Queen Mary
University of London

Discriminating music performers by timbre: On the relation between instrumental gesture, tone quality and perception in classical cello performance

CHUDY, M

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/18378

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

Discriminating music performers by timbre:

On the relation between instrumental gesture, tone quality and perception in classical cello performance

Magdalena Chudy

Thesis submitted in partial fulfilment
of the requirements of the University of London
for the Degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science,
Queen Mary University of London

2016

I, Magdalena Chudy, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: June 14, 2016

Details of collaboration and publications:
All collaborations and the author's previous publications related to this thesis are described in Section 1.3.

Abstract

Classical music performers use instruments to transform the symbolic notation of the score into sound which is ultimately perceived by a listener. For acoustic instruments, the timbre of the resulting sound is assumed to be strongly linked to the physical and acoustical properties of the instrument itself. However, rather little is known about how much influence the player has over the timbre of the sound — is it possible to discriminate music performers by timbre?

This thesis explores player-dependent aspects of timbre, serving as an individual means of musical expression. With a research scope narrowed to analysis of solo cello recordings, the differences in tone quality of six performers who played the same musical excerpts on the same cello are investigated from three different perspectives: perceptual, acoustical and gestural.

In order to understand how the physical actions that a performer exerts on an instrument affect spectro-temporal features of the sound produced, which then can be perceived as the player's unique tone quality, a series of experiments are conducted, starting with the creation of dedicated multi-modal cello recordings extended by performance gesture information (bowing control parameters). In the first study, selected tone samples of six cellists are perceptually evaluated across various musical contexts via timbre dissimilarity and verbal attribute ratings. The spectro-temporal analysis follows in the second experiment, with the aim to identify acoustic features which best describe varying timbral characteristics of the players. Finally, in the third study, individual combinations of bowing controls are examined in search for bowing patterns which might characterise each cellist regardless of the music being performed.

The results show that the different players can be discriminated perceptually, by timbre, and that this perceptual discrimination can be projected back

through the acoustical and gestural domains.

By extending current understanding of human-instrument dependencies for qualitative tone production, this research may have further applications in computer-aided musical training and performer-informed instrumental sound synthesis.

Acknowledgements

This work would not be possible without the support of many people around the globe whom, using this opportunity, I would like to thank with all my heart.

First of all, I would like to express my sincere gratitude to my supervisor, Simon Dixon, for his guidance, sound advice and friendly support through the course of my PhD. I would also like to thank my secondary supervisors, Anssi Klapuri and Andrew Robertson, for their inspirational feedback at different stages of my research work.

It was my privilege to be part of the Centre for Digital Music. Special thanks to Mark Sandler and Mark Plumbley for "taking me on board" and giving me the opportunity to pursue my research ideas. I am grateful to all my C4DM peers and senior colleagues for their warm and friendly welcome, instant help and advice when it was much needed and simply, for being such a vibrant and open-minded community.

Many thanks to Andrew Simpson and Asterios Zacharakis who not only shared their experience on designing perceptual studies but also readily offered their help and instruction when my own experiment was in preparation.

I would like to thank Music Technology Group researchers and Xavier Serra in particular for making the creation of the cello database possible and for being wonderful hosts during my stay in Barcelona. I am especially grateful to Alfonso Pérez-Carrillo, Marco Marchini and Panagiotis Papiotis who patiently acquainted me with the intricacies of the sensing electronics and then devotedly assisted me through all the recording sessions.

Statistical analyses carried out in this thesis would not fall into place without invaluable insights from Barbara Bogacka who took a genuine interest in my work and helped to clarify my research questions and designs.

I am utterly grateful to my fellow musicians who so enthusiastically participated in the experiments. It was my pleasure to work with ESMUC students and graduates: Carles, Josep-Oriol, Carlos, Laia and Marta. I truly enjoyed the recording sessions, beautiful music and post-session passionate discussions about timbre. I am also hugely indebted to my Polish musician friends who keenly devoted much of their private time to take part in the listening tests.

My special thanks go to WISE@QMUL and G.Hack team mates for all those inspirational seminars, hands-on workshops and artistic projects (not mentioning social events) which I had a great pleasure to be part of.

At last, my endless and most heartfelt thanks to my family and friends who supported me all the time and who always believed in me and were at hand, especially in those times when I was doubting myself.

This work was supported by a UK EPSRC DTA studentship EP/P505054/1 and the EPSRC funded OMRAS2 project EP/E017614/1.

Contents

List of Figures **11**

List of Tables **18**

List of Abbreviations **22**

1 Introduction **24**
1.1 Motivation and approach 25
1.1.1 Why the cello? 26
1.1.2 Perception 26
1.1.3 Tone acoustics and gesture controls 28
1.1.4 Relationship between gesture, tone quality and perception 29
1.2 Contributions 30
1.3 Collaborations and related publications 31
1.4 Thesis outline 32

2 Perceptual and acoustical aspects of timbre **35**
2.1 Introduction 35
2.2 On timbre definition 36
2.3 Perceptual studies of timbre 38
2.3.1 Timbre's interaction with pitch and dynamics 38
2.3.2 Transients' effect on timbre perception 39
2.3.3 The concept of timbre spaces 41
2.3.4 Acoustical correlates of timbre dimensions 42
2.4 Acoustic features in automatic instrument recognition 45
2.5 Single instrument timbre studies 47
2.6 Timbre as a means of musical expression 61
2.7 Methods for measuring perceptual attributes of timbre 63
2.7.1 Proximity rating 64
2.7.2 Multidimensional scaling 65
2.7.3 Semantic labelling: verbal attributes of timbre 66
2.8 Remarks and conclusions 70

3 The Cello: acoustic fundamentals and playing technique . . . 73
3.1 Introduction . . . 73
3.2 Acoustical properties of the cello . . . 75
3.2.1 Motion of the bowed string . . . 76
3.2.2 Resonances of the cello body . . . 78
3.2.3 The bow . . . 84
3.2.4 Cello sound spectra . . . 85
3.3 Playing technique . . . 88
3.3.1 Left hand technique . . . 88
3.3.2 Bowing technique: controlling tone quality . . . 90
3.4 Cello timbre in perception . . . 95
3.5 Summary . . . 98

4 Acquisition and analysis of bowing gestures . . . 99
4.1 Introduction . . . 99
4.2 Bowing machine studies . . . 100
4.3 Measuring bowing parameters in normal playing . . . 111
4.3.1 Systematic studies of the playable region . . . 111
4.3.2 Capturing bowing gestures for interactive performance, sound synthesis and bow stroke analysis . . . 117
4.4 Summary . . . 123

5 Experimental data collection for acoustical and gestural analysis . . . 124
5.1 Performers . . . 124
5.2 Data acquisition framework . . . 125
5.3 Studio recording setup . . . 129
5.4 Repertoire . . . 130
5.5 Recording session scenario . . . 132
5.6 Data processing . . . 134
5.7 Summary . . . 135

6 Perceptual evaluation of cello player tone via timbre dissimilarity and verbal attribute ratings . . . 136
6.1 Introduction . . . 136
6.2 Aim of the study and research goals . . . 139
6.3 Method . . . 140
6.3.1 Stimuli . . . 140

6.3.2 Participants 141
6.3.3 Procedure 142
6.4 Results and discussion 143
6.4.1 Effect of being a cellist or non-cellist on cello timbre perception 143
6.4.2 Inter-rater reliability analysis 144
6.4.3 Perceptual mapping of the players 146
6.4.4 Mapping the players into the verbal attribute space 150
6.5 Summary 159

7 Acoustical correlates of the perceptual dimensions 160
7.1 Introduction 160
7.1.1 *A priori* remarks 161
7.1.2 Research goals 162
7.2 Method 163
7.2.1 Acoustic feature extraction 163
7.2.2 ANOVA-based feature selection 164
7.3 Results and discussion 165
7.3.1 Factor analysis 165
7.3.2 Acoustical mapping of the players 174
7.3.3 Discriminating performers based on factor scores and acoustic features 186
7.3.4 Correlation between acoustical and perceptual dimensions 192
7.4 Summary 196

8 Identifying performer-specific bowing controls 198
8.1 Introduction 198
8.1.1 Research questions and *a priori* remarks 200
8.2 Method 201
8.2.1 Bowing data processing 202
8.2.2 Bowing data analysis 203
8.3 Results and discussion 203
8.3.1 Comparing general use of bowing controls across six musical contexts 203
8.3.2 Comparing the use of bowing controls amongst the players across six musical contexts 211
8.3.3 Correlation between bowing controls and acoustic features 215
8.4 Summary 220

9 Final notes and conclusions **222**
9.1 Final notes on the relation between gesture, tone quality and perception 222
9.2 Summary of contributions 228
9.3 Future work 234
9.4 Potential applications 238
9.5 A closing remark 238

Bibliography **240**

A Music Scores **265**
A.1 J.S. Bach – *3rd Cello Suite* 265
A.2 G. Fauré – *Élégie* 266
A.3 D. Shostakovich – *Cello Sonata* op. 40 267

B Acoustic Features **269**

C Experimental Data Examples **271**

List of Figures

1.1 The mechanical and physical processes behind a music performance. 26

2.1 Two-dimensional MDS solution for the mean dissimilarity ratings of eleven oboists. (From Fitzgerald, 2003) 51
2.2 Three-dimensional INDSCAL solution for the eleven oboists. (From Fitzgerald, 2003) 52
2.3 Dendrogram of HICLUS analysis of the dissimilarity ratings for the eleven oboists. (From Fitzgerald, 2003) 53
2.4 Three-factor configuration of the verbal attributes across the eleven oboists. (Adapted from Fitzgerald, 2003) 54
2.5 Three-dimensional clarinet timbre space and its mechanical and acoustical correlates. (From Barthet et al., 2010b) 57
2.6 Generlised PCA solution of mean VAME ratings across five pitches. For each pitch, relative positions of the four verbal attributes are indicated. (From Štěpánek, 2002) 59

3.1 Component parts of the cello in detail. (From Bynum and Rossing, 2010) 74
3.2 The cello bow: (a) the stick; (b) the tip; (c) the frog; (d) the screw; (e) the hair; (f) the lapping (wrap). (From Straeten, 1905) 75
3.3 The motion of a bowed string at successive times during the vibration cycle. (left) The bend races around an envelope; (right) the velocity of the string at different times in the vibration cycle. (From Rossing, 2010) 76
3.4 Displacement of bow and string at the point of contact. The points (a)-(h) correspond to the (a)-(h) steps shown in Figure 3.3. (From Rossing, 2010) 77
3.5 String velocity waveform at the bowing point. (Adapted from Woodhouse, 1997) 77

3.6 Waveform of time-varying transverse force exerted on the cello bridge by the open C string. The time period is approximately 15 ms. (From Richardson, 1999) 78
3.7 The spectrum of the ideal sawtooth waveform. (From Howard and Angus, 2009) 78
3.8 A cello response curve showing the input admittance (velocity amplitude per unit driving force) as a function of excitation frequency. Force was applied at the bridge in the bowing direction. The fundamental frequencies of the open strings are marked. (From Richardson, 1999) 80
3.9 Input admittance curves of high quality violin (top) and cello (bottom). (From Askenfelt, 1982, as cited by Rossing (2010)) . . 81
3.10 Average spectrum of a modern violin for 12 microphone positions spaced at 30° intervals around the instrument. (From Rossing, 2010) 86
3.11 Room-averaged sound spectra of a cello: (a) freely supported on rubber bands; (b) hand-held in playing position. (From Bynum and Rossing, 1997, as cited in Rossing (2010)) 86
3.12 Principal radiation directions of a cello at different frequencies: (left) in the vertical plane; (right) in the horizontal plane. (From Meyer, 2009) 87
3.13 Physical bowing parameters controlled by a violin player: bow velocity, bow position, bow force (the force pressing the bow against the string), and bow-bridge distance (as measured from bowing or contact point). (From Askenfelt, 1989) 91
3.14 Complementary bowing controls on the violin: bow tilt, bow inclination, and bow skewness. (From Schoonderwaldt and Demoucron, 2009) 91
3.15 Typical normal and abnormal playing conditions in the violin family related to bow force and bow position at constant bow velocity for sustained tones. A second set of coordinates refers to a cello A string bowed at 20 cm/sec. (From Schelleng, 1973) . . 93

3.16 Three-dimensional timbre space for 16 recorded instrument tones. Abbreviations for stimulus points: O1, O2 = oboes; C1, C2 = clarinets; X1, X2, X3 = saxophones; EH = English horn; FH = French horn; BN = bassoon; TP = trumpet; TM = trombone; FL = flute; S1, S2, S3 = cello tones. (Adapted from Gordon and Grey, 1978) 96
3.17 Two-dimensional timbre space for 16 complete tones, i.e. onset + remainder. (Adapted from Iverson and Krumhansl, 1993) 97

4.1 Three principal types of bowed string attacks. String velocity at the bowing point for prolonged (top), "perfect" (middle), and multiple flyback attacks (bottom). The violin open G string was played with a bowing machine using a normal bow. (From Guettler and Askenfelt, 1997) 104
4.2 Relation between bow acceleration, bow force and sound quality. (From Guettler, 2010) 105
4.3 Spectrum of string velocity for the three bow speeds. The spectrum (normalized to 0 dB for the 1st harmonic) was averaged over several strokes with constant bowing parameters. (From Guettler et al., 2003) 106
4.4 Two examples of anomalous low frequency (ALF) string-velocity waveforms with periods of about (a) twice and (b) three times the fundamental period T_1 (indicated by the vertical dashed lines). The bow velocity = 10 cm/s. A horizontal dashed line indicates nominal slip velocity v_S (Helmholtz motion). (From Schoonderwaldt, 2009b) 109
4.5 Anomalous low frequencies (ALF) in the Schelleng diagrams at bow velocities (a) 10 and (b) 15 cm/s. The numbers indicate the frequency in Hertz. Clusters of different types of ALF include: period doubling (around 150 Hz), period tripling (around 100 Hz), and pitch lowering by a semitone (around 270 Hz). The upper bow-force limits are indicated by solid lines. The nominal fundamental frequency of Helmholtz motion = 293 Hz (the open D string). (From Schoonderwaldt, 2009b) 110

4.6 Examples of bowing parameters measured in whole notes, half notes, and 16th notes performed by a violin player on the D string in *mf* dynamics. The parameters from the top: bow transverse position (x_B), bow velocity (v_B), bow force (F_B), and relative bow-bridge distance (β). (From Schoonderwaldt, 2009a) 114

5.1 The *Polhemus* system components: (a) the source of EMF (left), the sensor attached to the bow (upper right), the sensor marker (middle right), the sensor attached to the cello (bottom right); (b) and (c) the PC unit processing sensor data. 126
5.2 The *Polhemus* motion tracking components: (a) the sensor attached to the bow; (b) the sensor attached to the cello. 126
5.3 Equipment for bow force calibration: (a) bowing cylinder mounted on the *Transducer Techniques MDB-5* load cell, here with a 50 g precision weight for calibrating load values; (b) the *Transducer Techniques TMO-1* amplifier (right) and the *National Instruments USB-6009* A/D converter (left). 127
5.4 Screenshot of the VST plugin interface designed for controlling audio, bow motion and bow force data acquisition. 128
5.5 The recording studio layout. 129
5.6 Documenting the position of the microphone and the cellist. . . . 130
5.7 The pickup mounted on the bridge of Cello1 (a)–(b) and Cello2 (c)–(d). 131

6.1 2-D MDS maps of timbre dissimilarity ratings for each music excerpt. 148
6.2 2-D MDS map of timbre dissimilarity ratings averaged across six music excerpts. 149
6.3 2-D correspondence map for the 6x6 contingency table. The black dashed line drawn through the origin and Cellist 5 is used to determine which semantic labels were most associated with his timbre and to assess how often (relatively) each label appeared in his tone ratings (intersections with red dotted perpendiculars). The origin of the map represents the average or barycentre of both the Cellist and Timbre Attribute variables. 155

6.4 2-D correspondence map for the 36x6 contingency table. To preserve readability of the chart, only samples considered as less characteristic for each cellist's timbre were annotated with music excerpt labels. 157

7.1 Allemande. Scatter plots of mean factor scores (left) and mean acoustic features (right). 176
7.2 Bourrée. Scatter plots of mean factor scores (left) and mean acoustic features (right). 178
7.3 Courante. Scatter plots of mean factor scores (left) and mean acoustic features (right). 179
7.4 Élégie. Scatter plots of mean factor scores (left) and mean acoustic features (right). 181
7.5 Shost1. Scatter plots of mean factor scores (left) and mean acoustic features (right). 182
7.6 Shost4. Scatter plots of mean factor scores (left) and mean acoustic features (right). 184
7.7 All music styles combined. Three-factor solution. Scatter plots of mean factor scores (left) and mean acoustic features (right). . 185
7.8 All music styles combined. Two-factor solution. Scatter plots of mean factor scores (left) and mean acoustic features (right). . . . 186
7.9 The three-factor solution. Comparison of mean factor scores (a)–(c)–(e) and means of the highest loading acoustic features (b)–(d)–(f) across the cellists, ($N = 450$). 188
7.10 The two-factor solution. Comparison of mean factor scores (a)–(c) and means of the highest loading acoustic features (b)–(d) across the cellists, ($N = 450$). 192

8.1 Discriminant analysis on three bowing parameters across all six music pieces. Points represent discriminant scores on the 1st and 2nd discriminant functions for each note in the dataset grouped by Piece ($N = 426$). 207
8.2 Comparison of mean discriminant function scores (a)–(c)–(e) and the highest loading bowing controls (b)–(d)–(f) across the six pieces, ($N = 426$). 210

8.3 Discriminant analysis on three bowing parameters across all six music pieces. Points represent discriminant scores on the 1st and 3rd discriminant functions for each note in the dataset grouped by Cellist ($N = 426$). 212
8.4 Discriminant analysis on three bowing parameters across all six music pieces. Points represent discriminant scores on the 2nd and 3rd discriminant functions for each note in the dataset grouped by Cellist ($N = 426$). 213
8.5 Mean bowing parameters across the six cellists ($N = 426$): (a) bow-string distance (z_{bs}); (b) relative bow-bridge distance (β); (c) bow velocity (v_B) [cm/s]. 214
8.6 Shost1. The SubBand9Flux values plotted against relative bow-bridge distance and grouped by Cellist, with a least squares regression line marked ($N = 66$). 217
8.7 All music pieces combined. The SubBand10Flux values plotted against bow velocity and grouped by Cellist, with a least squares regression line marked ($N = 426$). 218
8.8 All music pieces combined. The SubBand10Flux values plotted against bow-string distance and grouped by Cellist, with a least squares regression line marked ($N = 426$). 219
8.9 All music pieces combined. The SubBand10Flux values plotted against relative bow-bridge distance and grouped by Cellist, with a least squares regression line marked ($N = 426$). 220

9.1 Three-dimensional acoustical space for the six cellists. Each point represents the acoustic features averaged across all music styles ($N = 426$). 223
9.2 Bowing control space for the six cellists. Each point represents z_{bs}, β and v_B parameters averaged across all music styles ($N = 426$). 224
9.3 The averaged SubBand10Flux values plotted against the second perceptual dimension coordinates of the six cellists. 225
9.4 Interrelations between the three performer domains. Bowing controls space mapped into (a) acoustical and (b) perceptual spaces; (c) mapping between acoustical and perceptual proximities. Each point represents the average distance between a pair of players across all music styles ($N = 15$). 227

C.1 Shost1. Comparison of bowing parameters extracted from the captured motion data of (a) Cellist 1 and (b) Cellist 2. The waveforms of the respective audio samples are shown in the background (in grey). Note the differences in the parameters' ranges between the two players. For Cellist 1, the means of bow velocity, bow-bridge distance and bow-string distance across notes were 18.27 cm/s, 10.66 cm and 0.71 respectively, compared to the corresponding values of 23.33 cm/s, 6.31 cm and 0.45 for Cellist 2. 272
C.2 Shost1. Comparison of spectrograms obtained from the audio samples of (a) Cellist 1 and (b) Cellist 2. The respective waveforms are shown in the upper plots. Instantaneous STFT power spectra were computed using 23.2-ms frames with 75% overlap and a 21.53 Hz frequency resolution. 273
C.3 Shost1. Comparison of long-term average spectra (LTAS) obtained from the audio samples of Cellist 1 (blue dashed line) and Cellist 2 (red line). Instantaneous STFT magnitude spectra were computed using 23.2-ms frames with 75% overlap and a 21.53 Hz frequency resolution. Note higher amplitudes of the frequencies from 550 Hz upwards combined with lower amplitudes of the frequency components around 300 Hz and below 100 Hz in the spectrum of Cellist 2. The corresponding value of spectral centroid averaged across notes was 1858.67 Hz compared to 1531.87 Hz for Cellist 1. 274

List of Tables

3.1 The cello normal modes and their frequencies compared to modal frequencies of a violin. Alternative labelling of the modes is given in parantheses. (Adapted from Rossing, 2010). 82

4.1 Overview of steady-state spectral effects when changing one bowing parameter at a time. (Adapted from Guettler, 2004, 2010) . . 108

6.1 Measures of internal consistency and agreement between the participants' ratings for each music fragment. 145
6.2 Goodness of fit measures for MDS solutions for six music excerpts. 146
6.3 Measures of inter-rater agreement in verbal attribute ratings across the six music excerpts. 151
6.4 Contingency table of timbre attribute votes across six music excerpts. The highest number of votes for each attribute is marked in bold. 152
6.5 Summary of two correspondence analyses. 153

7.1 Pitch and frequency range of the sound stimuli. 162
7.2 Feature subsets selected in ANOVA for pieces Allemande, Bourrée and Courante. 166
7.3 Feature subsets selected in ANOVA for pieces Élégie, Shost1. . . 167
7.4 Feature subset selected in ANOVA for Shost4. 168
7.5 Allemande. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 84.7% of total variance explained ($N = 78$). 169
7.6 Bourrée. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 81.3% of total variance explained ($N = 90$). 170
7.7 Courante. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 76.0% of total variance explained ($N = 120$). 171

7.8 Élégie. Factor analysis of audio features across all cellists. Factor loadings for two rotated solutions. 83.9% and 75.5% of total variance explained ($N = 36$). 172
7.9 Shost1. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 75.3% of total variance explained ($N = 66$). 173
7.10 Shost4. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 90.8% of total variance explained ($N = 60$). 173
7.11 The six music excerpts combined. Factor analysis of audio features across all cellists. Factor loadings for two rotated solutions. 81.2% and 74.8% of total variance explained ($N = 450$). 174
7.12 Allemande. Correlations between mean factor scores and between means of the highest loading features ($N = 6$). 175
7.13 Bourrée. Correlations between mean factor scores and between means of the highest loading features ($N = 6$). 177
7.14 Courante. Correlations between mean factor scores and between means of the highest loading features ($N = 6$). 177
7.15 Élégie. Correlations between mean factor scores and between means of the highest loading features ($N = 6$). 180
7.16 Shost1. Correlations between mean factor scores and between means of the highest loading features ($N = 6$). 183
7.17 Shost4. Correlations between mean factor scores and between means of the highest loading features ($N = 6$). 183
7.18 All music styles combined. Three-factor solution. Correlations between mean factor scores and between means of the highest loading features ($N = 6$). 184
7.19 Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for factor scores in the three-factor solution ($N = 450$). 187
7.20 Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for features in the three-factor solution ($N = 450$). 189
7.21 Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for factor scores in the two-factor solution ($N = 450$). 190

7.22 Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for features in the two-factor solution ($N = 450$). 190
7.23 Allemande. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 193
7.24 Bourrée. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 193
7.25 Courante. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 194
7.26 Élégie. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 194
7.27 Shost1. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 195
7.28 Shost4. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 195
7.29 All music styles combined. Three-factor solution. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 196
7.30 All music styles combined. Two-factor solution. Correlations between the perceptual and acoustical dimensions and features ($N = 6$). 196

8.1 Bowing control means and standard deviations grouped by music excerpt ($N = 426$). 204
8.2 Intercorrelations among the three bowing parameters. 205
8.3 Investigating differences in the use of bowing parameters across six musical contexts. Results of univariate ANOVAs for each bowing control ($N = 426$). 205
8.4 Results of discriminant analysis on three bowing parameters across all six music pieces ($N = 426$). 206
8.5 Discriminant function and correlation coefficients ($N = 426$). . . 208
8.6 Investigating differences in the use of bowing parameters between the cellists across six musical contexts. Results of univariate ANOVAs for each bowing control ($N = 426$). 215
8.7 Correlations between the three bowing parameters and perceptually linked acoustic features for each music excerpt and all excerpts combined. 216

9.1 Correlations between the perceptual dimensions and bowing controls for the six cellists. Bowing parameters averaged across all music styles ($N = 6$). 226
9.2 Correlations between the three performer domains based on calculated proximities between the cellists in gestural, acoustical and perceptual spaces ($N = 15$). 226

B.1 Frequency ranges of ten octave-scaled subbands. (From Alluri and Toiviainen, 2010) 269
B.2 Acoustic features and their definitions. Represented signal domains: T – temporal, S – spectral, ST – spectro-temporal. (Adapted from Eerola et al., 2012) 270

List of Abbreviations

ALF	Anomalous Low Frequency
ANOVA	Analysis of Variance
ART	Attack Rise Time
AT	Attack Time
CA	Correspondence Analysis
CLASCAL	Latent Class Weighted Euclidean Scaling
CONSCAL	Constrained Scaling
CV	Centroid Variability
DA	Discriminant Analysis
DI	Direct Input
DOF	Degrees Of Freedom
EMF	Electromagnetic Field
EXSCAL	Extended Two-Way Euclidean Scaling
FA	Factor Analysis
FM	Frequency Modulation
FSR	Force Sensing Resistor
HICLUS	Hierarchical Clustering
ICC	Intra-Class Correlation
IMU	Inertial Measurement Unit
INDSCAL	Individual Differences Scaling
IOI	Intertone Onset Interval
IRR	Inter-Rater Reliability
JND	Just Noticeable Difference
k-NN	k-Nearest Neighbours
KMO	Kaiser-Meyer-Olkin

LAT	Log Attack Time
LPCC	Linear Prediction Cepstral Coefficient
LTAC	Long-Time Average Centroid
LTAS	Long-Term Average Spectrum
MANOVA	Multivariate Analysis of Variance
MDS	Multidimensional Scaling
MFCC	Mel-Frequency Cepstral Coefficient
MIR	Music Information Retrieval
NN	Neural Network
OER	Odd/Even Harmonic Ratio
PAF	Principal Axis Factoring
PC	Personal Computer
PCA	Principal Component Analysis
PROXSCAL	Proximity Scaling
RF	Random Forests
RMS	Root Mean Square
SC	Spectral Centroid
SD	Spectral Deviation
SF	Spectral Flux
SS	Spectral Spread
STFT	Short-Time Fourier Transform
SV	Spectral Variation
SVD	Spontaneous Verbal Description
SVR	Support Vector Regression
VAME	Verbal Attribute Magnitude Estimation
VARR	Verbal Attribute Ranking and Rating
VST	Virtual Studio Technology
WLPCC	Warped Linear Prediction Cepstral Coefficient

Chapter 1

Introduction

The extent to which a classical performer can influence the resulting timbre of an instrument has been rarely considered in academic research, whether music acoustics, performance studies or Music Information Retrieval (MIR) applications. In contrast, the so-called "sound of a player" is a well-known phenomenon amongst musicians, related to that unique quality of tone which can be universally heard across different performances and which forms one of the most distinctive features of someone's musicianship.

This unique tone quality, however, has received little attention compared to other aspects of an expressive performance such as dynamics, articulation, tempo and timing, the individual variations of which have been employed to distinguish music performers (Widmer et al., 2003; Dillon, 2004; Stamatatos and Widmer, 2005; Tobudic and Widmer, 2005; Molina-Solana et al., 2008). For example, in a series of experiments Widmer et al. (Dixon et al., 2002; Zanon and Widmer, 2003; Saunders et al., 2004; Widmer and Zanon, 2004) used the extracted global tempo-loudness trajectories to analyse performances of six famous pianists and recognise the artists from their playing styles. Ramirez et al., on the other hand, employed sets of note-level descriptors to capture and then classify expressive trends of three jazz saxophonists (2007) and two violinists (2008). The selected descriptors included intra-note perceptual features based on energy envelope and spectral centroid (which can be considered as an attempt to capture timbral characteristics of the players), as well as inter-note contextual features related to pitch, duration and the note's position within the

melodic structure.

The unique tone quality, while being an integral means of musical expression, is embedded in the physical process of sound production and as such can be measured via acoustical properties of sound. The fact that a player's timbre depends heavily on their individual physical and perceptual abilities makes it automatically a good candidate for a discriminator, a kind of "timbral fingerprint" by analogy. It may act as a lower-level characteristic of a player, independent of the other expressive attributes. If individual timbre features can characterise a performer then timbre dissimilarities can be used for performer discrimination.

1.1 Motivation and approach

When an accomplished musician interacts with an instrument to produce sound he uses a range of technical skills developed through years of practising and mastering performance. Applying physical actions (instrumental gestures) to the instrument, he modulates its timbre creating a variety of sound "colours" that embody his musical intentions. All his efforts serve to convey the contents of a music score to a listener. The process of communication between a music performer and a listener involves then the transformation of the mechanical (gestural) input into the acoustical output of the instrument which then becomes the perceptual input for the cognitive process of music listening. A diagrammatic illustration of the transition between the performance domains is shown in Figure 1.1.

Since the listener is the main intended recipient of musical communication, his impressions and the overall experience of performance are what matters to a musician. In fact, it is the listener who is the ultimate censor of someone's musicianship and musical craft. Therefore, in order to understand what a player's timbre really is, this thesis starts the exploration of the phenomenon from the listener's perspective, looking for perceptual cues of timbral differences between musicians and searching further for the origins of these differences in acoustical (tone spectral characteristics) and gestural (performance controls) domains.

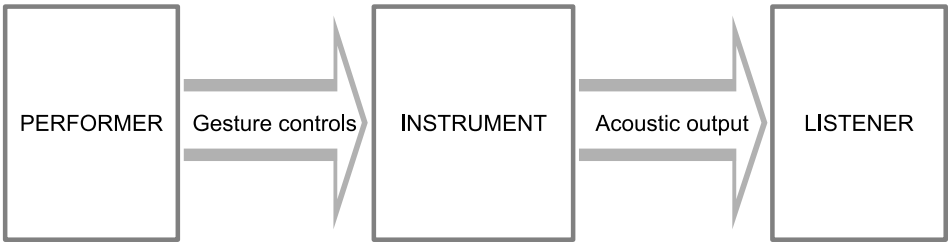


Figure 1.1: The mechanical and physical processes behind a music performance.

The following section explains why this research focuses on cello timbre, and Sections 1.1.2 to 1.1.4 describe the four main research questions (RQ1–RQ4) addressed in this thesis.

1.1.1 Why the cello?

The choice of the cello and its timbre as a case study stems from the fact of the author's being a professional cellist with years-long performing and teaching experience. The acute attention to tone quality and continuous search for richer timbral palette as a performer's expressive means have always been driving factors of her musicianship's growth. Detailed understanding of the mechanical and psychophysical processes behind the tone production and perception on bowed string instruments, derived from the author's musical expertise, has been beneficial for designing the experimental work conducted here.

Moreover, the advantage of studying timbres of bowed string and wind instruments (and of the cello in particular) rather than other acoustic instruments is that their sound production process relies on a continuous source of excitation (bowed string or blown air column) and as such provides a player with full control over the tone quality at any time point of the process.

1.1.2 Perception

RQ1: Can classical musicians be discriminated perceptually by timbre, i.e. by their tone quality?

The ability to discriminate and identify everyday sounds is deeply embedded in the human auditory system and is fundamental for our understanding and navigating throughout the surrounding environment. Humans use timbre to recognise sound sources, whether environmental (e.g. car horns, chirping birds, people's voices) or musical (e.g. acoustic instruments, singing voices). It is worth noticing that, while learning environmental sounds is a necessary part of the life adaptation process, becoming familiar with the sounds of orchestral instruments for example, is rather the matter of an individual being exposed to musical culture and/or education, if such is available. The influence of musical training was addressed in a number of perceptual studies which involved timbre dissimilarity ratings by both musically trained and untrained subjects (e.g. Marozeau et al., 2003; Handel and Erickson, 2004). For musicians, it is believed, the ability to recognise musical instruments extends to also differentiating between timbres of similar instruments (e.g. two violins), for no formal study has been attempted to examine the effect of musical training on perceiving the differences in sound quality between instruments of the same class. Beyond any doubt is the musicians' ability to tell the difference in sound properties of two instruments once they are given the opportunity to try them. When the task involves distinguishing between two players performing on the same instrument a higher level of musical expertise seems a necessary requirement.

The unique tone is an integral part of the individual playing style and, in a broader sense, of a performer's musical identity. However, can listeners perceive the tone quality itself as a discriminative feature of the player, regardless of other expressive attributes? One obvious way to test it is to present a group of expert listeners (for the task might be challenging for musically untrained subjects) with recordings of the same music fragments played by different performers on exactly the same instrument. Of a crucial importance is the length and contents of the tone samples presented to the subjects, whether they should be single notes or short musical motifs or phrases. Stimuli based on single notes provide a researcher with a material easy to control and manipulate which, however,

is not likely to capture the key characteristics of a player's timbral palette. Longer music fragments, on the other hand, though certainly constituting more representative tone samples, may induce the listeners to unintentionally focus on other music interpretation aspects such as phrasing and articulation rather than on the tone quality itself. Short musical motifs seem in this case best suited for the task as they offer a compromise between the scarcity and the excess of perceptual cues in single note stimuli and musical phrases respectively.

1.1.3 Tone acoustics and gesture controls

On a bowed string instrument such as the cello, the left-hand technique, responsible for vibrato and pitch changing by controlling the length of the string being played, and the bowing technique that controls the interaction of the bow hair and the string, form the playing apparatus. In general, the bowing controls are believed to contribute the most to the tone quality. Bowing parameters such as bow velocity, bow force and bow-bridge distance as well as bow position, bow tilt and bow inclination are constantly controlled by a player to obtain the desired timbre properties of musical sounds.

When comparing tone samples of various cello performers, the recorded audio signal does not provide direct access to performance parameters such as increasing values of the bow force during *forte* sections or the decreasing bow velocity when the final note is slowed down, for example[1]. Since these actions have immediate impact on the temporal evolution of the instrument's timbre, they are reflected in temporal, spectral and spectro-temporal features of the recorded sound. Therefore, this work explores acoustic features that can be interpreted in terms of physical actions of a player, and that can also capture the unique properties of the player's tone in order to answer the following question: *RQ2: Can classical musicians be discriminated by their acoustic parameters, i.e. acoustic characteristics of their tones?*

[1]However, first attempts at indirectly acquiring gesture controls from recorded audio signals have been made, see e.g. Traube (2004); Pérez and Wanderley (2015).

If listeners can differentiate between timbres of different performers based on recorded tone samples, and if the subsequent acoustical analysis reveals significant differences in their respective spectro-temporal characteristics, then, as the result of the existing physico-mechanical dependencies indicated in Figure 1.1, the exerted instrument control gestures must also be different for each performer.

In fact, the whole process of developing playing technique is strongly influenced by the performer's physique and requires long term training to achieve the mastery level. Adopting the principles of a particular playing school or being particularly influenced by some teachers are also important factors in shaping the musician's technical skills and individual preferences over the tone quality. If instrumental gestures are specifically adapted to suit the player's physique and preference for particular tone quality, then:

RQ3: Can classical musicians be discriminated by their mechanical parameters, i.e. gesture controls used?

1.1.4 Relationship between gesture, tone quality and perception

If playing technique itself can act as a discriminative feature of a performer, then the following question arises:

RQ4: To what extent do individual gesture controls determine the resulting tone quality of a player and subsequently affect the way his tone is perceived by a listener?

To explore the relationship between gestural, acoustical and perceptual domains of a player's timbre and consequently to answer the above question requires application of quantitative measures which are able to indicate both the presence and the strength of any examined relationships.

1.2 Contributions

This thesis is an exploratory investigation into the psychoacoustic phenomenon of timbral uniqueness characterising the tone of every classical musician, which has its roots in the individual playing technique. The work presented here contributes to the mainstream of research on musical timbre in the following ways:

- providing empirical evidence of tone quality being a distinctive feature of a classical performer, as perceived by a listener and exhibited by respective spectro-temporal characteristics and gesture control patterns
- identifying acoustic descriptors capable of capturing differences in tone quality of players recorded on the same instrument, which can be related to qualitative properties such as *brightness* and *roughness* of the tone, as well as to specific combinations of bowing controls exerted by the players
- providing a methodology for investigating player-dependent differences in timbre within a single instrument class, which can be applied to any continuously excited acoustic instrument (e.g. strings, winds), subject to changing measured performance controls being instrument-specific. The proposed methodology can also be indirectly applied to impulsively excited instruments (e.g. guitar, piano) provided that the extraction of both performance controls and acoustic features is adapted for the short-lived transients of the sound produced.
- extending our understanding of player-dependent aspects of sound production on acoustic instruments and their implications for future research on timbre in general and psychoacoustics of musical instruments in particular
- undertaking an interdisciplinary approach to exploration of a real-world psychoacoustic and psychophysical phenomenon by combining knowledge

and methods across disciplines such as timbre perception, signal processing, acoustics of bowed string instruments, string playing techniques, performance studies, and acquisition and analysis of bowing gestures

- creating a multi-modal database of solo cello recordings that contains timbrally diverse musical material in terms of instrument, musical context, articulation, dynamics and vibrato

1.3 Collaborations and related publications

The multi-modal database described in Chapter 5 was created in collaboration with the Music Technology Group based at Universitat Pompeu Fabra (Barcelona) during the author's research stay in May-July 2011. The recording sessions were carried out under the supervision of Alfonso Pérez-Carrillo and with the invaluable support of PhD students Marco Marchini and Panagiotis Papiotis. The author was also provided with dedicated software tools for extraction of bowing parameters from the recorded motion tracking data.

The perceptual experiment described in Chapter 6 was designed in collaboration with Andrew Simpson and Asterios Zacharakis, PhD students at Queen Mary University of London.

The publications listed below report on experiments conducted as precursors for the development of this thesis. The author was the main contributor under the supervision of Simon Dixon (the principal supervisor), and Alfonso Pérez-Carrillo (Chudy et al., 2013).

Peer-Reviewed Book Chapter

Chudy, M. and Dixon, S. (2013). Recognising Cello Performers Using Timbre Models. In Lausen, B., Van den Poel, D., and Ultsch, A., editors, *Algorithms from and for Nature and Life*, Studies in Classification, Data Analysis, and Knowledge Organization, pp. 511-518, Springer International Publishing.

Peer-Reviewed Conference Papers

Chudy, M., Pérez, A., and Dixon, S. (2013). On the relation between gesture, tone production and perception in classical cello performance. In *Proceedings of the 21st International Congress on Acoustics*, Montréal, Canada.

Chudy, M. and Dixon, S. (2010). Towards music performer recognition using timbre features. In *Proceedings of the 3rd International Conference of Students of Systematic Musicology*, pp. 45–50, Cambridge, UK.

Technical Report

Chudy, M. and Dixon, S. (2012). Recognising Cello Performers Using Timbre Models. Research Report EECSRR-12-01, Queen Mary University of London, UK. `http://www.eecs.qmul.ac.uk/~magdalenac/papers/Chudy_and_Dixon_EECSRR-12-01.pdf`

1.4 Thesis outline

The research work collated in this thesis can be divided into two parts, namely background and experimental. The background Chapters 2 to 4 provide an overview of literature relevant to an interdisciplinary investigation on performer-related facets of musical timbre spanning research areas such as timbre analysis and perception, psychoacoustics, cello acoustics, performance studies, bowing control acquisition and analysis. The experimental Chapters 5 to 9 report on the acquisition of the cello database, followed by three studies carried out on perceptual, acoustical and gestural data respectively, and concluding with general discussion and a summary. In particular:

Chapter 2 reviews a corpus of literature related to wide-ranging research on perceptual and acoustical aspects of musical timbre including topics such as the definition of timbre and its attributes, an introduction to timbre spaces and their acoustical correlates, timbre descriptors and their applications, the use of verbal attributes for timbre dissimilarity description, and respective methodologies.

Chapter 3 outlines acoustic principles of sound production on the cello in relation to the instrument's structural components and discusses the elements of playing technique responsible for control over tone quality.

Chapter 4 examines prior works on mechanics of bowing, tone production and playability of bowed string instruments. It reports on major findings made with the use of bowing machines and further evaluated in normal playing conditions (mainly on violin) which involved dedicated equipment for bowing motion tracking, and concludes with examples of bowing gesture capturing devices employed for interactive performances, sound synthesis and bowing technique analysis.

Chapter 5 describes details of the design and acquisition of multi-modal cello recordings which include bowing motion tracking data in addition to two audio streams captured from a bridge pick-up and ambient microphone.

Chapter 6 presents a perceptual experiment on tone samples of six cellists across six different musical contexts. Multidimensional scaling of timbre dissimilarity ratings and verbal attribute ratings combined with correspondence analysis are employed to obtain perceptual mappings of the players. Dissimilarity patterns in association with semantic labels are discussed.

Chapter 7 provides details of a series of acoustical analyses carried out on the same set of tone samples as in Chapter 6. An ANOVA-based approach to feature selection is applied to the initial set of 25 acoustic descriptors extracted at the note level from the audio signals. Factor analysis is used to obtain low-dimensional acoustic representations of each cellist. The results of MANOVA tests designed to discriminate between those acoustic representations are reported. Correlation analysis reveals three spectro-temporal features linked to perceptual differences in tone quality amongst the cellists.

Chapter 8 presents analyses of bowing controls extracted from the bowing motion data accompanying the recorded tone samples of the six cellists (same as in Chapters 6 and 7). By means of MANOVA and discriminant analysis, general use of bowing parameters across different music excerpts is studied.

Another MANOVA design is used to identify individual bowing patterns among the players. Relations between each bowing control and acoustic features most correlated with perceptual dimensions are examined.

Chapter 9 further discusses the links between perceptual, acoustical and gestural aspects of a player's timbre, and concludes the thesis with a summary of the findings and further directions for future work and potential applications.

Chapter 2

Perceptual and acoustical aspects of timbre

2.1 Introduction

Timbre is a fundamental element of music, although its role, whether in music structure, musical expression or music performance, seems not yet fully acknowledged, at least in the historical styles of Western music. Until the arrival of impressionist "colours" in works of Debussy and Ravel, and later on, of *Klangfarbenmelodie* (Schoenberg, Webern) where the role of timbre is finally elevated to the level of becoming an explicit means of music structure, "the chief function of timbre (...) has been that of carrier of melodic functions" and "the differences of timbre at different pitches and in different registers of instruments (...) have been treated as nuances" (Erickson, 1975, p. 12).

However, starting right from the very beginning of the music creation process, when a composer chooses for his piece a certain set of musical instruments (or sound sources in modern composition), he intentionally defines a space of timbres in which, according to his imagination and artistic vision, the subsequent music narration should unfold, and in which his artistic concept may emerge in its finest form. Further, within the piece, varying dynamics, tempo and articulation, which are means of music language for changing the character or mood of music, can first of all be perceived as variations in timbre, vivid examples of which are those moments when the sound intensity drops down to *piano* while timbre becomes soft or muffled, or when a phrase played *legato*

with warm, full-bodied sound changes to "crisp" and brilliant *staccato* motives. Finally, whenever a performer or a group of performers interprets the piece, they add the whole new timbral realisation to what is very often just implicit intentions of the composer.

This illustrates that, in fact, timbre functions at the core of music, being physically dependent on a sound source whether it is a traditional musical instrument or electroacoustic device. This also explicates the strong interest generations of researchers have taken at exploring and uncovering timbre's "elusive" nature (Schouten, 1968).

This chapter outlines some of the most salient findings about perceptual and acoustical aspects of timbre, which helped to broaden our understanding of the phenomenon. Starting with a brief description of timbre definition and related issues, the following sections give an overview of literature on timbre perception and methodological approaches to measuring perceptual attributes of timbre. Emphasis is placed on the experimental results and methods which may provide cues for performer-related investigation of cello timbre carried out in this thesis.

2.2 On timbre definition

Timbre, as a complex quality of sound has been studied thoroughly for decades. Its complexity is reflected in the fact that until now no precise definition of the phenomenon has been formulated, leaving space for numerous attempts at an exhaustive and comprehensive description.

The working definition provided by ANSI (1960, p. 45) describes timbre as an attribute of auditory sensation which enables distinguishing between two sounds having the same loudness, pitch and duration, that for example are played on two different musical instruments. However, the notion of timbre, is far more capacious than this simple distinction. Called in psychoacoustics *tone quality* or *tone color* (see Erickson, 1975, for some interesting remarks on that

matter), timbre not only categorises the source of sound (e.g. musical instruments, human voices) but also captures the unique sound identity of instruments or voices belonging to the same family (when comparing two violins or two dramatic sopranos for example). Interestingly enough, tones produced on just one instrument seem to possess their own timbres (Miller, 1909; Schaeffer, 1966).

Furthermore, when listening to tone samples or musical phrases by two performers who happened to play them on the same instrument, one can hear unique timbral features which distinguish one player from another, though they both operate within the timbral identity of the one instrument.
One could ask then, what really is timbre?

In his pioneer studies on musical timbre, Helmholtz (1877) already recognised that "the quality of the musical portion of a compound tone depends solely on the number and relative strength of its partial simple tones, and in no respect on their differences in phase", focusing primarily on spectral rather than on temporal aspects of musical tones. While adapting and expanding Helmholtz's theory, his followers, well into the 1960s, seemed constantly failing to acknowledge that temporal changes of spectral components are vital for tone quality (Risset, 1978) and that the transient parts of a sound can provide important clues for timbre identification (Young, 1960). A note added to the ANSI (1960) definition states that "timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus", which formally recognised the dynamic nature of timbre and its evolution over time.

Schouten (1968) proposed five major acoustic parameters that, in his opinion, can be sufficient to determine "the elusive attributes of timbre": its character ranging from "tonal" to "noiselike"; the spectral envelope; the time envelope in terms of rise, duration, and decay; the fluctuations of both spectral envelope (formant glide) and fundamental frequency (micro-intonation); and the onset of a sound differing notably from the steady state vibration. Erickson (1975, p. 6) found these "music-oriented concepts" suitable "for thinking about timbre,

whether noises, pitches, vocal sounds, traditional instrument sounds, electronic, or any other sounds". The proposed five dimensions "are fundamental to any discussion of timbre" (*ibid.*) and they formed a basis for a variety of acoustic descriptors developed mainly through perceptual studies of timbre.

2.3 Perceptual studies of timbre

2.3.1 Timbre's interaction with pitch and dynamics

As a generally adopted methodology for investigating psychoacoustic aspects of timbre, timbre analyses were conducted on single, isolated tones equalised in pitch, loudness and duration in order to give researchers, at least hypothetically, a full control over the experimental variable of timbre. However, it remained uncertain to what extent (if at all) the perception of timbre is invariant in presence of pitch, loudness or duration fluctuations. This inspired further investigations into different aspects of the pitch-timbre interaction (see Plomp and Steeneken, 1971; Krumhansl and Iverson, 1992; Handel and Erickson, 2001, 2004; Marozeau et al., 2003, for example).

A general conclusion was that timbre dissimilarities between musical instrument tones are perceived independently from differences in pitch for pitches varying within an octave and that this ability declines rapidly for notes more than one octave apart (Handel and Erickson, 2004). Steele and Williams (2006) replicated Handel and Erickson's study with some methodological refinements which included hiring musicians as well as non-musicians for the perceptual tasks. They showed that, although both groups exhibited a decline in accuracy of similarity ratings when the octave separation was increased, yet musicians were able to maintain above an 80% accuracy rate for tones at up to 2.5 octave difference in pitch. The result indicated that musical training is an important factor to consider when investigating *timbre invariance* across groups of listeners, however both Handel and Erickson (2004) and Steele and Williams (2006) agreed with Pitt (1994)'s conclusion on musicians' higher capability to separate

pitch and timbre changes. Interestingly, Marozeau et al. (2003); Marozeau and de Cheveigné (2007), who used musically trained and non-trained subjects, did not report any significant differences in dissimilarity ratings between the two groups (see Handel and Erickson, 2004, for general discussion).

No formal study has been undertaken to examine the salience of tone duration or the effect of change in dynamic level with respect to timbre perception of musical tones. Hajda et al. (1997) suggest that this is partly due to lack of "an empirical model which can predict a priori from acoustical information the perceptual loudness of complex time-variant tones" and that with current *a priori* methods only approximately "equal loudness" can be obtained. They conclude that, although "the perceptual loudness of complex time-variant tones varies with listener", with high correlations in timbral similarity judgements between subjects, "it is possible that minute differences in loudness do not significantly confound with timbre in the case of perceptual scaling".

2.3.2 Transients' effect on timbre perception

Once acknowledged, the influence of temporal cues on perception of timbre was also studied in detail. In a standard approach, the amplitude envelope of isolated tones was segmented into the onset, steady state and decay parts and the effect of each segment on either the identification (Clark et al., 1963; Berger, 1964; Saldanha and Corso, 1964; Wedin and Goude, 1972; Elliott, 1975) or similarity judgements (Iverson and Krumhansl, 1993) of musical instruments was investigated. The results showed that onsets seem vital for instrument recognition and, in most of cases, tones with only the attack part demonstrated similar identification accuracy to entire tones. Interestingly, however, onsets may not have the same salience for similarity judgements as Iverson and Krumhansl's study suggested. The ratings of "remainders" (tones with the onsets removed) were highly correlated with the ratings of complete tones and of the onset portions, indicating that the attributes salient for similarity judgements seem to be present throughout tones and may be different from acoustical attributes based

on which identification judgements are made (Iverson and Krumhansl, 1993).

Hajda et al. (1997) expressed great concern about Iverson and Krumhansl's results and conclusions. Firstly, they pointed out that the unbalanced choice of sound stimuli consisting of 13 continuant and 3 impulse instruments could bias the MDS solutions. Secondly, they referred to the definition of onset transient, set as the first 80 ms of each tone regardless instrument type, while amongst instruments chosen for the experiment such long attack can be observed only in flute, cello and violin, thus the distinction between the onset and steady state parts was not adequate for most of the stimuli (see Hajda et al., 1997, pp. 251–253 for further discussion).

In contrast to commonly employed isolated tones, Kendall (1986) examined the effect of different temporal segments of tones in instrument categorisation tasks using whole phrase versus single note contexts. His argumentation for the inclusion of psychomusical rather than psychoacoustic methodology stemmed from the facts that the latter "disregards the role of the listener; uses stimuli that are not normally apprehended in the normative musical contexts of a given culture; and disregards the role of the performer". For the purpose of the experiment, he defined the concept of instrument categorisation "as the ability of a listener, upon hearing the performance of one musical phrase, to match that phrase, with predictability beyond chance, with a different musical phrase performed by a different performer on a different instrument of the same class", assuming that the listener's ability to determine instrument class remains preserved across the variability due to different performers, instruments and instrument/performer interactions.

In his experiments, three musical phrases played *legato* were recorded on clarinet, violin and trumpet by two different performers on two different instruments per instrument class. Six temporal partitions of the recorded signals included normal, time-variant steady-state alone (with gaps and with elision), transients alone (with gaps), and static steady state with and without transients. A matching procedure described above was applied to collect answers

from musician and non-musician groups of subjects. In general, the mean response accuracy was significantly higher for whole phrases than for single notes. Based on the whole-phrase context results, Kendall concluded that "transients were neither sufficient nor necessary for the categorization of" the three instruments. The single-note context results, on the other hand, indicated that "transients were sufficient, but not necessary". However, for the single note part of the study, it is unclear how the single note stimuli were generated and further presented to the subjects, thus making comparisons with other isolated notes studies rather impracticable.

2.3.3 The concept of timbre spaces

Timbre is undoubtedly a multidimensional phenomenon (Plomp, 1970; Erickson, 1975). It can also be seen as a multidimensional realisation of a sound and can be graphically represented by a multidimensional "timbre space", where each sound is described by its spectral, temporal or spectro-temporal characteristics, and where its coordinates would correspond to perceptually intelligible sound attributes.

The concept of a timbre space was first applied by Plomp (1970) and further exploited, for example, in works of Wedin and Goude (1972); Miller and Carterette (1975); Grey (1977); Wessel (1979); Kendall and Carterette (1991); Iverson and Krumhansl (1993); McAdams et al. (1995); Lakatos (2000), who used either multidimensional scaling techniques (MDS) or factor analysis (FA) to process perceptual data. On the basis of dissimilarity judgements of sound stimuli, synthetic tones or tones of orchestral instruments (either natural or resynthesised) were mapped into two- or three-dimensional timbre spaces reflecting the perceptual distances between them. The next step consisted of correlating perceptual coordinates of each tone with its extracted acoustic parameters to interpret in physical terms its perceptual positioning.

In fact, the advances of multidimensional analysis provided researchers with

powerful tools for exploring the timbral relationships between stimuli (Donnadieu, 2007; McAdams, 2013) and subsequently enabled building adequate models of timbre comprehensive enough to cover a variety of musical instruments and instrument classes as well as to differentiate between possible timbral variants of one instrument in particular.

2.3.4 Acoustical correlates of timbre dimensions

In search of a model describing different instrument sounds, a number of MDS-based studies revealed continuous perceptual dimensions correlated with acoustic parameters, related to spectral, temporal and spectro-temporal properties of the sounds.

Amongst the first who used the MDS technique for perceptual representation of timbre was Grey (1977), Grey and Gordon (1978) who found timbre space dimensions correlated with the spectral centroid (spectral), spectral flux/attack synchronicity (spectro-temporal) and attack centroid (temporal) descriptors. They analysed 16 tones from 12 different instruments (3 cello samples represented the string family).

Iverson and Krumhansl (1993) diversified sound stimulus sets for testing the whole signals, onsets and sustained portions separately. For the stimuli consisting of 16 tones (15 instrument classes including violin and cello), they obtained a two-dimensional space spanned between spectral centroid and amplitude envelope.

Instead of natural sounds, Krumhansl (1989) and McAdams et al. (1995) used FM-synthesised simulations of instrument tones plus their hybrids, comparing sets of 21 and 18 timbres respectively (in both studies the string family was limited to a "bowed string" sample). Their experiments confirmed the correlation existing between the first dimension and the attack time descriptor and between the second dimension and spectral centroid, but they differed in interpretation of the third dimension. Krumhansl found it closely related to spectral flux, quantified later by Krimphoff et al. (1994) as spectral deviation,

while McAdams et al. also assigned the third coordinate with spectral flux but his descriptor did not correlate with the same higher specificities.

An exhaustive study of eleven natural continuant orchestral tones (10 instrument classes including violin) compared with their synthetic counterparts (three variants) was conducted by Kendall et al. (1999). They found only weak correlation between the rise times and MDS dimensions (as their stimulus set did not include impulse instruments) and concluded that for non-percussive signals time envelope characteristics are not primary in their perceptual differentiation. The obtained perceptual spaces correlated highly with spectral centroid (1st dimension) and spectral flux in terms of the mean coefficient of variation (2nd dimension). An alternative third dimension most often separated natural timbres from their synthetic variants.

Lakatos (2000) divided tones of natural orchestral instruments into continuant (winds and strings), impulsive (percussion) and all instruments combined stimuli sets (a total of 35 timbres represented 31 instrument classes including violin). Surprisingly, for all three timbre spaces derived from MDS analyses of similarity ratings, the acoustical correlates of Dimension 1 and 2 were identical, namely attack time and spectral centroid respectively.

A more recent study by Caclin et al. (2005), who employed purely synthetic sounds in order to fully control tones' acoustic properties, confirmed that attack time and spectral centroid are salient timbre parameters and they effectively explain the timbre space's first two dimensions. As for the third dimension, the results showed that spectral flux did not contribute as expected to differentiation between stimuli along this dimension. Instead, the authors proposed to interpret the variations in terms of spectral irregularity or spectrum fine structure.

McAdams et al. (2006) reviewed ten published timbre spaces from Grey (1977); Grey and Gordon (1978); Krumhansl (1989); Iverson and Krumhansl (1993); McAdams et al. (1995); Lakatos (2000) (all outlined above) by applying the same MDS technique (CLASCAL) to all data sets and extracting the same

set of acoustic features from all sounds (128 tones in total). Seventy two descriptors representing a wide range of temporal, spectral, and spectro-temporal properties of the acoustic signals were extracted from each tone. With the goal to identify the subset of acoustic descriptors that would most generalise prediction of timbral relations, they conducted correlation and cluster analyses which revealed four major descriptors: spectral centroid, spectral spread, spectral deviation, and temporal envelope (in terms of effective duration/attack time).

An interesting comparative analysis was conducted by Giordano and McAdams (2010) on 23 datasets from 17 published identification and dissimilarity rating studies. The aim was to quantify the extent to which mechanical properties of the sound source are associated with perceptual structures revealed in these studies, in other words, to what extent differences in the sound production mechanisms between instruments are reflected in the distances between sound stimuli within timbre spaces. Two mechanical properties were taken into account: the musical instrument family and excitation type. The results showed that in the identification tasks tones of instruments within the same family were significantly more often confused than were instruments from different families. These findings were consistent with cross-evaluation of dissimilarity ratings. Across the majority of the analysed datasets, tones generated by the same type of excitation or by instruments of the same family consistently clustered together and occupied the same region of the MDS space. Thus, dissimilarities in the mechanics of the sound source were associated with decreased identification confusions. In the discussion, the authors pointed out that, although the listeners' ability to differentiate between varying systems of sound production was positively validated, this ability was quantified independently of the acoustical correlates. Therefore, it remains "unclear what acoustical information listeners use to distinguish between families of musical instruments" (Giordano and McAdams, 2010).

Based on the findings from the perceptual studies, the standardised definitions of timbre descriptors were incorporated into MPEG-7 as part of the audio

data representation framework (ISO/IEC-15938-4, 2002). In addition to the basic spectral descriptors such as spectrum envelope, spectrum centroid, spectrum spread and spectrum flatness and basic signal parameters such as harmonicity and fundamental frequency, two timbral categories were formulated, namely Timbral Temporal descriptors which include log attack time and temporal centroid, and Timbral Spectral descriptors comprising harmonic spectral centroid, harmonic spectral deviation, harmonic spectral spread, harmonic spectral variation and spectral centroid (for a comprehensive review of the MPEG-7 audio standard including descriptors' definitions and applications refer to Kim et al., 2005).

Established definitions of timbre related descriptors have been also fully implemented in the Matlab environment, in a form of practical toolboxes released for the wider research community. Depending on application, various sets of temporal, spectral or spectro-temporal parameters can be now easily computed using, for example, MIRtoolbox (Lartillot et al., 2008) or Timbre Toolbox (Peeters et al., 2011) which both provide a relatively simple command line interface and a wealth of options for manipulation of the parameters' settings.

2.4 Acoustic features in automatic instrument recognition

Automatic recognition and classification of instrument sounds has become an important research topic in the Music Information Retrieval (MIR) domain, having direct applications in automatic music transcription, audio content segmentation and content-based searching.

The primary variables in instrument recognition strategies are the chosen set of features and relevant method of classification (an extensive review can be found in Herrera-Boyer et al. (2003)). Perceptual approaches require searching for acoustic features which offer the best explanation of perceptual dissimilarities

(as discussed in Section 2.3.4), while taxonomic approaches, labelling sounds according to a previously established taxonomy, concentrate on features which enable discrimination between instrument categories.

Numerous works address the task of instrument classification by exploring different variants of features in every possible combination. For example, Kostek (1995); Kostek and Wieczorkowska (1996) employed spectral characteristics derived from steady-state parts of sounds, such as MFCCs, spectral moments, formant frequencies, normalized frequency components, tristimuli, brightness, even and odd harmonic content, as well as a set of temporal characteristics extracted from the attack transients.

Jensen (1999) introduced a complete multi-level model of isolated instrument sounds. For instrument timbre modelling, he used the amplitude envelope and its attributes: the attack and release times, the relative amplitudes of the partials at the start of the release, and the attack curve form; the spectral envelope and its attributes: tristimuli, brightness, odd harmonic content, and irregularity. Additional features included shimmer (noise component, defined as the random fluctuation of the amplitude) and its attributes, jitter (another noise component defined as the random fluctuation of the fundamental frequency) and its attributes, and inharmonicity.

Eronen and Klapuri (2000) reported improved discrimination accuracy (compared to the results obtained by Martin and Kim (1998); Martin (1999) on the same dataset) using a combined set of 43 spectral and temporal features. The feature list included linear prediction cepstral coefficients (LPCCs) computed from both the onset and the remainder of the tone, rise and decay times, spectral centroid and its statistical moments, and fundamental frequency related parameters. In later work, Eronen (2001) showed that warped linear prediction cepstral coefficients (WLPCCs) as well as MFCC parameters and their derivatives outperformed LPCCs in classification experiments using samples from five different audio databases.

In order to reduce dimensionality of complex datasets and at the same time

to retrieve the most representative variables, Principal Component Analysis (PCA) is commonly used (Sandell and Martens, 1995; Jensen, 1999). Apart from PCA, discriminant analysis (DA) (Agostini et al., 2001) and rough sets (Kostek, 1995; Wieczorkowska, 1999) have been proved to be reliable data reduction methods.

2.5 Single instrument timbre studies

So far, the presented studies dealt with the tasks of differentiation, categorisation and classification (whether perceptually or automatically) of various, typically orchestral, instruments. In contrast, there have been only few studies which focused on exploring psychoacoustic aspects of timbre of just one instrument. Timbre-describing adjectives or semantic labels have been often their major means of investigation.

For example, Abeles (1979) investigated verbal attributes commonly used by musicians to describe timbre of clarinet. The initially collected 118 descriptors of clarinet tone quality were evaluated in a survey and reduced to the 40 most highly ranked by the survey respondents. Two experiments were conducted on the acquired data. For each study, three groups of subjects were recruited amongst clarinettists, other music majors and non-music majors. In the first study, sound stimuli consisted of 24 clarinet tones recorded by three players in four different registers (two samples per register). The subjects' task was to choose up to five descriptors most adequate for characterising the clarinet tones, from a pool of five descriptors randomly ordered and selected out of the previously prepared list of 40 highly ranked attributes. Factor analysis with Varimax rotation produced a three-factor solution which accounted for 50% of the total variance. Based on most correlated opposite attributes: *centered–pinched*, *clear–fuzzy* and *resonant–?*, factors were labelled as Shape, Density, and Depth respectively.

In the second study, subjects evaluated 66 pairs of clarinet tones collated from the same sample set. They were asked to mark which tone in the pair is

best represented by a randomly selected descriptor from a list of eight descriptors (*mellow*, *controlled*, *clear*, *penetrating*, *airy*, *complex*, *pleasing*, and *interesting*). Ranking data was analysed for the consistency of the subjects individually, within a subgroup, and for the agreement between subgroups. Abeles concluded that the results identified terms which may not be appropriate for describing clarinet timbre, however, they failed to identify a subset of most salient ones. On the other hand, the two groups of musically trained subjects (clarinettists and other music majors) had generally higher levels of within-group and between-group consistency than did the non-musicians in the choice of adjectives most suitable for clarinet description. This result came in agreement with the observations made by other researchers that musical training is an important factor for obtaining reliable perceptual data if musical timbre is under examination.

Melka (1994) reported a series of perceptual experiments investigating timbre and sound quality of tenor trombones. "In spite of the fact that the ability of any language to express the timbre of a sound by verbal categories satisfactorily is limited", Melka was interested in the "timbre vocabulary" of Czech professional orchestral trombonists, which he collected through a postal survey and post-listening interviews. From the postal survey asking to list pairs of words or word groups which have opposite meanings and are used by the player to describe the tonal qualities of tenor trombones, 52 different adjectives were acquired. Pairs of opposite attributes were derived and subsequently subjected to hierarchical clustering (HICLUS) which produced seven clusters.

In a listening test, ten subjects evaluated in pairwise comparisons the sound quality of eleven different models of tenor trombones which were presented in two musical contexts. Musical phrases where recorded at the same dynamic level by the same performer using the same mouthpiece. After each pair rating, subjects were asked to provide verbal explanation of their choice. The resulting vocabulary consisted of 117 terms. Melka reported that HICLUS applied to this verbal data produced cluster structures more distinct and consistent across musical contexts than the structures obtained from the postal survey.

As an alternative way of uncovering the underlying structure of timbre from verbal attributes, principal component analysis (PCA) with Varimax rotation was employed using frequencies of the adjectives from the vocabulary as dependent variables. A three-factor solution accounting for 71% of the total variance was obtained for each musical context. The PCA results appeared to be in a close agreement with the outcome of the HICLUS analyses. Depending on the context the first factor was related to either "softness/roundness vs rudeness/sharpness" or "wideness/roundness vs sharpness/narrowness" and the second factor most corresponded to the attributes "clearness vs veiling" or "clearness vs veiledness". The interpretation of the third factor was ambiguous as it tended to split into two subfactors.

Additional similarity judgements evaluating timbre differences in pairs of trombone tones in two musical contexts were collected from the same group of subjects. A non-metric Euclidean distance based multidimensional scaling was applied to both similarity ratings yielding three-dimensional spaces. The same vocabulary and the adjective frequencies acquired in the preference test were used to interpret the dimensions. Employing an adapted property fitting technique, four property axes were found, two of which corresponded closely in their interpretation to Factors 1 and 2.

Based on the combined results of all three multivariate procedures (HICLUS, PCA and MDS), Melka suggested that at least in the two studied contexts, the two-dimensional perceptual space of trombone timbre can be interpreted in terms of "roundness/softness vs sharpness/narrowness" or "wideness/roundness vs narrowness/sharpness" (first dimension) and "clearness/wideness vs veiledness" or "clearness/concreteness vs veiledness/not ringing" (second dimension).

Fitzgerald (2003) conducted a series of four perceptual experiments aimed to identify the acoustic cues in oboe tone discrimination. In particular, she was interested in revealing psychoacoustic aspects of timbre which depend on performer and which may lead to differentiation between various oboe players.

A sound corpus for Experiments 1 and 2 consisted of tones recorded at six

pitches (C4, F4, A4, A#5, C#5, F6) and two contrasting dynamic levels (*mf* and *ff*) by two professional oboists representing the English or American school of playing. A total of 24 tones were normalised in loudness and edited for equal duration with an artificial decay lasting 0.6 s. Thirty two subjects (trained musicians) rated dissimilarity between pairs of tones (60 pairs in total including identical pairs, each pitch set was evaluated separately). Dissimilarity data was subjected to two MDS analyses. Firstly, an unweighted, non-metric Euclidean distance model was applied to the mean dissimilarity ratings across subjects for the six pitches. Secondly, a weighted individual differences scaling model (INDSCAL) was obtained separately for each of the six pitches. In both cases, MDS produced two-dimensional spaces in which one dimension separated the tones by oboist and the other by dynamic level (with some confusion for pitches C4 and C#5). Additional repeated measures ANOVA on the subjects' ratings indicated significant differences between tones for different performers across the same and different loudness levels, as well as significant differences between tones across different loudness levels within a performer.

The same sound stimuli were evaluated by the same group of 32 subjects using verbal attribute magnitude estimation (VAME) (Kendall and Carterette, 1993a). The aim of this experiment was to investigate whether perceptual differences between the oboe tones can be captured and effectively described by means of verbal attributes. A selection of eight adjectives (*tremulous*, *nasal*, *brilliant*, *reedy*, *strong*, *ringing*, *light* and *rich*) was made based on Kendall and Carterette's evaluation. PCA (with Varimax rotation) of the averaged VAME ratings across all pitches revealed three main factors: Power, Vibrancy, and Pinched which accounted for 54% of the total variance. Relating them to the two-dimensional MDS space, Fitzgerald suggested that the Power factor could act as a label on Dimension 2 (differentiating the tones by dynamic level), whereas the Vibrancy factor could be used to differentiate between oboists (Dimension 1).

Similar three-factor solutions were obtained for individual pitches except for

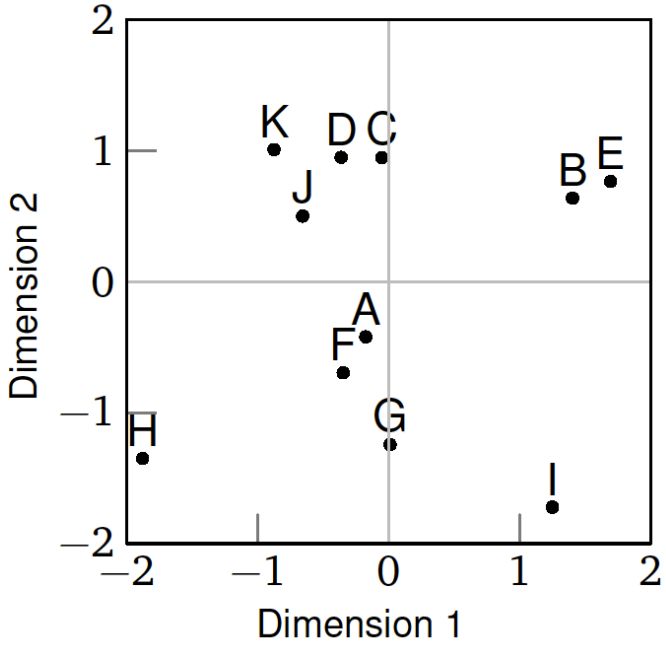


Figure 2.1: Two-dimensional MDS solution for the mean dissimilarity ratings of eleven oboists. (From Fitzgerald, 2003)

C#5 which loaded on four factors. Previously identified Power, Vibrancy and Pinched factors were relatively uniformly represented over pitches C4, F4, A4 and F6 via a consistent set of attributes loading positively, with some variations for the negative loadings. The least fitted three-factor model was obtained for A#5.

In Experiments 3 and 4, sound stimuli consisted of tones at the same pitch (A4), dynamic level (*ff*) and of equalised duration recorded by eleven oboists: two professionals (A, B) and nine students (C–K) (oboist B was influenced by the American school of playing). The study aimed to investigate perceptual differences across oboists and across schools of playing. Firstly, twenty two musically trained subjects were asked to make judgements of dissimilarity between pairs of tones (66 pairs in total including identical pairs). Classical unweighted MDS was performed on the dissimilarity ratings averaged across subjects, yielding a two-dimensional solution (Figure 2.1). To account for individual differences be-

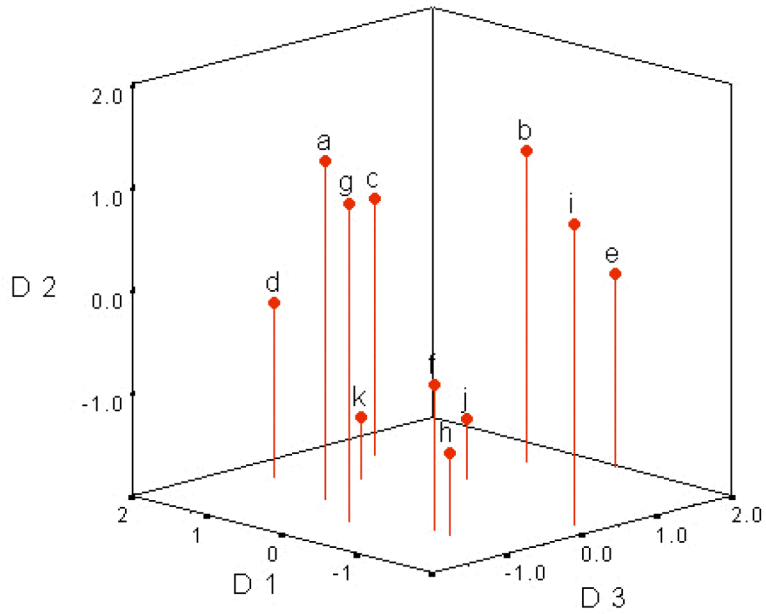


Figure 2.2: Three-dimensional INDSCAL solution for the eleven oboists. (From Fitzgerald, 2003)

tween subjects, INDSCAL was also performed on the subjects' individual ratings producing a three-dimensional configuration as an optimal solution (Figure 2.2). In comparison to the MDS results, HICLUS (complete-linkage) applied to the proximity ratings from each subject revealed two clusters clearly separating the oboist B (influenced by the American school) from the rest of oboists representing English school of playing (Figure 2.3). In the summary to Experiment 3, Fitzgerald concluded that both the MDS and HICLUS analyses produced similar results uncovering consistent similarities or differences between groups or pairs of players. The most interesting outcome from the HICLUS, showing oboist B clustered individually, strongly suggested further investigations into differences between schools of playing, as their influence seemed to be noticeable even in single isolated tones. On the other hand, subjects did not differentiate between professional and student oboists which could suggest that comparing just short tones may not be sufficient for the task.

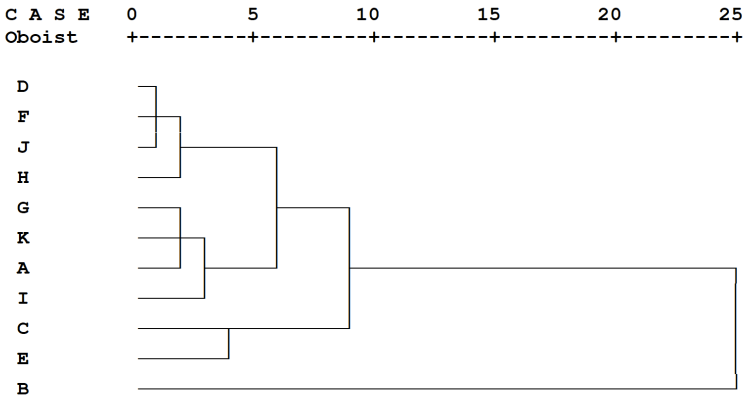


Figure 2.3: Dendrogram of HICLUS analysis of the dissimilarity ratings for the eleven oboists. (From Fitzgerald, 2003)

In Experiment 4, the same set of 22 subjects provided their VAME ratings of tones played by eleven oboists (the same tones used in Experiment 3). Three extra adjectives (*harsh*, *piercing* and *bright*) were added to the set of verbal attributes from the previous VAME study, which made a total of 11 attributes in the set. Similarly to Experiment 2, this study aimed to reveal perceptual differences between timbres of various performers which can be described through ratings of their verbal attributes.

PCA with Varimax rotation applied to the VAME ratings yielded a three-factor configuration accounting for 58.6% of the total variance (Figure 2.4). As in Experiment 2, factors Pinched, Power and Vibrancy were found to be the most representative for oboe timbre description. Comparing factor configurations from Experiment 2 and 4, Fitzgerald suggested that the reason for the Power factor (Factor 1) to account for the most variance in Experiment 2 was in varying dynamic levels that influenced the tones' perception the most. In contrast, in Experiment 4, where that influence was eliminated, subjects primarily focused on the degree of "oboeness" as reflected in loadings on the factor Pinched (Factor 1). She also reported that it was difficult to establish a relationship between the VAME and dissimilarity ratings, necessary for interpreting perceived timbral differences between the players with provided verbal attributes. She concluded

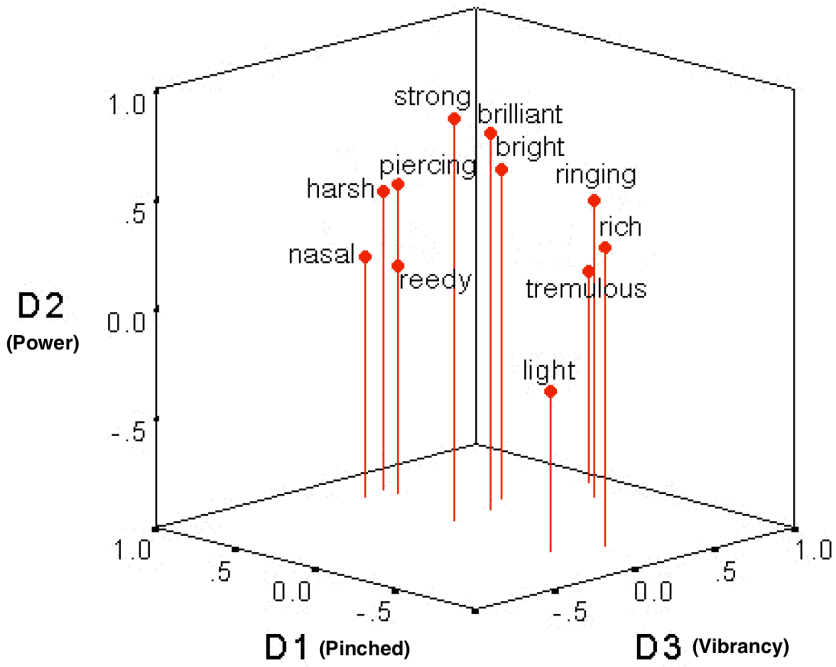


Figure 2.4: Three-factor configuration of the verbal attributes across the eleven oboists. (Adapted from Fitzgerald, 2003)

that either the selected adjectives were "not sensitive enough to describe the minute differences" or "the VAME rating task was too hard for subjects to discriminate over so many oboists in such detail".

In the last experiment, a set of acoustic features was extracted from the tones used in the perceptual studies in order to quantitatively explicate the physical dimensions of the oboe timbres under investigation. The acoustic features included spectral centroid (SC), spectral deviation (SD), spectral spread (SS), spectral variation (SV), spectral flux (SF), long-time average centroid (LTAC), centroid variability (CV), attack rise time (ART) and log attack time (LAT). Spectral and spectro-temporal parameters were extracted from 1-second long portions of the steady state of each tone. Frame-based instantaneous values of the features were subsequently averaged across time frames to obtain a single global value of each parameter. Two acoustical analyses were performed on the sound sets of 24 and 11 tones respectively.

Based on the PCA data and averaged VAME ratings from Experiments 2 and 4, factor scores for each oboist (or oboist/dynamic level condition) were calculated. These factor scores were then correlated with each of the global features obtained from the acoustical analysis. Results showed that for the three factors in Experiment 2, the Power factor best correlated with SC and SD, the Vibrancy factor with SS and SV, and the Pinched factor with SC and SV. In Experiment 4, significant correlations were found between factors: Pinched and SC/SV, Power and SD/SF, and Vibrancy and LTAC/SS. From the correlations between the averaged VAME ratings and acoustic features in Experiment 4, it was found that SC correlated positively and significantly with attributes such as *bright*, *harsh*, *piercing*, *nasal*, *brilliant* and *reedy*, which is in clear agreement with results of many timbre studies relating SC with a concept of perceptual *brightness*.

Fitzgerald's work deserves additional commentary not only because it significantly contributed to experimental research on musical timbre in general and on single instrument timbre in particular but, above all, because of its relevance to the development of this thesis. It has been, so far, the only work where performer–related facets of timbre were more thoroughly investigated, combining psychoacoustic and signal processing approaches. However, one important issue needs to be raised concerning the way the sound stimuli were designed. Both in Experiment 2 and 4, tone samples were recorded by the oboists on their own instruments and then subjected to dissimilarity rating. One might ask then whether resulting MDS timbre configurations (as illustrated in Figure 2.1 or Figure 2.2) actually reflect perceived dissimilarities between oboes rather than oboe players, thus undermining the validity of the presented results. Fitzgerald's standpoint was that "the combination of performer, reed and instrument should be treated as one complete mechanism" since "an oboist's reed and instrument are chosen and developed to suit the individual player whose choices have been influenced both by pedagogical, cultural and individual physical factors". Most instrumentalists would probably agree with this statement as

it stems from common practice and requirements of the profession. This indeed may sound more truly for oboe players considering the continuous necessity of reed scraping. However, regardless of a strong preference to always perform on his own instrument, any professionally trained musician possesses the skills and capabilities to perform enjoyably on any instrument of the same class, which is more than sufficient for a scientific experimental purpose. Therefore, if the research goal is to identify timbre cues which may contribute towards differentiating one performer from another, it is more than justified (if not recommended) to use tone or phrase samples registered on the same instrument by all players in question. Nevertheless, with a necessary reinterpretation of some of Fitzgerald's findings, her work still provides a wealth of new evidence extending our insight "in the micro–domain of oboe timbre".

Perceptual aspects of clarinet timbre in respect to two control parameters (related to the blowing pressure and the lip pressure on the reed) were explored by Barthet et al. (2010b) using multidimensional scaling and hierarchical clustering analysis of dissimilarity judgements. Sound stimuli for the experiments consisted of 15 short, sustained tones of E3 pitch generated by a physics-based synthesis model with varying blowing pressure and lip pressure values. Tones were subjectively equalised in loudness according to a reference signal. Sixteen musically trained subjects rated dissimilarity in pairs of non-identical tones (105 pairs in total) and were also asked to provide the criteria used for discriminating between stimuli. A non-metric MDS procedure yielded a three-dimensional perceptual space and a set of 21 acoustic descriptors extracted from the tones was employed to interpret the dimensions. It was found that coordinates of the timbre space were most correlated with the attack time or spectral centroid, tristimulus 2 and odd/even harmonic ratio descriptors (see Figure 2.5). Also, Dimensions 1 and 3 were highly and significantly correlated with the lip pressure and blowing pressure respectively (both correlations were positive). None of the control parameters correlated significantly with Dimension 2. Three distinct clusters of sounds obtained from HICLUS are also indicated in Figure 2.5. The

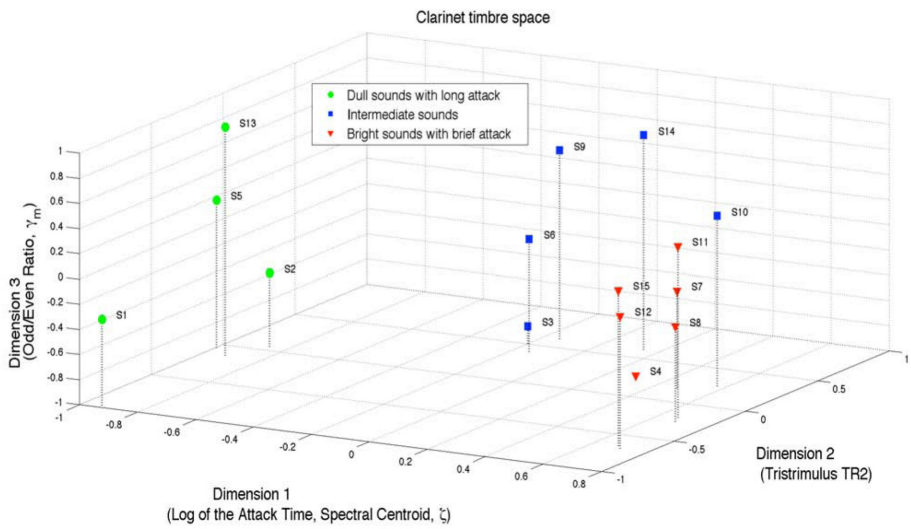


Figure 2.5: Three-dimensional clarinet timbre space and its mechanical and acoustical correlates. (From Barthet et al., 2010b)

first cluster (green markers) contained tones with smaller spectral centroid (SC) and longer attack time (AT), tones in the second cluster (in blue) had moderate values of both SC and AT descriptors, and tones with high SC (very bright) and short AT were gathered in the third cluster (red markers).

Qualitative analysis of verbal descriptions revealed three main criteria the subjects used for discriminating between varying clarinet timbres. They included categories such as Brightness, Attack and Tension. In particular, participants used words *bright*, *nasal* or *sharp* related to the brightness of the sounds, "softness of the attack" or "attack intensity" in relation to the dynamics of perceived onset transients, and attributes *soft* or *aggressive* to describe the sensation of tension in the sounds.

Anyone taking on the task of reviewing timbre related literature will be quickly struck by the fact that instruments from the bowed string family have been the least favoured amongst researchers investigating musical timbre. Štěpánek and his colleagues from the Prague-based Musical Acoustics Research Centre

have been amongst the very few who took up the challenge of examining violin timbre from the psychoacoustic perspective. From a long series of studies tackling different aspects of violin sound, results most relevant to this thesis are reported.

As a main sound corpus for their experiments Štěpánek chose violin tones of five different pitches (B3, F#4, C5, G5, D6) recorded by the same professional performer on twenty four violins of varying quality. Tones, played downwards *détaché*, *non vibrato*, at bow position *naturale* and in *mezzo forte* dynamics, were recorded in an anechoic chamber. The same loudness, pitch and tone duration were maintained during the session or otherwise later equalised. It is not fully clear whether the attack transients were removed from the signals, however, across their publications, the authors reported a few times using similar wording that "recordings of tones were subsequently manipulated to disable an influence of transient parts on perception" (Štěpánek and Otcěnášek, 2002)

In (Štěpánek et al., 1999), tones of 17 violins (at five pitches each) were evaluated in two listening tests. In the first one, 20 subjects (professional musicians) marked timbre dissimilarities in pairwise comparisons of all tones in each pitch set. The Euclidean distance based non-metric MDS of dissimilarity matrices yielded three-dimensional solutions for the pitches B3, F#4, C5, and two-dimensional solutions for G5 and D6 (no more details or illustrations of the obtained MDS spaces were provided). The second listening test included spontaneous verbal descriptions (SVD) of timbre differences in pairs of tones and judgements of preference of the perceived sound quality. This time 10 subjects evaluated recordings of eleven violins best represented in the perceptual spaces from the first experiment. The initial set of 267 collected words was reduced based on the overall frequency of occurrences (minimum 10 occurrences) and further subjected to correlation analysis in order to determine groups of relative/contradictory attributes. Finally, based on words with the highest overall frequency, four perceptual dimensions of violin timbre were identified (*soft–sharp*, *clear–damped*, *dark–bright*, *narrow*). Štěpánek et al. concluded that "the

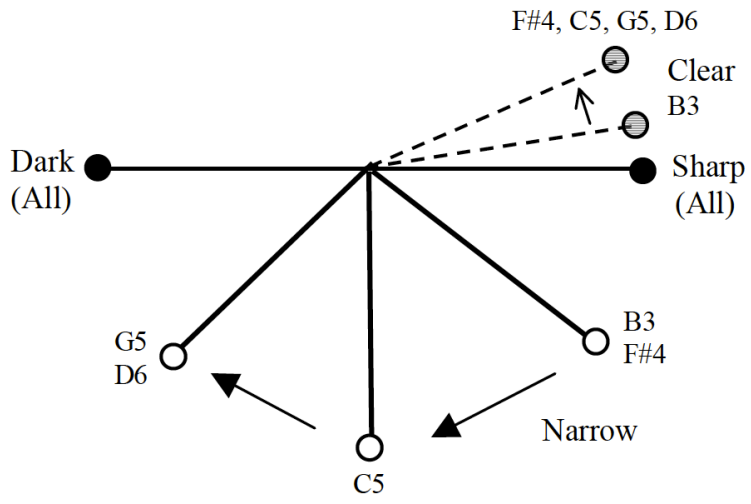


Figure 2.6: Generlised PCA solution of mean VAME ratings across five pitches. For each pitch, relative positions of the four verbal attributes are indicated. (From Štěpánek, 2002)

results are not definitive, but stable significant correlations of the frequency of occurrence of the words *soft* and *sharp* with a spectral centre of gravity, and *narrow* with a first harmonic level for all five tones support the existence of identified dimensions" (these "significant correlations" were not reported in the study).

In (Štěpánek, 2002) the same tones of eleven violins recorded at five pitches were evaluated according to four salient verbal attributes identified in (Štěpánek et al., 1999): *sharpness*, *clearness*, *darkness* and *narrowness*. Verbal attribute ranking and rating method (VARR) adopted from VAME was used to collect perceptual data. Eleven subjects (violin players and sound designers) ranked signals in each pitch set and then rated each tone on the magnitude scale from 0 to 10 according to the specified attribute. Principal component analysis (with Varimax rotation) of the mean ratings produced two-dimensional solutions for all pitch sets, summarised in Figure 2.6.

Analysis of correlations between mean ratings of verbal attributes indicated well established relationship between attributes *sharp*, *dark* and *clear* in all five tested pitches, where *dark* and *sharp* were the opposite attributes along

the same perceptual dimension. The perception of *narrowness* changed with pitch, from being positively correlated to *sharpness* for B3 and F#4 tones to become more closely related to *darkness* for pitches G5 and D6. Additional correlation analysis of VAME and perceived sound quality ratings revealed that better sound quality was most strongly associated with "darker" tones across all pitches except for G5, for which the *clearness* rather than *darkness* seemed to indicate a tone of good quality.

Spectral characteristics of violin tones in relation to verbal attributes: *sharp*, *dark* and *narrow* were examined in (Štěpánek and Otcěnášek, 2002; Štěpánek, 2004; Štěpánek and Otcěnášek, 2004, the results also reported in Štěpánek and Otcěnášek (2005)). Spectral features were calculated from the time-averaged power spectrum of the steady state of the sound and included amplitudes of individual harmonics (in dB), levels in critical bands (in Barks) and spectral centre of gravity (i.e. spectral centroid, in Hz). Eleven violins' tones of five different pitches (as used in the previous studies) were spectrally analysed and the obtained features were subsequently correlated with mean VAME ratings. For all pitches except for G5, higher levels of the fundamental were positively correlated (highly and significantly for B3, F#4 and C5) with the attribute *dark* and negatively with the attribute *sharp*. Stronger fundamental was also negatively correlated with the *narrowness* of the sound (highly and significantly for pitches B3, F#4 and G5). The perceived *sharpness* was found to correlate significantly and positively with spectral centroid for all pitches, again with exception of G5, and with larger amplitudes in higher critical bands (across all pitches) for band indexes varying between 18 and 24 depending on pitch.

In an additional series of experiments, Štěpánek and Otcěnášek (1999); Štěpánek et al. (2000) investigated spectral sources of the *rustle* attribute (also associated with words: *sandy*, *hissy*, or *dusty*), which appeared very often in verbal descriptions of D6 tones. The frequencies of the overall occurrence of word *rustle* and its synonyms acquired from spontaneous verbal descriptions were correlated with spectral characteristics of the signals. The results and

complementary listening tests suggested that higher amplitudes of frequency components either below the fundamental in the bands between 200 and 900 Hz or above 8 kHz (from the 7th harmonic onwards) may be responsible for the presence of *rustle* in the violin tone. It was also observed that the phenomenon occurred predominantly in lower quality instruments.

2.6 Timbre as a means of musical expression

The role of timbre in conveying the contents of music and particularly as a means of performer's expression has received considerably less attention than other performance parameters such as timing, dynamics, phrasing and articulation. Holmes (2011) suggests a few reasons for such a state of affairs. Firstly, timbre is by far the most difficult attribute to measure, following that to decompose the tone production process into some measurable variables becomes a challenge itself. Secondly, especially in Western music notation, there are relatively few indications as to what sort of timbral shape is desired for a particular motive, phrase or section, leaving space for performers to interpret the score freely. Thirdly, significant variability in perceptual judgements constantly raises the need for a reference point, i.e. what sounds *bright* to one listener may sound not so *bright* to another, posing the question of how the difference can be quantified or whether an objective scale can be established, at all. Lastly, expressive use of timbre has been exceptionally a domain of performers, particularly individual and ephemeral (Holmes, 2011), thus hard to capture empirically.

In an attempt to address the problem, Barthet et al. (2010a) investigated a set of acoustic factors accountable for expressiveness in clarinet performances. For that purpose, mechanical and expressive performances of two music excerpts were recorded in an anechoic chamber by one performer. Recordings were segmented into notes and a set of note–level descriptors was extracted. They included the timbre (attack time, spectral centroid, odd/even harmonic ratio), timing (intertone onset intervals), dynamics (root mean square envelope) and

pitch (fundamental frequency F0) parameters. A two-way ANOVA with the musician's expressive intentions and the notes as factors indicated a strong effect of the expressive intention on attack time (AT), spectral centroid (SC), odd/even harmonic ratio (OER), intertone onset interval (IOI) deviation and root mean square (RMS) envelope in both music excerpts. Significant interactions between the two factors also suggested that stronger variations in the timbre descriptors occurred depending on the position of the notes in the musical phrases. The authors concluded that "timbre, as well as timing and dynamics variations, may mediate expressiveness in the musical messages transmitted from performers to listeners".

To perceptually validate the obtained results, Barthet et al. (2011) investigated the effects of previously identified salient acoustic parameters on listeners' preferences. Using an analysis-by-synthesis approach, the same expressive clarinet performances were altered by reducing the expressive deviations from the descriptors. The alterations included SC freezing, i.e. partial removal of spectral flux, IOI deviation cancellation, i.e. replacing the effective IOIs with the nominal ones as given by score, and compression of the dynamics. From the recorded two excerpts only first phrases were selected as stimuli and subjected to the three alterations and their four combinations giving 8 sound files in total per excerpt (including the originals). Twenty musicians were asked to mark their preference in a pairwise comparison task. Each excerpt's stimulus set was assessed separately. A two-way repeated measures ANOVA was conducted on the mean preference scores across subjects to assess the effect of musical excerpt (two levels) and alteration (8 levels) as factors. No effect of musical excerpt was found while the effect of alterations was highly significant. *Post-hoc* multiple comparisons revealed that the SC freezing, i.e. removal of the spectral centroid variations, resulted in the greatest loss of musical preference. Surprisingly, the preference scores for IOI deviation cancellation or dynamic compression or these two alterations combined were still higher than the scores for spectrally altered samples. One would rather expect that removing timing deviations should be

the least preferred option. As a possible explanation for such high influence of SC freezing on the subjects' preferences, Barthet et al. suggested that altering the spectral centroid could affect the perceived timbre of the clarinet, i.e. its timbral identity and, by "causing the tones to be static and unlively", decrease the sound quality.

The outcomes of the two studies (Barthet et al., 2010a, 2011) have serious implications for further research into music performance. They empirically examined and proved that timbre variations play a fundamental role in expressive performance (at least equal to timing and dynamics variations) and as such they have a profound effect on the quality of musical communication between performers and listeners (see Holmes, 2011, for a review).

2.7 Methods for measuring perceptual attributes of timbre

From the corpus of literature reviewed over the previous sections, it becomes evident that a variety of methodological approaches can be employed for the task of studying such a complex variable as timbre. In methods adapted for measuring timbre's perceptual attributes, a group/groups of subjects is typically presented with a specific type of evaluation task including identification, classification (categorisation), matching, discrimination, proximity rating (similarity/dissimilarity rating) or semantic scaling (verbal attribute rating), which is executed upon hearing a set of sound stimuli (McAdams, 1993; Hajda et al., 1997).

Identification refers to the task of assigning a name or label to a sound stimulus according to its class or category, based on either the subject's *a priori* knowledge and experience (free identification) or a provided list of labels (forced identification). The number of hits and misses per category is usually stored in a confusion matrix and subsequent analysis of the confusions allows to determine features different stimuli may have in common (McAdams, 1993). Examples

of the identification technique can be found in earlier studies on timbre (Clark et al., 1963; Saldanha and Corso, 1964; Berger, 1964; Wedin and Goude, 1972).

In a classification task, subjects are asked to sort a set of sound stimuli into groups or classes which best represent their common features. In free classification, subjects can choose the number of classes they think is the most appropriate, while in other classification variants a list of predefined categories may be imposed.

Matching requires a listener to choose amongst the presented comparison stimuli the one which belongs to the same class or category as the model stimulus. This method has advantages over identification as it does not involve semantic labelling or require prior familiarity with the sound objects under investigation. A matching procedure was used for example by Kendall (1986) (see Section 2.3.2).

Discrimination refers to the task of subjectively differentiating between a pair of stimuli which differ in some controlled way. This method allows one to determine the so-called just noticeable difference (JND) within a set of stimuli where the level of modification is strictly controlled by an experimenter. Discrimination tasks were employed by Grey and Moorer (1977) to evaluate resynthesised tones against their original counterparts in terms of discriminability.

Methods such as proximity rating and verbal attribute rating including examples from the experimental literature are discussed in more detail in the following sections. These methods were selected for collecting perceptual data in the experiments described in Chapter 6.

2.7.1 Proximity rating

Proximity rating requires a subject to evaluate the level of similarity or dissimilarity between each pair of stimuli in a dataset. The number of pairs to rate is $n(n-1)/2$ or $n(n+1)/2$ if identical pairs are included. In a typical scenario, subjects mark their rating on a given scale, either continuous or Likert-type,

and the results are collected in individual proximity matrices. In the next step, proximity matrices are subjected to a multidimensional scaling procedure (individually or averaged across subjects) to obtain a graphical representation of perceptual distances between the stimuli.

From a psychometric point of view, proximity rating has an advantage of being independent from the subject's *a priori* knowledge or preconceptions about the stimuli being compared. Hajda et al. (1997) pointed out that the method has been proved feasible for a number of stimuli $8 < n < 25$, however, for a set of 25 stimuli it would mean rating 300 pairs – quite a substantial cognitive load on subjects. Despite this limitation, proximity rating has been applied in a considerable number of studies on timbre that laid foundations for the current understanding of the phenomenon. They include works of Plomp (1970); Wedin and Goude (1972); Wessel (1973); Grey (1975); Miller and Carterette (1975); Kendall and Carterette (1991); Kendall et al. (1995) and studies by Grey (1977); Grey and Gordon (1978); Krumhansl (1989); Iverson and Krumhansl (1993); McAdams et al. (1995); Kendall et al. (1999); Lakatos (2000); Caclin et al. (2005) outlined in more detail in Section 2.3.4. Dissimilarity ratings have also been employed in this thesis to investigate perceptual differences in tone quality within a group of cello players.

2.7.2 Multidimensional scaling

Multidimensional scaling (MDS) is often closely associated with dissimilarity ratings as a primary method for analysing proximity data. The concept behind MDS is to uncover the underlying structure hidden in the data and to help to establish quantitative relationships between the stimuli along potentially unknown dimensions or attributes. First introduced by Torgerson (1952), a classical MDS model (CMDS) and its further adaptations assumes proximities between objects in the original N-dimensional space to have metric properties, i.e. to be distances in the Euclidean sense, and attempts to reproduce them in a low-dimensional space. In reality, this assumption might be too restrictive when

proximities represent subjective human ratings of a psychological phenomenon which can not be measured in metric units. Non-metric or ordinal multidimensional scaling developed by Shepard (1962a,b) and Kruskal (1964a,b) overcomes this limitation allowing to interpret proximities in an ordinal sense, i.e. only the ranks of the distances are known. In the resulting low-dimensional space only these ranks are reproduced, not the distances themselves.

In contrast to the basic non-metric model which assumes that subjects use the same perceptual dimensions to compare objects, the weighted Euclidean model or INDSCAL (Carroll and Chang, 1970) weights these common dimensions differently by each subject. More complex models account also for dimensions or features that are specific to individual stimuli, called "specificities" (EXSCAL, Winsberg and Carroll, 1989) and different weights assigned to latent classes of listeners (CLASCAL, Winsberg and De Soete, 1993). The CONSCAL model by Winsberg and De Soete (1997) allows mapping between audio descriptors and the position of sounds along a perceptual dimension to be modelled for each listener. More details on applying the CLASCAL and CONSCAL models in the context of timbre research can be found in McAdams et al. (1995) and Caclin et al. (2005) respectively.

Across timbre studies, different MDS models were used in combination with dissimilarity ratings discussed in the previous section (2.7.1). For example, a simple non-metric MDS was employed by Plomp (1970), Iverson and Krumhansl (1993) and Kendall et al. (1999), INDSCAL by Grey (1977) and Kendall and Carterette (1991), EXSCAL by Krumhansl (1989) and CLASCAL by Lakatos (2000). For this study, the basic non-metric MDS technique was chosen to analyse dissimilarity ratings of six cello players' timbres.

2.7.3 Semantic labelling: verbal attributes of timbre

Verbal descriptions of musical timbre, though widely used over centuries by generations of musicians and composers to characterise desirable qualities of musical tones or phrases, only with the launch of scientific explorations of the

phenomenon have begun to be associated with particular shapes of harmonic spectra and their varying harmonic components. In his seminal work, Helmholtz (1877) for example described "simple" tones as *soft* and *pleasant* without any *roughness* but *dull* at low pitches, while "complex" musical tones are *rich* and *splendid* if they have more pronounced lower harmonics (up to the 6th), and also *sweet* and *soft* in absence of higher upper partials. Tones with only odd harmonics sound *hollow* and turn to *nasal* when a larger number of upper partials is also present. Higher amplitudes of harmonics beyond the 6th or 7th are found in tones perceived as *cutting* and *rough*, also *harsh* or *penetrating* (Helmholtz, 1877, pp. 118–119).

From these first adjectives describing timbre of tones, through introduction of semantic scales, more detailed explorations of verbal attributes followed resulting in works of Solomon (1958); Bismarck (1974a,b); Kendall and Carterette (1993a), for example. Across studies different techniques of acquiring initial sets of adjectives for the experiments included postal and electronic surveys, pre- and post-listening interviews, and spontaneous verbal descriptions of the stimuli. The semantic differential technique (Osgood et al., 1957) was commonly utilised to obtain ratings on verbal attributes of different instrument timbres. In this method, subjects are presented with preselected scales, each set up using polar adjectives (opposite-meaning terms) at the extremes, e.g. *dull–sharp*, and are asked to evaluate each stimulus along all bipolar dimensions. Verbal attribute data is most often subjected to factor analysis or principal component analysis to reduce the number of semantic dimensions to the most salient ones. Hajda et al. (1997) suggested also to use ANOVA for comparing means of groups of subjects, individual subjects (if repeated measures are used) or instruments.

Bismarck (1974a,b) employed the semantic differential and PCA to evaluate 35 spectrally shaped harmonic complex tones and noises along 30 adjective scales (tones were equalised in pitch and loudness). Two groups of subjects, musicians and non-musicians, rated the stimuli. Bismarck reported that out of 30 scales

just four were sufficient to describe the analysed timbres: *dull–sharp*, *compact–scattered*, *full–empty* and *colourful–colourless*. The first semantic dimension, relating to the attribute *sharpness*, accounted for most of the variance in the data (44%), followed by the second dimension (the attribute *compactness*) which explained 26% of the total variance. He also found that *sharpness* was primarily determined by "the frequency position of the overall energy concentration of the spectrum" (i.e. spectral centroid) and that *compactness* differentiated between noise and tone stimuli.

Kendall and Carterette (1993a) validated Bismarck's findings and took the investigation into verbal attributes of timbre a step further. A subset of eight bipolar adjectives from his experiments was used to rate ten natural wind instrument dyads by five musically untrained subjects. PCA on the verbal ratings produced one factor accounting for 89.4% of the variance, which in fact did not differentiate among dyad timbres. To improve the result, Kendall and Carterette proposed to use verbal attribute magnitude estimation (VAME), in which an antipode of an adjective was created with its negated version (*sharp–not sharp*) instead of using the opposite term (*sharp–dull*), thus the subjects' task was to rate the degree of a single attribute in each stimulus. The results of PCA yielded a three-factor structure (90.6% of the total variance), in which the first two factors grouped attributes such as *heavy*, *hard*, and *loud* vs *compact*, *dim*, and *pure* and *sharp* vs *complex* respectively. Further analysis showed that VAME ratings allowed to separate *loud*, *heavy* and *hard* alto saxophone dyads from all others but did not confirm the salience of the *sharp* attribute in discriminating wind instrument timbres, the result which was attributed to the likely cultural differences between German and English languages concerning the meaning of *sharp* in a musical context. Kendall and Carterette concluded that "the von Bismarck adjectives lacked ecological validity", and subsequently conducted a new series of experiments "searching for terms more musical and relevant for describing timbre".

In (Kendall and Carterette, 1993b), they collated 21 adjectives from Piston's *Orchestration* (1969) to be used in VAME ratings of 10 wind instrument dyads from the previous study. This time ten music majors were asked to rate the stimuli according to each attribute on a 0-100 scale. PCA of the mean verbal attribute ratings across dyads revealed four semantic factors, accounting for 86.34% of the variance: Power, Strident, Plangent[1], and Reed. Cross-correlations among attribute ratings were subjected to classical MDS which produced a three-dimensional solution. Dimension 1 (*strong* vs *weak*) corresponded to Factor 1 (Power), Dimension 2 (*nasal* vs *rich*) corresponded to Factor 2 (Strident) and Dimension 3 (*simple* vs *complex*) corresponded to Factor 3 (Plangent). Factor 4 (Reed) correlated with Dimension 2, corresponding to *nasal* vs *not nasal* attribute.

The VAME procedure was also incorporated in (Kendall et al., 1999, see Section 2.3.4) to explore the verbal characteristics of natural and synthetic single instrument tones. Eight highest loading attributes from PCA analysis of (Kendall and Carterette, 1993b) were selected: *strong*, *tremulous*, *light* (Factor 1), *nasal*, *rich* (Factor 2), *brilliant*, *ringing* (Factor 3) and *reedy* (Factor 4). Twenty two subjects, musicians and non-musicians, rated the magnitude of each attribute on a 100-point scale and the verbal data was subjected to PCA with Varimax rotation. Three semantic factors emerged: Power/Potency, Stridency/Nasal and Vibrato, which accounted for 83% of the variance. Cross-correlations of the physical measures and verbal attribute ratings showed that *nasality* correlated highly with long-time average spectral centroid and the first perceptual dimension (results on timbral similarity were reported in Section 2.3.4) while the second dimension correlated with spectral variability and only moderately with attributes *rich*, *brilliant*, and *tremulous*.

The eight attributes selected by Kendall et al. and the resulting semantic

[1] According to Kendall and Carterette's note, the attribute Plangent, meaning "*reverberant*, *ringing*, and *resonant*, tinged with *plaintiveness*", was created by Terrence Rafferty to describe the sound of Wynton Marsalis' trumpet.

factor structure formed a basis for Fitzgerald's experiments on verbal description of oboe timbre (discussed in detail in Section 2.5). She found them suitable for describing oboe timbre in general, but not specific enough to capture individual "oboe qualities" of compared performers.

Since its introduction, VAME has been utilised quite often in semantic studies on musical timbre, either investigating generalised attributes for describing timbres of different musical instruments (e.g. Disley et al., 2006; Zacharakis et al., 2012, 2014) or the timbral palette of a single instrument (e.g. Štěpánek, 2002; Fitzgerald, 2003).

The effectiveness of experimental design end experiment reproducibility in the context of perceptual judgements of timbre using verbal attributes was assessed by Darke (2005). In the study, twenty two musician subjects were asked to evaluate 15 sounds of pitched orchestral instruments against 12 adjectives using a VAME-like procedure, i.e. marking their judgement of "How Bright" is the sound on a 0–5 scale. In the discussion, he concluded that the results show "no conclusive evidence that subjects agree on how to effectively communicate timbral issues" and highlighted some potential causes of lower levels of agreement between subjects or within-subject consistency, which are often overlooked by experimenters and which might undermine credibility of the reported findings.

2.8 Remarks and conclusions

The examination of literature dealing with a variety of issues related to timbre perception revealed something quite "remarkable". That is, it appears that, across numerous perceptual studies, instruments of the bowed string family have been hugely underrepresented, since most stimulus sets typically included only one sample of either violin or cello (with exception of Grey (1977) who included three cello samples varied in playing technique). For comparison, in the same study, the woodwind family was represented by as many as eight different instruments! One could possibly argue that the principles of tone production for the strings are the same, so violin or cello can "stand for" the rest of the

family. However, does violin in any manner sound similar to double bass?

The implications of this state of affairs for the research here undertaken were not trivial. The major experimental studies in the field of timbre perception obtained perceptual spaces based on sound sets in which the strings were practically non-existent. One could ask to what extent the resulting timbre spaces can represent timbres not included in the stimuli. Thus, are these findings actually relevant to the strings?

Furthermore, the revealed acoustical correlates of timbre dimensions such as spectral centroid and attack time (which account for the brilliance of the tone and the rapidity of the attack respectively) and to a lesser extent measures of spectral fluctuation or irregularity over time seem to capture psychophysical differences between musical instruments quite effectively. However, are these descriptors reliable for characterising timbre of the strings and cello timbre in particular? A limited number of studies dealing specifically with timbre of bowed instruments from a psychoacoustic perspective have not as yet answered this question. The very few studies reviewed in this chapter have produced rather inconclusive results. Therefore, a set of acoustic descriptors examined by Eerola et al. (2012) on a broad range of instruments (110 in total) including 32 samples from the string family was taken into consideration for the acoustical analysis carried out in Chapter 7. Furthermore, since not much insight has been offered either, in regard to verbal descriptions of cello timbre, the selection of verbal attributes for the experiment described in Chapter 6 was mostly drawn from the vocabulary from the studies on violin.

The right choice of stimuli and participants in a perceptual study plays a fundamental role for the validity of the results. Since in this research a perceptual evaluation was planned as the first and defining experiment, methodologically vital decisions were made about participating subjects and selected sound samples based on the reviewed studies. Kendall (1986) found that the onset and decay portions of tones are not important for instrument identification when the tones are heard in a musical context, i.e. in a melody. This indirectly

implies that timbre identity of the stimulus should be possible to grasp to a similar extent (if not greater) in a musical phrase as it is for isolated tones. Note that Fitzgerald (2003) was able to discriminate between different oboes (oboe/performer combinations) based on single sounds. Therefore, capturing timbre differences/identities among a group of performers based on same musical phrases, all recorded on the same instrument, rather than on isolated tones seems more musically valid.

In regard to the choice of subjects for experiments, the importance of musical training had to be considered as a factor of higher reliability of the collected perceptual data; a fact quite often highlighted in the literature. Findings of Beal (1985) and Pitt (1994) suggested that musicians are able to separate pitch and timbre fluctuations and attend only to timbre dissimilarities, at least in single tone comparisons. Following Kendall's argument, perception of timbre identity (of a performer) is more likely to remain invariant for stimuli longer than just single tones. Therefore, expert listeners such as musicians should also be able to evaluate timbral differences when comparing the sequences of pitches, (i.e. short musical fragments) which come from the same instrument but are played by different performers.

Chapter 3

The Cello: acoustic fundamentals and playing technique

3.1 Introduction

The cello or *violoncello* belongs to the violin family of musical instruments which includes violin, viola, double bass and their predecessors. It is also a member of a wider class of so-called string or stringed instruments (chordophones) for which the primary source of vibration is one or more stretched strings. There are three different ways of setting a string into vibration: plucking (e.g. lutes, guitars, harpsichords, clavichords), striking (e.g. pianos) and bowing (e.g. the violin family). In any string instrument, energy from the vibrating string is transferred via the supporting bridge to the instrument body which acts as a resonator (or sound modifier, Howard and Angus (2009)) since the string itself can hardly produce any sound (Guillaume, 2006). Vibrations of the body can be categorised into *free* and *driven* (*ibid.*). The former occur when the body after receiving an initial impulse (e.g. plucking or striking) is left to vibrate freely without any further input, taking as an example the harpsichord, the piano, the guitar, also the violin played *pizzicato*. The latter occur when the sound is sustained by the player by bowing, as in case of bowed string instruments (the violin family), or blowing into the mouthpiece for wind instruments (*ibid.*). The sustained model of vibrations gives a player control over the quality of tone at any time point of the sound production process.

The cello shares the same construction principles with the rest of the violin

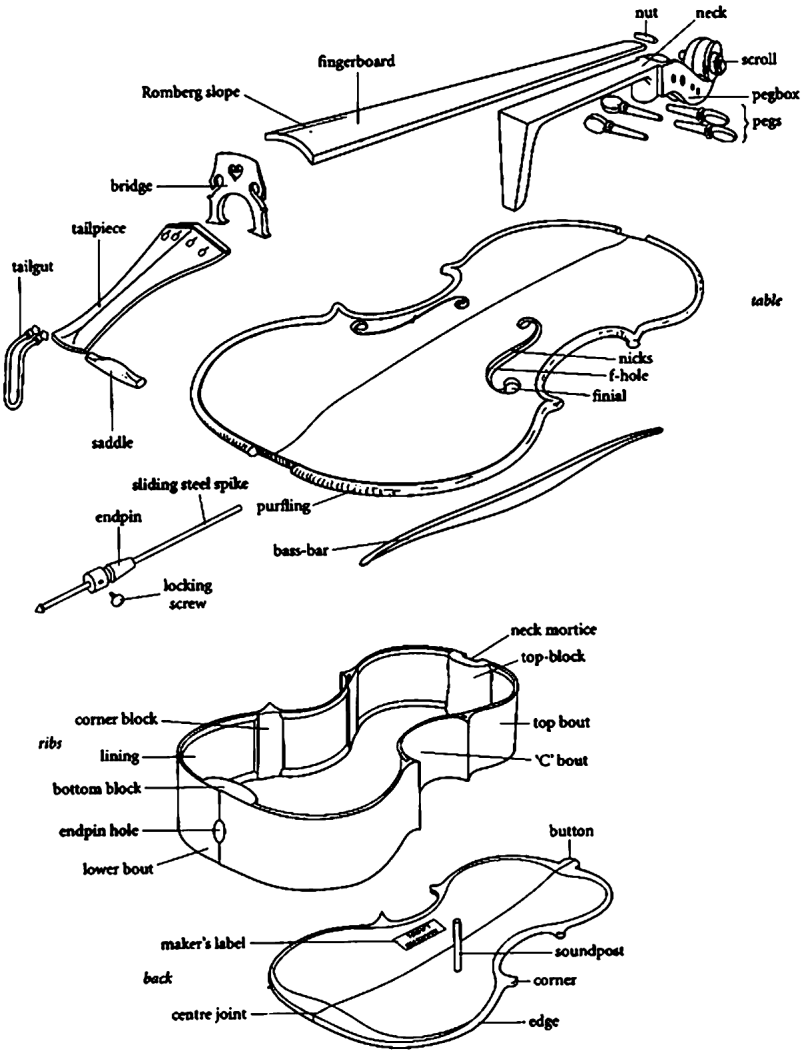


Figure 3.1: Component parts of the cello in detail. (From Bynum and Rossing, 2010)

family. Figure 3.1 illustrates an exploded view of its component parts in detail. In terms of building materials, the back plate, ribs and the neck, carved in one piece with the pegbox and scroll, are most often made of maple, while the top plate is generally made from spruce. The fingerboard is usually of ebony, and pegs, endpin and tailpiece can be made of ebony, rosewood, or boxwood.

The four strings of the cello are tuned to C2, G2, D3, and A3, resulting in a pitch range from C2 to C6 (65.4–1046.5 Hz) and beyond if using string

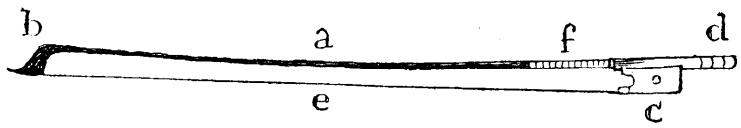


Figure 3.2: The cello bow: (a) the stick; (b) the tip; (c) the frog; (d) the screw; (e) the hair; (f) the lapping (wrap). (From Straeten, 1905)

harmonics (flageolet tones). This places the cello tuning an octave below the viola and a twelfth (octave plus a fifth) below the violin. Theoretically, being tuned a twelfth below the violin would require the cello body size three times larger than that of the violin to accommodate longer and lower-pitched strings (Richardson, 1999). In fact, the length and width of the cello body are closer to twice rather than to three times those of the violin and the compromise in size is achieved via increased rib height and relatively thinner construction to keep the resonances sufficiently low for bass enhancement (*ibid.*).

The cello bow (Figure 3.2) is slightly shorter than the violin and viola bows, thicker and less springy (Piston, 1969). The tip and stick of the bow are typically carved from one piece of pernambuco wood known for its unique combination of strength and resilience (Dilworth, 1999). Other possible materials include brazilwood used for inexpensive bows and carbon fibre which has become more and more popular over the last two decades. The bow hair is usually made of horsehair but synthetic (e.g. nylon) or metal threads are also in use. To secure the right amount of friction at the point of the bow and string contact, rosin is rubbed on the bow hairs.

3.2 Acoustical properties of the cello

A cello player generates the sound by drawing a bow perpendicularly across a string. Friction between the bow and the string sets the string into vibration. In particular, when the bow is moved across the string in either direction, the string is gripped and moved away from its equilibrium (so-called *stick* phase) until the string releases itself, moving past its equilibrium until the bow hairs

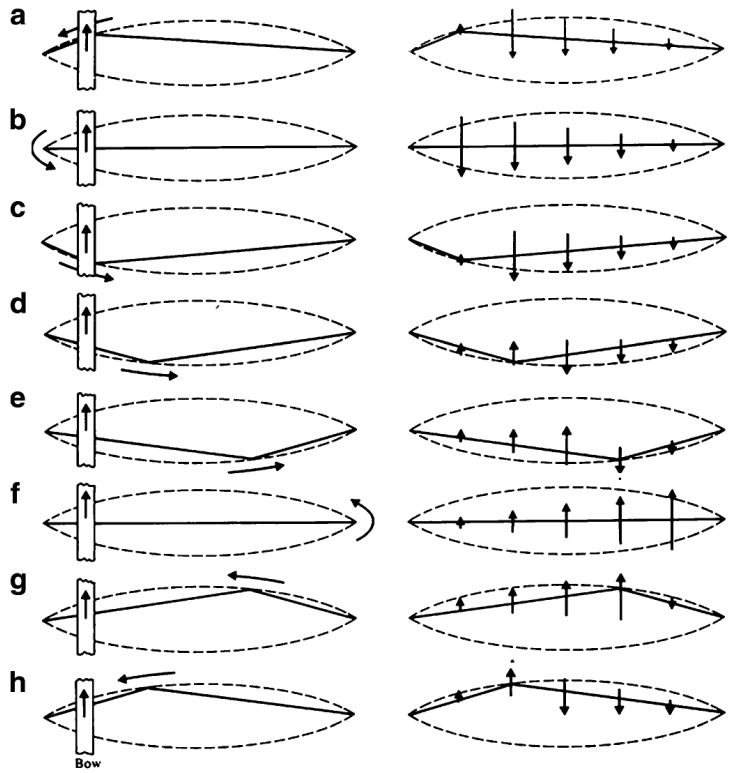


Figure 3.3: The motion of a bowed string at successive times during the vibration cycle. (left) The bend races around an envelope; (right) the velocity of the string at different times in the vibration cycle. (From Rossing, 2010)

grip it again to repeat the cycle (so-called *slip* phase). The stick-slip cycle is repeated continuously, i.e. stick-slip-stick-slip-stick etc. and hundreds of stick-slip cycles may occur while the player is moving the bow in just one direction (Jansson, 2002).

3.2.1 Motion of the bowed string

To the naked eye, the string appears to vibrate back and forth smoothly between two curved boundaries, much like a string vibrating in its fundamental mode. Helmholtz (1877) observed that, in fact, the string forms two straight lines with a sharp bend at the point of intersection (also called the Helmholtz corner). This bend travels along the envelope, which is made of two parabolic segments, concluding one round trip each period of the vibration as illustrated

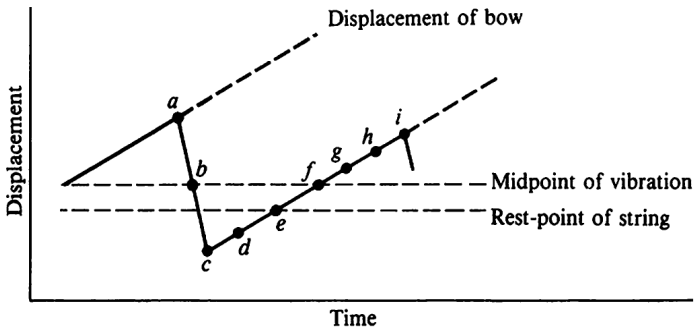


Figure 3.4: Displacement of bow and string at the point of contact. The points (a)-(h) correspond to the (a)-(h) steps shown in Figure 3.3. (From Rossing, 2010)

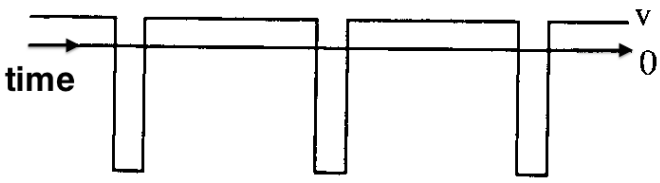


Figure 3.5: String velocity waveform at the bowing point. (Adapted from Woodhouse, 1997)

in Figure 3.3. When the bow moves in the other direction, the pattern is reversed (Howard and Angus, 2009). The motion of the string under a moving bow was named after its explorer as Helmholtz motion.

During the slip phase, as the bend passes the point of bowing, it triggers transitions between sticking and sliding frictions and the string makes a rapid return until it is caught by a different point on the bow (points a to c in Figure 3.4). During the stick phase, when the string is carried along by the bow hairs, it moves with the same velocity as the bow, i.e. the bow velocity (see points c to i in Figure 3.4). This results in a velocity waveform at the bowing point as shown in Figure 3.5.

The vibration of the string at the bridge results in a sawtooth force waveform applied to the bridge (see Figure 3.6). The spectrum of an ideal sawtooth waveform (Figure 3.7) contains all harmonics and their amplitudes decrease

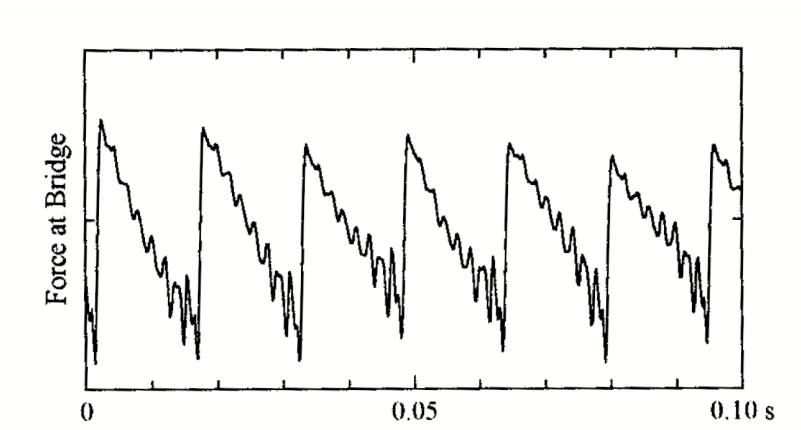


Figure 3.6: Waveform of time-varying transverse force exerted on the cello bridge by the open C string. The time period is approximately 15 ms. (From Richardson, 1999)

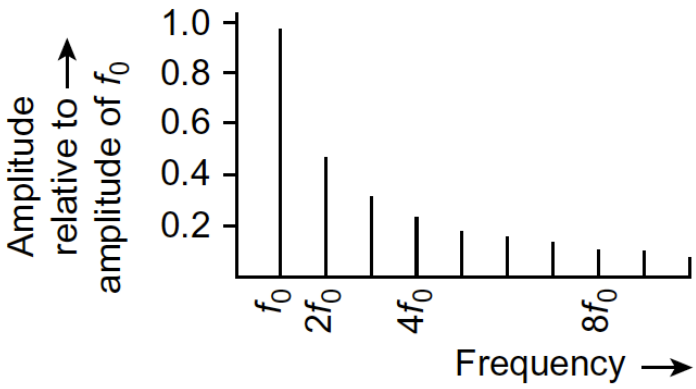


Figure 3.7: The spectrum of the ideal sawtooth waveform. (From Howard and Angus, 2009)

with ascending frequency as $(1/n)$, where n is the harmonic number.

3.2.2 Resonances of the cello body

When a bowed string is set into vibration, it produces a vibration force on the bridge, which is then transmitted via the bridge to the top plate and thereafter to the entire body of the cello. Once the complete body is in motion, its vibrations set the surrounding air into vibration resulting in the audible sound. Hence, the cello body acts as an effective sound radiator or acoustical amplifier and

modifier for the sound source provided by the bowed string (Jansson, 2002).

The sound quality and playability of a string instrument is determined by the vibrational properties of its body. While all component parts contribute to the sound modification process, in the case of the violin family, the output tone is shaped mainly by the coupled motions of the top plate (table), back plate, and enclosed air. The complex vibrations of the body are typically described in terms of normal modes of vibration or eigenmodes (Rossing, 2010). Being associated with structural resonances, the normal modes of violins or cellos have been classified according to the primary vibrating element as:

- Air modes ($A_0, A_1, A_2, ...$) related to substantial motion of the enclosed air;
- Top modes ($T_0, T_1, T_2, ...$) indicating motion primarily of the top plate;
- Body modes ($C_0, C_1, C_2, ...$) in which the top and back plates move similarly.

  (after Fletcher and Rossing, 1998)

One way of measuring how an instrument vibrates or radiates sound at different frequencies is to measure its frequency response. The frequency response can be expressed in terms of the mobility (or mechanical admittance) when an applied sinusoidal force, for example at the bridge, is observed as a velocity at some other point, or in terms of radiance when the pressure of the radiated sound is captured with a microphone (*ibid.*). An example of bridge input admittance for a cello is shown in Figure 3.8. Determined by the instrument construction, the unique details of the response curve form an "acoustical fingerprint" which in turn determines the sound quality and playability of a particular instrument (Richardson, 1999).

Peaks in the response curve correspond to mechanical resonances of the body, i.e. modal frequencies. The lowest mode of acoustical importance, A_0, often referred to as "air resonance" or "f-hole resonance" involves both structural vibrations and significant air displacement through the f-holes (Figure 3.9, lower

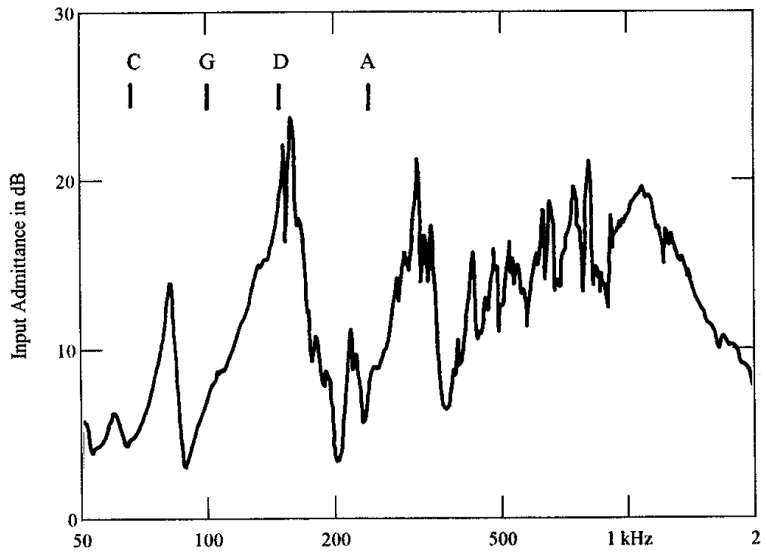


Figure 3.8: A cello response curve showing the input admittance (velocity amplitude per unit driving force) as a function of excitation frequency. Force was applied at the bridge in the bowing direction. The fundamental frequencies of the open strings are marked. (From Richardson, 1999)

pane). In cellos, this mode occurs typically around 90-100 Hz, close to the frequency of the open G string (98 Hz). Two other air modes, A_1 and A_2 (not indicated in Figure 3.9), in which the air in the cavity interacts strongly with the top and back plates, appear at around 200 Hz and 300 Hz respectively (Rossing, 2010).

In regard to the body modes, for example, a bending mode of the entire cello, B_1, has been observed at 57 Hz and is thought to contribute to the "feel" of the instrument, though it radiates very little sound. A rather symmetrical mode, C_1, was found to be not a good radiator either. In contrast, two important radiating resonances, C_2 and C_3 occur around 140 Hz and 220 Hz respectively (*ibid.*). Other reported frequencies range from 132 to 185 Hz for C_2 and from 185 to 219 Hz for the C_3 mode. The C_2 mode is also labelled as T_1 to indicate the contribution of strong top plate motions. Another peak in the input admittance curves near 195 Hz is the C_4 mode, which, although prominent, does not radiate very well (*ibid.*). Note, that as the cello modes are designated the same labels as

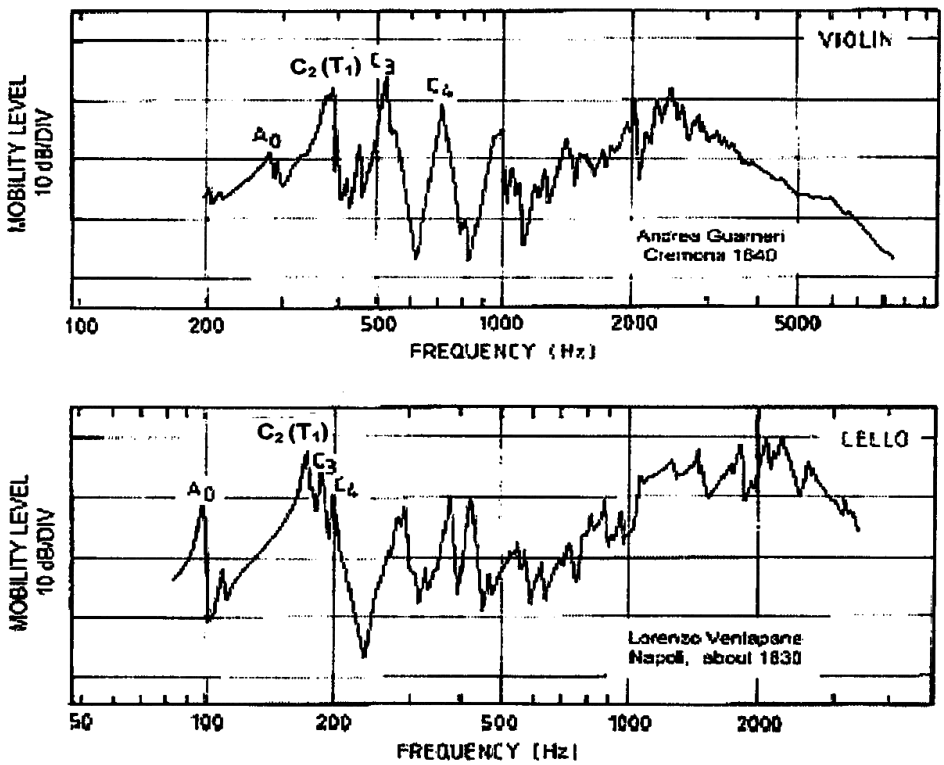


Figure 3.9: Input admittance curves of high quality violin (top) and cello (bottom). (From Askenfelt, 1982, as cited by Rossing (2010))

the respective modes of the violin, it gives the C_4 mode a lower frequency than the C_3 mode and may cause confusions with the modes' labels in Figure 3.9. In addition to the peaks related to particular structural resonances (normal modes of vibration), there is also a formant-like region observed between 800-1200 Hz, the so-called *bridge hill*, the shape of which acts as one indicator of the acoustical quality of the instrument (*ibid.*).

From the examination of input admittances of 24 violins of different qualities, Alonso Moral and Jansson (1982) found the A_0, T_1, C_3, and C_4 modes to be the most salient low-frequency modes of the violin. Instruments with the highest quality scores tended to have uniformly high levels of admittance for these modes as well as a rapid increase in admittance from 1.4 to 3 kHz. No such investigation has been reported on the cello, however, Askenfelt (1982), in his study on eigenmodes and tone quality of the double bass, compared input

Table 3.1: The cello normal modes and their frequencies compared to modal frequencies of a violin. Alternative labelling of the modes is given in parantheses. (Adapted from Rossing, 2010).

Mode	**Freq. (Hz)**	**Ratio to violin**
B_1 (C_1)	57	0.33
A_0	102	0.36
C_2 (T_1)	144	0.32
C_1 (C_2)	170	0.42
C_4	195	0.29
A_1	203	0.43
C_3	219	0.39
A_3	277	0.25
A_2	302	0.37

admittance curves of a high quality violin and cello (see Figure 3.9). As in the case of the violin, prominent peaks in the lower frequency range of the cello curve, corresponding to the four major resonances, were evident, followed by a marked rise in admittance starting around 1 kHz. Unlike the violin, the cello's T_1, C_3, and C_4 modes clustered together forming a second dominant peak after a relatively pronounced peak of the air resonance A_0.

It is worth noting that, while research on body vibrations of violins has been carried out over the last 150 years and improved greatly with advances in optical holography and digital computers, relatively few studies have been undertaken on the body vibrations of cellos (e.g. Firth, 1974; Langhoff, 1995; Bynum and Rossing, 1997). For example, Rossing et al. (1998), who compared normal modes of violins and cellos, also found the modes in a cello to be quite similar to the corresponding modes of a violin, although shifted in frequency. Modal frequencies in a cello occurred at 0.25 to 0.43 times the corresponding mode frequencies in a violin (see Table 3.1).

Acting as a transmission element between a vibrating string and an instrument body, the bridge plays a crucial role in the sound generation process and resulting tone quality of the instrument. In particular, the bridge's main function is to transform a horizontal force from the string to a couple of vertical

forces at the bridge feet (Jansson, 2002). Similarly to the instrument body, the bridge has its unique resonances, for example, at least two significant in-plane modes in case of the violin and three in case of the cello or the double bass. The first two cello resonances occur around 1 kHz and 2 kHz respectively (Richardson, 1999).

The influence of the bridge resonances can be seen on the input admittance curve. Askenfelt (1982), who compared admittance curves of high quality violins, cellos and double basses, concluded that the observed steep slope of the curve at frequencies above the major modal peaks derives from the principal bridge resonance. Jansson (2002) also reported that a boast in higher frequencies is mainly due to the bridge contribution, exhibited in the admittance curve as the bridge hill.

In bridge making, every single element starting from the choice of wood, and shape to the precise details of thickness, height, and overall proportions is of a key importance, since a minute change to the bridge can have dramatic consequences to the tone quality of an instrument (Rossing, 2010). Therefore, it is no wonder that the bridge has been always a subject of special attention from instrument makers and string players.

To this point, the discussion about the vibrational characteristics of the cello body was centred around vibrational effects of component parts such as the top and back plates, the bridge and the air cavity. Other parts such as the neck and fingerboard or, hidden inside the body, the soundpost and bass bar, however structurally important, contribute to the acoustical output of the instrument to a much lesser extent. In addition, the influence of the glues and especially the varnish on the final sound quality has been long debated with all eyes turned on the world's most valuable string instruments made by Antonio Stradivari. His violins and cellos, famously regarded for their unique timbre, have inspired generations of researchers hoping to unlock their acoustical secrets.

3.2.3 The bow

There is one more crucial element to the already complicated acoustics of the cello, that is, the bow. It is a generally accepted opinion that a bow acts as an extension of a string player's right hand and that in the hand of a skilful musician it becomes a powerful tool of musical expression. Askenfelt (1992) suggested that the quality of a bow can be assessed in view of: (1) playing properties, the way the bow can be controlled by a player, and (2) tonal qualities, the influence of the bow on the tone quality, and that the two quality aspects can be effectively defined by the distributions of mass and stiffness along the bow stick. He further proposed to characterise the playing properties in terms of parameters such as "the position of the centre of gravity, the centre of percussion (with respect to an axis through the frog), and resistance to bending for a well defined load", while the tonal properties seemed to be related to the normal modes of the bow including transverse vibrations of the bow stick (bending modes), and longitudinal resonances in the bow hair. In a series of experiments, Askenfelt explored the normal modes of the bow stick and assembled bow across a set of seven violin bows ranging from poor to excellent quality. Mode frequencies and damping ratios were compared to establish the correlation between the modal properties of each bow and its quality rating.

It was found that in the freely suspended violin bow stick (without the frog and the bow hair) around 12 pronounced transverse modes occur in the frequency region up to 2 kHz with approximate frequencies at 60, 160, 300, 500, 750, 1000, 1300, and 1700 Hz for the eight lowest modes. In comparison, the only empirical data about modal frequencies of cello bows comes from Schumacher (1975)'s early study, who observed that the cello bow modes were shifted in frequency about 30% in respect to those of the violin.

The obtained damping ratios for the modes of the free violin bow stick ranged from 0.2 to 0.6% (percentage of critical damping) with a slight increase with mode frequency. For the assembled bow, the mode frequencies decreased by 1-7% while the damping ratios doubled. An additional mode was found in

the assembled bow, identified as the lowest transversal mode of the bow hair, with the frequency within 60-75 Hz for normal bow hair tension, which coupled to the lowest mode of the bow stick at about 60 Hz. When the bow hair was rested on the string, a new "bouncing" mode occurred with a frequency dropping from 30 Hz to 6 Hz for the resting point at the tip and at the frog respectively.

Comparing tonal quality ratings with respect to acoustical properties of the bows under investigation, Askenfelt reported that no clear differences in the mode frequencies were found between bows of good and poor quality. In contrast, the measured damping ratios suggested that good bows have lower damping below 1 kHz.

3.2.4 Cello sound spectra

The acoustical output from the instrument, i.e. the sound we hear, is the result of the sound input being modified by the acoustic properties of the instrument itself (Howard and Angus, 2009). In case of the cello, vibrations of a bowed string, or plucked if played *pizzicato*, are convolved with combined structural resonances of the entire body and the bow, and radiated via the surrounding air medium. The process is "influenced by shadowing effects from the player" (Woodhouse, 1997) which include the way the string is excited, i.e. bowing characteristics (see Section 3.3.2) and a damping effect of the cellist's body in a normal playing position.

The output spectrum, i.e. the spectrum of the radiated sound, can be measured in terms of sound pressure captured at the microphone position, per-unit-force applied at the bridge, so-called *radiativity* (Rossing, 2010). The resulting frequency response depends on how the bridge is excited (using different driving points and different directions), the method of excitation including bowing machines and electromagnetic bridge drivers, and the position of the microphone/s. Figure 3.10 shows an example of the averaged magnitude spectrum of a good quality, modern violin measured in an anechoic chamber at twelve microphone positions, spaced evenly around the instrument, in response to an

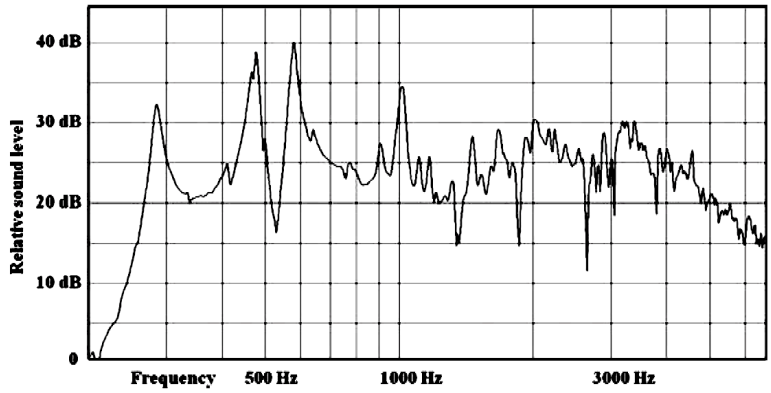


Figure 3.10: Average spectrum of a modern violin for 12 microphone positions spaced at 30° intervals around the instrument. (From Rossing, 2010)

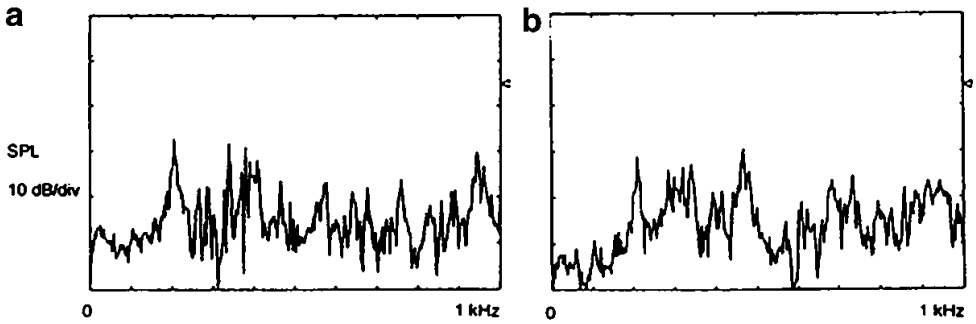


Figure 3.11: Room-averaged sound spectra of a cello: (a) freely supported on rubber bands; (b) hand-held in playing position. (From Bynum and Rossing, 1997, as cited in Rossing (2010))

impact hammer tapping the bass corner of the bridge (*ibid.*). No systematic attempts have been made to measure and compare radiativity of different quality cellos. A more general example of cello spectra, both freely vibrating and held by a cellist, is given by Bynum and Rossing (1997) (see Figure 3.11).

The directional characteristic of the radiated sound depends primarily on the frequency component. Meyer (2009) investigated directional radiation patterns of all orchestral instruments including cello. As illustrated in Figure 3.12, the cello tends to radiate more broadly toward both the sides and front at the lower frequencies, while at the higher frequencies it exhibits much more directionality. The indicated radiation areas are within 3 dB of the sound maximum value

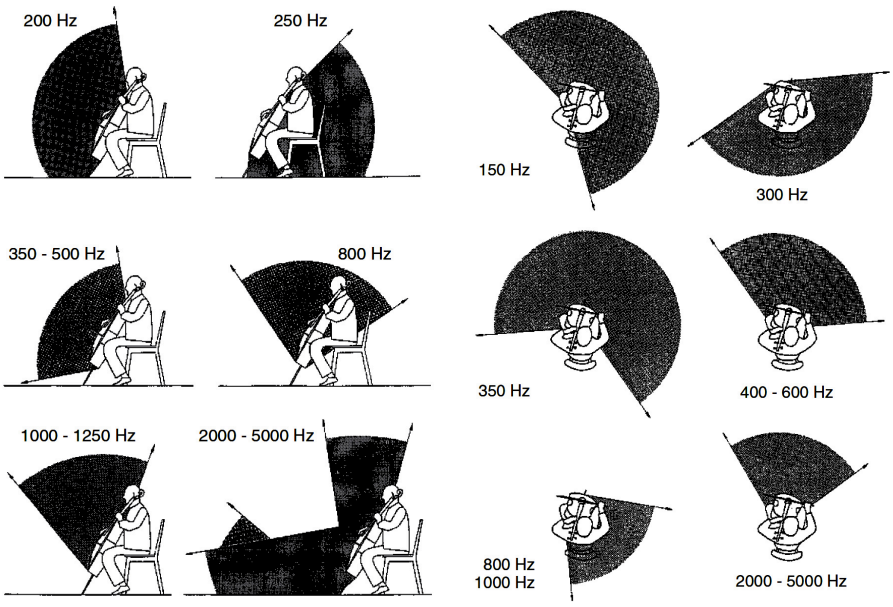


Figure 3.12: Principal radiation directions of a cello at different frequencies: (left) in the vertical plane; (right) in the horizontal plane. (From Meyer, 2009)

averaged over the measured frequency range. Meyer also observed that below the air resonance (roughly 110 Hz) the radiated power level drops at a rate of 6 dB/octave and causes the fundamentals of the C string to have lower intensities. In the frequency region between about 200 and 2000 Hz, the power level fluctuates by 5 dB from the steady 6 dB drop/octave due to structural resonances of the cello body. Above approximately 2000 Hz the radiation behaviours vary with the cello registers. The spectra of the lower or middle registers drop with a slope of about 16 dB/octave while in the upper register this drop decreases to a value of 10 dB/octave.

From the point of view of the listener, the complex radiation patterns and characteristic declines in the power level across frequency regions result in that the perceived timbre of cello sound very much depends on the listener's position in the audience (Meyer, 2009) .

3.3 Playing technique

It is beyond the scope of this thesis to detail all intricate aspects of cello playing technique which gradually evolved throughout the mid to late 18th and 19th centuries in parallel with developments in the instrument and the bow. An interested reader is referred to early 20th century treatises by Straeten (1905); Krall (1913); Alexanian (1922) to start with, followed by more recent studies and handbooks on modern cello technique by Eisenberg (1957); Pleeth (1982); Mantel (1995), and Potter (1996), for example. Since the focus of this study is placed on timbral characteristics of a cello player as seen from perceptual, acoustical, and, most importantly, gestural perspectives, only those technique elements related to the process of controlling tone quality are discussed.

3.3.1 Left hand technique

Left hand technique is responsible for changing the pitch by controlling the length of the string being played, i.e. stopping the string closer to the bridge results in higher-pitched sound, because the vibrating string length has been shortened. To achieve a tone of a clear pitch from the very start of the note, requires left hand fingers to be fully coordinated with the bowing movements of the right hand. In particular, it involves so-called finger articulation or *percussion* as introduced by Pablo Casals. A properly "articulated" finger, whether it hits the fingerboard to stop the string at a higher pitch or is lifted off the string to lower the pitch, allows the current string vibrations to be cut off abruptly, and with coordinated actions of the bow to excite the string again practically at the same moment. This precise action of the left hand is crucial for obtaining the Helmholtz motion in the bowed string.

Important components of finger articulation include finger weight, finger pressing force, and finger dropping and lifting speed (Suchecki, 1982). As finger weight is determined by the player's physical characteristics only finger force and speed are controlled and adjusted accordingly by the player. Optimally, the entire pressing force of the finger should come from the combined weights of

forearm and upper arm which simply rest on the fingertip so that no extra force is required. In that way, the finger acts as a support for the rest of the hand and changing a finger means changing the supporting point, which results in almost effortless movement along the fingerboard while switching between pitches. The left fingers' speed, on the other hand, relies on individual motor skills and agility of the left hand combined with strength and flexibility developed via dedicated exercises over the course of the technique forming period.

Most would agree that a beautiful cello tone is undeniably associated with playing with *vibrato*. "The vibrato is one of the most active factors of the 'fullness' of tone-color" says Straeten in his treatise on cello playing. Indeed, the use of varied vibrato introduces a new wealth of colours to the instrument timbre and a well-developed vibrato technique is considered an essential element of a modern cellist's skill.

Potter (1996) describes the cello vibrato as "a bouncing, somewhat rotary movement back and forth, parallel to the fingerboard, produced by the left forearm from the elbow, with the wrist acting only as a part of the whole vibrato unit (unlike the violin vibrato)". Two vibrato parameters, which remain in a close relationship, are controlled by a player: the amplitude (or extent) of the movement and its speed. For example, too large amplitude forces to lower the speed and the tone becomes moaning-like. On the other hand, too high speed decreases the amplitude to such an extent that the resulting tone sounds feverish or nervous (Suchecki, 1982). The vibrato amplitude is naturally reduced and the speed increased accordingly in higher positions, where the string in play is shortened. In respect to sound intensity, loud playing requires a wider vibrato than does playing soft tones. The speed of the vibrato, under certain constraints related to the amplitude, is generally more a matter of personal taste and temperament (Potter, 1996).

3.3.2 Bowing technique: controlling tone quality

Bowing technique or right hand technique is a key technical component on bowed string instruments, allowing sound production by controlling the interaction of the bow hair and the string. It is also the major determinant of expressiveness, similarly to the breath of a wind instrument player or voice emission of a singer. Musical elements such as tempo, dynamics, articulation and tone quality depend directly on the bowing technique used.

Bowing could generally be described as a complex combination of the upper arm, forearm and palm movements in reference to the bowing point on the string, with instantaneously adjusted relative position, speed, and centre of gravity (weight) of the whole hand. The entire sequence of minute displacements of the right hand parts is devoted to control what happens at the point of contact between the string and the bow hair in every single bow stroke. So what actually happens there?

The bow is drawn perpendicularly across the string at a certain distance from the bridge, i.e. at the bowing point. Drawing the bow is performed at a certain velocity, i.e. bow velocity, and with a certain pressure against the string, the so-called bow force (see Figure 3.13). The right combination of the bow force, bow speed and bowing point triggers a "healthy" Helmholtz motion in the bowed string, resulting in a clear and rich tone, the ultimate target of any string player. It needs to be mentioned here that there are also other bowing controls such as bow tilt, bow inclination, and bow skewness (see Figure 3.14), complementary to the major bowing parameters, the role of which in tone production has not been yet thoroughly investigated.

However, there are several physical and mechanical conditions to what can be the right combination of the bowing parameters. First of all, when the bow is pulled in the *down bow* direction, i.e. from the frog towards the tip, the bowing point, from a point supporting the hand weight (which near the frog is in such excess that some of that weight needs to be virtually lifted off the string), gradually, and as the bows moves, becomes a lever point for the hand

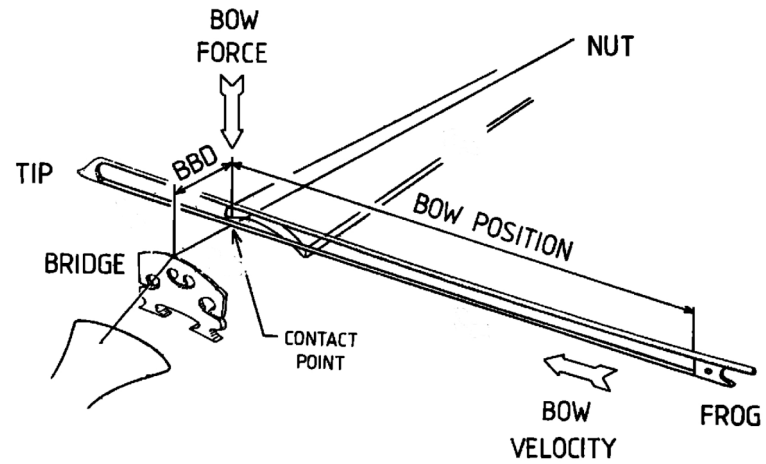


Figure 3.13: Physical bowing parameters controlled by a violin player: bow velocity, bow position, bow force (the force pressing the bow against the string), and bow-bridge distance (as measured from bowing or contact point). (From Askenfelt, 1989)

(which in turn needs to add an extra weight when it reaches the tip); all that to maintain a stable, uniform quality tone throughout the bow stroke. One can easily guess that exactly the opposite process takes place when the bow is pulled *up bow*, i.e. from the tip towards the frog. That seemingly basic skill

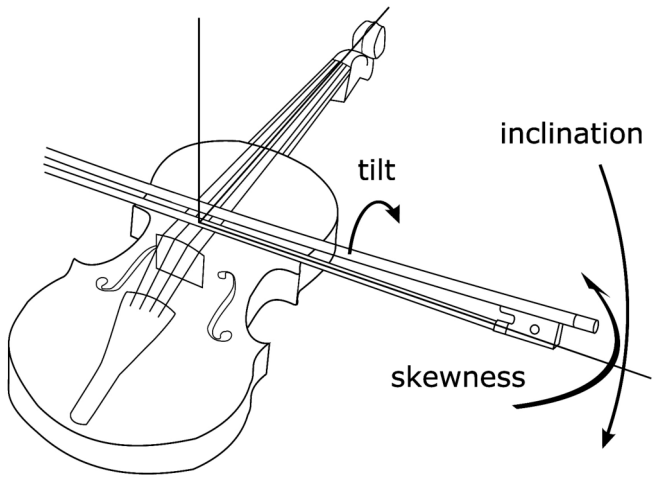


Figure 3.14: Complementary bowing controls on the violin: bow tilt, bow inclination, and bow skewness. (From Schoonderwaldt and Demoucron, 2009)

requires from a beginner player at least a couple of months of diligent practising to master it. Interestingly, when bow force for a whole-bow stroke is measured and plotted against time (see Chapter 4), it is shown as a sort of plateau with some increase and decrease at the bow endings but with no sign of that hand weight balancing which smoothly runs in the background.

Secondly, it is generally recommended that the bowing point should be approximately midway between the end of the fingerboard and the bridge for the lowest three positions and be moved closer to the bridge when playing in higher positions. Suchecki (1982) gives slightly more precise guidelines suggesting that for good quality tone the optimal bowing point lies at $L/9$ or $L/10$ from the bridge, where L denotes the length of the string in vibration. He argues that in the resulting tone the proportion between lower and higher harmonics and their amplitudes is well balanced. Thirdly, to generate a tone of good quality, bow velocity or bow speed needs to be proportional to the bow pressure against the string, i.e. the greater the pressure (bow force), the higher the bow velocity (Suchecki, 1982).

These practical recommendations have been passed down through generations of string players and have stemmed from the accumulated performing practice and teaching experience. The player's bowing "arsenal", i.e. bow force, bow velocity, and bow-bridge distance (bowing point), further referred to as bowing parameters, and their relationship in respect to the resulting string spectra have been more systematically studied relatively recently, drawing on the earlier pioneering works of Helmholtz (1877) and Raman (1918). For example, Schelleng (1973)'s results showed that the string player "can not select the bowing parameters freely" and that "for a specific value of the bow velocity the 'bow pressure' must be selected within a permitted working range" in relation to bow-bridge distance to obtain a proper tone (Jansson, 2002). The empirically obtained bow force limits were elegantly presented in the form of the so-called Schelleng diagram (see Figure 3.15), which has been referred to often in research on bowing parameters ever since. The diagram indicates the

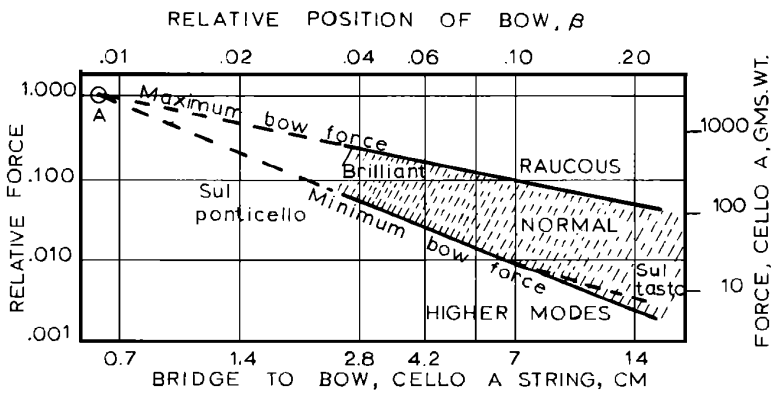


Figure 3.15: Typical normal and abnormal playing conditions in the violin family related to bow force and bow position at constant bow velocity for sustained tones. A second set of coordinates refers to a cello A string bowed at 20 cm/sec. (From Schelleng, 1973)

force range required for maintaining the Helmholtz motion as function of the bow's position on the vibrating string for a given bow speed, with a reference to perceived timbre attributes. Further discussion on the bowing parameters' interrelations and also on methods for their acquisition and measurement is provided in Chapter 4.

From the cello player's perspective nothing better summarises fundamentals of good tone production than Potter (1996)'s *Some remarks on tone quality*:

1. To maintain a consistently smooth and even quality of tone, keep the contact point of the bow on the string steady (halfway between the bridge and end of the fingerboard), without letting it shift as the bow is drawn. The bow should travel in a line parallel to the bridge.
2. Develop good "bow distribution" and management by dividing the bow strokes properly in relation to the particular note values and tempo involved.
3. Don't change the bow speed during any one stroke unless a change of dynamic strength is called for.
4. A thin, dull, or raspy quality of tone may be due to any one or

a combination of the following: (1) Drawing the bow too close, or too far, over the end of the fingerboard. (2) Not stopping the string with sufficient firmness of the left fingers. (3) Using too great a bow speed in relation to the bow pressure employed.

5. (...) explore the advantages of playing with the contact point of the bow on the string closer to the bridge. This enriches the tonal quality, and heightens the sonority, when playing long and sustained tones in *forte*, due to the greater number of overtones available near the bridge. In order to achieve tonal clarity and articulation when playing in higher positions (particularly on the upper strings) it is especially necessary to bow closer to the bridge because of the shortened string length. When the bow is travelling more rapidly, however, one cannot bow as close to the bridge as when playing long sustained tones.

6. (...) explore the tonal resources available when bowing near, or over the end of the fingerboard. There, other factors being equal (bow speed, style, tone quality involved), a lighter and more delicate tone, of much less intensity, can be produced for playing soft passages (*sur la touche* or *sul tasto*).

The sensitive player is ever mindful that control and variety of tone quality and colour are very important areas of musical study and achievement. (p. 63)

And yet another statement from one of the great violin pedagogues of the early 20th century, Carl Flesch (1939), who expressed his concern with tone quality:

> "The technique of tone production represents the noblest portion of the collective technique of violin playing. Pure tone is the most valid interpreter of emotions. Yet it should never cease to be only a means..."

3.4 Cello timbre in perception

The timbre of the cello is regarded as one of the most beautiful amongst the whole orchestra set and its quality is often compared to the human voice. For example, the famous Russian opera singer Feodor Ivanovich Shalyapin who lived at the turn of the 19th and 20th centuries and himself possessed a deep and expressive bass voice was used to say that "one should sing as sings the cello".

From the psychoacoustic perspective, cello timbre has been relatively little explored in comparison with its smaller "sibling", the violin. The very few perceptual studies which actually used cello tone samples in the experiments (Grey, 1977; Grey and Gordon, 1978; Gordon and Grey, 1978; Iverson and Krumhansl, 1993, see Section 2.3.4) provided a preliminary insight into the perceptual positioning of the cello amongst the other orchestral instruments which was then interpreted in terms of spectro-temporal characteristics.

For example, in all three studies by Grey and Gordon, three cello tones played *sul ponticello*, normal bowing, and muted *sul tasto* (labelled S1, S2, and S3 respectively), represented the bowed string family. The resulting timbre space is shown in Figure 3.16. As indicated by respective psychophysical correlates of the timbre dimensions (revised by McAdams et al., 2006), a cello tone may have the properties of narrow spectral bandwidth and a concentration of low-frequency energy when played *sul tasto* (S3) and change towards a much wider spectral bandwidth with less energy concentrated in the lowest harmonics for tones played *sul ponticello* (S1). At the same time, the spectra of the cello tones strongly fluctuate over time and the upper harmonics seem to be rather temporally independent in their patterns of attacks and/or releases (higher spectral flux). Cello tones also possess characteristic high-frequency, low-amplitude, most often inharmonic energy which precedes the full attack of the note (higher attack centroid). One might find the above described spectro-temporal properties not fully representative for the entire cello timbre since they were derived from just three tones of the same pitch, resynthesised with considerable simplifications with respect to the originals. It would be also quite interesting to

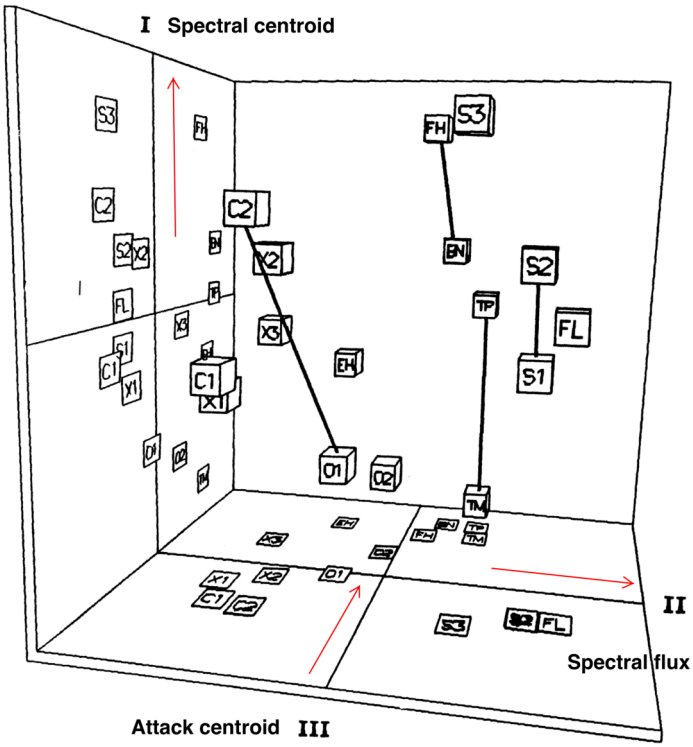


Figure 3.16: Three-dimensional timbre space for 16 recorded instrument tones. Abbreviations for stimulus points: O1, O2 = oboes; C1, C2 = clarinets; X1, X2, X3 = saxophones; EH = English horn; FH = French horn; BN = bassoon; TP = trumpet; TM = trombone; FL = flute; S1, S2, S3 = cello tones. (Adapted from Gordon and Grey, 1978)

compare the obtained psychoacoustic representations of cello tones with that of other bowed string instruments if they had been present in the stimuli.

For comparison, Iverson and Krumhansl (1993) used both violin and cello samples in their experiments. The resulting MDS solution for similarity judgements on the complete tones is presented in Figure 3.17. The vertical dimension, highly correlated with spectral centroid (centroid frequency), corresponds to the perceived *brightness* of the sounds, similarly to the vertical dimension of the Gordon and Grey's space. Interestingly, according to the position along this axis, the cello sounded brighter than the violin (respective centroid frequencies 2853 and 2035 Hz), which might be quite surprising. One would rather expect the opposite. In this case, the explanation may come from the fact that the

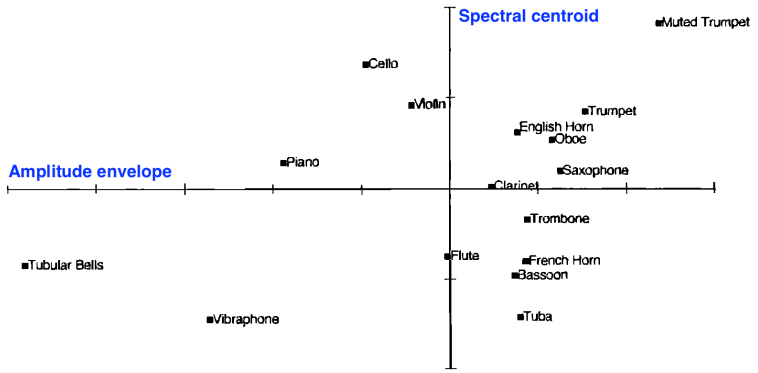


Figure 3.17: Two-dimensional timbre space for 16 complete tones, i.e. onset + remainder. (Adapted from Iverson and Krumhansl, 1993)

pitch of each tone in the stimulus set was chosen to be C4 (262 Hz), and if the tones of violin and cello were recorded in their default positions, then the violin C4 was captured on the lowest string G and the cello C4 on the highest string A. The difference in timbre and especially in *brightness* between these two strings is substantial and may even cause one to confuse the two instruments with each other. This is just an example of how important in that kind of experimental scenario is the right choice of the sound stimuli so that they can be truly representative of each instrument class under investigation while maintaining a reasonable dataset size. On the other hand, such a compromise can be difficult to achieve considering that only the use of isolated tones provides a researcher with ability to experimentally control the tones' properties other than timbre.

In performance practice the entire pitch range of the instrument is typically divided into low, middle and high registers to account for different frequency regions. Such a seemingly natural partition, however, does not reflect timbre differences between the pitches. Specifically for the bowed string instruments, it is possible to play the same note on two, three or even four different strings, thus allowing a performer to choose from various timbral and aesthetic characteristics. Suchecki (1982) proposed an extra grouping according to psychoacoustic properties of the tones. On cello, for example, *dark* tones are generally obtained on the two lowest strings C and G and partially on D, while a *bright* sound is

typical for the string A and for the higher positions of the strings D and G. He gave more detailed verbal descriptions of tones in each of the registers. In general, tones of the low register sound *free* and *mellow* and easily respond to even minute changes in dynamics and vibrato. These features become less pronounced in the higher positions of the string C due to the shortened string length. The middle register tones also resonate well but their timbre is *dull* and the dynamic range is suppressed. This register provides a player with the richest palette of timbral nuances since its tones are the most responsive to vibrato and timbre manipulations. Tones of the high register, on the other hand, are generally *bright* and have the largest dynamic range. They are also quite easy to manipulate in terms of timbre and vibrato changes. If played in the higher positions of the strings D and G they become *dull* and less sounding. In addition to the three main pitch regions, Suchecki distinguishes also the highest register the tones of which resonate shorter and require intense vibrato in *p-mp* dynamics. While the dynamic range is still quite wide, at the same time timbre nuances available to a player are limited.

3.5 Summary

In this chapter, fundamentals of cello acoustics and playing technique have been outlined with the emphasis on the facets most related to the sound quality and timbre of the cello tone. The resulting tone quality is primarily determined by the very complex acoustical characteristics of the instrument itself including structural resonances of the body, the bridge, and the bow, the choice of strings and rosin. However, in search of a beautiful tone, a cello player can choose from a multiplicity of possible combinations of bowing parameters those most corresponding to his physics and technical skills to control and manipulate the instrument's acoustical output.

Chapter 4

Acquisition and analysis of bowing gestures

4.1 Introduction

The mechanics behind an expressive or virtuosic performance has always intrigued scientists keen to unlock the secret of a virtuoso's beautiful tone or their phenomenal playing technique. It is not then surprising that the first attempts of capturing empirical data from expressive performance and particularly specific performance parameters in piano playing date back to the end of 19th century (Binet and Courtier, 1895).

An interest in the mechanics of bowing, first manifested in pioneering works of Helmholtz (1877), triggered a series of scientific explorations of the phenomenon since the dawn of the 20th century (e.g. Raman, 1918, 1920). First attempts at measuring physical variables of the bowing process were made using bowing machines which allowed researchers to systematically examine variables' ranges and their interrelations (see Section 4.2). Results on mechanically bowed violins revealed the existence of physical limits on the combinations of bowing parameters which can produce steady-state Helmholtz motion, hence a good quality tone. Once the theoretical relations between the main bowing parameters were established, a natural step further was to validate them in normal playing conditions. Various dedicated equipment has been built to capture bowing gestures from string players and systematic analyses of bowing motion data

followed (see Section 4.3.1). Advances in computer technology and electronics provides musicians and composers with practically unlimited resources for experimenting with new musical instruments, interfaces, and controllers in the search for alternative forms of musical expression. The prospect of novel ways of interacting with such traditional instruments as violin or cello via gesture-informed digital processes has opened a new area for scientific explorations. Applications of bowing motion capture for gesture-controlled musical interfaces and augmented bowed string instruments or virtual environments for learning practice, as well as for gesture-based sound synthesis are discussed in Section 4.3.2.

To systematise the terminology used in the next sections, bowing controls or bowing parameters are defined as follows: **bow position** (x_B): the current transverse position of the bow-string contact point in relation to the frog; **bow velocity** (v_B): the velocity of the bow transverse to the strings; **bow force** (F_B): the normal force of the bow pressing against the string; **relative bow-bridge distance** (β): the distance along the string from the bridge to the bow-string contact point (x_B), relative to the effective length of the string. i.e. the length of the string in vibration.

4.2 Bowing machine studies

First experiments employing a bowing machine for investigating the mechanical conditions necessary for obtaining a steady violin tone were conducted by Raman (1920). He observed that bow speed is the main resource of the violinist to alter the intensity of tone and that with the increase of bow speed, while the other factors remain the same, bow force has to increase too. He also found that the minimum bowing force varies in inverse proportion to the square of bow-bridge distance.

Raman's theoretical and empirical foundations of minimum bow force were extended by Schelleng (1973) who formulated the upper bow force limit and

defined playable regions (known as the Schelleng diagram) by systematic measurements of bow force versus bow-bridge distance at fixed bow velocities using a bowing machine (see Figure 3.15 in Chapter 3). In the simplest terms, the minimum and maximum bow force limits were found to be proportional to v_B/β^2 and v_B/β respectively. Bow force limits interpreted from the viewpoint of the player mean that when bow force is too small a "surface sound" is produced due to bow-string friction being insufficient to keep hold of the string while the Helmholtz corner is travelling between bow and nut (McIntyre and Woodhouse, 1978). This in turn causes two or more slips to occur per cycle (instead of one) preventing the fundamental vibration to be developed properly. The resulting tone contains mainly higher partials. On the other hand, too large a bow force causes bow-string friction to interfere to such an extent that the Helmholtz corner arriving from the nut cannot pass and the slip phase of the string does not occur. Once the Helmholtz motion fails, a "raucous scratching" is the result (Schelleng, 1973).

In terms of controlling the sound volume, Schelleng, following Raman's formulations, stated that the output sound pressure is proportional to v_B/β. He also pointed out that bow force has little effect on volume and acts "primarily as the catalytic agent that makes possible a correct reaction between bow speed and bow position". Both relationships were later confirmed empirically in Bradley (1976)'s experiments with a bowing machine.

Subsequent evaluations of Schelleng's model include measurements of the maximum bow force for unstopped notes on a variety of violin strings by Schumacher (1994) and bowing machine experiments on a cello D string at a single velocity by Galluzzo (2003). Although they introduced more realistic friction functions into the model, both studies, in general reproduced Schelleng's findings. Galluzzo's methodical experiments deserve more attention, as they were conducted on a full-sized cello bowed with a rigid point-contacting perspex rod, followed by tests with the use of a real bow for comparison. Generally, bow force limits obtained with the bow strongly resembled those measured for the

perspex rod, suggesting that the observed bow force values were not affected by the finite width and compliance of bow hair.

Schoonderwaldt et al. (2007, 2008) proposed to reformulate Schelleng's upper and lower bow-force limits to account for variations in the friction coefficient δ (the difference between the coefficients of static and dynamic friction), especially at small values of v_B and large values of β. According to the modified model, both the maximum and minimum bow force converge to a finite minimum when bow velocity approaches zero.

Schelleng's and the modified bow force limits were validated in a systematic investigation by means of a bowing machine using a normal violin bow and standard violin strings. Most common types of string motion which occur depending on the bowing parameter combination were classified to empirically determine Shelleng diagrams. The observed string motion types included: (1) Helmholtz motion, characterised by one slip and stick phase per fundamental period; (2) multiple slipping, due to insufficient bow force; (3) raucous motion, when Helmholtz motion is broken (no slip phase) due the bow force excess; (4) anomalous low frequency (ALF), a special condition when a too large bow force prevents the slip phase and the bow hair becomes a quasi-termination point for the string which is forced to vibrate with fundamental frequencies lower than the natural first mode frequency (Hanson et al., 1994); and (5) S-motion, characterised by a single slip phase per fundamental period and a strong presence of secondary waves caused by reflections between the bow and the bridge (Schoonderwaldt, 2009b).

The results based on wide ranges of bow force and bow-bridge distance measured at four bow velocities suggested a good agreement between the empirically obtained Schelleng diagrams and Schelleng's definitions of the playable region. The so-called Schelleng's triangle, corresponding to Helmholtz motion, was surrounded from the top by regions of raucous motion and anomalous low frequencies at higher bow forces and from the bottom by a region of multiple slipping at lower forces (Schoonderwaldt et al., 2008). However, in terms of

defining the bow-force limits, a better fit with the empirical data was found for the modified upper limit which takes into account the friction coefficient δ varying with bow velocity. More importantly, while the fitted lower bow force limit did not deviate significantly from being proportional to $1/\beta^2$, it showed no dependence on bow velocity, in contrast to Schelleng's estimation. It was also demonstrated that the effect of string damping on minimum bow force (for example, by stopping the string with a finger) was much stronger than theoretically inversely proportional, leading the theoretical model to an underestimation of the minimum bow force (almost one order of magnitude difference). In line with earlier studies (e.g. Woodhouse, 1993), Schoonderwaldt et al. (2008) also pointed out that Raman's and subsequently Schelleng's theoretical model of the minimum bow force did not account for phenomena such as Helmholtz corner rounding (Cremer, 1972, 1973) and ripple, which occur in a bowed string as a result of string bending stiffness and internal/external losses and the wave reflections between the bow and the bridge (or between the bow and the nut/finger) during the stick phase, respectively.

If established Helmholtz motion in a bowed string is necessary for generating sound of a good quality, a "perfect" pre-Helmholtz motion transient is crucial for obtaining a clean start of each tone. A"perfect" transient refers to establishing one stick-slip transition per period of the fundamental frequency as quickly as possible (Guettler, 1992). Bowing parameters such as acceleration, bow force and bow-bridge distance are primary controls during the attack phase.

McIntyre and Woodhouse (1979); Cremer (1982) and Guettler (1992) provided the first insights into starting transients of a bowed string by means of theoretical models and computer simulations. Further investigations by Guettler and Askenfelt (1995, 1997), involving bowing machine experiments, identified three possible characteristics of the pre-Helmholtz transient depending on the combination of bowing parameters. Besides an ideal attack with Helmholtz triggering from the very start, they observed attacks with *prolonged* periods (delayed triggering) or with a division of the period into two or several parts,

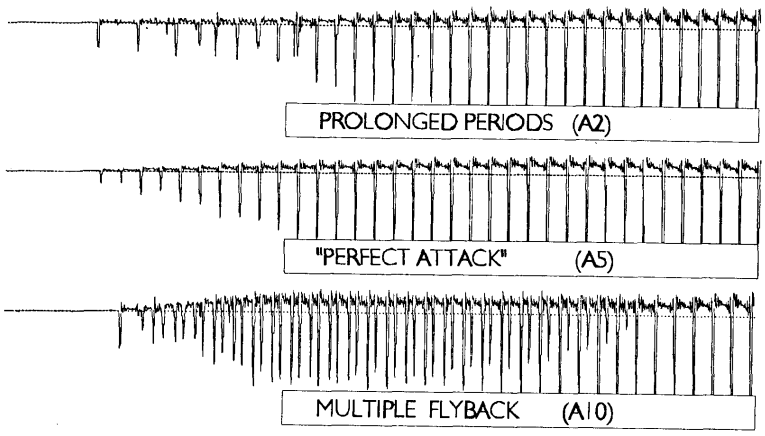


Figure 4.1: Three principal types of bowed string attacks. String velocity at the bowing point for prolonged (top), "perfect" (middle), and multiple flyback attacks (bottom). The violin open G string was played with a bowing machine using a normal bow. (From Guettler and Askenfelt, 1997)

so-called *multiple flyback* (see Figure 4.1). In terms of audible effects, the latter two were described as *choked/creaky* and *loose/slipping*, respectively (Guettler and Askenfelt, 1997). An additional perceptual study on a group of string players revealed that the allowed duration of transients in attacks categorised as acceptable is rather limited. For violin, the acceptance limits were 50 ms for prolonged periods and 90 ms for multiple flyback, corresponding to 10 and 18 nominal periods of the fundamental frequency (an open G string), respectively. Interestingly, the same approximate limits in ms were obtained for simulated viola attacks, however the number of nominal periods was reduced due to the lower fundamental frequency of the viola string. The results suggested that the acceptance limits if expressed in absolute terms, i.e. in ms, do not depend on the fundamental frequency. This implies that for lower ranged instruments such as viola, cello and double bass the number of nominal periods before reaching Helmholtz triggering has to be significantly smaller, thus, the range of available bowing parameter combinations securing a "perfect" attack is much narrower.

As one step further, the conditions for establishing Helmholtz motion during a tone onset were formulated by Guettler (2002) and examined by means of computer models with bow force and bow acceleration as main operands.

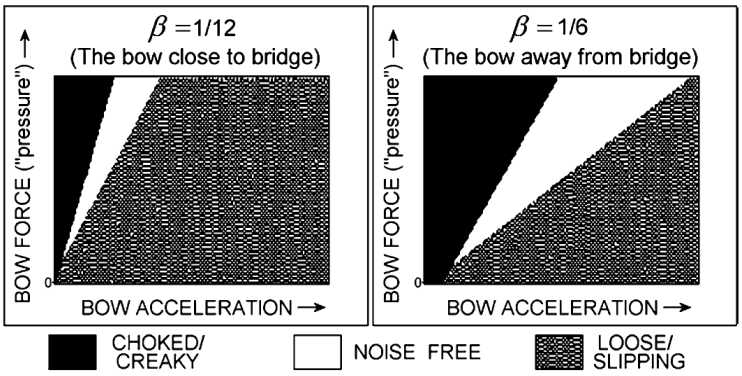


Figure 4.2: Relation between bow acceleration, bow force and sound quality. (From Guettler, 2010)

Computed parameter spaces described the relation between bow acceleration and bow force for a given bow-bridge distance in respect of three perceptual categories of sound produced. As Figure 4.2 indicates, for the bowing point closer to the bridge, the noise-free attack region is much smaller, and generally less acceleration is required for the same amount of force exerted. Guettler's simulations and resulting diagrams were validated by Galluzzo (2003) in a series of experiments on a cello, bowed mechanically with a perspex rod and with a normal bow. The results showed that, in the experimentally obtained so-called Guettler diagrams, areas with occurrences of Helmholtz motion were roughly of a wedge-like shape, and their size shrank and position shifted upwards as β decreased, similarly to the triangular patterns reported by Guettler. However, Galluzzo reported that "the change in the appearance of the Guettler diagram as β was decreased was not gradual", and, for particular β values, multiple fly-back motion or S-motion occurred where the normal Helmholtz triggering was expected.

It seems that Schelleng (1973) first related alterations of bow force with specific changes in sound spectra which are then perceived as shifts in tone colour, e.g. from *brilliant* toward soft sounding *sul tasto* as bow force decreases (Figure 3.15). Guettler et al. (2003) pointed out that it does not stem clearly from the Schelleng diagram whether producing brighter tone is the working

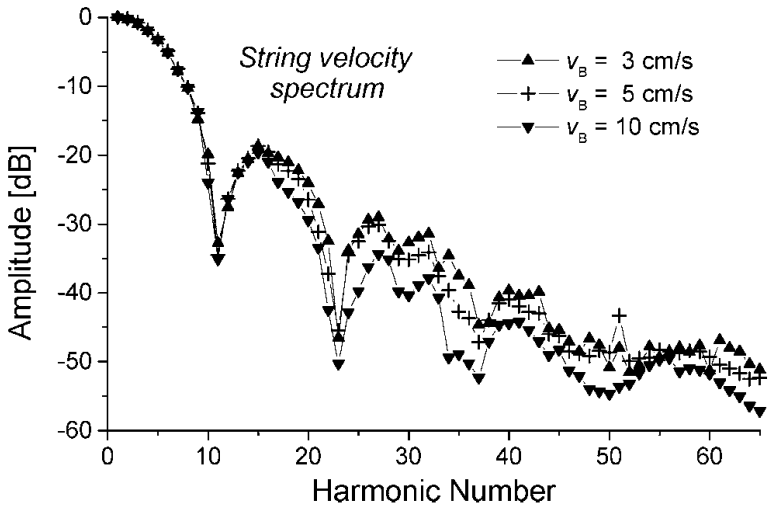


Figure 4.3: Spectrum of string velocity for the three bow speeds. The spectrum (normalized to 0 dB for the 1st harmonic) was averaged over several strokes with constant bowing parameters. (From Guettler et al., 2003)

of higher bow force or bowing closer to the bridge as the two parameters are coupled, while the common practice simply recommends to move towards the bridge in the quest for a more brilliant sound. In an attempt to validate this empirically, a series of computer simulations with the largely varying bowing point while bow force and bow velocity remained fixed was conducted. Guettler et al. (2003) found that the general spectral envelope of the force on the bridge showed no significant changes or trends across simulated conditions except for minor local deviations for node frequencies, in contradiction to what was commonly thought about the role of bow-bridge distance. Instead, results of an experiment on an open violin D string bowed with three velocities at fixed bow force and bowing point suggested that increasing the bow speed reduced the amplitudes of upper harmonics (from 16th to 65th) between 1.3 and 5.2 dB on average (Figure 4.3). Guettler et al. concluded that it is bow velocity rather than bow-bridge distance which influences the output spectrum most when bow force is kept constant. This holds especially at low speeds within the Helmholtz regime.

Other bowing components such as changing the width of the bow hair, in normal playing combined with tilting the bow, have been also found to affect the string spectrum. Extending the earlier work of Pitteroff and Woodhouse (1998), Schoonderwaldt et al. (2003) in experiments using a bowing machine showed that with the decreased bow hair width higher harmonics were boosted considerably. The observed gain in amplitudes above the 20th harmonic ranged from 3 to 6 dB. The effect was more pronounced for higher bow velocities and bow forces. A 45° tilt (leaning the bow stick towards the fingerboard) combined with the reduction of the bow hair width also gave a consistent boost of the higher partial amplitudes for both playing closer and further away from the bridge. Schoonderwaldt et al. noted that tilting evidently improved the quality of the attacks, which is in full agreement with musical practice especially for bowing near the frog. In fact, during the *down-bow* stroke, the bow stick remains tilted well until the middle of the bow length and is gradually lifted to its upright position (perpendicular to the string) towards the tip to help capturing the string with the full hair width. The reversed order of events takes place during the *up-bow* stroke. Considering the results of the companion study (Guettler et al., 2003), Schoonderwaldt et al. rightly concluded that "there might be a substantial combined effect on the spectral slope by bringing the bow closer to the bridge while simultaneously increasing the bowing force, lowering the bowing velocity, and adjusting tilting". Table 4.1 compiled by Guettler (2004, 2010) summarises these and earlier findings, outlining parameters that affect the string spectrum when playing within the Helmholtz regime.

The influence of the bowing parameters on the spectral content and pitch of the violin tone was further investigated by Schoonderwaldt (2009b). The spectral content was measured by means of spectral centroid, a timbre descriptor associated with perceptual *brightness* of sound, and strongly related to bow force, a primary contributor in sharpening the Helmholtz corner. In addition, the conditions for pitch flattening, anomalous low frequencies (see Figure 4.4) and other, higher types of string motion were examined. Pitch flattening was

Table 4.1: Overview of steady-state spectral effects when changing one bowing parameter at a time. (Adapted from Guettler, 2004, 2010)

Parameter value increased	**Effect on tone color (spectral profile)**
Bow force (F_B)	Increased sharpness/brilliance
Bow velocity (v_B)	Decreased sharpness/brilliance
Tilting of bow-hair ribbon with respect to the string (only if tilted toward the fingerboard)	Increased sharpness/brilliance (moderate effect only)
Width of bow-hair ribbon	Decreased sharpness/brilliance (moderate effect only), and increased noise due to partial slipping across the hair ribbon during stick intervals (particularly when bowing near the bridge)
Length of string (with constant bending stiffness and impedance but with the fundamental frequency decreasing)	Increased sharpness/brilliance (relative to the fundamental frequency)
Finger-pad damping	Decreased sharpness/brilliance
Relative bowing position (β)	Only local spectral deviations – no general trend except increased slipping noise due to the increased slipping intervals

first described by McIntyre et al. (1977) as an audible effect of the note going flat, typically by a small fraction of a semitone, as the bow force exceeds a given level, especially when playing with a low bow speed at a moderate distance from the bridge. The phenomenon was systematically explored in studies by McIntyre and Woodhouse (1979); Schumacher (1979); McIntyre et al. (1983); Boutilion (1991) and was found to be related to a hysteresis occurring along the sticking-slipping cycle due to the Helmholtz corner rounding.

In Schoonderwaldt's study, computed values of spectral centroid and pitch level were mapped onto empirically obtained Schelleng diagrams (based on a wide range of relative bow-bridge distance β and bow force F_B combinations performed at four bow velocities v_B on a bowing machine). As predicted, among the three bowing parameters, bow force was found to be the major determinant

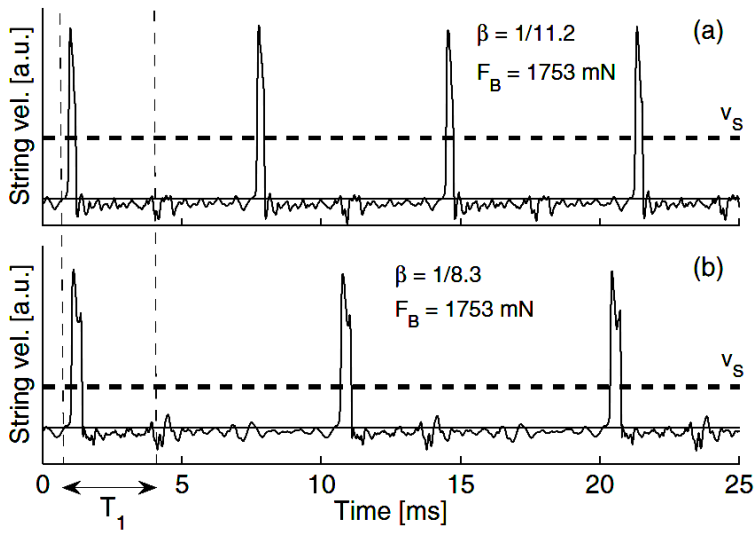


Figure 4.4: Two examples of anomalous low frequency (ALF) string-velocity waveforms with periods of about (a) twice and (b) three times the fundamental period T_1 (indicated by the vertical dashed lines). The bow velocity = 10 cm/s. A horizontal dashed line indicates nominal slip velocity v_S (Helmholtz motion). (From Schoonderwaldt, 2009b)

of the spectral centroid values which increased steadily with increasing bow force, at least within the playable region. On the other hand, the influence of bow-bridge distance and bow velocity was rather minor. There was a weak tendency of spectral centroid to increase with β and decrease with increasing v_B, as regression analysis revealed. The pitch flattening effect was evident at high bow forces approaching the upper bow force limit and more pronounced at higher bow velocities and large bow-bridge distances. It was shown that the 5–10 cent flattening regions followed approximately the slope of the upper bow-force limit, hence the line separating areas with excess of 5–10 cent flattening could be considered as a practical upper bow-force limit (Schoonderwaldt, 2009b). The observed dependence of pitch flattening on bow-bridge distance had a somewhat irregular nature, to the extent that for middle range β occasional pitch sharpening occurred just before bow force reached the critical value and the pitch flattening was triggered. In addition, in the regions above the upper bow-force limit, raucous motion and other nearly periodic motions such

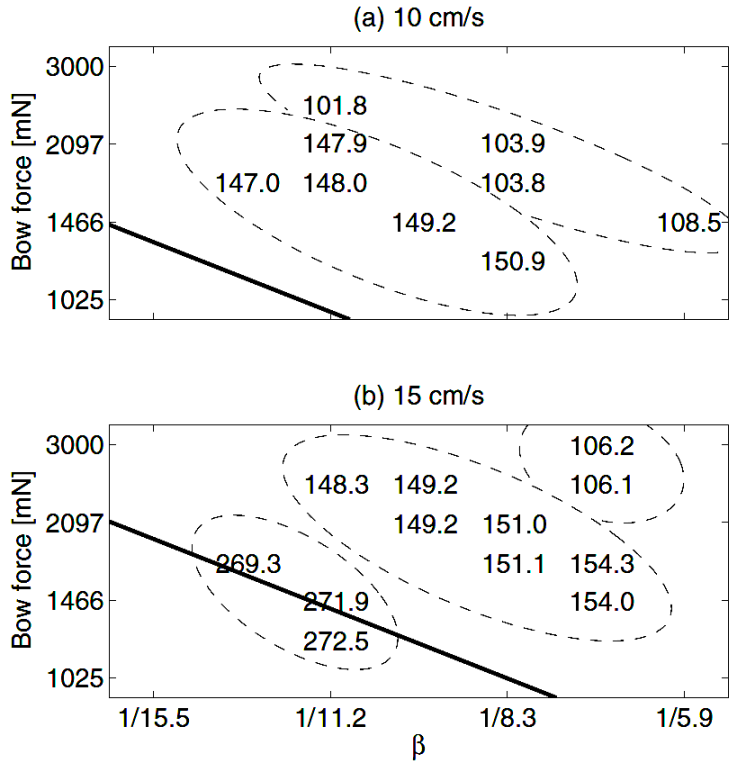


Figure 4.5: Anomalous low frequencies (ALF) in the Schelleng diagrams at bow velocities (a) 10 and (b) 15 cm/s. The numbers indicate the frequency in Hertz. Clusters of different types of ALF include: period doubling (around 150 Hz), period tripling (around 100 Hz), and pitch lowering by a semitone (around 270 Hz). The upper bow-force limits are indicated by solid lines. The nominal fundamental frequency of Helmholtz motion = 293 Hz (the open D string). (From Schoonderwaldt, 2009b)

as anomalous low frequencies (ALF) and S-motion were observed. Typically, there were ALF motions with doubling or tripling of the periods, and to a lesser extent motions with only a semitone lowering in pitch. In all cases, their frequency clearly depended on bow-bridge distance, i.e. increased with larger β, as can be seen in Figure 4.5. Schoonderwaldt concluded that bow force is the violinist's main control of the spectrum higher frequency content, hence the control over the *brilliance* of the tone, while bow-bridge distance and bow velocity act as indirect control parameters providing the player with access to a suitable bow-force range constrained by the bow force lower and upper limits.

4.3 Measuring bowing parameters in normal playing

4.3.1 Systematic studies of the playable region

Askenfelt (1986, 1989) was the first to extract bowing parameters in violin playing. He used a thin resistance wire placed among the bow hair to measure the instantaneous transverse bow position and bow-bridge distance (the latter added in Askenfelt, 1989), and four strain gauges mounted at the frog and tip of the bow to capture the instantaneous normal bow force at the point of bow-string contact. In both studies, bow velocity was derived from the bow position signal. In addition, the output dynamic level was estimated by means of the vibration level measured by an accelerometer placed on the top plate, close to the left bridge foot. Two professional violinists who played short exercises on the same violin using the same adapted bow were recorded.

The main bowing parameters (bow velocity, bow-bridge distance and bow force) and the vibration level were analysed in respect to the note duration (whole, half and quarter notes), dynamics (*forte*, *mezzo forte*, *piano*) and different bowing patterns including sustained *detaché* notes, scales, *crescendo-diminuendo*, *sforzando*, and *saltellato* (a type of bouncing bow stroke). All the note examples were performed on the open G string. Aiming to extract typical values that occur in violin playing, Askenfelt also reported on individual strategies in the use of the parameters between the two players. For example, comparing values of bowing parameters in whole notes played *mezzo forte* (values averaged over 10 s of music signal), both violinists seemed to bow with the same force while one of them played relatively further from the bridge and with higher bow velocity producing the vibration level at least 1 dB lower than that of the second player. On the other hand, the similar vibration levels obtained for *forte* notes were the result of completely opposite bowing strategies: relatively larger bow force applied closer to the bridge in combination with lower bow velocity against smaller bow force combined with higher bow velocity and

larger bow-bridge distance. However, Askenfelt reported that "due to an unfortunate loss of absolute calibration in a range of +/- 2 dB, the vibration levels..." for the player who in this scenario used much larger bow force "...have been shifted arbitrarily to give the same level in *forte* as for the other player", thus the actual output levels might have been quite different and remained in actual correspondence to the individual combination of bowing parameters exhibited by the player. With no other statistical measures provided (except for range values), it can not be determined whether the differences in the parameters between the violinists were significant, with one exception. For the above mentioned *forte* whole notes, provided range values of bow force for each player did not overlap implying that the observed mean difference of at least 0.5 N was significant.

With respect to bow velocity and bow-bridge distance contribution to the output dynamic level, Askenfelt's findings were consistent with those of Schelleng. He confirmed that bow velocity and bow-bridge distance are the player's main controls of the sound level. As observed, the vibration level is proportional to the v_B/β ratio. Bow force, although increasing with increasing dynamics (increasing v_B or reducing β requires a higher bow force), does not contribute to the amplitude of sound. It mainly regulates the sound's harmonic content.

Askenfelt also reported typical values of the bowing parameters as observed in the study. For example, bow-bridge distance ranged from 10 to 50 mm, compared to 1–4 mm and 55–60 mm in *sul ponticello* and *sul tasto* playing respectively. Bow velocity varied between 0.2 and 1 m/s with the occasional decrease to 0.1 m/s or increase to 1.5 m/s. Typical values of bow force ranged within 0.5–2 N. The lowest force of 0.15 N was observed with the bow resting on the string at the tip, while the highest force was about 3 N. Askenfelt noted that none of the parameters was stationary at any given time point, being continuously adapted by the players.

The coordination and control of bowing parameters in violin and viola performance was studied by Schoonderwaldt (2009a). Optical motion capture for

tracking the position and orientation of the bow and the violin combined with a bow-force sensor and an accelerometer mounted on the frog (Schoonderwaldt and Demoucron, 2009) were used to record bow velocity, bow-bridge distance and bow force of three violin and three viola players. Various experimental settings included different note durations (whole, half and 16th notes) played at three dynamic levels (*forte*, *mezzo forte*, *pianissimo*), as well as four varying *crescendo-diminuendo* patterns performed on half notes. Examples of bowing control parameters captured in the three note-length conditions are presented in Figure 4.6. With the aim to extend Askenfelt's findings limited to the open G string, the note sequences were performed on all four strings each stopped at a musical fourth above the open string. The data was collected using the same instrument and bow combinations except for one viola player who chose to play a smaller viola.

The obtained bowing parameters were analysed on steady parts of notes, excluding transients corresponding to bow changes (cut off margins before and after the bow change were 200 and 50 ms for long and 16th notes respectively). The minimum bow force threshold was set at 0.01 N. The results across four strings, with contrasting note durations (whole vs 16th notes) and dynamic levels (*forte* vs *pianissimo*), although being generally in agreement with the findings of Askenfelt (1989), showed also stronger contrasts in control strategies and larger differences between the extreme values of the bowing parameters. For example, in violin performance, bow velocity ranged from about 0.05 to 2 m/s, bow force from 0.1 to 2–2.5 N, and bow-bridge distance within 11–63 mm (corresponding to β values 1/22–1/4) for the stopped string (15–84 mm for the open string). There were clear differences in the use of bowing parameters across strings. Bow force was generally higher on the lowest string G, combined with slightly larger bow-bridge distance to account for higher characteristic impedance and internal damping of the string. Interestingly, in *forte* notes, bow force measured on the E string was also higher compared to the middle strings D and A, but was not accompanied by the shift in bow-bridge

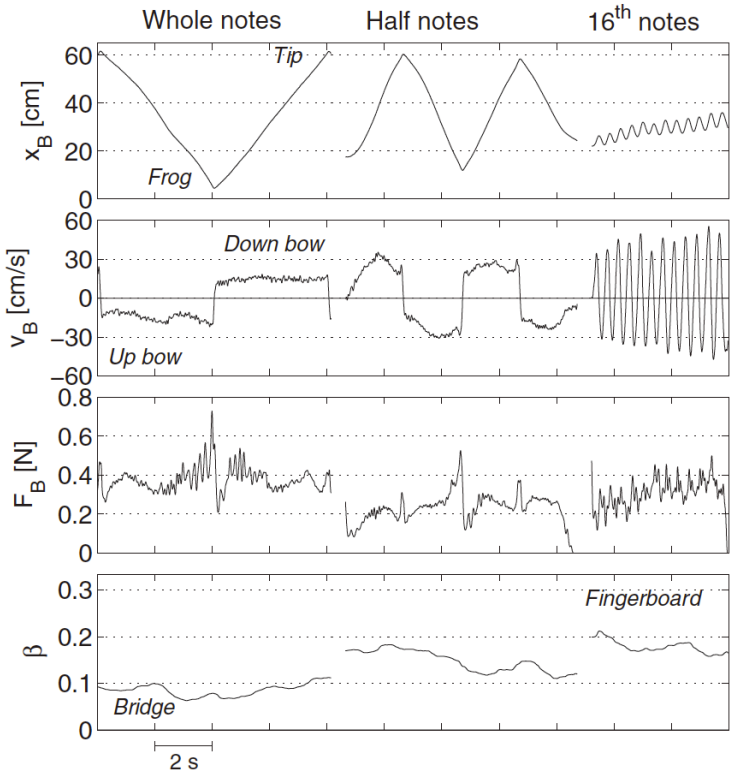


Figure 4.6: Examples of bowing parameters measured in whole notes, half notes, and 16th notes performed by a violin player on the D string in *mf* dynamics. The parameters from the top: bow transverse position (x_B), bow velocity (v_B), bow force (F_B), and relative bow-bridge distance (β). (From Schoonderwaldt, 2009a)

distance. On the other hand, no substantial differences in bow velocity between the strings were noted except for 16th notes in *forte* and *mezzo forte* showing a significant increase of the parameter from the lowest towards the highest string.

In terms of controlling dynamic levels, a typical "trade-off" (Askenfelt, 1989) between bow velocity and bow-bridge distance was observed. In the whole notes, the dynamics was primarily dependent on bow-bridge distance, while in 16th notes bow velocity was the main control parameter. From the aggregated data across all conditions, it became evident that an increase in measured sound levels is proportional to an increase in v_B/β ratio, confirming Askenfelt (1989)'s results. Bow force was also clearly highly correlated with v_B/β but its role in

setting dynamic levels lies with boosting higher harmonics which then affect the perceived loudness.

The strongest differences between violin players occurred in the use of bow force in *forte* and *mezzo forte* notes independently of the note length. This was combined with rather similar bow-bridge distances in *forte* and more diverse values of the parameter in *mezzo forte*, exhibiting stronger tendency to reduce β in playing with higher forces and vice versa. For example, the violinist who on average used the highest bow forces in all conditions tended to play the closest to the bridge. Interestingly, in *pianissimo*, where the bow forces observed among the players were alike, the individual differences in bow-bridge distance were most pronounced. As for bow velocity, there were minor differences in the whole and half notes at all dynamic levels between the players, in contrast to the 16th note condition where the differences were substantial. Significant differences in bow force observed for the same condition were kept in proportion, i.e. the violinists who used relatively higher velocities played with larger bow forces, which would be expected in accordance with the established relation of bow force being proportional to bow velocity.

Unfortunately, no further details were given of how the players' individual bowing strategies were reflected in spectral centroid, the audio feature extracted from sound recordings. As one of the acoustic characteristics of tone quality, it could provide a preliminary insight into timbral differences between the violinists. Instead, only general dependencies of the feature on the main bowing parameters were presented. By means of multiple linear regression, it was shown that on the lowest string G bow force was the most dominant factor in controlling spectral centroid, followed by bow velocity and bow-bridge distance. In general, spectral centroid increased with increasing bow force and only slightly with bow-bridge distance, and decreased with increasing bow velocity, which came in agreement with the earlier experiments on a bowing machine (Schoonderwaldt, 2009b). Towards the higher strings, the spectral centroid's dependency on bow force and then on bow velocity gradually diminished resulting in a very weak

interrelation between bowing parameters and spectral centroid for the E string. Schoonderwaldt suggested that other factors such as *vibrato* and damping effect of the finger could cause spectral centroid fluctuations without a direct relation to bowing parameters. Generally, spectral centroid increased from the lowest to the highest string, partly due to increase in pitch, and partly due to change in mechanical and acoustical properties of the strings themselves. In addition, the results aggregated across conditions indicated that spectral centroid increased with the dynamics, from *pp* to *f*.

Studies by Askenfelt (1986, 1989), Schoonderwaldt and Demoucron (2009) and Schoonderwaldt (2009a) aimed at systematic exploration of bowing parameters that occur in normal violin or viola playing. By means of dedicated equipment, bowing parameters of the players were captured in a variety of bowing scenarios and further analysed in reference to Schelleng's bow force limits. These experiments provided empirical evidence that those limits were generally respected, and as the players changed the dynamic levels they moved along the Schelleng diagram keeping the coordinated parameters within the limit contours (Schoonderwaldt, 2009a).

No such studies have been so far attempted on the cello. However, following Galluzzo (2003)'s earlier findings on a mechanically bowed cello and their great resemblance with those obtained on the violin, one may expect the revealed bow force limits for establishing and maintaining Helmholtz motion in normal violin playing, as they are expressed in relative terms, to be universal for other bowed string instruments, except for observing increased absolute values of the three bowing parameters due to longer and thicker strings.

4.3.2 Capturing bowing gestures for interactive performance, sound synthesis and bow stroke analysis

In parallel to the systematic studies, sensor-based devices capturing bowing gestures have been developed for augmented interactive performance and gesture-driven sound synthesis, and for analysis of standard bowing techniques. Hypercello (Machover, 1992) was one of the earlier attempts to create an interface which via gesture-controlled real-time modifications of the digitally processed sound would provide a player with new ways of musical expression. It consisted of an electric cello and a dedicated bow fitted with custom sensors to track right hand wrist angle in two dimensions (flexion and deviation), the right hand index finger pressure on the bow (as a representation of bow force), bow position in two dimensions, i.e. transverse position and bow-bridge distance, and left hand finger position on strings. The raw sensor information together with captured string loudness and pitch tracking were used to detect higher level cues such as note attacks, wrist tremolos, bowing style, and bow range, which were then combined in different cello-modes and linked to specific sound manipulations (Machover, 1992). Other augmented instruments and novel interfaces equipped with the ability to measure movements of the bowing hand for controlling sound effects or sound synthesis included projects such as Celletto (Chafe, 1988), BoSSA (Bowed-Sensor-Speaker-Array) complemented with R-Bow (Trueman and Cook, 1999), the eviolin (Goudeseune, 2001), vBow (Nichols, 2002), or the Overtone Violin (Overholt, 2005).

Extending the concept of Hypercello (and other Hyperinstruments), the Hyperbow project (Young, 2001, 2002, 2003) focused on an augmented violin bow. To measure bow transverse position, a system based on electric field sensing was integrated into a commercial carbon fibre bow. It comprised a resistive strip spanning the length of the bow hair attached to the stick and a simple electrode antenna placed behind the bridge of the violin. Foil strain gauges were mounted at the middle of the stick to detect downward and lateral strains in the bow

stick, while 3-D bow acceleration was captured by means of two accelerometers. The accelerometers and the electronics for wireless data transmission were accommodated on a printed circuit board mounted on the frog.

The Hyperbow controller was essentially designed for professional musicians to be used in new music performance scenarios (e.g. Tod Machover's *Toy Symphony*, Patrick Nunn's *Gaia Sketches* and *MODES* by Artem Vassiliev for Hyperbow adapted for cello, Young et al., 2006). To serve also as a research tool for measuring violin bowing technique, the original design was adjusted to incorporate units for measuring the orientation of the bow in relation to the violin (Young, 2007). It was done by adding 3-D angular velocity sensing (by means of gyroscopes) to the existing 3-D acceleration sensing on the bow to create a 6DOF inertial measurement unit (IMU). A similar 6DOF IMU was added to the violin. The initial electric field bow position sensing subsystem was also expanded to include four receive antennas (one for each violin string). The violin part of the system was implemented on a commercial electric violin. For applicability and comparability of the gesture data, the sensors' output was calibrated in SI units.

The above setup was used to collect gesture and audio data from eight violinists performing six different bowing techniques on each of the four violin strings. Registered bow strokes included accented *détaché*, *détaché lancé* (*détaché* with unaccented, distinct breaks between notes), *louré* (gently pulsed *legato* notes executed in one bow stroke, also known as *portato*), *martelé* (notes with a "pinched" attack followed by a quick release, executed on-string), *staccato*, and *spiccato*. For gesture analysis, eight sensor data were selected: the downward and lateral forces; x, y, z acceleration; and angular velocity about the x, y, and z axes. The results of principal component analysis (to reduce the dimensionality of the data) combined with k-Nearest Neighbours (k-NN) classification showed that the gesture data captured by the implemented system was sufficient to discriminate between common bowing techniques as well as to indicate similarities and differences between the players in executing the same

bow strokes.

IRCAM's Augmented Violin (Rasamimanana, 2004) was largely inspired by the Hyperbow and Hypercello projects. Like the Hyperbow, the motion sensing system was added to a conventional carbon fibre bow. The bow position (transverse and with respect to the bridge) was acquired by means of the electromagnetic sensor (a magnetic tape fixed along the stick) combined with an antenna placed behind the violin bridge. Two accelerometers mounted on the electronics board (attached to the frog) measured bow velocity fluctuations in three dimensions. The forefinger pressure on the bow stick captured by a force sensing resistor (FSR) represented the downward pressure of the bow hair on the string (bow force). The bow data was sent via a radio frequency (RF) transmitter.

As reported by Rasamimanana et al. (2005), the augmented violin was employed to record three types of bow strokes (*détaché*, *martelé* and *spiccato*) in two tempi (60 and 120 bpm) and three dynamic levels (*pianissimo*, *mezzo forte*, *fortissimo*) from two violinists playing scales on each of the four strings separately. From the collected gesture data only accelerometer signals were selected for the analysis, and derived features such as the minimum and maximum acceleration per note/stroke were used to model bow stroke classes. A k-NN classification procedure on acceleration parameters generally yielded high recognition rates, especially for the whole database and cross-player scenarios, with some confusions occurring for particular stroke-type–dynamic-level conditions in one classification scenario. Rasamimanana et al. concluded that bow acceleration can be considered as a reliable parameter for characterising and recognising different bowing techniques and subsequently can also be related to continuous sound characteristics and/or perceptual features of the player. With the implemented real-time bow stroke analysis and classification software module, the augmented violin became a central component of the interactive gesture-controlled composition *BogenLied* (Bevilacqua et al., 2006).

Another real-time system for classification of violin bow strokes was developed under the CyberViolin project (Peiper et al., 2003). Its aim was to measure and identify different bowing articulations and provide a violinist with objective real-time feedback in a visual form within a dedicated virtual environment. The gesture capture was based on an electromagnetic (EMF) motion tracking system with two sensors attached to the violin and the frog of the bow, respectively. A set of motion parameters computed from the raw sensor data included bow sensor position at the beginning and the end of the stroke, its average speed, the stroke length, as well as frequency of bow change, acceleration or deceleration within a stroke, continuity of motion between strokes, bow position (middle, upper, lower), number of changes in a single coordinate, and lack of movement within a stroke. These bowing features fed a decision tree algorithm in both training and classification modes in order to model and then discriminate between five types of articulation such as *détaché*, *martelé*, *spiccato*, *legato*, and *staccato*. In spite of promising performance, Peiper et al. reported that the accuracy of the system had some limitations at that stage of development, mainly due to insufficient sensor precision and sampling frequency. The concept of visual feedback for a string player was taken to a new level with the introduction of the i-Maestro project[1]. Designed as an interactive environment for teaching and learning of bowing technique and posture, the system provides multi-modal feedback based on analysis, visualisation and sonification of 3-D motion capture data (Ng et al., 2007a,b). In later stages, the system has been supplemented with new components including score following and annotation, symbolic music representation, and audio analysis and processing (Ng, 2008).

Bowing gesture acquisition based on EMF motion tracking technology was also exploited by Maestre et al. (2007); Maestre (2009) and Pérez et al. (2007, 2008); Pérez (2009) for gesture-driven sound synthesis applications. The sensing system was complemented with a bow force measuring component based on two strain gauges mounted on the frog of the bow to detect the hair ribbon

[1] `www.i-maestro.org`

deflection in respect to the current bow position (Guaus et al., 2007) similarly to solutions proposed earlier by Askenfelt (1986) and Demoucron et al. (2006). The obtained strain gauge signals were calibrated into Newtons. Later, the calibration method was modified (Guaus et al., 2009) to compensate for the changes in the bow tension (affecting strain gauge readings) during a long recording session. The calibration data was used to train Support Vector Regression (SVR) models employed to predict the real pressing force (in Newtons) based on the same input parameters from a real recording compared with the strain gauge measurements. Aiming at a reduced intrusiveness of the measurement process, further development of the method (Marchini et al., 2011) led to the total elimination of strain gauges in favour of bow force estimation based solely on the motion and calibration data.

Two 6DOF (3-D position and orientation) sensors attached to the violin and the bow provided the raw motion data for computation of bowing parameters: bow transverse position, bow transverse velocity, bow acceleration, bow-bridge distance, bow tilt, bow inclination, bow skewness, and bow-string distance (a measure of bow hair deflection under the pressing force at the point of contact on the string). Additional parameters included estimated string being played and left hand finger position. The system was employed to create a dedicated database of thoroughly annotated multi-modal recordings covering most common violin playing contexts. The acquired bowing and audio data was used to build generative timbre models for gesture-informed concatenative synthesiser (Pérez, 2009; Pérez et al., 2012) and to analyse and model bowing parameter contours for physical modelling and sample-based synthesis (Maestre, 2009; Maestre et al., 2010). More recently, lower level ensemble performance data obtained via the sensing/recording framework extended to four instruments of a string quartet (two violins, viola, and cello) were used to investigate interdependencies among musicians and build computational models of ensemble expressive performance (Marchini et al., 2013, 2014).

Multi-modal recordings of six cellists collected in 2011 by means of a cello-adapted version of the above sensing system (see details in Chapter 5) formed the basis for the development of this thesis exploring timbre differences between string players with respect to their executed bowing gestures.

Pardue and McPherson (2013); Pardue et al. (2015) proposed to use near-field optical reflectance sensors as an "inexpensive, portable and non-intrusive" alternative to the bowing motion tracking systems based on electric/electromagnetic field sensing or optical motion capture. They argued that "the ideal tracking system" should allow the performer to install it on his/her own instrument with minimal adjustments required, be available for the use in any on-stage real-time scenario, offer satisfactory spatial and temporal resolution, and above all, provide all these qualities at a reasonable cost. To address these requirements, their system, built on four near-field optical sensors attached to the bow stick, used the reflected infrared light to measure the distance between the bow hair and the stick, i.e. the amount of bow hair deflection under the pressing force. From the acquired optical data, bowing controls such as bow transverse position and bow force were estimated. In (Pardue et al., 2015), the system, which combined the bowing motion capture module with sensor-based fingerboard tracking, was evaluated in gesture-informed note onset detection, with a view to future real-time applications. The experiment involved classifying three different types of onsets including off-string attack, on-string bow change and finger change in three musical contexts. The aimed accuracy of the system for real-time performance was set at 10 ms after the ground-truth label. The best accuracy for early onset detection, i.e. within the targeted 10 ms, was obtained for off-string attacks (68%) followed by finger changes (56%). For bow changes, Pardue et al. reported only 19% of correct detections. They concluded that, in comparison with motion capture systems, the proposed optical sensing approach provides the player with more flexibility and freedom of movement, while offering high resolution of millimetre distance measurement (here, of bow stick-hair distance) and possibility of processing optical data in real-time.

4.4 Summary

In this chapter, different methods for acquiring and analysing bowing controls have been outlined. They included systematic studies with the use of bowing machines as well as with the employment of custom sensor-based systems tracking a player's bowing movements. The results revealed strong dependence of output sound spectra on the main bowing parameters and led to defining the playable region spanned between bow force and bow-bridge distance for a given bow velocity. The conditions for perfect Helmholtz motion triggering and its steady-state maintenance were also discussed.

Few studies (Askenfelt, 1989; Young, 2007; Schoonderwaldt, 2009a) have made an attempt to compare bowing techniques of different players. Their results across most common bowing articulations suggested that in some cases differences between individual bowing parameter ranges or parameter combinations can be substantial. However, no details were provided whether the observed differences in the use of bowing controls had any significant effect on the resulting tone spectra, or perceived tone quality of the players. This lack of established link between the individual gesture and tone colour motivated a series of investigations carried out in the scope of this thesis in search of individual bowing strategies which can characterise a player and become a mechanical determinant of his/her unique timbre.

Chapter 5

Experimental data collection for acoustical and gestural analysis

Collecting relevant data which then enables a thorough investigation into a scientifically stated problem is a crucial step of every exploratory research process. This experimental work was no exception and a new database of cello recordings, enhanced by measurements of bowing control parameters via capturing performance gestures, was required in order to understand how the physical actions that a performer exerts on an instrument affect spectro-temporal features of the sound produced, which then can be perceived as the player's unique tone quality. Such a database was created in collaboration with the Music Technology Group based at Universitat Pompeu Fabra (Barcelona) where a dedicated sensing/recording system was available for that purpose.

The following sections describe in detail audio and sensor data acquisition, recording conditions and equipment used, musical content of the database and the gesture data structure.

5.1 Performers

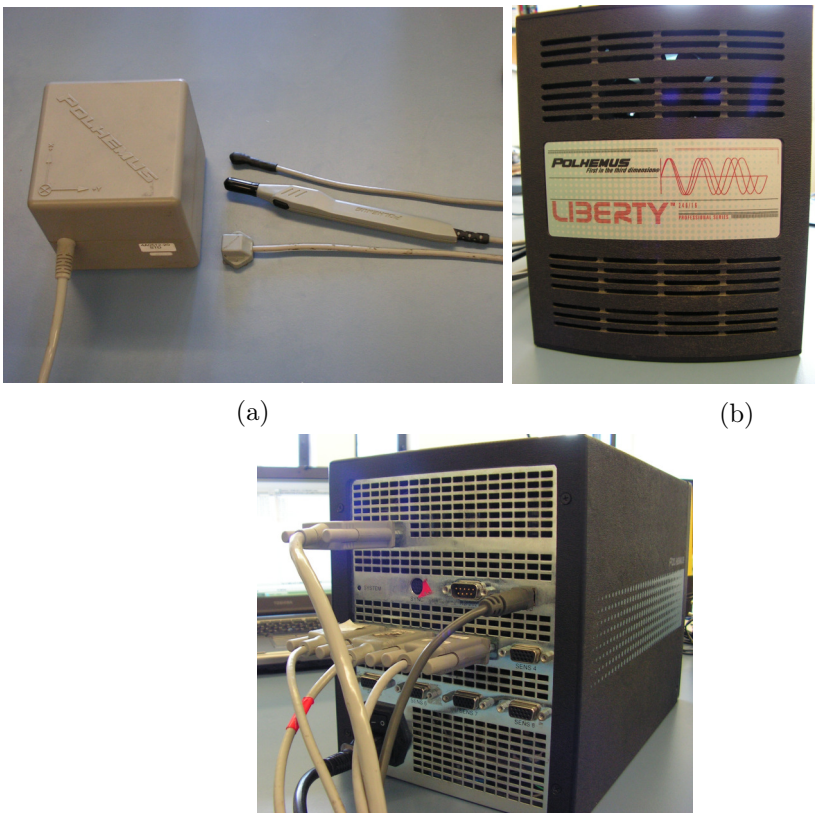
Six advanced cellists participated in the recordings. Five of them were students and graduates from ESMUC (Escola Superior de Música de Catalunya, Barcelona) and the sixth was the author of this report. A carefully chosen repertoire was recorded on two different instruments, both of a good luthier class,

using the same high quality bow. The first cello (further as Cello1) was borrowed for the experiment and none of the participating cellists was the owner, the second cello (further as Cello2) and the bow used in all recordings belonged to the author. The recording sessions were held in a professional studio located at Roc Boronat Campus of Universitat Pompeu Fabra where all required equipment was available and proper recording conditions secured.

5.2 Data acquisition framework

The motion tracking and audio capturing system, built originally for studying instrumental gestures in violin performance (full descriptions can be found in Maestre, 2009; Pérez, 2009), was designed to acquire three data streams: (i) audio signals from a bridge pickup and ambient microphone, (ii) bowing motion coordinates from sensors, and (iii) load cell values during a bow force calibration procedure.

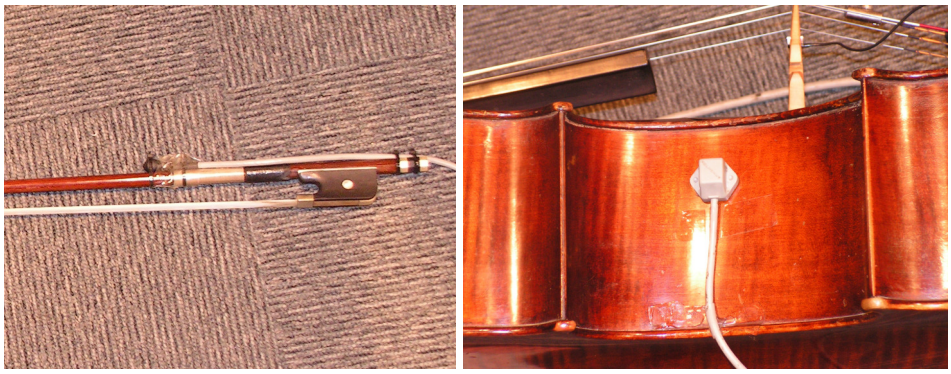
The bowing motion tracker was built on the *Polhemus Liberty* commercial unit, a six-degree of freedom (6DOF) tracking system based on electromagnetic field sensing (EMF). Its four components included: a transmitting source, a sensor marker, and a couple of receiving wired spheric sensors, one fixed to the bow stick at the frog part, the second attached to the side plate of the cello (see Figures 5.1–5.2). The guaranteed electromagnetic field radius was approximately 1.5 cubic meters and extra precautions had to be undertaken while arranging a space around as the field itself was very sensitive to any metallic objects placed nearby. The sensor marker was used to calibrate cello and bow coordinates in a 3-D sensing space with help of a reference plane. In particular, each string was marked at the bridge, at the beginning of the fingerboard and at the nut points, and the bow hair ribbon was marked at four points, two taken at the frog and two taken at the tip. The sensor data was captured at a sampling rate $s_r = 240$ Hz and synchronised via a PC unit with the audio streams. The data synchronisation did not work perfectly and additional manual corrections were necessary.



(a) (b)

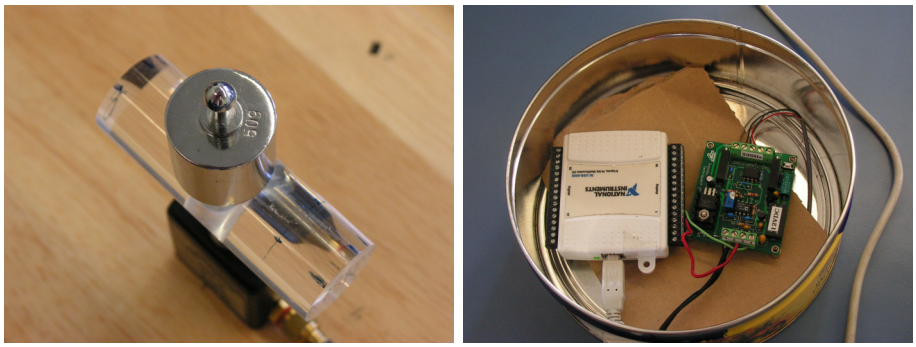
(c)

Figure 5.1: The *Polhemus* system components: (a) the source of EMF (left), the sensor attached to the bow (upper right), the sensor marker (middle right), the sensor attached to the cello (bottom right); (b) and (c) the PC unit processing sensor data.



(a) (b)

Figure 5.2: The *Polhemus* motion tracking components: (a) the sensor attached to the bow; (b) the sensor attached to the cello.



(a) (b)

Figure 5.3: Equipment for bow force calibration: (a) bowing cylinder mounted on the *Transducer Techniques MDB-5* load cell, here with a 50 g precision weight for calibrating load values; (b) the *Transducer Techniques TMO-1* amplifier (right) and the *National Instruments USB-6009* A/D converter (left).

Taking into account that the results of subsequent comparative timbre analysis could be biased by recording artefacts, a crucial part of the data collection was to acquire high quality audio signals. Having the recording setup strictly preserved through all the sessions, two reference sound sources were captured. The first audio signal came from a high class ambient mono microphone located in the near-field to the cello, the second was a signal from a bridge pickup. This signal was of special importance for investigating timbral content of the acquired sound. It captures the spectral content of the sound source, i.e. the vibrating string, modified by the resonances of the bridge itself but unaffected by the resonances of the cello body and, most importantly, unaffected by reverberating characteristics of the recording studio or minute differences in the positioning of the player at the microphone. Employing the pickup signal allowed more direct observation of the relation between bowing controls used by a player and resulting instantaneous changes in spectral characteristics of the sound produced.

Since bow force is one of the fundamental bowing parameters related to timbre, along with bow velocity and bow-bridge distance, the third component of the framework was developed to measure bow force data acquired by means of a load cell, so that bow pressing force can be estimated. A dedicated facility

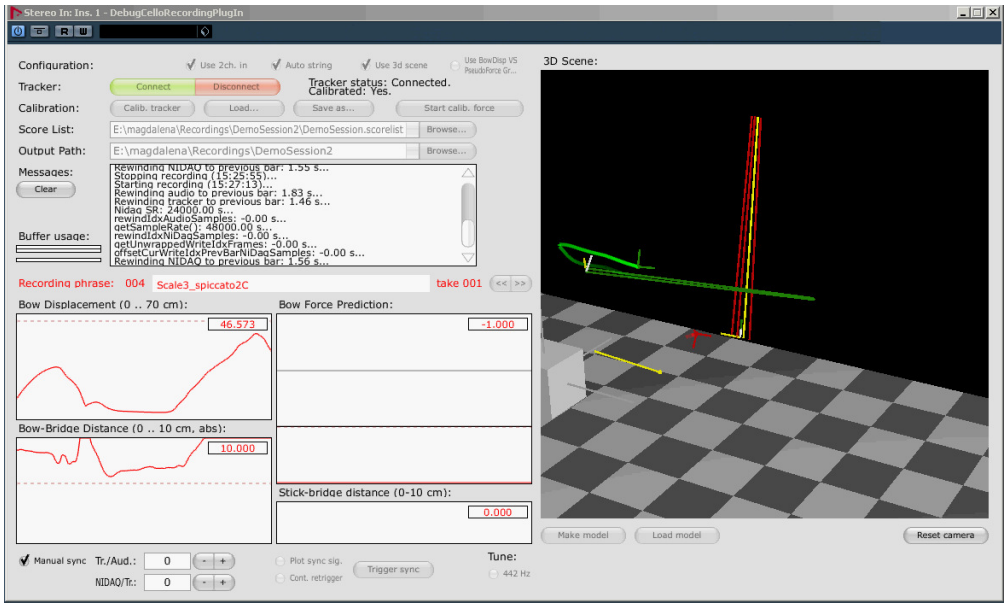


Figure 5.4: Screenshot of the VST plugin interface designed for controlling audio, bow motion and bow force data acquisition.

was built for this purpose using the *Transducer Techniques MDB-5* load cell connected to the *Transducer Techniques TMO-1* amplifier and the *National Instruments USB-6009* A/D multifunctional converter (as shown in Figure 5.3). In the bow force calibration procedure, a cylinder mounted on the load cell was bowed by the experimenter and bow force range was sampled at different points along the bow length. The load values were recorded simultaneously with bowing motion coordinates from sensors, and translated into Newton units before further application.

The overall process of audio/motion data acquisition was synchronised and controlled via a dedicated VST plugin built into the *Steinberg Nuendo 4* recording software. While capturing the data, a real time 3-D representation of the position of the cello and bow was visualised on the screen together with some bowing parameters such as bow displacement and bow-bridge distance (see Figure 5.4).

Figure 5.5: The recording studio layout.

5.3 Studio recording setup

Since acoustical conditions of a recording session affect the quality and spectral content of the music signal and consequently also timbre features extracted from a digital recording, to preserve all the recording conditions from the first to the last session was a matter of fundamental importance. This required to maintain "untouched": (i) the studio layout, i.e. positions of the microphone, the player's chair and the overall studio arrangement (Figure 5.5), (ii) the recording equipment (microphone, bridge pickup, DI box and appropriate cables), (iii) the recording console settings (dynamic levels for the microphone and pickup), once set for the first player remained identical for all participants.

The recording equipment specifications were as follows: (i) an ambient microphone – model Neumann U87Ai P48, (ii) a piezo-ceramic bridge pickup – model Fishman V-100, (iii) a DI box – model BSS Audio AR-133 for balancing a signal from the pickup connected through Pinanson Audio Patch Cord EC 605, (iv) a recording console – model Yamaha 02R96.

To obtain an absolute dynamic level of the recorded ambient sounds, a 1 kHz/80 dB sine tone, intended to serve as a reference level for all recordings, was recorded before each session. The sine signal was emitted from a Roland DS-5 Bi-Amp monitor placed on the cellist's chair at approximately 1 m distance from the microphone. The signal level was measured at the position of the microphone using a digital sound level meter – model CESVA SC-2c.

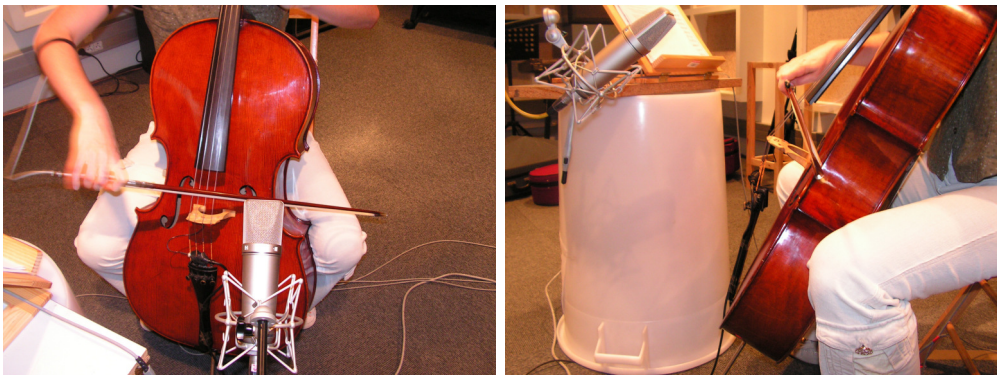


Figure 5.6: Documenting the position of the microphone and the cellist.

For the ambient sound recording, the same sitting position and distance of the cellist from the microphone were maintained through all the sessions, to ensure that the captured sound reflected the players' natural sound intensity. For this purpose appropriate photo documentation was carried out by taking pictures of the player position on a chair and the way he/she held the cello (Figure 5.6).

Since two cellos were used in the recordings, it was necessary to take care about the pickup placement as the cellos' bridges differed significantly in their shapes. Taking into consideration that the mounting place will affect the spectral content of the pickup signal, the position of the pickup had to be chosen carefully. Once decided for each cello it was recreated throughout sessions with help of the photo documentation (Figure 5.7).

Finally, to ensure intonation consistency of the recorded musical material, the instruments were tuned to a reference 440 Hz (A4) tone using Yamaha Chromatic Tuner YT-250 and checked through the course of each session.

5.4 Repertoire

The musical repertoire for the experiments was chosen to represent three different types of musical expressivity based on the Baroque, Romantic and contemporary music styles. The selection of excerpts aimed to encourage a large range of articulation and dynamic variation in the data. The length of the

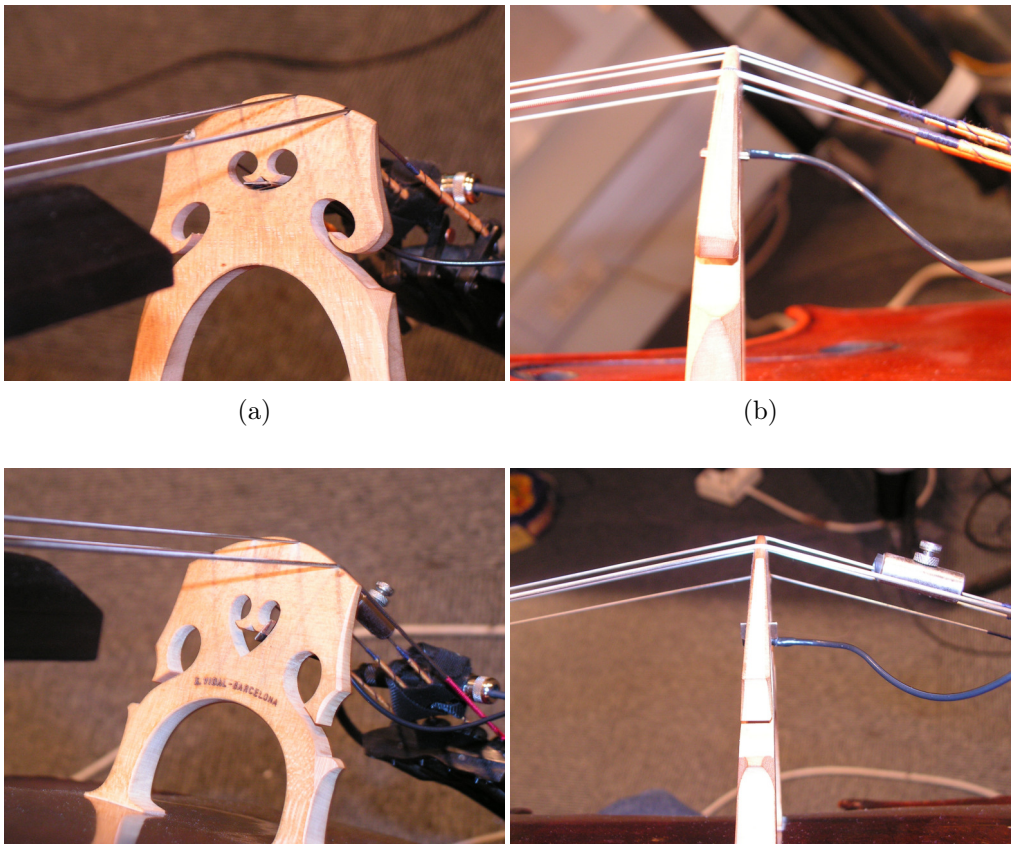


Figure 5.7: The pickup mounted on the bridge of Cello1 (a)–(b) and Cello2 (c)–(d).

chosen music fragments ranged (depending on the context) from a few bars (a phrase) to the whole movement, avoiding unnecessary fatigue of participants when a number of repetitions was required. To provide additional reference data for "timbre objectivity" and observe timbre changes depending on a performer playing within or without the musical context, a scale in four articulation variants was recorded by each player on both cellos. The resulting collection of tone samples featured timbre diversity in the following aspects:

1. instrument

   the entire music material was recorded on two different cellos to provide an insight into the ability of the players to adapt to physical constraints of an instrument while achieving a desired quality of tone;

2. musical context
   music fragments comprised examples of three different music styles to ensure capturing timbre identities of the players when they operated within various musical aesthetics;

3. articulation
   since timbre features vary significantly with articulation, to capture detailed timbral palettes of the cellists, recorded samples of scales and Baroque music included different articulation variants;

4. dynamics
   similarly to articulation, dynamics is a significant factor in shaping spectral content of the produced sound; therefore Bach's *Bourrée II* and Fauré's *Élégie* were recorded at two dynamic levels;

5. vibrato
   vibrato is known to play a key role in perception of instrumental timbre; to enable investigation into the vibrato effect on timbral characteristics of the players, all fragments of Bach's *Suite* were recorded in two variants: with vibrato and *non vibrato*.

5.5 Recording session scenario

Each recording session was carried out according to the following scenario:

I. Preparation

1. studio layout check

2. microphone, pickup and console setting check

3. recording of a 1 kHz/80 dB reference signal

II. Sensing system setup (Cello1)

1. Fixing the sensors on the cello and the bow

2. System calibration

3. Bow force calibration

III. Music recording (respective music scores attached in Appendix A)

1. D-major scale – 3 octaves (played upwards and downwards using identical fingering), 4 articulation variants:

   (a) the whole bow – 4 notes *legato*, at tempo ♩ = 80 bpm

   (b) the lower 2/3 part of the bow – a combination of 2 notes *legato* and 2 notes *détaché*, at tempo ♩ = 160 bpm

   (c) at the bow gravity point – *spiccato*, at tempo ♩ = 320 bpm

   (d) the upper 2/3 part of the bow – punctuation rhythm, at tempo ♩ = 80 bpm

   Tempo was given to the players before the start and during the recording using a metronome light signal.

2. Bach – *3rd Cello Suite* – each excerpt recorded in 3 bowing variants: (i) the participant's own bowing and articulation variant, (ii) the same but performed *non vibrato*, (iii) universal bowing variant identical for all participants with *vibrato* applied only to climax notes

   (a) *Prélude* (bars 1–6)

   (b) *Allemande* (bars 1–4)

   (c) *Courante* (bars 1–8)

   (d) *Bourrée II* (bars 1–8), recorded twice with a change in dynamics $mf \rightarrow p$

3. Fauré – *Élégie* (bars 2–22)

4. Shostakovich – *Cello Sonata* op. 40

   (a) 1st Movement (bars 1–53)

(b) 4th Movement (bars 17–39)

IV. Photo documentation of the player position with the cello.

V. Repeat points I.2 and II-IV For Cello2.

The total duration of recorded music material (including repetitions) was approximately 1 h 10 min per participant.

5.6 Data processing

Before any further analysis the two audio streams (captured from the microphone and the bridge pickup) were checked for synchronisation errors with the sensor data and manually corrected where required.

From the acquired bowing motion coordinates a set of bowing controls and some auxiliary parameters were computed (definitions and derivation methods described in Maestre, 2009). The bowing controls included: (i) **bow-bridge distance** *bb_dist*, (ii) **bow position** (bow transverse position) *bow_pos*, (iii) **bow transverse velocity** *bow_vel*, (iv) **bow acceleration** *bow_acc*, (v) **bow-bridge angle** *bb_angle*, (vi) **bow tilt** *bow_tilt*, (vii) **bow inclination** *bow_incl*, (viii) **bow-string distance** *bs_dist*, and (ix) **string estimation** *string_est*.

The absolute bow-bridge distance values were translated into values of the β parameter, i.e. bow-bridge distance relative to the length of the string in vibration (the effective length) determined by the fingering position:

$$\beta = \frac{bb_dist}{fingerPos} \tag{5.1}$$

where the respective fingering position or finger-bridge distance was calculated as follows:

$$fingerPos = \frac{L_s f_s}{f_0} \tag{5.2}$$

where L_s is the total length of the string being played, f_s is the fundamental frequency of the open string (in Hz), and f_0 is the instantaneous fundamental frequency extracted from the audio signal (in Hz). The instantaneous fundamental frequency f_0 was computed using the *Timbre Toolbox* (Peeters et al.,

2011) from either the pickup or the microphone signal, depending on further applications.

For each cellist and each cello, bowing parameters such as bow-string distance, bow position and bow tilt, and readings of the load cell values, all captured during the force calibration procedure, were employed to model real bow force. As opposed to a direct acquisition method based on a dual strain gage device mounted on the frog of the bow (as applied in Maestre, 2009, for example), this alternative model of bow force estimation calculates bow force using only sensor data (Marchini et al., 2011, see also Section 4.3.2). In the modelling phase, the bow-string distance (a simplified physical model of the hair ribbon and string deflection under bow pressing force, so-called *pseudo-force*), the bow position and tilt act as predictors in a regression model of the respective load cell values, computed using non-linear multiple regression techniques such as support vector regression (SVR) or random forests (RF). The resulting model is then applied to the sensor data acquired from music recordings to estimate real bow force based on the same three bowing parameters.

5.7 Summary

This chapter provided details of creating a new database of cello recordings which included measurements of performance gestures for subsequent extraction of bowing control parameters. This was possible thanks to the existing audio capturing and motion tracking framework developed specifically for research in instrumental gesture in violin playing conducted by the Music Technology Group (Barcelona).

The newly created database aimed at collecting an extended set of tone samples by different players, timbrally diverse in instrument, musical context, articulation, dynamics, and vibrato. This timbrally rich material formed the basis for a series of experiments in exploring performer-dependent facets of musical timbre, the results of which are presented in the next three chapters of this thesis.

Chapter 6

Perceptual evaluation of cello player tone via timbre dissimilarity and verbal attribute ratings

This chapter describes the first of the three major studies conducted within the scope of this thesis, that aim in understanding the relation between gesture, tone quality and perception in cello playing. Commencing the investigation from a listener's point of view, timbre dissimilarity ratings of six cellists' tone samples across various musical contexts are first analysed to obtain timbral maps in which the relationships between the tones, as perceived by the listeners, can be studied. Then, the association between semantic labels and the cellists' tones is examined via verbal attribute ratings, providing an auxiliary interpretation of perceptual dimensions.

6.1 Introduction

This investigation starts at the point where a performer's mastery of tone encounters its ultimate judge – a listener, for in the ears of the listener it is decided whether a performer's tone, resulting from a complex sequence of most subtle physical actions, does or does not become "a vector of musical expression" (Barthet et al., 2011). What it implies is that, although the main principles of playing technique remain the same, the actual "implementation" across different playing schools and then within individuals varies to a great extent. However, only when listening to a performance one can assess and possibly appreciate

the differences, as it is the resulting music experience that matters to a listener, not the technique itself. Following that argument, for exploring the nature of these differences, it was necessary to begin the study with perceptual evaluation of tone samples of the six cellists in question to reveal whether listeners can discriminate between the cellists' timbres in the first place.

A number of (dis)similarity studies established proximity ratings as a standard way to investigate timbre (e.g. Grey, 1977; Kendall and Carterette, 1991; McAdams et al., 1995). However, the majority of them focused on perceptual discrimination between various musical instruments or between tones of a single instrument rather than on exploring what may underlie the perception of timbral dissimilarities between different performers. Such an attempt was reported by Fitzgerald (2003) who in two separate experiments compared single tones of two oboists at different pitch and dynamic levels and tones performed at the same pitch and dynamic by eleven oboists respectively. The results suggested that listeners were able to discriminate between different player/oboe combinations based on single tones of less than 2 seconds in duration (see Section 2.5 for discussion).

Using single notes in similar experimental scenarios seems advantageous for a researcher as it gives full control over the stimuli presented to the subjects in terms of pitch, loudness and duration. However, can a single tone become a representative of the whole timbral palette each player possesses? From the author's standpoint the entire player's timbre cannot be evaluated based on just one note. What is more, since the tone quality serves as a channel of a performer's musical expression, it also needs to be examined within a musical context. A study by Barthet et al. (2010a) clearly shows how timbre descriptors differ when expressive and inexpressive performances of a musical phrase by one performer are compared at the note level. Therefore, for the analyses to follow, the sound corpus consisted of short musical phrases, in order to give a fuller insight into the timbral individuality of each cellist.

Amongst numerous studies on perceptual aspects of timbre, dissimilarity

ratings are often accompanied by verbal attribute evaluation to provide the researcher with the semantic meaning of the revealed timbre dimensions. The common goal is to identify those semantic labels which best describe perceived differences between musical instruments or, to a lesser extent, tones of just one instrument. To the author's best knowledge, a study by Fitzgerald (2003) is the only work which used verbal attributes to explore qualitative differences between timbres from different performers. Verbal magnitude estimation (VAME) ratings on either eight or eleven attributes (preselected as most adequate for oboe timbre) were employed to differentiate between two or eleven oboists respectively. In both cases, the results were significant and proved the attributes to be suitable for the task.

The evaluation of cellists' tones in terms of verbal attributes was also incorporated in the current investigation to aid the interpretation of the perceptual dimensions which emerged from dissimilarity ratings. Among semantic labels already identified by research exploring psychophysical properties of bowed string instruments' timbres (see Štěpánek, 2002; Štěpánek and Moravec, 2005; Štěpánek and Otcěnášek, 2005, for example) three attributes were selected, namely *bright*, *rough* and *tense*. Timbre properties such as *brightness* and *roughness* had already existing acoustical correlates and were widely used to characterise sounds of various instruments including those from the string family. The *tension* attribute was proposed here as a property more related to internal tension in the sound as a means of musical expression, but also as a property related to sound resonance and projection. In general, tense tones sound more condensed/focused, and may often be a bit damped and less resonant.

If the selected attributes seem to characterise slightly negative aspects of a player's tone, it is worth mentioning that, from a psychological perspective, it is easier for listeners to detect or evaluate what is less associated with a beautiful sound. Also, when using unidimensional semantic scales, one can determine the adjective's opposite and place it at the other extreme of the scale, as in the

case of the semantic differential method, or quantify how much of an attribute is present in a certain tone on a "adjective−not adjective" labelled scale, as in VAME ratings. In this study, neither of the two methods was employed since they both assume that no reference value exists for each evaluated attribute. Subjects themselves are expected to define their own reference value for each stimulus, and by ultimately averaging across subjects a global estimation of the attribute is obtained. Instead, a pairwise differential approach was proposed, which required to evaluate the difference in attribute magnitude between two tone samples, rather than assigning an absolute value to each sample separately. This way, in every pair of two samples, each sample became a reference to the other and, once all tones were pairwise compared, an inequality relation was established which ordered the samples according to the attribute.

6.2 Aim of the study and research goals

A series of experiments was designed to perceptually evaluate tones of six cellists playing on the same instrument six short musical phrases which varied in music style and/or genre (these six phrases are often referred to later as "musical contexts").

At first, dissimilarity ratings by a group of expert listeners aimed to reveal whether the six cellists can be perceptually discerned within each music excerpt as well as whether dissimilarity patterns are consistent across different musical contexts.

Further, verbal attribute differential judgements of the same set of tones aimed to uncover whether the six cellists can also be perceptually discriminated through qualitative differences between their timbres, whether these qualitative differences varied with musical context, and ultimately whether they can be interpreted into perceptual dimensions.

6.3 Method

6.3.1 Stimuli

To evaluate timbral differences between tone samples of the six cello players, six short music excerpts (approximately 3–7 seconds long) representing three different music styles were selected from the database described in Chapter 5. They included:

(i) three Baroque fragments from the 3rd *Cello Suite* by Bach: *Allemande* (bar 2: notes 7–19), *Bourrée II* (bars 1–2: notes 1–15, the version in *piano* dynamics), *Courante* (bars 1–3: notes 1–20); from three available interpretation variants the 3rd one was used in which cellists followed identical bowing indications and were asked to use vibrato only on the melodically most important notes, i.e. climax notes,

(ii) one Romantic fragment from *Élégie* by Fauré (bar 6: notes 22–27),

(iii) and two contemporary fragments from the 20th century *Cello Sonata* by Shostakovich: Movement I (bars 3–5: notes 5–15) and Movement IV (bars 17–20: notes 1–10).

The respective music scores containing the fragments listed above are given in Appendix A. The rationale behind the excerpt selection was to choose fragments which: best capture tone differences of the players in various musical contexts, are concise in terms of length to reduce the cognitive load on listeners, and where possible, break perception of phrase, which was intended to encourage the listeners to focus specifically on tone quality dissimilarities rather than on differences in the musical phrase execution between the cellists.

To enable perceptual comparison of the cellists' tones, the selected audio files were extracted only from the players' ambient microphone recordings made on Cello1 (for details see Chapter 5). Finally, the resulting 36 music samples were manually equalised in terms of loudness.

Listening tests were carried out in quiet room conditions and the stimuli were presented to all participants at a fixed comfortable listening level using the same laptop and high quality headphones.

6.3.2 Participants

Twenty Polish speaking experienced musicians participated in the experiment: ten cellists, five violinists, three viola players and two pianists (12 females and 8 males). The age range was 30–55 (M = 42.65, SD = 7.41). Fifty percent reported that have been working professionally for 15 to 24 years, 35% – between 5 and 14 years and remaining 15% have been in the profession for 30 to 34 years. The period of study in a music academy was not included. Amongst twenty participants there were three music academy teachers, thirteen music school teachers and ten orchestra players (the numbers overlap as six persons worked in at least two job roles).

A rationale behind employing also other string players such as violinists and viola players, instead of just focusing on cellists (considered as "expert" listeners), was that all three instruments, being members of bowed string family, share the same bowing technique principles crucial for production of a good quality tone. The violinists and viola players chosen for this study had long-term experience of performing in a string quartet or other string ensembles in which timbre homogeneity across instrument parts is fundamental for clarity of the harmonic structure. Therefore, their ability to evaluate cello timbre was unquestionable and regarded as an expert one. If timbral differences between analysed tone samples are evident and can be perceived by cellists they are likely to be also perceived as such by other expert musicians, which can help to generalise the findings. With a similar rationale in mind, the two pianists were asked to participate in the tests as they both were music academy piano teachers with extensive experience of accompanying cello players of all levels and performing piano parts across the entire cello repertoire. The effect of major instrument on participants' ratings was tested before commencing any further

analysis (see Section 6.4.1).

6.3.3 Procedure

The participants' task was to rate perceived timbre dissimilarity between two versions of the same music excerpt. The two versions were played one after another with 500 ms of silence between them. Each time, the versions came from two different cellists however through the course of the whole experiment listeners were intentionally not informed that they were comparing different performers or that music samples were recorded on the same cello.

Pairs of cellists were presented to the listeners in a random order across randomised music fragments. The players' order within each pair was also randomly permuted, i.e. whether Cellist 1 was played after Cellist 2 or the other way around. In total, there were 90 pairs to rate (6 music excerpts x 15 pairs per excerpt in order to compare 6 players with each other). Timbre dissimilarities between the cellists were rated on a continuous scale from 0 to 10 where 0 indicated *No difference* and 10 *Very different* (*Nie ma różnicy* and *Bardzo różne* in Polish).

In the second part of the test, participants were asked to evaluate the difference between the players in each pair using verbal attributes *bright*, *rough* and *tense* (*jasny*, *szorstki* and *napięty* in Polish). Their task was to mark which of the two presented samples sounds brighter, rougher or tenser. They could rate one, two, all three attributes or none. In this way, they were free to decide if any of the attributes was applicable for the evaluated pair.

Verbal attribute ratings were recorded as ternary votes ([-1], [1] or [0]), where [-1] indicated that it was the first player in the pair who sounded brighter or rougher or more tense, a [1] vote was attributed to the second player in the pair when his tone was perceived as brighter, rougher or tenser, and a [0] score for any of the attributes was used to mark its irrelevance.

Apart from evaluating timbre differences, listeners were asked to express their preference for one of the versions in each pair. The preference was marked

on a bipolar continuous scale from -5 to 5 where negative numbers indicated preference for the first player in the pair and positive numbers preference for the second, [0] was reserved for the case when neither was preferred. The extremes and the middle point of the scale were annotated with labels *Strongly prefer the 1st*, *Strongly prefer the 2nd* and *No preference* respectively (*Zdecydowanie 1-sza*, *Zdecydowanie 2-ga* and *Nie mam zdania* in Polish).

Each listening test was preceded by a short training session in which participants familiarised themselves with the interface and the task. To give them an overview of possible timbral and stylistic variants in the stimuli a set of 15 (out of 36) randomly selected music samples was also played as a part of training.

Both, the experimental procedure and graphical user interface were created and operated in the Matlab environment.

6.4 Results and discussion

6.4.1 Effect of being a cellist or non-cellist on cello timbre perception

When selecting participants for the perceptual study it was anticipated that for other string players and pianists who specialised in cello repertoire, the ability to evaluate cello timbre would not generally differ from that of the cellists. With the ratings collected it was now necessary to check whether "being a cellist" or "non-cellist" actually affected the way the listeners rated perceived differences in cello tone samples.

For this, two separate Multivariate Analyses of Variance (MANOVAs) were carried out on the entire dataset with Instrument as the independent grouping variable and dissimilarity ratings of fifteen pairs of cellists as 15 dependent variables. In the first MANOVA, the groups of the Instrument variable were defined as Cello and Other and in the second analysis as Cello, Violin, Viola and Piano to enable more detailed investigation into perceptual differences between various instrumentalists in case such differences occurred. Prior to the analyses, the data was screened for the assumption of normality and outliers but no

deviations were detected.

The two MANOVAs yielded "not significant" results (Wilks' $\Lambda = .82$, $F(15, 104) = 1.47$, $p < .128$, $\eta^2 = .17$[1] and Wilks' $\Lambda = .68$, $F(45, 303.8) = .94$, $p < .58$, $\eta^2 = .12$[1]respectively) indicating that, across all evaluated pairs of cello tone samples, other instrumentalists did not differ in their ratings when compared to cellists. The follow-up ANOVAs for each dependent variable (each pair of cellists in evaluation) with a Bonferroni adjustment of alpha levels for multiple tests ($p < .05/15 \approx .0034$) also proved that there was no significant differences in ratings between the groups.

Finally, a series of mixed design ANOVAs was conducted to test whether being a cellist or non-cellist might have affected dissimilarity ratings of the same pair of cellists when compared in various musical contexts. In this particular design, the dissimilarity ratings of each pair were treated as a single dependent variable, six music excerpts were the levels of Piece within-subjects effect (since each participant had to rate the same pair of cellists on six different occasions) and Instrument was a between-subjects factor with two (Cello and Other) or four (Cello, Violin, Viola and Piano) levels respectively. In total, thirty ANOVAs were carried out, two per each pair of cellists but no significant interaction between the Piece and Instrument variables was found.

To summarise, the above analysis demonstrated, as it was predicted, that instrumentalists other than cellists can exhibit similar ability to evaluate cello timbre hence they can be admitted as "expert" listeners together with cellists.

6.4.2 Inter-rater reliability analysis

As the next step, timbre dissimilarity ratings were examined in terms of the inter-rater reliability (IRR) (for an excellent tutorial on this topic refer to Hallgren, 2012). In particular, the degree to which the participants agreed in their ratings across pairs of cellists was assessed using a two-way random, absolute

[1] *partial* η^2 reported; in one-way designs *partial* η^2 equals η^2 therefore the reported values indicate 17% and 12% of total variance explained respectively (Tabachnick and Fidell, 2007)

Table 6.1: Measures of internal consistency and agreement between the participants' ratings for each music fragment.

Music Excerpt	**Mean Inter-Subject** r	**Cronbach's alpha**	**Absolute agreement**	F **statistics**	**Sig.**
Allemande	.17	.79	.73	$F(14,252) = 4.67$	.0001
Bourrée	.13	.70	.66	$F(14,210) = 3.37$	.0001
Courante	.31	.89	.83	$F(14,266) = 8.87$	.0001
Élégie	.12	.65	.54	$F(14,210) = 2.85$	.001
Shost1	.16	.76	.65	$F(14,238) = 4.24$	.0001
Shost4	.32	.89	.85	$F(14,238) = 9.17$	.0001
Overall	.26	.86	.82	$F(89,1513) = 7.08$	.0001

agreement, average-measures intra-class correlation (ICC) (McGraw and Wong, 1996).

To improve reliability measures, for each music fragment except for Courante, one to four participants were removed due to their negative subject-total correlation. This coefficient measures how the rating of each respondent correlates with the total rating across all respondents from which his own input is subtracted. The resulting mean inter-subject correlation, Cronbach's alpha, intraclass correlation coefficient for absolute agreement and respective F statistics are shown in Table 6.1.

According to commonly-cited thresholds for ICC values (Cicchetti, 1994), at least a good level of agreement (ICC $>$ .60) was observed for all excerpts except for Élégie, indicating that timbre dissimilarity was evaluated quite similarly across raters. As for Élégie, only fair agreement between the ratings suggested the existence of considerable discrepancies in perceived dissimilarities that could in turn reduce statistical power of subsequent analyses. However, when calculated across all excerpts combined together (with two participants removed), the resulting overall ICC was well above .75 (see Table 6.1), indicating an excellent agreement between the raters. Therefore, after careful consideration, the Élégie ratings were retained for the experiments to follow and the outcomes were interpreted with caution.

Table 6.2: Goodness of fit measures for MDS solutions for six music excerpts.

Music excerpt	**Normalized Raw Stress**	**Kruskal's Stress-1**	**Dispersion Accounted For**
Allemande	.006	.08	.99
Bourrée	.005	.07	.99
Courante	.000	.005	1.00
Élégie	.007	.08	.99
Shost1	.014	.12	.99
Shost4	.000	.01	1.00
Overall	.006	.08	.99

6.4.3 Perceptual mapping of the players

The individual dissimilarity ratings were organised into 6x6 distance matrices collating perceived timbral differences between the six cellists. Since no significant differences in cello timbre perception were detected between groups of raters and the inter-rater agreement was at least at a good level, the individual distance matrices were then averaged across raters for each music excerpt as well as across all excerpts, resulting in seven aggregated dissimilarity matrices. Non-metric multidimensional scaling (MDS) was employed to obtain perceptual mappings for each music fragment and for six music fragments combined together (all models were computed using PROXSCAL in the IBM SPSS software implementation). Following the MDS outcome diagnostic steps, including examination of the Shepard diagrams for residuals, the scree plots (the stress values plotted against the number of dimensions) for dimensionality and the Kruskal's Stress-1 values for goodness-of-fit evaluation, only 2-D solutions were considered for further analysis. Table 6.2 summarises respective measures of the models' overall fit. Figure 6.1 displays yielded timbral maps in all six music excerpts and the overall timbral map is presented in Figure 6.2.

In general, two-dimensional solutions were more reliable in fitting perceived timbre dissimilarities between the players into a common space. The respective Kruskal's Stress-1 values for all music fragments were below the accepted

threshold for a good fit, i.e. $\leq .15$ (Kruskal and Wish, 1978), and the dispersion measure, i.e. the amount of variance explained by the model, was $\geq .99$ in all cases. Amongst one-dimensional solutions, only the Courante and Shost4 models met the requirements for a good fit.

When looking at the obtained two-dimensional perceptual maps, it is clearly visible that the players occupy distinct positions within the space, however, the distribution patterns vary depending on musical style and/or genre. One exception can be found in the Shost4 map (Figure 6.1f) where Cellists 2 and 4 are located very close to each other, which may suggest that their tone samples in this particular musical context were perceived rather as similar. From the inspection of the MDS representations, one can observe that in the excerpts with mixed articulation (mostly *staccato* notes with short *legato* passages) and performed in *mezzo forte*, such as Allemande, Courante and Shost4, timbre dissimilarities between the cellists seem to a greater extent distributed along just one dimension, in comparison with Bourrée, Élégie and Shost1 which contain only *legato* notes played in *mezzo piano* or *piano*. In the latter cases, points representing the cellists in the perceptual spaces are more evenly distributed along the two dimensions, which may indicate that at least two factors or attributes played an important role in the task of discriminating between the cellists' timbres.

A clear two-factor perceptual space emerges also from Figure 6.2, representing the dissimilarity ratings aggregated over all six musical contexts. One can notice that Cellists 1 and 2 as well as Cellists 3 and 6 seem to cluster into two subgroups based on some tone attributes, although within the clusters the points are well separated from each other. The dimensions themselves are, however, as in case of any MDS solution, meaningless. Moreover, the actual orientation of axes is arbitrary and can be subjected to rotation, translation, dilation, or reflection to facilitate their interpretation (Borg and Groenen, 1997; Borg et al., 2013). In fact, interpreting a dimension in an MDS map refers to the task of assigning an attribute or property (based on some prior knowledge about

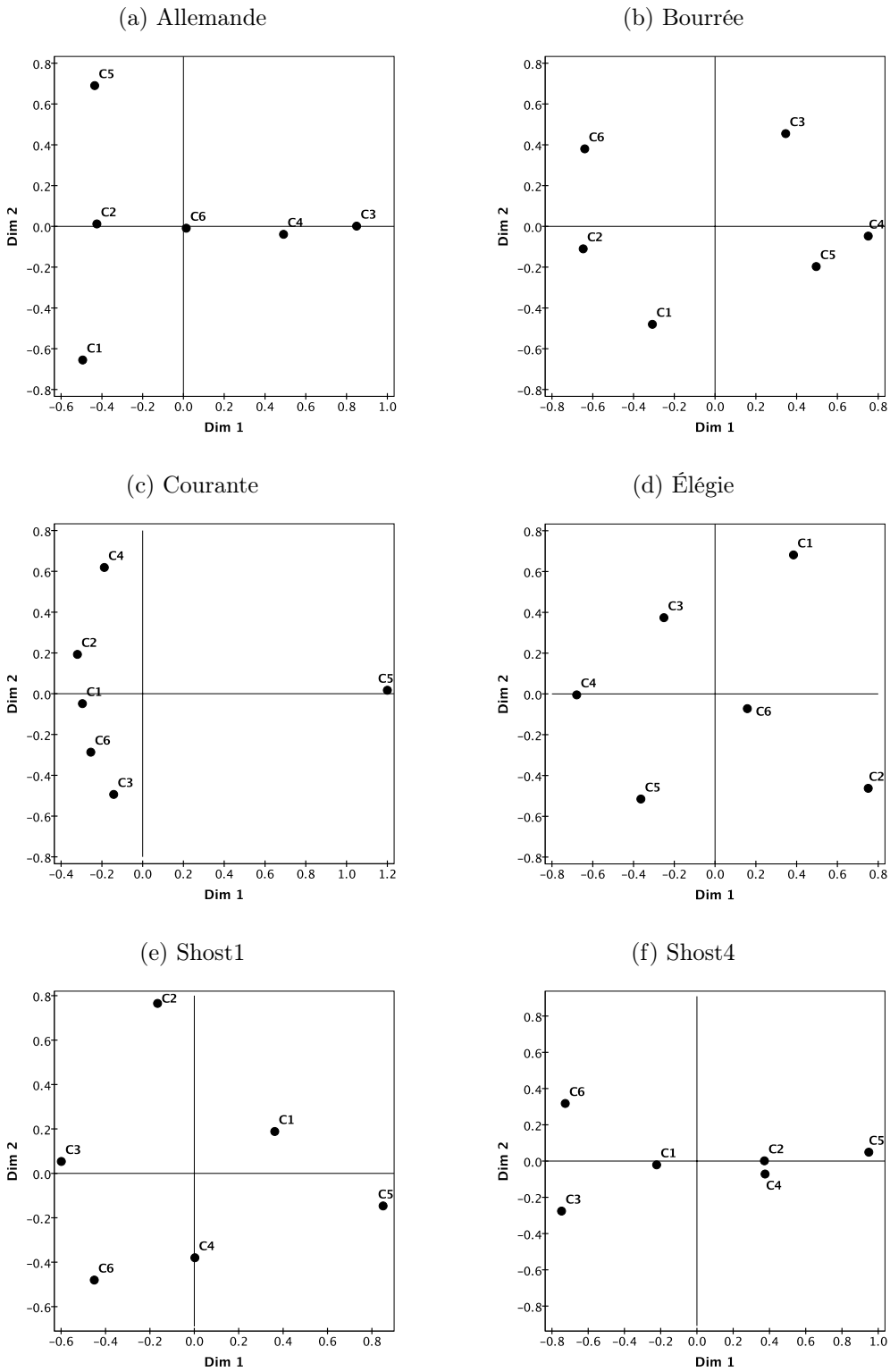


Figure 6.1: 2-D MDS maps of timbre dissimilarity ratings for each music excerpt.

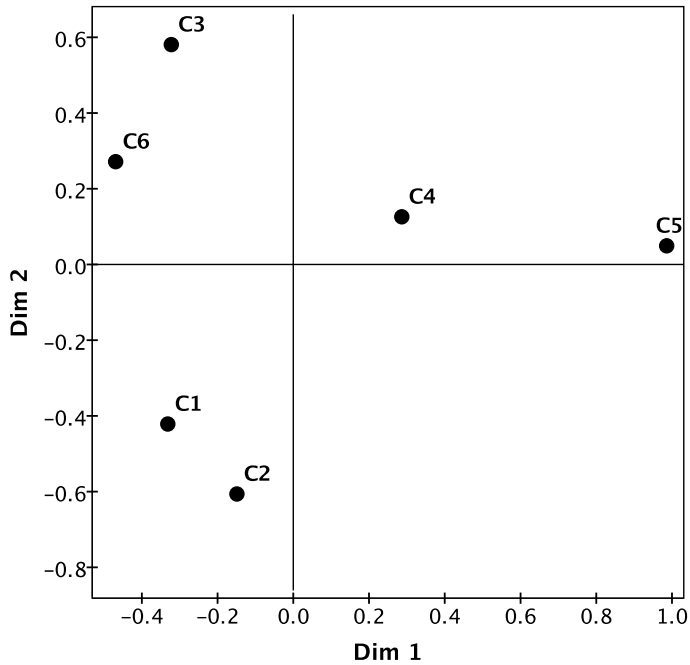


Figure 6.2: 2-D MDS map of timbre dissimilarity ratings averaged across six music excerpts.

the items represented in the map) such that the items can be qualitatively or quantitatively ordered according to that property.

At this point of the analysis no *a priori* information about specific features or attributes which might discriminate the cellists' tones was available to preliminarily interpret the dimensions. Instead, the analysis of verbal attribute ratings was designed to identify which attributes or characteristics could be assigned to the MDS dimensions and "explain" the perceived dissimilarities between the cellists. It is also important to remember that for each MDS map (when comparing music excerpts for example) these attributes may be different.

Although without currently knowing what the dimensions might particularly mean and whether the orientation of the axes will need further transformations to aid the interpretation, yet some interesting observations can be reported. For instance, in Courante (Figure 6.1c), the tone sample of Cellist 5 seems to be highly distinguished from the other cellists' tones based on Dimension

1, while in turn their timbres seem to be well differentiated along Dimension 2. In Allemande (Figure 6.1a), the second dimension contrasts Cellists 5 and 1, as they occupy its positive and negative extremities respectively, with the other cellists who seem to be distinguished from each other by Dimension 1. In Shost4 (Figure 6.1f), it is mostly the first dimension which separates the cellists' tones except for Cellists 3 and 6 who are differentiated by Dimension 2. As for the remaining three music excerpts and the overall dissimilarity map the division of the tone samples between the dimensions is less clear. For example, in Bourrée (Figure 6.1b) and Élégie (Figure 6.1d), the first dimension might separate Cellists 3, 4 and 5 from Cellists 1, 2 and 6, and this could be also true for the overall map (Figure 6.2) if to rotate the axes about 45 degrees anticlockwise. Similar transformation of the axes might be needed in the case of the Shost1 excerpt (Figure 6.1e) if the first dimension is meant to distinguish Cellists 1, 2 and 5 from Cellists 3, 4 and 6.

Nevertheless, the most important outcome of the MDS analysis is that it confirms that the cellists, even if recorded in identical music pieces and on the same instrument, did sound different and that these differences were audible across various musical styles and genres.

6.4.4 Mapping the players into the verbal attribute space

Correspondence analysis (CA) can be used to interpret the dimensions obtained from MDS. This exploratory technique (Greenacre, 1984; Benzécri, 1992) is designed to examine the association between objects and a set of descriptive characteristics or attributes specified by the researcher. Similarly to factor analysis (FA) or principal component analysis (PCA), CA attempts to explain the variance in a model and decompose this variance into a low-dimensional representation. However, CA explores categorical data and determines which category variables are associated with one another, whereas the two former techniques, while designed for interval measurements, extract which variables explain the largest amount of variance in the data set (Doey and Kurta, 2011).

Table 6.3: Measures of inter-rater agreement in verbal attribute ratings across the six music excerpts.

Attribute	**Mean Inter-Subject** r	**Krippendorff's alpha**
Brighter	.29	.19
Rougher	.24	.15
Tenser	.16	.10

In the current study CA was employed to answer the question whether perceived dissimilarities between the six cellists can be characterised in terms of verbal attributes and visualised in a verbal attribute space.

The ratings of the three timbre attributes were first checked for inter-rater reliability across the six music excerpts using Krippendorff's alpha, recently proposed by Hayes and Krippendorff (2007), as the standard reliability measure. The obtained alpha values were, however, disappointingly low (the alpha exceeding .60 is the commonly accepted minimum level of agreement between raters) and did not much improve after removing participants with the negative subject-total correlations from the Brighter and Tenser categories (see Table 6.3). This result posed a question whether subsequent correspondence analysis of the verbal attributes should be carried out. Considering the purely exploratory nature of CA and the fact that this technique mainly provides a graphical representation of cross tabulations (contingency tables) in order to uncover relationships among categorical variables (Yelland, 2010), it was decided to proceed with the analysis while keeping in mind that any possibly interesting outcomes should be interpreted with caution.

Two separate experiments were designed on the verbal attribute data. The first one aimed at exploring the relationship between semantic labels and perceived tones of the six cellists throughout the whole dataset, i.e. regardless of the music performed. The goal of the second experiment was to examine the associations between semantic labels and the cellists' timbres comparing different music excerpts.

To carry on with analyses, the verbal attribute ratings were collected in two

Table 6.4: Contingency table of timbre attribute votes across six music excerpts. The highest number of votes for each attribute is marked in bold.

	Timbre Attribute						
Cellist	Brighter	Rougher	Tenser	LessBright	LessRough	LessTense	Total
1	229	168	**230**	164	247	143	1181
2	**266**	228	185	132	198	206	1215
3	224	282	223	156	153	144	1182
4	146	137	145	243	282	213	1166
5	80	135	103	**345**	**298**	**285**	1246
6	237	**330**	229	142	102	124	1164
Total	1182	1280	1115	1182	1280	1115	7154

contingency tables of 6x6 and 36x6 sizes. For the first table, rows and columns represented Cellist and Timbre Attribute variables respectively. In addition to obtaining votes on players whose tones were perceived as brighter, rougher or tenser, cellists who sounded less bright, rough or tense in comparison were also recorded to complement the acquired information. Subsequently, the Attribute variable was split into six categories (see Table 6.4). For the second table, each Cellist category was split further into 6 subcategories to represent timbre attribute votes in each of the six music fragments.

In order to correctly understand a potential relationship between the perceptual dimensions and semantic labels, it needs to be recalled that the listeners' judgements were relative, i.e. made by weighting the presence of a particular attribute in the sound between two players instead of rating it on an absolute scale. Hence, each cell of Table 6.4 simply contains the number of times the player was perceived as having the tone feature either more or less pronounced.

Based on the number of votes collated in Table 6.4, one can conclude that the timbre of Cellist 2 was most often perceived as Brighter, Cellist 6 as Rougher, and Cellist 1 as Tenser. In the same time, according to the ratings, the tone of Cellist 5 was the least bright, rough and tense, followed by the brighter, slightly rougher, and tenser tone of Cellist 4. The timbre of Cellist 3 was placed around the midpoint of the attribute ranking, being not particularly bright, but showing some tendency towards roughness and to a lesser extent, towards tension.

Table 6.5: Summary of two correspondence analyses.

Model	**Dimension**	**Inertia [%]**	**[%] of Inertia accounted for**
6x6	1	8.8	84.6
	2	1.2	11.3
	Total	10.4	95.9
36x6	1	13.9	67.3
	2	4.6	22.4
	3	1.4	6.8
	Total	20.7	96.5

Two correspondence analyses (with the symmetrical normalisation option) were carried out using the above mentioned contingency tables. The χ^2 measure of association was significant in both cases ($\chi^2(25) = 747.16$ and $\chi^2(175) = 1478.45$, $p < .0001$) justifying the hypothesis that the two variables (Cellist and Timbre Attribute) are related. The respective measures of the association's strength; Cramér's V = .145 and .203 (Cramér, 1999), were both significant at $p < .0001$, but being less than 0.3 indicated that although the observed relationship was not due to chance, it was also not strong.

The two CAs yielded five-dimensional solutions (the maximum number of dimensions is one less than the smaller of the number of rows or columns). Similarly to PCA, the eigenvalues, called Inertia in CA, represent the percentage of variance explained by each dimension and reflect its relative importance, with the first dimension always being the most important. In the current solutions two and three dimensions were retained accounting for most of the inertia in the models (95.9% and 96.5% respectively).

From Table 6.5 it can be seen that the amount of inertia explained is small (10.4% and 20.7% in total), once again indicating the fact that the correspondence between Cellists and perceived attributes of their tone, while significant, is rather weak. However, the relevant literature does not provide precise recommendations on how to accurately evaluate CA models in terms of the amount of inertia accounted for. Mazzocchi (2008), for example, states that the total inertia above 20% is regarded as acceptable for an adequate representation,

while Murtagh (2005) suggests that values of about 50% of the captured inertia are not uncommon in CA and do not necessarily lead to an inadequate model. Though a common recommendation is that even in cases when a correspondence map is of a low quality, i.e. when it explains just 20% or less of the inertia, it is still capable to show the strongest patterns evident in the data.

As for the current solutions, the smaller amount of inertia explained is likely due to higher levels of disagreement between the raters, as indicated by Krippendorff's alphas. On the other hand, the result may have also its explanation in the limited number of attributes selected for describing timbre dissimilarities, capable only of capturing certain acoustic features from the players' rich timbral palette. This was partly confirmed in the respondents' comments. For some compared pairs of players it was reported that none of the proposed attributes was adequate to characterise the perceived difference in timbre.

Figures 6.3 and 6.4 present a 2-D visualisation of each CA solution. The attribute-based maps show the relative proximities of both players and verbal attributes. Categories that are similar to each other appear close to each other in the plots. Therefore, it is easy to see which categories of variables Cellist and Timbre Attribute are similar to each other or which categories of the two variables are related. However, the distance between two points representing two different variables on the map cannot be interpreted directly, i.e. in terms of the Euclidean distance. In order to compare two such points, i.e. one of the Cellist category with one of the Timbral Attribute category, one needs to draw a line from each object to the center of the map (which represents the average profile or barycentre of each category) and assess the angle between the two lines. If the angle is very small it suggests that the categories are (positively) associated and, if they are both far from the center of the map, their association is relatively strong. If the angle is more than 90 degrees, it indicates a negative association while a right angle suggests that no relationship exists.

The association between the two objects can only be quantified in terms of relative frequencies (Yelland, 2010), i.e. relative to the average category profile

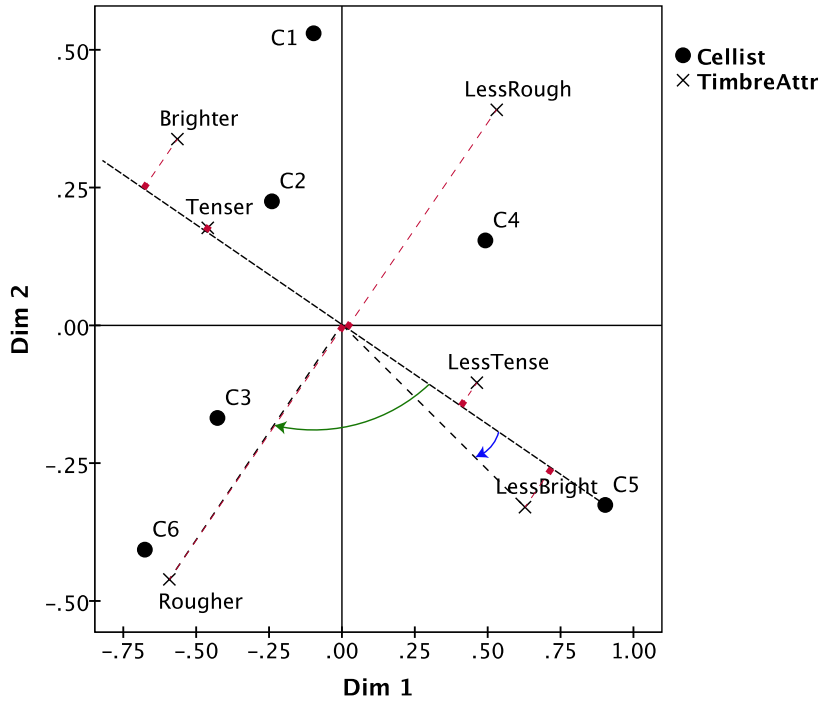


Figure 6.3: 2-D correspondence map for the 6x6 contingency table. The black dashed line drawn through the origin and Cellist 5 is used to determine which semantic labels were most associated with his timbre and to assess how often (relatively) each label appeared in his tone ratings (intersections with red dotted perpendiculars). The origin of the map represents the average or barycentre of both the Cellist and Timbre Attribute variables.

represented by the centre of the CA map. For example, the tone of Cellist 5 is strongly related to the LessBright attribute and was perceived as less bright more often than the average for all the cellists, as indicated by the acute angle between the two lines (marked by the blue arrow in Figure 6.3). To examine whether one label appeared more often than another in his timbre ratings, one needs to project the Timbre Attribute points onto a line crossing through the origin and the point corresponding to the cellist in question (as illustrated in Figure 6.3) and examine the intersections. From the ordering of the labels' intersection points on the Cellist 5 axis, from the closest to the furthest, it can be seen that the LessBright label was assigned to him more frequently than any other attribute. Conversely, his sound was perceived as Brighter much

less often than the average and the attribute Brighter was the least frequently assigned to his tone. Interestingly, the angle between the lines of Cellist 5 and either of the Rougher or LessRough attributes is close to a right angle suggesting no correspondence between the cellist and the two labels. Since the respective perpendiculars dropped from both attributes intersect the Cellist 5 line around the origin, it would also suggest that attributes Rougher and LessRough were assigned to his tone less frequently than, for example, the LessTense label but relatively more often the Tense attribute. This somehow contradicts the rating data collated in Table 6.4 according to which Cellist 5 was rated as (note the attributes' order) the least Bright, Rough, and Tense amongst the cellists. Discrepancies between a contingency table and respective CA solution are often due to the fact that as a large amount of information is compressed, this causes distortions in the resulting CA model (Q-Software, 2015). It is also worth to remember that the solution depicted in Figure 6.3 accounted only for 10.4% of inertia.

Similar analyses carried out for each cellist suggest that, for example, Cellist 6's tone was mostly associated with the attribute Rougher, as was the tone of Cellist 3 except that his association with the label was less pronounced (the representing point lies much closer to the barycentre). Also, the attribute Rougher was relatively most frequent in the ratings of the two cellists compared to the other labels. As for general Brightness and Tension attributes, they seem to occur in both cellists' samples no more than at the average level. On the other hand, labels Brighter and Tenser seem most closely associated with the tone of Cellist 2, with being Brighter as most frequent. In the case of Cellist 1, his respective point does not lie in close proximity to any particular attribute, which could indicate that his tone's features were not accurately captured by the available verbal attributes. Cellist 1's tone can be only tentatively labelled as Brighter, Tenser and LessRough. In regard to Cellist 4, his tone was mostly associated with the LessRough attribute and it was also the most frequent label in his timbre ratings, followed by, according to the CA map, the LessTense and

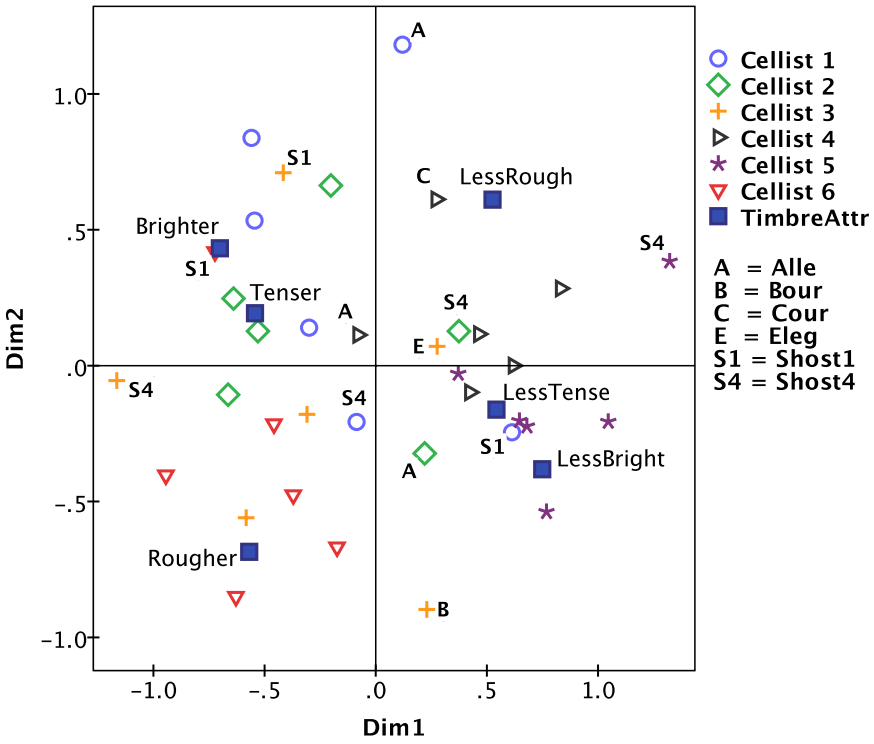


Figure 6.4: 2-D correspondence map for the 36x6 contingency table. To preserve readability of the chart, only samples considered as less characteristic for each cellist's timbre were annotated with music excerpt labels.

LessBright labels. Once again it needs to be pointed out, that the resulting from CA graphical representation of the data deviates, in some cases quite substantially, from what can be inferred from the frequencies in Table 6.4 about the association between categories of the Cellist and Timbre Attribute variables.

More detailed illustration of how tone samples of the cellists were perceived in terms of timbre attributes is provided in Figure 6.4. One can see that tone samples of Cellist 5 indeed cluster mostly around the LessBright and LessTense labels with exception of the Shost4 excerpt, which was labelled rather as LessRough. The samples of Cellist 4, in majority attributed as LessTense, were also perceived as LessBright and LessRough (except for Allemande). Cellist 6's tone samples, as it already appeared in Figure 6.3, most strongly corresponded to the Rougher attribute, with one exception for Shost1 which was more often labelled as Brighter and Tenser.

The observed associations of Timbre Attributes with tones samples of Cellists 1, 2 and 3 were much more diverse, suggesting that these players might have adjusted their timbres more deliberately to differentiate between various music styles or genres. For example, the fragments of Bourrée, Courante and Élégie performed by Cellist 1 were most frequently perceived as Brighter and Tenser in contrast to his Shost1 which was LessBright and LessTense, while his Allemande and Shost4 were rather attributed as LessRough and Rougher respectively. The Courante, Élégie and Shost1 interpretations of Cellist 2 were most often associated with the Tenser and Brighter labels, unlike his Allemande and Shost4, while his Bourrée was equally weakly associated with the Brighter and LessRough attributes.

The timbre of Cellist 3 seemed to be more elusive in its characteristics than it can be described with just three verbal attributes, as evidenced by the wider distribution of his tone samples across different timbre attributes when compared to the mappings of other cellists. For instance, his Allemande was most often related to the Rougher label and so his Courante, though for Courante this association appeared much weaker. At the same time, his two Shostakovich fragments were more frequently assigned as Brighter and Tenser, his Élégie as LessRough, LessBright or LessTense, and his Bourrée somewhat between Rougher and LessBright, however, in these cases the association with the attribute was not strong.

To interpret the dimensions in a correspondence map, one should look first at so-called *contributions* or loadings onto each factor (dimension) that CA provides for each element of the map. Contributions larger than the average usually indicate those elements which are important for a given factor (Abdi and Williams, 2010). In the current study, the second CA solution, regarded as a more accurate representation of the verbal attribute ratings, was employed for interpreting the dimensions. Subsequently, respective contributions of the Timbre Attribute categories were investigated, as the aim of this study was to map the cellists' timbres into the verbal attribute space. The largest loadings for

most positively and negatively located points were obtained for labels LessBright and Bright respectively, hence the first dimension can be tentatively named Brightness. The second dimension can be then named Roughness since the greatest contribution to this dimension were the Rougher and LessRough labels.

At this point, it needs to be recalled that while some conclusions have been drawn about the correspondence between Cellists and Timbre Attributes, one must keep in mind that the presented models explained only up to 20.7% of the variance in the data, thus can only be treated as exploratory aids. Nevertheless, the CA solutions provided the first insight into the meaning of the MDS dimensions and possible semantic labels for the acoustical correlates discussed in Chapter 7.

6.5 Summary

Timbre dissimilarity ratings of six cellists' tone samples were collected and subjected to the MDS procedure to obtain perceptual maps. Two-dimensional MDS solutions revealed that the players' tones were perceived by the listeners as distinct from each other within and across various musical contexts, except for Cellists 2 and 4 in the Shost4 excerpt where they seemed to sound quite similar.

Verbal attributes such as *bright*, *rough* and *tense* were used to characterise timbre dissimilarities between the cellists. Correspondence analysis of the attribute ratings was employed to obtain perceptual mappings of the players into the verbal attribute space. Although the amount of variation in the data explained by the CA solutions was not high and the results can only be considered as auxiliary, they seemed to identify two perceptual dimensions, namely Brightness and Roughness, which can partially explain in qualitative terms timbral differences between the players.

Chapter 7

Acoustical correlates of the perceptual dimensions

The two-dimensional MDS solutions discussed in Chapter 6 provided perceptual positioning of the cellists within timbral spaces across six various music styles and genres. In each space the players were well separated from each other, implying that their timbres were perceived as distinctly different. Verbal attribute ratings on the other hand gave a preliminary insight into the meaning of the resulting perceptual dimensions.

In this chapter, subsets of preselected acoustic features which best describe varying timbral characteristics of the players are identified. Factor analysis is employed to reduce the dimensionality of feature vectors and obtain compact acoustic representations of each cellist. The relation between acoustical and perceptual dimensions is then studied to reveal which spectro-temporal components of a cellist's tone play the most important roles in perceptual discrimination.

7.1 Introduction

Employing acoustic features seems a straightforward way to capture salient characteristics of the players and explain the nature of existing dissimilarities. However, the majority of timbre studies (as discussed in Chapter 2) concentrated their efforts on finding audio descriptors that can efficiently describe acoustical properties of different musical instruments, and very few investigated features

which can also discriminate tones of different players performing on the same instrument (e.g. Fitzgerald, 2003). None of them dealt with performer-dependent aspects of timbre on a bowed string instrument.

The lack of relevant references made the starting point for further analysis somewhat difficult since the choice of adequate audio features was crucial. After careful consideration, it was decided to give a closer look to the audio feature set proposed and exploited by Alluri and Toiviainen (2010) and Eerola et al. (2012). In both studies, significant relationships were found between spectro-temporal descriptors and perceptual dimensions of either verbal attributes of polyphonic timbres or affect ratings of various instrument sounds. Despite the fact that the sound stimuli used did not come from the same instrument class, both the methods employed and the findings were useful for developing a methodology for the current study.

To obtain a reliable and exhaustive acoustic representation of each performer, which then can be mapped against dissimilarity ratings, was another important issue to resolve. One needs to remember that participants in the perceptual study listened to and evaluated each tone sample, i.e. short music fragment, as a whole, regardless of the number of notes it comprised, whether six or twenty. They, however, could have paid attention to individual timbral details in the passage to form their final judgements. In more technical terms, any change in pitch (which may also be associated with a change of the string being played), articulation (various bow strokes being used) and dynamics (varying intensity levels) within a passage affected an instantaneous spectro-temporal profile of the sound and would need to be accounted for in the extracted features. To fulfil this requirement, it was necessary to compute relevant audio descriptors at the note level.

7.1.1 *A priori* remarks

In regard to variation in pitch, the pitch range covered by the music samples (details in Table 7.1) slightly exceeds the first two octaves, comprising low to

Table 7.1: Pitch and frequency range of the sound stimuli.

Music excerpt	Number of notes	Pitch range	Frequency range [Hz]
Allemande	13	C2–G3	65.41–196.00
Bourrée	15	D3–E♭4	146.83–311.13
Courante	20	C2–D4	65.41–293.66
Élégie	6	G3–E♭4	196.00–311.13
Shost1	11	E3–F4	164.81–349.23
Shost4	10	A2–E♭4	110.00–311.13

middle frequencies which are the most characteristic for cello timbre. Within this range one would expect the cello sound to be warm, rich in harmonics and vibrant or velvety, depending on music context. Exploring how each of the cellists manipulated the given instrument timbre to shape musical phrases and how the resulting quality of tone can be described in terms of acoustic features was one of the goals of the following experiments.

In terms of minute changes in dynamic levels across notes within each analysed passage, they were considered to be an integral part of timbral manipulations executed by a player to shape musical phrases and as such, to be captured in acoustic features extracted at the note level.

7.1.2 Research goals

The study here undertaken aimed at:

- identifying salient acoustic features that can help to differentiate between performers; in particular, evaluating the set of spectro-temporal descriptors derived from (Alluri and Toiviainen, 2010) for the purpose of discriminating between the players performing on the same instrument
- finding low-dimensional acoustic characterisations of performers
- investigating the relationships between perceptual and acoustical dimensions of a player's timbre.

7.2 Method

A series of experiments was designed to address each of the research goals stated above. Firstly, acoustic feature extraction and selection was needed to obtain multidimensional spectro-temporal characteristics of the cellists at the note level for the six music excerpts. Secondly, a dimension reduction technique, factor analysis, was employed to uncover an underlying structure of acoustical dimensions and form a compact acoustic representation of each player. Thirdly, low-dimensional representations of the players were compared in order to find timbral dissimilarities by means of correlation analysis and MANOVA tests. Finally, respective acoustical and perceptual dimensions were correlated to find which acoustic factors may best explain perceived timbre differences amongst the cellists.

7.2.1 Acoustic feature extraction

The main objective of the initial feature selection was to focus on those features which have already proved to be effective in discriminating various instruments' timbres and which might be able to discriminate between cellists' tones where the same fragments of music were recorded on the same instrument in identical acoustical conditions. From the parameters characterising spectral, temporal and spectro-temporal aspects of timbre proposed by Alluri and Toiviainen (2010) and Peeters et al. (2011), a set of features was selected. They included primarily spectral and spectro-temporal parameters following that temporal descriptors such as Attack Time for example were found inadequate for the task of discriminating between different cellists' tones (Chudy and Dixon, 2010). The final set comprised twenty four time-varying descriptors computed using the magnitude STFT plus one time-domain parameter. The full list of the features and their descriptions is given in Appendix B.

Thirty six music samples (6 per player) used previously in the perceptual experiment were segmented into notes. In total 450 notes were obtained (75 per player) and subsequently subjected to a feature extraction procedure. For

each note 25 features were computed using 25-ms frames with 75% overlap (MIRtoolbox 1.5, Lartillot et al. (2008) and Timbre Toolbox 1.4, Peeters et al. (2011) were used, both toolboxes implemented in the Matlab environment). The median value across all frames was taken to form a compact representation of each feature. The median is considered a more robust measure of central tendency and better suitable for summarising any time-varying audio descriptor in a single value (Peeters et al., 2011). The resulting feature datasets consisted of 13, 15, 20, 6, 11 and 10 25-feature vectors per cellist for Allemande, Bourrée, Courante, Élégie, Shost1 and Shost4 excerpts respectively.

7.2.2 ANOVA-based feature selection

Before employing factor analysis to define acoustical dimensions of the timbre spaces a pruning of the 25 acoustic features was necessary. Bearing in mind that the ability to discriminate between players is the main goal, the aim of this step was to choose those descriptors which demonstrate strong and significant variability across cellists and weak or moderate variability across pitches. However, it was anticipated that there might be a strong interaction between cellist and music excerpt affecting acoustic feature values (due to changes in tempo, articulation and dynamics).

To first investigate the differences across the players and across the excerpts and whether the interaction between the two factors is significant, a mixed between-within-subjects ANOVA was conducted for each acoustic feature using Cellist (6 levels) as a within-subjects variable and Piece (6 levels) as a between-subjects variable. In this repeated measures scenario, notes were treated as subjects exposed to six different conditions, i.e. being performed by six different players, and the median values of a measured acoustic feature were treated as scores.

To proceed with the analysis, all features were screened for the assumption of normality. In cases where this assumption was violated the data was adjusted using a Box-Cox transformation (Box and Cox, 1964). The remaining features

were standardised using the Z-score transformation in order to unify the scales across the data.

For all 25 spectro-temporal descriptors a significant main effect of Cellist was found at $p < .0005$ except for SpectVariation ($p < .001$) and SubBand6Flux ($p < .045$). As predicted there was a significant factor interaction for the descriptors at p levels ranging from .0005 to .047 except for SpectVariation for which no interaction was found. This result suggested that all 25 features could be effectively used for factor analysis to define the acoustical dimensions across the six music fragments. However, the observed significant effect of Piece on the feature variations across the players had to be taken into consideration. It might affect a pruning outcome so that for each excerpt a different subset of features would be selected. Therefore, to locate the source of Cellist-Piece interaction according to a standard follow-up design, a one-way repeated measures ANOVA was applied to each of the six music datasets separately. Once again, the analysis was carried out for each descriptor individually and the resulting pruned subsets of acoustic features are presented in Tables 7.2–7.4.

Based on preliminary tests, to improve factorability of the data, further feature pruning was performed resulting in the following features being removed: Roughness and Irregularity from Bourrée, Roughness from Courante and Élégie, Roughness, Irregularity and Rolloff85 from Shost1, and Rolloff95 from Shost4 feature subsets.

7.3 Results and discussion

7.3.1 Factor analysis

With the aim of reducing the dimensionality of acoustic feature vectors representing timbral characteristics of the players, and above all, to find a small set of underlying constructs which can effectively characterise the acoustic data with a minimal loss of information, factor analysis was carried out on the six music excerpt datasets. Principal axis factoring (PAF) was chosen as a factor extraction

Table 7.2: Feature subsets selected in ANOVA for pieces Allemande, Bourrée and Courante.

Music excerpt	**# of notes**	**Features**	**F statistics**	**Sig.**
Allemande	13	HighFreqEnergy	$F(5, 60) = 6.27$	.0005
		Rolloff85	$F(5, 60) = 3.05$	.016
		Spread	$F(5, 60) = 11.50$	.0005
		Skewness	$F(5, 60) = 2.87$	.022
		Kurtosis	$F(5, 60) = 5.64$	.0005
		Flatness	$F(5, 60) = 8.44$	.0005
		SubBand1Flux	$F(5, 60) = 5.79$	.0005
		SubBand2Flux	$F(5, 60) = 2.78$	.025
		SubBand3Flux	$F(5, 60) = 3.47$	.008
		SubBand7Flux	$F(5, 60) = 4.20$	.002
		SubBand9Flux	$F(5, 60) = 2.56$	.036
Bourrée	15	Centroid	$F(2.62, 36.72)^* = 5.21$	.006
		HighFreqEnergy	$F(5, 70) = 8.14$	.0005
		Rolloff85	$F(2.50, 34.94)^* = 7.67$	.001
		Skewness	$F(2.48, 34.76)^* = 4.75$	.01
		SpectEntropy	$F(5, 70) = 11.69$	.0005
		Roughness	$F(5, 70) = 8.87$	.0005
		Irregularity	$F(5, 70) = 8.41$	.0005
		ZeroCrossings	$F(5, 70) = 4.80$	.001
		SpectralFlux	$F(5, 70) = 5.00$	.001
		SubBand1Flux	$F(5, 70) = 15.09$	.0005
		SubBand2Flux	$F(2.41, 33.72)^* = 5.62$	.005
		SubBand3Flux	$F(2.71, 37.96)^* = 17.16$	.0005
		SubBand4Flux	$F(3.27, 45.75)^* = 12.14$	.0005
		SubBand5Flux	$F(5, 70) = 9.94$	.0005
		SubBand6Flux	$F(5, 70) = 9.59$	.0005
		SubBand7Flux	$F(5, 70) = 12.86$	.0005
		SubBand8Flux	$F(5, 70) = 11.63$	.0005
		SubBand9Flux	$F(5, 70) = 4.23$	.002
		SubBand10Flux	$F(2.67, 37.34)^* = 5.58$	.004
		SpectDeviation	$F(5, 70) = 10.47$	.0005
		SpectVariation	$F(5, 70) = 2.72$	.026
Courante	20	Spread	$F(5, 95) = 4.52$	.001
		Roughness	$F(5, 95) = 7.59$	.0005
		SpectralFlux	$F(3.62, 68.85)^* = 4.38$	.004
		SubBand1Flux	$F(5, 95) = 10.15$	.0005
		SubBand2Flux	$F(2.87, 54.62)^* = 7.12$	.0005
		SubBand3Flux	$F(3.22, 61.11)^* = 3.13$	.029
		SubBand4Flux	$F(5, 95) = 2.47$	.038
		SubBand6Flux	$F(5, 95) = 2.80$	.021
		SubBand7Flux	$F(5, 95) = 6.57$	.0005
		SubBand8Flux	$F(5, 95) = 4.28$	.001
		SubBand9Flux	$F(5, 95) = 3.67$	.004
		SubBand10Flux	$F(5, 95) = 5.35$	.0005
		SpectDeviation	$F(5, 95) = 3.49$	.006

*Greenhouse-Geisser correction for Sphericity

Table 7.3: Feature subsets selected in ANOVA for pieces Élégie, Shost1.

Music excerpt	**# of notes**	**Features**	F **statistics**	**Sig.**
Élégie	6	HighFreqEnergy	$F(5, 25) = 3.73$	.012
		Rolloff85	$F(5, 25) = 3.15$	.0241
		Spread	$F(5, 25) = 2.83$	.037
		Skewness	$F(5, 25) = 3.09$	.026
		Kurtosis	$F(5, 25) = 3.55$	.015
		Roughness	$F(5, 25) = 3.06$	.027
		SpectralFlux	$F(2.08, 10.40)^* = 4.32$	.042
		SubBand1Flux	$F(1.97, 9.84)^* = 31.49$	.0005
		SubBand2Flux	$F(5, 25) = 6.26$	.001
		SubBand3Flux	$F(5, 25) = 6.78$	.0005
		SubBand4Flux	$F(5, 25) = 6.46$	.001
		SubBand7Flux	$F(5, 25) = 6.00$	.001
		SubBand8Flux	$F(2.05, 10.28)^* = 4.43$	.040
		SubBand9Flux	$F(5, 25) = 3.01$	.029
		SubBand10Flux	$F(5, 25) = 2.63$	.048
		SpectVariation	$F(2.08, 10.40)^* = 4.29$	.043
Shost1	11	Centroid	$F(2.17, 21.69)^* = 12.74$	.0005
		HighFreqEnergy	$F(5, 50) = 12.28$	.0005
		Spread	$F(5, 50) = 23.18$	.0005
		Skewness	$F(5, 50) = 17.50$	.0005
		Kurtosis	$F(5, 50) = 20.42$	.0005
		Rolloff95	$F(5, 50) = 18.01$	.0005
		Rolloff85	$F(5, 50) = 12.36$	.0005
		SpectEntropy	$F(5, 50) = 9.47$	.0005
		Flatness	$F(2.33, 23.30)^* = 11.89$	.0005
		Roughness	$F(5, 50) = 7.44$	.0005
		Irregularity	$F(5, 50) = 3.23$	.013
		ZeroCrossings	$F(5, 50) = 7.50$	.0005
		SpectralFlux	$F(5, 50) = 4.82$	.001
		SubBand1Flux	$F(2.39, 23.86)^* = 15.09$	.0005
		SubBand2Flux	$F(1.75, 17.47)^* = 12.99$	.001
		SubBand3Flux	$F(5, 50) = 22.87$	.0005
		SubBand4Flux	$F(5, 50) = 12.88$	.0005
		SubBand5Flux	$F(2.20, 22.01)^* = 3.35$	.05
		SubBand7Flux	$F(5, 50) = 6.70$	.0005
		SubBand8Flux	$F(5, 50) = 12.78$	.0005
		SubBand9Flux	$F(5, 50) = 16.99$	.0005
		SubBand10Flux	$F(5, 50) = 17.07$	.0005
		SpectDeviation	$F(5, 50) = 2.89$	.023
		SpectVariation	$F(2.62, 26.21)^* = 3.99$	.022

*Greenhouse-Geisser correction for Sphericity

method and the initial factor structures (based on eigenvalues > 1 criterion) were adjusted using Varimax orthogonal rotation (with Kaiser normalisation).

Table 7.4: Feature subset selected in ANOVA for Shost4.

Music excerpt	**# of notes**	**Features**	F **statistics**	**Sig.**
Shost4	10	Centroid	$F(5, 45) = 3.33$	.012
		HighFreqEnergy	$F(5, 45) = 3.86$	.005
		Spread	$F(5, 45) = 3.07$	.018
		Rolloff95	$F(5, 45) = 3.98$	.005
		Rolloff85	$F(5, 45) = 7.14$	.0005
		SpectEntropy	$F(5, 45) = 4.25$	.003
		Flatness	$F(5, 45) = 4.42$	.002
		SubBand8Flux	$F(2.36, 21.28)^* = 3.36$	.047

*Greenhouse-Geisser correction for Sphericity

The advantage of the PAF method over traditional principal component analysis (PCA) comes from the fact that during factor extraction the shared variance of a variable is partitioned from its unique variance and error variance to reveal the underlying factor structure, thus only shared variance appears in the solution. Since PCA does not discriminate between shared and unique variance, principal components are calculated using all of the variance of the measured variables, and all of that variance is included in the solution (Costello and Osborne, 2005). For that reason PAF is regarded as a truly exploratory factor analysis technique.

To ensure that the obtained solutions were valid and significant, the correlation and multicollinearity levels were assessed. The strength of intercorrelations between the features was checked by means of Bartlett's test of sphericity. In all cases the test was highly significant ($p < .0005$) supporting the validity of the performed factor analyses. The presence of multicollinearity was tested via the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, the values of which ranged from .69 to .89, exceeding the recommended minimum value of .60 (Kaiser, 1974).

The number of yielded factors and subsets of features loading on the factors varied within the datasets (two-, three- and four-factor solutions were obtained). In two cases the number of retained factors was reduced to improve interpretability of the results: Élégie – from the initial four factors (83.9% of the

Table 7.5: Allemande. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 84.7% of total variance explained ($N = 78$).

	Factors		
	1	**2**	**3**
Variance explained	31.4%	29.4%	23.9%
Flatness	**.967**		
Kurtosis	-.923		
Spread	.897		
Skewness	-.813	-.514	
HighFreqEnergy		**.951**	
Rolloff85		.946	
SubBand7Flux		.787	.465
SubBand1Flux			**.902**
SubBand2Flux			.758
SubBand3Flux			.752
SubBand9Flux		.567	.641

total variance explained) to a three-factor solution (75.5% of the total variance explained) and Shost1 – from the initial four factors (81.5% of the total variance explained) to a three-factor solution (75.3% of the total variance explained). In general, the total variance explained by the obtained solutions was high, ranging from 75.3% to 90.8%.

From the factor analysis results collated in Tables 7.5–7.10 one can observe that two- and three-factor structures suggest the existence of two or three underlying acoustical dimensions which can roughly be described as Brightness (high frequency energy content plus noisiness measures), Spectral Variation or Spectral Flux (variation of the spectrum components over time) and Spectral Shape (parameters of the spectrum distribution). Depending on music excerpt, Spectral Flux may be divided into more detailed subdimensions representing variations over time of the low, medium and high frequency regions.

In Allemande (Table 7.5), the three emerging acoustical dimensions are Spectral Shape, Brightness and Spectral Flux of the low- and high-frequency ranges. Descriptors SubBand7Flux and SubBand9Flux, which capture variations in the 1.6–3.2 kHz and 6.4–12.8 kHz frequency bands, being associated with the high frequency content, also contribute to the Brightness factor.

Table 7.6: Bourrée. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 81.3% of total variance explained ($N = 90$).

	Factors		
	1	**2**	**3**
Variance explained	39.6%	30.3%	11.5%
SubBand9Flux	**.950**		
SubBand10Flux	.909		
SubBand8Flux	.875	.347	
SubBand3Flux	.860	-.321	
SubBand4Flux	.772	-.313	.367
SpectDeviation	.771	-.321	
SubBand1Flux	.765		
SubBand2Flux	.760	-.334	
SubBand7Flux	.759		
SubBand6Flux	.720		.392
SubBand5Flux	.695		.485
Centroid		**.964**	
Rolloff85		.956	
Skewness		-.939	
HighFreqEnergy		.932	
SpectEntropy		.867	
Zerocross	-.301	.709	
SpectVariation		.342	**.858**
SpectralFlux	.380		.796

In Bourrée (Table 7.6), the Spectral Flux dimension has been split into spectral variation across all 10 subbands (Factor 1) and the overall spectral variation (Factor 3). An additionally examined two-factor solution which explained 74.5% of the total variance (6.8% less than the three-factor one) yielded all spectral flux and variation descriptors merged into Factor 1 accounting for 43.4% of the variance explained by the model. In both solutions, Factor 2 is clearly associated with the Brightness dimension having the highest loadings from HighFreqEnergy and Rolloff85 (two-factor structure), and Centroid and Rolloff85 descriptors (three-factor structure).

In Courante (Table 7.7), due to the preceding feature selection step, none of the high frequency content or spectral distribution descriptors (with the exception of Spectral Spread) was subjected to factorisation. As a result, the obtained three factors represent spectral fluctuations in low, high and medium frequency

Table 7.7: Courante. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 76.0% of total variance explained ($N = 120$).

	Factors		
	1	**2**	**3**
Variance explained	31.8%	29.7%	14.5%
SubBand3Flux	**.948**		
SubBand1Flux	.853		
SubBand2Flux	.831		
SpectralFlux	.740		
SubBand4Flux	.687		
SubBand9Flux		**.938**	
SpectDeviation		.794	
Spread		.780	
SubBand8Flux		.772	.556
SubBand10Flux	.579	.727	
SubBand7Flux		.335	**.839**
SubBand6Flux			.729

regions with additional contributions to Factors 1 and 2 from the overall spectral variation parameters such as Spectral Flux and Spectral Deviation.

The initial four-factor model in Élégie (Table 7.8) indicated the existence of four acoustical dimensions, namely Spectral Flux of the medium- and high-frequency ranges, Spectral Flux of the low-frequency region and the overall spectrum, Brightness, and Spectral Shape. Subjected to further dimension reduction, Factor 1 and Factor 3 merged into one dimension characterising fluctuations in the high frequency range (SubBands 7-10) and high frequency content (HighFreqEnergy and Rolloff85) while Factors 2 and 4 remained practically the same (with the exception of SubBand4Flux which now loads higher on Factor 2 and Factor 4 axis being inverted).

Similar to the four-factor Élégie model, a division of the Spectral Flux dimension into the mid plus high frequency regions and low plus the overall frequency regions can be observed in the Shost1 excerpt (Table 7.9). These subdimensions are represented by Factors 2 and 3. On the other hand, Factor 1 is a combination of Brightness and Spectral Shape constructs with the highest loading from the Centroid descriptor.

Table 7.8: Élégie. Factor analysis of audio features across all cellists. Factor loadings for two rotated solutions. 83.9% and 75.5% of total variance explained ($N = 36$).

	Factors				Factors		
	1	**2**	**3**	**4**	**1**	**2**	**3**
Variance explained	23.7%	23.7%	19.0%	17.5%	31.7%	26.2%	17.5%
SubBand9Flux	**.907**		.304		.840		
SubBand10Flux	.900				.675	.466	
SubBand8Flux	.737		.527		.901		
SubBand7Flux	.672		.627		**.918**		
SubBand4Flux	.657	.647			.370	.770	
SubBand2Flux		**.812**				.824	
SpectVariation		.811				.732	
SpectralFlux		.793			.306	.746	
SubBand3Flux		.783		-.315		**.841**	.329
SubBand1Flux		.604				.634	
HighFreqEnergy			**.894**		.649		
Rolloff85			.808	.375	.748		-.428
Spread				**.906**			**-.876**
Kurtosis	-.308		-.353	-.881	-.485		.864

The simplest two-factor structure was obtained for the Shost4 excerpt (Table 7.10) and, what is more interesting, only seven acoustic features accounted for 90.8% of the variance in the data, i.e. timbral variation between the players. The two factors are associated with Brightness and Spectral Shape.

Compared to the two- or three-dimensional structures characterising distinct acoustical spaces of the six music excerpts, the three-factor solution based on the whole dataset, i.e. across all music fragments, looks quite similar (to improve factorability, the Irregularity descriptor was excluded from this run of factor analysis). As can be seen from Table 7.11, the Spectral Shape dimension is no longer present, being largely merged with Brightness (Factor 1). The second dimension represents spectral fluctuations across all 10 subbands while the overall spectral variation is captured by Factor 3. Following further dimension reduction, the emerged acoustical structure became simpler. The obtained factors Brightness and SubBand 1-10 Flux were sufficient to explain 74.8% of the variance in the data (see Table 7.11).

Table 7.9: Shost1. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 75.3% of total variance explained ($N = 66$).

	Factors		
	1	**2**	**3**
Variance explained	34.8%	23.8%	16.7%
Centroid	**.967**		
Skewness	-.922	-.354	
Rolloff95	.901	.348	
Kurtosis	-.896	-.378	
SpectEntropy	.813		.432
Flatness	.796		
HighFreqEnergy	.783		
Spread	.739	.375	
ZeroCrossings	.726		
SubBand9Flux	.433	**.880**	
SubBand10Flux		.861	
SubBand8Flux	.458	.841	
SubBand5Flux		.840	
SubBand7Flux	.360	.703	
SpectDeviation		.689	
SubBand4Flux		.612	.571
SpectralFlux			**.818**
SubBand3Flux			.781
SpectVariation	.316		.768
SubBand2Flux			.728
SubBand1Flux			.586

Table 7.10: Shost4. Factor analysis of audio features across all cellists. Factor loadings for the rotated solution. 90.8% of total variance explained ($N = 60$).

	Factors	
	1	**2**
Variance explained	65.2%	25.6%
HighFreqEnergy	**.954**	
Rolloff85	.943	
SpectEntropy	.932	
Centroid	.913	.350
SubBand8Flux	.779	
Spread		**.992**
Flatness	.662	.709

Table 7.11: The six music excerpts combined. Factor analysis of audio features across all cellists. Factor loadings for two rotated solutions. 81.2% and 74.8% of total variance explained ($N = 450$).

	Factors			Factors	
	1	**2**	**3**	**1**	**2**
Variance explained	39.0%	27.1%	15.1%	40.1%	34.7%
Centroid	**.989**			**.976**	
Rolloff85	.963			.952	
Rolloff95	.956			.971	
Skewness	-.952			-.974	
HighFreqEnergy	.942			.919	
Kurtosis	-.904			-.937	
Flatness	.855			.876	
ZeroCrossings	.840			.788	
SpectEntropy	.818		.403	.752	
Spread	.714		-.303	.758	
SubBand10Flux		**.908**			.909
SubBand9Flux	.464	.846		.579	.719
SubBand5Flux		.803	.408		**.911**
SubBand6Flux		.765	.335		.828
SubBand8Flux	.551	.762		.651	.647
SubBand4Flux		.760	.344		.841
SpectDeviation		.755		.385	.573
SubBand7Flux	.330	.710		.404	.717
SubBand3Flux	-.426	.628	.455	-.377	.805
SubBand1Flux	-.376	.586	.473	-.336	.774
SubBand2Flux	-.434	.555	.517	-.405	.775
SpectVariation			**.826**		.664
Roughness		.332	.794		.683

7.3.2 Acoustical mapping of the players

To first illustrate how the cellists are positioned within the acoustical dimensions, the averages of the factor scores obtained from the factor analyses described in the previous section were calculated per player across notes in each music excerpt dataset. The same averaging procedure was then repeated for the acoustic features most correlated with the factors. The resulting mean factor scores and the feature means were visualised using scatter plots (Figures 7.1-7.6) in order to investigate dis/similarities between the players in terms of acoustical characteristics. The respective correlation values between the mean factor scores and between the feature means are presented in Tables 7.12-7.17. The

Table 7.12: Allemande. Correlations between mean factor scores and between means of the highest loading features ($N = 6$).

Factors	Spectral Shape	Brightness	**Features**	Flatness	HighFreq Energy
Brightness	-.70		HighFreqEnergy	-.48	
SubBand 1-3 Flux	.54	-.78*	SubBand1Flux	.06	-.67

$^{*}p < .05$

factors are named according to the interpretations discussed in Section 7.3.1.

A strong negative and significant correlation between SubBand 1-3 Flux and Brightness observed for the Allemande (Table 7.12) confirms the fact that brighter tone samples have on average less fluctuation in the low frequency range (compare positioning of Cellist 2 and Cellists 3 and 4 in Figure 7.1e). Interestingly, similar results were obtained by Alluri and Toiviainen (2010) in their study on polyphonic timbres where stimuli perceived as *bright* tended to have less fluctuation in the lower frequency regions (SubBand1) and more fluctuation in the higher frequency regions (SubBand6 and SubBand7).

This tendency is less vivid (according to the lower and not significant correlation value) when the players were mapped into the space of the respective acoustic features, i.e. SubBand1Flux and HighFreqEnergy (Figure 7.1f). Cellist 2 still has the brightest tone with the lowest variation level in the lower frequency region while for example the tone of Cellist 1 being also one of the brightest fluctuates in this frequency range as much as the tone of Cellist 4.

The factors Brightness and Spectral Shape also exhibit a relatively high although not significant negative correlation. Cellists 3 and 4, possessing the least bright tones, have at the same time the highest Spectral Shape values (Figure 7.1a). The relationship between the two factors might be more easily interpreted in terms of acoustic features. The Flatness descriptor, which loaded the most on Spectral Shape, is used to discriminate between noisy and tonal signals, having values close to 1 for flat spectra (white noise). Looking at Figure 7.1b, one can see that Cellist 3 has the least bright tone and has a very similar Flatness value to Cellist 2 who sounds the brightest. This may suggest

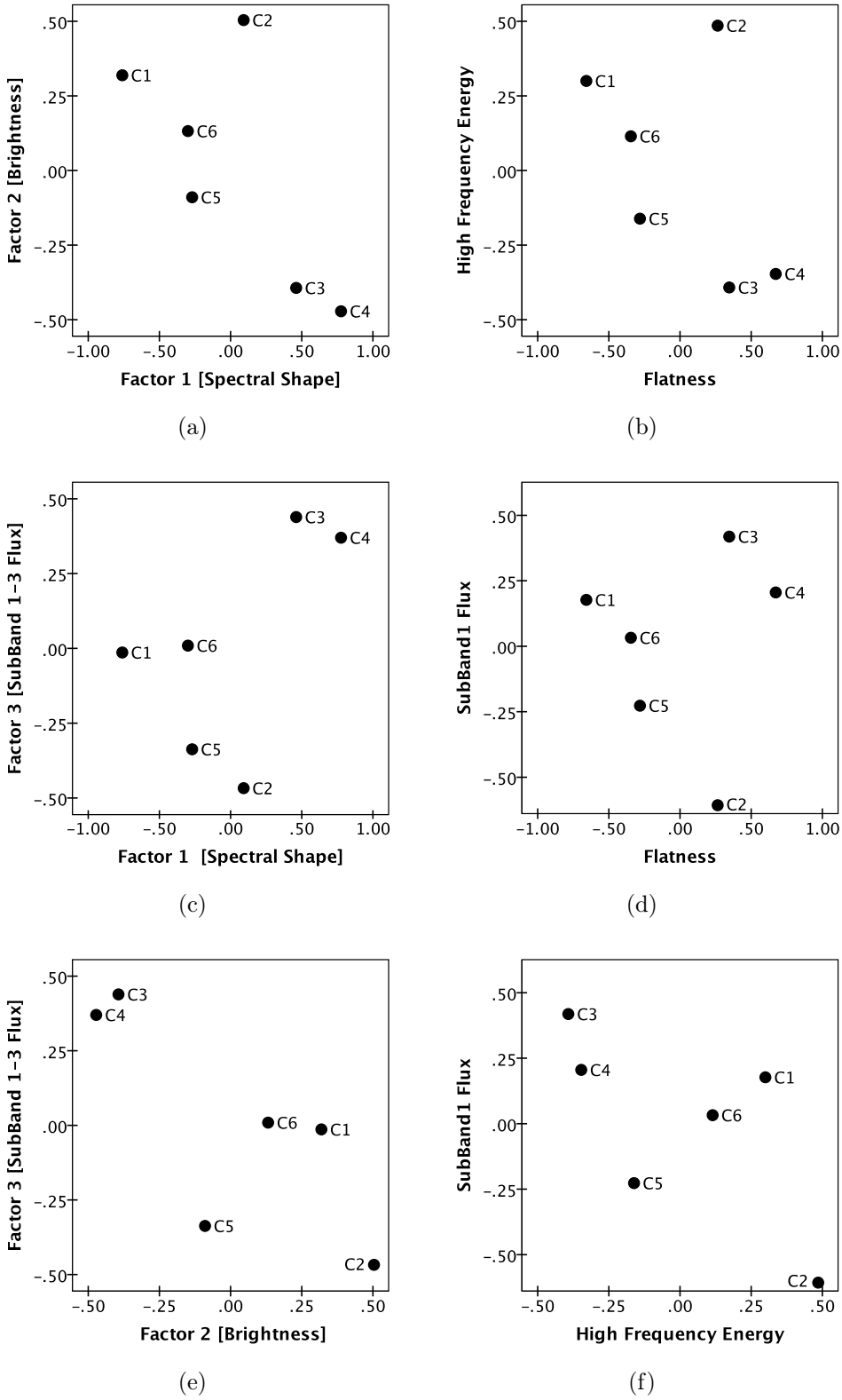


Figure 7.1: Allemande. Scatter plots of mean factor scores (left) and mean acoustic features (right).

Table 7.13: Bourrée. Correlations between mean factor scores and between means of the highest loading features ($N = 6$).

Factors	SubBand 1-10 Flux	Brightness	**Features**	SubBand9 Flux	Centroid
Brightness	.22		Centroid	.17	
TotalSpectralFlux	.93**	.38	SpectVariation	.87*	.60

*$p < .05$, **$p < .005$

that while the two players have distinct spectral envelope slopes the level of noise components in their tones is comparable.

Since SubBand 1-10 Flux and overall Spectral Flux capture similar timbral characteristics they are also positively and significantly correlated as one can see from Table 7.13 and Figure 7.2c, displaying the acoustical mapping of the cellists for the Bourrée. A similar trend can be observed between respective acoustic features, i.e. SubBand9Flux and Spectral Variation (Figure 7.2d), except for Cellist 2 whose timbre has the least fluctuation in this particular frequency subband but much more variation across the entire spectrum. The case of Cellist 2 is even more interesting taking into account that his tone is also the brightest one (Figure 7.2b) and one would expect to see more fluctuations in the higher frequency band as is the case for Cellists 6 and 1.

In regard to the Courante, the only strong and significant correlation (Table 7.14) was found between SubBand3Flux and SubBand7Flux features (Figure 7.3d) and to a lesser extent (but not significant) between respective acoustic factors such as SubBand 1-4 Flux and SubBand 6-7 Flux (Figure 7.3c). This tendency might seem interesting considering the frequency ranges between which

Table 7.14: Courante. Correlations between mean factor scores and between means of the highest loading features ($N = 6$).

Factors	SubBand 1-4 Flux	SubBand 8-10 Flux	**Features**	SubBand3 Flux	SubBand9 Flux
SubBand 8-10 Flux	.19		SubBand9Flux	.31	
SubBand 6-7 Flux	.61	-.51	SubBand7Flux	.78*	.04

*$p < .05$

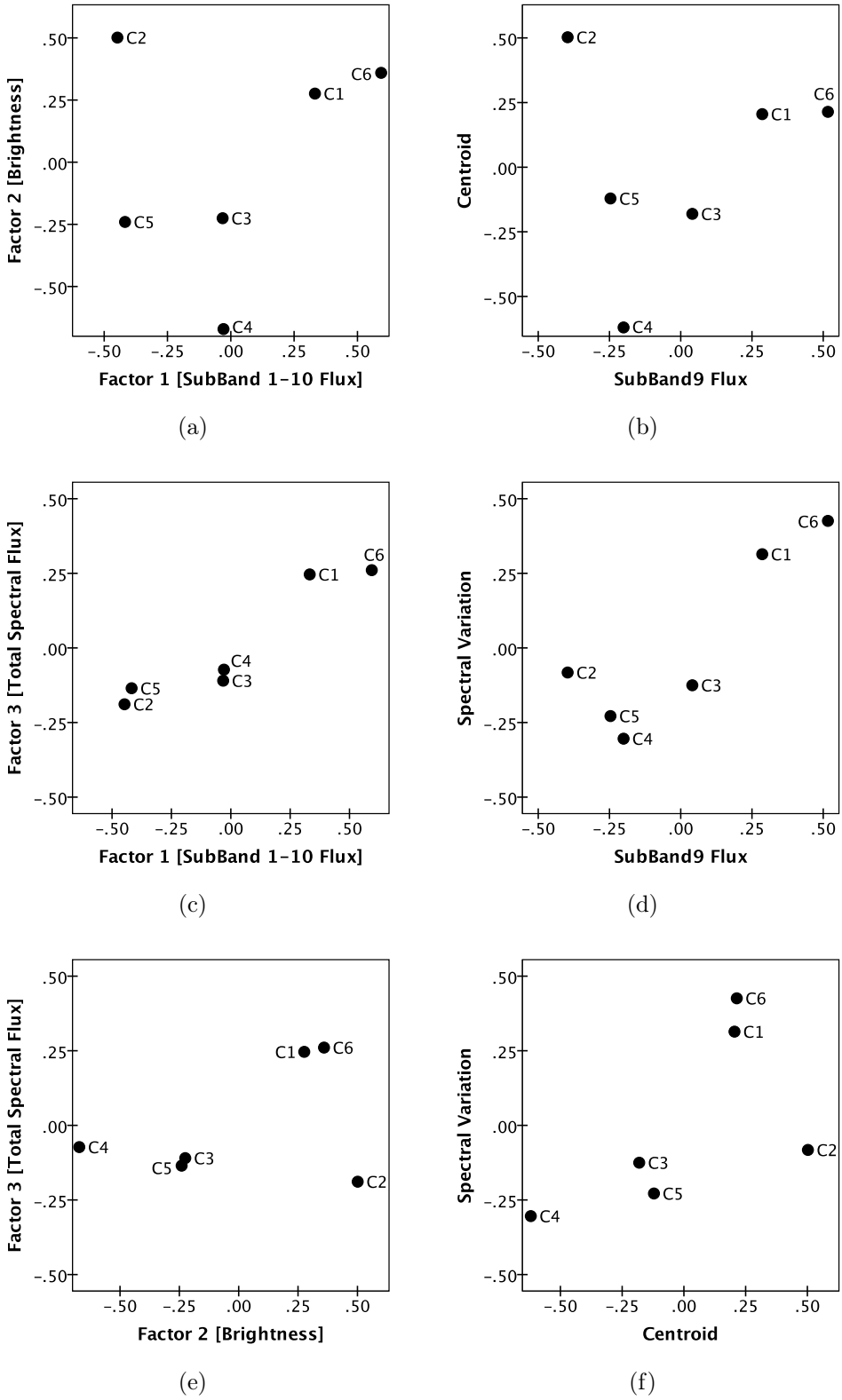


Figure 7.2: Bourrée. Scatter plots of mean factor scores (left) and mean acoustic features (right).

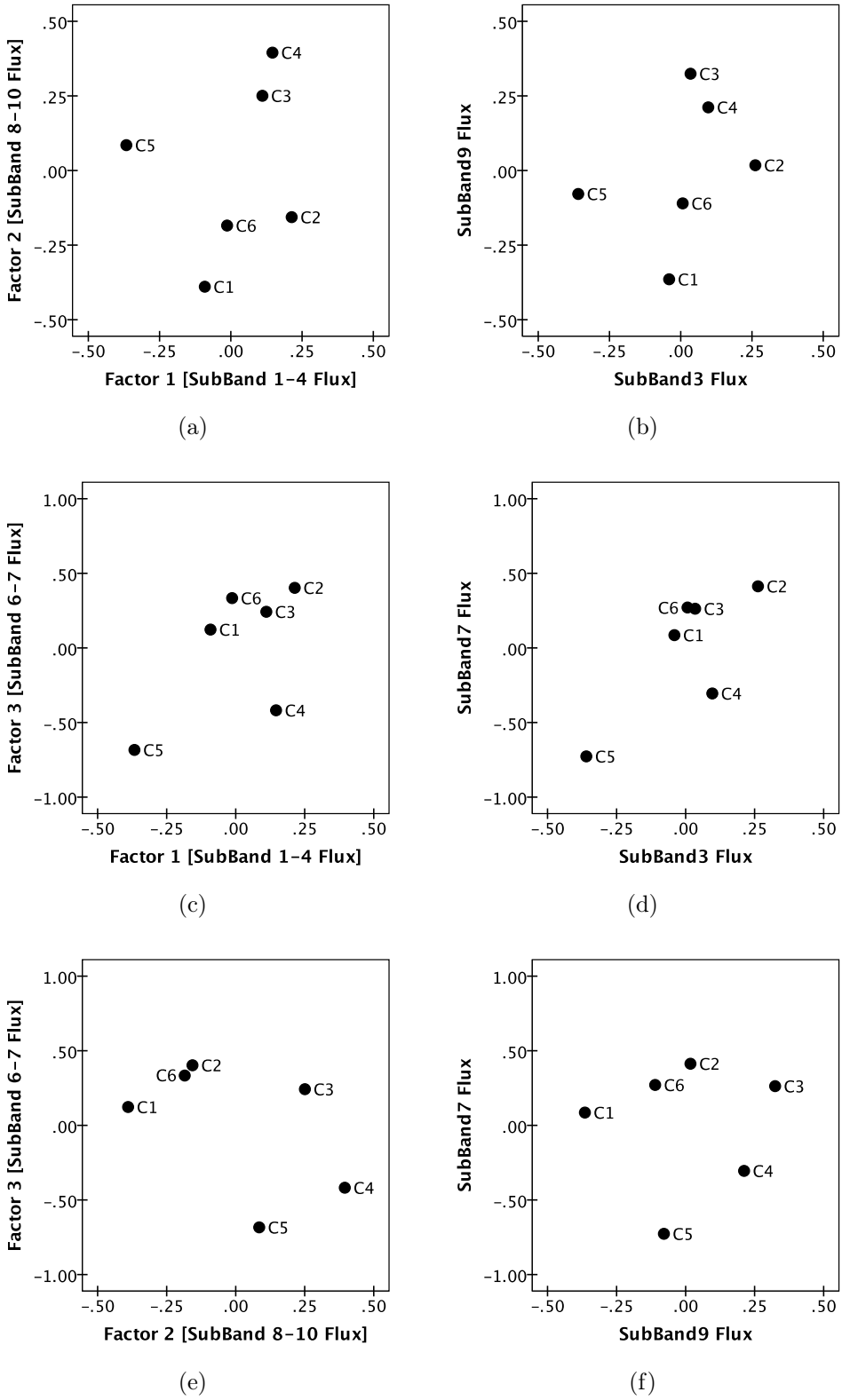


Figure 7.3: Courante. Scatter plots of mean factor scores (left) and mean acoustic features (right).

Table 7.15: Élégie. Correlations between mean factor scores and between means of the highest loading features ($N = 6$).

Factors	SubBand 7-10 Flux	SubBand 1-4 Flux	**Features**	SubBand7 Flux	SubBand3 Flux
SubBand 1-4 Flux	.08		SubBand3Flux	.73*	
Spectral Shape	.31	-.25	Spread	-.03	.12

*$p < .05$

the amount of variation was positively correlated. Tones with higher fluctuation in the 100–200 Hz frequency band also showed higher fluctuation in the range of 1.6–3.2 kHz. Interestingly, Cellist 4's tone deviates from this tendency, despite having at the same time relatively high fluctuation in the 6.4–12.8 kHz frequency band, i.e. SubBand9Flux (see Figure 7.3b). The opposite side of the trend is occupied by Cellist 5, whose tone samples seem quite distinct from the rest of the players.

Similarly to the Courante, a positive and significant correlation between SubBand3Flux and SubBand7Flux features was observed in Élégie (Figure 7.4b, Table 7.15). This time it is Cellist 6 who occupies the positive extremity of the trend having the highest fluctuation in both frequency regions in opposition to Cellist 5 who is again located on the other side of the trend. Moreover, Cellist 5 has also the lowest Spread value compared to other cellists (Figures 7.4d–7.4f). At the same time Cellist 6's tone, having the strongest fluctuations in SubBand3Flux or in the lower frequency subbands (SubBand 1-4 Flux), also has relatively low Spread or relatively high Spectral Shape values (Figures 7.4e–7.4f).

In regard to the Shost1 excerpt (Figure 7.5b) there is a positive (although not significant, see Table 7.16) relationship between Centroid and SubBand9Flux descriptors. It implies that brighter sounds also have more fluctuation in the higher frequency region, i.e. 6.4–12.8 kHz. The exception here is Cellist 2 who has the brightest tone but only moderate fluctuation in this subband. Similarly, Cellist 1 deviates from the positive trend between SubBand9 Flux and the overall Spectral Flux (Figure 7.5f). Compared to other players his tone fluctuates

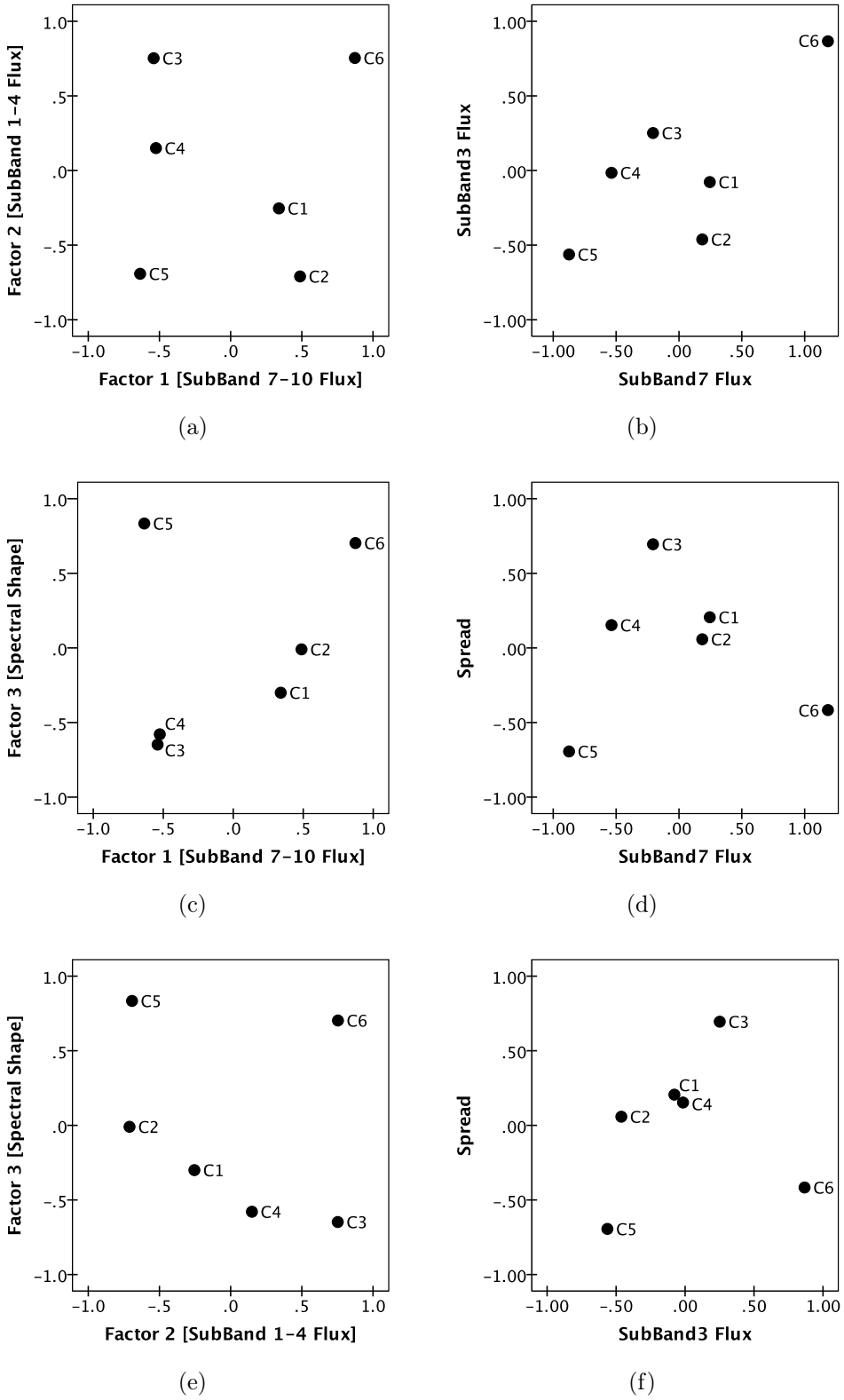


Figure 7.4: Élégie. Scatter plots of mean factor scores (left) and mean acoustic features (right).

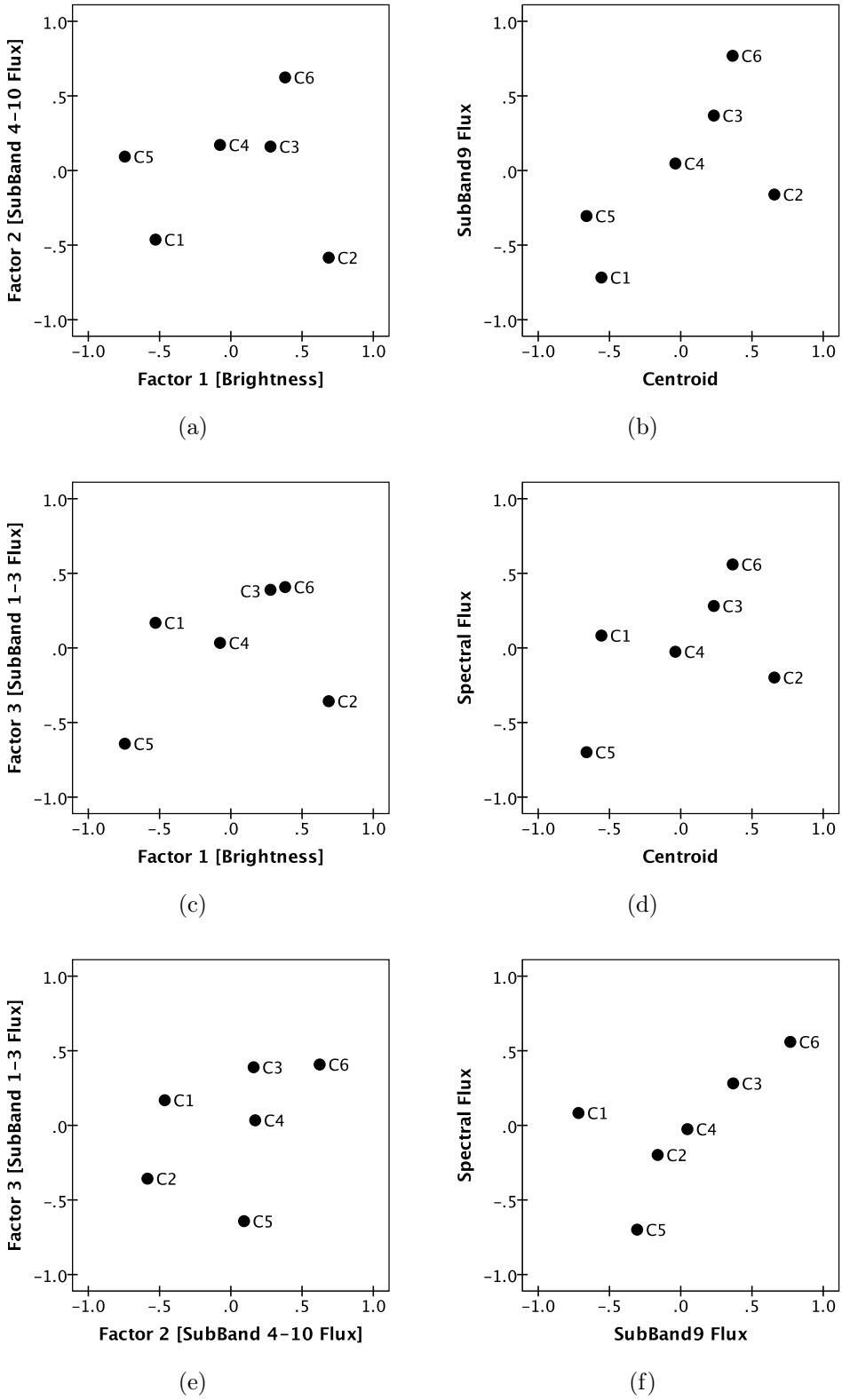


Figure 7.5: Shost1. Scatter plots of mean factor scores (left) and mean acoustic features (right).

Table 7.16: Shost1. Correlations between mean factor scores and between means of the highest loading features ($N = 6$).

Factors	Brightness	SubBand 4-10 Flux	**Features**	Centroid	SubBand9 Flux
SubBand 4-10 Flux	.03		SubBand9Flux	.63	
SubBand 1-3 Flux	.35	.42	Spectral Flux	.49	.64

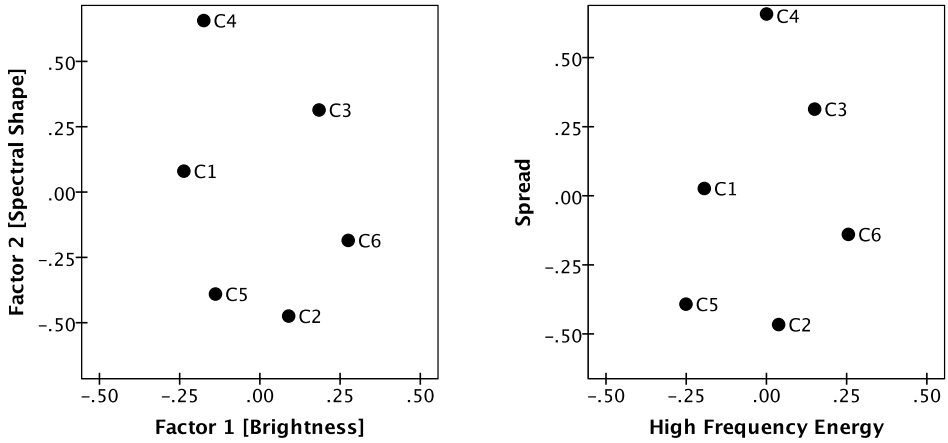
moderately across all frequencies, having at the same time the least fluctuation in the region of 6.4–12.8 kHz. The two tendencies observed between the acoustic features are not present between the respective factors but one interesting feature emerges from the examination of all six acoustical spaces (Figures 7.5a–7.5f). It can be seen that the position of Cellist 4 in all cases is mostly central and therefore may represent the average tone characteristics.

No trends between acoustic factors or respective features were observed in the Shost4 excerpt (see Table 7.17). From Figure 7.6 one can notice that the positioning of the cellists in both spaces is quite similar. The players' tones are well separated from each other although there is higher variation between the players in terms of Spectral Spread than for the HighFreqEnergy descriptor.

Figure 7.7 illustrates the positioning of the cellists in the factor and highest loading feature spaces obtained from factorising the entire dataset, i.e. the six music excerpts combined together. According to Factor 2 (SubBand 1-10 Flux) and the respective SubBand10Flux descriptor the players' tones clustered into two groups based on the amount of fluctuation across various frequency subbands (or in the 12.8–22.05 kHz frequency range in particular). The tones could be then characterised by either high or low level of the subband flux with no mid level represented. For example, Cellists 2 and 6 having the brightest

Table 7.17: Shost4. Correlations between mean factor scores and between means of the highest loading features ($N = 6$).

Factors	Brightness	**Features**	HighFreqEnergy
Spectral Shape	-.25	Spread	.21



(a) (b)

Figure 7.6: Shost4. Scatter plots of mean factor scores (left) and mean acoustic features (right).

timbres (Figure 7.7a) had at the same time the lowest and the highest spectral variation over time respectively. Interestingly, spectral fluctuations in the subbands did not correlate with the total spectral flux (Factor 3) as it was the case for Bourrée (compare Figures 7.7e and 7.2c and respective Tables 7.18 and 7.13). The only pronounced relationship between two factors (r = .70, n.s.) was found for Brightness and TotalSpectralFlux (Figure 7.7c). On average, the brighter the tone, the more fluctuating was its overall spectrum, with exception of Cellist 1 whose relatively less bright timbre was most varying over time.

The resulting acoustical mapping of the cellists based on the two-factor solution looks somewhat similar to the map obtained for Factors 1 and 3 in the three-dimensional space. As can be seen from Figure 7.8, Brightness and

Table 7.18: All music styles combined. Three-factor solution. Correlations between mean factor scores and between means of the highest loading features ($N = 6$).

Factors	Brightness	Total SpectralFlux	**Features**	Centroid	Spect Variation
TotalSpectralFlux	.70		SpectVariation	.58	
SubBand 1-10 Flux	.05	-.46	SubBand10Flux	.16	.13

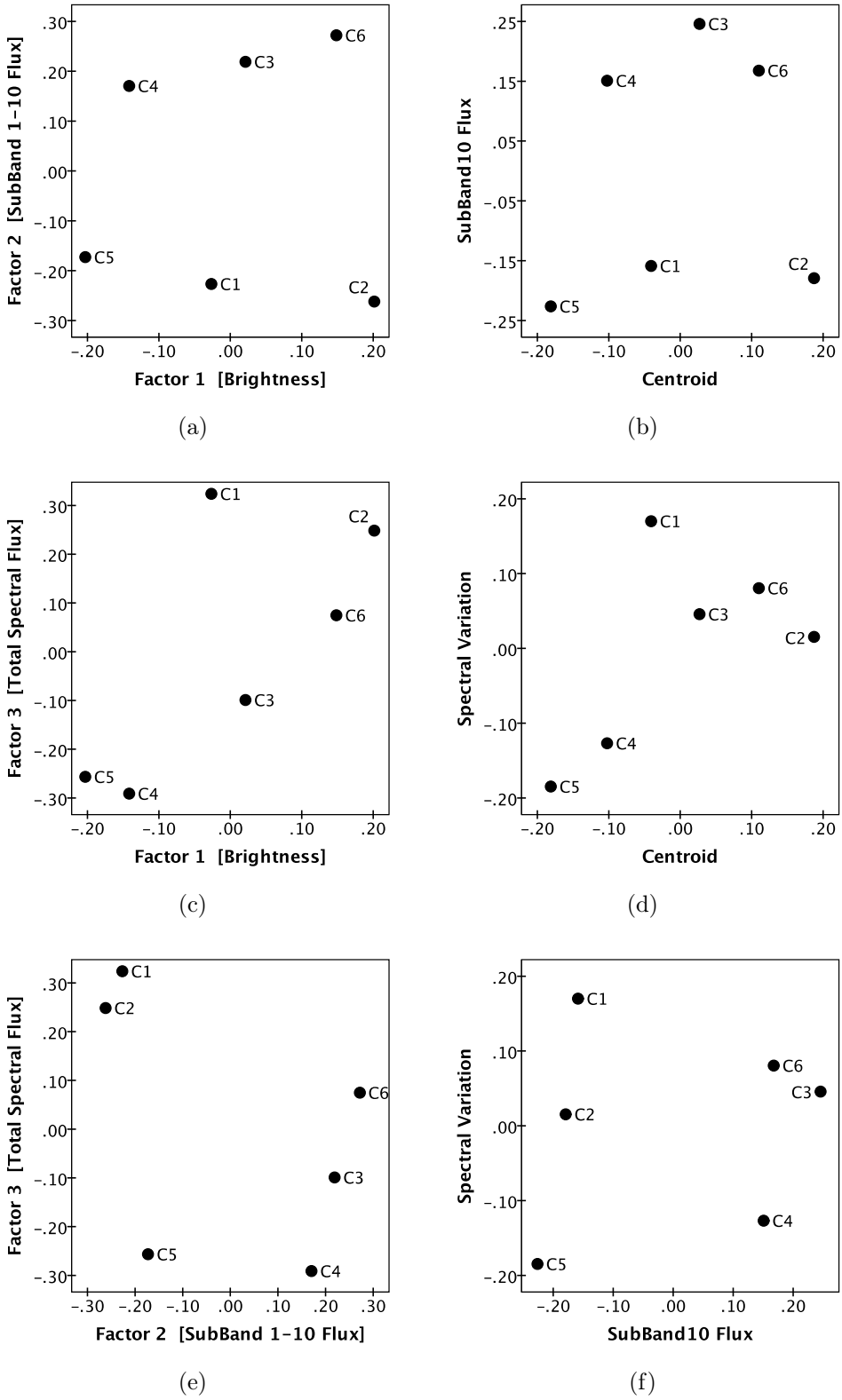
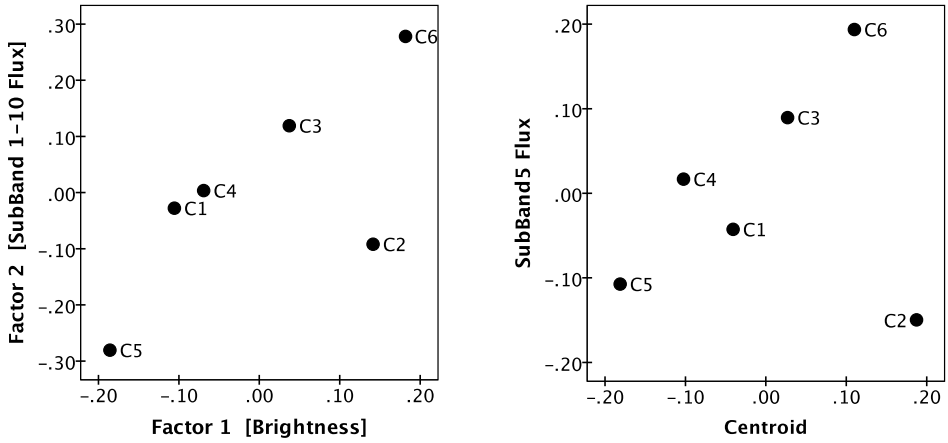


Figure 7.7: All music styles combined. Three-factor solution. Scatter plots of mean factor scores (left) and mean acoustic features (right).



(a) (b)

Figure 7.8: All music styles combined. Two-factor solution. Scatter plots of mean factor scores (left) and mean acoustic features (right).

SubBand 1-10 Flux dimensions are positively correlated (r = .70, n.s.) with some shifts in coordinates of the players for the acoustic feature mapping. The observed tendency suggest that brighter tones had also more strongly fluctuating spectra, except for Cellist 2 whose timbral characteristics notably deviated from this trend.

7.3.3 Discriminating performers based on factor scores and acoustic features

In Section 7.3.1 a series of factor analyses conducted on acoustic feature subsets across and over six music excerpts identified two- or three-dimensional acoustical spaces in which varying timbral characteristics of the cellists could be effectively described (as shown in Section 7.3.2). These spacial characterisations, however, did not yet provide an answer to the question of whether and to what extent the players' tones can be acoustically discriminated regardless of the music performed.

It was already demonstrated that each of the 24 spectro-temporal descriptors exhibited significant variations across the players when tested on the whole

Table 7.19: Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for factor scores in the three-factor solution ($N = 450$).

Factor	F **statistics**	**Significance**	**Effect size****
1 [Brightness]	$F(4.46, 329.74)^* = 8.02$	$p < .0005$	$\eta^2 = .10$
2 [SubBand 1-10 Flux]	$F(4.30, 318.32)^* = 15.47$	$p < .0005$	$\eta^2 = .17$
3 [Total Spectral Flux]	$F(5, 370) = 18.14$	$p < .0005$	$\eta^2 = .20$

*Huynh-Feldt correction for Sphericity, **partial η^2 reported

dataset, i.e. including all notes from the six excerpts. Consequently, it was projected that the two- or three-factor structures, obtained from 24-feature spaces, should also retain those variations at a significant level. Therefore, to examine whether factor scores and respective highest loading descriptors are sufficient to differentiate between the cellists' tones, four one-way repeated measures MANOVAs were carried out for the two- and tree-factor solution scores and their correlated features. In these designs the factors or descriptors served as the dependent variables, and the players comprised the six-level within-subjects variable. The data was screened for the assumptions of multivariate and univariate normality and outliers but no violations were detected. A visualisation of the factor scores and descriptors across the cellists in the three-factor solution is presented in Figure 7.9.

According to the first two MANOVA results, three-dimensional timbral characteristics of the players differed significantly at $p < .0005$ (Wilks' $\Lambda = .31$, $F(15, 60) = 9.05$, effect size *partial* $\eta^2 = .69$ for factor scores and Wilks' $\Lambda = .33$, $F(15, 60) = 8.08$, effect size *partial* $\eta^2 = .67$ for the most correlated features). In both cases, the η^2 measure represents the variance accounted for by the best linear combination of dependent variables (Tabachnick and Fidell, 2007), i.e. factor scores or features, indicating 69% and 67% of variance explained respectively. Univariate ANOVA tests for each factor and feature, with a Bonferroni adjustment of alpha levels for multiple tests ($p < .05/3 \approx .017$), also proved significant variations between the players (Tables 7.19 and 7.20), with Factor 3 showing the strongest variations followed by Factors 2 and 1

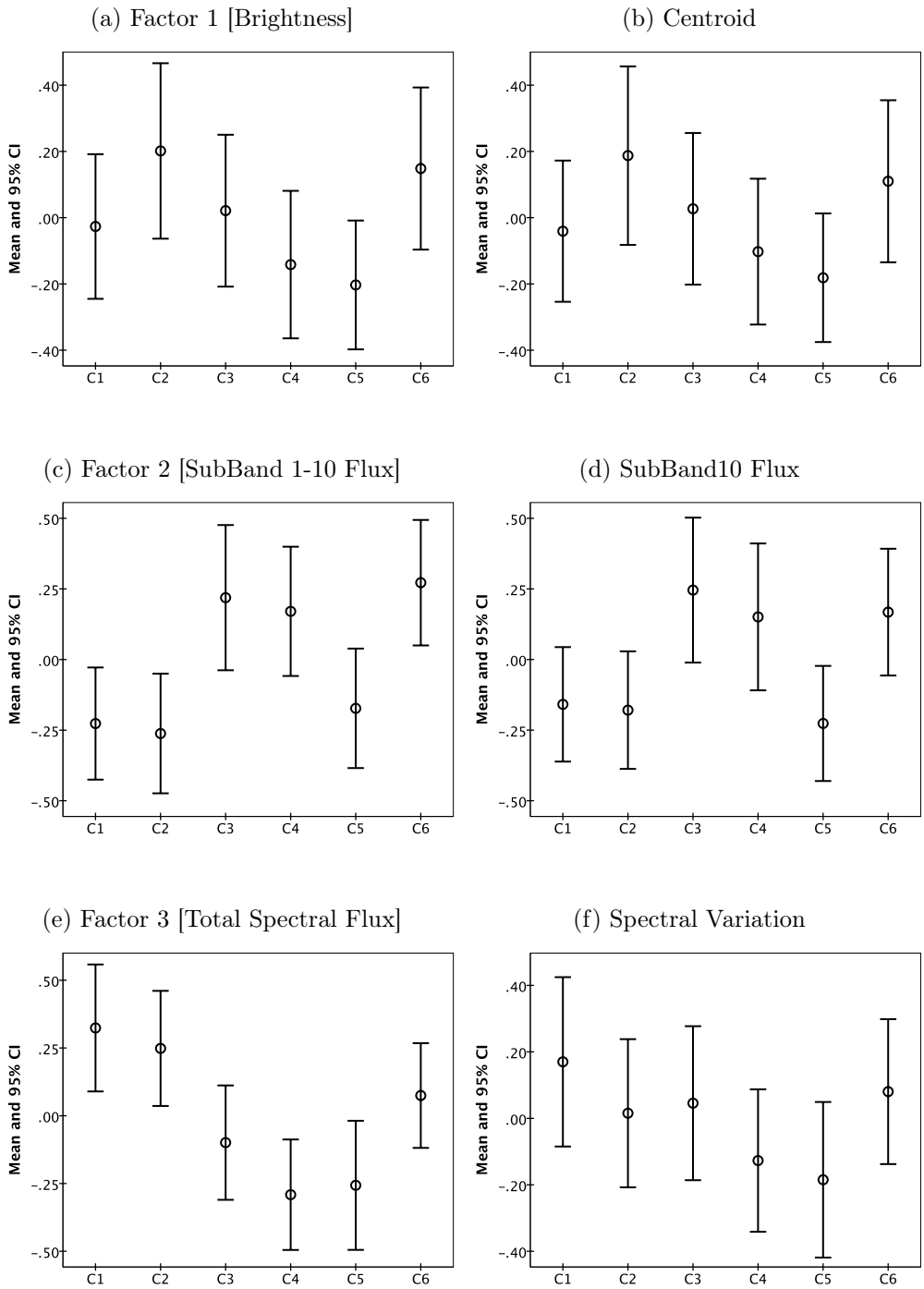


Figure 7.9: The three-factor solution. Comparison of mean factor scores (a)–(c)–(e) and means of the highest loading acoustic features (b)–(d)–(f) across the cellists, ($N = 450$).

Table 7.20: Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for features in the three-factor solution ($N = 450$).

Feature	F **statistics**	**Significance**	**Effect size****
Centroid	$F(4.09, 302.64)^* = 7.02$	$p < .0005$	$\eta^2 = .09$
SubBand10Flux	$F(4.26, 315.28)^* = 12.50$	$p < .0005$	$\eta^2 = .14$
SpectVariation	$F(4.49, 332.54)^* = 5.92$	$p < .0005$	$\eta^2 = .07$

*Huynh-Feldt correction for Sphericity, **partial η^2 reported

(note the respective F and *partial* η^2 values). As for the acoustic features, the strongest Cellist effect was indicated for the SubBand10Flux descriptor followed by Centroid and SpectVariation. For both factors and features, the observed Cellist effect was at least of a medium size, i.e. $\eta^2 \geq .06$ according to Cohen (1988)'s guidelines[1].

A Bonferroni *post-hoc* comparison showed (Figure 7.9e) that in terms of Factor 3 (TotalSpectralFlux) the difference between Cellist 1 and Cellists 3, 4 and 5 was significant at $p < .0005$. Cellist 2 differed significantly from Cellist 3 at $p < .004$ and from Cellists 4 and 5 at $p < .0005$. Cellist 4 and 5's tones, with the least fluctuating overall spectra, also differed significantly from Cellist 6 at $p < .001$ and $p < .002$ respectively. For Factor 2 (SubBand 1-10 Flux), it was already demonstrated in Section 7.3.2 that Cellists 1, 2 and 5 had significantly less spectral fluctuations across various frequency subbands in comparison with Cellists 3, 4 and 6 at $p < .0005$ (see Figure 7.9c). In terms of Factor 1 (Brightness) Cellist 2 had significantly brighter tone than Cellists 4 and 5 ($p < .016$ and $p < .0005$), and so had Cellist 6 ($p < .009$ and $p < .0005$). There was also a significant difference ($p < .043$) between Cellists 5 and 3.

In regard to the acoustic features, the differences between the players although significant were less pronounced. Similarly to Factor 2, Cellists 1, 2 and 5 had significantly less spectral fluctuations in SubBand10Flux (Figure 7.9d) compared to Cellists 3, 4 and 6 (p levels ranged from .0005 to .023). The least

[1]The η^2 values between .01 and .06 suggest a small effect while $\eta^2 \geq .14$ indicates a large size effect

Table 7.21: Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for factor scores in the two-factor solution ($N = 450$).

Factor	F **statistics**	**Significance**	**Effect size****
1 [Brightness]	$F(4.62, 341.55)^* = 6.83$	$p < .0005$	$\eta^2 = .08$
2 [SubBand 1-10 Flux]	$F(4.53, 335.38)^* = 17.04$	$p < .0005$	$\eta^2 = .19$

*Huynh-Feldt correction for Sphericity, **partial η^2 reported

bright tone of Cellist 5 (as indicated by the Centroid descriptor, see Figure 7.9b) differed significantly from tones of Cellists 2, 3 and 6 at $p < .001$, $p < .038$ and $p < .0005$ respectively, while the difference between the brightest tone of Cellist 2 and second most "dark" tone of Cellist 4 was significant at $p < .049$. As for SpectVariation, the only significant differences found were between Cellist 1 and Cellists 4 and 5 ($p < .0005$ and $p < .005$), and between Cellist 5 and Cellist 6 at $p < .003$.

Note. For a number of significant results reported above, confidence intervals of the respective means shown in Figure 7.9 overlap which might suggest something contradictory. In fact, it is the 95% confidence interval for the difference between two group means, not containing zero, which indicates the significant difference. The overlap of confidence intervals between two significantly different means x_1 and x_2 occurs when their difference $x_1 - x_2$ is:

$$1.96(SE_1 + SE_2) > x_1 - x_2 > 1.96\sqrt{SE_1^2 + SE_2^2} \tag{7.1}$$

where SE_1 and SE_2 are respective standard errors.

Table 7.22: Investigating acoustical differences between the cellists across six musical contexts. Results of univariate ANOVAs for features in the two-factor solution ($N = 450$).

Feature	F **statistics**	**Significance**	**Effect size****
Centroid	$F(4.09, 302.64)^* = 7.02$	$p < .0005$	$\eta^2 = .09$
SubBand5Flux	$F(4.31, 318.75)^* = 5.64$	$p < .0005$	$\eta^2 = .07$

*Huynh-Feldt correction for Sphericity, **partial η^2 reported

The MANOVAs conducted on the two-factor solution confirmed that two-dimensional timbral characteristics were also sufficient to discriminate between the players, Wilks' $\Lambda = .41$, $F(15, 60) = 9.28$, $p < .0005$, effect size *partial* $\eta^2 = .59$ for factor scores and Wilks' $\Lambda = .60$, $F(15, 60) = 4.33$, $p < .0005$, effect size *partial* $\eta^2 = .40$ for the most correlated features (59% and 40% of variance explained respectively). Follow up univariate ANOVAs, with a Bonferroni adjustment of alpha levels for multiple tests ($p < .05/2 = .025$), indicated that each of the two factors and descriptors varied significantly between the cellists, with medium to large effect sizes (Tables 7.21 and 7.22). Significant main effects of Cellist are illustrated in Figure 7.10 (refer to *Note* on the previous page when comparing results).

Post-hoc Bonferroni adjusted comparisons of Factor 2 (SubBand 1-10 Flux) scores showed that both the most and the least spectrally fluctuating tones of Cellists 6 and 5 were significantly different from the others (except for Cellists 3 and 2) at p levels ranging from .0005 to .002 (see Figure 7.10c). Another significant difference ($p < .041$) was found between Cellists 2 and 3. Evidently less pronounced variations were observed for the SubBand5Flux descriptor (Figure 7.10d). The only significant differences were found between Cellist 6 and Cellists 1, 2 and 5 ($p < .009$, $p < .006$ and $p < .001$). In terms of Factor 1 (Brightness), the brightest tone of Cellist 6 differed significantly from tones of Cellists 1, 4, and 5 at $p < .001$, $p < .03$ and $p < .0005$ respectively, while the differences between the least bright tone of Cellist 5 and tones of Cellists 2 and 3 were significant at $p < .003$ and $p < .045$ (Figure 7.10a). Interestingly, according to the Centroid results (Figure 7.10b), it was Cellist 2 who possessed the brightest tone, significantly different from those of Cellists 4 and 5 ($p < .049$ and $p < .001$) which were the least bright in comparison. Cellist 5's tone, on the other hand, differed significantly from brighter tones of Cellists 3 and 6 at $p < .038$ and $p < .0005$ respectively.

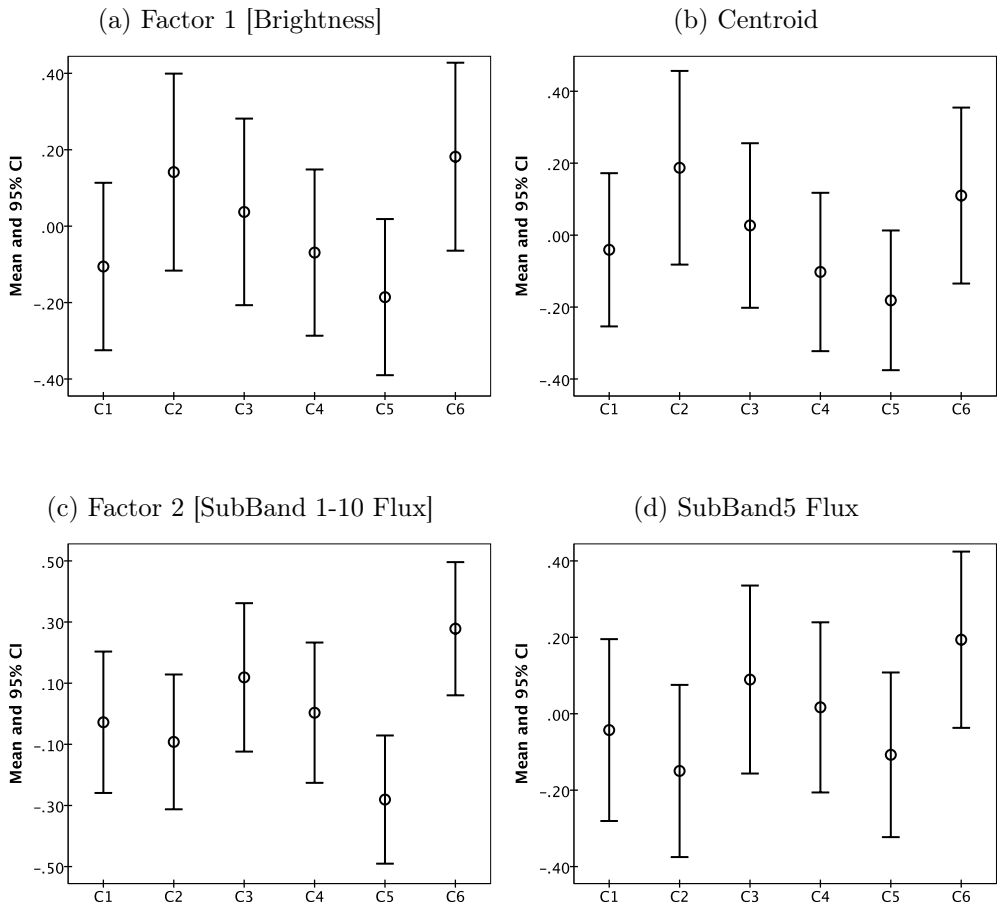


Figure 7.10: The two-factor solution. Comparison of mean factor scores (a)–(c) and means of the highest loading acoustic features (b)–(d) across the cellists, ($N = 450$).

7.3.4 Correlation between acoustical and perceptual dimensions

To finally interpret the perceptual dimensions, the axis coordinates of the cellists obtained from the MDS models from Chapter 6 were correlated with the mean factor scores and then with the mean acoustic features. Tables 7.23-7.28 collate correlation analysis results for each of the six music excerpts.

In Allemande (Table 7.23), the second perceptual dimension could be effectively represented by Brightness or Low Frequency Flux factors following their strong negative correlation with each other (as discussed in Section 7.3.2). The

Table 7.23: Allemande. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
Spectral Shape	.24	.63	Flatness	.25	.48
Brightness	-.16	-.97**	HighFreqEnergy	-.19	-.97**
SubBand 1-3 Flux	-.46	.87*	SubBand1Flux	-.57	.80*

$^{*}p < .05$, $^{**}p < 0.001$

Table 7.24: Bourrée. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
SubBand 1-10 Flux	-.24	.30	SubBand9Flux	-.14	.42
Brightness	-.71	.63	Centroid	-.66	.59
Total Spectral Flux	-.39	.19	SpectVariation	-.53	.51

two factors can be replaced by the respective highest loading acoustic features such as HighFreqEnergy and SubBand1Flux with a minimal loss of information. As the first perceptual dimension was found to not correlate significantly with any of the acoustical dimensions or features it could be tentatively related to SubBand1Flux due to its strongest yet non-significant correlation coefficient.

In Bourrée, neither of the two perceptual dimensions correlated significantly with any of the factors or acoustic features. As one can see from Table 7.24 they could be best explained by Centroid and Spectral Variation features.

The first perceptual dimension in Courante correlated highly and significantly with both SubBand3 and SubBand7 Flux (see Table 7.25). Following that the two features were also highly and significantly correlated with each other (see Section 7.3.2), either of them could be used to explain variations of timbre characteristics represented by this dimension. On the other hand, the second perceptual dimension could only to some extent be interpreted using SubBand7Flux or the SubBand 6-7 Flux factor.

In Élégie, except for the higher frequency flux factor, none of the remaining factors or acoustic features correlated significantly with the perceptual dimensions (see Table 7.26). While SubBand 7-10 Flux might explain the second

Table 7.25: Courante. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
SubBand 1-4 Flux	-.77*	.17	SubBand3Flux	-.87*	.21
SubBand 8-10 Flux	.37	.22	SubBand9Flux	.02	-.01
SubBand 6-7 Flux	-.83*	-.47	SubBand7Flux	-.86*	-.38

$^*p < .05$

Table 7.26: Élégie. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
SubBand 7-10 Flux	.40	.79*	SubBand7Flux	.48	.55
SubBand 1-4 Flux	.32	-.43	SubBand3Flux	.35	-.12
Spectral Shape	-.55	.48	Spread	.67	-.36

$^*p < .05$

dimension of the perceptual space, the first dimension was found to be moderately correlated with Spectral Shape. In terms of acoustic features, the two perceptual dimensions could be interpreted (allowing some interpretive margin) using HighFreqEnergy feature (instead of Spread) which loaded the highest on Factor 3 in the four-factor solution and which correlated more highly with the first dimension ($r = .72$, $p < .052$) and using the SubBand7Flux descriptor for the second dimension.

In regard to the Shost1 excerpt, both perceptual dimensions correlated highly and significantly with acoustic factors. Dimension 1 was found to be related to Brightness or Low Frequency Flux and Dimension 2 to Mid-High Frequency Flux (Table 7.27). As for acoustic features, they all correlated highly with the first dimension (with Centroid having the highest correlation coefficient) but weakly and not significantly with the second dimension. Looking back onto acoustic features that had the highest loadings in the four-factor solution it was found that the second perceptual dimension could be explained by descriptors such as ZeroCrossings ($r = .76$, $p < .05$) or SubBand5Flux ($r = -.75$, $p < .05$).

Table 7.27: Shost1. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
Brightness	-.87*	.37	Centroid	-.85*	.33
SubBand 4-10 Flux	-.36	-.79*	SubBand9Flux	-.80*	-.37
SubBand 1-3 Flux	-.73*	-.19	Spectral Flux	-.83*	-.11

$^{*}p < .05$

Table 7.28: Shost4. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
Brightness	-.46	.69	HighFreqEnergy	-.66	.60
Spectral Shape	-.63	-.44	Spread	-.64	-.41

Since none of the factors or acoustic features correlated significantly with any of the two perceptual dimensions in Shost4 (Table 7.28), the perceptual axes can only be tentatively interpreted with the Spectral Shape factor or equivalent Spread descriptor for Dimension 1 and Brightness or HighFreqEnergy for Dimension 2. Taking into consideration that the HighFreqEnergy descriptor correlates in fact more highly with the first dimension makes the interpretation of the perceptual dimensions ambiguous.

Finally, the mean factor scores and respective highest loading acoustic features obtained from the factor analysis of the entire dataset, i.e. the six music fragments combined, were correlated with the perceptual dimensions. Tables 7.29 and 7.30 show the correlation coefficients for the three- and two-factor solution respectively. In case of the three-factor structure, Dimension 1 correlated highly and significantly with Brightness (Factor 1) or the Centroid descriptor while Dimension 2 was found to correlate highly and significantly with SubBand 1-10 Flux (Factor 2) or SubBand10Flux. In fact, the first perceptual axis could be more effectively explained by SpectVariation following its higher correlation coefficient. As for the two-factor structure, only the SubBand 1-10 Flux factor correlated significantly with Dimension 1 and none of the factors with Dimension 2. In terms of acoustic features which correlated highly and significantly,

Table 7.29: All music styles combined. Three-factor solution. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
Brightness	-.78*	-.25	Centroid	-.76*	-.24
SubBand 1-10 Flux	-.31	.87*	SubBand10Flux	-.45	.80*
Total Spectral Flux	-.68	-.66	SpectVariation	-.91**	-.20

$^{*}p < .05$, $^{**}p < .01$

Table 7.30: All music styles combined. Two-factor solution. Correlations between the perceptual and acoustical dimensions and features ($N = 6$).

Factors	Dim 1	Dim 2	**Features**	Dim 1	Dim 2
Brightness	-.72	.04	Centroid	-.76*	-.24
SubBand 1-10 Flux	-.81*	.48	SubBand5Flux	-.53	.75*

$^{*}p < .05$

the two perceptual dimensions could be then interpreted using Centroid and SubBand5Flux respectively.

The above findings provide a significant input to the contribution of this work as they confirm the perceptual importance of the Brightness factor or equivalent descriptors of higher frequency content in the spectrum in discriminating not only between various orchestral instruments (refer to Section 2.3.4) or tones of just one instrument (see Section 2.5) but also between subtleties of different players' timbres performing on the same instrument. The role of the second factor, being the indicator of spectral variations over time across different frequency subbands or spectral fluctuations in particular subbands, comes also in agreement with the previous timbre studies (though different definitions of spectral flux/variation were employed depending on the research context) which indicated spectral flux as one of the major acoustical correlates of the revealed timbre spaces.

7.4 Summary

Tone samples of the six cellists used in the perceptual experiment were acoustically analysed in order to explain the source of timbral differences between

the players revealed by perceptual ratings. ANOVA based feature selection was applied to 25 initially extracted acoustic features to obtain subsets of features best capturing variations between the players depending on music style and character. Factor analysis of respective feature subsets revealed two, three or four acoustical dimensions best describing spectral characteristics of the cellists. The highest correlating features in each factor solution were selected to facilitate the interpretation of the acoustical dimensions. The emerging three main factors included Brightness, Spectral Shape and Spectral Variation or Spectral Flux which tended to split into Spectral Flux of particular frequency regions.

Results of the MANOVA tests conducted on the factor scores and the most correlated features across the entire dataset showed that the cellists can be discriminated based on their low-dimensional acoustic characteristics. Finally, the players' mean factor scores and feature values were correlated with the players' perceptual coordinates to find possible relationships. For the factor solutions across the six excerpts, the Brightness factor (and respective HighFreqEnergy and Centroid descriptors) was found to correlate most strongly with the perceptual dimensions followed by Spectral Flux of lower or higher frequency regions (and respective SubBand1Flux or SubBand3Flux and SubBand7Flux or SubBand9Flux features). The correlation analysis of the factor solution on the entire dataset revealed Brightness and SubBand 1-10 Flux factors to be the most linked with perceptual dimensions. In terms of features, however, SpectVariation and SubBand10Flux descriptors appeared to be the strongest acoustical correlates of the first and second perceptual dimension respectively.

Chapter 8

Identifying performer-specific bowing controls

In the previous two chapters it was shown that the players can be perceptually discriminated by listeners and that perceived dissimilarities have their source in significantly different acoustic characteristics of each player. The naturally emerging question is: what did the cellists do in terms of performance gestures to obtain such different tone effects?

To provide an answer, the combination of bowing controls is first analysed across music excerpts to explore how the bowing parameters were adapted in response to varying music scores. The individual bowing techniques are then compared in search for bowing patterns which might characterise a player regardless of the music being performed. Finally, a relation between characteristic bowing controls and acoustic features is established to examine to what extent manipulating performance gesture affects spectral content of the sound played.

8.1 Introduction

As already mentioned in Chapter 3, it is the bowing technique that is crucial for controlling the quality of sound on a bowed string instrument such as cello. It determines each subtle interaction between the bow hair and the string, giving an accomplished string player numerous ways of shaping the spectrum of a desired sound.

To compare bowing techniques of different players in search of the source of

their distinct tone properties, capturing their performance gestures is necessary. But what actually can be measured?

Bowing control parameters (bowing controls) are so far the only measurable variables of the complex bowing process which is fully controlled by a player. They are able to capture what is directly exerted on the instrument, i.e. the mechanics of the bowing process (discussed in more detail in Section 3.3.2). They include major controls such as bowing speed (further referred to as bow velocity), bow pressing force (or "bow pressure" as often called by musicians, further as bow force) at the point of the bow-string contact, and bow-bridge distance (the distance from the bowing point to the bridge). They may also include some auxiliary controls (refer to Figure 3.14) such as bow tilt (bow-string angle), bow inclination, and bow skewness (bow-bridge angle). Yet, all these parameters are very much dependent on the actual bow position (bow displacement), i.e. the current bowing point position between the frog and the tip.

Bowing variables have been first measured and systematically examined using bowing machines (see Section 4.2) followed by the use of dedicated motion tracking equipment to capture bowing gestures in normal playing scenarios (as detailed in Section 4.3.1). They uncovered physical limits to bowing parameter combinations available to a string player, in order to trigger and sustain Helmholtz motion in a bowed string, crucial for production of a good quality tone.

However, as they are designed to capture what is happening at the bowing point, bowing controls do not account for a performer's physique, aspects such as body height (affecting sitting position at the instrument), weight and height together with instantaneously adjusted relative position, speed and centre of gravity of the right hand, and the way that a player holds his bow (whether tightly or loosely, allowing the bow to vibrate freely). Neither are they able to show us how his technique developed over years, what playing school he may belong to nor how long and intensively he was practising to reach a master

technical level. They are rather the instantaneous resultants of all the above mentioned factors being in action. Therefore, for further development of this work, bowing controls measured in live performance are considered a gestural extension of the player and treated as a whole as his gestural identity.

8.1.1 Research questions and *a priori* remarks

1. What are the major differences in the use of bowing controls between the music excerpts which vary in style and genre?

   The six music fragments chosen for this study come from three distinct music styles, i.e. they represent Baroque, Romantic and contemporary music. Within each style the selected excerpts also vary in terms of genre or character taking as examples three baroque dances: Allemande, Courante and Bourrée from Bach's *3rd Suite* or different characters of 1st and 4th movements from Shostakovich's *Sonata*. These stylistic differences translate at the music score level into differences in tempo, articulation and dynamics which in turn have direct impact on the choice of bowing controls. For example, amongst the six excerpts the cheerful Courante is performed in the fastest tempo in opposition to the lyrical slow-paced Élégie. One may then expect that as the cellists adapt their bow velocity to music tempo there will be strong variations of the parameter between the pieces.

   Articulation indications are at the first place linked to bow pressing force or, as it is the case here, to bow-string distance. For example, larger values of the parameter may occur for *staccato* or *marcato* notes (Allemande, Courante, Shost4) and smaller values for phrases played *legato* (Bourrée, Shost1). At the same time, the bow-string distance oscillations are affected by dynamic levels. On average, for notes performed in *forte*, bow-string distance may be larger than for notes in *piano*. In this study, Allemande, Courante and Shost4 were performed in *mezzo forte*, Élégie and Bourrée in *piano*, and Shost1 in *mezzo piano*. Moreover, the string on which the notes are performed also has a compounding effect on bow-string distance.

Typically, the lower and thicker the string being played, the larger the bow-string distance.

In this particular experimental scenario, due to short music excerpts (reduced in length to facilitate the perceptual study) and by consequence small sample sizes, the effect of string and the effect of dynamics were considered as inherent parts of the musical piece effect and their aggregated impact is not evaluated. This, however, may constitute an interesting topic for a follow-up study conducted on broader and more diverse bowing data available from the multi-modal cello database described in Chapter 5.

2. Do the cellists differ in their choice of bowing controls when adapting for changes in tempo and articulation? Are there any individual bowing preferences regardless of performed music?

   Preliminary observations obtained from an earlier study (Chudy et al., 2013, not discussed in this thesis) suggest the existence of individual strategies especially in regard to the choice of bowing distance from the bridge. Amongst the six cellists two of them exhibit definitely antithetic preferences for this parameter which then are balanced by appropriate changes in the other bowing controls.

3. To what extent are the individual bowing controls related to acoustic features characterising the player's timbre?

   Having perceptual and acoustical dimensions linked together, the naturally occurring conclusion is that spectro-temporal characteristics of a player's tone must have their source in performer-specific bowing controls.

8.2 Method

In order to address to the above stated questions the experimental study was designed as follows.

8.2.1 Bowing data processing

The same music samples of six cellists representing different music styles and genres used in perceptual and acoustical studies (Chapters 6 and 7) were analysed. For each music excerpt (6 per cellist, 36 excerpts in total), a set of bowing parameters was computed from acquired bowing motion coordinates (details can be found in Section 5.6). The bowing controls included: (i) bow-bridge distance relative to the string length and fingering position (β), (ii) bow transverse velocity (v_B), and (iii) bow-string distance (z_{bs}), a measure of hair ribbon and string deflection under the bow pressing force used in this study as a simplified model of real bow force (*pseudo-force*).

Due to sensor instability at the edges of the sensing magnetic field, some motion coordinate readings were affected, causing substantial discrepancy in bow-string distance measurements across the recording sessions. As a consequence, it became impossible to compare bow-string distance values between the players in an absolute manner. Furthermore, the obtained bow-string distance or *pseudo-force* was intended as an auxiliary parameter for real bow force modelling. This operation involved bow force data acquired by means of a load cell, together with *pseudo-force*, bow position and tilt (all captured in the force calibration procedure) which were entered into a regression model based on Support Vector or Random Forests methods. Since the computed models did not produce satisfying results (due to erroneous bow-string distance measurements), it was finally decided to use the *pseudo-force* parameter itself as an approximation of working bow force after necessary normalisations.

The normalisation procedure consisted of finding the minimum and the maximum of the parameter across all excerpts per performer and rescaling all values to the [0,1] range so that they are comparable to those of the other players. It is important to note here that the standard bow-string distance values on cello can range from 0 to about 1.5 cm (the upper limit depends on hair ribbon tension, i.e. the lesser the tension the larger hair ribbon deflection may occur, yet within certain physical limits). In cases of, for instance *staccato* articulation, the bow is

lifted from the string at the end of each note and the bow-string values become negative (for more details on how the *pseudo-force* parameter was calculated refer to Marchini et al., 2011). The normalised parameter includes those cases so bow-string distance values near zero may indicate that the bow was actually off the string.

To proceed further, as in the case of the audio samples, the obtained bowing controls were segmented into notes to get more detailed insight into the parameters changes over time. To summarise the sequences of parameters regardless of the varying lengths of notes, the median values were calculated as representations of each control per note. The final bowing datasets consisted of 13, 15, 20, 6, 11 and 10 3-parameter vectors per cellist for Allemande, Bourrée, Courante, Élégie, Shost1 and Shost4 excerpts respectively.

8.2.2 Bowing data analysis

With such data the following three major experiments were conducted. Firstly, multivariate analysis of variance (MANOVA) combined with discriminant analysis (DA) were employed to study general use of bowing controls across music pieces. Secondly, individual bowing patterns among the players were identified by means of a repeated measures MANOVA and three follow up ANOVAs. Finally, correlation analysis was performed to examine possible relationships between acoustic feature and bowing control dimensions.

8.3 Results and discussion

8.3.1 Comparing general use of bowing controls across six musical contexts

In order to investigate whether there were significant changes in the use of the three bowing parameters depending on musical context, ANOVA-based analysis was carried out. Since players adjust all bowing controls simultaneously to maintain a desired quality of tone, it was therefore justified to test the effect of musical context on the three bowing parameters combined using a multivariate

Table 8.1: Bowing control means and standard deviations grouped by music excerpt ($N = 426$).

		z_{bs}		v_B $[cm/s]$		β	
Music excerpt	***N***	Mean	SD	Mean	SD	Mean	SD
Allemande	78	.59	.16	26.77	9.39	.17	.03
Bourrée	90	.36	.13	18.97	8.11	.20	.06
Courante	96	.59	.12	34.67	11.80	.15	.03
Élégie	36	.38	.16	13.41	4.89	.22	.05
Shost1	66	.48	.22	22.10	6.01	.17	.05
Shost4	60	.48	.17	16.03	9.38	.18	.05

design (MANOVA). In this scenario, the three bowing parameters across all six cellists served as the dependent variables, and the music excerpts comprised the six-level independent variable. If the musical context effect proves to be significant, further investigation can reveal the major differences in bowing controls across excerpts and whether they were related to tempo and articulation markings in the music scores.

Before conducting MANOVA, the bowing data was checked for the assumptions of univariate and multivariate normality, linearity, univariate and multivariate outliers, and multicollinearity but no serious violations were detected. However, preliminary screening of sample variances for each bowing control across music excerpts indicated that the assumption of homogeneity of variance-covariance matrices might be violated since the ratio of largest to smallest variance for bow-string distance exceeded 12:1 (Tabachnick and Fidell, 2007). To decrease the disproportion in group sizes between the largest (Courante – 120 notes in total) and the smallest (Élégie – 36 notes in total) datasets and to ensure robustness of the test, it was decided to remove from the Courante dataset 4 notes out of the 20 available per player. The removed notes were pitches 2, 8, 14, and 20, which had the largest bow velocity values. Table 8.1 summarises resulting group sizes together with respective descriptive statistics. The bivariate correlations for the bowing controls across all 426 notes are presented in Table 8.2.

Table 8.2: Intercorrelations among the three bowing parameters.

Parameter	z_{bs}	β	v_B
z_{bs}	–	-.43***	.20***
β		–	-.23***
v_B			–

***$p < .001$, $N = 426$

According to the MANOVA results, strategies in the use of combined bowing parameters significantly differed between musical excerpts (Pillai's Trace = .61, $F(15, 1260) = 21.61$, $p < .0005$), however the best linear combination of dependent variables accounted for only 20% of variance (effect size *partial* $\eta^2 = .20$). Univariate ANOVA tests for each bowing control, with a Bonferroni adjustment of alpha levels for multiple tests ($p < .05/3 \approx .017$), also showed significant variations between the pieces (Table 8.3), with the bow velocity showing the strongest effect followed by the bow-string distance and bow-bridge distance parameters (in each case an effect size was large, i.e. $\eta^2 \geq .14$).

Instead of analysing the effect of musical context on each bowing parameter separately, which would be a standard follow-up, it was more revealing to take advantage of the multivariate ability of MANOVA to discriminate between bowing strategies observed across music pieces. That is because MANOVA is statistically identical to discriminant analysis. To test whether mean differences among groups on a combination of dependent variables are likely to have occurred by chance, MANOVA creates a linear combination of measured dependent variables so that a new dependent variable maximally separates the groups,

Table 8.3: Investigating differences in the use of bowing parameters across six musical contexts. Results of univariate ANOVAs for each bowing control ($N = 426$).

Bowing parameter	F **statistics**	**Significance**	**Effect size****
z_{bs}	$F(5, 163.65)^* = 38.07$	$p < .0005$	$\eta^2 = .25$
v_B	$F(5, 178.75)^* = 53.59$	$p < .0005$	$\eta^2 = .39$
β	$F(5, 159.37)^* = 15.82$	$p < .0005$	$\eta^2 = .15$

*Welch's adjustment for Homogeneity of Variances, **partial η^2 reported

Table 8.4: Results of discriminant analysis on three bowing parameters across all six music pieces ($N = 426$).

Function	**Eigenvalue**	**% of Variance**	**Wilks' Λ and χ^2 statistics**	**Effect size***
1st	.99	88.8	$\Lambda = .45,\ \chi^2(15) = 339.64$	$\eta^2 = .50$
2nd	.08	7.5	$\Lambda = .89,\ \chi^2(8) = 50.87$	$\eta^2 = .08$
3rd	.04	3.7	$\Lambda = .96,\ \chi^2(3) = 17.14$	$\eta^2 = .04$

*partial η^2 reported

and ANOVA run on this new dependent variable tests hypotheses about group means (Tabachnick and Fidell, 2007). The linear combinations of dependent variables, called discriminant functions, are the core of discriminant analysis. The discriminant function coefficients are, in fact, regression weights and they represent exactly how dependent variables are combined to maximally discriminate between groups.

With three bowing parameters as dependent variables and six music excerpts as the levels of Piece main effect, three discriminant functions were found, the first two functions significant at the level $p < .0005$ and the third function significant at the level $p < .001$. As one can see from Table 8.4, the first discriminant function has the highest proportion of variance shared between the independent variable and first multivariate combination of dependent variables and provides the best separation among the musical excerpts based on the three bowing controls combined (note a very large effect size). The second discriminant function is orthogonal to the first and best separates the pieces on the basis of associations not used in the first function (about 7% of shared variance, a medium effect size). The third discriminant function (orthogonal to the former two), although sharing less than 4% of variance (a small effect size), is also important since its coefficients represent another combination of the bowing controls, not accounted for by the first two functions, offering an additional perspective on the bowing strategies. Figure 8.1 shows how individual notes (in terms of the combined bowing controls) are distributed along the first two discriminant functions. Each point on the chart represents a discriminant score calculated

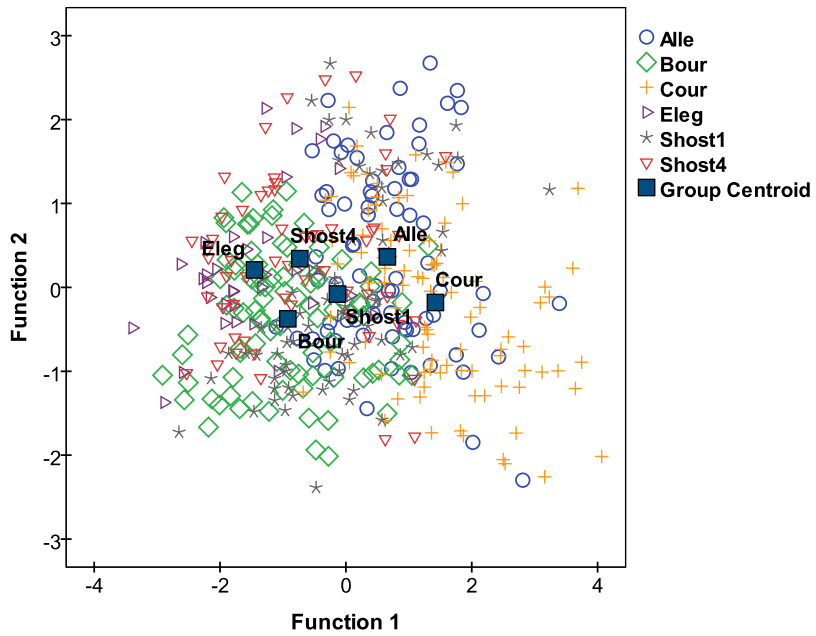


Figure 8.1: Discriminant analysis on three bowing parameters across all six music pieces. Points represent discriminant scores on the 1st and 2nd discriminant functions for each note in the dataset grouped by Piece ($N = 426$).

by multiplying the three bowing control values by their respective discriminant function coefficients (the coefficient values can be found in Table 8.5). The "group" centroids, also marked on the chart, are multivariate means of discriminant scores for each excerpt.

Discriminant scores on individual functions formed new multivariate composite variables which were subjected to further ANOVA analyses. Significant main effects of Piece on the first, $F(5, 420) = 82.92$, $p < .0005$, *partial* $\eta^2 = .50$, and on the second, $F(5, 168.05)$[1] $= 8.16$, $p < .0005$, *partial* $\eta^2 = .08$, multivariate composites are visualised in Figure 8.2. *Post-hoc* comparisons of Function 1 scores (Hochberg's GT2 test was chosen due to unequal group sizes) showed that Allemande and Courante differed significantly from each other and the rest of excerpts at $p < .0005$. Shost1 was also significantly different from the other excerpts at $p < .0005$ with exception of Shost4 for which the difference

[1]Welch's adjustment for Homogeneity of Variances

Table 8.5: Discriminant function and correlation coefficients ($N = 426$).

Bowing parameter	Standardized Coefficient w_{ij} for # Function			Correlation Coefficient r_{ij} for # Function		
	1st	2nd	3rd	1st	2nd	3rd
z_{bs}	.54	.90	.19	.54	.83	−.17
v_B	.82	−.46	.36	.78	−.53	.31
β	−.18	.32	.998	−.39	.03	.92

was found significant at $p < .013$. Shost4 differed significantly from Élégie at $p < .01$ but there was no difference in Function 1 scores between Shost4 and Bourrée nor between Bourrée and Élégie.

As for the second multivariate composite, *post-hoc* Games-Howell comparisons (used due to unequal group sizes and violated equality of variances) showed significant differences between Bourrée and Allemande, Élégie, and Shost4 at $p < .0005$, $p < .007$ and $p < .0005$ respectively. Also Courante differed significantly from Allemande at $p < .016$ and Shost4 at $p < .019$. There was also significant main effect of Piece found on the third multivariate composite, $F(5, 162.58)$[2] $= 3.36$, $p < .006$, *partial* $\eta^2 = .04$. However, the amount of variation explained was very small (3.7%) and *post-hoc* Games-Howell comparisons indicated only two significant differences, between Shost1 and Élégie at $p < .017$ and between Shost1 and Allemande at $p < .045$.

While ANOVAs conducted on the multivariate composites proved the existence of substantial variations in the use of combined bowing controls across the pieces, discriminant functions provided direct explanation of the source of these variations. In addition to the standardised discriminant coefficients, Table 8.5 contains also correlation coefficients which are equivalent to loadings in factor analysis (FA) and which constitute correlations between dependent variables and discriminant functions. Similarly to FA, the loadings are employed to facilitate interpretation of the results.

[2]Welch's adjustment for Homogeneity of Variances

As the first discriminant function reveals, it is bow velocity that had the highest impact (indicated by the largest coefficient ($r_{12} = .78$) on discriminating between music excerpts. Since it depends directly on tempo of a piece, any increase or decrease in bow velocity involves immediate adaptations of bow force and bow-bridge distance exerted by a performer. As suggested by the two other loadings on that function, an increase in bow velocity is combined with an increase in bow-string distance ($r_{11} = .54$) and a decrease in bow-bridge distance ($r_{13} = -.39$). In agreement with *a priori* observations, the positive extreme of the function was occupied by Courante while Élégie took the opposite one. It is important here to remember, before interpreting the next correlation coefficients, that each subsequent discriminant function explains variations not accounted for by the previous functions. The second discriminant function captured major differences between the pieces based on bow-string distance ($r_{21} = .83$). As two other coefficients indicated, larger bow-string distance was combined with reduced bow velocity ($r_{22} = -.53$) and almost no change in bow-bridge distance ($r_{23} = .03$). Along this bowing control dimension, the extremes belonged to Allemande and Bourrée. Finally, the third discriminant function differentiated between the music excerpts mainly based on bow-bridge distance ($r_{33} = .92$). In those cases larger bow-bridge distance was usually correlated with moderately increased bow velocity ($r_{32} = .31$) and slightly smaller bow-string distance ($r_{31} = -.17$).

Figure 8.2 compares the distribution of discriminant function scores across the music excerpts with their highest loading bowing controls. As one can notice, significant differences in bow velocity observed between the pieces (8.2b) are mostly repeated by the first discriminant function (8.2a) with only slight shift in the score means due to the combined influence of two other bowing parameters. When looking at bow-string distance means (8.2d), there is evident separation between the pair Allemande and Courante, which comprise mainly *staccato* notes mixed with a few short *legato* passages, all performed in *mezzo forte*, and Élégie and Bourrée, which group together as they contain only legato

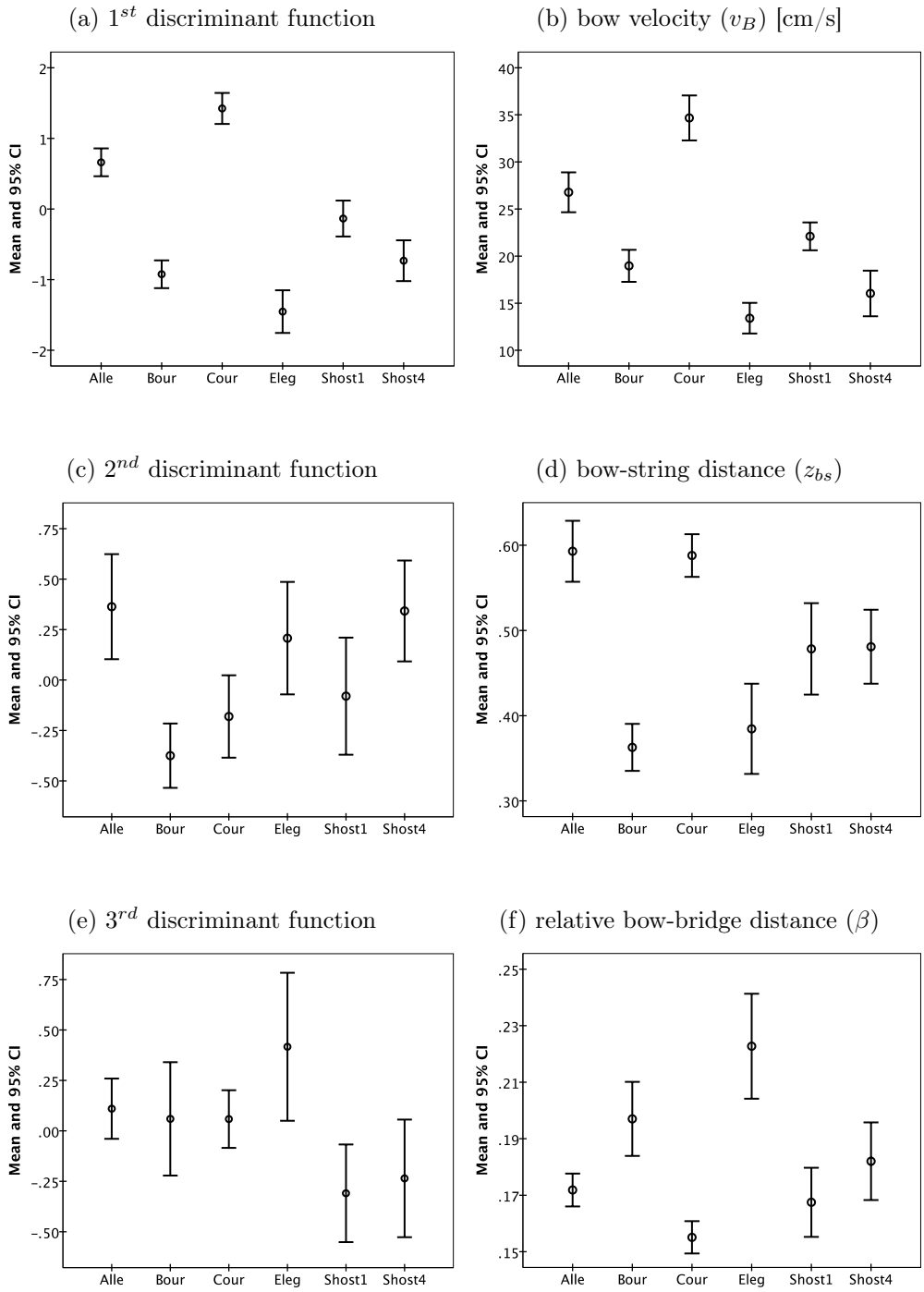


Figure 8.2: Comparison of mean discriminant function scores (a)–(c)–(e) and the highest loading bowing controls (b)–(d)–(f) across the six pieces, ($N = 426$).

notes played with the whole bow, in *piano* dynamics. Surprisingly, Shost1 and Shost4 excerpts are located in between although they have contrasting articulation and dynamics (*legato* passages played in *mezzo piano* in opposition to sharp staccato notes in *mezzo forte*). Significant differences in bow-string distance are, however, not strongly reflected in the second discriminant function (8.2c) since, in majority, they have been already accounted for by a linear combination of all three bowing controls in the first function. A similar scenario can be observed in the bow-bridge distance distribution (8.2f). The third discriminant function (8.2e) indicates significant differences in bow-bridge distance combined with bow velocity and bow-string distance that have not been captured by the first two functions.

These are clear examples of how bowing controls have been adjusted by the performers to execute tempo and articulation indications included in the music scores and confirms *a priori* predictions on the effect these two elements of music performance have on the bowing technique used.

8.3.2 Comparing the use of bowing controls amongst the players across six musical contexts

In the preceding section it was demonstrated how general bowing strategies varied with musical context. The next step was to investigate how the cellists individually adapted their bowing controls independent of the music performed. The first clue was provided by the third discriminant function which separated the six musical pieces mainly based on relative bow-bridge distance. Although significant, the discriminative power of the function was small. However, when the discriminant scores on that function, i.e. the third multivariate composite, were plotted against the first composite variable and grouped by Cellist it revealed interesting "bowing" behaviours of the players. As illustrated in Figure 8.3 by the cellists' centroids, on average Cellist 4 played the furthest from the bridge regardless of the music performed, followed by Cellist 1 and Cellist 3 who played at moderate distances, and finally Cellist 5, Cellist 6 and Cellist 2

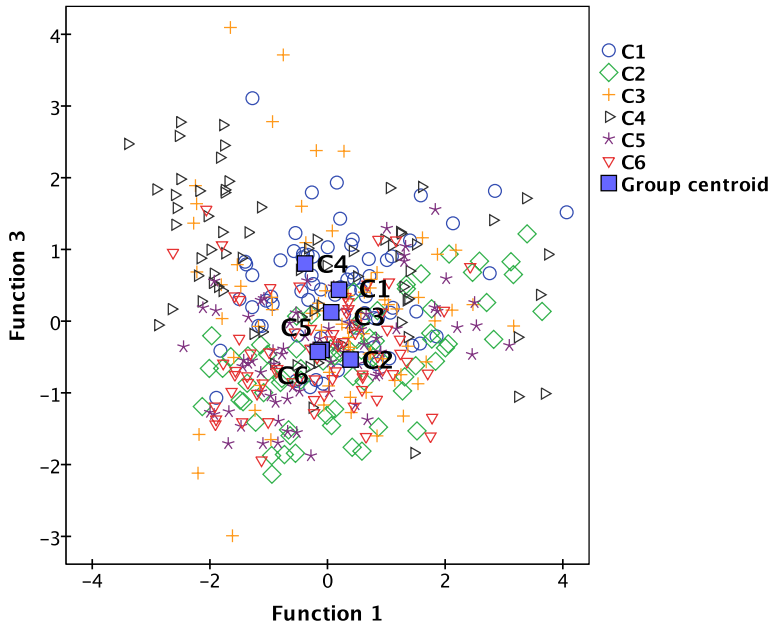


Figure 8.3: Discriminant analysis on three bowing parameters across all six music pieces. Points represent discriminant scores on the 1st and 3rd discriminant functions for each note in the dataset grouped by Cellist ($N = 426$).

who played closest to the bridge. At the same time, differences in choice of the bowing distance from the bridge across the players were associated with differentiated bow-string distance parameter as shown in Figure 8.4. Higher levels of bow pressing force were typical for Cellist 1 and Cellist 6 in opposition to Cellist 4 and Cellist 5 who on average were using less bow force.

To investigate whether different bowing strategies initially observed amongst the players were valid and significant, a one-way repeated measures MANOVA design was applied to the six cellists' bowing control datasets combined together. Similarly to the analysis conducted in Chapter 7, notes were treated as subjects exposed to six different conditions, i.e. being performed by six different players, and the three bowing controls served as dependent variables. An alternative option to examine individual bowing patterns in each excerpt separately using multivariate analysis was also considered. However, due to the limited number of cases, i.e. notes per player in each dataset, resulting in insufficient degrees of

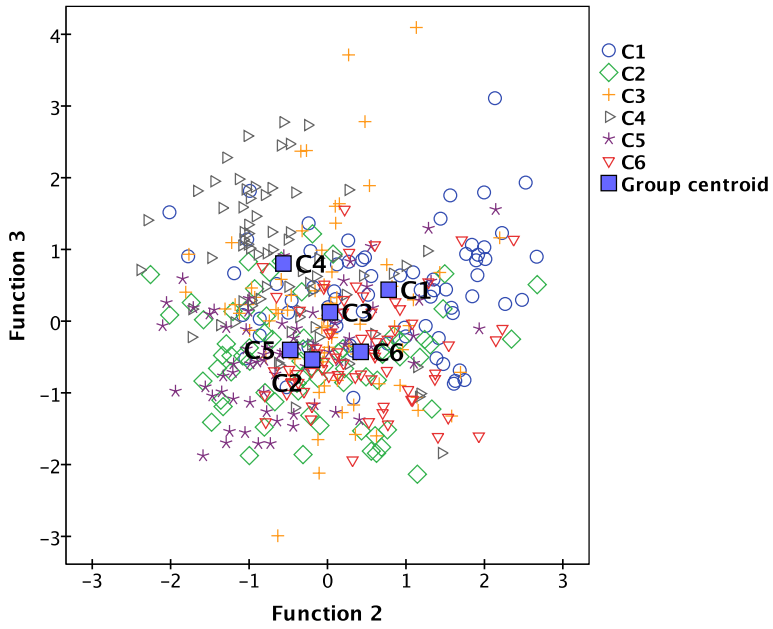


Figure 8.4: Discriminant analysis on three bowing parameters across all six music pieces. Points represent discriminant scores on the 2nd and 3rd discriminant functions for each note in the dataset grouped by Cellist ($N = 426$).

freedom for the error component, such analysis was not possible.

Before proceeding with MANOVA, preliminary evaluation of underlying normality assumptions did not reveal any substantial anomalies, and the *a priori* level of significance was set at .05. A visualisation of the three bowing parameters across the cellists presented in Figure 8.5 suggested notable Cellist effect in bowing control distributions, especially in regard to the bow-string and bow-bridge distance parameters.

MANOVA results confirmed the above findings, yielding a significant main effect of Cellist on the three bowing controls combined, Wilks' $\Lambda = .12$, $F(15, 56) = 28.05$, $p < .0005$, *partial* $\eta^2 = .88$ (i.e. 88% of variance explained). Follow up univariate ANOVAs, with a Bonferroni adjustment of alpha levels for multiple tests ($p < .05/3 \approx .017$), indicated that each of the three controls differed significantly between the cellists (Table 8.6), with the bow-string distance showing the strongest variations followed by the bow-bridge distance and bow velocity

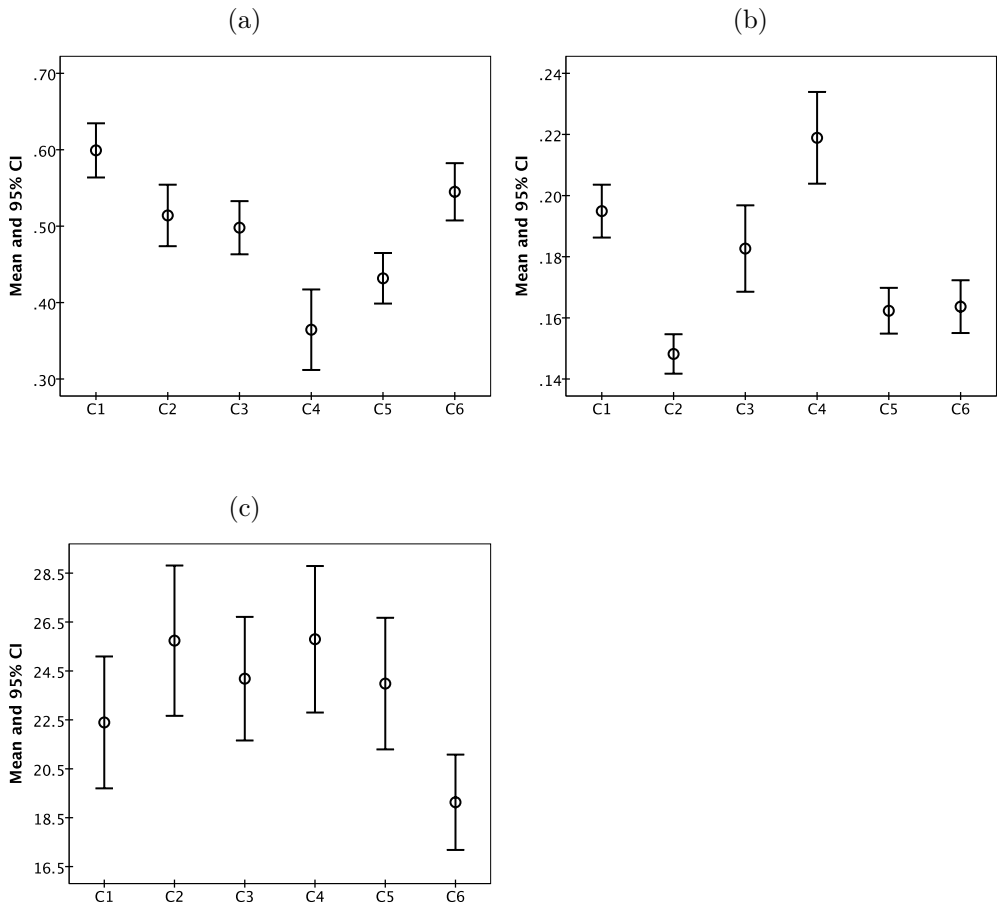


Figure 8.5: Mean bowing parameters across the six cellists ($N = 426$): (a) bow-string distance (z_{bs}); (b) relative bow-bridge distance (β); (c) bow velocity (v_B) [cm/s].

parameters (note respective very large and medium effect sizes).

A Bonferroni *post-hoc* comparison demonstrated that in terms of bow-string distance (Figure 8.5a) Cellist 1 used significantly the largest bow force compare to the others ($p < .0005$) except for Cellist 6 for whom the difference was significant at $p < .031$. Cellist 2 playing at moderate bow force levels differed from the others at $p < .0005$ except for Cellists 3 and 6, and Cellist 4 played with significantly the smallest bow force ($p < .0005$) except for Cellist 5 for whom the difference was significant at $p < .007$. At the same time, Cellist 4 played significantly the furthest from the bridge ($p < .0005$) when compared to

Table 8.6: Investigating differences in the use of bowing parameters between the cellists across six musical contexts. Results of univariate ANOVAs for each bowing control ($N = 426$).

Bowing parameter	F **statistics**	**Significance**	**Effect size****
z_{bs}	$F(3.12, 218.30)^* = 53.33$	$p < .0005$	$\eta^2 = .43$
β	$F(2.77, 193.81)^* = 42.47$	$p < .0005$	$\eta^2 = .38$
v_B	$F(4.22, 295.49)^* = 8.92$	$p < .0005$	$\eta^2 = .11$

*Greenhouse-Geisser correction for Sphericity, **partial η^2 reported

the other players (Figure 8.5b), while Cellist 2 performed significantly closest ($p < .0005$) with exception of Cellists 5 and 6 for whom, although they played relatively close to the bridge, the difference with Cellist 2 was still significant at $p < .002$ and $p < .001$ respectively. Cellist 3, using mid bow-bridge distances, differed significantly from the others at $p < .0005$ except for Cellists 1 and 5, and there was no significant difference in the control use between Cellists 5 and 6. Finally, the only significant difference in bow velocity ($p < .0005$) was indicated for Cellist 6 (Figure 8.5c) who on average played at slower tempi (and subsequently used lower v_B) compared to the rest of the players (at $p < .001$ compared to Cellist 5).

8.3.3 Correlation between bowing controls and acoustic features

In Chapter 7, sets of acoustic features were extracted from tone samples of the cellists in order to find their timbre characteristics in each music excerpt performed. Subsets of preselected spectro-temporal descriptors were then subjected to factor analysis to obtain a number of acoustic factors best describing the source of timbral differences between the players. The mean factor scores and the highest loading features were correlated with the perceptual coordinates for players to reveal which of the spectral characteristics most affected the listeners' perception so they were able to perceive the players' tones as distinctly different.

Since the resulting timbre characteristics seem to be strongly dependent on

Table 8.7: Correlations between the three bowing parameters and perceptually linked acoustic features for each music excerpt and all excerpts combined.

Music excerpt	Parameter	Acoustic feature		
		Flatness	HighFreqEnergy	SubBand1Flux
Allemande	z_{bs}	-.24*	.24*	.54***
$N = 78$	β	.03	.06	.24*
	v_B	.46***	-.04	-.38***
		SubBand9Flux	Centroid	SpectVariation
Bourrée	z_{bs}	.39***	.06	.08
$N = 90$	β	-.29***	-.22*	-.11
	v_B	.49***	-.43***	-.003
		SubBand3Flux	SubBand9Flux	SubBand7Flux
Courante	z_{bs}	-.26**	.03	.07
$N = 96$	β	.21*	-.34***	-.20*
	v_B	-.22*	.44***	.19
		SubBand7Flux	SubBand3Flux	HighFreqEnergy
Élégie	z_{bs}	.40**	-.02	.46**
$N = 36$	β	-.06	-.13	.10
	v_B	.22	-.23	-.07
		Centroid	SubBand9Flux	SubBand5Flux
Shost1	z_{bs}	.21*	.49***	.31**
$N = 66$	β	-.43***	-.58***	-.29**
	v_B	.40***	.54***	.53***
		HighFreqEnergy	Spread	
Shost4	z_{bs}	.32**	-.05	
$N = 60$	β	-.17	.16	
	v_B	.49***	.40***	
		Centroid	SubBand10Flux	SpectVariation
All combined	z_{bs}	-.06	.39***	.38***
$N = 426$	β	-.14**	-.29***	-.12**
	v_B	.00	.53***	.21***

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

the bowing technique used by the players, the next step was to explore whether such a relationship truly exists and to what extent the choice of bowing controls affects the spectral content of the sound. For each of the six music excerpts and

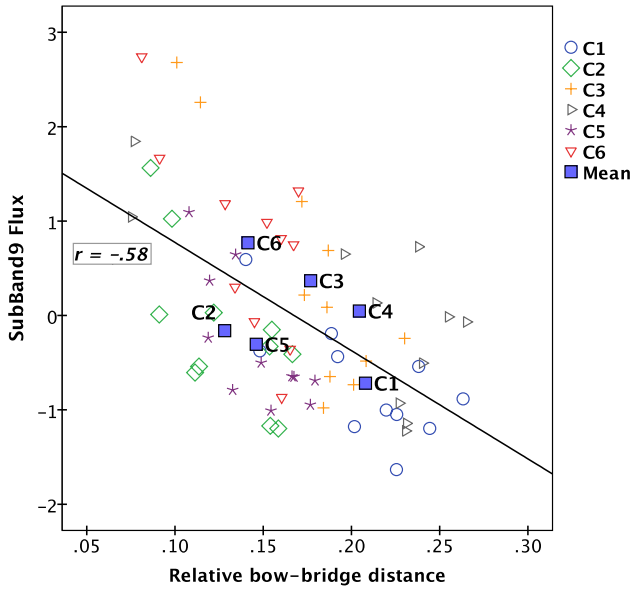


Figure 8.6: Shost1. The SubBand9Flux values plotted against relative bow-bridge distance and grouped by Cellist, with a least squares regression line marked ($N = 66$).

for all excerpts combined together, the three bowing controls were correlated with those acoustic features which were most related to the perceptual coordinates of the cellists, as shown in Section 7.3.4. It can be seen from Table 8.7 that for each music excerpt at least one spectral descriptor can be effectively explained by the linear combination of three bowing parameters as suggested by moderately large and significant correlation weights. For example, in Allemande, a cello tone with stronger fluctuations in SubBand 1 (SubBand1Flux) is likely to be a result of (or at least co-occur with) the performer playing his passage with reduced bow velocity, slightly further from the bridge and with increased bow pressure. Similarly, in Bourrée, any stronger fluctuations in SubBand 9 (SubBand9Flux) characterise the tone of cellists who played substantially faster, closer to the bridge and with larger bow pressure. An exemplary intercorrelation between a spectro-temporal descriptor and bowing parameter is illustrated in Figure 8.6. It is important to note that the relationships between particular acoustic features and bowing controls revealed for each music context

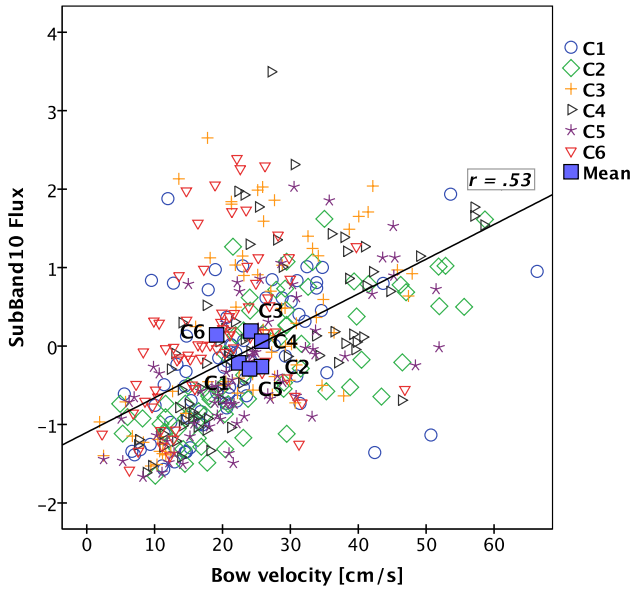


Figure 8.7: All music pieces combined. The SubBand10Flux values plotted against bow velocity and grouped by Cellist, with a least squares regression line marked ($N = 426$).

cannot be generalised as they are results of performance gesture choices made in relation to expressive elements such as tempo, articulation and dynamics to interpret a particular music score.

More general insight into bowing parameters and acoustic characteristics dependencies can be gained from the results of correlation analysis performed on the entire dataset (shown in the last row of Table 8.7). Interestingly, while both perceptually linked spectro-temporal descriptors, SubBand10Flux and Spect-Variation, can be predicted using a linear combination of three bowing controls (note the significant, but varying in magnitude, correlation coefficients), no stronger relationship between any bowing control and Centroid has been found. This result is a bit surprising knowing that, in Shost1 for example, Centroid was moderately and significantly correlated with bow velocity and bow-bridge distance and weakly with bow-string distance (refer to Appendix C to see an illustration of two different combinations of bowing controls and their effect on the resulting tone spectra, as observed in Shost1). In terms of bowing technique

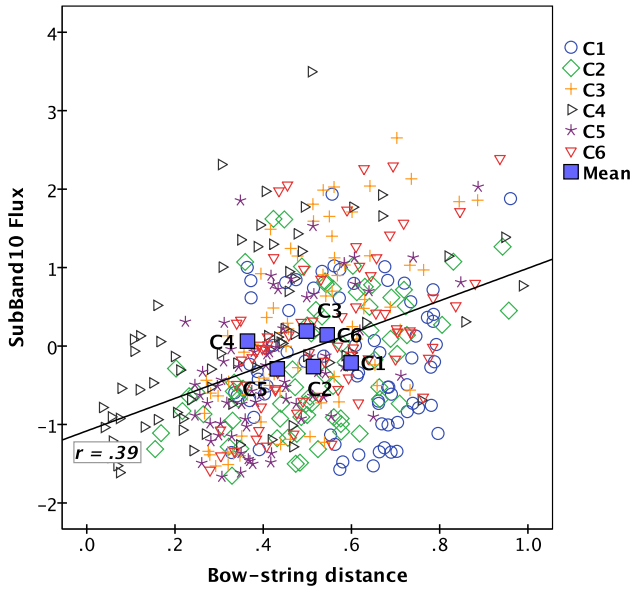


Figure 8.8: All music pieces combined. The SubBand10Flux values plotted against bow-string distance and grouped by Cellist, with a least squares regression line marked ($N = 426$).

applied, as suggested by correlation weights quite similar in magnitude to those of SubBand9Flux in Bourrée, stronger variations in SubBand 10 might occur due to playing with greater bow speed, closer to the bridge and with larger bow pressure.

Figures 8.7–8.9 illustrate the relation between SubBand10Flux and the three bowing parameters across the entire dataset. Comparing note clusters of the Cellists one can see that although they generally tend to overlap, some differences between the players are also observed. They are slightly more noticeable when looking at the cluster means. The least average difference between Cellists was found for bow velocity followed by larger dissimilarities in the use of bow force (bow-string distance) and bow-bridge distance. In regard to SubBand10Flux, as its values may suggest, the timbres of Cellists 3, 4 and 6 had on average more fluctuations in this subband than those of Cellists 1, 2 and 5.

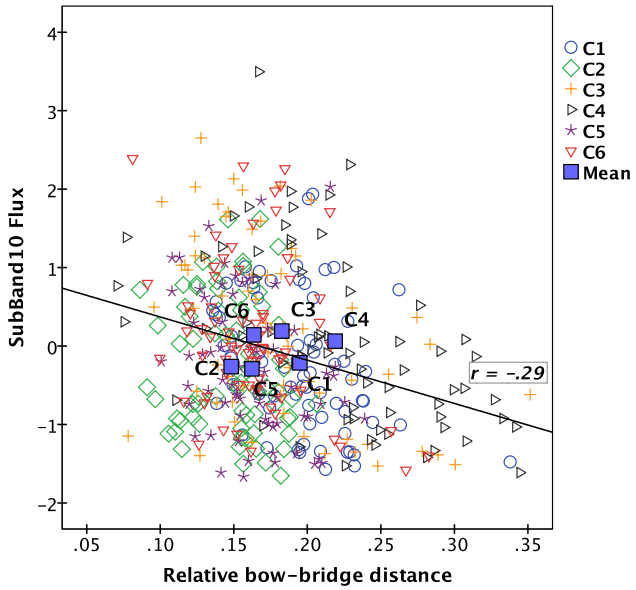


Figure 8.9: All music pieces combined. The SubBand10Flux values plotted against relative bow-bridge distance and grouped by Cellist, with a least squares regression line marked ($N = 426$).

8.4 Summary

By using multivariate analysis of variance it was possible to track general bowing strategies of the six cellists in the recorded musical fragments which varied in terms of music style and genre. The musical markings of the tempo, articulation and dynamics related to each interpreted score, when executed by the players, had a significant effect on their choice of bowing controls. The results suggest that main bowing parameters such as bow-string distance, bow-bridge distance and bow velocity simultaneously controlled by each cellist were first adapted for changes in tempo followed by changes in articulation and dynamics. It was shown that significant differences between the music excerpts had their source in substantial variations in bow velocity followed by lesser variations in the bow-string and bow-bridge distance parameters.

Additional multivariate analysis revealed that, apart from general adaptations to the requirements of each music score observed among the players, there

were strong individual differences in relation to the bowing controls used, regardless of the music performed. These were related primarily to the choice of bowing distance from the bridge and bow-string distance, i.e. bow pressing force by approximation.

Finally, interrelationships between each bowing parameter and acoustic features most correlated with perceptual dimensions were examined for each music excerpt and for all excerpts combined together. The results indicated that only a moderate proportion of spectro-temporal descriptor variations could be explained by a linear combination of the three bowing parameters. What it suggests is that a simple correlation measure may not be sufficient to describe the mapping between gestural input of a player and acoustical output of an instrument and a more complex model of the relationship may be required.

Chapter 9

Final notes and conclusions

This chapter concludes the thesis providing additional comments on the revealed links between perceptual, acoustical and gestural aspects of a player's timbre (Section 9.1), followed by a summary of the main findings in Section 9.2 and further directions for future work and potential applications in Sections 9.3–9.4.

9.1 Final notes on the relation between gesture, tone quality and perception

The experiments carried out over the course of this work aimed at answering the research questions stated in Chapter 1, i.e. whether classical musicians can be discriminated: i) perceptually, by timbre dissimilarity, ii) acoustically, by measured sound characteristics of their tones, iii) gesturally, by bowing controls used, and iv) whether any quantitative interrelations between the perceptual, acoustical and gestural domains exist.

As the results in Chapters 6–8 demonstrated, timbres of the six cellists were generally perceived as distinctly different and the revealed two perceptual dimensions seemed qualitatively linked to the levels of brightness and roughness in the players' tones. Acoustically, timbral differences between the cellists, regardless of music performed, were best observed in the three-dimensional space spanned between SubBand10Flux, Centroid and SpectVariation descriptors, with SubBand10Flux being the strongest discriminator (see Figure 9.1). In terms of performance gesture, i.e. the bowing mechanics behind the actual tone production, there were found combinations of bowing parameters specific for each player,

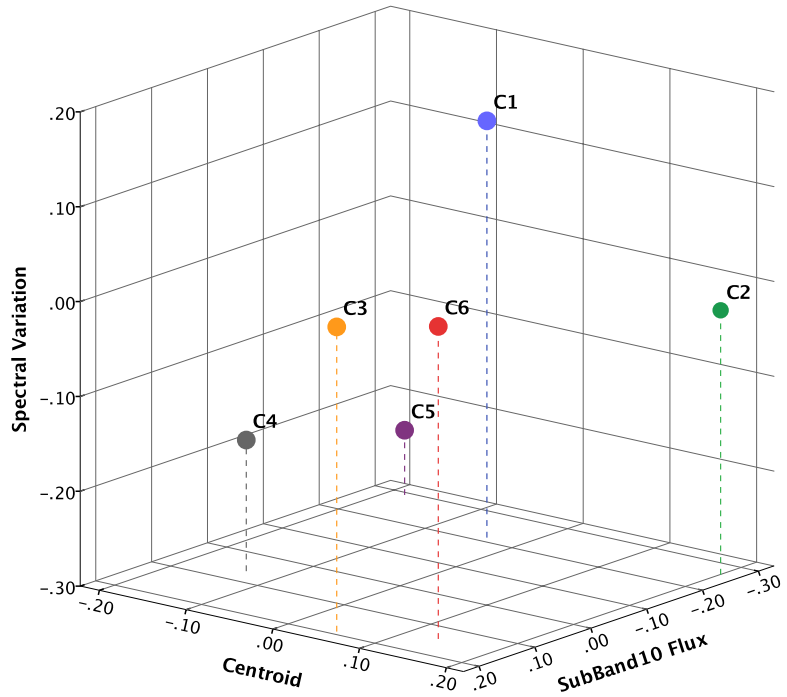


Figure 9.1: Three-dimensional acoustical space for the six cellists. Each point represents the acoustic features averaged across all music styles ($N = 426$).

which can be traced across different music contexts (see Figure 9.2). Finally, it was examined how these player specific combinations of bowing controls translated into his acoustic characteristics, and then into his perceptually distinctive timbre.

The correlation analyses in Section 8.3.3 indicated that, particularly for the aggregated data (across pieces and players), none of the bowing controls correlated specifically strongly with any of the acoustic descriptors to become its "mechanical" determinant. In the majority of cases, it was the combination of the three parameters, with usually one parameter loading slightly higher, which controlled the spectral contents of the tone.

A particularly interesting result was obtained for the Centroid descriptor. Based on the cello aggregated data, it seemed practically independent of any

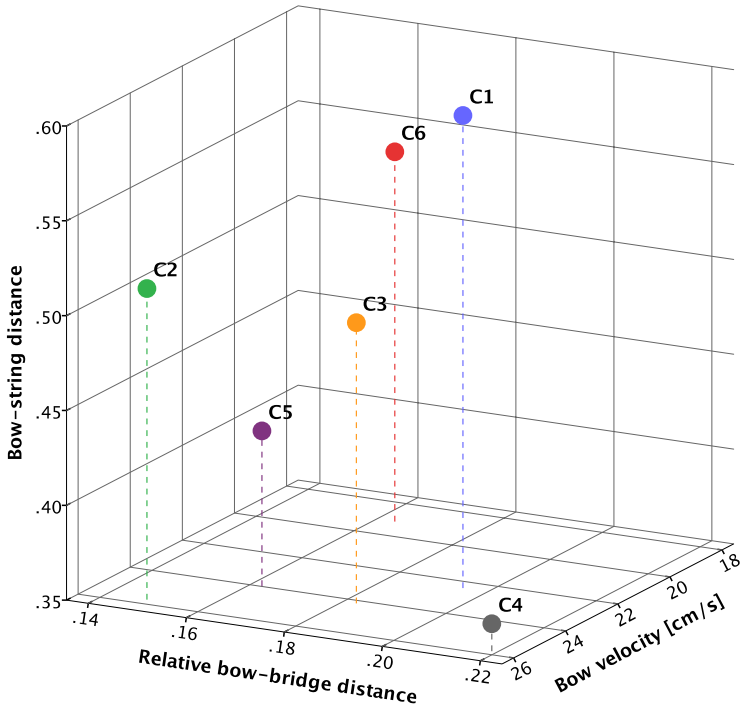


Figure 9.2: Bowing control space for the six cellists. Each point represents z_{bs}, β and v_B parameters averaged across all music styles ($N = 426$).

bowing control, though it was found to be the second main acoustical discriminator between the cellists, as the statistical analyses carried in Chapter 7 revealed. This result, however, generally agrees with Schoonderwaldt's study on violin playing (2009a), which showed the effect of bowing parameters on the spectral centroid to consequently diminish, from being substantial on the violin lowest string G to minor on the highest string E. His suggested explanation of the phenomenon is that, since the higher strings have lower characteristic impedance and internal damping, the damping which occurs due to fingering may play an increasing role in shaping the spectrum. He also suggests that *vibrato* "might cause additional fluctuations in spectral centroid without a direct relation with the bowing parameters" (Schoonderwaldt, 2009a). If this is the case, then, indeed, when analysing the spectral contents across the entire

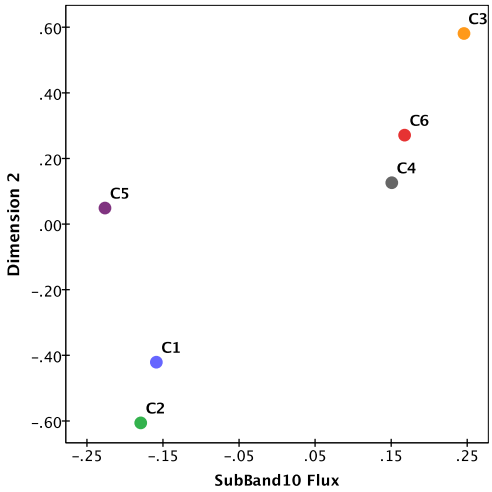


Figure 9.3: The averaged SubBand10Flux values plotted against the second perceptual dimension coordinates of the six cellists.

cello dataset, which comprised a mixture of different pitches played on all four strings, with and without *vibrato*, and with different articulations, the effect of bowing controls on Centroid could possibly no longer be observed.

The revealed strong links between acoustical and perceptual dimensions (see Table 7.29) imply that Dimension 1 or *tone brightness* can be effectively explained by SpectVariation and Centroid and Dimension 2 by the SubBand10Flux descriptor. This can be illustrated by Cellists 4 and 5's positioning in both spaces. The two players' timbres, having relatively low contents of higher components in the spectrum and less varying spectrum over time, are well separated from the others, and are also perceptually discerned as less bright (see Figures 6.2 and 6.3). The distinction between Cellists 4 and 5 themselves can be attributed to differences in the amount of fluctuation in the highest frequency band (SubBand10Flux) or tentatively to differences in the amount of roughness. However, in the perceptual space, it is more evident in their positioning along the first (varying brilliance) rather than along the second dimension. Figure 9.3 shows the mapping between SubBand10Flux and Dimension 2's coordinates of the cellists. While the respective data points of others seem to follow more or

Table 9.1: Correlations between the perceptual dimensions and bowing controls for the six cellists. Bowing parameters averaged across all music styles ($N = 6$).

Parameter	Dim 1	Dim 2
z_{bs}	-.71	-.33
β	-.00	.25
v_B	.44	-.25

Table 9.2: Correlations between the three performer domains based on calculated proximities between the cellists in gestural, acoustical and perceptual spaces ($N = 15$).

Space	gestural	acoustical	perceptual
gestural	–	-.35	-.31
acoustical		–	.80***
perceptual			–

***$p < .001$

less a straight line ($r = .80$, $p \leq .05$), Cellist 5 deviates from the general trend.

From analysing correlations between coordinates of the cellists in the bowing control space and their perceptual counterparts, the obtained correlation weights (shown in Table 9.1) suggest that there was a relatively strong (though not significant) negative relationship between Dimension 1 and bow-string distance and moderate (also non-significant) positive correlation with bow velocity but no dependence on bow-bridge distance whatsoever. Since Dimension 1 was qualitatively linked to the perceived brightness of cello timbre (Chapter 6), this result is consistent with Guettler et al. (2003), Schoonderwaldt et al. (2003) and Schoonderwaldt (2009b)'s earlier studies on violin, in which brilliance of the tone, or higher harmonic contents in acoustical terms, was found to increase with bow force and to decrease with increasing bow velocity but was not affected to any noticeable level by varying bowing point. In regard to Dimension 2, the observed perceptual differences between the cellists along this dimension can not be directly attributed to any of the bowing controls nor to their combination (at least based on simple linear regression), as weak and non-significant correlation weights indicate.

To this point, it was shown that music performers can be discriminated by

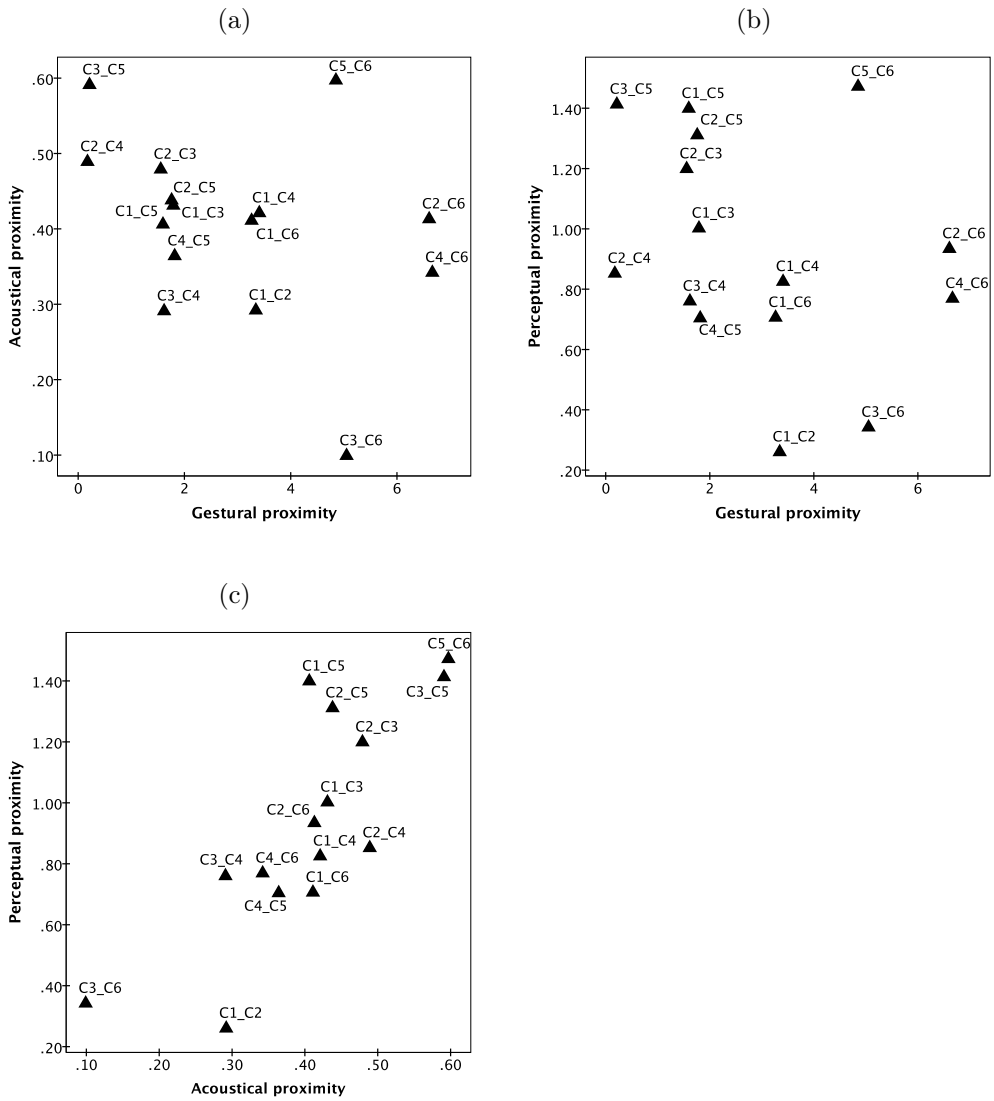


Figure 9.4: Interrelations between the three performer domains. Bowing controls space mapped into (a) acoustical and (b) perceptual spaces; (c) mapping between acoustical and perceptual proximities. Each point represents the average distance between a pair of players across all music styles ($N = 15$).

gesture, as well as acoustically and perceptually, and that there were meaningful interrelations found between the three domains. Consequently, it was interesting to examine to what extent the observed players' dissimilarities (distances) remained preserved across the domains. Figure 9.4 illustrates the relationships between the performer domains based on calculated proximities between the cellists in each domain. Figures 9.4a–9.4b show the bowing control space mapped

into respective acoustical and perceptual spaces. One can see that gestural proximities do not translate into acoustical or perceptual ones in a linear manner, which weak and non-significant correlation coefficients confirm (see Table 9.2). In contrast, there is a strong and significant linear relation observed between acoustical and perceptual domains (Figure 9.4c) which map into each other more accurately ($r = .80$, coefficient of determination $R^2 = .64$ indicating 64% of explained variance).

What these results imply is that the transformation from the gestural input into acoustical output is much more complex and relying on just three bowing controls to account for all the gestural variations might simply be not sufficient. It may also suggest searching for regression models other than linear, capable to capture the transition between the two domains more efficiently (see Future work section). On the other hand, sound qualities translate into what is ultimately perceived by listeners reasonably well. It can be accredited to the fact that acoustical correlates of perceptual dimensions have long been studied (see Chapter 2) and the revealed spectral and spectro-temporal features are well defined. It can also be attributed to the two-stage feature selection process involving ANOVA and Factor Analysis (see Sections 7.2.2 and 7.3.1) which enabled determination of acoustic features best discriminating between the players.

9.2 Summary of contributions

In this thesis player-dependent aspects of musical timbre have been investigated from perceptual, acoustical and gestural perspectives. While focusing on cello timbre, the objective was to find individual characteristics of a player in each performance domain and to examine whether these characteristics can be projected into each other across domains.

The investigation started with the collection of multi-modal solo cello recordings which included motion tracking data for extracting bowing control parameters (Chapter 5). This dedicated dataset comprises tone samples of six

advanced cello players captured on two different instruments played with the same bow in controlled recording conditions. The recorded audio tracks include both ambient near-field microphone and bridge pickup signals which allow to carry out comparative acoustical analyses. The database provides timbrally diverse musical material in terms of instrument characteristics (two different cellos), musical context (scales and three different music styles: Baroque, Romantic, and contemporary), articulation (varied articulation in recorded scales and Baroque music excerpts), dynamics (varied dynamic levels in Bach's *Bourrée* and Fauré's *Élégie*), and vibrato (*con vibrato* and *non vibrato* variants of all Baroque fragments). The accompanying bowing control data includes extracted main bowing parameters such as bow-string distance (approximation of bow force), bow velocity and bow-bridge distance, as well as auxiliary controls such as bow transverse position, bow acceleration, bow-bridge angle (skewness), bow tilt, bow inclination, and string estimation. Considering a certain margin for detected measurement errors, the collected gesture data provides essential details about individual bowing techniques of the players, which can then be analysed and compared.

With the database created, a perceptual experiment was designed, which aimed at revealing whether listeners can discriminate between the cellists' tones and whether the observed timbral differences (if any) can be described in semantic terms (Chapter 6). The stimuli consisted of six short music samples, varying in music style and genre, extracted from each player's set of ambient recordings on Cello1. In the experiment, twenty expert subjects were presented with pairs of samples of an identical music excerpt performed by two different cellists. Their task was to rate perceived timbre dissimilarity on a 0–10 continuous scale. The same group of subjects was also asked to evaluate the qualitative difference between the players in each pair using verbal attributes such as *bright*, *rough* and *tense*. Differential judgements were collected by weighting the presence of an attribute in compared samples.

The obtained results revealed that each cellist's timbre was perceptually distinct in every music fragment as well as on average across varying music styles, and that the timbral distinction could be attributed, along the first axis of the underlying two-dimensional perceptual space, to the perceived level of brilliance or brightness of tone. The second perceptual axis was more difficult to interpret. Verbal attribute ratings indicated *tone roughness* as a second discriminator (after *brightness*), however there were disagreements between the two solutions in positioning the cellists along this dimension.

In terms of methodology, the suitability of semantic differential judgements in combination with correspondence analysis (CA) for qualitative evaluation of tone quality was examined. Subjects' voting on tone samples with stronger presence of a particular attribute had the advantage of having the players directly ranked according to that attribute, without necessarily quantifying its magnitude in reference to some arbitrarily set up maximum and minimum levels (individually by each subject in fact) as in VAME ratings. A total number of votes on each attribute per player formed a measure of its respective strength to be compared across the cellists. Despite higher levels of disagreement in ratings between the subjects, which resulted in decreased reliability of CA solutions, the application of correspondence analysis offered an interesting alternative to a standard VAME ratings plus factor analysis (FA) or principal component analysis (PCA) approach as well as provided a graphical representation of the association revealed between the cellists' timbres and their semantic descriptions.

In respect of the choice of subjects in designing listening experiments on bowed string instruments, the results showed that there was no difference between string players (whether cellists, violinists or viola players) in their ability to perceptually evaluate timbre subtleties of a bowed string instrument such as cello, and suggested that they can be employed interchangeably as expert listeners in perceptual studies on the strings. This ability may also extend to pianists who specialise in the strings' repertoire.

A series of acoustical analyses on the cellists' tone samples selected for the perceptual experiments followed in Chapter 7, aiming to identify salient features which best capture varying timbral characteristics of the players and can facilitate their discrimination. The initial set of twenty five temporal, spectral and spectro-temporal descriptors came mainly from the audio feature set proposed by Alluri and Toiviainen (2010) which, among others, included Spectral Flux calculated in ten octave-scaled subbands of the spectrum. Frame-based vectors of features were extracted at the note level from a total of 36 music samples. The median value per note was computed to obtain a compact representation of each descriptor.

Thus prepared, the acoustic feature sets were further subjected to a feature selection process. This step was crucial in order to determine descriptors most effectively capturing variability across cellists regardless of varying pitch. Since a significant interaction was found between cellist and music excerpt, ANOVA based feature subset selection was run separately for each music fragment.

To uncover an underlying structure of acoustical dimensions and form a compact acoustical representation of each player, factor analysis (FA) was conducted on each excerpt's feature subset as well as on the six excerpts combined together. Advantage was taken of principal axis factoring (PAF) as the factorisation method which extracts the shared variance of a variable partitioned from its unique variance. The results revealed that up to three factors were needed to describe varying timbral characteristics of the cellists. They included indicators of: high frequency energy content plus noisiness (Brightness), the amount of variation of the spectrum components over time (Spectral Variation or Spectral Flux), and the spectrum distribution (Spectral Shape). For the factor solutions obtained on the combined dataset, the factors Brightness and Spectral Shape merged into one dimension, spectral fluctuations across all ten subbands became the second dimension and the overall spectral flux was captured by Dimension 3. MANOVA designs were applied to derived factor

scores and respective highest loading descriptors in order to examine the possibility of acoustically differentiating between the cellists regardless of the music performed. In all cases significant differences were found between the players' spectral characteristics, whether two- or three-dimensional. For the three-dimensional timbre space in particular, Total Spectral Flux was the strongest discriminator amongst the factors and SubBand10Flux amongst the correlated features. Correlations calculated between acoustical and perceptual dimensions suggested that the Brightness factor (or combined Centroid and SpectVariation descriptors) may explicate perceived nuances of *tone brightness* or brilliance while factor SubBand 1-10 Flux (or SubBand10Flux respectively) may account for *tone roughness*.

The next phase of this investigation (Chapter 8) focused on exploring differences in performance gesture, and bowing control parameters in particular, which were hypothesised to be the source of substantial differences found in spectral characteristics of the six cellists in question. The same six music excerpts were investigated in terms of bowing controls derived from the motion tracking data which accompanied each player's audio recordings. The bowing parameters, extracted at the note level with parallel to the note-based audio features, included bow-bridge distance (relative to the string length and fingering position), bow transverse velocity and bow-string distance (the approximation of bow pressing force, so-called *pseudo force*).

The first experiment employed MANOVA combined with discriminant analysis (DA) to study general use of bowing controls across music pieces. The results showed a strong tendency amongst the players to simultaneously adapt the three parameters in order to execute musical markings of the tempo, articulation and dynamics across the interpreted scores. Bow velocity was the most varying control (likely related to differences in tempo between the music excerpts) followed by bow-string distance (which may indicate *staccato* vs *legato* played phrases, for example) and bow-bridge distance.

In the second experiment, another MANOVA design tested the existence of individual bowing patterns among the players, independent of the performed music. Significant differences were revealed in the use of three controls combined, as well as for each parameter separately. The bow-string distance exhibited the strongest between-cellist variations followed by bow-bridge distance and bow velocity. These results suggest that, though each music score has its particular requirements for the playing technique of a performer to be accordingly adapted for, and each musician responds to these requirements differently, certain technical features of his execution remain remarkably constant across the interpreted scores, at least for a period of time.

Finally, the relationship between bowing controls and acoustic descriptors was examined by means of correlation analysis. It revealed that all three "mechanical" inputs have to be accounted for when predicting complex spectro-temporal characteristics of the sound produced. This leads to the conclusion that, since at least two, if not three, acoustic features are needed to describe sound qualities of a player's tone, a more complex model is required to capture the mapping between the two domains.

The overall discussion (Section 9.1) brought up another important finding. Three acoustical correlates of perceptual dimensions were able to explain 64% of the perceived dissimilarity between the cellists. Though one would certainly wish this value to be higher, the outcome suggests that the initial feature selection procedure resulted in a subset of spectral features reasonably well fitted for the task and relatively easy to interpret.

It is worth noticing that the results presented in Chapters 6–8 were obtained for a small group of six players recorded on the same cello. One might ask whether these results can be representative for a larger sample or, statistically speaking, for a population of cellists in general. For other randomly selected six players recorded on exactly the same cello, one may expect to obtain similar 3-D acoustical space, as the selection of spectro-temporal descriptors best characterising varying timbres of the players is largely determined by the

acoustical properties of the instrument itself. With the sample size increased, clusters of players in respective gestural, acoustical and perceptual spaces may occur, suggesting the existence of within-group similarities due to factors not yet accounted for. For example, physiological aspects such as body height and weight, pedagogical considerations such as schools of playing and teachers, years of musical practice or cultural background may play increasing role in explicating individual differences in tone quality between the players, as captured via bowing controls and acoustic features and finally perceived by the listeners.

9.3 Future work

While a number of research goals set at this study's commencement have been achieved, several directions for further developments in this research area have also emerged.

Finding mappings from gestural to acoustical and from acoustical to perceptual domains

In this study, a simple correlation measure was used to investigate the relationship between the three performer domains, and it was demonstrated that such relationships exist. The next step might involve modelling those relationships by means of predictive models, and a simple linear regression is one obvious choice to start with.

In the case of mapping between performance gesture and acoustical output, results indicated that all three bowing controls contribute to shaping of tone spectra. A multiple regression model can be used to predict a player's acoustic characteristics from the bowing inputs. However, since a single spectral feature is not sufficient to differentiate between the cellists' timbres, a multivariate multiple regression model seems to be a better choice. Depending on the model accuracy more advanced methods can be tested, such as Bayesian multivariate linear regression, support vector regression (SVR), or other machine learning techniques.

Pérez et al. (2012)'s study is an example of machine learning application. He employed neural networks (NN) to estimate spectral energy in forty frequency bands from a given set of performance controls. The obtained regression model was applied to improve a sample-based synthesiser with gesture control driven spectral transformations. A similar approach can be used to predict acoustic characteristics of a player based on bow-string distance, bow velocity and bow-bridge distance as the model inputs. The prediction accuracy of the above mentioned methods may also be improved by adding extra input parameters including auxiliary bowing controls such as bow tilt and bow acceleration and other parameters such as estimated pitch and finger position (*ibid.*).

Although the mapping between the cellists' coordinates in the acoustic feature and perceptual spaces seems to strongly and significantly resemble a linear relationship, and a simple linear or multivariate regression would likely produce a reliable model, there is still a room for further improvements of the model's accuracy. This can involve tailoring the initial feature dataset as well as changing the feature selection methodology. For the latter, for example, instead of using ANOVA-based evaluation of each acoustic descriptor's discriminative ability, so-called *wrapper* methods including sequential selection and heuristic search algorithms might be examined for their suitability to select an optimal feature subset.

Performer classification based on gesture controls and acoustic features

With the three acoustic descriptors identified as best discriminating between the six cellists' tones and with individual bowing strategies revealed, validating their discriminative power in classification experiments seems a natural step forward. The database described in Chapter 5 provides a wealth of audio and gesture material (only a small fraction was used in this thesis) for designing optimal training and test datasets. A number of classification methods drawn from supervised and unsupervised learning might be suitable for the task. The

emphasis would be on those procedures which provide more explicit interpretation of the classification output, i.e. in terms of detected bowing patterns or unique spectral characteristics.

Comparable study of the cellists' bowing gestures and their timbres captured on the second cello

The multi-modal database of cello recordings (Chapter 5) comprises samples of the same musical repertoire recorded on another cello. Since each cello is physically different, i.e. cellos vary in size and shape, building materials, woodworking technology, the applied glues and varnish, its acoustical properties are also unique (see Chapter 3). In other words, the physical properties of the instrument determine its tone quality. Another aspect of the instrument's quality is viewed in terms of playability, i.e. the ease of playing and the acoustical responsiveness to the player's musical intentions. In each case, the player needs to adapt his technique to bring out the best in the instrument. This poses the question to what extent this adaptation takes place, and whether any player-specific, though independent of the instrument being played, gesture controls can be observed.

The initial investigation was carried out in (Chudy et al., 2013). The study analysed samples of a D-major scale played in two articulation variants on both cellos. The averages of bow-bridge distance, bow velocity and estimated bow force, compared across the cellists, showed that, regardless of the individual bowing techniques being adapted for each instrument, there were also, at least for two cellists, cross-instrumental consistencies in their choice of bowing controls. This result encourages a full scale study to be undertaken, aiming at a comparison of individual bowing techniques across recorded samples of the collected music repertoire.

In parallel to examining bowing techniques of the cellists, Chudy et al. (2013) also analysed their timbral features. Although the results were inconclusive, they indicated the difference in brightness between the two cellos (based on

harmonic spectral centroid) and, consequently, a shift in higher frequency contents of each cellist's spectral characteristics. With the SpectVariation, Centroid and SubBand10Flux descriptors, identified as best capturing spectral variability of the players (Chapter 7), an additional study might reveal whether a player whose tone was determined as the least bright (in acoustical terms) on one cello would be also the least bright on another, and whether there is more general resemblance in the cellists' positioning between the two cello spaces.

Investigating differences in the execution of bow strokes

The cello database provides also a rich material for studying individual differences in executing articulation markings. This can be performed on captured bowing controls as well as on extracted acoustic descriptors. Pérez (2013) characterised the differences between bouncing (off-string) and on-string bow strokes in terms bowing parameters and audio features. Bow velocity and bow force were found to be major factors discriminating note attack types and controlling the note's sustain and release segments. His findings can be directly applied for comparing and evaluating bowing techniques of the six cellists, which can eventually lead to better characterisation of their timbral identities.

Analysis of the overall preference in relation to tone quality

In the perceptual experiment described in Chapter 6, in addition to timbre dissimilarity and verbal attribute ratings, participants were also asked to mark their overall preference for one (or no) tone sample in each evaluated pair of cellists.

Although the preference data has not yet been fully analysed, the preliminary result indicated that Cellist 4 was the most preferred performer, both in terms of preference magnitude and frequency, across compared music styles and genres. A closer examination of the cellist's acoustic features might give some clues about the origins of such preference, and whether they can be linked to a superior quality of tone. If this holds true, a further examination of the player's bowing

technique may have pedagogical implications for musical training and musical instrument instruction.

9.4 Potential applications

Pérez (2009) and Maestre (2009) proposed to enhance sample-based synthesis of the violin using estimated bowing contours and gesture driven spectral transformations. They retrieved bowing contours or trajectories from bowing control temporal curves via Bézier cubic curve segmenting and Gaussian mixture modelling. The obtained models were used to generate synthetic contours matching the indications of the synthesised music score.

One possible extension to their approach would be to create a database of gesture trajectories (bowing contours) of different performers and apply their estimated bowing grammars for sample-based sound synthesis tailored with performer-specific spectral shaping.

Interactive systems such as i-Maestro (Ng and Nesi, 2008) have already taken advantage of motion capture technologies combined with real-time audio analysis for musical training and instrument instruction purposes. The i-Maestro project's audio analysis component extracts audio descriptors such as pitch and loudness and timbre parameters such as noisiness and brilliance. They can be visualised along with captured bowing gesture data for detailed inspection of the performance. Based on the perceptually-informed 3-D acoustic characterisation of a player's timbre proposed in this thesis, the system's usability might benefit from visualising the player's timbral "trajectory" for better control over the tone quality.

9.5 A closing remark

Findings presented in this thesis shed light on an often overlooked aspect of performing on acoustic instruments. WHO is playing the instrument does make

a difference, not only in terms of much-studied expressive parameters such as timing and dynamics, but also in terms of the quality of sound. Whether an exceptional tone quality makes an exceptional performer is a question for another investigation.

Bibliography

Abdi, H. and Williams, L. (2010). Correspondence analysis. In Salkind, N. J., editor, *Encyclopedia of Research Design*. SAGE Publications, Inc.

Abeles, H. (1979). Verbal timbre descriptors of isolated clarinet tones. In *Bulletin of the Council for Research in Music Education*, volume 59, pages 1–7. University of Illinois Press.

Agostini, G., Longari, M., and Pollastri, E. (2001). Musical instrument timbres classification with spectral features. In *Proceedings of 2001 IEEE 4th Workshop on Multimedia Signal Processing*, Cannes, France.

Alexanian, D. (1922). *Theoretical and practical treatise of the Violoncello*. A. Z. Mathot, Paris.

Alluri, V. and Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception*, 27(3):223–242.

Alonso Moral, J. and Jansson, E. V. (1982). Input admittance, eigenmodes, and quality of violins. *Quarterly Progress and Status Report*, 23(2-3):60–75. Dept. of Speech, Music, and Hearing, Royal Institute of Technology (KTH), Stockholm.

ANSI (1960). *American Standard Acoustical Terminology, Definition 12.9, Timbre*. New York.

Askenfelt, A. (1982). Eigenmodes and tone quality of the double bass. *Quarterly Progress and Status Report*, 23(4):149–174. Dept. of Speech, Music, and Hearing, Royal Institute of Technology (KTH), Stockholm.

Askenfelt, A. (1986). Measurement of bow motion and bow force in violin playing. *Journal of the Acoustical Society of America*, 80(4):1007–1015.

Askenfelt, A. (1989). Measurement of the bowing parameters in violin playing. II: Bow-bridge distance, dynamic range, and limits of bow force. *Journal of the Acoustical Society of America*, 86(2):503–516.

Askenfelt, A. (1992). Observations on the dynamic properties of violin bows. *Quarterly Progress and Status Report*, 33(4):43–49. Dept. of Speech, Music, and Hearing, Royal Institute of Technology (KTH), Stockholm.

Barthet, M., Depalle, P., Kronland-Martinet, R., and Ystad, S. (2010a). Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Perception*, 28(2):135–154.

Barthet, M., Depalle, P., Kronland-Martinet, R., and Ystad, S. (2011). Analysis-by-synthesis of timbre, timing, and dynamics in expressive clarinet performance. *Music Perception*, 28(3):265–278.

Barthet, M., Guillemain, P., Kronland-Martinet, R., and Ystad, S. (2010b). From clarinet control to timbre perception. *Acta Acustica united with Acustica*, 96(4):678–689.

Beal, A. L. (1985). The skill of recognizing musical structures. *Memory and Cognition*, 13:405–412.

Benzécri, J. P. (1992). *Correspondence Analysis Handbook*. Marcel Dekker, New York.

Berger, K. W. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America*, 36(10):1888–1891.

Bevilacqua, F., Rasamimanana, N., Fléty, E., Lemouton, S., and Baschet, F. (2006). The augmented violin project: Research, composition and performance report. In *Proceedings of the 6th International Conference on New Interfaces for Musical Expression (NIME'06)*, Paris, France.

Binet, A. and Courtier, J. (1895). Recherches graphiques sur la musique. *L'Année Psychologique*, 2(1):201–222.

Bismarck, G. v. (1974a). Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30:159–172.

Bismarck, G. v. (1974b). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, 30:146–159.

Borg, I. and Groenen, P. J. F. (1997). *Modern Multidimensional Scaling: Theory and Applications*. Springer-Verlag, New York.

Borg, I., Groenen, P. J. F., and Mair, P. (2013). *Applied Multidimensional Scaling*. Springer-Verlag, New York.

Boutilion, X. (1991). Analytical investigation of the flattening effect: The reactive power balance rule. *Journal of the Acoustical Society of America*, 90(2):754–763.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, 26(2):211–252.

Bradley, J. S. (1976). Effects of bow force and speed on violin response. *Journal of the Acoustical Society of America*, 60(1):274–275.

Bynum, E. and Rossing, T. D. (1997). Holographic studies of cello vibrations. In *Proceedings of the International Symposium on Musical Acoustics (ISMA 1997)*, Edinburgh, UK.

Bynum, E. and Rossing, T. D. (2010). Cello. In Rossing, T. D., editor, *The Science of String Instruments*. Springer-Verlag, New York.

Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, 118(1):471–482.

Carroll, J. D. and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, 35(3):283–319.

Chafe, C. (1988). Celletto. `http://chrischafe.net/portfolio/celletto-pieces`, `https://ccrma.stanford.edu/~cc/shtml/cellettoMusic.shtml`.

Chudy, M. and Dixon, S. (2010). Towards music performer recognition using timbre features. In *Proceedings of the 3rd International Conference of Students of Systematic Musicology*, pages 45–50, Cambridge, UK.

Chudy, M. and Dixon, S. (2012). Recognising cello performers using timbre models. Technical report, C4DM, Queen Mary University of London, UK.

Chudy, M. and Dixon, S. (2013). Recognising cello performers using timbre models. In Lausen, B., Van den Poel, D., and Ultsch, A., editors, *Algorithms from and for Nature and Life: Classification and Data Analysis*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 511–518. Springer International Publishing.

Chudy, M., Pérez, A., and Dixon, S. (2013). On the relation between gesture, tone production and perception in classical cello performance. In *Proceedings of the 21st International Congress on Acoustics*, Montréal, Canada.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4):284–290.

Clark, M., Luce, D., Abrams, R., Schlossberg, H., and Rome, J. (1963). Preliminary experiments on the aural significance of tones of orchestral instruments and on choral tones. *Journal of the Audio Engineering Society*, 11(1):45–54.

Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 2nd edition.

Costello, A. B. and Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7).

Cramér, H. (1999). *Mathematical Methods of Statistics*. Princeton University Press.

Cremer, L. (1972). The influence of "bow pressure" on the movement of a bowed string. Part I. *Catgut Acoustical Society Newsletter*, 18:13–19.

Cremer, L. (1973). The influence of "bow pressure" on the movement of a bowed string. Part II. *Catgut Acoustical Society Newsletter*, 19:21–25.

Cremer, L. (1982). Consideration of the duration of transients in bowed instruments. *Catgut Acoustical Society Newsletter*, 38:13–18.

Darke, G. (2005). Assesment of timbre using verbal attributes. In *Proceedings of the 2nd Conference on Interdisciplinary Musicology, CIM'05*, Montréal, Canada.

Demoucron, M., Askenfelt, A., and Causse, R. (2006). Mesure de la pression d'archet des instruments à corde frottée. application à la synthèse sonore. In *Proceedings of the 8th French Congress on Acoustics*.

Dillon, R. (2004). *On the recognition of expressive intentions in music playing: a computational approach with experiments and applications*. PhD thesis, University of Genoa.

Dilworth, J. (1999). The bow: its history and development. In Stowell, R., editor, *The Cambridge Companion to the Cello*, pages 28–36. Cambridge University Press.

Disley, A. C., Howard, D. M., and Hunt, A. D. (2006). Timbral description of musical instruments. In Baroni, M., Addessi, A. R., Caterina, R., and Costa, M., editors, *Proceedings of the 9th International Conference on Music Perception and Cognition*, pages 61–68, Bologna, Italy.

Dixon, S., Goebl, W., and Widmer, G. (2002). The Performance Worm: Real time visualisation of expression based on Langner's tempo-loudness animation. In *Proceedings of the International Computer Music Conference, ICMC'02*, pages 361–364, Gothenburg, Sweden.

Doey, L. and Kurta, J. (2011). Correspondence analysis applied to psychological research. *Tutorials in Quantitative Methods for Psychology*, 7(1):5–14.

Donnadieu, S. (2007). Mental representation of the timbre of complex sounds. In Beauchamp, J. W., editor, *Analysis, Synthesis, and Perception of Musical Sounds. The Sound of Music*, pages 272–319. Springer-Verlag, New York.

Eerola, T., Alluri, V., and Ferrer, R. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, 30(1):49–70.

Eisenberg, M. (1957). *Cello Playing of Today.* The Strad, London.

Elliott, C. A. (1975). Attacks and releases as factors in instrument identification. *Journal of Research in Music Education*, 23(1):35–40.

Erickson, R. (1975). *Sound Structure in Music.* University of California Press.

Eronen, A. (2001). Comparison of features for musical instrument recognition. In *Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 753–756, Istanbul, Turkey.

Eronen, A. and Klapuri, A. (2000). Musical instrument recognition using cepstral coefficients and temporal features. In *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 753–756, New Paltz, New York, USA.

Firth, I. M. (1974). The wolf tone in the cello: Acoustic and holographic studies. *Quarterly Progress and Status Report*, 15(4):42–56. Dept. of Speech, Music, and Hearing, Royal Institute of Technology (KTH), Stockholm.

Fitzgerald, R. A. (2003). *Performer–dependent dimensions of timbre: identifying acoustic cues for oboe tone discrimination.* PhD thesis, School of Music, University of Leeds.

Flesch, C. (1939). *The Art of Violin Playing. Book One.* Carl Fischer, New York, 2nd rev. edition. English translation by Martens, F. M.

Fletcher, N. H. and Rossing, T. D. (1998). *The Physics of Musical Instruments.* Springer-Verlag, New York.

Galluzzo, P. M. (2003). *On the playability of stringed instruments.* PhD thesis, Trinity College, University of Cambridge.

Giordano, B. L. and McAdams, S. (2010). Sound source mechanics and musical timbre perception: Evidence from previous studies. *Music Perception*, 28(2):155–168.

Gordon, J. W. and Grey, J. M. (1978). Perception of spectral modifications on orchestral instrument tones. *Computer Music Journal*, 2(1):24–31.

Goudeseune, C. (2001). *Composing with parameters for synthetic instruments.* PhD thesis, University of Illinois at Urbana-Champaign.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis.* Academic Press, London.

Grey, J. M. (1975). *An exploration of musical timbre.* PhD thesis, Stanford University.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5):1270–1277.

Grey, J. M. and Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, 63(5):1493–1500.

Grey, J. M. and Moorer, J. A. (1977). Perceptual evaluations of synthesized musical instrument tones. *Journal of the Acoustical Society of America*, 62(2):454–462.

Guaus, E., Blaauw, M., Bonada, J., Maestre, E., and Pérez, A. (2009). A calibration method for accurately measuring bow force in real violin performance. In *Proceedings of the 35th International Computer Music Conference*, Montréal, Canada.

Guaus, E., Bonada, J., , Pérez, A., Maestre, E., and Blaauw, M. (2007). Measuring the bow pressing force in a real violin performance. In *Proceedings of the International Symposium on Musical Acoustics (ISMA 2007)*, Barcelona, Spain.

Guettler, K. (1992). The bowed string computer simulated – some characteristic features of the attack. *Catgut Acoustical Society Journal*, 2(2):22–26.

Guettler, K. (2002). On the creation of the Helmholtz motion in bowed strings. *Acta Acustica united with Acustica*, 88(6):970–985.

Guettler, K. (2004). Looking at starting transients and tone coloring of the bowed string. *Journal of ITC Sangeet Research Academy*, 18:60–76.

Guettler, K. (2010). Bows, strings, and bowing. In Rossing, T. D., editor, *The Science of String Instruments*. Springer-Verlag, New York.

Guettler, K. and Askenfelt, A. (1995). What is a proper start of a bowed string? *Quarterly Progress and Status Report*, 36(2-3):97–106. Dept. of Speech, Music, and Hearing, Royal Institute of Technology (KTH), Stockholm.

Guettler, K. and Askenfelt, A. (1997). Acceptance limits for the duration of pre-Helmholtz transients. *Journal of the Acoustical Society of America*, 101(5):2903–2913.

Guettler, K., Schoonderwaldt, E., and Askenfelt, A. (2003). Bow speed or position–Which one influences spectrum the most? In *Proceedings of the*

Stockholm Music Acoustics Conference (SMAC'03), pages 67–70, Stockholm, Sweden.

Guillaume, P., editor (2006). *Music and Acoustics: From Instrument to Computer*. ISTE Ltd, London, UK.

Hajda, J. M., Kendall, R. A., Carterette, E. C., and Harshberger, M. L. (1997). Methodological issues in timbre research. In Deliège, I. and Sloboda, J., editors, *The Perception and Cognition of Music*, pages 253–306. Psychology Press, Hove, East Sussex, UK.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1):23–34.

Handel, S. and Erickson, M. (2001). A rule of thumb: The bandwidth for timbre invariance is one octave. *Music Perception*, 19(1):121–126.

Handel, S. and Erickson, M. (2004). Sound source identification: The possible role of timbre transformations. *Music Perception*, 21(4):587–610.

Hanson, R. J., Schneider, A. J., and Halgedahl, F. W. (1994). Anomalous low-pitched tones from a bowed violin string. *Catgut Acoustical Society Journal*, 2(6):1–7.

Hayes, A. F. and Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1):77–89.

Helmholtz, H. (1877). *On the Sensations of Tone: As a Physiological Basis for the Theory of Music*. Longmans, Green & Co., London, 4th English edition. (1912). Translated by Ellis, A. J.

Herrera-Boyer, P., Peeters, G., and Dubnov, S. (2003). Automatic classification of musical instrument sounds. *Journal of New Music Research*, 32(1):3–21.

Holmes, P. A. (2011). An exploration of musical communication through expressive use of timbre: The performer's perspective. *Psychology of Music*, 40(3):301–323.

Howard, D. M. and Angus, J. A. S. (2009). *Acoustics and Psychoacoustics*. Elsevier, Oxford, UK, 4th edition.

ISO/IEC-15938-4 (2002). Information technology − multimedia content description interface − part 4: Audio.

Iverson, P. and Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, 94(5):2595–2603.

Jansson, E. (2002). *Acoustics for violin and guitars makers*. Dept. of Speech, Music, and Hearing, Royal Institute of Technology (KTH), Stockholm, Sweden. Published online.

Jensen, K. (1999). *Timbre models of musical sounds*. PhD thesis, University of Copenhagen.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39:31–36.

Kendall, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception*, 4(2):185–213.

Kendall, R. A. and Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, 8(4):369–404.

Kendall, R. A. and Carterette, E. C. (1993a). Verbal attributes of simultaneous wind instrument timbres: I. von Bismarck's adjectives. *Music Perception*, 10(4):445–467.

Kendall, R. A. and Carterette, E. C. (1993b). Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's "orchestration". *Music Perception*, 10(4):469–501.

Kendall, R. A., Carterette, E. C., and Hajda, J. M. (1995). Perceptual and acoustical attributes of natural and emulated orchestral instrument timbres. In *Proceedings of the International Symposium on Musical Acoustics*, pages 596–601.

Kendall, R. A., Carterette, E. C., and Hajda, J. M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, 16(3):327–363.

Kim, H.-G., Moreau, N., and Sikora, T. (2005). *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Wiley, Chichester, UK.

Kostek, B. (1995). Feature extraction methods for the intelligent processing of musical signals. In *Proceedings of the 99th AES Convention*, New York, USA.

Kostek, B. and Wieczorkowska, A. (1996). Study of parameter relations in musical instrument patterns. In *Proceedings of the 100th AES Convention*, Copenhagen, Denmark.

Krall, E. (1913). *The Art of Tone-Production on the Violoncello*. The Strad. John Leng & Co., London.

Krimphoff, J., McAdams, S., and Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. *Journal de Physique*, 4 (C5):625–628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In Nielzen, S. and Olsson, O., editors, *Structure and Perception of Electroacoustic Sound and Music*, pages 43–53, Elsevier, Amsterdam.

Krumhansl, C. L. and Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3):739–751.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2):115–129.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional scaling*, volume 11 of *Quantitative Applications in the Social Sciences*. SAGE Publications, Inc.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics*, 62(7):1426–1439.

Langhoff, A. (1995). Modal analysis of violin, viola and cello compared to the acoustical spectrum. In *Proceedings of the International Symposium on Musical Acoustics (ISMA 1995)*, pages 286–290, Dourdan, France.

Lartillot, O., Toiviainen, P., and Eerola, T. (2008). A Matlab toolbox for Music Information Retrieval. In Preisach, C., Burkhardt, H., Schmidt-Thieme, L., and Decker, R., editors, *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*, pages 261–268. Springer-Verlag.

Machover, T. (1992). Hyperinstruments: A progress report 1987-1991. Technical report, MIT Media Laboratory.

Maestre, E. (2009). *Modeling Instrumental Gestures: An Analysis/Synthesis Framework for Violin Bowing*. PhD thesis, Music Technology Group, Universitat Pompeu Fabra, Barcelona.

Maestre, E., Blaauw, M., Bonada, J., Guaus, E., and Pérez, A. (2010). Statistical modeling of bowing control applied to violin sound synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(4):855–871.

Maestre, E., Bonada, J., Blaauw, M., Pérez, A., and Guaus, E. (2007). Acquisition of violin instrumental gestures using a commercial EMF tracking device. In *Proceedings of the 33rd International Computer Music Conference (ICMC'07)*, Copenhagen, Denmark.

Mantel, G. (1995). *Cello Technique: Principles and Forms of Movement*. Indiana University Press.

Marchini, M., Papiotis, P., Pérez, A., and Maestre, E. (2011). A hair ribbon deflection model for low-intrusiveness measurement of bow force in violin performance. In *Proceedings of the 11th International Conference on New Interfaces for Musical Expression (NIME'11)*, pages 481–486, Oslo, Norway.

Marchini, M., Ramirez, R., Papiotis, P., and Maestre, E. (2013). Inducing rules of ensemble music performance: A machine learning approach. In Luck, G. and Brabant, O., editors, *Proceedings of the 3rd International Conference on Music & Emotion*, Jyväskylä, Finland.

Marchini, M., Ramirez, R., Papiotis, P., and Maestre, E. (2014). The sense of ensemble: a machine learning approach to expressive performance modelling in string quartets. *Journal of New Music Research*, 43(3):303–317.

Marozeau, J. and de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre. *Journal of the Acoustical Society of America*, 121(1):383–387.

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America*, 114(5):2946–2957.

Martin, K. D. (1999). *Sound-Source Recognition: A Theory and Computational Model*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Martin, K. D. and Kim, Y. E. (1998). Musical instrument identification: A pattern recognition approach. In *Proceedings of the 136th meeting of the Acoustical Society of America*, Norfolk, VA, Canada.

Mazzocchi, M. (2008). *Statistics for Marketing and Consumer Research*. SAGE Publications Ltd.

McAdams, S. (1993). Recognition of sound sources and events. In McAdams, S. and Bigand, E., editors, *Thinking in Sound: The Cognitive Psychology of Human Audition*, pages 148–198. Oxford University Press.

McAdams, S. (2013). Musical timbre perception. In Deutsch, D., editor, *The Psychology of Music*, pages 35–67. Academic Press, 3rd edition.

McAdams, S., Giordano, B. L., Susini, P., Peeters, G., and Rioux, V. (2006). A meta-analysis of acoustic correlates of timbre dimensions (A). *Journal of the Acoustical Society of America*, 120(5):3275.

McAdams, S., Winsberg, S., Donnadieu, S., Desoete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3):177–192.

McGraw, K. O. and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1):30–46.

McIntyre, M. E., Schumacher, R. T., and Woodhouse, J. (1977). New results on the bowed string. *The Catgut Acoustical Society Newsletter*, 28:27–31.

McIntyre, M. E., Schumacher, R. T., and Woodhouse, J. (1983). Oscillations of musical instruments. *Journal of the Acoustical Society of America*, 74(5):1325–1345.

McIntyre, M. E. and Woodhouse, J. (1978). The acoustics of stringed musical instruments. *Interdisciplinary Science Reviews*, 3(2):157–173.

McIntyre, M. E. and Woodhouse, J. (1979). On the fundamentals of bowed-string dynamics. *Acustica*, 43(2):93–108.

Melka, A. (1994). Methodological approaches to the investigation of musical timbre. *Journal de Physique IV*, pages 569–576.

Meyer, J. (2009). *Acoustics and the Performance of Music*. Springer-Verlag, New York, USA, 5th edition.

Miller, D. C. (1909). The influence of the material of wind-instruments on the tone quality. *Science*, 29(735):161–171.

Miller, J. R. and Carterette, E. C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America*, 58(3):711–720.

Molina-Solana, M., Arcos, J. L., and Gomez, E. (2008). Using expressive trends for identifying violin performers. In *Proceedings of the 9th International Conference on Music Information Retrieval, ISMIR'08*, pages 495–500, Philadelphia, PA, USA.

Murtagh, F. (2005). *Correspondence Analysis and Data Coding with Java and R*. Chapman & Hall/CRC.

Ng, K. (2008). Technology-enhanced learning for music with i-Maestro framework and tools. In *Proceedings of the EVA London 2008 Conference*, pages 177–187.

Ng, K., Larkin, O., Koerselman, T., and Ong, B. (2007a). 3D motion data analysis and visualisation for string practice training. In *Proceedings of the EVA London 2007 Conference*, pages 20.1–20.8.

Ng, K., Larkin, O., Koerselman, T., Ong, B., Schwarz, D., and Bevilacqua, F. (2007b). The 3D Augmented Mirror: Motion analysis for string practice training. In *Proceedings of the 33rd International Computer Music Conference*, volume II, pages 53–56, Copenhagen, Denmark.

Ng, K. and Nesi, P. (2008). i-Maestro framework and interactive multimedia tools for technology-enhanced learning and teaching for music. In *Proceedings of the International Conference on Automated solutions for Cross Media Content and Multi-channel Distribution, AXMEDIS '08*, pages 266–269, Florence, Italy.

Nichols, C. (2002). The vBow: a virtual violin bow controller for mapping gesture to synthesis with haptic feedback. *Organised Sound*, 7(2):215–220.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. University of Illinois Press.

Overholt, D. (2005). The overtone violin. In *Proceedings of the 5th International Conference on New Interfaces for Musical Expression (NIME'05)*, pages 34–37, Vancouver, Canada.

Pardue, L., Harte, C., and McPherson, A. (2015). A low-cost real-time tracking system for violin. *Journal of New Music Research*, 44(4):305–323.

Pardue, L. and McPherson, A. (2013). Near-field optical reflectance sensing for violin bow tracking. In *Proceedings of the 13th International Conference on New Interfaces for Musical Expression (NIME'13)*, pages 363–368, Seoul, South Korea.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., and McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, 130(5):2902–2916.

Peiper, C., Warden, D., and Garnett, G. (2003). An interface for real-time classification of articulations produced by violin bowing. In *Proceedings of the 3rd Conference on New Interfaces for Musical Expression (NIME'03)*, Montréal, Canada.

Pérez, A. (2009). *Enhancing Spectral Synthesis Techniques with Performance Gestures using the Violin as a Case Study*. PhD thesis, Music Technology Group, Universitat Pompeu Fabra, Barcelona.

Pérez, A. (2013). Characterization of bowing strokes in violin playing in terms of controls and sound: Differences between bouncing and on-string bow strokes. In *Proceedings of the 21st International Congress on Acoustics*, Montréal, Canada.

Pérez, A., Bonada, J., Maestre, E., Guaus, E., and Blaauw, M. (2007). Combining performance action with spectral models for violin sound transformation. In *Proceedings of the 19th International Congress on Acoustics*, Madrid, Spain.

Pérez, A., Bonada, J., Maestre, E., Guaus, E., and Blaauw, M. (2008). Score level timbre transformations of violin sounds. In *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland.

Pérez, A., Bonada, J., Maestre, E., Guaus, E., and Blaauw, M. (2012). Performance control driven violin timbre model based on neural networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(3):1007–1021.

Pérez, A. and Wanderley, M. M. (2015). Indirect acquisition of violin instrumental controls from audio signal with hidden Markov models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(5):932–940.

Piston, W. (1969). *Orchestration.* Victor Gollancz Ltd, London, 5th edition.

Pitt, M. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5):976–986.

Pitteroff, R. and Woodhouse, J. (1998). Mechanics of the contact area between a violin bow and a string. Part III: Parameter dependence. *Acta Acustica united with Acustica*, 84(5):929–946.

Pleeth, W. (1982). *Cello.* Macdonald & Co., London.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In Plomp, R. and Smoorenburg, G. F., editors, *Frequency Analysis and Periodicity Detection in Hearing*, pages 397–414, Leiden. Sijhoff.

Plomp, R. and Steeneken, H. J. M. (1971). Pitch versus timbre. In *Proceedings of the 7th International Congress on Acoustics*, volume 3, 20-H-7, pages 377–380, Budapest.

Potter, L. J. (1996). *The Art of Cello Playing: A Complete Textbook Method for Private or Class Instruction.* Alfred Music, 2nd edition.

Q-Software (2015). Q Research Software: Correspondence Analysis. `http://wiki.q-researchsoftware.com/wiki/Correspondence_Analysis`.

Raman, C. V. (1918). On the mechanical theory of the vibrations of bowed strings and of musical instruments of the violin family, with experimental verification of the results. *Bulletin of the Indian Association for the Cultivation of Science*, 15:1–158.

Raman, C. V. (1920). Experiments with mechanically-played violins. *Bulletin of the Indian Association for the Cultivation of Science*, 6:19–36.

Ramirez, R., Maestre, E., and Pertusa, A. (2007). Identifying saxophonists from their playing styles. In *Proceedings of the 30th AES International Conference*, Saariselkä, Finland.

Ramirez, R., Pérez, A., and Kersten, S. (2008). Performer identification in celtic violin recordings. In *Proceedings of the 9th International Conference on Music Information Retrieval, ISMIR'08*, pages 483–488, Philadelphia, PA, USA.

Rasamimanana, N. (2004). Gesture analysis of bow strokes using an augmented violin. Master's thesis, Université Pierre et Marie Curie, Paris VI, France.

Rasamimanana, N., Fléty, E., and Bevilacqua, F. (2005). Gesture analysis of violin bow strokes. In Gibet, S., Courty, N., and Kamp, J.-F., editors, *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop (2005), Revised Selected Papers*, pages 145–155. Springer-Verlag, Berlin.

Richardson, B. (1999). Cello acoustics. In Stowell, R., editor, *The Cambridge Companion to the Cello*, pages 37–51. Cambridge University Press.

Risset, J. C. (1978). Musical acoustics. In Carterette, E. C. and Friedman,

M. P., editors, *Handbook of Perception*, volume 4, Hearing, pages 521–566, New York. Academic Press Inc.

Rossing, T. D., editor (2010). *The Science of String Instruments*. Springer-Verlag, New York.

Rossing, T. D., Roberts, M., Bynum, E., and Nickerson, L. (1998). Modal analysis of violins and cellos. In *Proceedings of the 16th International Congress on Acoustics*, Seattle, USA.

Saldanha, E. L. and Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, 36(11):2021–2126.

Sandell, G. J. and Martens, W. L. (1995). Perceptual evaluation of principal-component-based synthesis of musical timbres. *Journal of the Audio Engineering Society*, 43(12):1013–1028.

Saunders, C., Hardoon, D. R., Shawe-Taylor, J., and Widmer, G. (2004). Using string kernels to identify famous performers from their playing style. In *Proceedings of the 15th European Conference on Machine Learning Machine, ECML'04*, pages 384–395, Pisa, Italy.

Schaeffer, P. (1966). *Traité des Objets Musicaux*. Éditions du Seuil.

Schelleng, J. C. (1973). The bowed string and the player. *Journal of the Acoustical Society of America*, 53(1):26–41.

Schoonderwaldt, E. (2009a). The player and the bowed string: Coordination of bowing parameters in violin and viola performance. *Journal of the Acoustical Society of America*, 126(5):2709–2720.

Schoonderwaldt, E. (2009b). The violinist's sound palette: Spectral centroid, pitch flattening and anomalous low frequencies. *Acta Acustica united with Acustica*, 95(5):901–914.

Schoonderwaldt, E. and Demoucron, M. (2009). Extraction of bowing parameters from violin performance combining motion capture and sensors. *Journal of the Acoustical Society of America*, 126(5):2695–2708.

Schoonderwaldt, E., Guettler, K., and Askenfelt, A. (2003). Effect of the width of the bow hair on the violin string spectrum. In *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03)*, pages 91–94, Stockholm, Sweden.

Schoonderwaldt, E., Guettler, K., and Askenfelt, A. (2007). Schelleng in retrospect – a systematic study of bow force limits for bowed violin strings. In *Proceedings of the International Symposium on Musical Acoustics (ISMA 2007)*, Barcelona, Spain.

Schoonderwaldt, E., Guettler, K., and Askenfelt, A. (2008). An empirical investigation of bow-force limits in the Schelleng diagram. *Acta Acustica united with Acustica*, 94(4):604–622.

Schouten, J. F. (1968). The perception of timbre. In Kohasi, Y., editor, *Reports of the 6th International Congress on Acoustics*, volume 1, GP-6-2, pages 35–44, Tokyo, Japan.

Schumacher, R. T. (1975). Some aspects of the bow. *Catgut Acoustical Society Newsletter*, 24:5–8.

Schumacher, R. T. (1979). Self-sustained oscillations of the bowed string. *Acta Acustica united with Acustica*, 43(2):109–120.

Schumacher, R. T. (1994). Measurements of some parameters of bowing. *Journal of the Acoustical Society of America*, 96(4):1985–1998.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I. *Psychometrika*, 27(2):125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part II. *Psychometrika*, 27(3):219–246.

Solomon, L. N. (1958). Semantic approach to the perception of complex sounds. *Journal of the Acoustical Society of America*, 30(5):421–425.

Stamatatos, E. and Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artificial Intelligence*, 165(1):37–56.

Steele, K. M. and Williams, A. K. (2006). Is the bandwidth for timbre invariance only one octave? *Music Perception*, 23(3):215–220.

Štěpánek, J. (2002). Evaluation of timbre of violin tones according to selected verbal attibutes. In *Proceedings of the 32nd International Acoustical Conference, European Acoustics Association (EAA) Symposium*, Banská Štiavnica, Slovakia.

Štěpánek, J. (2004). Spectral sources of basic perceptual dimensions of violin timbre. In *Proceedings of the 7th French Congress on Acoustics and 30th DAGA Conference*, pages 103–104, Strasbourg, France.

Štěpánek, J. and Moravec, O. (2005). Verbal description of musical sound timbre in Czech language and its relation to musicians profession and performance quality. In *Proceedings of the 2nd Conference on Interdisciplinary Musicology, CIM'05*, Montréal, Canada.

Štěpánek, J. and Otcěnášek, Z. (1999). Rustle as an attribute of timbre of stationary violin tones. *Journal of the CATGUT Acoustical Society*, 3(8):32–38.

Štěpánek, J. and Otcěnášek, Z. (2002). Spectral sources of selected features of violin timbre. In *Proceedings of the 6th French Congress on Acoustics*, pages 695–698, Lille, France.

Štěpánek, J. and Otcěnášek, Z. (2004). Interpretation of violin spectrum using psychoacoustic experiments. In *Proceedings of the International Symposium on Musical Acoustics (ISMA2004)*, Nara, Japan.

Štěpánek, J. and Otcěnášek, Z. (2005). Acoustical correlates of the main features of violin timbre perception. In *Proceedings of the 2nd Conference on Interdisciplinary Musicology, CIM'05*, Montréal, Canada.

Štěpánek, J., Otcěnášek, Z., and Melka, A. (1999). Comparison of five perceptual timbre spaces of violin tones of different pitches. In *Proceedings of Joint Meeting of the 137th ASA Meeting, 2nd Convention of the EAA: Forum Acusticum and 25th DAGA Conference*, Berlin, Germany.

Štěpánek, J., Otcěnášek, Z., and Moravec, O. (2000). Analytical and perceptual detection of rustle in stationary violin tones. In *Proceedings of the 5th French Congress on Acoustics*, pages 433–436, Laussanne, Switzerland.

Straeten, E. v. d. (1905). *The Technics of Violoncello Playing*. The Strad. John Leng & Co., London, 2nd edition.

Suchecki, R. (1982). *Wiolonczela od A do Z (The Cello from A to Z)*. Polskie Wydawnictwo Muzyczne, Kraków, Poland. In Polish.

Tabachnick, B. G. and Fidell, L. S. (2007). *Using Multivariate Statistics*. Pearson, Boston, 5th edition.

Tobudic, A. and Widmer, G. (2005). Learning to play like the great pianists. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05*, Edinburgh, Scotland.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 17(4):401–419.

Traube, C. (2004). *An Interdisciplinary Study of the Timbre of the Classical Guitar*. PhD thesis, McGill University, Montréal, Canada.

Trueman, D. and Cook, P. R. (1999). Bossa: The deconstructed violin reconstructed. In *Proceedings of the International Computer Music Conference*, Beijing, China.

Wedin, L. and Goude, G. (1972). Dimension analysis of instrumental timbre. *Scandinavian Journal of Psychology*, 13(1):228–240.

Wessel, D. L. (1973). Psychoacoustics and music: A report from Michigan State University. *PACE: Bulletin of the Computer Arts Society*, pages 1–2.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3(2):45–52.

Widmer, G., Dixon, S., Goebl, W., Pampalk, E., and Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine*, 24(3):111–130.

Widmer, G. and Zanon, P. (2004). Automatic recognition of famous artists by machine. In *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'04*, pages 1109–1110, Valencia, Spain.

Wieczorkowska, A. (1999). Rough sets as a tool for audio signal classification. In Ras, Z. W. and Skowron, A., editors, *Proceedings of the 11th International Symposium on Foundations of Intelligent Systems (ISMIS'99)*, pages 367–375. Springer-Verlag, Berlin.

Winsberg, S. and Carroll, J. D. (1989). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika*, 54(2):217–229.

Winsberg, S. and De Soete, G. (1993). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, 58(2):315–330.

Winsberg, S. and De Soete, G. (1997). Multidimensional scaling with constrained dimensions: CONSCAL. *British Journal of Mathematical and Statistical Psychology*, 50(1):55–72.

Woodhouse, J. (1993). On the playability of violins. Part II: Minimum bow force and transients. *Acustica*, pages 137–153.

Woodhouse, J. (1997). Stringed instruments: Bowed. In Crocker, M. J., editor, *Encyclopedia of Acoustics*, volume 4, pages 1619–1626. John Wiley & Sons, Inc.

Yelland, P. M. (2010). An introduction to correspondence analysis. *The Mathematica® Journal*, 12.

Young, D. (2001). New frontiers of expression through real-time dynamics measurement of violin bows. Master's thesis, Massachusetts Institute of Technology.

Young, D. (2002). The Hyperbow controller: Real-time dynamics measurement of violin performance. In *Proceedings of the 2nd Conference on New Instruments for Musical Expression (NIME'02)*, Dublin, Ireland.

Young, D. (2003). Wireless sensor system for measurement of violin bowing parameters. In *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03)*, pages 111–114, Stockholm, Sweden.

Young, D. (2007). *A Methodology for Investigation of Bowed String Performance Through Measurement of Violin Bowing Technique*. PhD thesis, Massachusetts Institute of Technology.

Young, D., Nunn, P., and Vassiliev, A. (2006). Composing for Hyperbow: A collaboration between MIT and the Royal Academy of Music. In *Proceedings of the 6th Conference on New Interfaces for Musical Expression (NIME'06)*, Paris, France.

Young, R. W. (1960). Musical acoustics. In *McGraw-Hill Encyclopedia of Science and Technology*, page 661. McGraw-Hill.

Zacharakis, A., Pastiadis, K., and Reiss, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception*, 31(4):339–358.

Zacharakis, A., Pastiadis, K., Reiss, J. D., and Papadelis, G. (2012). Analysis of musical timbre semantics through metric and non-metric data reduction techniques. In Cambouropoulos, E., Tsougras, C., Mavromatis, P., and Pastiadis, K., editors, *Proceedings of the ICMPC–ESCOM 2012 Joint Conference: 12th International Conference on Music Perception and Cognition and 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, pages 1177–1182, Thessaloniki, Greece.

Zanon, P. and Widmer, G. (2003). Recognition of famous pianists using machine learning algorithms: First experimental results. In *Proceedings of the 14th Colloquium on Musical Informatics, CIM'03*, Florence, Italy.

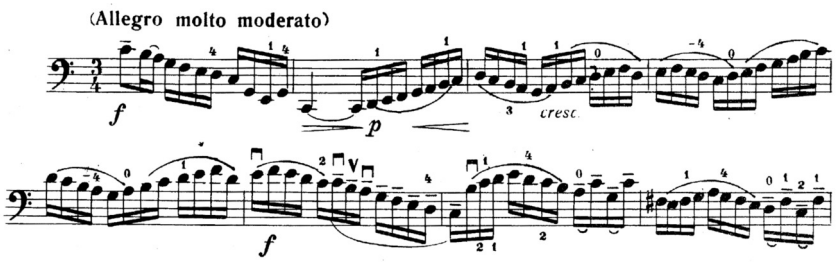
Appendix A

Music Scores

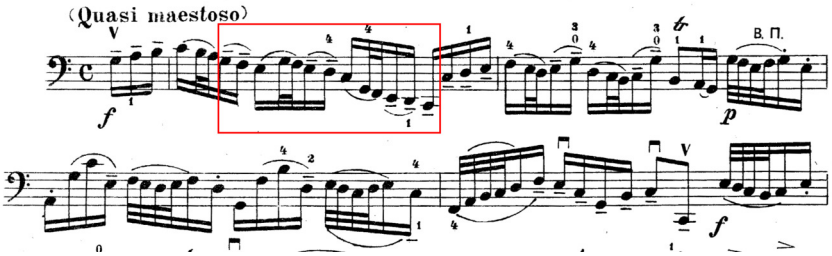
Red rectangles indicate the extracted music samples used for perceptual evaluation in Chapter 6 and then in acoustical and bowing gesture analyses in Chapters 7 and 8.

A.1 J.S. Bach – *3rd Cello Suite*

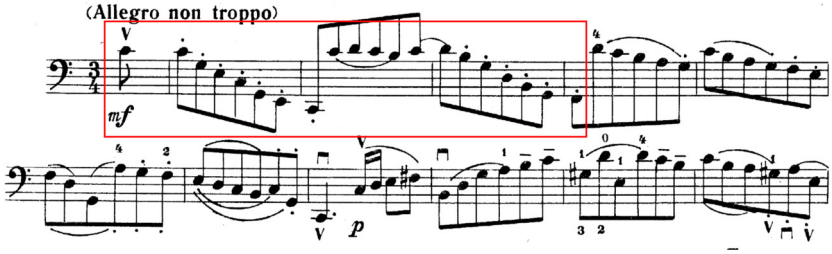
Prélude, bars 1–6



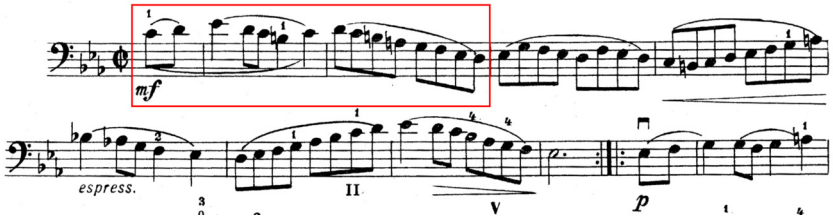
Allemande, bars 1–4



Courante, bars 1–8

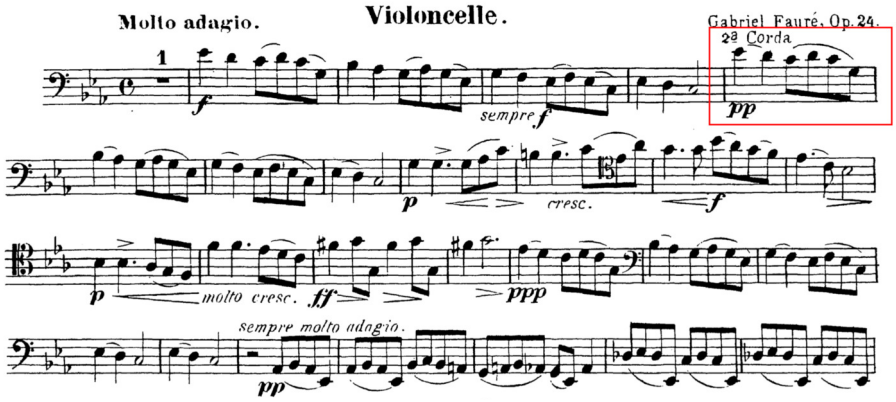


Bourrée II, bars 1–8



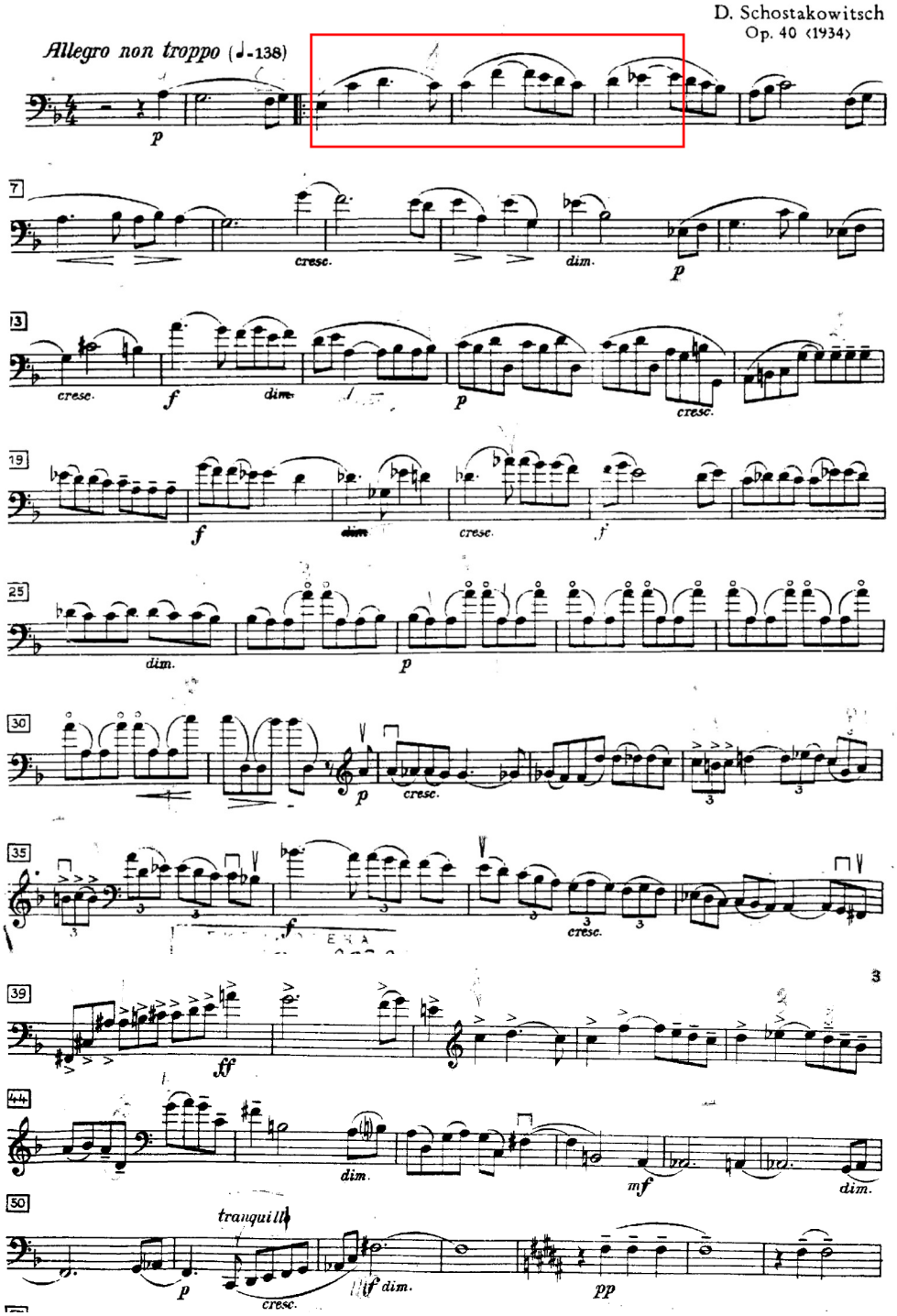
A.2 G. Fauré – *Élégie*

Élégie, bars 2–22

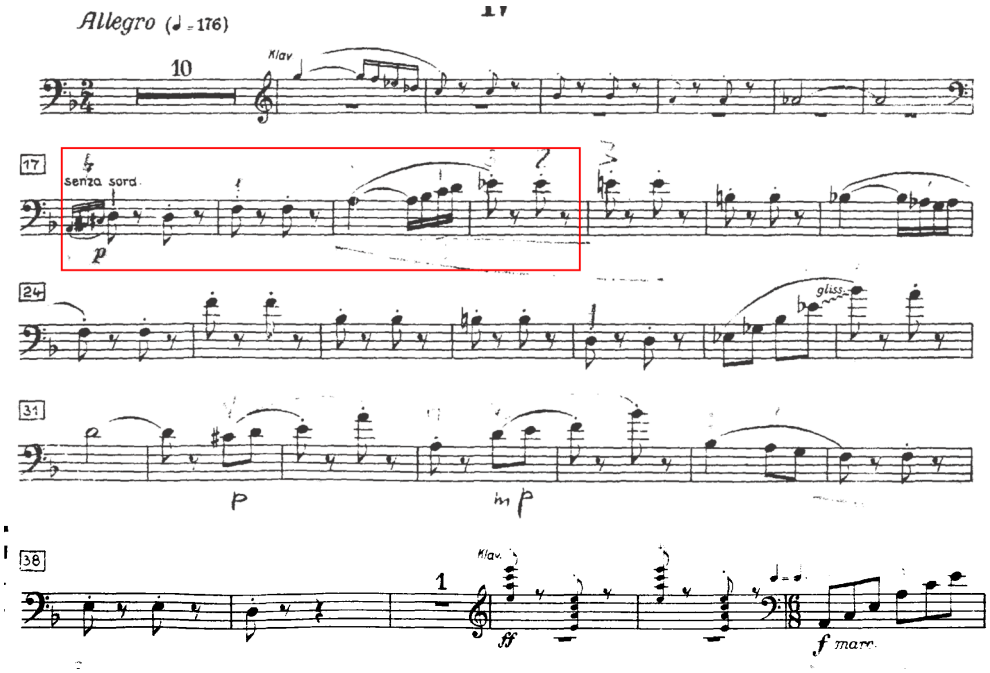


A.3 D. Shostakovich – *Cello Sonata* op. 40

I movement, bars 1–53



IV movement, bars 17–39



Appendix B

Acoustic Features

Table B.1: Frequency ranges of ten octave-scaled subbands. (From Alluri and Toiviainen, 2010)

SubBand No.	**Frequency Range**
SubBand No. 1	0–50 Hz
SubBand No. 2	50–100 Hz
SubBand No. 3	100–200 Hz
SubBand No. 4	200–400 Hz
SubBand No. 5	400–800 Hz
SubBand No. 6	800–1600 Hz
SubBand No. 7	1600–3200 Hz
SubBand No. 8	3200–6400 Hz
SubBand No. 9	6400–12800 Hz
SubBand No. 10	12800–22050 Hz

Table B.2: Acoustic features and their definitions. Represented signal domains: T – temporal, S – spectral, ST – spectro-temporal. (Adapted from Eerola et al., 2012)

D	Feature	Definition	Interpretation
T	Zero-Crossing Rate (ZeroCrossings)	Number of time-domain zero crossings*	A simple indicator of noisiness
S	High Frequency Energy (HighFreqEnergy)	Percent of the spectral energy above 1500 Hz frequency*	High frequency energy content
S	Spectral Roll-off 95 (Rolloff95)	The frequency below which 95% of the total spectral energy is contained*	High frequency energy content
S	Spectral Roll-off 85 (Rolloff85)	The frequency below which 85% of the total spectral energy is contained*	High frequency energy content
S	Spectral Entropy (SpectEntropy)	Measure of disorder of the spectrum*	Discriminates noise from harmonic content
S	Spectral Centroid (Centroid)	Geometric center of the amplitude spectrum*	Spectral distribution descriptor
S	Spectral Spread (Spread)	Standard Deviation of the spectrum*	Spectral distribution descriptor
S	Spectral Skewness (Skewness)	Skewness of the spectrum*	Spectral distribution descriptor
S	Spectral Kurtosis (Kurtosis)	Kurtosis of the spectrum*	Spectral distribution descriptor
S	Spectral Flatness (Flatness)	Ratio between the geometric and the arithmetic mean of the spectrum*	Discriminates noise from harmonic content
S	Spectral Irregularity (Irregularity)	Measure of variation of the successive peaks of the spectrum* (Jensen, 1999)	
S	Spectral Deviation (SpectDeviation)	Measure of variation of the successive peaks of the spectrum* (Krimphoff et al., 1994)	
ST	Roughness	Estimation of the sensory dissonance*	
ST	Spectral Variation (SpectVariation)	Correlation based measure of change between the consecutive spectral frames**(Peeters et al., 2011)	Represents the amount of variation of the spectrum over time
ST	Spectral Flux (SpectralFlux)	Euclidean distance based measure of change between the consecutive spectral frames*	Represents the amount of variation of the spectrum over time
ST	SubBand No.1-10*** Flux (SubBand1Flux,..., SubBand10Flux)	Fluctuation of frequency content in ten octave-scaled sub-bands of the spectrum* (Alluri and Toiviainen, 2010)	

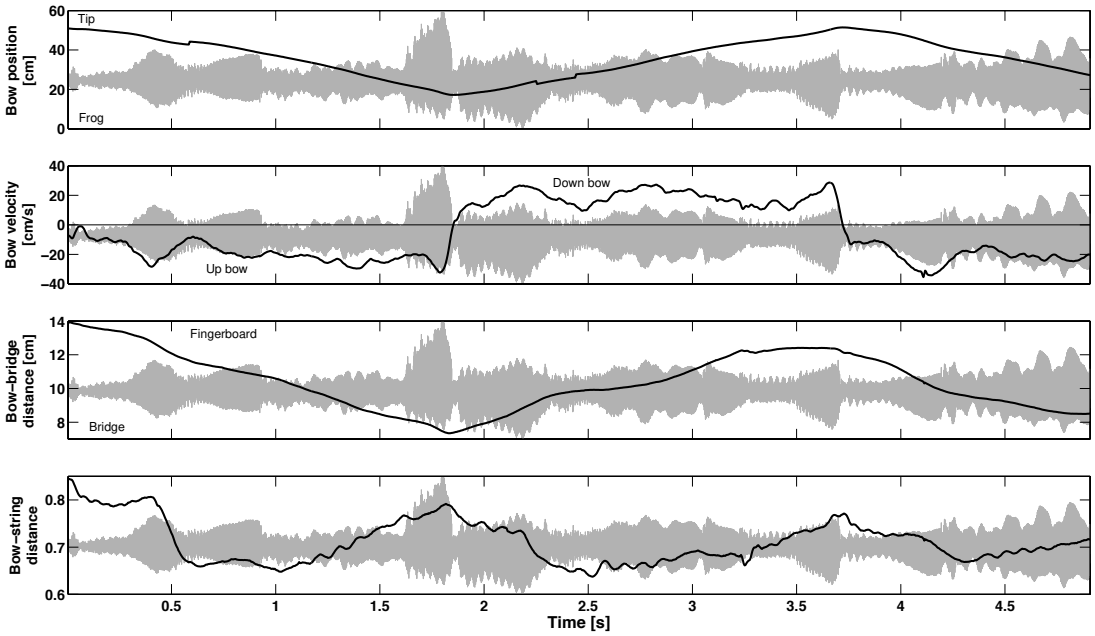
*MIRtoolbox 1.5 (Lartillot et al., 2008), **Timbre Toolbox 1.4 (Peeters et al., 2011), ***SubBand frequency ranges are given in Table B.1

Appendix C

Experimental Data Examples

Figure C.1 shows two different bowing control combinations observed between the players in the recorded cello database (Chapter 5). The presented bowing parameters of Cellists 1 and 2, captured in the Shost1 excerpt, are complemented with spectrograms (Figure C.2) and long-term average spectra (LTAS) (Figure C.3) of the respective audio signals to illustrate the effect the individual bowing controls had on spectral contents of the cellists' tones.

(a)



(b)

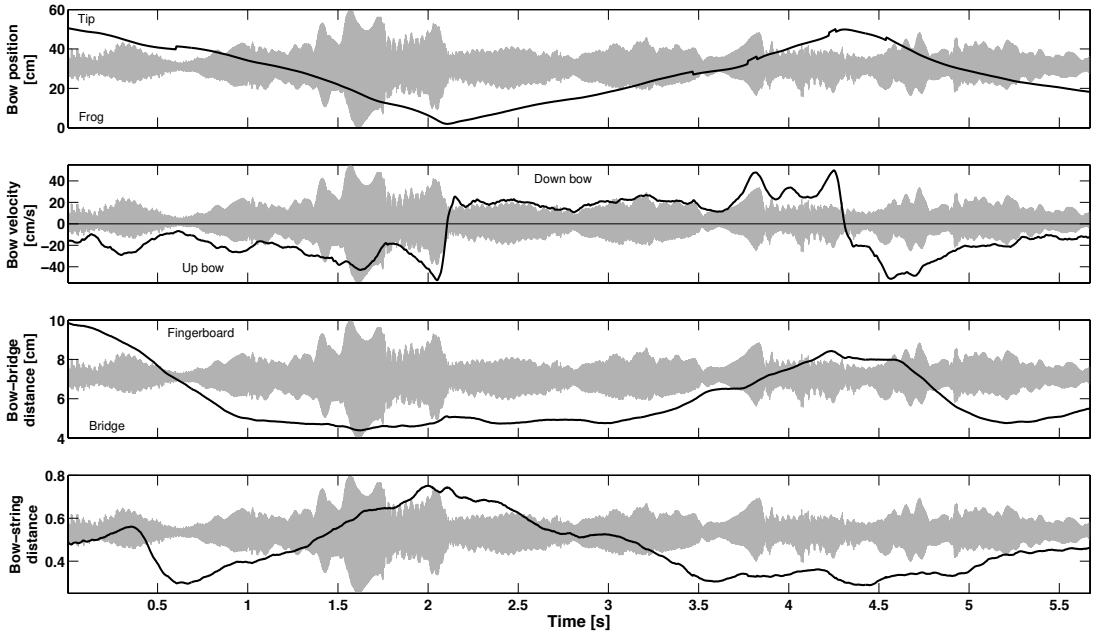
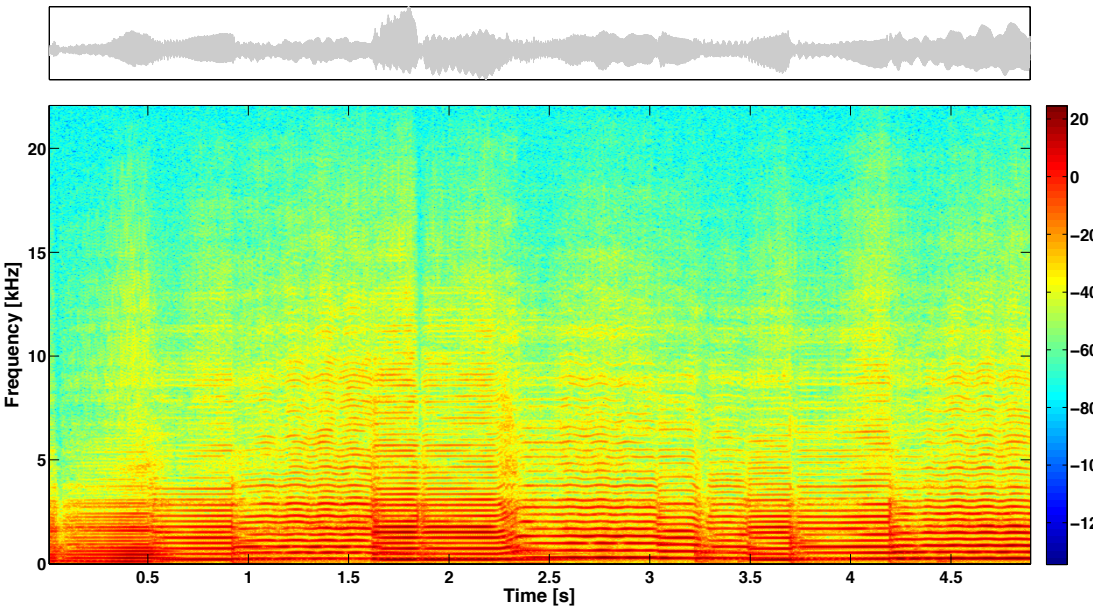


Figure C.1: Shost1. Comparison of bowing parameters extracted from the captured motion data of (a) Cellist 1 and (b) Cellist 2. The waveforms of the respective audio samples are shown in the background (in grey). Note the differences in the parameters' ranges between the two players. For Cellist 1, the means of bow velocity, bow-bridge distance and bow-string distance across notes were 18.27 cm/s, 10.66 cm and 0.71 respectively, compared to the corresponding values of 23.33 cm/s, 6.31 cm and 0.45 for Cellist 2.

(a)



(b)

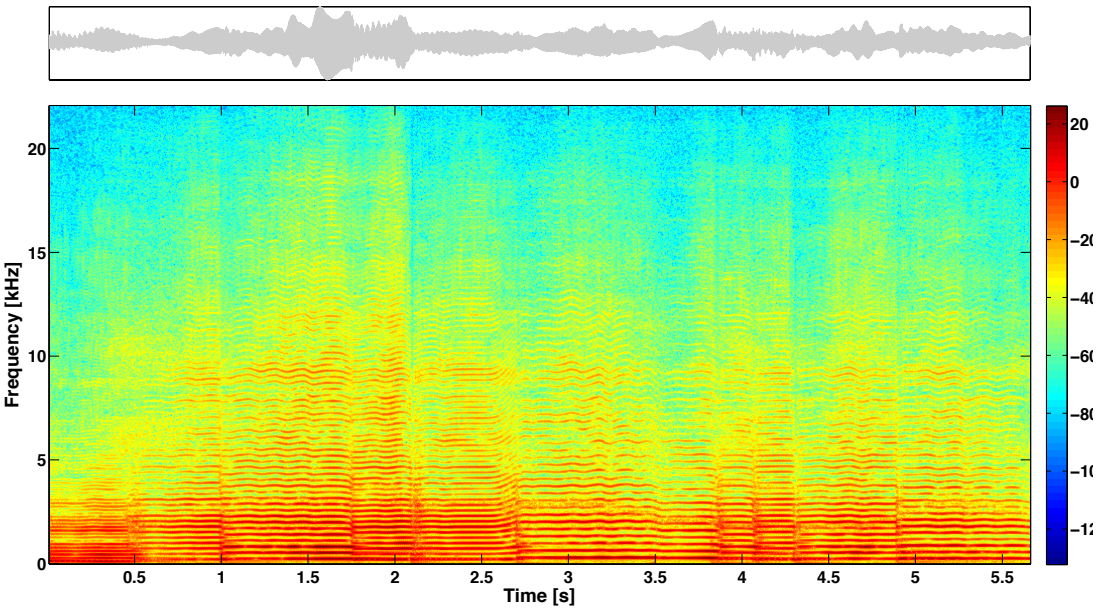


Figure C.2: Shost1. Comparison of spectrograms obtained from the audio samples of (a) Cellist 1 and (b) Cellist 2. The respective waveforms are shown in the upper plots. Instantaneous STFT power spectra were computed using 23.2-ms frames with 75% overlap and a 21.53 Hz frequency resolution.

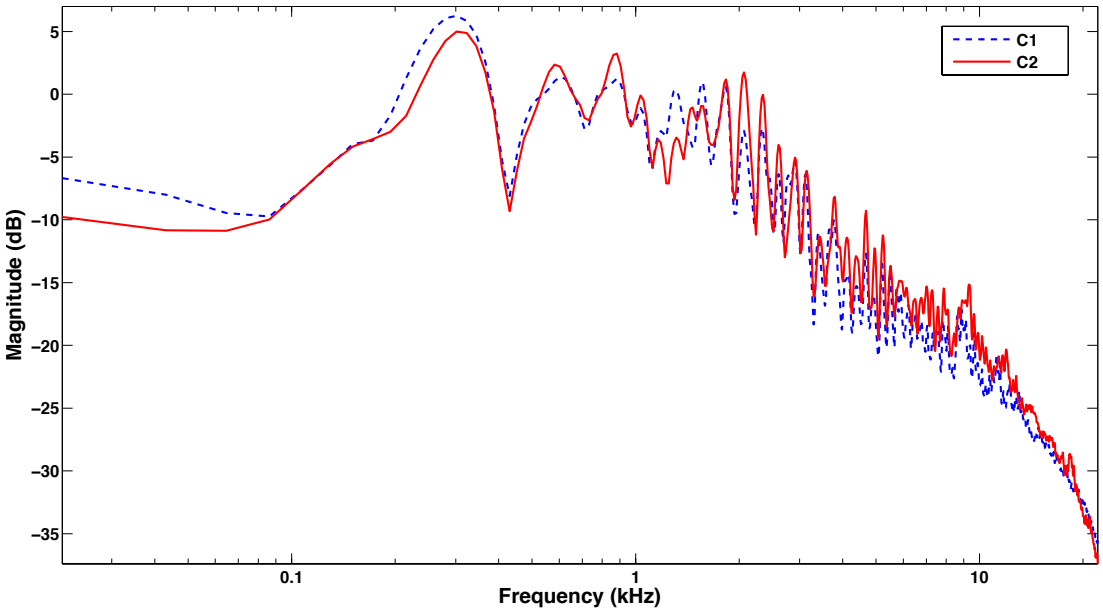


Figure C.3: Shost1. Comparison of long-term average spectra (LTAS) obtained from the audio samples of Cellist 1 (blue dashed line) and Cellist 2 (red line). Instantaneous STFT magnitude spectra were computed using 23.2-ms frames with 75% overlap and a 21.53 Hz frequency resolution. Note higher amplitudes of the frequencies from 550 Hz upwards combined with lower amplitudes of the frequency components around 300 Hz and below 100 Hz in the spectrum of Cellist 2. The corresponding value of spectral centroid averaged across notes was 1858.67 Hz compared to 1531.87 Hz for Cellist 1.