filter, dBSPL(A).

The instrument input level of the preamplifier in the audio interface was adjusted for every instrument according to prior testing and taking into account the sound pressure level of the instrument. Every listening test and its values were documented.

The aim of the listening test was to ensure uniformity of testing conditions. When measuring similar SPL dB(A) in the musical instruments, values with a tolerance within +/-10dB can be assumed to be similar. Furthermore, it is possible to level the gain input
and headphones output of the audio interface when necessary, depending on the sound pressure level of the instrument. For instance, it was clear that a triangle performer needed more headphone and microphone level than a trumpet performer for acoustical reasons, regardless of the dynamic characteristics of the score.

**Headphones**

For the delayed audio signal, isolation headphones by Vic Firth SIH1 with 24dB (manufacturer information) \[10\] and passive noise reduction were used to minimize the effect of room acoustics and external sounds. The actual measured value, which is around 10dB, was estimated in Appendix C. The headphones are designed to reduce the level of external sound. In addition, the Vic Firth SIH1 stereo headphones reduce ambient noise from instruments \[214\], which traduces in a control of the environment and room acoustics for the listening test.

The circum-aural (around the ear) closed earphones \[10\] selected for the listening experiment enclose the pinna surrounding the surface of the head \[22\]. The headphones are also known as supra-aural earphones. The pinna is pressed by the earphone and the transducer is close to the pinna \[1\].

Direct contact with the head occurs via compliant cushions. The foam cushions automatically self-align for better comfort \[214\]. The selected earphones touch the pinna without compressing it. These headphones deliver sound isolation and at the same time the sound produced by the instrument can still be heard by the performer. The headphones enabled the sending of their own delayed feedback to musicians reducing the volume of the acoustical non-delayed original sound.

It is also a fact that the air distance introduces additional latency due to the transmission path between the microphone and the reproduction channel, i.e. loudspeaker and the ears \[51\]. However, this delay is even smaller than the delay introduced by air propagation between instrument and microphone. The listening experiment auditory stimuli were monaural. The same audio signal was sent to each headphone ear channel. Other audio signal configurations such as stereo, binaural or dichotic (different signals to each ear), were not considered. Mono signals provided the necessary amount of information for the test subjects. Studies have shown that the monaural presentation process does not
deteriorate with age in comparison with the binaural processing [215].

The decision to use isolation circumaural headphones instead of noise-reducing headphones was based on the technical characteristics of the latter. Noise-reducing headphones with active noise cancellation could dramatically change the delayed signal or even cancel it. With noise-reducing headphones, a microphone picks up the local acoustic background noise. This background noise is phase reversed and added to the original signal, as a result, the background noise is eliminated [121]. These headphones may also add extra nonlinear latency due to signal processing. On the other hand, passive noise cancellation headphones are intrusive and may disturb the performance. With these headphones, the ear canal is blocked reducing the acoustic background level that enters the ear canal [121]. In order to prevent any hearing damage, special care was taken regarding the latency signal returned to the headphones prior to every test. Signal return level at the audio interface was controlled by avoiding loud audio signals. As stated before, test subjects were allowed to take breaks whenever they felt necessary.

The headphone level was controlled through the “HP level” on the audio interface which establishes the level for the delayed audio signal and for the audio metronome. The audio metronome level can also be controlled through the slides in the section “Metronome Set-up” of the DAW as shown in the Appendix B.

The use of isolation headphones enabled the listening test to be conducted in different rooms or enclosed spaces. However, it was necessary to provide a noise-free environment in order to eliminate possible measurement bias. According to the recommendations of the American National Standard included in the ‘‘Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms’’ [1], avoiding all ambient noise in a room is not possible. Nevertheless, there are maximum permissible ambient noise levels. Measurements of ambient noise below the references given in the standard provide a good means of identifying a test room as suitable for the listening test. Appendix F presents a list of all the rooms where the different instruments were recorded. All room SPL measurements were under 40dB(A). Information on room dimensions, reverb time (RT60) and sound level pressure (dB SPL) is listed for every single room.
4.1.4 Procedure

Latency was introduced to the musicians through headphones and perceived as delay. This latency ($L_d$), which is the variable delay in the total latency equation ($L_t$), is generated progressively in increments of 10ms beginning with no latency (delay 0ms) up to 300ms. While there is no real 0ms latency, this refers to the absence of delay in the DAW template. In other words, the delay plug-in is bypassed in the template $L_d=0\text{ms}$. The only existing latencies were the computer and audio interface latency ($L_c$) and the latency produced due to air transmission ($L_a$). Both $L_c$ and $L_a$ are considered constants.

The audio interface latency value was less than 15ms for both the Mac and PC. A description of the method used to establish the audio interface latency value, the double FFT (Fast Fourier Transform) method, is described in Appendix E.

A simple delay plug-in was inserted in the audio channel of the Cubase DAW template that records the microphone signal. By means of automation, it was possible to generate a template that increases the delay (latency) in 10ms steps continuously after every 4 bars. The musician played the 2-bar score. In the first pass, the latency was set to 0ms. Because the score is a loop, the first repetition was also with a $L_d$ value of 0ms. At the moment when the second repetition of the score began, latency was increased by 10ms for each second repetition, up to an amount of 300ms. It was expected that the performance would be disrupted before the value of $L_d=300\text{ms}$ was achieved.

Running the test

Before performing the listening test, random values were generated for the five different BPM tempi using the software MATLAB and its function randperm. The room was measured with respect to noise levels. A measurement with a sound pressure level was enough to test if the room was suitable for running the listening test. Any room under 40dB-SPL(A), which is the level that equals a library or a quiet residence [89, 170], is considered sufficient to run the test. As stated before, it was impossible to avoid all noise in a room [1].
The following steps were followed for the listening test:

1. An introductory explanation for the test subject was handed out. Information regarding participation, test duration, test procedure and data gathering for the listening test, listening test information and listening test consent was presented before the listening test (see Appendix G and H). The test subject has to introduce the categorical and numerical data in the questionnaire (questions 1 to 9 see Appendix M).

2. The musician played in the standing or sitting position in front of the microphone. The microphone was previously placed in the best position according to the instrument sound radiation pattern and the sound pressure level (see Appendix F).

3. Sound pressure level (SPL) and the distance between instrument and microphone were measured and written down as presented in the listening test results sheet of Appendix M. In addition, the input level of the audio interface and the headphones gain were both measured (in dB).

4. The first part of the test was the measurement to determine normal hearing ability in order to discard any test subject with hearing impairments.
5. Participants were required to play the score as seen on the HD LCD monitor. The test for each tempo was finished in the exact moment of a performance breakdown\(^5\) as explained in item four of the test procedure in Appendix G. The screen was placed in such a way that sound reflections of the instrument were avoided. The sound emitted from the musical instrument was tracked and recorded while being performed. This audio signal was returned directly to the musician’s headphones. Participants were asked if the headphone levels and set-up were optimal. One of the three metronomes was always active during the performance.

6. A different metronome was used for each of the three trials using the following order:
   - An aural metronome (headphones).
   - A visual metronome (HD LCD monitor).
   - Both metronomes at the same time (HD LCD monitor and headphones).

   Each trial consisted of seven attempts: the different BPM values from 60 BPM to 240 BPM in 30 BPM increments presented in a randomized order. These were the levels of the factor tempo.

7. During the test, the automated DAW template ran independently and increased the latency time every 4 bars (every second score repetition). After a solo performer breakdown, the latency value was notated. The test was concluded for that attempt. Each new trial with a new tempo was announced through the talk-back microphone.

8. After finishing the experiment, musicians answered the last question (see Appendix M), the recording was re-listened to and values were confirmed.

9. Headphones were always cleaned after every listening test.

The audio signal produced by the musicians and captured by the microphone during the listening test was always recorded for further data evaluation and to minimize measurement errors produced throughout the process of writing up the data.

\(^5\) A performance breakdown or performance disruption is when the musician stops playing the musical instrument as defined in the Methodology chapter.
4.2 Pilot Test

In order to test the methodological approach, a pilot test was carried out to obtain a first set of results and highlight any potential flaws in the design of the experiment. Test subjects for the pilot test included 10 musical performers testing 12 different musical instruments types including chordophones, membranophones and aerophones.

The first goal of the pilot test was to look ahead for possible issues or problems during testing. The set-up is the same as described in the methodology. Listening sessions were 30 minutes long.

As a result of the pilot test, two methodology issues were found. Any changes resulting from these issues were effectively implemented for the main test.

- The latency values obtained for tempo 60 BPM were considered too ambiguous. For further listening experiments the minimum tempo value was set at 90 BPM.

- The latency values obtained for tempo 240 BPM were contradictory. For some instrumentalists it was anatomically difficult to play this fast. The top musical tempo value for further listening experiments was set to be 210 BPM.

The ambiguity of results for the 60 BPM tempo probably lies in the internal division time (mental division) that some test subjects might use for this very slow tempo. As a matter of fact, after the listening test and personal communication, some test subjects said that they played at 60 BPM as if it were at 120 BPM. With regard to the faster tempo of 240 BPM, it was impossible for some musicians to even start the performance at this level and their playing was based not on musical reading but on memory.

4.2.1 Data Collection

Each test included answering the questions in the questionnaire (Table 3.1) relating to the practice of music as well as several trials with different musical tempo values. The number of trials depended on the amount of time necessary to collect the data. A maximum of 30 minutes per experiment was recommended based on previous research [56, 22] to avoid test subject fatigue. Data was transcribed manually after reading the latency value results and performance breakdown from the digital audio workstation.
4.2.2 Data Analysis

For every musical tempo there were three trials, one trial per metronome: aural, visual or both. The results of each trial included a time value, i.e. the latency value in milliseconds. The arithmetical mean or average of the latency value of the three trials for each BPM is the value used in the data analysis. Collecting three different time readings enabled the comparison of significant differences between latency values for the same musical tempo. It could indicate discrepancies in the measurement method.

After averaging each latency time value for the different musical tempi, it was possible to plot a chart showing latency time for each musical tempo and instrument group as in Figure 4.16. For the pilot test, the only relevant statistical measure was the arithmetic mean of the latency values obtained. Being a pilot test, the data of a total of 12 musical instruments was gathered. Any statistical analysis is only descriptive. However, it was possible to estimate possible trends and results on the basis of the information obtained.

Pilot test results description

With all test subjects undergoing the same test procedure, results were analysed in the open source software environment for statistical computing named R, using R Studio as an integrated development environment (IDE), thus enabling source code editing, debugging and automation. Ten subjects playing different musical instruments were tested. Table 4.3 is a summary of the musical instruments and the associated instrument groups of the test subjects evaluated.

<table>
<thead>
<tr>
<th>Instrument group</th>
<th>Musical instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerophones</td>
<td>Alto saxophone, Tenor saxophone, Transverse flute, Trumpet in B</td>
</tr>
<tr>
<td>Chordophones</td>
<td>Violin, Classical guitar, Acoustic guitar</td>
</tr>
<tr>
<td>Membranophones</td>
<td>Snare drum</td>
</tr>
</tbody>
</table>

Table 4.3: Pilot test musical instruments
Table 4.4 shows the latency values and performance breakdown for the different instruments that were tested. It is important to highlight the fact that NA refers to “not available” values. There are different reasons for this. For instance, the first listening tests with violins did not take into account higher BPM values. For the first tests, only aural or visual metronomes were tested. The last test subjects were the only subjects to undergo three different measurements with the aural, visual and aural-visual metronomes.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Group</th>
<th>Metronome</th>
<th>60 BPM</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
<th>240 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violin</td>
<td>Chordophones</td>
<td>Aural</td>
<td>195</td>
<td>150</td>
<td>120</td>
<td>90</td>
<td>80</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Violin</td>
<td>Chordophones</td>
<td>Aural</td>
<td>210</td>
<td>150</td>
<td>150</td>
<td>120</td>
<td>120</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Aural</td>
<td>180</td>
<td>160</td>
<td>140</td>
<td>110</td>
<td>90</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Aural</td>
<td>180</td>
<td>170</td>
<td>130</td>
<td>120</td>
<td>90</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Visual</td>
<td>165</td>
<td>95</td>
<td>110</td>
<td>80</td>
<td>85</td>
<td>80</td>
<td>75</td>
</tr>
<tr>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Visual</td>
<td>85</td>
<td>90</td>
<td>100</td>
<td>95</td>
<td>95</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Aural</td>
<td>100</td>
<td>100</td>
<td>110</td>
<td>90</td>
<td>80</td>
<td>80</td>
<td>75</td>
</tr>
<tr>
<td>Trumpet in B</td>
<td>Aerophones</td>
<td>Aural</td>
<td>120</td>
<td>160</td>
<td>130</td>
<td>110</td>
<td>75</td>
<td>60</td>
<td>55</td>
</tr>
<tr>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Aural</td>
<td>200</td>
<td>140</td>
<td>140</td>
<td>130</td>
<td>120</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Visual</td>
<td>230</td>
<td>180</td>
<td>130</td>
<td>120</td>
<td>130</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Aural</td>
<td>190</td>
<td>180</td>
<td>140</td>
<td>140</td>
<td>130</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Aural</td>
<td>160</td>
<td>120</td>
<td>85</td>
<td>75</td>
<td>60</td>
<td>50</td>
<td>NA</td>
</tr>
<tr>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Visual</td>
<td>120</td>
<td>100</td>
<td>50</td>
<td>65</td>
<td>20</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Tenor saxophone</td>
<td>Aerophones</td>
<td>Aural</td>
<td>250</td>
<td>125</td>
<td>100</td>
<td>90</td>
<td>80</td>
<td>65</td>
<td>50</td>
</tr>
<tr>
<td>Transverse flute</td>
<td>Aerophones</td>
<td>Aural</td>
<td>170</td>
<td>140</td>
<td>90</td>
<td>65</td>
<td>65</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Visual</td>
<td>300</td>
<td>175</td>
<td>145</td>
<td>280</td>
<td>45</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Visual</td>
<td>300</td>
<td>240</td>
<td>300</td>
<td>230</td>
<td>215</td>
<td>140</td>
<td>165</td>
</tr>
<tr>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Aural</td>
<td>300</td>
<td>260</td>
<td>300</td>
<td>290</td>
<td>210</td>
<td>210</td>
<td>150</td>
</tr>
<tr>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Both</td>
<td>80</td>
<td>125</td>
<td>100</td>
<td>110</td>
<td>110</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Visual</td>
<td>60</td>
<td>100</td>
<td>110</td>
<td>90</td>
<td>115</td>
<td>80</td>
<td>140</td>
</tr>
<tr>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Aural</td>
<td>100</td>
<td>165</td>
<td>120</td>
<td>110</td>
<td>95</td>
<td>75</td>
<td>35</td>
</tr>
<tr>
<td>Acoustic guitar</td>
<td>Chordophones</td>
<td>Both</td>
<td>170</td>
<td>185</td>
<td>220</td>
<td>110</td>
<td>70</td>
<td>55</td>
<td>35</td>
</tr>
<tr>
<td>Acoustic guitar</td>
<td>Chordophones</td>
<td>Visual</td>
<td>110</td>
<td>165</td>
<td>140</td>
<td>290</td>
<td>260</td>
<td>NA</td>
<td>60</td>
</tr>
<tr>
<td>Acoustic guitar</td>
<td>Chordophones</td>
<td>Aural</td>
<td>130</td>
<td>175</td>
<td>300</td>
<td>190</td>
<td>130</td>
<td>115</td>
<td>65</td>
</tr>
</tbody>
</table>

Table 4.4: Latency values for the pilot test

The values from Table 4.4 were used to generate the first latency ($L_d$) values, providing information about the possible tendencies that could be expected. As stated before, the data was insufficient for making any statistical assumptions.
Experimental Procedure

Pilot test forecast and trends

The preliminary outcomes confirmed the results of previous research with respect to the inverse relationship between latency and tempo [18] which can be defined as latency adaptive tempo (LAT), see Figure 4.15. Figure 4.16 demonstrates how a slower tempo resulted in greater latency values for all of the instrument groups.

Figure 4.15: Latency ($L_d$) vs. tempo in BPM

Figure 4.16: Latency ($L_d$) vs. instrument groups
Figure 4.17 presents a comparison of the latency values ($L_d$) for chordophones and aerophones at different tempi. The range under the curve is where the instrument is playable without a breakdown in the performance. Note the different ranges with respect to the different instrument groups.

These previous results allowed a general forecast of the possible outcomes. As shown in Figure 4.16 and Figure 4.17, different instrument groups appear to have different latency breakdown values. Musicians playing chordophones were able to tolerate increased latency at the same musical tempo than musicians playing aerophones. It is, however, necessary to test more instruments in order to make general assumptions about the behaviour of the different musical instrument groups.

4.2.3 Interpretation of the Results

The majority of the musicians had to undergo the listening test three times, each time with a different metronome (aural, visual and both). At the end of the experiment, 21 latency values per instrument were manually noted: three values per trial and seven different trials according to each musical tempo from 60 to 240 BPM.

The audio data of every session was recorded in the audio format BWF for further analysis and to verify the manually noted values. The broadcast wave format is a commonly used
audio format for broadcast and sound studio distribution without any data compression and has the data extension .wav [229].

Avoiding the first 60 BPM measurement was found to be convenient for the final listening test. Musicians rated 60 BPM as too slow in the first trials, and this value produced very different results between the musicians (aerophones), even when playing the same instrument. Figure 4.17 clearly shows the widespread latency tolerance values (more than 40ms) for the tempo 60 BPM.

From musical research it is known that musicians subdivide time into shorter intervals when time intervals are slower than 60 BPM [17]. Research done by Halsband et al. [115], showed that there is evidence of spontaneous grouping in piano players when performing quarter note pulse-beat patterns. Subjects can reprogram these structures.

It was noted that some test subjects played even shorter intervals for tempi up to 90 BPM. The measurement of latency in 60 BPM displayed different values within the three measurements.

Tempi up to 600 BPM are common in amadinda xylophone music of Uganda, and in Western music such as bebop jazz, where tempi of 240 to 300 BPM are normal [17]. However, the tempo value of 240 BPM was avoided in further experiments. Musicians had constant difficulties performing at this tempo. Even when using the metronome, contradictory values were obtained for very high and very low tempi (60 and 240 BPM respectively) in the pilot test.

Another necessary correction was the elimination of question 6 in the questionnaire (Table 3.1). Musicians have difficulty in answering questions relating to defining the expertise level between amateur and professional. For the main test, it was considered sufficient to classify musicians as either professional musicians or students.

Two new questions were added to the questionnaire. The first related to metronome affinity, i.e. how proficient the musician is with respect to performance with the help of a metronome. It is important to establish a possible bias in the results due to previous experience with the musical instrument when using a metronome. The second question related to proficiency in more than one instrument. Some musicians may be proficient in
more than one instrument and this information is also important when determining possible outliers. In the questionnaire there are no open questions\(^6\). Item 11 (Notes), however, enables an open answer from the musician with regard to the listening test. The answers may help to understand the personal views and the listening test experiences from the musician’s perspective. The final questionnaire is found in Appendix I.

As a result of the pilot test, some flaws were detected and corrected for the final version of the listening test. Musicians mostly preferred the aural metronome over the visual metronome. However, no changes in relation to this issue are necessary. There is no significant influence of the metronome (visual or aural) relative to the latency values obtained.

4.2.4 Adjustments and Modifications

The following list presents a summary of the improvements and adjustments necessary for some items of the pilot test in order to generate a better main listening test:

- Musical tempi tested were changed to start from 90 BPM up to 210 BPM in 30 BPM intervals.
- The question about defining the musical level of expertise between amateur and professional was withdrawn from the questionnaire.
- A question on the performance of additional musical instruments was introduced.

4.3 Experiment

After evaluating the pilot test, necessary adjustments were introduced for the final test. The set-up was mainly the same as that used in the pilot test except for the audio interface (RME Fireface UCX) and the computer (ThinkPad PC). The audio interface settings for the new interface matched the results obtained with the RME Fireface 400 (see Appendix D). For the PC, the virtual MIDI software loopMIDI was used (see Appendix A).

\(^6\) The methodological approach is only quantitative. The questionnaire is only to gather numerical and categorical data.
4.3.1 Data Collection

Data was gathered throughout the duration of the experiment. As the musicians performed the instrument, the experimenter noted the latency values of performance breakdown in a template (see Appendix N). Afterwards, the recorded data was re-listened to compare results. The numerical values obtained were rounded off to numbers without a decimal separator and transferred to a data-sheet table. The experimenter remained in the same room as the test subject throughout the experiment. Musicians did not have direct eye contact with the experimenter. However, the experimenter was able to observe every test subject and their performance. This made it possible to reject some of the subjects, for e.g. those subjects that did not perform the music using the score but closed their eyes and learned the score by heart.

“Not available” values (NA) were introduced in the final data table when the following issues occurred:

1. Incorrect notes were performed throughout the performance. From the literature review [168], it is known that, with latency, musicians perform the wrong notes. Thus, where only false notes were performed throughout the attempt, the noted value was NA.

2. Performing the entire test at the wrong tempo. It is also well known that variations in tempo are to be expected due to latency [94, 106, 116]. However, if the tempo is wrong throughout the trial, the notated value was NA.

3. If a rest during the performance lasted more than 8 bars (4 times the score). In this case, the noted value was also NA.

Each of the 18 variables of the experiment, both numerical and categorical, were introduced in the columns of the data-sheet. The observations, in this case the test subjects, are the rows. Each test subject/instrument has three rows, each row for the result of each metronome measurement (aural, visual and both). Besides numerical data obtained during the listening test, categorical data was gathered from the answers to the questionnaire.
Tables 4.5 and 4.6 summarize the information from the listening test and the questionnaire. Answers to questions 6 to 10 provide the core of the numerical data for the experiment. Information gathered through questions 1 to 5 will help to define correlations between musical instruments and the latency values obtained.

<table>
<thead>
<tr>
<th>Item</th>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>participant</td>
<td>numerical</td>
<td>number of the participant</td>
</tr>
<tr>
<td>2</td>
<td>age</td>
<td>numerical</td>
<td>participant age</td>
</tr>
<tr>
<td>3</td>
<td>metronome hours</td>
<td>numerical</td>
<td>weekly hours of musical practice using metronome</td>
</tr>
<tr>
<td>4</td>
<td>years of experience</td>
<td>numerical</td>
<td>number of years playing the instrument</td>
</tr>
<tr>
<td>5</td>
<td>hours of practice</td>
<td>numerical</td>
<td>weekly hours of musical practice</td>
</tr>
<tr>
<td>6</td>
<td>90 BPM</td>
<td>numerical</td>
<td>breakdown latency value for the tempo 90 BPM ($L_d$)</td>
</tr>
<tr>
<td>7</td>
<td>120 BPM</td>
<td>numerical</td>
<td>breakdown latency value for the tempo 120 BPM ($L_d$)</td>
</tr>
<tr>
<td>8</td>
<td>150 BPM</td>
<td>numerical</td>
<td>breakdown latency value for the tempo 150 BPM ($L_d$)</td>
</tr>
<tr>
<td>9</td>
<td>180 BPM</td>
<td>numerical</td>
<td>breakdown latency value for the tempo 180 BPM ($L_d$)</td>
</tr>
<tr>
<td>10</td>
<td>210 BPM</td>
<td>numerical</td>
<td>breakdown latency value for the tempo 210 BPM ($L_d$)</td>
</tr>
</tbody>
</table>

Table 4.5: List of numerical variables

The categorical variables presented in Table 4.6 will help to subdivide musical instruments into different classifications such as musical instrument group and sound generation methods. Further information regarding gender and expertise as well as information about the preferred metronome may simplify the categorization of the information gathered. The variables describe general information relating to the test subjects and the musical instrument performed. This information enables a better analysis of the numerical data.
<table>
<thead>
<tr>
<th>Item</th>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>instrument</td>
<td>categorical</td>
<td>musical instrument performed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>musical instrument group:</td>
</tr>
<tr>
<td>12</td>
<td>group</td>
<td>categorical</td>
<td>1. Chordophones, 2. Aerophones</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Membranophones, 4. Idiophones</td>
</tr>
<tr>
<td>13</td>
<td>gender</td>
<td>categorical</td>
<td>participant gender:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Male 2. Female</td>
</tr>
<tr>
<td>14</td>
<td>expertise</td>
<td>categorical</td>
<td>participant expertise:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1. Student or 2. Professional</td>
</tr>
<tr>
<td>15</td>
<td>sound</td>
<td>categorical</td>
<td>instrument sound generation:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Idiophones: a. Struck</td>
</tr>
<tr>
<td>16</td>
<td>metronome</td>
<td>categorical</td>
<td>preferred metronome in the experiment:</td>
</tr>
<tr>
<td></td>
<td>preference</td>
<td></td>
<td>1. Aural, 2. Visual or 3. Both</td>
</tr>
<tr>
<td>17</td>
<td>metronome</td>
<td>categorical</td>
<td>metronome used in the experiment:</td>
</tr>
<tr>
<td>18</td>
<td>other</td>
<td>categorical</td>
<td>other musical instruments performed besides the main instrument</td>
</tr>
<tr>
<td></td>
<td>instrument</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6: List of categorical variables

The final data is presented in Table 4.7, with results in milliseconds. The $(L_d)$ data obtained for each of the three metronomes (aural, visual and both) was averaged. The table summarizes the information regarding the musical instruments, the group and the sound generation method. The table presents a definitive list of subjects. For the results, note that the piano and piano (upright) are considered the same type of musical instrument. This assumption is based on the dBSPL(A) measurements as listed in Appendix F and on the fact that both pianos have the same keyboard and were played with one hand.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Instrument</th>
<th>Instrument group</th>
<th>Sound generation</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Piano</td>
<td>Chordophones</td>
<td>Struck</td>
<td>216.00</td>
<td>172.33</td>
<td>186.33</td>
<td>181.00</td>
<td>145.33</td>
</tr>
<tr>
<td>2</td>
<td>Piano</td>
<td>Chordophones</td>
<td>Struck</td>
<td>174.67</td>
<td>178.00</td>
<td>162.67</td>
<td>132.00</td>
<td>100.33</td>
</tr>
<tr>
<td>3</td>
<td>Cello</td>
<td>Chordophones</td>
<td>Bowed</td>
<td>280.00</td>
<td>241.67</td>
<td>132.33</td>
<td>97.67</td>
<td>71.00</td>
</tr>
<tr>
<td>4</td>
<td>Cello</td>
<td>Chordophones</td>
<td>Bowed</td>
<td>234.00</td>
<td>227.33</td>
<td>218.67</td>
<td>166.00</td>
<td>195.33</td>
</tr>
<tr>
<td>5</td>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Plucked</td>
<td>131.00</td>
<td>100.33</td>
<td>94.00</td>
<td>96.33</td>
<td>61.67</td>
</tr>
<tr>
<td>6</td>
<td>French horn</td>
<td>Aerophones</td>
<td>Lip reed</td>
<td>300.00</td>
<td>272.33</td>
<td>158.67</td>
<td>107.67</td>
<td>91.00</td>
</tr>
<tr>
<td>7</td>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
<td>300.00</td>
<td>259.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>8</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
<td>288.67</td>
<td>254.33</td>
<td>300.00</td>
<td>288.67</td>
<td>234.33</td>
</tr>
<tr>
<td>9</td>
<td>Trumpet in B</td>
<td>Aerophones</td>
<td>Lip reed</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>10</td>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Struck</td>
<td>145.00</td>
<td>137.33</td>
<td>93.50</td>
<td>63.33</td>
<td>77.67</td>
</tr>
<tr>
<td>11</td>
<td>Piano (upright)</td>
<td>Chordophones</td>
<td>Struck</td>
<td>264.00</td>
<td>244.00</td>
<td>176.00</td>
<td>153.67</td>
<td>129.67</td>
</tr>
<tr>
<td>12</td>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Struck</td>
<td>71.00</td>
<td>74.67</td>
<td>67.33</td>
<td>60.00</td>
<td>71.00</td>
</tr>
<tr>
<td>13</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
<td>223.33</td>
<td>177.67</td>
<td>180.33</td>
<td>137.67</td>
<td>114.33</td>
</tr>
<tr>
<td>14</td>
<td>Transverse flute</td>
<td>Aerophones</td>
<td>Air reed</td>
<td>113.00</td>
<td>86.00</td>
<td>78.00</td>
<td>76.67</td>
<td>19.00</td>
</tr>
<tr>
<td>15</td>
<td>Trombone</td>
<td>Aerophones</td>
<td>Lip reed</td>
<td>123.33</td>
<td>86.33</td>
<td>97.00</td>
<td>63.00</td>
<td>NA</td>
</tr>
<tr>
<td>16</td>
<td>Timpani</td>
<td>Membranophones</td>
<td>Struck</td>
<td>91.33</td>
<td>131.00</td>
<td>96.00</td>
<td>48.00</td>
<td>22.00</td>
</tr>
<tr>
<td>17</td>
<td>Trombone</td>
<td>Aerophones</td>
<td>Lip reed</td>
<td>177.67</td>
<td>116.00</td>
<td>88.67</td>
<td>79.50</td>
<td>73.33</td>
</tr>
<tr>
<td>18</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
<td>280.00</td>
<td>266.33</td>
<td>177.00</td>
<td>166.67</td>
<td>132.00</td>
</tr>
<tr>
<td>19</td>
<td>Triangle</td>
<td>Idiophones</td>
<td>Struck</td>
<td>244.67</td>
<td>227.50</td>
<td>131.50</td>
<td>60.50</td>
<td>45.00</td>
</tr>
<tr>
<td>20</td>
<td>Tenor saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
<td>110.33</td>
<td>120.33</td>
<td>103.67</td>
<td>101.00</td>
<td>76.67</td>
</tr>
<tr>
<td>21</td>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Plucked</td>
<td>277.33</td>
<td>264.00</td>
<td>248.00</td>
<td>192.00</td>
<td>172.33</td>
</tr>
<tr>
<td>22</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
<td>183.67</td>
<td>181.00</td>
<td>143.00</td>
<td>117.33</td>
<td>109.00</td>
</tr>
<tr>
<td>23</td>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Struck</td>
<td>73.00</td>
<td>117.33</td>
<td>84.50</td>
<td>104.67</td>
<td>94.00</td>
</tr>
<tr>
<td>24</td>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
<td>132.00</td>
<td>182.67</td>
<td>78.00</td>
<td>128.33</td>
<td>100.00</td>
</tr>
<tr>
<td>25</td>
<td>Triangle</td>
<td>Idiophones</td>
<td>Struck</td>
<td>178.33</td>
<td>169.33</td>
<td>142.67</td>
<td>154.00</td>
<td>126.50</td>
</tr>
<tr>
<td>26</td>
<td>Marimba</td>
<td>Idiophones</td>
<td>Struck</td>
<td>242.33</td>
<td>228.67</td>
<td>175.00</td>
<td>166.00</td>
<td>179.00</td>
</tr>
<tr>
<td>27</td>
<td>French horn</td>
<td>Aerophones</td>
<td>Lip reed</td>
<td>262.67</td>
<td>175.00</td>
<td>228.00</td>
<td>209.67</td>
<td>140.33</td>
</tr>
<tr>
<td>28</td>
<td>Double bass</td>
<td>Chordophones</td>
<td>Bowed</td>
<td>156.30</td>
<td>158.67</td>
<td>150.67</td>
<td>153.00</td>
<td>139.00</td>
</tr>
<tr>
<td>29</td>
<td>Harp</td>
<td>Chordophones</td>
<td>Plucked</td>
<td>276.67</td>
<td>214.50</td>
<td>149.33</td>
<td>138.67</td>
<td>120.33</td>
</tr>
<tr>
<td>30</td>
<td>Bassoon</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
<td>260.50</td>
<td>181.33</td>
<td>207.67</td>
<td>83.67</td>
<td>35.00</td>
</tr>
<tr>
<td>31</td>
<td>Tenor saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
<td>114.00</td>
<td>59.33</td>
<td>57.67</td>
<td>50.50</td>
<td>57.67</td>
</tr>
</tbody>
</table>

Table 4.7: Summary of the latency value results ($L_d$)
4.4 Summary

The experimental procedure, the set-up and the salient characteristics of the listening test are presented. The first results and trends from the pilot test are analysed and discussed. The chapter concludes with a summary of the data gathered from the listening test. This information forms the basis for the statistical analysis and exploration in the next chapter.
Chapter 5

Findings and Analysis

In this chapter, the data analysis and results are presented. The results were obtained from the data gathered in the listening test. Table 5.1 shows the descriptive statistical results for all test subjects according to tempo in BPM. The values obtained for this table included all three different metronomes (aural, visual and aural-visual). Latency values for the data gathered that is used in all plots, tables and numerical operations are referred as latency ($L_d$) and are measured in milliseconds as described in Equation 3.1, unless otherwise stated.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>86.00</td>
<td>88.00</td>
<td>87.00</td>
<td>86.00</td>
<td>84.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>7.00</td>
<td>5.00</td>
<td>6.00</td>
<td>7.00</td>
<td>9.00</td>
</tr>
<tr>
<td>min</td>
<td>49.00</td>
<td>29.00</td>
<td>43.00</td>
<td>27.00</td>
<td>5.00</td>
</tr>
<tr>
<td>max</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>range</td>
<td>251.00</td>
<td>271.00</td>
<td>257.00</td>
<td>273.00</td>
<td>295.00</td>
</tr>
<tr>
<td>median</td>
<td>211.00</td>
<td>173.00</td>
<td>145.00</td>
<td>127.50</td>
<td>108.00</td>
</tr>
<tr>
<td>mean</td>
<td>205.59</td>
<td>180.06</td>
<td>156.52</td>
<td>140.56</td>
<td>122.83</td>
</tr>
<tr>
<td>SE.mean</td>
<td>8.53</td>
<td>8.43</td>
<td>8.30</td>
<td>8.02</td>
<td>8.18</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>16.95</td>
<td>16.75</td>
<td>16.51</td>
<td>15.94</td>
<td>16.27</td>
</tr>
<tr>
<td>var</td>
<td>6251.54</td>
<td>6251.11</td>
<td>5998.81</td>
<td>5527.00</td>
<td>5621.30</td>
</tr>
<tr>
<td>std.dev</td>
<td>79.07</td>
<td>79.06</td>
<td>77.45</td>
<td>74.34</td>
<td>74.98</td>
</tr>
</tbody>
</table>

Table 5.1: Descriptive statistics for the tempo in BPM

The purpose of Table 5.1 is to present an overview and a statistical description for all data gathered (see Appendix K.1). The columns of Table 5.1 are the five different tempi in BPM. The rows list the number of values (nbr.val.); the number of null values (nbr.null); the number of non available or missing values (nbr.na.); the minimum and the maximum
values; the range; the median; the mean; the standard error of the mean (SE mean) which is used to estimate the standard deviation of the sampling distribution of the mean, i.e. a measure of the variability between the means of the different samples [93]; the confidence interval of the mean at a p-level of 95% (CI.mean 0.95); the variance and the standard deviation.

Results in Table 5.1 consistently show the linear inverse relationship between latency and tempo. This tendency is clear when observing the median and mean. In addition, the variance also diminishes towards faster tempi, indicating a tendency to homogeneity of latency values \( L_d \). The slower the tempo the higher the median value. The numerical mean value for the results presented in Table 5.1 is comparable to the median. The standard error of the mean (SE.mean) in Table 5.1 has similar values for each tempo, i.e. around 8.5. The confidence interval at 95% for the mean is around 17. For the lower band of the confidence interval, a value of around 17 (different for each tempo) should be subtracted from the mean, while for the upper band a value of around 17 (different for each tempo) should be added to the mean. The real value of the mean with a confidence interval of 95% is likely to be found in the range between these lower and upper bands. The standard deviation, which is the square root of the variance, is around 74 and 79 depending on the tempo. Variance and standard deviation define how well the mean represents the data [92]. The accuracy of the mean may increase with the number of observations. The different numerical values of the descriptive statistics represent time ( \( L_d \), latency in milliseconds), with the exception of the first three rows.

The following bar plots and histograms, Figures 5.1, 5.2 and 5.3, summarize the information relating to the number of test subjects according to age, gender, expertise, instrument groups, sound generation methods and additional instruments performed.
Figure 5.1: Age and gender distribution

From Figure 5.1, one can easily see that the majority of the test subjects were younger, between 18 and 35 years old, and the mean and median were 25.2 and 22 years old respectively. Music students made up around four-fifths of the total number of participants. It is not surprising that availability and openness to new research issues plays a role for this group of subjects. With respect to gender, there are considerably more male test subjects (20) compared with female test subjects (11). Availability was the only reason for this result. The choice of test subjects was purposive but not convenient.

A total of 40 test subjects took part in the listening test. The data of 31 subjects representing 17 different musical instruments was analysed and plotted. Some listening tests did not fulfil the requirements for inclusion in the analysis. The main reasons for withdrawing the data of a specific listening test were:

- Memorizing the score (playing by heart or not looking at the score).
- Inability to play while listening to a metronome.
- Huge variations in tempo (more than 30 BPM) between the metronome and the performed score.
- Long rests (more than eight bars).
Including the data of the test subjects that did not fulfil the requirements may translate into a bias in the results and data gathered in the listening test.

The instrument groups in Figure 5.2 are not equally distributed, i.e. there is an unbalanced design. Chordophones and aerophones outnumber the membranophones and idiophones. Again, availability of the test subjects played the main role in this. However, the distribution reflects the distribution of musicians playing in an orchestra relative to the different instrument groups. Chordophones and aerophones, which are the strings, woodwinds and brass instruments, are the largest groups in the modern orchestra with regard to the number of musicians [3]. On the other hand, membranophones and idiophones belong mainly to the percussion instruments. This group has a relatively small number of musicians in comparison with the groups of chordophones and aerophones.

Figure 5.3 shows which other additional instruments the test subject performs in addition to their main musical instrument. The legend on the right of the bar plot lists the main instrument. In addition to the main musical instrument performed, almost half the test subjects either play the piano or no instrument at all. Professional musicians primarily perform only their main instrument.
Summary

- 31 subjects were tested. The majority were young, male European musicians.

- Latency values $L_d$ in milliseconds decreases as tempo increases.

- Aerophones and chordophones are the most represented musical instrument groups with respect to the number of instruments.

- Musicians play either no second instrument or mainly the piano as a second instrument.

Discussion

For this research the number of test subjects is higher compared with former investigations regarding the latency issue. Boley and Lester [152] tested 19 musicians playing 6 different musical instruments. In research done by Bartlette et al. [21] the number of participants is even reduced. Similar to former research, the amount of aerophones and chordophones was higher in number in relation to the number of participants.

The inverse relationship latency vs. tempo is confirmed as presented by Barbosa [18]. However, the numerical data and comparisons presented in this research enable a more differentiated picture of this relationship. As described in Table 5.1, the similar results
of the variance of the lower tempi 90 BPM and 120 BPM compared with the other three tempi indicates a stronger influence of the instrument in the lower tempi.

The inclusion of the question regarding the performance of an additional instrument has shown that musicians either play only the main instrument or the piano and indicates only the preferences of music students and musicians. This result is conforms with normal practice. Proficient musicians dedicate time exclusively to the performed instrument. On the other hand, the piano is considered nowadays a versatile and may be the best known instrument [3].

5.1 Exploratory Data Analysis (EDA)

The first step in obtaining information from the data is to establish possible relationships between the different numerical variables. Using Cartesian coordinates, it is possible to create a scatter plot matrix displaying the values of pairs of variables. Such a plot visually displays several correlations between multiple variables and trends are easy to identify. A scatter plot is mainly used to determine if there is any correlation between variables [59].
Figure 5.4: Scatter plot matrix for all numerical variables as indexed in Appendix K.1
In the scatter plot matrix (rotated through 90 degrees) of Figure 5.4, all numerical variables are plotted against each other. The different variables are shown in the diagonal line from the top left to bottom right of the plot. The relationships can be appreciated in pairs, e.g. age and metronome hours or as column 2 and row 1. All relationships in the upper right-hand area are mirrored in the lower left-hand, e.g. metronome hours and age or column 1 and row 2. The plot shows different numerical values at the bottom right-hand side.

The numerical values of Figure 5.4 are: $L_d$ from 0 to 300 in milliseconds (column 5 to 9), age values in years (column 1), years of experience (column 3), metronome hours in hours (column 2) and practice hours in hours (column 4).

The linear trends in Figure 5.4 are easy to identify for the metronome data (90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM). The data for 90 BPM vs. 120 BPM and 120 BPM vs. 90 BPM show a linear relationship with a relatively high variance. On the other hand, for the relationships between 150 BPM, 180 BPM and 210 BPM, the variance is only high with increasing index numbers.

Averaging the data of all metronomes (aural, visual and both) has no influence on the linear trends. Figure 5.5 presents basically the same information as Figure 5.4, the only difference being a reduction in the amount of observations to one-third due to the average. However, all tendencies are still observable. Plots under the diagonal are mirrored with respect to the plots above the diagonal. The information content remains the same.

Different relationships can be visually established between the numerical variables (column 1 to 4) and the numerical variables (column 5 to 9). Between the variables of column 1 to 4, the only relationship is between column 3, years of experience, and row 1, age. An older person is normally expected to have had more years of experience playing the instrument. This correlation is visually displayed. Further relationships are shown for the numerical values, for example column 6 and row 5, where the linear relationship is clear.

With respect to Figure 5.5, it is important to state at this point that for further results presented in this chapter, tables and figures including information on tempi (90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM) and instruments are the average of the measures of the three different metronomes (aural, visual and both), except where otherwise specified.
Figure 5.5: Scatter plot matrix of all numerical variables averaging the metronome data indexed in Table 4.7
Summary

- Data visualization is not affected by averaging the latency values $L_d$ of the three different metronomes.
- Age and years of experience are correlated.
- All numerical variables related to tempo in BPM correlate with each other.

Discussion

Observing the Figures 5.4 and 5.5, in particular the scatter plots for the relationships between the different tempi (columns 5 to 9), it is clear that the shape remains the same even when averaging the data. The data can be averaged without changing the outcome visually. The five different tempi measurements correlate with each other, this correlation is also the product of measuring the same variable ($L_d$). As explained before, there is a trivial relationship between age and years of experience which is easy to visualize (column 3, row 1).

5.1.1 Testing for Normality of the Data Distribution

In order to evaluate and further analyse the data, it is necessary to estimate if the distribution of the data is normal or not. An initial approach is to plot the data. Graphical methods for testing a theoretical distribution include the histogram, the box plot and the Q-Q plot. Plotting is an easy way to compare the data with respect to normality [59]. The first step is to compare a normal random sample with the same number of elements with the distribution of the data.

From the random distribution in Figure 5.6, the shape of all four diagrams is clear: the raw data is randomly distributed across the whole plot. The histogram shows that all possible values are present and the frequency of the middle values (100 to 200) confirm a normal distribution.
As expected, the box plot in Figure 5.6 indicates the common shape for a random distribution: the median is the middle value and the second and third quartiles, the 25% and 75% values respectively, match these values. The Q-Q plot also indicates also unequivocally that the values are normally distributed and within the confidence interval denoted with the dashed line.

Figure 5.6: EDA for a random distribution

Figure 5.7: EDA for the data distribution of the research
Findings and Analysis

The data gathered and presented in Figure 5.7 shows a very different outcome. The raw data in the upper left-hand plot shows an inhomogeneous distribution which is far from random. It is possible to identify a decaying behaviour of the data. This is explained through the inverse relationship between latency and tempo. The horizontal axis of this plot is the number of observations, while the vertical axis is the latency value. The outliers are the data on the ceiling which is the 300ms value or the latency value limit of the experiment.

The histogram displays the higher frequency of the 300ms value. This value is the product of outliers in the data. In other words, the latency values of the musicians who could play all the way up to 300ms regardless of the tempo in BPM. These are clearly seen in both plots.

The third plot, the lower left-hand box plot, represent the complete data and indicates the border values of latency between 0 and 300ms. The Q-Q plot indicates without any doubt that the distribution is not normal.

The implication of not having a normal distribution for the data gathered is important and forces the consideration of two possible alternatives in order to make inferences from the data. The first possibility is to transform the data to obtain normality. By means of a transformation, parametric statistical tests can be applied. The second choice is to use non-parametric tests that work on the principle of ranking but also have less power than the parametric tests [93]. The use of non-parametric tests may imply the impossibility of detecting statistically significant effects [93].

A normal distribution is expected after transforming the original data which enables the use of parametric statistics for the data evaluation. Two of the most common transformations for positive data are the logarithm (log) and the square root transformations [93]. In Figure 5.8 and Figure 5.9 the exploratory data analysis of each transformation is presented. The effect of the transformation is easy to see in the raw data distribution plot (upper left for both plots).
Both transformations failed to achieve a normal distribution when compared to the random data in Figure 5.6. The logarithm transformation as shown in Figure 5.9 is a good fit, however, normality is not achieved. The confidence interval is violated. Furthermore, when using transformations, it is important to be aware that the transformation of the data might also change the hypothesis being tested and the risk of choosing a wrong transformation may be high [93]. For further inferential data analysis, the use of non-
parametric tests in this study is unavoidable.

To confirm whether the data deviates from a normal distribution, the Shapiro-Wilk test enables a comparison between the data and a normally distributed set of scores having the same mean and standard deviation [93]. Table 5.2 shows the results of the Shapiro-Wilk normality test. For significant p-values (p < 0.05), it can be concluded that the distribution is not normal. For a normal distribution the W-value is very near to 1.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>W value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.99673</td>
<td>0.4726</td>
</tr>
<tr>
<td>Experiment</td>
<td>0.92221</td>
<td>3.834e-14</td>
</tr>
<tr>
<td>Sqrt (Exp.)</td>
<td>0.95812</td>
<td>1.003e-09</td>
</tr>
<tr>
<td>Log (Exp.)</td>
<td>0.93301</td>
<td>5.444e-13</td>
</tr>
</tbody>
</table>

Table 5.2: Shapiro test results for normality

The results presented in Table 5.2 indicate that the data of the experiment and both the square root and logarithm transformations are not normally distributed. Therefore, the next step is to establish further relationships between the variables of the experiment based on a visual analysis.

Further transformations in order to enable the comparability of data and establish relationships between variables are not necessary. The systematic research and experiments presented in Chapter 3 and Chapter 4 allow for the comparison of results between all musical instruments. The purpose of the previous transformations and their results presented in Figures 5.8 and 5.9 was only directly concerned with the non-normal distribution of the totality of the gathered data.

Knowing that the data distribution is non-normal, the next step is to define relationships between the different variables of the listening test. For the calculation of correlation coefficients, two non-parametrical methods are available: Spearman’s method ($r_s$) and Kendall’s correlation coefficient ($\tau$). Because of the relatively small data set, Kendall’s ($\tau$) is considered the best choice for calculating the correlation coefficients for the different relationships of the data [93, 117].
Findings and Analysis

For the different metronome results, the influence of gender can be explored in Figure 5.10. For an easy visual analysis, differences were indicated using colours\(^1\).

In Figure 5.10, the linear relationship between 90 BPM and 120 BPM is the same as that in the first scatter plot matrix of Figure 5.4. The tempi 150 BPM, 180 BPM and 210 BPM also display a linear relationship. On the other hand, it is clear from the graphical information and colour that gender plays no role with respect to the ability to cope with latency. Both genders are equally distributed across the different plots in the matrix.

Figure 5.10 shows multivariate numeric data. The lower diagonal is a matrix with scatter plots. The diagonal shows the densities and confirms the assumption that gender is not a relevant issue with regard to the ability to perform a musical instrument while listening to oneself with added latency. The shape of the functions on the diagonal are similar for both gender at the different tempi. The differences in peak values might be a result of the number of observations for male and female subjects. In the upper diagonal, correlation

\[^1\text{Horizontal and vertical axis for Figures 5.10 and 5.11 are latency } (L_d) \text{ in milliseconds. Data in the diagonal shows only the shape of the data distribution.}\]
values for male and female are shown.

Figure 5.11 shows the latency data \((L_d)\) for each metronome. The multivariate numerical data was split according to the three different metronomes. The values of the aural metronome are the red dots, for the visual metronome blue dots and for the results of the aural and visual metronome the dots are green.

According to the scatter plots and density plots, the metronome in its three different variants has no influence on the data distribution. In other words, there is no measurable effect of the type of metronome used with respect to latency.

The correlation results supplied by the upper diagonal numerically confirm the assumption of the absence of influence of the type of metronome used. In other words, the effect on the results when using one type of metronome rather than another is irrelevant. The correlation coefficients are, in most cases, numerically similar.
Findings and Analysis

Summary

- The distribution of the data gathered is non-normal, even after applying square root and logarithmic transformations.
- Gender plays no role in the ability to cope with latency.
- The influence of the different types of metronomes (aural, visual and both) is minimal or not existent with respect to the latency ($L_d$) results.

Discussion

In order to proceed with an inferential statistical analysis, it is mandatory to analyse and define the distribution of the data gathered. Previous research done by Boley and Lester [152], Carot [5] and Pfordresher [181] presented descriptive results and assumed normal distributions. The data gathered for the present research is non-normal distributed, it implies the use of non-parametric statistical methods. A possible reason for this distribution may be the reduced number of test subjects.

Gender is not relevant regarding the issue of latency and musical instruments as seen in Figure 5.10. This is a confirmatory result also found in DAF research done by Gates [106].

Bartlette et al. [21] recommended to incorporate visual cues into the test environment and it is well known that visual information interacts together with aural information as Smith presented in his dissertation [209]. The use of three different kinds of metronomes (aural, visual and aural-visual) was planned and developed based on the knowledge gathered from researches such as Posner [191] and the need for a holistic view proposed by Lanier [148]. The gathered results indicate visually that using one or another metronome produces no significant effect regarding the latency ($L_d$) results, which is a novel outcome. It could be explained through the high adaptability of humans as proposed by Altisnon [143].

5.1.2 Correlation Analysis

At the beginning of Chapter 5.1, Exploratory Data Analysis (EDA) and the scatter plot matrix for all numerical values was presented in Figure 5.4. It was possible to examine the existence of any correlations within the data. A quantitative confirmation is now presented in Figure 5.12 where the correlation values for all numerical variables are displayed
Findings and Analysis

in a matrix. The correlation values can be either positive or negative. The range of the colour scale, which supports the numerical values goes from -1 (blue) through 0 (white) up to 1 (red). According to the numerical values, there is no meaningful correlation between latency and age. Moreover, latency does not correlate with hours of practice of the instrument (practice_hours) and hours of practice with a metronome (metronome_hours) at the different tempi of the experiment. In addition, the number of years of experience is also not correlated with any of the breakdown latency values obtained at each tempo in BPM.

Figure 5.12: Correlation matrix summary for all variables

Figure 5.12 indicates correlations only between age and years of experience, and such a relationship can be considered logical and trivial to the experiment. The older the subject, the more years he or she has played the instrument. The metronome hours of practice and the hours of total practice are also moderately correlated. This is also an expected logical result.
Findings and Analysis

Figure 5.13: Correlation matrix plot for latency values

Correlations exist between the different latency values as shown in Figure 5.13. This tendency was observed in the scatter plots matrices of Figure 5.4 and Figure 5.5. As previously stated, there are two sections which are highly correlated. The first section corresponds to the 90 BPM and the 120 BPM tempi and the second section to the 150 BPM, 180 BPM and 210 BPM tempi.

A more exhaustive analysis\(^2\) of the correlations is shown in Figure 5.14. In addition to being a standardized measure, the correlation coefficient is considered a common method to describe observed effects. For the range between -1 and 1, the size of the effect can be defined as small (+/- 0.1), medium (+/- 0.3) and large (+/- 0.5) [93].

In Figure 5.14, correlation values are not rounded and it is also possible to visualize the correlation with the confidence interval and the variation of data. The red line is the fitted linear regression and the grey area around it shows the 95% confidence level interval zone. The kernel density estimation (KDE) for the latency \((L_d)\) results of each tempo is displayed by the density plots in the diagonal. The KDE is a smoothed version of the histogram [92]. For 90 BPM, latency \((L_d)\) is proportionally distributed across the latency x-axis with a peak near 300ms. For the tempi 120 BPM, 150 BPM, 180 BPM and 210 BPM the peaks moves in the direction of lower latency \((L_d)\) values. The faster the tempo, the lower the latency \((L_d)\) musicians are able to cope with without breaking down the performance.

\(^2\) Horizontal and vertical axis for Figure 5.14 are latency \((L_d)\) in milliseconds. In the upper left of the vertical axis is density.
The y-axis in Figure 5.14 is standardized from 0 to 300ms, i.e. the latency $L_d$ values. On the contrary, the y-axis for the density plots (KDE) on the diagonal does not have the same scale for every kernel density estimation plot. Figure 5.15 shows the real standardized density value scaling. For this summary plot, only the shapes of the density plot can be compared.

The size of the effect between adjacent tempi are large (+/- 0.5), especially for 90 BPM and 120 BPM and for 150 BPM, 180 BPM and 210 BPM, even when using the Kendall’s correlation coefficient ($\tau$). This coefficient has a more mathematically conservative approach and is 66-75% smaller than other correlation coefficients, e.g. Spearman’s ($r_s$) and Pearson’s ($r$) [93].

The linear inverse relationship between latency and tempo is known from Barbosa [18], Farner [90] and Schuett [205], to quote some authors. Results presented in this study confirm this relationship. Additionally, it is shown that the relationship between latency and tempo is variable, with latency values ($L_d$) varying to a greater degree at slower (90 BPM) tempi. Tempo is not the only cause of latency. There may be several measured
and unmeasured variables affecting results, such as note resolution or performed pattern as reported by Carót [42], Lester [152] and Gurevich [113]. In the present study, however, the correlation shows that tempo is indeed a very important variable related to latency. At the same time the direction of causality for the listening test is clear. In other words, different tempi induced different latency values. It is also well known from research done by Chafe [54] and Farner [90] that latency values may also slow or accelerate a performance.

Summary

After analysing the information provided by the correlation matrix and the plot, it is possible to state that:

- There is no correlation between the hours of practice of the musical instrument and latency values $L_d$. The use of a metronome is also irrelevant.
- Years of experience of playing the instrument do also not correlate with the latency values $L_d$.
- The inverse relationship between latency ($L_d$) and tempo is observed and confirmed.
- In addition to this and based on the correlation coefficients, large effects are observed for adjacent tempi (e.g. between 90 BPM and 120 BPM or 180 BPM and 210 BPM).
- The effects observed are stronger for the tempo pairs 90 BPM and 120 BPM, 150 BPM and 180 BPM and 180 BPM and 210 BPM. Less stronger effects are present between the tempo pair 120 BPM and 150 BPM.

Discussion

There is small or no correlation between the latency values for the different tempi and the years of experience, the hours of practice or even the hours of practice using a metronome to play the musical instrument. It indicates that those variables are not related to the latency values $L_d$. On the other hand, the different tempi 90 BPM to 210 BPM and their respective latency values $L_d$ are highly related with variations between the different tempo pairs. The existence of such a relationship is known from Barbosa [18], the internal relationship between the different adjacent tempi and the visual and numerical information are new outcomes. The result is interesting and challenges the assumption of Rottondi et
al. [198]. She stated that the ability for delay compensation may be directly related to performer dexterity.

5.1.3 Density Plots for Latency Values

Figure 5.15 shows the kernel density estimation (KDE) for the data over the continuous range of latency ($L_d$). The diagonal in Figures 5.10 and 5.11 is actually a kernel density estimation.

In Figure 5.15, peaks are displayed where the concentration of latency ($L_d$) values is high. Different tempi in BPM are represented using different colours. The use of a kernel density estimation plot enables the determination of the shape of the distribution, even when the data is minimal. At this point it is important to mention that the KDE plots, as in Figure 5.15, as well as the basic line plots in Figure 5.16, imply continuous values of latency. However, the latency values were only gathered at discrete points, i.e. the five different tempi of 90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM.

Figure 5.15: Density plot for all instruments according to the different tempi (average of all 3 metronomes)

In Figure 5.15, it is clear that the assumptions made with regard to the visual results of the correlation scatter plots for tempi 90 BPM and 120 BPM, and for those from 150 BPM to 210 BPM are also shown here. For slower tempi such as 90 BPM or 120 BPM,
the distribution across the latency \((L_d)\) shows two peaks, with the most salient around 275ms for 90 BPM and around 175ms for 120 BPM. On the other hand, for faster tempi beginning at 150 BPM the magnitude of the peak increases when the latency values decrease. For 150 BPM this is around 160ms and for 180 BPM and 210 BPM this is around 90 to 100ms latency \((L_d)\).

Summary

- Peaks in the KDE plot are clear for the tempi 150 BPM, 180 BPM and 210 BPM.
- Latency \(L_d\) distribution across the range 0 to 300ms depends strongly on tempo.

Discussion

The higher the tempo, the more musicians breakdown the performance at similar latency values \(L_d\), that is the explanation for the narrow and higher peaks. Narrow peaks mean that substantially more musicians disrupted the performance at an specific \(L_d\), which translates in a higher peak or a major density as seen in Figure 5.15. The effect of the tempo is crucial and the effect was already known from research done by Chew et al. [63] and by Carôt [5]. The new information presented here is the distribution of lower tempi (90 BPM and 120 BPM). The influence of the instrument and the wide range of latency values \(L_d\) are visually different when compared with the narrow distribution of higher tempi (150 BPM, 180 BPM and 210 BPM).

5.1.4 Latency Measurements on Musical Instruments

The next plots show a graphical summary of the information of latency, tempo and musical instruments for all the data gathered. The box plots in Figure 5.16 show the distribution of latency versus tempo for all instruments, the three different metronomes of the experiment being averaged. The value of the median decreases as the tempo increases. There is a clear linear relationship as is also shown in previous research [42, 113, 152]. The comparison between the different tempi is relevant. For 90 BPM and 120 BPM the interquartile range between 25% and 75% is large. On the contrary, for 150 BPM, 180 BPM and 210 BPM the interquartile range is not so wide. This information confirms the visualization of the density plot in Figure 5.15.
In summary, at slower tempi the ability of musicians to cope with increasing latency may have a high variance and may be also more dependent on the instrument performed. The influence of the musician is obvious and very different latency ($L_d$) values are possible. For tempi around 150 BPM and faster, the interquartile range (IQR) decreases significantly in comparison to the tempi 90 BPM and 120 BPM. The musical instrument may still play a role but is not as significant as it is at slower tempi.

Table 5.3 shows the descriptive statistic results for all test subjects according to tempo in BPM averaged for the three metronomes.
Findings and Analysis

### Table 5.3: Descriptive statistics for tempo in BPM (average of all 3 metronomes)

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>31.00</td>
<td>31.00</td>
<td>31.00</td>
<td>31.00</td>
<td>30.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>min</td>
<td>71.00</td>
<td>59.33</td>
<td>57.67</td>
<td>48.00</td>
<td>19.00</td>
</tr>
<tr>
<td>max</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>range</td>
<td>229.00</td>
<td>240.67</td>
<td>242.33</td>
<td>252.00</td>
<td>281.00</td>
</tr>
<tr>
<td>median</td>
<td>216.00</td>
<td>178.00</td>
<td>149.33</td>
<td>128.33</td>
<td>104.67</td>
</tr>
<tr>
<td>mean</td>
<td>200.81</td>
<td>180.78</td>
<td>155.04</td>
<td>135.39</td>
<td>117.78</td>
</tr>
<tr>
<td>SE.mean</td>
<td>13.50</td>
<td>11.88</td>
<td>12.43</td>
<td>12.35</td>
<td>12.89</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>27.58</td>
<td>24.26</td>
<td>25.39</td>
<td>25.23</td>
<td>26.35</td>
</tr>
<tr>
<td>var</td>
<td>5652.24</td>
<td>4374.16</td>
<td>4789.50</td>
<td>4730.38</td>
<td>4980.85</td>
</tr>
<tr>
<td>std.dev</td>
<td>75.18</td>
<td>66.14</td>
<td>69.21</td>
<td>68.78</td>
<td>70.58</td>
</tr>
<tr>
<td>IQR</td>
<td>138.83</td>
<td>109.50</td>
<td>88.33</td>
<td>82.50</td>
<td>68.41</td>
</tr>
</tbody>
</table>

The results in Table 5.3 are very similar to those in Table 5.1 with respect to median, mean and standard deviation. On the other hand, the standard error mean is almost double the value of that in Table 5.1 and the confidence interval (95%) is more than 10 points larger at each tempo. A possible reason for this is the effect of averaging the values of the three different metronomes.

The IQR is the interquartile range, a measurement of statistical dispersion which is the difference between the upper and lower quartiles (75th and 25th percentiles). This measurement indicates how variable the results are, and this can be seen in Figure 5.16. The first two tempi, 90 BPM and 120 BPM, vary more in comparison to the other tempi. For musical instrument groups and sound generation methods, the IQR is defined as the latency tolerance range (LTR).

For every plot in the following sections, latency is assumed as \((L_d)\), the simulated latency value in the listening test. The latency values \((L_a)\) and \((L_c)\), which are the latency values of the distance through air and through analogue to digital to analogue conversion, respectively, are considered constants and are not included in any of the plots.
Figure 5.17 shows the latency ($L_d$) vs. tempo for every musical instrument analysed in this research. The plot summarizes the latency adaptive tempo (LAT) for the different instruments as proposed by Barbosa [18] and presented in Chapter 2.4. However, the plots in Figure 5.17 are the product of a listening test with a systematic methodology which is replicable and was performed under similar conditions for every test subject.

Latency value differences between musical instruments are easy to recognize. Furthermore, some instruments of the same type have similar latency ($L_d$) values, with exceptions being the alto saxophone, classical guitar and violin. With these instruments, two musicians were able to perform even when the delay increased to 300ms during the experiment. Some of these higher values may be considered outliers.

This assumption of similarity in values ($L_d$) can be graphically confirmed for piano, triangle and snare drum as shown in Figure 5.18. Some instruments have a larger interquartile range, especially the chordophones and aerophones.
Findings and Analysis

Figure 5.18: Box plots of the latency values ($L_d$) for each musical instrument

The plots of Figure 5.17 and Figure 5.18 are based on latency value ($L_d$) and suggest continuity across the latency range 0 to 300ms. However, the latency values ($L_d$) were only gathered at discrete points at the five different tempi of 90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM.

Summary

- The interquartile range (IQR) for latency $L_d$ is wider for slower tempi such as 90 BPM and 120 BPM and narrower for the tempi 150 BPM to 210 BPM.

- The mean and median of the latency values $L_d$ are higher for slower tempi and lower for faster tempi.

- Different musical instruments have different latency ($L_d$) values.

- The same musical instruments (e.g. piano compared to piano) often have similar latency ($L_d$) values.
Discussion

The inverse relationship latency vs. tempo already known from Barbosa et al. [18, 19] is presented with far more detail indicating the interquartile range for different tempi. This information is an easier way to visualize the results of the correlation matrix of Figure 5.14. Furthermore, Figure 5.17 is a visual evidence, indicating that different musical instruments enable different latency values ($L_d$). On the other hand, similar instruments often produce similar latency values ($L_d$).

Latency measurements for the instrument groups

In the following section the information is analysed based on the four musical instrument groups: the aerophones, chordophones, membranophones and idiophones. A graphical analysis of the latency tolerance range for the different groups is also presented.

Figure 5.19 plots latency ($L_d$) vs. tempo for the four different instrument groups. Membranophones and idiophones are in the lower latency ranges, and it appears that musicians performing membranophones are not able to cope with high latency values. The data for the aerophones and chordophones is not so clear. Within the aerophones, there is a distinction between lower latency range instruments and two other instruments, the
alto saxophone and trumpet in B, which are able to perform even when latency values are high (up to 300ms). The effect of the trumpet in B can easily be identified as an outlier with a value of 300ms. Chordophone values are widespread throughout the whole range of latency and are mainly above 100ms.

Figure 5.20 reinforces the premise observed in Figure 5.19. The range is more easily identified for the idiophones and membranophones while it is not so clear for the chordophones and aerophones. The latency tolerance range (LTR) can be easily visualized for each musical instrument group based on Figure 5.20. However, outliers may increase the range of the measurement making differences between groups unobservable.

![Figure 5.20: Latency ($L_d$) vs. instrument group (average of all 3 metronomes)](image)

A better approximation would be to base the latency tolerance range (LTR) on the interquartile range (IQR) of the different musical instrument groups as described in Equation 3.2.
Figure 5.21: LTR according to musical instrument group

Figure 5.21 shows a clearer picture of the results of the aerophones where the latency tolerance range is very wide. This is in contrast to the other three groups which have smaller latency tolerance ranges. This is confirmed by Figure 5.22.

In addition, Figure 5.21 shows the bigger picture of the findings with respect to the four musical instrument groups. Here, the differences between the groups with respect to ability to cope with latency can be seen. These differences are calculated using the latency tolerance range in Equation 3.2 (see Chapter 4). In other words, the LTR varies for the different musical instrument groups.

The differences in the interquartile ranges are also directly related to the number of instruments per group that took part in the listening test. Figure 5.22 shows a further visual analysis. The LTR for musical instruments and different tempi demonstrates two tendencies. Firstly, with exception of the membranophones, the faster the tempo the lower the latency values. Secondly, the latency tolerance range for different tempi varies more for some groups than for others. For the aerophones as well as the chordophones, the LTR tends to diminish towards faster tempi. For the idiophones and membranophones, the LTR increases for both the faster and slower tempi.
Figure 5.22: LTR with respect to musical instruments group and tempi

In Figure 5.22 the differences between the four groups are very clear. Variations in the latency tolerance range (LTR) for the different tempi can be visually observed. In the next section, the LTR values and additional statistical information are presented in table form.

**Density plots for latency values and instrument groups**

This section presents the descriptive statistics for the four different musical instrument groups: aerophones, chordophones, membranophones and idiophones. Tables 5.4, 5.5, 5.6 and 5.7 summarize the numerical information used for the LTR in Figure 5.22. In addition to the tables, the density plot (KDE) for each instrument group is presented in Figures 5.23, 5.24, 5.25 and 5.26.
Table 5.4 summarizes the visual information for the aerophones’ LTR plot in figure 5.22. The latency tolerance range is very large which means that the differences between the instruments of the aerophones group are very large, especially at the tempi of 90 BPM and 150 BPM. This is not due to the presence of outliers but rather the structure of the instruments themselves as well as the performance of the musicians. Using breathing as well as the embochure to generate the sound may play a significant role in this result.

Mean and median values are very dissimilar for higher tempi, even with differences up to 50 points. The tendency for aerophones to show lower latency values at faster tempi is an expected result. However, these variations are not so evident. The median and mean for 90 BPM and 120 BPM differ by less than 12 points. This variation increases as tempo increases, explaining the increase in the confidence interval. In summary, the instruments of the aerophone group are very different with regard to latency.
Findings and Analysis

Figure 5.23: Density plot for the aerophones at different tempi (average of all 3 metronomes)

Figure 5.23 confirms the results of Table 5.4. The peak value for the tempi 150 BPM to 210 BPM where performance breaks down is around 100ms. For the values corresponding to 90 BPM, the distribution presents two peaks at around 110ms and 300ms with similar densities. Table 5.24 is a summary of the descriptive statistics for the group of the chordophones.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
<td>13.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>min</td>
<td>131.00</td>
<td>100.33</td>
<td>94.00</td>
<td>96.33</td>
<td>61.67</td>
</tr>
<tr>
<td>max</td>
<td>288.67</td>
<td>266.33</td>
<td>300.00</td>
<td>288.67</td>
<td>234.33</td>
</tr>
<tr>
<td>range</td>
<td>157.67</td>
<td>166.00</td>
<td>206.00</td>
<td>192.33</td>
<td>172.67</td>
</tr>
<tr>
<td>median</td>
<td>234.00</td>
<td>214.50</td>
<td>176.00</td>
<td>153.00</td>
<td>129.67</td>
</tr>
<tr>
<td>mean</td>
<td>229.68</td>
<td>206.17</td>
<td>178.33</td>
<td>155.44</td>
<td>132.67</td>
</tr>
<tr>
<td>SE.mean</td>
<td>14.91</td>
<td>13.71</td>
<td>14.69</td>
<td>13.73</td>
<td>13.15</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>32.49</td>
<td>29.88</td>
<td>32.02</td>
<td>29.92</td>
<td>28.66</td>
</tr>
<tr>
<td>var</td>
<td>2890.24</td>
<td>2444.54</td>
<td>2807.22</td>
<td>2451.34</td>
<td>2248.69</td>
</tr>
<tr>
<td>std.dev</td>
<td>53.76</td>
<td>49.44</td>
<td>52.98</td>
<td>49.51</td>
<td>47.42</td>
</tr>
<tr>
<td>LTR</td>
<td>93.67</td>
<td>66.33</td>
<td>37.00</td>
<td>34.07</td>
<td>36.33</td>
</tr>
</tbody>
</table>

Table 5.5: Descriptive statistics for the chordophones

For the chordophones, mean and median values at each tempo are similar, even though the number of observations is higher than for the aerophones. As observed in Figure 5.22, the latency tolerance range is narrower and more constant at tempi 150 BPM up to 210
BPM. The confidence interval is also half the value obtained for the aerophones. It is plausible to assume that the chordophones are a more homogeneous group compared to the aerophones with respect to latency. Differences between individual instruments are not so pronounced.

Figure 5.24: Density plot for the chordophones at different tempi (average of all 3 metronomes)

Figure 5.24 shows the differences between the different tempi for the chordophones. For tempi between 150 BPM, 180 BPM and 210 BPM the peaks of the latency breakdown values are identifiable and the curve is narrow. Latency values for 90 BPM and 120 BPM are wider and there are no pronounced peaks.
Findings and Analysis

Table 5.6: Descriptive statistics for the membranophones

<table>
<thead>
<tr>
<th></th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>min</td>
<td>71.00</td>
<td>74.67</td>
<td>67.33</td>
<td>48.00</td>
<td>22.00</td>
</tr>
<tr>
<td>max</td>
<td>145.00</td>
<td>137.33</td>
<td>96.00</td>
<td>104.67</td>
<td>94.00</td>
</tr>
<tr>
<td>range</td>
<td>74.00</td>
<td>62.67</td>
<td>28.67</td>
<td>56.67</td>
<td>72.00</td>
</tr>
<tr>
<td>median</td>
<td>82.17</td>
<td>124.17</td>
<td>89.00</td>
<td>71.67</td>
<td>74.33</td>
</tr>
<tr>
<td>mean</td>
<td>95.08</td>
<td>115.08</td>
<td>85.33</td>
<td>74.00</td>
<td>66.17</td>
</tr>
<tr>
<td>SE.mean</td>
<td>17.26</td>
<td>14.10</td>
<td>6.49</td>
<td>12.58</td>
<td>15.49</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>54.92</td>
<td>44.88</td>
<td>20.65</td>
<td>40.04</td>
<td>49.31</td>
</tr>
<tr>
<td>var</td>
<td>1191.14</td>
<td>795.66</td>
<td>168.39</td>
<td>633.19</td>
<td>960.33</td>
</tr>
<tr>
<td>std.dev</td>
<td>34.51</td>
<td>28.21</td>
<td>12.98</td>
<td>25.16</td>
<td>30.99</td>
</tr>
<tr>
<td>LTR</td>
<td>32.25</td>
<td>25.91</td>
<td>13.91</td>
<td>31.67</td>
<td>23.00</td>
</tr>
</tbody>
</table>

Table 5.6 summarizes the information for the membranophones. The assumptions are based on only four instruments. The values of the mean and median in Table 5.6 are similar for all tempi, with maximum differences of 13 points. On the other hand, the confidence interval is wider for the two extreme tempo values (90 BPM and 210 BPM) decreasing from these extremes to the lowest value at the tempo of 150 BPM. The latency tolerance range as observed in Figure 5.22 is narrow, tending to increase for extreme tempo values of 90 BPM and 210 BPM and decrease towards the middle tempo value of 150 BPM. An important value is the maximal latency value ($L_d$) of 145ms for 90 BPM. This value could imply that test subjects performing on membranophones are less able to perform the instrument while listening to their own latency when compared to the aerophones and chordophones.
Figure 5.25: Density plot for the membranophones at different tempi (average of all 3 metronomes)

Figure 5.25 is a visual representation of the information in Table 5.6. The distribution of the density plot shows more than one peak for every tempo. One reason is the small amount of data, with only four observations. On the other hand, it is clear that the ability to play membranophones without breaking down the performance is not as good as when compared with aerophones or chordophones.

Table 5.7 summarizes the information for the three observations from the idiophone group.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>min</td>
<td>178.33</td>
<td>169.33</td>
<td>131.50</td>
<td>69.50</td>
<td>45.00</td>
</tr>
<tr>
<td>max</td>
<td>244.67</td>
<td>228.67</td>
<td>175.00</td>
<td>166.00</td>
<td>179.67</td>
</tr>
<tr>
<td>range</td>
<td>66.33</td>
<td>59.33</td>
<td>43.50</td>
<td>105.50</td>
<td>134.67</td>
</tr>
<tr>
<td>median</td>
<td>221.78</td>
<td>208.50</td>
<td>149.72</td>
<td>126.83</td>
<td>117.06</td>
</tr>
<tr>
<td>mean</td>
<td>21.73</td>
<td>19.59</td>
<td>13.04</td>
<td>33.35</td>
<td>39.16</td>
</tr>
<tr>
<td>CI mean 0.95</td>
<td>93.51</td>
<td>84.27</td>
<td>56.12</td>
<td>143.48</td>
<td>168.49</td>
</tr>
<tr>
<td>var</td>
<td>1416.93</td>
<td>1150.86</td>
<td>510.40</td>
<td>3336.08</td>
<td>4600.68</td>
</tr>
<tr>
<td>std.dev</td>
<td>37.64</td>
<td>33.92</td>
<td>22.59</td>
<td>57.76</td>
<td>67.83</td>
</tr>
<tr>
<td>LTR</td>
<td>33.17</td>
<td>29.67</td>
<td>21.75</td>
<td>52.75</td>
<td>67.33</td>
</tr>
</tbody>
</table>

Table 5.7: Descriptive statistics for the idiophones
In Table 5.7, mean and median results vary by up to 30 points. Variation in the confidence interval is high and may be produced by the small number of observations. Contrary to the other three musical instrument groups, the latency tolerance range from Figure 5.22 increases significantly at faster tempi (180 BPM and 210 BPM). This may indicate strong differences between the musical instruments of the idiophone group.

Figure 5.26: Density plot for the idiophones at the different tempi (average of all 3 metronomes)

Figure 5.26 is a visual representation of the tendencies in Table 5.7. There is more than one peak for the latency value of performance breakdown. Moreover, for 210 BPM there is no peak for the distribution across the latency ($L_d$)-axis. This result is based on a small number of observations.

Before explaining the above results, it is necessary to clarify to what extent generalizations are allowed based on the gathered data. The kernel density estimation (KDE) plots for all musical instruments are presented in Figure 5.15, and these outcomes can be generalized and supported by the total data. On the other hand, when analysing the different musical instrument groups, the data for aerophones, chordophones, membranophones and idophones represents only a fraction of the total data. In other words, the data for the KDE plots for the musical instrument groups is reduced. This reduction is clear in Figure
Findings and Analysis

5.25 and in Figure 5.26. Tables 5.6 and 5.7 provide a numerical confirmation of this data reduction.

In Figure 5.23 latency values ($L_d$) for tempi 90 BPM and 120 BPM are concentrated around 110 to 120ms. The faster the tempi, the lower the latency value ($L_d$). For tempi 180 BPM and 210 BPM these values are concentrated around 80 to 95ms. It is possible to observe peaks at higher latency values, these are mainly due to musicians who were able to perform the musical instruments with latency values up to 300ms.

Figure 5.24 shows a slightly different result for the chordophones as compared to the aerophones. There is mainly one peak, with exception of the blue line (180 BPM) and the yellow line (120 BPM). The peaks show the latency value with the higher concentration of results. It is clear that the slower the tempi, the higher the latency value the musician is able to perform at without a breakdown. For 90 BPM this lies around 265ms, for 120 BPM the peak is around 230ms and 175ms, for 150 BPM the peak is at 175ms, for 180 BPM the peak is around 150ms and for 210 BPM it is 125ms. It is expected that those musicians, who were able to perform the score with latencies up to 300ms were responsible for the second peak for the 180 BPM curve.

Summary

- Musical instruments can be visually characterised with respect to the latency values ($L_d$) based on the four musical instrument groups (chordophones, aerophones, membranophones and idiophones).

- The LTR enables comparisons of the latency values ($L_d$) for musical instrument groups at different tempi and is dependent on the number of observations.

Discussion

Both approaches are new in research on latency and musical instruments. First of all, it has been shown visually in Figure 5.19 that it is possible to group instruments according to the main musical instrument groups. Furthermore, this grouping is congruent regarding latency values ($L_d$). Secondly, the proposed measure, the latency tolerance range (LTR) shows clearly in Figures 5.21 and 5.22 the differences for latency values ($L_d$). This information enables to summarise latency values, musical instrument groups and musical tempo at once.
Latency measurements for sound generation techniques

In this section, the same approach used to analyse the musical instrument groups is used for the eight different sound generation methods. It is important to clarify that the only change for the analysis is in relation to the aerophones with three different sound generation methods used (air reed, lip reed and mechanical reed) and the chordophones, also with three different sound generation techniques (bowing, plucking and striking). Both the idiophones and the membranophones in this research were always struck. Therefore, for the membranophones and the idiophones the previous description of the musical instrument group is the same for the sound generation techniques. The sound generation techniques were not mixed between instrument groups. In other words, a struck idiophone is not the same as a struck membranophone and is also different to a struck chordophone. The sound generation techniques approach is group related rather than instrument related.

Figure 5.27 shows the latency adaptive tempo (LAT) as defined by Barbosa [18]. The plot provides new information for the bowed chordophones. These are identifiable as a group. The different sound generation methods have different latency values and ranges are visually identifiable.

![Figure 5.27: Latency ($L_d$) vs. tempo for the different sound generation techniques](image-url)
Figure 5.28 attempts to show the latency tolerance range (LTR). Due to extreme values and outliers it is visually difficult to identify a specific range.

Figure 5.28: Latency ($L_d$) vs. sound generation technique (average of all 3 metronomes)

The latency tolerance range in Equation 3.2 is calculated mainly for the latency values $L_d$ relating to musical instrument groups. However, this calculation can be also be applied to the different sound generation techniques for different musical instruments as shown in Figure 5.29 and Figure 5.30. Figure 5.29 shows that, in addition to the struck membranophones and struck idiophones, the struck chordophones have a small interquartile range. Although the method of sound generation is the same (striking), the latency tolerance range (LTR) is different.
Findings and Analysis

Figure 5.29: LTR according to the sound generation technique

Figure 5.30 shows the LTR information according to the sound generation techniques and tempi. Values for membranophones and idiophones are the same as those in Figure 5.22. On the other hand, additional information for chordophones and aerophones is observed. The LTR for lip reed and mechanical reed in Figure 5.30 (the yellow and green bars) differs for the tempi 90 BPM and 120 BPM, while for 150 BPM mechanical reed and lip reed instruments have a similar LTR behaviour. The data obtained for the air reed (transverse flute) is insufficient to determine a range. Bowed, plucked and struck chordophones have a common range except at the tempo 90 BPM, where differences between bowed and plucked chordophones are obvious.
Figure 5.30: LTR according to the sound generation technique and tempi

Figure 5.30 summarizes the descriptive statistical information presented in Tables 5.8 to 5.13. The number of observations is greatly reduced. KDE plots for each sound generation technique are not generated. Table 5.8 presents the information of the transverse flute, which is the only air reed instrument of this study. The only information available is the median and the mean, which are the same. These values describe the normal behaviour with regard to the ability to cope with latency. The faster the tempo, the lower the latency value for breakdown in musical performance.

Table 5.8: Descriptive statistics for air reed sound generation

Table 5.9 is a summary of the descriptive statistics for musical instruments using lip reed for sound generation, such as French horn, trombone and trumpet in B. Mean and median show differences up to 43 points. The influence of the values obtained for the trumpet
in B is evident. The trumpet in B results can be considered outliers. Range variations increase parallel to the faster tempi. In other words, differences between latency values are more pronounced at faster tempi. The confidence interval mean is very wide, a result that reinforces the information about range and standard deviation. On the other hand, the LTR is almost the same for the tempi 150 BPM to 180 BPM.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>4.90</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>min</td>
<td>123.33</td>
<td>86.33</td>
<td>88.67</td>
<td>63.00</td>
<td>73.33</td>
</tr>
<tr>
<td>max</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>range</td>
<td>176.67</td>
<td>213.67</td>
<td>211.33</td>
<td>237.00</td>
<td>226.67</td>
</tr>
<tr>
<td>median</td>
<td>262.67</td>
<td>175.00</td>
<td>158.67</td>
<td>107.67</td>
<td>115.67</td>
</tr>
<tr>
<td>mean</td>
<td>232.73</td>
<td>189.93</td>
<td>174.47</td>
<td>107.67</td>
<td>115.67</td>
</tr>
<tr>
<td>SE.mean</td>
<td>35.32</td>
<td>42.03</td>
<td>40.15</td>
<td>44.93</td>
<td>51.60</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>98.06</td>
<td>116.69</td>
<td>111.47</td>
<td>124.74</td>
<td>164.20</td>
</tr>
<tr>
<td>var</td>
<td>6236.58</td>
<td>8831.63</td>
<td>8059.20</td>
<td>10093.03</td>
<td>10648.93</td>
</tr>
<tr>
<td>std.dev</td>
<td>78.97</td>
<td>93.98</td>
<td>89.77</td>
<td>100.46</td>
<td>103.19</td>
</tr>
<tr>
<td>LTR</td>
<td>122.33</td>
<td>156.33</td>
<td>131.00</td>
<td>130.17</td>
<td>93.67</td>
</tr>
</tbody>
</table>

Table 5.9: Descriptive statistics for lip reed sound generation

Table 5.10 indicates the behaviour of instruments using the mechanical reed for sound generation, such as saxophones (alto and tenor) and bassoon. Median and mean are very different and the cause may be the outliers resulting from one of the alto saxophones. The LTR value is narrower compared to the group of the lip reed sound generation instruments and varies greatly according to the tempi. The confidence interval mean shows a large dispersion of latency values around the mean. Figure 5.27 clearly displays the origin of the dispersion, namely one of the alto saxophone instruments.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>min</td>
<td>110.33</td>
<td>59.33</td>
<td>57.67</td>
<td>50.50</td>
<td>35.00</td>
</tr>
<tr>
<td>max</td>
<td>300.00</td>
<td>259.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>range</td>
<td>189.67</td>
<td>199.67</td>
<td>242.33</td>
<td>249.50</td>
<td>265.00</td>
</tr>
<tr>
<td>median</td>
<td>132.00</td>
<td>181.33</td>
<td>103.67</td>
<td>101.00</td>
<td>76.67</td>
</tr>
<tr>
<td>mean</td>
<td>183.37</td>
<td>160.53</td>
<td>149.40</td>
<td>132.70</td>
<td>113.87</td>
</tr>
<tr>
<td>SE.mean</td>
<td>40.21</td>
<td>33.52</td>
<td>45.65</td>
<td>43.69</td>
<td>47.75</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>111.64</td>
<td>93.07</td>
<td>126.75</td>
<td>121.30</td>
<td>132.57</td>
</tr>
<tr>
<td>var</td>
<td>8084.26</td>
<td>5618.92</td>
<td>10449.97</td>
<td>9541.59</td>
<td>11400.03</td>
</tr>
<tr>
<td>std.dev</td>
<td>89.91</td>
<td>74.96</td>
<td>102.08</td>
<td>97.69</td>
<td>106.77</td>
</tr>
<tr>
<td>LTR</td>
<td>146.50</td>
<td>62.33</td>
<td>129.87</td>
<td>44.67</td>
<td>42.33</td>
</tr>
</tbody>
</table>

Table 5.10: Descriptive statistics for mechanical reed sound generation
Findings and Analysis

Table 5.11 is the summary for the bowed chordophones such as the double bass, cello and violin. Median and mean values are very similar with differences of up to only 10 points. Due to the large number of observations compared to other sound generation methods, the effect of outliers is minimal. Variation in the confidence interval and standard deviation are not obvious. The LTR is stable and tends to decrease towards faster tempi. These effects can be observed in Figure 5.30.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
<td>7.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>min</td>
<td>156.50</td>
<td>158.67</td>
<td>132.33</td>
<td>97.67</td>
<td>71.00</td>
</tr>
<tr>
<td>max</td>
<td>288.67</td>
<td>266.33</td>
<td>300.00</td>
<td>288.67</td>
<td>234.33</td>
</tr>
<tr>
<td>range</td>
<td>132.17</td>
<td>107.67</td>
<td>167.67</td>
<td>191.00</td>
<td>163.33</td>
</tr>
<tr>
<td>median</td>
<td>234.00</td>
<td>227.33</td>
<td>177.00</td>
<td>153.00</td>
<td>132.00</td>
</tr>
<tr>
<td>mean</td>
<td>19.44</td>
<td>16.01</td>
<td>21.91</td>
<td>23.34</td>
<td>20.92</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>47.56</td>
<td>39.18</td>
<td>53.60</td>
<td>57.11</td>
<td>51.18</td>
</tr>
<tr>
<td>var</td>
<td>2644.06</td>
<td>1794.68</td>
<td>3358.96</td>
<td>3813.70</td>
<td>3062.37</td>
</tr>
<tr>
<td>std.dev</td>
<td>51.42</td>
<td>42.36</td>
<td>57.96</td>
<td>61.76</td>
<td>55.34</td>
</tr>
<tr>
<td>LTR</td>
<td>76.5</td>
<td>68.67</td>
<td>52.66</td>
<td>38.83</td>
<td>55.50</td>
</tr>
</tbody>
</table>

Table 5.11: Descriptive statistics for bowed sound generation

The results in Table 5.12 give evidence of the very small number of observations for plucked sound generation. Very different values were obtained for the same type of instrument (the classical guitar). Mean and median are the same due to the small amount of information. On the other hand, values such as the confidence interval contain minimal information. The LTR variation for tempi 90 BPM up to 150 BPM is very small.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>min</td>
<td>131.00</td>
<td>106.33</td>
<td>94.00</td>
<td>96.33</td>
<td>61.67</td>
</tr>
<tr>
<td>max</td>
<td>277.33</td>
<td>264.00</td>
<td>248.00</td>
<td>192.00</td>
<td>172.33</td>
</tr>
<tr>
<td>range</td>
<td>146.33</td>
<td>163.67</td>
<td>154.00</td>
<td>95.67</td>
<td>110.67</td>
</tr>
<tr>
<td>median</td>
<td>276.67</td>
<td>214.50</td>
<td>149.33</td>
<td>138.67</td>
<td>130.33</td>
</tr>
<tr>
<td>mean</td>
<td>192.94</td>
<td>163.78</td>
<td>142.33</td>
<td>118.11</td>
<td>118.11</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>209.40</td>
<td>208.51</td>
<td>193.79</td>
<td>119.09</td>
<td>137.54</td>
</tr>
<tr>
<td>var</td>
<td>7105.44</td>
<td>7045.18</td>
<td>6085.48</td>
<td>2298.11</td>
<td>3065.48</td>
</tr>
<tr>
<td>std.dev</td>
<td>84.29</td>
<td>83.94</td>
<td>78.01</td>
<td>47.94</td>
<td>55.37</td>
</tr>
<tr>
<td>LTR</td>
<td>73.17</td>
<td>81.83</td>
<td>77.00</td>
<td>47.83</td>
<td>55.33</td>
</tr>
</tbody>
</table>

Table 5.12: Descriptive statistics for plucked sound generation

The summary of descriptive statistics for the struck chordophones (piano) is presented in Table 5.13. Mean and median are no more than 20 points apart. The LTR is narrow
and decreases with increasing tempi up to 150 BPM. For the tempi 180 BPM and 210 BPM the LTR remains stable. The confidence interval is larger for slower tempi and very narrow for the 150 BPM values. Figures 5.27 and 5.30 (the C.Struck chart) include a visual summary of the data for the piano.

<table>
<thead>
<tr>
<th>Tempo</th>
<th>90 BPM</th>
<th>120 BPM</th>
<th>150 BPM</th>
<th>180 BPM</th>
<th>210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr.val</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>nbr.null</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>nbr.na</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>min</td>
<td>174.67</td>
<td>172.33</td>
<td>162.67</td>
<td>132.00</td>
<td>100.33</td>
</tr>
<tr>
<td>max</td>
<td>264.00</td>
<td>244.00</td>
<td>186.33</td>
<td>181.00</td>
<td>145.33</td>
</tr>
<tr>
<td>range</td>
<td>89.33</td>
<td>71.67</td>
<td>23.67</td>
<td>49.00</td>
<td>45.00</td>
</tr>
<tr>
<td>median</td>
<td>216.00</td>
<td>178.00</td>
<td>176.00</td>
<td>153.67</td>
<td>129.67</td>
</tr>
<tr>
<td>mean</td>
<td>218.22</td>
<td>198.11</td>
<td>175.00</td>
<td>155.56</td>
<td>125.11</td>
</tr>
<tr>
<td>SE.mean</td>
<td>25.81</td>
<td>23.00</td>
<td>6.85</td>
<td>14.18</td>
<td>13.19</td>
</tr>
<tr>
<td>CI.mean.0.95</td>
<td>111.06</td>
<td>98.97</td>
<td>29.47</td>
<td>61.00</td>
<td>56.75</td>
</tr>
<tr>
<td>var</td>
<td>1998.81</td>
<td>1587.37</td>
<td>140.78</td>
<td>602.93</td>
<td>521.81</td>
</tr>
<tr>
<td>std.dev</td>
<td>44.71</td>
<td>39.84</td>
<td>11.86</td>
<td>24.55</td>
<td>22.84</td>
</tr>
<tr>
<td>LTR</td>
<td>44.67</td>
<td>35.83</td>
<td>11.83</td>
<td>24.50</td>
<td>22.50</td>
</tr>
</tbody>
</table>

Table 5.13: Descriptive statistics for struck sound generation (chordophones)

The information for the struck sound generation of membraphones and idiophones has already been presented in Tables 5.6 and 5.7. For the membranophones (struck sound generation) the median increases by around 40 points from 90 BPM to 120 BPM, which is contrary to the inverse linear relationship between tempi and latency. A possible explanation for this result may be the very short attack time of the membranophones (in this research, the snare drum and timpani).

Idiophones such as the triangle and marimba are both struck to produce sound. However, the behaviour of these instruments seems to be very heterogeneous with regard to the ability to cope with latency. The marimba player was able to perform with latency by playing louder. On the other hand, the strategy of the triangle performers was to mute the strokes (see Appendix L). Needless to say, triangle performers are not mainly trained in the triangle but in percussion instruments in general.

Summary

- Grouping western musical instruments according to the sound generation method enables a more precise visual classification.
• The concept of latency tolerance range (LTR) can be applied to the different sound generation methods. However, more data is necessary in order to obtain reliable results.

Discussion

The latency tolerance range can be expanded to make comparisons between different sound generation methods, which translate in comparisons within each musical instrument group (aerophones and chordophones). However, bigger amounts of data are necessary for reliable results.

5.1.5 Metronome: Aural, Visual and Both

All the previous information is based on the averaged data of the three metronomes used in the listening experiment: aural, visual and both. It was assumed that the difference in the results of using one or another metronome was not significant. This assumption can be visually confirmed in Figure 5.31. It is evident that the lines representing the different metronomes are very similar for the majority of musical instruments tested in this research. A numerical test is presented in the next section.
Findings and Analysis

Figure 5.31: Latency values ($L_d$) according to the three metronomes (aural, visual and both) for each instrument.

Figure 5.31 separates the information of latency values according to the metronome used for each instrument type. The majority of the plots show very similar values for the different metronomes with the exception of the marimba, bassoon and alto saxophone. In the case of the marimba (only one instrument) the different metronomes did indeed have an effect on the ability to cope with latency. For the alto saxophone the reason is mainly due to averaging between two alto saxophones, and one of them can be considered an outlier. For the bassoon there are two possible explanations. Firstly, the use of a visual metronome was very difficult for the musician. Secondly, the breath pattern of the musician changed after each trial. At the last trial with the metronome (both), an increase of the latency values at the lower tempi was perceived.
Figure 5.32: Metronome results (aural, visual and both) for all instruments

A complete average of all latency measurements at all tempi for the three different metronomes is presented in Figure 5.32. Interestingly, when using both aural and visual metronomes at the same time (green line), musicians were able to perform with larger latencies in their headphones signal.

A difference is observed between the aural and the visual metronome around 150 BPM. At this tempo the use of a visual metronome allows higher latency values than when compared to the aural metronome, even though the majority of test subjects had a strong preference for the aural metronome in comparison to the visual metronome, as shown in the bar plot of Figure 5.33.
Findings and Analysis

Figure 5.33: Bar graph of metronome preference of the test subjects according to the instrument group

Figure 5.33 indicates a strong preference of test subjects for the aural metronome. More than half the musicians stated that the aural metronome was easy to follow. The visual metronome was only preferred by one cello player.

Summary

- The three different metronomes used in the listening test (aural, visual and both) have no obvious effect when visually compared with each other. Data gathered for the three metronomes can be averaged.

- The majority of musicians were more comfortable using the aural metronome.

Discussion

As shown in Figure 5.31 the three different metronomes can be averaged in most of the cases. This constitutes a very important result regarding the methodology proposed even knowing that the majority of musicians preferred the aural metronome (Figure 5.33). Both results are not contradictory. On the other hand, it shows the high adaptability of
Findings and Analysis

humans to different conditions as discussed by Scharer in her dissertation [203].

5.1.6 Cluster Visual Analysis

By plotting all the data, without averaging for the different metronomes, it is possible to display the information in a scatter plot matrix split into the four instruments groups or into the different sound generation methods.

![Figure 5.34: Musical instrument group clusters for 90 BPM to 210 BPM](image)

For this research the visual cluster analysis\(^3\) presented in Figure 5.34 and Figure 5.35 is an additional analysis of the data distribution in order to explore and confirm the presence of information in clusters. The clusters represent the instrument musical groups (Figure 5.34) and the sound generation methods (Figure 5.35). Further data may be necessary to draw conclusions. However, tendencies are recognizable and some patterns are identifiable.

\(^3\) Horizontal and vertical axis for Figures 5.34 and 5.35 are latency ($L_d$) in milliseconds. Data in the diagonal shows only the shape of the data distribution (KDE).
In Figure 5.34 the clusters for the four musical instrument groups are easily identified by their different colours. The diagonal with the density plots is a summary of the KDE plots from the previous section. For this kind of numerical multivariate plot, the scale of the density plot is not the same for all KDE plots and only the distribution shape is relevant. In the upper diagonal Kendall’s correlation coefficients ($\tau$) are an indicator for comparing the mathematical relationship of the distributions and confirming the results of medium (+/-0.3) and large (+/-0.5) effects for adjacent tempi.

![Figure 5.35: Sound generation method clusters for 90 BPM to 210 BPM](image)

In Figure 5.35 the four main instruments groups are split into eight subgroups according the sound generation method. The amount of information provided in the visual patterns is insufficient to draw conclusions from the data. However, the clusters indicating different sound generation methods are distributed in specific areas of the plot. Moreover, struck membranophones are identifiable as a sound generation method where the ability to cope with latency is very reduced. The density information on the diagonal as well as the correlation coefficient cannot be used to analyse the data. Information is based on less than three observations for some of the sound generation methods.
Summary

- A further qualitative analysis using the information of the musical instrument groups and the sound generation methods may be interesting. However, more observations are necessary to identify obvious patterns.

Discussion

The generated plots in Figure 5.34 and 5.35 outline the tendencies of clustering with regard to musical instrument groups and sound generation. Even with the amount of data, the clusters are identifiable. However more data is necessary to obtain a differentiated picture.

5.2 Inferential Analysis

As seen in the previous section, several relationships became obvious in the data plots. A further step is to evaluate these relationships in a quantitative way. Before choosing a statistical analysis test it is necessary to summarize the information of the listening test.

In Chapter 3 the research design proposed for the experiment was the repeated-measures design. In this design the same subjects take part in every condition of the experiment or provide data at different measurement points [93]. This design implies that the relationship between pairs of experimental conditions is comparable or, in other words, the dependence level between the different experimental conditions is similar [93, 26], therefore, the samples are dependent. Furthermore, when the research design enables the comparison of specific characteristics, the samples are also dependent [85]. When attempting to determine latency ($L_d$), subjects were tested for performing at different tempi (90 BPM to 210 BPM).
Purposive sampling was chosen in the research design. Subjects were musicians who were able to read a score. Regardless of the non-probabilistic sampling, the analysis of inferential statistics may be applied [82]. However, care has to be taken with any generalization of the outcomes. As stated in Chapter 3, the purposive sample may represent a random sample of a virtual population [26]; subjects were not chosen for convenience, but rather for their ability to play a score, and this selection was purely random. In this context, it is possible to assume random samples [26]. Statements may be referred to the virtual population.

Summary

With the information displayed in Table 5.14, in addition to the assumption of non-normality of data distribution as presented in Chapter 5.1, the choice for a statistical test can be simplified. In Table 5.14 the description of the characteristics of the data applies to the total amount of numerical gathered data.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution</td>
<td>non-normal</td>
</tr>
<tr>
<td>Scale</td>
<td>interval (ratio with 0 point)</td>
</tr>
<tr>
<td>Samples</td>
<td>dependent</td>
</tr>
<tr>
<td>Sampling design</td>
<td>non-probabilistic (purposive)</td>
</tr>
</tbody>
</table>

Table 5.14: Summary of the characteristics of the data

5.2.1 Metronomes and Tempi

Figure 5.36 shows the box plots for the five different tempi of the listening test (90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM). For each tempo, the distribution of latency values is similar for each of the three different metronomes used (aural, both and visual).
Findings and Analysis

In Appendix J, the Tables J.1, J.2, J.3, J.4 and J.5 show the values obtained for each tempo using a specific metronome. In order to establish whether there is any difference in the data when using an aural, visual or aural-visual metronome, a statistical test should be performed. With a non-normal distribution, a non-parametric should be used. The only test that can be considered for the data distribution assumptions in this case is the Friedman test.

For the Friedman test data is organized in a matrix. Table J.1 gives the results obtained with the three different metronomes at the tempo 90 BPM. The test requires the elimination of every row having missing values (NA). Note that the number of subjects varies between 24 and 25. NA values were omitted.

Each of the five different tempi are represented in a different table. The Friedman test uses the numerical information of each table. A null hypothesis can be formulated. The null hypothesis for the test states that there is no observable effect produced by any of the different metronome types (aural, visual or both). The p-value is significant if \( p < 0.05 \), otherwise the null hypothesis is supported.

Every subject in Tables J.1 to J.5 representing every musical instrument was assumed to be an independent block for testing differences between the three different metronomes.
(aural, visual and both). The aural, visual and aural-visual metronomes are the three different effects to be tested. Subjects are the rows of the matrix used to introduce the information in the Friedman test’s mathematical procedure. Each metronome (aural, visual and both) represents a different column in the matrix. Table 5.15 shows the results of the Friedman test for the data distribution at each tempo for the three different metronomes.

<table>
<thead>
<tr>
<th>Tempo (BPM)</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>1.6750</td>
<td>2</td>
<td>0.43280</td>
</tr>
<tr>
<td>120</td>
<td>4.7368</td>
<td>2</td>
<td>0.09363</td>
</tr>
<tr>
<td>150</td>
<td>7.3895</td>
<td>2</td>
<td>0.02485</td>
</tr>
<tr>
<td>180</td>
<td>4.9451</td>
<td>2</td>
<td>0.08437</td>
</tr>
<tr>
<td>210</td>
<td>5.0112</td>
<td>2</td>
<td>0.08163</td>
</tr>
</tbody>
</table>

Table 5.15: Friedman test results for the three different metronomes at each tempi

Relevant values are the chi-square statistic ($\chi^2$), the p-value and the df (degrees of freedom). The df is always 2, which is the column number minus 1. There are three different metronomes, one per column. The p-value is significant for $p < 0.05$. For all p-values higher than 0.05, the effect of the different metronome types (aural, visual and both) is meaningless. Only for the metronomes at tempo 150 BPM is $p < 0.05$. For all other tempi, the different metronomes have no effect on the latency values. Thus, the null hypothesis is rejected. Figure 5.11 is a visual confirmation of this result.

For the tempo at 150 BPM, a post-hoc test is necessary to identify if there is actually any difference between metronomes. The post-hoc test chosen for the pairwise comparison is based on the friedmanmc$^4$ function of the software R. Results are labelled as TRUE or FALSE depending on the statistical significance of the comparisons [93]. Table 5.16 shows the results. The post hoc test only found significant differences for the visual metronome compared with the aural-visual metronome at 150 BPM.

---

$^4$ The friedmanmc function is the chosen post-hoc test for all inferential analysis in Chapter 5.
Findings and Analysis

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>obs.dif.</th>
<th>critical dif.</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>4.5</td>
<td>17.26323</td>
<td>FALSE</td>
</tr>
<tr>
<td>1-3</td>
<td>13.5</td>
<td>17.26323</td>
<td>FALSE</td>
</tr>
<tr>
<td>2-3</td>
<td>18</td>
<td>17.26323</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Table 5.16: Post hoc Friedman test for metronomes at 150 BPM

Summary

- Averaging the three different metronomes is a valid procedure. In 15 pairwise comparisons (three different metronomes at five different tempi) only a significant difference for the tempo 150 BPM between the visual and aural-visual metronome was found.

Discussion

A non-parametrical inferential analysis with subsequent Post hoc tests shows the non-existence of an effect in the results due to the use of an aural, visual or aural-visual metronome. Results obtained with one of those metronomes can be averaged.

5.2.2 Hypothesis Testing

In order to accept or reject the null hypothesis ($H_0$) formulated in the Introduction, it is necessary to analyse the numerical information of the test. At the beginning of this section the characteristics of the data were summarized. The Friedman test was the analytical tool chosen.

The formulated hypothesis $H_0$ and the alternative hypothesis $H_1$ presented at the beginning of this work are:

1. **Null hypothesis $H_0$: The role of the musical instrument is not relevant with respect to the ability to cope with latency.**

2. **Alternative hypothesis $H_1$: The ability to cope with latency when performing music is directly related to the musical instrument played.**
The information in Table 4.7 is used to analyse the role of the instrument and the latency measured at five different tempi. It is possible to transpose the matrix, in other words, swap the row corresponding to the factor participant-musical instrument and the columns with the factor tempi. The Friedman test is similar to a two-way ANOVA, taking account of the effect of two possible factors at the same time [71]. For each calculation, the Friedman test can evaluate the significance of just one factor by dividing the data into blocks [71]. This means that the information in Table 4.7 is organized in a matrix and transposed. The results of the Friedman test are shown in the Tables 5.17 and 5.18. NA values were omitted. The data of similar instruments (e.g. piano, cello) was not averaged and each piano or cello represents a column of the matrix. The blocks are the different tempi and every tempo is the average of three different metronomes.

Two errors associated with the null hypothesis results are Type I & II errors. In Type I errors the null hypothesis is wrongly rejected, while in Type II errors the null hypothesis is mistakenly not rejected. Type I errors can occur due to [117]:

- Very small sample populations.
- Selection of an incorrect statistical test.

As stated before, a non-normal distribution implies the use of a non-parametrical test, such as the Friedman test. The results of the test are described in Table 5.17. The Friedman test involves an analysis of variance by ranks and is performed if the samples were obtained through a repeated measures design [87], as in the case of this research. It is important to be aware that the Friedman test may produce inaccurate p-values for small sample sizes [87].

<table>
<thead>
<tr>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.523</td>
<td>30</td>
<td>2.208e-09</td>
</tr>
</tbody>
</table>

Table 5.17: Friedman test results supporting the alternative hypothesis $H_1$

The p-value is very significant. The null hypothesis $H_0$: The role of the musical instrument is not relevant with respect to the ability to cope with latency, can be rejected and the alternative hypothesis $H_1$: The ability to cope with latency when performing music is directly related to the musical instrument played, is supported.
However, it is not possible to gather more information from this result. It is necessary to run a post-hoc test to obtain more detailed information. Post-hoc tests are pairwise comparisons. The performance of the post-hoc procedures is directly related to conditions such as normality of the data distribution, differences in the population variances and differences in the group sizes (unbalanced design) [93]. The data of the listening test is non-normally distributed and the design is unbalanced. In other words, each musical instrument type (e.g. piano, cello, triangle, etc.) has a different number of observations and some instruments, such as the violin were tested more often (4 violins) compared to the tests on, for e.g., the marimba (1 marimba).

The next step after calculating the p-value for the global hypothesis is to compare the pairs against each other. For this comparison, the significance level is adjusted using a Bonferroni correction [87], which is considered an accurate method in order to avoid Type I errors due to its conservative approach [93]. For this reason, the post-hoc test may not indicate significant differences in pairwise comparisons, even when the p-value (see Table 5.17) is very significant [87]. Results for all pairwise comparisons are listed in Appendix J. The comparison values where the difference is significant (TRUE) are listed in Table 5.18.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>obs.dif.</th>
<th>critical dif.</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-12</td>
<td>109.5</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>7-14</td>
<td>100.0</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>7-31</td>
<td>107.5</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>8-12</td>
<td>105.0</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>8-31</td>
<td>103.0</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>9-12</td>
<td>113.5</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>9-14</td>
<td>104.0</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>9-31</td>
<td>111.5</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
<tr>
<td>12-21</td>
<td>100.0</td>
<td>99.59828</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Table 5.18: Post-hoc test significant results for the pairwise comparisons

The comparisons are between columns. For this specific case, the number of the column matches the number of the participant. As expected, the statistical power of the Friedman test compared to other parametric tests, such as ANOVA, is not sufficient to detect every
relevant effect, especially when using the post-hoc test. An alternative, as suggested by M. Gardener [105], is to present a box plot of the data. Figure 5.37 shows the values obtained for every test subject (musical instrument).

Figure 5.37: Latency ($L_d$) vs. musical instrument (average of all 3 metronomes)

Figure 5.37 represents every single instrument evaluated in the listening test. The differences between the instruments even for the same type of instrument (i.e. the piano compared to the other pianos) are easy to identify.

From the Table 5.18, significant differences exist between the instruments 7 and 12 (alto saxophone and snare drum), 7 and 14 (alto saxophone and transverse flute), 7 and 31 (alto saxophone and tenor saxophone), 8 and 12 (violin and snare drum), 8 and 31 (violin and tenor saxophone), 9 and 12 (trumpet in B and snare drum), 9 and 14 (trumpet in B and transverse flute), 9 and 31 (trumpet in B and tenor saxophone), and 12 and 21 (snare drum and classical guitar). Thus, the snare drum, the alto saxophone and the trumpet in B are significantly different in their characteristics when compared to the other instruments. A possible explanation might be the higher latency ($L_d$) values obtained with these instruments.
There are some obvious limitations to the statements. The major drawback is the unbalanced design. The number of musical instruments is not the same for every comparison. This has an effect on the calculation of the p-value. On the other hand, the p-value is extremely low, meaning there is a real difference between musical instruments with regard to the ability to cope with latency. According to the data, the differences of some pairwise comparisons are larger than others and some are very significant (Table 5.18).

<table>
<thead>
<tr>
<th>Subject</th>
<th>Instrument</th>
<th>Instrument group</th>
<th>Sound generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Piano</td>
<td>Chordophones</td>
<td>Struck</td>
</tr>
<tr>
<td>2</td>
<td>Piano</td>
<td>Chordophones</td>
<td>Struck</td>
</tr>
<tr>
<td>3</td>
<td>Cello</td>
<td>Chordophones</td>
<td>Bowed</td>
</tr>
<tr>
<td>4</td>
<td>Cello</td>
<td>Chordophones</td>
<td>Bowed</td>
</tr>
<tr>
<td>5</td>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Plucked</td>
</tr>
<tr>
<td>6</td>
<td>French horn</td>
<td>Aerophones</td>
<td>Lip reed</td>
</tr>
<tr>
<td>7</td>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
</tr>
<tr>
<td>8</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
</tr>
<tr>
<td>9</td>
<td>Trumpet in B</td>
<td>Aerophones</td>
<td>Lip reed</td>
</tr>
<tr>
<td>10</td>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Struck</td>
</tr>
<tr>
<td>11</td>
<td>Piano (upright)</td>
<td>Chordophones</td>
<td>Struck</td>
</tr>
<tr>
<td>12</td>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Struck</td>
</tr>
<tr>
<td>13</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
</tr>
<tr>
<td>14</td>
<td>Transverse flute</td>
<td>Aerophones</td>
<td>Air reed</td>
</tr>
<tr>
<td>15</td>
<td>Trombone</td>
<td>Aerophones</td>
<td>Lip reed</td>
</tr>
<tr>
<td>16</td>
<td>Timpani</td>
<td>Membranophones</td>
<td>Struck</td>
</tr>
<tr>
<td>17</td>
<td>Trombone</td>
<td>Aerophones</td>
<td>Lip reed</td>
</tr>
<tr>
<td>18</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
</tr>
<tr>
<td>19</td>
<td>Triangle</td>
<td>Idiophones</td>
<td>Struck</td>
</tr>
<tr>
<td>20</td>
<td>Tenor saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
</tr>
<tr>
<td>21</td>
<td>Classical guitar</td>
<td>Chordophones</td>
<td>Plucked</td>
</tr>
<tr>
<td>22</td>
<td>Violin</td>
<td>Chordophones</td>
<td>Bowed</td>
</tr>
<tr>
<td>23</td>
<td>Snare drum</td>
<td>Membranophones</td>
<td>Struck</td>
</tr>
<tr>
<td>24</td>
<td>Alto saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
</tr>
<tr>
<td>25</td>
<td>Triangle</td>
<td>Idiophones</td>
<td>Struck</td>
</tr>
<tr>
<td>26</td>
<td>Marimba</td>
<td>Idiophones</td>
<td>Struck</td>
</tr>
<tr>
<td>27</td>
<td>French horn</td>
<td>Aerophones</td>
<td>Lip reed</td>
</tr>
<tr>
<td>28</td>
<td>Double bass</td>
<td>Chordophones</td>
<td>Bowed</td>
</tr>
<tr>
<td>29</td>
<td>Harp</td>
<td>Chordophones</td>
<td>Plucked</td>
</tr>
<tr>
<td>30</td>
<td>Bassoon</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
</tr>
<tr>
<td>31</td>
<td>Tenor saxophone</td>
<td>Aerophones</td>
<td>Mechanical reed</td>
</tr>
</tbody>
</table>

Table 5.19: List of the musical instruments of the test

Table 5.19 lists of all the musical instruments in the listening test together with the number (test subject), musical instrument group and sound generation method. Instruments in italic and bold are those with significant differences in the post-hoc test of Table 5.18. This summary is useful for further inferential analysis.

Comparisons between similar instruments

Box plots showing the results between similar instruments may further help with the analysis of the results. Figure 5.38 summarizes the results of the cello (2), French horn (2),
Findings and Analysis

piano (3), snare drum (3), tenor saxophone (2), triangle (2) and trombone (2). Results for each tempo from 90 BPM to 210 BPM are given. Variations within the same type of instruments are identifiable through the box plot sizes. For the tempi 90 BPM and 210 BPM the box plots of the majority of the instruments are large, while for the middle tempi, especially for 150 BPM, box plots are narrow. It can be assumed that the ability to deal with latency increases for tempi around 150 BPM.

Figure 5.38: Latency range box plots for the same type of instruments with similar latency values

Figure 5.39 compares instruments of the same type for the different tempi where latency values vary widely. The box plots have a very large range indicating very different results for the same type of instrument. In the case of alto saxophone (2), classical guitar (2) and violin (4), the role of outliers may be decisive towards these results.
Summary

- The alternative hypothesis ($H_1$) is supported: *The ability to cope with latency when performing music is directly related to the musical instrument played.*

- Pairwise comparisons in a post-hoc test showed significant differences between some instruments. Results are confirmed with a visual approach using box plots.

Discussion

The assumption that different musical instruments might have an influence regarding latency is well documented qualitatively in research done by Boley and Lester [152], Kleimola [141] and Rottondi et al. [199], just to cite a few. In this research, the numerical data supports the alternative hypothesis. Furthermore, the statistical analysis establishes a quantitative relationship for significant differences between different musical instruments.
5.2.3 Tempi Effect Inferential Analysis

A last analysis is the comparison of the effects produced by the different tempi at 90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM. Table 5.20 lists the instruments (rows) and latency values (columns) at the different tempi. As for the other Friedman test matrices, rows with missing values (NA) were omitted, in this case the trombone, instrument number 15, as in Table 4.7. Before the Friedman test, the null hypothesis can be formulated as: tempo has no significant effect on latency results.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Avg. 90 BPM</th>
<th>Avg. 120 BPM</th>
<th>Avg. 150 BPM</th>
<th>Avg. 180 BPM</th>
<th>Avg. 210 BPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>216.00</td>
<td>172.33</td>
<td>186.33</td>
<td>181.00</td>
<td>145.33</td>
</tr>
<tr>
<td>2</td>
<td>174.67</td>
<td>178.00</td>
<td>162.67</td>
<td>132.00</td>
<td>100.33</td>
</tr>
<tr>
<td>3</td>
<td>280.00</td>
<td>241.67</td>
<td>132.33</td>
<td>97.67</td>
<td>71.00</td>
</tr>
<tr>
<td>4</td>
<td>234.00</td>
<td>227.33</td>
<td>218.67</td>
<td>166.00</td>
<td>195.33</td>
</tr>
<tr>
<td>5</td>
<td>131.00</td>
<td>100.33</td>
<td>94.00</td>
<td>96.33</td>
<td>61.67</td>
</tr>
<tr>
<td>6</td>
<td>300.00</td>
<td>272.33</td>
<td>158.67</td>
<td>107.67</td>
<td>91.00</td>
</tr>
<tr>
<td>7</td>
<td>300.00</td>
<td>259.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>8</td>
<td>288.67</td>
<td>254.33</td>
<td>300.00</td>
<td>288.67</td>
<td>234.33</td>
</tr>
<tr>
<td>9</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
<td>300.00</td>
</tr>
<tr>
<td>10</td>
<td>145.00</td>
<td>137.33</td>
<td>93.50</td>
<td>83.33</td>
<td>77.67</td>
</tr>
<tr>
<td>11</td>
<td>264.00</td>
<td>244.00</td>
<td>176.00</td>
<td>153.67</td>
<td>129.67</td>
</tr>
<tr>
<td>12</td>
<td>71.00</td>
<td>74.67</td>
<td>67.33</td>
<td>60.00</td>
<td>71.00</td>
</tr>
<tr>
<td>13</td>
<td>223.33</td>
<td>177.67</td>
<td>180.33</td>
<td>137.67</td>
<td>114.33</td>
</tr>
<tr>
<td>14</td>
<td>113.00</td>
<td>86.00</td>
<td>78.00</td>
<td>76.67</td>
<td>19.00</td>
</tr>
<tr>
<td>15</td>
<td>91.33</td>
<td>131.00</td>
<td>96.00</td>
<td>48.00</td>
<td>22.00</td>
</tr>
<tr>
<td>16</td>
<td>177.67</td>
<td>116.00</td>
<td>88.67</td>
<td>79.50</td>
<td>73.33</td>
</tr>
<tr>
<td>17</td>
<td>280.00</td>
<td>266.33</td>
<td>177.00</td>
<td>166.67</td>
<td>132.00</td>
</tr>
<tr>
<td>18</td>
<td>244.67</td>
<td>227.50</td>
<td>131.50</td>
<td>60.50</td>
<td>45.00</td>
</tr>
<tr>
<td>19</td>
<td>110.33</td>
<td>120.33</td>
<td>103.67</td>
<td>101.00</td>
<td>76.67</td>
</tr>
<tr>
<td>20</td>
<td>277.33</td>
<td>264.00</td>
<td>248.00</td>
<td>192.00</td>
<td>172.33</td>
</tr>
<tr>
<td>21</td>
<td>183.67</td>
<td>181.00</td>
<td>143.00</td>
<td>117.33</td>
<td>109.00</td>
</tr>
<tr>
<td>22</td>
<td>73.00</td>
<td>117.33</td>
<td>84.50</td>
<td>104.67</td>
<td>94.00</td>
</tr>
<tr>
<td>23</td>
<td>132.00</td>
<td>182.67</td>
<td>78.00</td>
<td>128.33</td>
<td>100.00</td>
</tr>
<tr>
<td>24</td>
<td>178.33</td>
<td>169.33</td>
<td>142.67</td>
<td>154.00</td>
<td>126.50</td>
</tr>
<tr>
<td>25</td>
<td>242.33</td>
<td>228.67</td>
<td>175.00</td>
<td>166.00</td>
<td>179.67</td>
</tr>
<tr>
<td>26</td>
<td>262.67</td>
<td>175.00</td>
<td>228.00</td>
<td>209.67</td>
<td>140.33</td>
</tr>
<tr>
<td>27</td>
<td>156.50</td>
<td>158.67</td>
<td>150.67</td>
<td>153.00</td>
<td>139.00</td>
</tr>
<tr>
<td>28</td>
<td>276.67</td>
<td>214.50</td>
<td>149.33</td>
<td>138.67</td>
<td>120.33</td>
</tr>
<tr>
<td>29</td>
<td>260.50</td>
<td>181.33</td>
<td>207.67</td>
<td>83.67</td>
<td>35.00</td>
</tr>
<tr>
<td>30</td>
<td>114.00</td>
<td>59.33</td>
<td>57.67</td>
<td>50.50</td>
<td>57.67</td>
</tr>
<tr>
<td>31</td>
<td>114.00</td>
<td>59.33</td>
<td>57.67</td>
<td>50.50</td>
<td>57.67</td>
</tr>
</tbody>
</table>

Table 5.20: Experimental data for all tempi

Table 5.21 presents the results. The p-value is significant. A further post-hoc test is necessary to estimate which pairwise comparisons are relevant.

<table>
<thead>
<tr>
<th>χ²</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>71.57</td>
<td>4</td>
<td>1.058e-14</td>
</tr>
</tbody>
</table>

Table 5.21: Friedman test results for the influence of the different tempi
The results of the post-hoc test are shown in Table 5.22 for all pairwise comparisons. The null hypothesis is rejected and the alternative hypothesis that tempo has an effect on latency is supported. Comparisons of side by side columns (e.g. 1 and 2, 2 and 3, 3 and 4 and 4 and 5) are not significant. This absence of significance between adjacent tempi may be strongly related to the large and medium Kendall’s correlation coefficients ($\tau$) in Figures 5.14. The effect of pairwise comparisons of adjacent tempi (e.g. 90 BPM and 120 BPM) is not significant. On the other hand, all other pairwise comparisons are very significant. In short, while the effect produced at 90 BPM is not so different from the effect produced at 120 BPM, the effect of the tempi 150 BPM, 180 BPM and 210 BPM are significantly different compared to the effect of 90 BPM. The same can be stated for each tempo compared to the other tempi.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>obs.dif</th>
<th>critical dif.</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>18.5</td>
<td>34.379</td>
<td>FALSE</td>
</tr>
<tr>
<td>1-3</td>
<td>44.5</td>
<td>34.379</td>
<td>TRUE</td>
</tr>
<tr>
<td>1-4</td>
<td>65.5</td>
<td>34.379</td>
<td>TRUE</td>
</tr>
<tr>
<td>1-5</td>
<td>89.0</td>
<td>34.379</td>
<td>TRUE</td>
</tr>
<tr>
<td>2-3</td>
<td>26.0</td>
<td>34.379</td>
<td>FALSE</td>
</tr>
<tr>
<td>2-4</td>
<td>47.0</td>
<td>34.379</td>
<td>TRUE</td>
</tr>
<tr>
<td>2-5</td>
<td>70.5</td>
<td>34.379</td>
<td>TRUE</td>
</tr>
<tr>
<td>3-4</td>
<td>21.0</td>
<td>34.379</td>
<td>FALSE</td>
</tr>
<tr>
<td>3-5</td>
<td>44.5</td>
<td>34.379</td>
<td>TRUE</td>
</tr>
<tr>
<td>4-5</td>
<td>23.5</td>
<td>34.379</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

Table 5.22: Post-hoc test for all pairwise comparisons for tempi 90 BPM to 210 BPM

The 30 BPM interval between tempi proved to be a good choice. Firstly, the interval includes five steps from 90 BPM up to 210 BPM which is an overview of all relevant musical tempi for any kind of musical composition. On the other hand, the effect of the tempi on the breakdown latency values is significant. In addition, a shorter interval would lead to a longer experiment duration time.

Summary

- The five different tempi values chosen for the listening test (90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM) satisfactorily cover the whole range of tempi present in western music.
Discussion

The interval of 30 BPM between tempi enables adjacent tempi to be similar with no significant statistical effect (see Table 5.22). On the other hand, comparing each tempo value with the others (except the adjacent) has a statistically significant effect (TRUE difference as presented in Table 5.22). This translates into a time effective listening test which enables to frame the whole range of musical tempi in western music.

5.3 Questionnaire Answers and Observations During the Test

Prior to the listening test, subjects were asked to answer some questions relating to their practice of the instrument. This categorical and numerical data was analysed at the beginning of the chapter. In addition to the categorical data, test subjects had the option to answer two questions after the listening test.

The first question covered the preferred metronome during the listening test. The answers are given in Figure 5.33. The second question was an open question for comments or notes and some test subjects answered this question. The information collected is presented in this section. During the listening test, the author also wrote some notes which are included in this section. The original transcriptions can be found in Appendix N.

A list of the most relevant comments from the test subjects can be listed as follows (transcriptions do not use the exact words):

- The test is difficult, concentration is necessary (piano).
- The visual metronome is very hard to follow and perform with at the same time (piano).
- The visual metronome is a distraction (guitar).
- A strategy to cope with latency was to listen to the original sound of the instrument (cello).
Findings and Analysis

- When using the aural-visual metronome most test subjects focused and used only the aural metronome. However, some musicians used the visual metronome as a starting reference and only thereafter the aural metronome.

- When the last note of the score merges with the first, it is impossible to play (French horn, flute).

- After some time it was easy to play (violin test subject with 300ms average). Data, in this case, can be considered as an outlier.

- The aural metronome was easy to follow (trumpet in B).

- Latency is very annoying (snare drum).

- The body can be used to avoid the latency drawbacks by feeling the strike (snare drum) or the sound of the instrument in the neck (violin).

- A strategy to compensate for latency was to feel the finger at the string (guitar with higher $L_d$).

- The musician was aware of the habituation effect due to the instrument characteristics. Performing with a visual metronome was very difficult (marimba).

- Some musicians had the impression that the metronome for every specific tempo was not steady within the listening test. This is a possible consequence of the delayed audio.

- Musicians playing aerophones have an additional issue while playing. With the delayed feedback in the headphones and listening to the breath, the performance is difficult (bassoon). However, the strategy adopted is to use the foot and the metronome to create a flow and perform the score.

- Having experience with contemporary musical works that use latency as a fundamental part of the performance enables the playing with latency without breaking down, even for higher latencies up to 300ms.

Being located very near to the musicians, the author was able to make some observations regarding strategies to cope with latency and other general issues. These included:

- The majority of the musicians used the foot to keep the tactus.
Findings and Analysis

- Musicians playing chordophones tried to avoid the effect of latency by playing *vibrato*. The higher the latency, the more *vibrato*.

- Musicians who could cope well with latency (300ms) were used to playing with latency, having either a vast home recording studio experience or having developed a strategy to somehow memorize rhythm and body patterns.

- When playing the triangle at faster tempi, musicians tended to mute the strokes to cope better with latency.

- Playing louder was a strategy chosen to avoid the latency issue.

Musicians and musical instruments are a unit when performing music. In addition, sensory, psychological and environmental issues influence the manner in which the instrument is played to a large degree. Having analysed the qualitative and quantitative results of this research, a clearer understanding of the playing of musical instruments and the influence of latency in non-collaborative performances is obtained.

5.4 Summary

The analysis of the quantitative data shows that there is a strong relationship between the musical instruments and latency. The alternative hypothesis is supported using statistical data. In addition, the measure of the latency tolerance range (LTR) enables the comparison between different musical instrument groups with regard to latency. The next section reviews the findings, the limitations and the contribution of this research.
Chapter 6

Conclusion

6.1 Summary

This research develops a new controllable, reliable and replicable method to investigate the relationship between western musical instruments and the ability of musicians to cope with latency in non-collaborative performances. In addition, a measurement has been developed and verified through a comprehensive statistical analysis. The outcome shows that the latency tolerance is subject to the musical instrument performed and the tempo of the performance.

Chapter 1

In the first chapter the problem related to the latency issue and musical performances is presented and necessity of a new research approach outlined. The research question, sub-questions and the hypothesis are defined. Finally, the structure of the different chapters of the dissertation is exposed.

Chapter 2

The literature review chapter introduces the most significant previous research. This research has been mainly engaged with the question of how collaborative performances are affected by latency. Up until the present, quantitative research addressing the latency issue, including the role of the musical instrument and its influence, has been rare. The majority of the studies are qualitative. To conclude this chapter, the different approaches to measure latency are presented and discussed, in particular the clarification of termi-
nology in particular the concept of “cope with latency”. The necessity of a new and standardised measure with regard to the latency tolerance for different musical instruments is outlined.

Chapter 3

The methodological approach for this research is discussed in this chapter. The measure of latency tolerance range (LTR) was developed and included. Based on the research question, a methodology is chosen and explained. The quantitative methodological approach integrates multidisciplinary branches already established in psychophysics, psychology and psychoacoustics. The result is a procedure with a listening test as the core of the experiment based on previously designed approaches and concepts analogous to musical perception, delayed auditory feedback (DAF), latency adaptive tempo (LAT) and perceptual attack tempo (PAT).

Chapter 4

In this chapter, the experimental procedure is described. All elements necessary to the conduction of the experiment such as participants, stimuli, the physical set-up and the procedure itself are presented. At the end, the pilot test and its results are analysed and further changes for the final experiment are described. The numerical results of the final test are attached to the end of this chapter, these results are the basis for the statistical approach in Chapter 5.

Chapter 5

The Findings and Analysis chapter presents the descriptive and inferential statistical results of the study. Results obtained confirm previous findings and the inverse relationship of latency vs. tempo is confirmed for a vast number of western musical instruments. The importance of sensory information as addressed by DAF experiments is witnessed and documented. In addition to the confirmation of these results, new issues are also detected:
The inverse relationship between latency and tempo is quantitatively different for each instrument as seen in Figure 5.16. In addition, this relationship is group dependent according to the data as seen in Figure 5.19.

The sound generation method of each instrument as shown in Figure 5.27 is an additional alternative for grouping musical instruments with respect to the relationship between latency and tempo.

For validating the methodology, a null hypothesis $H_0$ is tested and rejected based on non-parametric inferential statistical methods. The alternative hypothesis $H_1$ is supported based on the inferential analysis of the gathered data.

$H(1)$: The ability to cope with latency when performing music is directly related to the musical instrument played.

In other words, the observed effect is not random and the musical instrument is de facto relevant. The post-hoc test confirmed some salient differences between pairwise comparisons, especially for instruments belonging to different musical instrument groups and for instruments with very dissimilar latency values within the musical instrument groups.

This work shows that performing on different instruments leads to different latency performance breakdown values. The musician and the musical instrument are a unit that is difficult to separate. However, the control methods used in the experiment of this research, such as the score, the metronome (aural, visual and both) and the listening test conditions, allowed for the hypothesis testing. Nevertheless, every musician is unique as a human being and musical performing abilities are an integral part of this uniqueness.

Two main issues influence the generalizability of results from the proposed method. Firstly, the number of samples and secondly the sample population. With a larger number of samples, the distribution data tends towards a normal shape based on the assumptions of the central limit theorem of statistics. However, discussions about what constitutes a large enough number of samples have not been resolved. On the contrary, there is no agreement regarding this issue. Independent of how the discussion continues, economical and logistical issues will always influence the acquisition of samples. In addition, the expected effect is not small and the exploratory data analysis shows huge differences between different western musical instruments and the latency issue. For this research, the
quantitative conclusions are based on the data of 31 test subjects. Furthermore, samples represent a virtual random population.

The measurement of the latency tolerance range (LTR) is a quantitative approach. It characterizes ranges of playability of classical western musical instruments when listening to self-delay (latency) according to tempi and musical instrument groups. Its numerical character makes it very dependent on the data sample gathered. However, a lower number of observations enables general conclusions to be drawn. Moreover, a further analysis including sound generation methods was outlined (see Figure 5.30). The latency tolerance range enables the comparison of differences between musical instrument groups and even between musical instrument sound generation methods.

Results are summarised and related back to research literature. Some results are confirmatory. Nevertheless, the obtained results expand the knowledge with regard to the issue of latency and its relationship with western musical instruments.

6.2 Limitations

The data distribution is non-normal, therefore non-parametrical tests were used in order to accept or reject the hypothesis. As explained in Chapter 5, the statistical power is diminished. Moreover, the sample population implies an unbalanced design. Different musical instruments were available but not always in the same number according to instrument type. However, conservative post-hoc tests were applied and only very significant differences were presented.

As explained in Chapter 3, controlled experiments have an impact on external validity. The results can be generalized to the virtual population, based on the number of samples, representing mainly young European music students. In addition, only the results of classical western musical instruments were analysed. This delimitation was necessary, in order to approach and develop some of the control mechanisms of the experiment.
6.3 Contribution

The main contribution of this research is the development and testing of a method which enables the systematic measurement of latency in non-collaborative performances using a quantitative approach.

Using the proposed methodology, it was possible to establish relationships between the classical western musical instruments corresponding to the musical groups of chordophones, aerophones, membranophones and idiophones. Furthermore, general tendencies regarding the different sound generation methods were revealed and results for the different tempi used in music performance from 90 BPM to 210 BPM were presented. The selection of 30 BPM intervals was a good trade-off between the time duration of the listening test and common musical practice from slow to fast tempi.

The collection of categorical parameters such as gender, hours of musical practice with or without a metronome as well as years of experience by means of a questionnaire is a further step in the proposed method. The influence of these parameters on the ability of musicians to deal with latency is minimal or not present as shown in Figure 5.12.

A relationship between musical instruments and the ability to cope with latency was determined. In addition to answering the research question, this work present further findings with regard to this relationship.

The contribution of this research can be summarised as follows:

- A new method of measurement involving latency measurements on musicians is presented and developed, its results have been analysed.
- The Latency Tolerance Range (LTR) is defined as a new measure. This measure enables comparisons regarding musical tempi and latency values up to 300ms between different musical instrument groups.
- The influence of the different metronomes (aural, visual and aural-visual) is minimal. Results obtained using one or the other metronome are similar and without any significant statistical effect.
- Different musical instruments enable different latency values. The relationship latency vs. tempo varies depending on the musical instrument performed.
• The whole range of musical playability regarding musical tempo in beats per minute (BPM) can be framed from lower tempi around 90 BPM to higher tempi around 210 BPM using intervals of 30 BPM.

The presented method is new and is defined to avoid ambiguities in the measurement process. In addition, the set-up is easy and can be replicated in different environments. The control variables such as the metronome and the score are defined to eliminate any possible bias caused by internal or external influences. It has been shown that the use of different kind of metronomes (aural, visual and aural-visual) has no measurable effect with regard to the results of the experiments. On the one hand, it enables the averaging of results within a single listening test. On the other hand, it indicates that by sensory research a more holistic approach is necessary.

The relationship between latency vs. tempo for musical instruments is not only confirmed. Moreover, the relevance of the musical instrument in this relationship is shown in a descriptive and numerical approach. Similarities and differences are exposed in relation to the musical instrument groups and the sound generation methods. The quantitative methodology approach enables the development of the Latency Tolerance Range (LTR) measure. This measure is a complete new numerical description of the relationship between musical instrument groups and latency. The tempi from 90 BPM to 210 BPM specify the range of relevant musical tempi in western music. The 30 BPM increment is the best compromise to evaluate transitions between tempi.

6.4 Discussion

Numerical results are assumed to be continuous. However, discrete measures of five different tempi (90 BPM, 120 BPM, 150 BPM, 180 BPM and 210 BPM) permitted conclusions regarding the issue of latency in non-collaborative performances to be made for all results concerning all musical instruments in the sample population.

At faster tempi beginning at 150 BPM up to 210 BPM, the results for performance breakdown of western musical instruments are not so widespread when compared to the results at 90 BPM and 120 BPM. This relationship is also observed based on the Kendall’s correlation coefficients (τ) presented in Figure 5.14.
According to the results of variance in latency measurements, homogeneity in the musical groups of chordophones, membranophones and idiophones is evident. Previous research has already considered chordophones as a homogenous group on the basis of research on psychology. The number of samples per group also plays an important role. On the contrary, the bigger variance in the results of the aerophones group may be produced by two characteristics relevant to the performance of the instrument: firstly, the embouchure and secondly, breathing. Both characteristics are intrinsically related to the musician and may ameliorate the role of the musical instrument itself.

As expected, the use of aural metronomes is preferred over other alternatives, such as visual or the combination of aural and visual as presented in Figure 5.33. Deviation from the beat is considered a salient characteristic of good musicianship. The metronomes used in the listening test ameliorate the effect of tempo individuality.

Beyond the numerical results obtained for the latency tolerance range for musical instrument groups based on non-collaborative performances, further implications for musical practices over networks can be expected. Based on the key findings presented in the previous section, a comparison between musical instrument groups is possible. In other words, by comparing the different LTR values for different musical instrument groups, it can be estimated -based on the lowest LTR- which network latency may be appropriate when performing as an ensemble. This is a relevant information for composers of network music. Knowing that the influence of the different metronomes (aural, visual and aural-visual) is minimal, researchers can elaborate listening tests combining these cues and develop further investigations with regard to the adaptability of musicians. Tempi ranges play an important role in the relationship musical instrument and latency. The results of this study may help software engineers to develop tools. Producers, recording engineers and musicians may be able to identify technical issues sooner and elaborate a better planning of musical projects involving networks e.g. satellite links, monitoring or remote studio productions.

The awareness of the differences between musical instruments regarding latency enables a better scheduling for musical session arrangements that include different musical instrument groups. Furthermore, the latency measurement method presented in this work can be used and adapted for experimental research involving musicians and different stimuli.
6.5 Outlook

The methodology presented was evaluated on classical western musical instruments. It is possible to adapt some of the control mechanisms, such as the score and the range of musical instruments that form part of the listening test. In doing so, the external validity of the results can be increased.

While performing on musical instruments, visual, aural and tactile sensory feedback is extremely important. The metronome used as a control mechanism in this research is based on aural and visual cues. It may be interesting to include a metronome based on tactile cues. Nowadays, vibrating metronome bracelets or more general wearable metronomes are available and it might be worth knowing how reliable a tactile metronome performs as a control method. In addition, the holistic approach can be further expanded.

The simplified acoustic instrument model presented in Chapter 2 may be the first step to modelling western musical instruments with regard to the latency issue. Variables such as room and direct sound can be controlled in order to analyse the full influence of the body parts (fingers, hands and feet), the embouchure in aerophones and, in some cases, the playing tool. A next step may be to expand the cluster visual analysis outlined in the Findings and Analysis Chapter. Larger data samples are necessary to tackle this task. This could be helpful for the classification and recognition of patterns and characteristics that allow the further development of mathematical models intended to a more accurate calculation of the latency tolerance range with regard to specific musical tempi. Besides that, the results can be used as suggestions for the set-up of quality of service (QoS) values in network components. Software engineers could enhance and implement further control items in network gear such as audio interfaces or network switches. Digital audio workstations (DAW) may enable to include latency values for live monitoring or remote session recordings.

The author is convinced that the use of network technologies in music continues expanding. This research provides a solid scientific basis for future investigation. Supplementary research in order to understand external influences on the issues of latency affecting the performance of music should be further addressed.
Bibliography


BIBLIOGRAPHY


[158] Teemu Mäki-Patola and Perttu Hämäläinen. Latency tolerance for gesture controlled continuous sound instrument without tactile feedback. *Laboratory of


251


[201] A. A. Sawchuk, Elaine Chew, R. Zimmermann, C. Papadopoulos, and C. Kyriakakis. From remote media immersion to distributed immersive performance. In


