Exploring the perception of violin qualities: student- vs. performance-level instruments, strings and soundpost height

Lei Fu

Music Technology Area

Department of Music Research, Schulich School of Music

McGill University

Montreal, Canada

March 2020

A dissertation submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

The violin has reached its current, highly refined form over centuries through empirical methods, with some of the most valuable instruments being made more than 300 years ago. It is thus perhaps remarkable that very little is understood about violin quality and what particular aspects of the instrument are most important in the perception of violin qualities. This thesis reports three perceptual experiments that were designed to help better understand the relationship between a violin's perceived qualities and its physical structure.

Previous studies [Saitis et al., 2012; Fritz et al., 2012b, 2014] have shown a general lack of agreement among players in terms of violin preference when evaluating instruments intended for intermediate to advanced players ($1300 CAD and higher). The first experiment of this thesis explored whether there would be greater perceptual agreement when comparing violins meant for entry-level vs. advanced players, whether there would be significant perceptual differences between these two categories of violins and whether some structural vibration characteristics could be found to explain these differences. The results showed that performance violins were on average rated significantly higher than student violins in terms of preference and the three attribute criteria *clarity*, *richness* and *balance*.

The second and third studies of this thesis investigated the origin of the disagreement among players through two specific modifications to the violin. The second study investigated how different strings affect the perception of violin qualities. Two violins of the same make with similar sound quality and playability were employed. They were both strung with Dominant strings initially. Subjects played the violins, described and rated the difference (on eight criteria - *responsiveness*, *power*, *resonance*, *brightness*, *clarity*, *richness*, *balance* and overall quality) between the two violins during a session labeled D1-D2. Subsequently, the strings of violin 2 were changed to a different brand (Kaplan or Pro-Arté), unbeknownst to the players, and players had to re-evaluate the differences between the two violins (session D1-K2 or D1-P2). Results showed no significant differences between the experimental conditions except that the *brightness* difference ratings obtained in D1-D2 were found to be significantly higher than those in D1-P2.

The third study involved both playing and listening (using recorded sounds) experiments to investigate how changes in soundpost height (for a fixed soundpost position) affect the perceptual qualities of the violin and what is the threshold of change below which players and

luthiers do not perceive differences. A height-adjustable carbon fibre soundpost was employed. During the playing experiment, subjects played, in a first phase, a provided violin on which the soundpost height was modified by the experimenter in order to find their optimal soundpost height. Then, in a second phase, the experimenter varied the soundpost height randomly in ten trials (including cases where no change was made) within a range of approximately ± 0.1 mm around their optimal height. The results showed that the subjects' optimal soundpost heights varied from 0.132 mm to 0.616 mm relative to the original soundpost height (53 mm), reasonably well inside the extreme soundpost heights that were tested (0 mm and 0.66 mm, relative to the original height). The smallest height variation that could be recognized above chance level was about 0.044 mm for players and 0.088 mm for makers. During the listening experiment, subjects listened to 16 pairs of recordings through a computer interface and were asked, for each pair, whether the violin set-up was the same or different. The results showed that subjects in the listening experiment could differentiate soundpost heights with a difference of about 0.066 mm at better than chance levels.

Résumé

Le violon a atteint sa forme actuelle après plusieurs siècles de développements empiriques et certains des instruments les plus renommés actuellement ont été fabriqués il y a plus de 300 ans. Il est ainsi assez remarquable que notre compréhension de la qualité d'un violon et de ce qui y contribue reste limitée. Cette thèse décrit trois expériences perceptives qui ont été mises en place pour mieux comprendre les relations entre les qualités perçues d'un violon et sa structure physique.

Des travaux antérieurs [Saitis et al., 2012; Fritz et al., 2012b, 2014] ont montré que les musiciens n'étaient généralement pas d'accord entre eux en termes de préférence lors de l'évaluation de violons au-dessus d'un certain prix (1300 $ CAD). La première étude perceptive s'est intéressée à examiner si l'accord entre les musiciens lors d'essais comparatifs était plus grand pour des violons d'entrée de gamme que pour des violons « de concert », si des différences perceptives significatives pouvaient être mises en évidence entre les deux catégories de violons et s'il existait des caractéristiques vibro-acoustiques qui pourraient expliquer ces différences. Les résultats montrent que les violons de concert ont été évalués, de manière significative, plus positivement que les violons d'étude, aussi bien en termes de préférence que sur les critères *clarté*, *richesse* et *équilibre*.

Les deux autres études ont exploré les sources de désaccord entre musiciens à travers deux modifications spécifiques du violon. Comment les cordes influencent la perception des qualités d'un violon fut l'objet de la deuxième étude. Deux violons, de qualité sonore et de jouabilité très similaires ont été utilisés. Ils étaient initialement montés avec des cordes Dominant. Les participants devaient jouer les deux violons puis décrire et évaluer les différences au cours d'une session intitulée D1-D2. Ensuite, à l'insu des participants, les cordes du violon 2 étaient changées pour un autre modèle (Kaplan ou Pro-Arté) et les musiciens devaient réévaluer les différences entre les deux violons (session D1-K2 ou D1-P2). Les résultats n'ont pas montré de différences significatives entre les diverses conditions expérimentales, à l'exception des évaluations de différence de *brillance* qui sont significativement plus élevées dans la session D1-D2 que dans la session D1-P2.

La troisième étude consistait en des tests de jeu mais aussi d'écoute (pour explorer comment des modifications de la hauteur de l'âme (pour une position donnée de l'âme) influencent les qualités perceptives d'un violon et quel est le seuil de modification en-dessous duquel luthiers et violonistes ne perçoivent pas de différences. A cet effet, une âme en fibre de carbone de longueur ajustable a été utilisée. Durant le test de jeu, les participants commençaient par évaluer les qualités d'un violon donné, dont la longueur de l'âme était modifiée par un expérimentateur, afin de déterminer la longueur optimale pour chacun d'entre eux. Dans un deuxième temps, l'expérimentateur modifiait, ou non, la longueur de l'âme, dans une plage de ± 0,1 mm autour de la longueur optimale, et les participants devaient dire, si le réglage avait été modifié ou pas. Les résultats montrent que la longueur optimale pour chaque sujet varie entre 0.132 mm et 0.616 mm de plus que la longueur de départ de l'âme (53 mm), un intervalle bien compris entre les valeurs extrêmes qui ont été testées (de 0 à 0.66 mm). La plus petite variation de longueur pouvant être détectée au-dessus du seuil de chance est de 0,044 mm pour les musiciens et 0,088 mm pour les luthiers. Durant le test d'écoute, les sujets devaient écouter 16 paires d'enregistrements via une interface sur ordinateur et spécifier, pour chaque paire, si le réglage du violon était le même ou non pour les deux enregistrements. Les résultats montrent que la différence de longueur pouvant être détectée au-dessus du seuil de chance est d'environ 0,066 mm.

Acknowledgements

First and foremost, I would like to thank my co-supervisors Gary Scavone and Claudia Fritz for providing inspiring guidance, constant support, constructive feedback, and spending considerable time on discussing experiment design, solving problems occurred during the experiment, interpreting the data and revising all manuscripts and this thesis with great enthusiasm. I would like to show my great appreciation to Gary for suggesting this interesting research topic to me, providing advising and helping me ensure my essential measurements worked smoothly. I'm tremendously grateful to Claudia Fritz for her guidance in violin perception experiment design and various statistical analysis. I had the pleasure of working with her over a three-month trip to l'équipe Lutheries-Acoustique-Musique (LAM) in Sorbonne University and three violin acoustics workshops. Thank her for facilitating my visiting to LAM and life in Paris as well as the translation of the thesis abstract.

I would like to extend my gratitude to the professors in Music Technology Area, especially Stephen McAdams for discussing with me about my experiment data analysis and Ichiro Fujinaga for his guidance and encouragement at the beginning of the life in Montreal. I express my warm thanks to visiting Prof. Xin Wang for discussing with me about my experiment design and many topics related. I'm also grateful to Prof. Caroline Traube, Jim Woodhouse for important inputs before my formal experiments. I would like to sincerely thank Prof. Ruiqing Jia and Ailin Zhang for their initial encouragement for entering this interesting study area and continuous concern.

I gratefully thank my colleagues in CAML and Music Technology Area for their various valuable helping: Cornor, Douglas, Harish, Jason, Johnty, Lingxiao, Martha, Negin, Quim, Song, visiting professor Zijin Li, Wen, Yaolong, Yinan and Yong. Special thanks are due to Esteban Maestre for helping with my measurements and modal analysis, Charalampos Saitis for statistics analysis and Tanor Bonin for discussions about my experiment design. Many thanks to Darryl Cameron for his technical support and to Yves Méthot for his assistance in conducting my experiments in CIRMMT.

Attending the violin acoustics workshops in Oberlin in the summers of 2017 and 2018, and in Ghent in the Spring of 2018 has been a great learning experience. I want to thank the luthiers, musicians, and researchers for fruitful discussions and their participation in my experiments. My sincere thanks go to Fan Tao for providing strings for my research and facilitating my participation. I am also grateful to Colin Gough, George Bissinger, George Stoppani and Joseph Curtin for discussing with me about my research direction. Gratitude also goes to violinist Nathan Giem for making the recording and all luthiers who helped with installing and adjusting the soundpost and the radiation measurements.

Through generous funding by the Graduate Mobility Award provided by Schulich School of Music, McGill, I was able to visit LAM in Sorbonne University during the spring of 2018 and worked closely with Claudia. Thanks are due to many people, especially my colleagues in LAM for facilitating my visiting, providing technical support, discussing with me about my experiment, participating in pilot study and helping with measurements. Special thanks to violin maker Paul Noulet and violinist Lea for my pilot study recordings. Thanks to Senior and Junior Pal Molnar for generously loaning the Anima Nova height-adjustable soundpost, as well as for instructions on the installation and use of the soundpost. Thanks are also due to Hanqi Zhu, Honge Yusheng, Tingting Geng, Xiao Xiao and Yan Huang for the friendship.

My gratitude also goes to the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), the Natural Sciences and Engineering Research Council of Canada (NSERC), the Hector and Ada Ma China Scholarship Council Fellowship and the China Scholarship Council (CSC) for providing funding for this research. I want to thank the Schulich School of Music for loaning violins. I'm also grateful to violin maker Tom Wilder for lending violins for the string experiment, helping with the installation of the soundpost to the violin, and giving important suggestions for my research direction. Thanks to violinist Zhi Xu for discussing with me about violin experiment design. Thanks to violinist Liqiong Liu for helping with instrument selection. Thanks to all the violinists and luthiers who participated in the experiments.

I would like to thank my friends for spending precious time together: Carli, Danlan, Jiaying, Junqian, Shunyao, Siting, Wenyang, Xiaohua, Xin, Yang, Yiqin and Zhibo. Thank Yinan for discussing the perceptual experiment design with me. Thank my violin teacher Julie and Yiyun. Thank Yixing for encouragement, accompany and all the positive energy. At last, my greatest appreciation goes to my parents for their precious love, to my dear grandfather and my brother for their support.

Table of Contents

Contribution of Authors

This thesis, and the research to which it refers, is the candidate's own original work except for commonly understood and accepted ideas or where explicit reference to the work of other people, published, or otherwise, is made. The dissertation is formatted as a monograph comprising seven chapters and includes contents from the following conference publications:

1. Chapter 3 : Fu, L., Scavone, G., Fritz, C. (2019). "Player evaluation of performance and student violins". In Proceedings of 26th International Congress on Sound and Vibration (Montreal, Canada).
2. Chapter 4 : Fu, L., Scavone, G., Fritz, C. (2018). "How different strings affect violin qualities". In Proceedings of 176th Meeting of Acoustical Society of America & 2018 Acoustics Week in Canada. **35**, 035003 (Victoria, Canada).
3. Chapter 5 : Fu, L., Fritz, C., Scavone, G. (2019). "Perception of violin soundpost height differences". In Proceedings of International Symposium on Music Acoustics 2019, 450-457 (Detmold, Germany).

The candidate was responsible for every step involved in designing and conducting all experiments in this dissertation, as well as analyzing the collected data and preparing manuscripts for the publications listed above. Gary P. Scavone and Claudia Fritz, a thesis co-advisor working at Lutheries-Acoustique-Musique in Sorbonne Université as well as Centre National de la Recherche Scientifique (Paris, France) contributed with guidance in experimental design, data analysis and intepretation of the results, as well as review of the manuscripts. For the study in chapter 5, Gary Scavone participated as an experimenter modifying the soundpost height. For the study in chapter 6, Claudia Fritz helped with the organization of the violin sound recording session and the pilot study; Gary Scavone developed the computer interface for the listening test. Gary Scavone provided necessary funding, laboratory equipment and space.

Chapter 1

1 Introduction

1.1 Motivation and Objectives

The "quality" of a violin is the general overall impression that people perceive related to the violin. It depends on many factors according to players, luthiers and researchers, including the physical characteristics and vibrational properties of the plates [Hutchins, 1981], the vibrational properties of the strings, the bridge, soundpost, bassbar and the properties of the varnish [Schelleng, 1968]. As it relates to the human perception, the violin quality can also be influenced by the player's expertise, preference, mood, etc. Thus, it is necessary to employ formal psychoacoustic experimental protocols for violin quality evaluation in an effort to control for variables such as experiment venue, player expertise, or visual identification of instrument make. A violin's overall quality evaluation will depend on many different aspects of its sounding or playing behaviour, such as *richness*, *response* or *projection*, and these "sub-characteristics" are generally referred to as "qualities." Correlating the changes in the physical structure and specific dynamic behaviour of the violin with the perceptual quality/qualities of the violin has been a long-standing goal of violin research.

After nearly 350 years of research, the functioning of the violin is now fairly well understood, including the stick-slip Helmholtz motion of the bowed string [Helmholtz, 1863; 1954; Raman, 1918], the influence of the physical properties of the string on the Helmholtz motion [Cremer, 1984], the boundary conditions to sustain a stable Helmholtz motion [Schelleng, 1973] and the transients for its creation [Guettler, 2002a; 2002b]. The vibration behavior of the violin body determines the sound quality, intensity as well as playability to a large extent. With the development of new experimental measurement systems and computational methods, our knowledge about the violin body has increased significantly. From the use of Chladni patterns, holographic visualization techniques, modal analysis, finite element modeling [Rodgers, 2001;

Gough, 2015], as well as statistical analysis methods, we have been able to understand more about the signature modes in the low frequency range, the bridge hill in the middle frequency range and their origins, as well as the high frequency response of the violin body and other components.

Along with the in-depth understanding of the violin and the development of various measurement methods, scientists have been working together with makers to search for the "secret" of "good" violins. They have measured the admittance and/or radiation characteristics of the distinguished old Italian violins or violins with different qualities to search for the similarities shared among each group. Computed Tomography (CT) scanning has also been applied to extract the exact geometry or wood density on a large number of good violins, though it is not practically possible to replicate a particular instrument by copying geometrical properties because of variations in wood properties. For instance, Bissinger [2008] realized the CT scan of 17 violins for density-shape material information. He conducted a study over about 10 years seeking out possible "robust" parameters that could be used to identify quality relationships in violins. A wide range of vibration and radiation measurements were carried out on these 17 violins, which were quality-rated by a professional violinist and Bissinger himself from "bad" to "excellent". Little difference was found between the very best violins and the worst (when properly setup) except the radiation of Helmholtz-like cavity mode A0 was significantly stronger for the excellent violins. It is uncertain whether these results were reliable or generalizable as their assessments about the violin qualities were based on a few violinists or only the subjectivity of the authors. These largely inconclusive results also led Bissinger to remark: "Perhaps a contrarian viewpoint about quality might be useful here? What truly defines violin excellence? If the answer is truly excellent violinists, then the reliability-reproducibility of their psychoacoustic judgments must draw more attention. It would seem illogical to expect violinists who pride themselves on their personal sound not to prefer certain violins over others because they are better at creating that sound".

Thus, in recent years, several scholars have conducted well-controlled perceptual evaluations of violin qualities. Saitis et al. [2012, 2015] performed a series of experiments to investigate violinists' evaluation process. It was found that violinists were self-consistent while evaluating violins, however, the lack of agreement between different players was significant. The violinists tended to agree with each other on "*richness*" and "*dynamic range*" as important criteria for determining the preference of violins. These researchers also found that the players were better able to discriminate between violins in playing tasks than in listening tasks. Fritz et al. [2012b, 2014, 2017] conducted a series of experiments investigating players' and listeners' preference among new and old violins. In the study of 2014, there was general agreement on one or two violins: one new violin was chosen as the first or second favorite violin by 8 out of 10 subjects, and one old violin was rejected by 9 out of 10 players, but otherwise, many violins were selected almost as many times among the favorites as among the rejected violins. Combing the results from the previous violin evaluation experiments, we may wonder whether any violin could be considered as a good violin, and whether the violin quality is just a question of preference. Could we find a limit of "goodness" under which players would agree that the violins are not good? If so, what are the differences between those "bad" violins and the "good" ones in terms of perception (which criteria are relevant to tell them apart) and in terms of vibratory response? Thus, the first study in this thesis sought to assess whether the lower quality Suzuki violins would be consistently distinguished by violinists from the better-quality violins under more controlled conditions and whether there would be agreement regarding the qualities of those instruments that the subjects found less desirable. Bridge admittance measurements were also performed to search for the differences between the two types of violins.

What factors affect the perception/quality of the violin? Violins can differ on a large number of criteria, thus it is not easy to understand where the disagreement among players originates from. One approach to this problem is to make a modification on a given instrument and then ask players to re-evaluate it. The components of a violin that can be changed include the bridge, soundpost, tailpiece and strings. Of these, however, only the strings are normally changed by players themselves and they tend to agree that strings can make a big difference. Thus, the second study in this thesis was to investigate the effect that different types of strings can have on the perception of the violin. In the third study, we decided to investigate the influence of another modification, which could be done more quickly, and which could lead to a quantification of the perceived differences through objective measurements, and not just through verbal data (which is difficult to analyze as the vocabulary used by players is very large). Taking advantage of a height-adjustable soundpost, we decided to explore the influence of the height of the soundpost on the perceived quality of the violin. The detailed experiment design will be described in the next subsection.

1.2 Content and Structure of Thesis

This thesis consists of seven chapters including this introduction. Chapter 2 begins with the basic knowledge about the violin, introducing the structure of the violin and the commonly used bridge admittance measurement. Then the previous studies correlating mechanical characteristics to the violin quality are summarized. Subsequently, the violin quality evaluation studies employing formal psychoacoustic procedures performed by different researchers in recent years including playing tests, listening tests and linguistic analysis are presented. At the end of this chapter, previous research about the strings and soundpost of the violin is reviewed.

In Chapter 3, a perceptual experiment examining how violinists differentiate between performance and student (entry-level) violins, and their level of agreement, is reported. Bridge admittances were also measured to search for corresponding objective violin quality parameters. A pool of six violins of "different qualities" was assembled: 3 performance violins and 3 student violins. Nine violinists participated in this experiment. Among them, three violinists described themselves as professional violinists. Subjects were scheduled individually. We asked the subjects to rank and rate the 6 violins according to preference as well as five attribute criteria on a continuous scale from 0 to 5. They were also asked to provide written responses to questions related to their evaluation criteria after each rating scale. Quantitive analysis about the players' preference and criteria ratings of the violins as well as the agreement between players were performed. Verbal responses collected from open-ended questionnaires were also analyzed to investigate the criteria that violinists value during the violin quality evaluation in addition to the quantitive results.

In Chapter 4, the second study investigating how different strings affect the violin quality was carried out through a perceptual experiment. Two violins of the same make with similar sound quality and playability and three types of strings (Dominant strings, Kaplan strings, and Pro-Arté strings) were employed. The two violins were both strung with Dominant strings initially (session labeled D1-D2). Subjects played the violins, described and rated the difference between the two violins (violin 2 compared to violin 1) according to eight criteria. Subsequently, the strings of violin 2 were changed to a different type. Subjects rated the difference between the two violins again. In Oberlin, nine subjects compared Dominant and Kaplan strings in two sessions (called respectively D1-D2 and D1-K2). In Montreal, ten subjects compared Dominant, Kaplan and Pro-

Arté strings in three trials (D1-D2, D1-K2 and D1-P2). Statistical analysis of the differences between the experimental conditions in either place was conducted.

In Chapter 5 and 6, we explore the perception of soundpost height differences through a playing test and a listening test, respectively. A height-adjustable carbon fibre soundpost was employed. Thirteen violinists and six luthiers participated in the playing experiment and thirteen violinists and eight luthiers participated in the listening experiment. During the playing experiment, subjects played a provided violin on which the soundpost height was modified by the experimenter in order to find their optimal soundpost height. Then, within a range of approximately ± 0.1 mm around their optimal height, the experimenter varied the soundpost height randomly in ten trials (including cases where no change was made). Subjects played the violin and compared it with the previous setting to decide whether it was the same setup or not. The variation of the optimal soundpost height among all subjects, players and makers was analyzed. The perceptual sensitivity of the soundpost height differences around the optimal soundpost height was estimated as well as the comparison between the results of players and makers. During the listening experiment, subjects performed pairwise comparisons of the recordings through a computer interface. The pairs of recordings included identical recordings, different recordings at the same soundpost height and recordings at different soundpost heights. The recordings were made on a different violin from the playing experiment. The perceptual sensitivity of the soundpost height differences was estimated besides the comparison between the results of players and makers.

Chapter 7 concludes and discusses the main findings in the three studies and makes suggestions for future research.

Chapter 2

2 Literature Review and Research Questions

2.1 Introduction

The violin has reached its current, highly refined form over centuries through empirical methods. Many efforts have been made to understand the working mechanism of the violin through vibrational and acoustical measurements. It is a long-term goal to correlate the physical measurements of the violin with its perceptual quality. By comparing the mechanical characteristics of the violin with perceptual evaluations, scientists hope to distinguish the good instruments from the inferior ones. Formal psychoacoustic evaluations have provided many important additions to our comprehension of violin quality in recent years.

Section 2.2 of this chapter reviews relevant knowledge about the violin structure and acoustic measurements of the violin. Several experiments have attempted to correlate the mechanical characteristics to the violin quality, which are outlined in Section 2.3. Section 2.4 describes the experiments of violin quality evaluation from three aspects: playing test, listening test and linguistic analysis. Section 2.5 reviews relevant research about factors that could influence the violin quality, including strings and the soundpost which we explored further in this thesis. Finally, Section 2.6 discusses the research questions of this thesis in the context of this literature review.

2.2 Basic Knowledge of the Violin

2.2.1 The Structure of the Violin

The most prized old Italian violins were made by the "masters" Antonio Stradivari (1644-1737), Giuseppi Guarneri (1678-1744) and his family, the Amati family, Jacobus Stainer, etc... [Hutchins and Benade, 1997].

It is generally accepted that the first violin emerged in the early sixteenth century, combining the body shape of lira da braccio (a variant of the medieval fiddle) and the stringing arrangement and tuning manner of the rebec (a pear-shaped back with a neck smoothly merged) by an unknown Italian maker. The result was a three-stringed instrument. A fourth string with a higher pitch than the initial three strings was added by about 1550. They were tuned as the modern violin: G3-D4-A4-E5. The oldest surviving instruments are from Cremona [Campbell et al., 2014]. The early development of the violin lasted more than 100 years. Subsequently, the violin was manufactured, modified and refined according to the demands of players and makers, its tone and playing qualities reaching the height of excellence in the early 18th century [Hutchins and Benade, 1997]. In 1704, Stradivari arrived at a violin model that became a prototype for himself as well as subsequent luthiers up to modern times [Campbell et al., 2014].

In the late eighteenth and early nineteenth centuries, the innovation of the violin shifted from Italy to France. In response to the rise in frequency of the musical note A and the demand for a more powerful sound in bigger orchestras, various structural modifications were introduced to the violin: increase of the string length, which also required a longer neck and higher bridge, and a heavier bassbar as the string tension also increased to achieve the traditional pitches. All these changes resulted in a so-called modern setup [Curtin and Rossing, 2010]. Further innovations happened later, such as the chin rest on the violin and materials of the strings: metal wound gut or synthetic (mainly nylon) cores or single steel strand replaced the traditional gut strings. These changes in strings remarkably increased playability ("playability" correlates to the mechanical interactions between the instrument and the player [Zhang, 2015; Woodhouse, 1993a, 1993b]) and stability, which contributed to brilliance and power further [Curtin and Rossing, 2010]. The old Italian violins we know today were repaired and rebuilt in the modern style to adapt to all the changes.

Figure 2.1 shows the essential components of the violin as we know it today. The bridge stands in line with the notches of the f-holes under the tension of the strings. The soundpost is squeezed in between the two plates, and usually under and below the treble foot of the bridge. Vibrations of the string that are initiated by an applied force and motion of the player's bow are transmitted via the bridge to the top plate and then to the complete violin body. The violin body and the air cavity work as an amplifier and filter, through which the vibration energy is able to radiate into the surrounding air and ultimately reach the ear of listeners. Thus, the vibration behavior of the violin body determines the intensity (or sound radiation) and timbre of the violin sound (or sound quality), as well as playability to a large extent [Fletcher and Rossing, 1998; Saitis, 2013; Woodhouse, 1993a; 1993b]. There has been extensive research on the vibrational characteristics of the violin body. The vibration of the violin body and related measurement methods will be introduced in the next section.

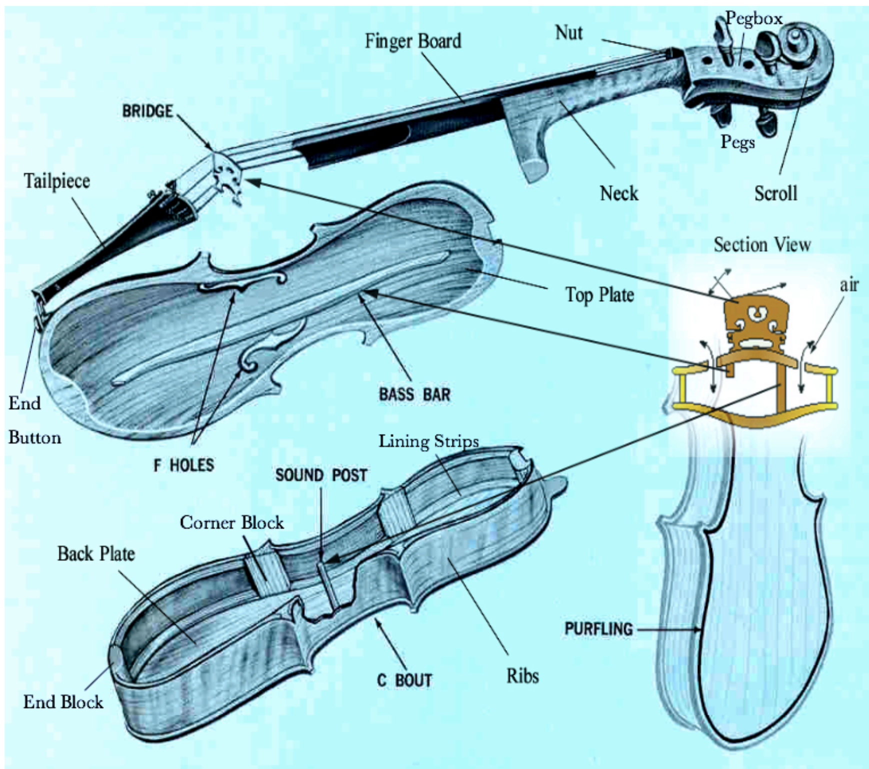


Figure 2.1 Exploded view of violin showing the components by Hutchins [Hutchins, 1967; Zhang, 2015].

2.2.2 Bridge Admittance Measurement

All vibrating structures will exhibit some number of normal modes of vibration. Each normal mode can be characterized by a natural frequency, a mode shape formed by nodal lines (along which the vibrations have minimum amplitude) and anti-nodes (where the amplitude of the vibrations is at maximum), a damping factor and a radiation pattern and strength [Gough, 2007; Woodhouse, 2014; Curtin and Rossing, 2010]. In the low frequency range of the structural vibration (below 1kHz for violins), it is easier to distinguish individual modes. The first few modes, often called the "signature modes", are well separated from each other and are considered very important to the violin sound [Bissinger, 2008]. Similar mode shapes can be detected in most normal violins. However, as the frequency increases, the modal overlap factor (a ratio of damping bandwidth to the modal spacing) increases as well, with more than one mode contributing significantly at each specific frequency, and thus the study of different modes seems less useful and different analysis methodologies may be employed [Woodhouse, 2014].

The signature mode shapes of the violin can be visualized by Chladni plate vibrations (usually for unassembled plates), holographic interference and modal analysis techniques. The Chladni patterns are formed by powder or particles of thin aluminum flake bouncing up and down on the vibrating plate (excited acoustically, electromagnetically or with a bow drawn across an edge), slowly gathering together at the nonvibrating nodal areas, thereby outlining the nodal lines of the vibration mode [Hutchins, 1981]. Time-averaged holographic interferometry allows the visualization of both the nodal lines and the anti-nodal areas of the violin modes through laser beam interference patterns [Jansson et al., 1970]. Modal analysis can obtain detailed modal parameters by measuring the frequency responses [Marshall, 1985]. The technique typically involves the application of an impulsive force at one point and the measurement of acceleration or velocity responses at a large number of points on the surface of the violin, from which mode shapes and frequencies can be deduced.

The ratio of the induced velocity (acceleration) to the applied force is known as the mechanical admittance or mobility (accelerance). Input, or bridge, admittance is typically measured at one corner of the violin bridge. The driving force is provided by a miniature impulse hammer at one top corner of the bridge along the bowing direction of the nearest string and the resulting velocity is measured on the same or other top corner by a laser vibrometer. The bridge admittance is the most common and easiest way to characterize the acoustical properties of the violin body [Gough, 2007]. It includes the essential information about the energy transfer between the string and the violin body [Cremer, 1984]. Figure 2.2 demonstrates the input admittance of a Guarneri violin measured on the bass bar side of the bridge [Alonso Moral and Jansson, 1982]. Some important resonance peaks below 1 kHz in the figure are identified and labelled with the corresponding names: each peak represents a signature mode. In the high frequency range, it is not possible to recognize individual modes as explained before, however, the frequency response shows a “hill” like feature formed by a cluster of peaks at around 2.5 kHz, which is called the “bridge hill” [Jansson and Niewczyk, 1999].

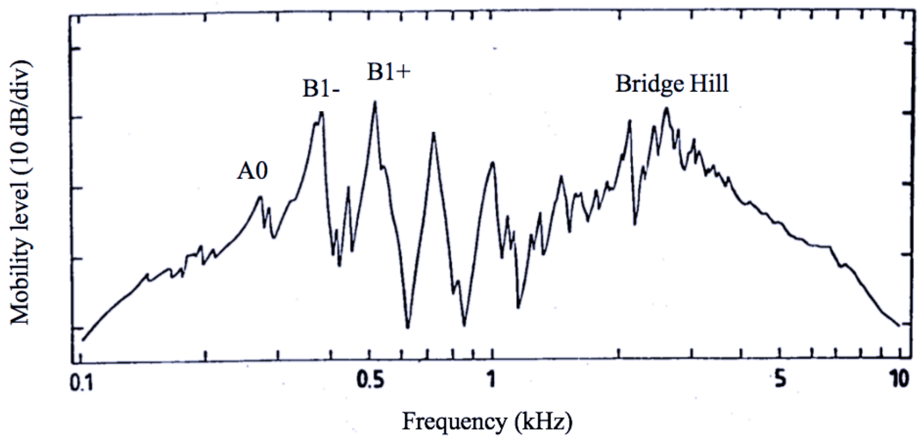


Figure 2.2 Bridge admittance of a Guarneri violin by Alonso Moral and Jansson [1982].

The signature modes in the low frequency region can be recognized in most normal violins with similar shapes and usually in the same order: (1) cavity mode A0, known as the “main air resonance” or “Helmholtz resonance”, often occurs around 280 Hz, is the lowest frequency dominant radiator and radiates primarily through the f-holes; (2) cavity mode A1, the 1st longitudinal mode, which is not identified in Figure 2.2, is a strong radiator for the large instruments but only for certain violins [Bissinger and Keiffer, 2003]. It is often at around 460 Hz; (3) corpus mode CBR (central bout rhomboid) existing around 400 Hz, which is not recognized in Figure 2.2 as well, is a strongly vibrating while weakly radiating mode; (4) main wood resonance modes B1- (around 480 Hz) and B1+ (around 550 Hz) are the 1st corpus bending modes and the lowest strong corpus radiators [Bissinger, 2005]. The A1 and CBR modes are not identified in the plot because it is not possible to know which peaks correspond to which modes without a full modal analysis or mode visualization.

2.3 Correlating Mechanical Characteristics with the Violin Quality

As partly described in the previous section, the working mechanism of the violin has been studied by many scientists for almost 350 years [Hutchins, 1997]. Knowledge about the physics of the violin including the string, the body, the bridge and the radiated sound are summarized in [Cremer, 1984]. This research helped people understand the correlation between the physical measurements, mechanical characteristics and vibrational or acoustical properties of the violin. Several studies have attempted to relate the mechanical characteristics and the vibrational or acoustical properties to the violin quality.

Meinel [1957] recorded the response curves of sound pressure of each string on a 1715 Stradivari violin (a concert violin which was thought to have a fascinating and fine tone quality) using a bowing machine. The sound pressure was measured at about one meter away. He found several significant characteristics regarding the superb violin timbre: (1) High amplitudes at the low frequency ranges which he thought could lead to agreeably sonorous sounds and carry well; (2) Low amplitudes at high frequencies above about 3 kHz that would allow a fine, pure response and a harmonious softness; (3) Low amplitudes near 1.5 kHz prevented the sound from being nasal; (4) Strong resonance between 2 kHz and 3 kHz (the bridge hill range) gave the sound an agreeable, pitchy and dull brightness. The author then applied the measurement to a broad range of violins including 6 good old Italian violins, 6 good and 6 mediocre modern violins and similar frequency response features were observed on the best violins. However, when he examined 4 violins of "bad" qualities, large deviations were found from the average curve of the tested old Italian violins.

Gabrielsson and Jansson [1979] investigated the long time average spectra (LTAS) of the recordings of 22 violins and their relations to the tonal quality ratings. During the 1975 instrument exhibition held by the Scandinavian Violin Maker Association, two professional violinists evaluated 103 violins from loudness equality and timbre of all notes and strings through playing specified scales and listening to each other's playing. Two three-octave scales: A major and A flat major were played evenly and slowly on each violin. Thereafter 22 violins were selected by the authors representing different tonal quality ratings. Another violin player was asked to make the recordings on each of the 22 violins in a reverberation chamber. The recordings consisted of three tone scales over three octaves started from the open G-string. The scales were played détaché with a tempo of about 60 bpm and as loud as possible. Different analyses implied that higher tonal quality ratings corresponded to "strong" frequency responses approximately from 200 Hz to 600 Hz and from 1.5 kHz to 3 kHz; "weak" frequency response around 1 kHz and above 3 kHz.

Alonso Moral and Jansson [1982] selected 77 violins from a violin makers' competition and asked two professional violinists to rate them for volume, brilliance of tones, evenness and playability. Then 24 violins were selected from the original pool covering different ranges of ratings and grouped in three classes according to their quality ratings. Together with an old concert Italian violin "Andrea Guarneri" (its input admittance has been shown in Figure 2.2), the researchers measured the input bridge admittances from both bassbar and soundpost sides of the bridge for these violins. It was found that the resonance peaks B1-, B1+, C4 (another eigenmode around 700 Hz labelled by Jansson) and a bridge hill around 3 kHz correlated strongly with the tonal quality. Higher average levels of the three signature mode frequencies, smaller discrepancies between the average level and single peak levels and higher slope from 1.4 to 3 kHz indicated better violin qualities.

Hutchins (1989) measured the A1 and B1+ modes frequencies of over 100 violins and found that the frequency spacing between the A1 and B1+ modes were closely correlated to the tone and playing qualities of the violins which were described by their owners-players. Violins with frequency differences between the A1 and B1+ modes over 100 Hz were too harsh and hardly playable; frequency differences less than 20 Hz were easily played and softly sounding while with little projection or power; frequency differences in the 40-70 Hz range were used by soloists and the 55-70 Hz range corresponded to the more powerful ones; frequency differences below 40 Hz were more played by chamber music violinists and were very easily played.

Dünnwald [1991] reported frequency response curves for a large set of violins using a different measurement technique. Violins were excited at the bridge by sinusoidal vibrations and the radiated sound was measured with one microphone which was placed at the typical concert listeners' positions in an anechoic chamber. He measured approximately 700 violins including 53 old Italian violins, 75 old master violins, 300 master violins made after 1800, roughly 180 factory made violins and 42 amateur maker made violins. He found four important frequency bands for assessing the violin sound quality: (1) the frequency range of 190-650 Hz is important for lower overtones (i.e. the location of the signature modes); (2) The sound will be boxy and nasal if the response in the frequency range of 650-1300 Hz is too strong; (3) Good radiation and brilliance depends on the frequency range of 1300-4200 Hz; (4) a low amplitude in the band of 4200-6400 Hz is responsible for the creation of a clear sound, otherwise the sound will be very harsh. These conclusions were similar to [Meinel, 1957] (though later perceptual tests, described in the next section, contradicted some of these conclusions). Dünnwald also regarded the higher radiation level of the Helmholtz resonance (A0 mode) as a criterion of violins with good tone quality.

Jansson [1997] later performed bridge admittance measurements on 25 violins of soloist quality from the Järnåker Foundation of the Royal Swedish Academy of Music. Similar conclusions were obtained as in [Alonso Moral and Jansson, 1982]. Those high quality violins had a dominating, high level of the B1+ peak and a noticeable broad peak and a phase step at 2.5 kHz. The conclusions agreed with his earlier experiment [Alonso Moral and Jansson, 1982] and in line with Dünnwald's frequency band features for not nasal, not harsh and clear tone [Dünnwald, 1991].

More recently, Bissinger [2008] conducted a study in which he measured a wide range of vibrational and sound radiation characteristics of 17 violins. A professional player rated 12 violins, while Bissinger himself rated the other 5, from bad to excellent quality. Bissinger found that there were no significant quality differentiators between the 17 violins, with the exception of the Helmholtz-like cavity mode A0. The radiation of this mode was significantly stronger for good than for bad violins.

2.4 Violin Quality Evaluation

The studies reported in the previous section explored the relationship between the violin quality and the signature mode characteristics or properties of bridge admittance or sound radiation measurements over different frequency ranges. Regarding the descriptions of the violin quality, most were based on the evaluation of the authors or a very small number of players, and no formal perceptual experiments were performed, thus it is uncertain whether the correlation results are reliable or generalizable. A formal perceptual experiment should typically involve a sufficient number of subjects, and the influences of parameters such as visual condition and the choice of bow should be controlled.

In recent years, scholars have conducted more formally controlled perceptual evaluations of violin qualities. These experiments included playing tests or listening tests. Playing tests allow players to explore the instrument with intimate contact, thus allowing them to try different playing techniques and play in different registers. In addition, scientists can study the vibrotactile feedback from playing tests. Listening tests have various forms: subjects listening to recordings/synthesized sounds played through the computer or subjects listening to live performance by designated players behind a screen. Recordings or synthesized sounds allow experimenters to adjust different parameters, which are not easy, possible or repeatable when using playing tests, based on different experimental purposes. While synthesized sounds may lack naturalness, researchers have to choose different test formats through comprehensive consideration. There is also a branch of violin quality evaluation classified as linguistic or semantic analysis, which is mainly concerned with the verbal descriptions used by subjects about violin quality.

2.4.1 Playing Tests

One somewhat recent formal scientific perceptual experiment on the violin was performed by Inta et al. [2005]. This experiment studied the effect of ageing and playing on the violin through three years of tracking. Two similar violins were employed, which were constructed "in parallel" at the beginning and were evaluated in both listening and playing tests in a concert hall. Afterwards, one of the violins was stored in a museum under controlled conditions and not played regularly, while the other was played regularly by a professional musician. After three years, listening and playing tests were carried out again without any adjustment on the two violins. Four days later, minor adjustments were made to the regularly played violin and the perceptual evaluations were repeated. The conditions for the evaluations were the same except that the subjects were not blindfolded for the first test, as the two violins looked quite similar. Listeners were not blindfolded. The same bow was employed across all tests. Players rated the violins from poor to excellent on a scale from 0 to 10 for 8 criteria: *evenness*, *responsiveness*, *dynamic range*, *speaking ability*, *brightness*, *warmth*, *distinctive character* and *playability*. Listeners rated the instruments similarly for 5 criteria: *clarity*, *projection*, *distinctive character*, *warmth* and *evenness*. The playing and listening tests showed no significant differences in all three evaluations, i.e., the effect of three years of playing on the violin quality was small.

Saitis et al. [2012, 2015] performed a series of experiments investigating the consistency and agreement of violinists when making violin quality evaluations. In the first study (Saitis et al., 2012) two violin playing tests were carried out and they made quantitative analyses of experienced violinists preference judgements. Players used their own bow and wore dark sunglasses. As well, the lighting in the experiment room was reduced so that the violins could not be visually identified. In the first experiment, 20 skilled players participated. They were asked to rank 8 violins during two identical sessions, which occurred on different days. During each session, the subjects ranked the violins according to their own preference five times over five trials. The results showed that players consistently ranked the same violins in terms of preference in different trials and on different days. The lack of agreement between different individuals however was significant. In the second experiment, the origin of lack of agreement between players was examined. Another 13 skilled violinists evaluated 10 violins according to 5 specific criteria (*easy to play*, *response*, *richness*, *balance* and *dynamic range*) as well as preference on continuous scales. Each violin was rated 3 times in 3 trials. The specific evaluation criteria were developed based on results from the verbal descriptions collected in the first experiment. It was shown that the players tend to agree to some extent on *richness* and *dynamic range* as criteria for determining preference. In 2015, the scholars conducted a new experiment to further examine the evaluation of *richness* and *dynamic range* from playing versus listening tasks [Saitis et al., 2015]. Sixteen skilled string players took part in this experiment. It was found that the players were better able to discriminate between violins in playing tasks than in listening tasks. In the playing test, players became more self-consistent and there was more agreement between players when the playing task was more focused, i.e., specific notes in specific registers.

There is a long history of comparison between the highly-priced old Italian violins and new violins made by contemporary luthiers. The comparisons were often performed through listening tests in less scientific contexts. Recently, several studies have been conducted to investigate the premise of the superior tonal quality of the old Italian violins in formal scientific experiments. In Fritz et al.'s studies [Fritz et al., 2012b, 2014], the researchers designed two experiments in double blind conditions to examine musicians' preference between old and distinguished Italian violins and new violins made by professional violin makers. The first study was conducted in a hotel room and 21 experienced violinists of various levels participated. Three old Italian violins and three new violins were employed. The second study took place at two venues: the home of a professional string player and a 300-seat concert hall. Ten soloists participated to evaluate six old Italian and six new violins. The studies found that the violinists could not tell old violins from new ones at better than chance levels. And a general preference for new violins was shown within the results. These results are a challenge to conventional wisdom. It implies that future research might best focus on how violinists evaluate instruments, what specific qualities they are most concerned with and how these qualities relate to physical characteristics of the instruments, whether old or new.

Wollman et al. [2014] studied the violin quality evaluation from a different aspect: how vibrotactile and auditory feedback affect the quality assessment of the violins. Fifteen violinists evaluated three violins on four criteria (*loud and powerful*, *pleasure*, *rich sound* and *alive and responsive*) in two conditions: a regular playing condition and a condition called "active listening" in which the participants listened to a professional violinist playing next to them (behind a screen) while fingering the score on an isolated neck which could vibrate (or not) similarly to the neck of the violin being played. The results demonstrated that the presence of vibration affected the judgment of the criterion of *loud and powerful.* In the listening test with vibrotactile feedback, violins were rated more positively with original vibration level at the isolated neck than with half the level for all criteria except for *alive and responsive*. No firm conclusions were drawn in the comparison between the playing and listening tests, however, the criteria were more highly rated in the listening test than playing test.

More recently, Fritz et al. [2016] studied the influence of the violin model (Stradivari, del Gesu, ...) on the quality evaluation of the violins employing a free sorting task. Twenty-one violinists were asked to freely play and sort 9 violins (5 of them were Stradivari model, 3 were del Gesu models, and 1 innovative design) into an unconstrained number of groups based on their perceived similarity. Contrary to the conventional belief, the results suggested no universal descriptors can be applied to a specific violin model.

2.4.2 Listening Tests

Fritz et al. [2007, 2010b, 2012a] performed a series of experiments using "virtual violins," enabling an identical performance to be replayed using different violin body responses, so that the relationship between acoustical characteristics of violins and perceived qualities could be better explored. A piezoelectric force sensor mounted on the bridge was used to record the representative force waveforms of the strings from real playing on a violin. In [Fritz et al., 2007], the recorded waveforms then were applied to a violin computer model using digital filters corresponding to the admittance curves of different real violins. The frequencies or amplitudes of single modes or frequency bands that Dünnwald [1991] proposed were changed in "virtual violins" and played to listeners. The correlations between different frequency ranges and violin sound properties that Dünnwald proposed were not confirmed in this listening test. Listeners' thresholds in detecting these changes were also characterized. In [Fritz et al., 2010b], a series of listening tests were conducted to explore the influence of the vibrato magnitude and damping level of the violin resonance modes on the perception of violin notes. Synthesized, recorded sounds or live performances were employed. The results showed that the vibrato magnitude and the damping level were independent perceptual dimensions. Another series of listening tests examined the effect of the vibrato and body damping on the judgments of *liveliness* and preference of the sound of single notes. It was found that the preference judgments were more consistent between subjects than the *liveliness* judgments. And no clear relationship between the vibrato magnitude and the *liveliness* ratings were found. The use of the word *liveliness* was found to be used inconsistently across participants. Thus, in the subsequent publication [Fritz et al., 2012a], the researchers explored the verbal descriptions employed by performers when describing the distinctive timbres of different violins. The collected descriptors were then used to correlate with acoustical modifications of "virtual violin sounds" in five frequency bands: 190-380 Hz, 380-760 Hz, 760-1520 Hz, 1520-3040 Hz, and 3040-6080 Hz through listening tests. The results showed that an increase in *harshness* corresponded to increased level in band 4 (1520-3040 Hz); increased *brightness* and *clarity* corresponded to increased levels in bands 4 (1520-3040 Hz) and 5 (3040-6080 Hz). Those results again were not consistent with what had been proposed by Dünnwald.

Projection is another important perceptual dimension for violinists and in general, instruments with better projection are thought to be preferable [Curtin and Schleske, 2003]. Loos [1995] performed a series of experiments examining the projection of violin sounds through both physical measurements and listening tests. Six music students played their own six violins of different quality (€2.5-20 K) in a small concert hall (900 m^2). Single notes and musical excerpts were played and recorded in the ears of players employing ear microphones at distances of both one and 12 meters from the violin. The author found the differences in sound pressure level and loudness of different violins are larger under the ear and at 1 m distance than between 1 and 12 m away. Listening tests were also conducted through A-B comparisons of six single notes with vibrato. Listeners compared the perceived nearness of each pair of sounds. Strong low harmonics seemed to enhance the perceived nearness. Fritz et al. [2016] continued the comparison between the old Italian violins and new master violins in terms of projection. The authors performed two experiments in Paris and New York separately. Each of the experiments was organized in a concert hall. Several soloists were invited to play behind an acoustically transparent screen. A group of listeners were presented with pairwise comparisons consisting of a performance on a Stradivari and a new violin. Three new violins and three old violins by Stradivari were involved in the experiments. Listeners were divided into two groups comparing the violins in terms of projection and preference separately. The results showed that the new violins were considered to project better and were more preferred than the Stradivari by listeners, and they could not discriminate the new and old violins at better than chance levels. The loudness under the ear of players rated by themselves were generally consistent with the projection ratings by listeners.

2.4.3 Linguistic Analysis

In the previous reported playing and listening tests, the authors often present some criteria like *liveliness* and *brightness* to ask for perceptual judgments from the subjects. In trying to build connections between the perceptual qualities and physical measurements of the violin, it may be a necessary step to explore how violinists describe the perceptual properties of the violin. Several studies were thus conducted to study the descriptors that violinists use when evaluating violins.

In [Fritz et al., 2010a], researchers employed a situated and cognitive approach for the violin quality evaluation. Three professional French violinists were invited to assess three violins of different qualities. They were first asked “what is a good violin” and “what is a bad violin”, followed by a playing test evaluating the three violins and a listening test (violins performed by another violinist) together with semi directed interviews. The verbal responses were analyzed and compared between all tasks. It was found that violinists used the same linguistic resources when evaluating the three violins. From semantic analysis and preference ranking, the three violinists were highly consistent. Two different objects were identified in the musicians’ expressions: the violin and the sound. Violinists described more about the interaction between the player and the instrument in the playing test than in the listening test.

In [Fritz et al., 2012a], the diverse verbal descriptions of the distinctive timbres of different violins used by performers were collected and analyzed using multidimensional scaling. Sixty-one descriptors related to the timbre of the violin were collected from 19 violinists (native English speakers) and from ten recently published volumes of “The Strad” magazine. Those descriptors were then arranged by 15 experienced violinists (native English speakers) on a two-dimensional map: words with similar meanings placed close together. The analysis of multidimensional scaling demonstrated consistent use of many words among violinists. The identified dimensions of verbal descriptions were all related to properties that were considered as good or bad in terms of evaluative aspects.

To demonstrate the words that violinists used to describe the violin quality more thoroughly, spontaneous preference descriptions collected during the experiments reported in [Saitis et al., 2012] were subsequently conceptualized by the authors in [Saitis et al., 2017]. The collected free verbal linguistic expressions were categorized according to semantic proximities, and the acoustical interpretation of the semantic categories-descriptors was proposed by the researchers. This is an important step to translate the semantics of violinists' descriptions into hypotheses that link the perceptual judgments to physical characteristics of violins. Eight semantic categories of violin quality concepts emerged: *richness*, *texture*, *resonance*, *projection*, *response*, *clarity*, *balance* and *interest*. From the perspective of musicians, it was found that they not only focus on the sound produced, but also the interaction with the instrument during playing. The items collected and interpreted in this article can be used in future violin evaluation experiments. Then the authors also present a model that explained how the dynamic behavior of a violin correlates to the perceptual quality in the mind of the player. The results can help us understand more about how violinists evaluate violins, and what words are used frequently. By analyzing the violinists' verbal responses during the violin evaluation experiments, these studies formulated a common framework of semantic descriptors used by violinists to describe perceptual aspects of violins.

2.5 Other Factors Affecting Violin Quality

In this section, previous research related to violin strings and the soundpost are briefly reviewed, as these are topics that are relevant to the research reported in later chapters.

2.5.1 Strings

As has been described in Section 2.2.1, the traditional gut strings installed on the violin were gradually replaced by metal wound gut or synthetic (mainly nylon) cores or single steel strand. The acoustic properties of the string [Fletcher and Rossing, 1998], the stick-slip Helmholtz moiton of the bowed string [Schelleng, 1973], and the relationship between them are fairly well understood today even though the correlation between the acoustic properties of the strings and their perceived sound quality and playability on violins has not yet been properly investigated.

Pickering [1985, 1986] carefully measured the physical properties of some violin strings that were widely used. From what he measured, the elasticity (a measure proportional to the Young's modulus) of a steel string is three times greater and seven times greater than a synthetic string and a gut string, respectively. The frequency of a newly installed string drops after its initial tuning. Pickering [1986] measured the time that different types of strings take to stabilize: steel strings take a few minutes, synthetic strings need about 8 hours, while gut strings could require as much as 48 hours.

Firth [1987] measured the inharmonicity of different brands of strings and then correlated them to the preferences of players. The results, however, were surprising: the strings with the lower inharmonicity were ranked low in player preference, but this could probably be attributed to other factors of the strings. Those strings he measured all had a gut core, nylon overwrap and an outer wrap of aluminum or silver. By using a scanning electron microscope, he was also able to study their construction details.

2.5.2 Soundpost

The soundpost (SP) of a violin is an essential component of the instrument. According to luthiers, subtle changes to the soundpost dimensions or position can result in significant variations in the violin sound and playing qualities. The soundpost is typically made of the same wood as the top plate, and it is a cylinder of approximately 0.7 g, 6 mm diameter and a bit longer than 50 mm [Bissinger, 1995]. It provides structural support between the top and back plates and also a means of adjustment in the assembled instruments. As stated by Savart in 1840 [Savart, 1840], the soundpost can help transmit the vibrations from the top plate to the back plate. Through properly interpreted experiments, he also proved that the first acoustical purpose of the soundpost is to introduce asymmetry to the violin.

Jansson et al. [1970], Schelleng [1971], Bissinger [1995] and Gough [2017, 2018] studied the function of the soundpost through comparison between the violin with soundpost and without soundpost. Jansson et al. [1970] employed hologram interferometry to study the resonances of the violin body. They designed an artificial immovable soundpost for observing the interferograms of the plates. A nodal line or a nodal area appeared on the interferograms of the plates around the position of the soundpost when the soundpost was in place and the resonance frequencies increased with a soundpost compared to without a soundpost. The appearance of the modes changed more on the top plate than the back plate when the soundpost was installed. Double exposure holograms on the complete instrument while pressing strings against the fingerboard showed that the maximum deformation of the back plate is at the soundpost. Schelling [1971] approximated the violin body as a closed cigar box and the soundpost as immovable to explain the effect of the soundpost in enhancing the sound radiation. Without the soundpost, the strongest radiating mode is not excited. Also, he explained that the appearance of a new body mode with the soundpost installed depends on the adjacent modes without soundpost that do not have a null at the soundpost position. He then abandoned the assumption of the immovable soundpost and found that the admittance of the contact point of the soundpost and back plate is the smallest compared to the top plate and the ribs, i.e., it is unnecessary to assume all motion of the back plate is ascribed to the soundpost. Bissinger [1995] employed a modal analysis method to test an unvarnished violin. The peaks in the accelerance spectra did not show a substantial shift in frequency with soundpost or without soundpost, and the large peaks usually stayed large. About one-third of the peaks in the no-SP spectrum did not correlate easily to the SP spectrum, and the correlation reliability generally dropped with increasing mode frequency. Using the modal analysis data, he calculated the radiation efficiency of the violin. He observed a very considerable radiation efficiency enhancement of SP over no-SP in the region of 500-800 Hz, in which there are some very important peaks in the response or radiativity curves. Overall, the average radiation efficiency increased by 17% with the soundpost installed. Simulated response curves and Fourier spectra of bowed slide tones of this violin showed that removing the soundpost weakened the frequency response as well as the overall acoustical response from 0 to 2 kHz, and this effect is much more substantial in the frequency range of 400 to 800 Hz. Gough [2017, 2018] studied the function of the soundpost using COMSOL shell structure finite element (FEA) computations. He found that the soundpost and bassbar both can break the symmetry of the empty violin body shell and introduces asymmetric coupled modes in opposite directions. Thus, between them, there is a symmetry-breaking competition. They both influence the tonal balance of the violin over the whole playing range and the intensities of the radiated sound strongly.

Saldner et al. [1996] studied the action of the soundpost by employing a TV-holography technique to visualize the modal patterns of an unvarnished violin in real time, while also measuring the bridge admittance of the violin. They compared the violin without soundpost, with soundpost in normal position and with soundpost 10 mm closer to the centerline. Through the bridge admittance measurements, they found that the magnitude of the B1- peak stays about the same with soundpost in normal position or 10 mm closer toward the centerline; the frequency of the B1- peak increases by 25 Hz (5%) when moving the soundpost closer to the centerline. The magnitude of the B1+ mode however increased considerably when shifting the soundpost closer to the centerline, with the frequency of the B1+ mode remaining about the same. In observing the holographic vibration distributions for the B1- mode, they found a similar frequency shift as in the bridge admittance measurements. Compared to the no-SP condition, they found that the main vibrations in the top plates are shifted to the opposite side of the soundpost. There are small vibrations or a nodal line at the soundpost position. The soundpost makes it possible for the symmetric vibration modes to be excited by the bridge.

Jansson [2004] measured the bridge admittance to compare the violin with soundpost and without soundpost as well. He found that the magnitude of the "bridge hill" (BH) is the highest with soundpost, while without the soundpost, the magnitude of a peak at approximately 550 Hz is the highest. He also explored the effect of the soundpost position on the violin timbre. The soundpost was moved closer toward the bridge or further away from the bridge, and closer to the centerline or towards the nearby f-hole by 5 mm. The BH was attenuated when the soundpost was moved closer to the bridge, and the timbre turned sharper according to the author; the BH was increased when the soundpost was moved away, and the timbre became softer. The magnitude of the B1+ peak increased with the soundpost moved towards the centerline, and the timbre turned darker; the magnitude of the B1+ peak decreased with the soundpost moved towards the nearby f-hole, and the timbre became lighter. However, no formal perceptual evaluation of the violin timbre variation was conducted.

2.6 Research Questions

As mentioned in previous sections, Saitis et al. [2012; 2015] performed a series of experiments to investigate violinists' evaluation process. It was found that violinists were self-consistent while evaluating violins, however, there was significant lack of agreement between different players. Before the formal experiments, Saitis et al. [2012] conducted a pilot study to select instruments for the experiments. They found that the musicians could easily discriminate entry-level Suzuki instruments. Thus, those violins were omitted from consideration as it was felt they would skew the consistency of the results. On hindsight, these authors began to wonder how the musicians could consistently distinguish the Suzuki violins from "good" violins. Are there specific aspects of the Suzuki violins that most violinists might agree make them of lower quality? If the answer is yes, it might be possible to correlate those qualities to acoustical characteristics and physical measurements of the violins. Therefore, the first study in this thesis sought to assess whether the entry-level Suzuki violins would be consistently distinguished from the better quality violins under more controlled conditions and whether there would be agreement regarding the qualities of those instruments that the subjects found less desirable. Bridge admittance measurements were also performed to search for the differences between the two types of violins.

In Section 2.5.1, we mentioned several studies that compared the physical properties of some commercial strings of different brands/materials. However, how different strings can affect the perceptual qualities of the violin has not yet been well studied. Thus, the second study in this thesis investigated the influence of different strings on the violin quality through a perceptual experiment. Three types of strings were employed. Considering the popularity of use among violinists and the prices of the strings, we did not include steel strings (except for the E strings) or gut strings. Instead, only synthetic core with metal wound strings of different prices were used: Kaplan strings costing around $108, Dominant strings about $78, and Pro-Arté strings around $49.

Previous studies on violin soundposts were presented in Section 2.5.2. Most of them focused on physical or acoustical aspects and were concerned with the role of the soundpost (installed vs. removed) or general trends in its positioning. How the soundpost affects the perceptual qualities of the violin, however, has not been fully investigated. Hence, the third study in this thesis focused on this question. In designing a perceptual study to evaluate the influence of the soundpost (as reported in Chapters 5 and 6 of this thesis), several practical constraints had to be addressed. First, it is not possible for a violin to be played under full tension without a soundpost, as it would likely be damaged. Second, it is extremely difficult to specify repeated position changes of a traditional soundpost with sufficient accuracy and speed during a playing experiment. Therefore, the third study in this thesis was designed to investigate correlations between a change in height of the soundpost and variations of the perceived quality of the violin through both playing and listening tests. The availability of an easily adjustable carbon fiber soundpost was crucial to these studies.

Chapter 3

3 Player Evaluation of Performance and Student Violins

3.1 Introduction

It has been a long-standing goal for scientists to correlate the properties of physical structure and specific dynamic behaviour of the violin to its perceptual qualities. Scientists' attempts to quantify the characteristics of "excellent" and "bad" violins through physical measurements alone have been largely inconclusive, in large part because they didn't involve a formal psychoacoustic evaluation process of the violin quality. Thus, in recent years, several scholars have conducted controlled perceptual evaluations of violin qualities. Saitis et al. [2012, 2015] performed a series of experiments to investigate violinists' evaluation process. Fritz et al. [2012b, 2014, 2017] conducted several experiments investigating players' and listeners' preference among new and old violins. Within their results, the lack of agreement between different players in terms of violin preference and quality ratings, however, was significant. It should be noted that violins used in these studies were generally of intermediate-level and higher (only three were valued at less than $10K, with the cheapest at $1.3K).

In the process of selecting instruments for the first study of Saitis et al. [2012], informal tests seemed to show that musicians could easily discriminate entry-level Suzuki instruments from more advanced-level violins. For this reason, the Suzuki violins were excluded from further consideration by those authors. Given the subsequent lack of agreement, however, the results with the entry-level instruments became more intriguing. The study reported in this chapter was designed to investigate whether there would be more agreement among players in comparing entry-level Suzuki instruments to more advanced ones and whether particular distinctive qualities of the less preferred instruments might be distinguishable in bridge admittance measurements.

Detailed materials and methods of the experiment are presented in Section 3.2. Sections 3.3 and 3.4 summarize the findings of the two phases of the experiment separately. Section 3.5 displays and analyzes the bridge admittance measurements of the test violins.

3.2 Materials and Methods

This section describes the details of the experiment. It includes the general design of the experiment, the details about the test violins, controls of the experiment, the characteristics of the participants, and the detailed procedure.

3.2.1 General Design

The goal of this experiment is to examine whether there is agreement on less desirable features of violins among violinists, and whether they agree on what the less desirable features are. The experiment consisted of two phases. The first phase allowed the violinists to rate all violins on a continuous scale from 0 to 5 based on their own preference. After the preference rating, several open questions were given to the subjects to answer in order to determine how different violinists evaluate violins. During the second phase, the subjects were asked to rate each violin on a continuous scale from 0 to 5 for *responsiveness, resonance, clarity, richness,* and *balance*.

3.2.2 Test Instruments

A pool of three performance violins (labeled P1, P2 and P3) and three entry-level violins (labeled S1, S2 and S3) from Schulich School of McGill (SSM) was assembled (see Table 3.1). The performance violins were from a set of higher quality instruments donated to the SSM over the years while the student violins came from a collection of Suzuki violins used by music education students. They were not played on a regular basis, especially the performance violins. While scientific studies [Fritz et al., 2012b, 2014; Saitis et al., 2015] may suggest that this should not influence the individual evaluations, players may argue that this could lower the perceived quality of these instruments. However, it should certainly not influence inter-individual agreement. Two violinists participated in the selection process for the test violins. Of the six violins chosen, it was suggested by the violinists that violin P1 be adjusted. Thus, violin P1 was sent to a luthier, who adjusted the soundpost, bridge and installed new strings before the experiment. The participants in the experiment were given the option to either use a provided shoulder rest (Kun Original model), use their own shoulder rest, or not use a shoulder rest at all.

Table 3.1 Violins used in the experiment along with preference score averaged across subjects (continuous rating scale from 0 to 5; two-sided 95% confidence interval of the mean in square brackets). The score of the most preferred violin (P2) is indicated in bold and the least preferred violin (S1) in italics.

Violin	Origin	Luthier	Year	Estimated Price	Preference score
P1	Unknown	Lorraker	1989	$14.1K	3.37 [2.36, 4.38]
P2	Unknown	Unknown	Unknown	$8K	**3.47 [2.15, 4.78]**
P3	Italy	Nicolas	Unknown	$2.4K	3.06 [1.82, 4.29]
S1	Unknown	Unknown	Unknown	$750	*1.42 [-0.04, 2.88]*
S2	Unknown	Unknown	Unknown	$750	2.34 [1.31, 3.38]
S3	Unknown	Unknown	Unknown	$750	1.71 [0.42, 3.00]

3.2.3 Participants

Nine violinists took part in this experiment (6 females, 3 males; 6 native English speakers, 2 native Chinese speakers and 1 native Catalan speaker; average age = 30 yrs, SD = 14 yrs, range = 20-55 yrs). They had at least 12 years of violin experience (average years of violin playing = 22 yrs, SD = 11 yrs, range = 12-40 years; average years of violin training = 14 years, SD = 4 yrs, range = 8-23 yrs; average hours of violin practice per week = 19 hrs, SD = 13 hrs, range = 0-35 hrs). The estimated prices of their own violins range from $10K to $20K, and they were paid for their participation. Three violinists described themselves as professional violinists. One of the players had a master's degree in music performance (MMus), 3 had bachelor's degrees (BMus, B.A.), 1 had a conservatory degree, and 5 were undergraduate students in music performance. They reported playing various musical styles [classical (100%), folk (22%), baroque (22%), jazz/pop (44%), contemporary (22%) and electronic (11%)] and in various types of ensembles [chamber music (67%), symphonic orchestra (89%), solo (67%), private violin teacher (11%) and electronic/indie/R&B (11%)].

3.2.4 Controls

The possible effect of visual information, such as the style of the violin, the colour of the varnish, identifying marks of the violin, may cause preference biases in the evaluation process. In order to eliminate this possible influence and also ensure the safety of the players and instruments, the subjects were provided dark sunglasses and the light level in the room was significantly reduced.

As in several previous studies [Saitis et al., 2012; Fritz et al., 2012b], we considered the bow to be an extension of the player and asked the subjects to use their own bows. A common bow across all violinists might also trigger a quality debate. The violinists were also asked to bring their own violins with them, in case they wanted to use it as a reference during the tests.

This experiment took place in a diffusive sound space (walls treated with diffusive panels) in order to minimize the effects of room reflections on the direct sound from the instruments. The area of the room was approximately 27 m^2, and the reverberation time was approximately 0.18 s.

3.2.5 Detailed Procedure

This experiment was organized in two phases and lasted around one hour. Subjects were scheduled individually. The experimenter was constantly present in the room for instructing and taking notes for the subjects. Before the experiment, the subjects answered a questionnaire and signed the consent form. Then they were given instructions about the experiment. Before the experiment, the six violins were assigned a letter from A to F randomly, to avoid presentation order effects; the letter was written on a small piece of paper, which was then stuck on the scroll of each violin. The violins were ordered from A to F and placed on a table along with the subject's own violin. During the first phase, the subjects were given up to 25 minutes to play all six violins, and compare and rate the violins from least preferred to most preferred on a continuous scale from 0 to 5. The continuous scale was printed on a sheet of paper, with the numbers 0 to 5 labeled on the scale. Main scale marks were denoted above the six numbers, and nine minor tick marks were denoted between every two numbers, in an even distribution. Above the 0 and 5 graduation lines, phrases of "Least Preferred" and "Most Preferred" were indicated, respectively. Subjects were asked to rate each violin from 0 to 5 by making a vertical line on the scale and label each line with the letter of the violin. While the design of the scale in this way was intended to provide subjects with a clear delineation of the range, on hindsight it may have contributed to some subjects using it more as a system for ranking. Subjects were free to play the instruments in any manner and any order. They were also encouraged to comment out loud when assessing the violins, and the experimenter took notes of the subjects' comments. They were instructed to follow their own strategy imaging that they were choosing violins for themselves at a violin shop. They were allowed to play their own violins whenever it seemed useful. Upon completing the first phase of the experiment, subjects were asked to provide written responses to a set of very general open-ended (in order to avoid confining the answers into pre-existing categories) questions as follows:

A1. *How and based on which criteria did you make your rankings/ratings?*

A2. *Why did you choose the violin ranked as the most-preferred?*

A3. *Why did you choose the violin ranked as the least-preferred?*

A4. *In general, what distinguished the less-preferred violins from the more-preferred violins?*

A5. *Do you have any comments or remarks about the task you were involved in? To what extent was wearing sunglasses disturbing?*

After finishing the first phase of this experiment, subjects were given five criteria for assessment of each violin: *responsiveness, resonance, clarity, richness* and *balance*. These criteria were selected from previous studies [Fritz et al., 2012b, 2014; Saitis et al., 2012, 2017], with the aim of choosing the most common terms covering a diverse range of violin qualities, while also needing to limit the number of criteria to minimize subject fatigue. Subjects were given 5 minutes to evaluate each criterion and rate the six violins on a continuous scale from 0 to 5, which was the same as the preference rating in Phase 1. To ensure all subjects had a common interpretation of the rating scales, each criterion was presented with a descriptive phrase, together with an explanatory text, referring to [Saitis et al., 2012, 2017]:

- **RESPONSIVENESS:** *Responsiveness* describes how fast the violin can respond to different bowing techniques by the violinist, and how easier the violinist can control the playing process and the played sound. Expressions that violinists may use to describe *responsiveness* such as "easy to play", "responsive", "comfortable," "has a broad dynamic range", or "hard to play", "heavy", "slow", etc. The subject might consider the violins from least responsive to most responsive, and rate them on a continuous scale from 0 to 5.

- **RESONANCE:** *Resonance* describes sustain time after bowing has stopped. Violinists may use “powerful”, “open”, “ringing”, “loud”, “responsive” or “muted”, “weak”, “tight”, etc. to describe the violin in terms of resonance. The subject might consider the violins from least resonant to most resonant, and rate them on a continuous scale from 0 to 5.
- **CLARITY:** A sound is described as “clear” when perceived as lacking audible artifacts when played, such as wolf notes, “buzzing”, or a slow buildup of energy during attacks and transients. Violinists may use “clear”, “pure”, “clean” or “scratchy”, “muddy”, “whistles”, etc. to describe the sound in terms of clarity. The subject might consider the violins from least clear to most clear, and rate them on a continuous scale from 0 to 5.
- **RICHNESS:** *Richness* refers to the presence of overtones in the sound, or the perceived number of partial frequencies present in a violin note. Violinists may use “rich”, “(with many) colors”, “(with many) overtones”, “deep”, “full”, “thick”, or “hollow”, “simple”, “inexpressive” etc. to describe the violin sound in terms of *richness*. The subject might consider the violins from least rich to most rich, and rate them on a continuous scale from 0 to 5.
- **BALANCE:** *Balance* refers to the relative similarity of sound or physical response of the violin across notes and strings of the instrument. Violinists may use “even”, “consistent”, “stable”, or “uneven”, “unstable”, etc. to describe the violin sound in terms of balance. The subject might consider the violins from least responsive to most responsive, and rate them on a continuous scale from 0 to 5.

After rating of each criterion, subjects were given a question to answer in written form:

B1. *Do you have specific comments or remarks about the “balance (each criterion)” of the violins? Was there a particular behavior in the violin rated as least balanced or the one rated as most balanced?*

After rating all five criteria, subjects were asked to answer two optional questions:

C1. *Do you have any other further comments or remarks about the violins?*

C2. *Would you like to change the preference ranking after rating these criteria?*

3.3 Detailed Analyses and Results of Phase 1

This section provides the subjects' preference ratings of the violins and verbal responses from Phase 1. First, the comparison of preference ratings of the six violins are presented. Then, the levels of inter-individual consistency in the preference ratings were measured. The analysis approach of inter-individual consistency is the same as that used by [Saitis, 2013], including the calculation of concordance correlation coefficient between each pair of subjects, inter-individual consistency for each subject and a cluster analysis based on the concordance correlations between subjects for potential grouping of the subjects. This section also studies the rating difference between performance violins and student violins. The relationship between the rating difference and subject characteristics were also analyzed.

3.3.1 Overall Preference Ratings of the Violins

The overall preference rating results of the violins by each subject are reported in Table 3.2. The number of times each violin was rated as most preferred and least preferred is shown in Table 3.3. Violin P2 was rated as most preferred the most times, and violin S1 was rated as least preferred the most times. The across-subjects average preference scores are shown in Figure 3.1. Error bars of two-sided 95% confidence interval (CI; all CIs are two-sided 95% intervals through this chapter) of the means are also displayed. The observed mean rating score of violin S1 was markedly below the other violins. To determine whether the preference ratings between the six violins were statistically different, we first conducted a Shapiro-Wilk test to measure the distribution of the preference ratings for each violin by all subjects. The results showed that the preference ratings of violin P2, S1 and S3 were not normally distributed; the preference ratings of the other three violins were normally distributed. We performed a repeated-measures ANOVA to test the difference between the preference ratings of the six violins as this test was considered quite robust with a few violations of normality and the result showed that the mean preference ratings differed statistically between the six violins: $F\ (5, 40) = 2.626$, $p = 0.038$. However, post hoc tests using the Bonferroni correction revealed no statistically different preference ratings between any two violins.

From Table 3.2, we can see that subjects 2, 3, 6, 7, and 8 made ordinal judgments, instead of specifying more precise perceptual distances between the violins. Those different rating strategies employed by different subjects may have an influence on the mean ratings of the violins

(Figure 3.1), i.e., increasing or decreasing the mean ratings of violins. And the rating value for any violin at the same ranking may vary among different subjects, which could be due to subjects considering the scale differently. ANOVA allows the researcher to remove the influence of using different parts of the scale [Lawless and Heymann, 2010].

Table 3.2 Overall preference ratings of the violins by each subject. The most preferred violin and the least preferred violin of each subject are indicated in bold and in italics, respectively.

Subject \ Violin	P1	P2	P3	S1	S2	S3
1	3.32	**4.68**	2.33	3.61	0.73	*0.40*
2	3.00	1.00	4.00	2.00	**5.00**	*0.00*
3	1.00	**5.00**	4.00	*0.00*	3.00	2.00
4	3.03	3.58	4.70	4.15	*2.52*	**4.97**
5	**5.00**	*0.93*	1.49	3.52	2.44	4.00
6	4.00	**5.00**	3.00	*0.00*	2.00	1.00
7	4.00	**5.00**	3.00	*0.00*	2.00	1.00
8	2.00	4.00	**5.00**	*0.00*	3.00	1.00
9	**5.00**	2.00	0.00	*-0.50*[a]	0.41	1.00

a: Subject 9 wrote the violin label (one letter of A to F) of S1 outside the scale, so a negative rating appeared. This appearance occurred several times in Phase 2 ratings of this subject.

Table 3.3 Number of times each violin was rated as most preferred and least preferred.

Violin	P1	P2	P3	S1	S2	S3
Times most preferred	2	4	1	0	1	1
Times least preferred	0	1	0	5	1	2

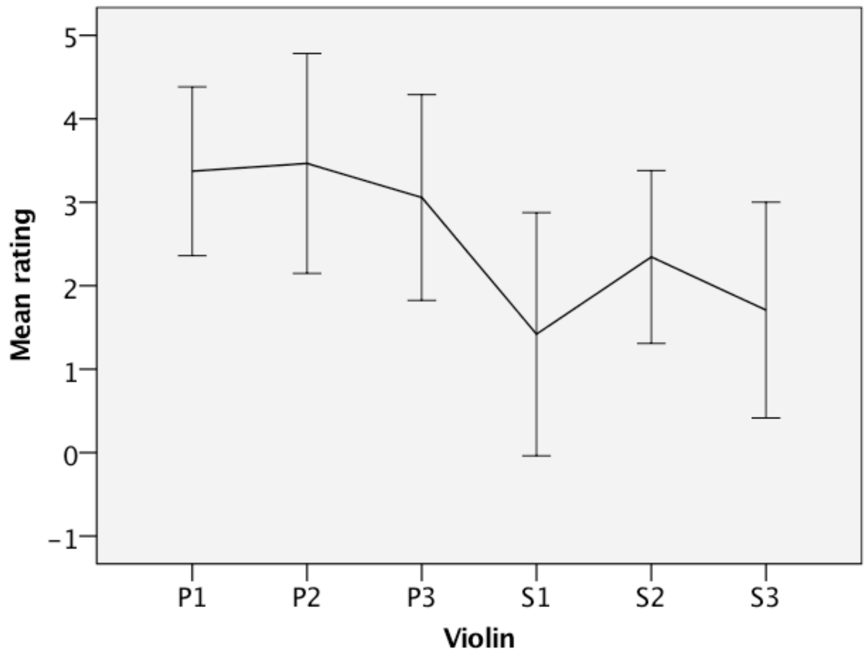


Figure 3.1 Across-subjects average of the overall preference score for each violin (error-bar = 95% confidence interval of the mean).

3.3.2 Concordance Correlation between Subjects

Inter-individual consistency was measured as the concordance correlation between preference ratings of different participants. The concordance correlation coefficient is a special case of Pearson's correlation coefficient, which is introduced by Lin [1989]. It is defined as:

$$\rho_c(A,B) = \frac{2rs_A s_B}{s_A^2 + s_B^2 + (\bar{A} - \bar{B})^2}$$

where the pair of samples (A, B) are independently selected from a bivariate population. $\bar{A}$ and $\bar{B}$ are their means, s_A^2 and s_B^2 are their variances and r is the Pearson product-moment correlation coefficient. Lin's ρ_c measures departures from the equality lines with slopes $\pm 45^{\circ}$: $\rho_c(A,B)$ = 1 and -1 mean that $A = B$ and $A = -B$, respectively, and $\rho_c(A,B)$ = 0 indicates there's no association between A and B. ρ_c does not assume linear relationships, whereas which is the premise of Pearson's correlation coefficient.

Figure 3.2 displays the histograms for all the ρ_c computed between the preference ratings of every two subjects. Subjects 6 and 7 showed perfect consistency: ρ_c (6, 7) = 1. The second highest concordance correlation coefficient was between subject 3 and 8: 0.886, which was significantly higher than 0 with a two-tailed 95% CI of [0.398, 0.983]. The calculation of the confidence interval of ρ_c can be found in Lin [1989]. The third or fourth highest concordance correlation coefficient was between subject 6 or 7 and subject 8: 0.714, which was close enough to be significantly higher than 0, as the corresponding two-tailed 95% CI was [-0.084, 0.954]. Subjects 6, 7 and 8 had described themselves as professional violinists. Subject 3 was a 4th year undergraduate student in music performance, who later continued to pursue a graduate degree. The other concordance correlations between subjects were not significantly higher than 0. Overall, the mean concordance correlation was very low: 0.115, 95% CI was [-0.031, 0.260], which was not significantly higher than 0: $t(35) = 1.602$, $p = 0.118$. Generally, large inter-individual variation in the preference for violins between all subjects existed, but the inter-individual concordance correlations among professional musicians were considerably high.

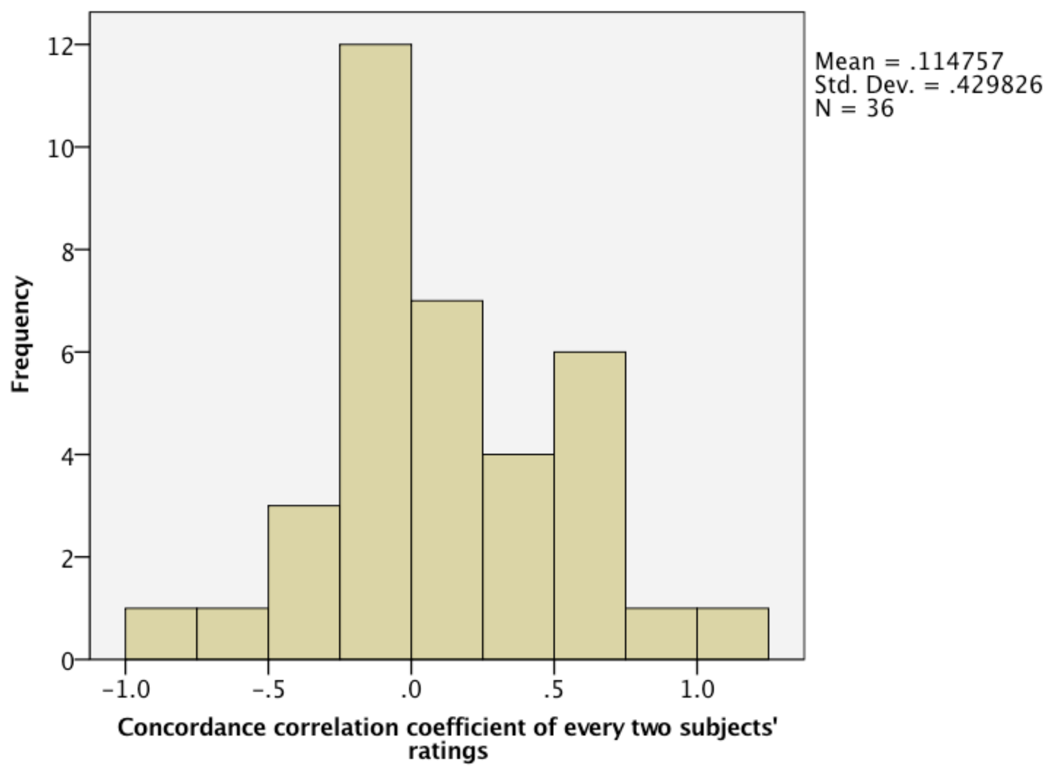


Figure 3.2 Distribution of all 36 concordance correlation coefficients of every two subjects' ratings.

Further, we computed the inter-individual consistency for each subject, in order to examine the relationship between the preference ratings by one subject and the other subjects. The inter-individual consistency was defined as the mean of the ρ_c between the preference ratings of this subject and those of the other eight subjects [Saitis, 2013]. It is shown in Figure 3.3. According to the definition of the concordance correlation, the inter-individual consistency for each subject varies from -1 to 1 as well. The inter-individual consistency for subject 2 was small, and subject 4 and 5 had negative inter-individual consistencies. By looking at their preference rankings in Table 3.4, we could find that the most preferred violin of subject 2 was a student violin, which was different from the other subjects except subject 4. Subjects 4 and 5 included two student violins among their three most preferred instruments, which was different from the other subjects.

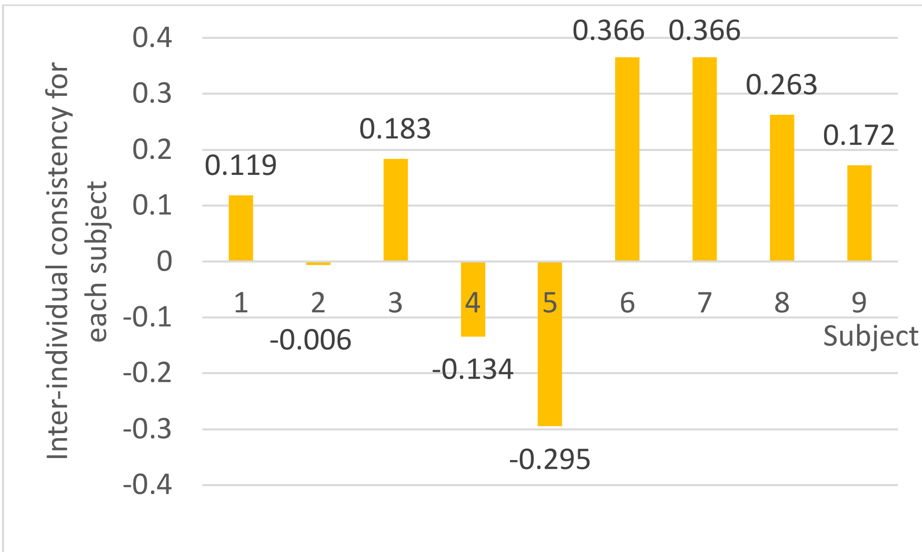


Figure 3.3 Inter-individual consistency for each subject. 1 corresponds to perfect consistency, 0 corresponds to no consistency and -1 corresponds to perfect anti-consistency.

A clustering method (hierarchical cluster analysis, average linkage) was used to detect potential grouping of agreement in the preference ratings (see Figure 3.4), in order to examine whether subjects with similar backgrounds made similar preference ratings. The y-axis represents the distance calculated based on the concordance correlation coefficient. For the first stage clusters (formed by subjects directly, e.g., subject 6 and 7), the y-axis represents the distance between the subjects, which is defined by 1 minus the corresponding concordance correlation coefficient. For the groups formed by lower stage clusters, the y-axis represents the average distance between all pairs of subjects in any two clusters. For example, the distance between the cluster {6, 7} and the cluster {3, 8} is the average distance of subjects 6 and 3, subjects 7 and 3, subjects 6 and 8, and subjects 7 and 8. The solid lines that connect subjects or clusters imply that there are significant positive concordance correlations between every two subjects in the cluster. We can see that the preference rating distance between the subjects 6, 7, 3 and 8 are small, which confirms the high inter-individual consistency among professional musicians (subjects 6, 7 and 8) as well as subject 3. Table 3.4 shows the violin preference profiles corresponding to the resulting clusters based on a cut-off value of 0.2.

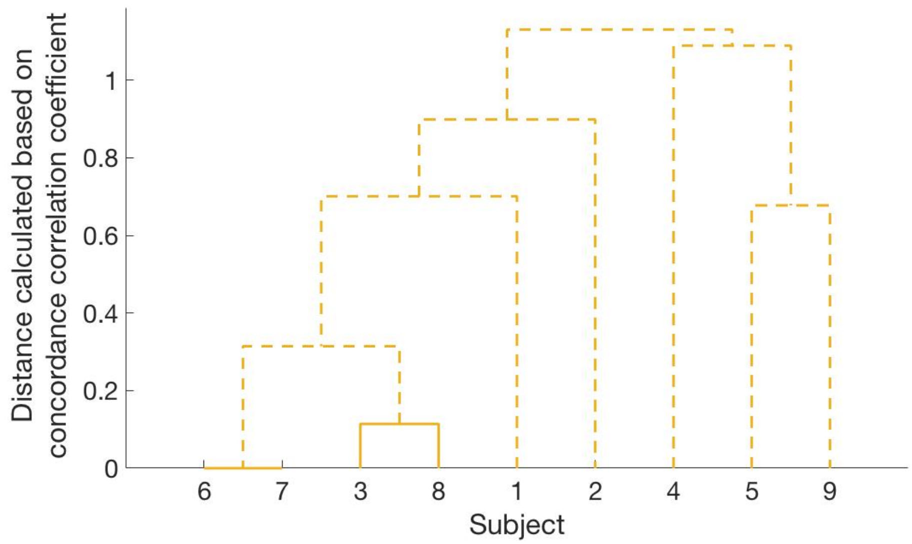


Figure 3.4 Hierarchical cluster analysis on subject-specific preference profiles. The solid or dashed lines that connect subjects and clusters indicate their respective correlations.

Table 3.4 Violin preference profile for each cluster of subjects. The dashes between two violins indicate reversed orderings (ex., for cluster {3,8}, P1, S3 for subject 3, but S3, P1 for subject 8).

Cluster {Subjects}	Least → most preferred					
{6,7}	S1	S3	S2	P3	P1	P2
{3,8}	S1	P1-S3		S2	P3-P2	
1	S3	S2	P3	P1	S1	P2
2	S3	P2	S1	P1	P3	S2
4	S2	P1	P2	S1	P3	S3
5	P2	P3	S2	S1	S3	P1
9	S1	P3	S2	S3	P2	P1

3.3.3 Preference Ratings of Performance Violins and Student Violins

The ratings of performance violins and student violins are compared in this section. Table 3.5 displays the mean preference ratings of performance violins, student violins and the rating difference between the two types of violins by each subject. The rating difference was defined by the mean preference rating of performance violins minus the mean preference rating of student violins. On average, the mean preference rating of performance violins by each subject was 3.30, 95% CI = [2.80, 3.80]; the mean preference rating of student violins by each subject was 1.82, 95% CI = [0.93, 2.71]; the mean rating difference between the performance and student violins by each subject was 1.47, 95% CI = [0.42, 2.53]. To test whether the mean preference ratings between the performance and student violins by each subject were statistically different, we first conducted a Shapiro-Wilk test on the distribution of the rating difference between the mean preference ratings of the performance and student violins by each subject. The result showed that the rating differences were normally distributed, thus the paired-samples *t*-test was performed. According to the paired-samples t-test, the mean preference ratings of the performance and student violins by each subject were significantly different: $t(8) = 3.221$, $p = 0.012$. This is important and interesting, because if you look at the individual violin (last column of Table 3.1 and Figure 3.1), there is no significant differences between any two violins. But if you group them in performance/student type, then there is a significant difference between the two groups.

Table 3.5 Mean preference ratings of performance violins, student violins and the rating difference between the two types of violins by each subject.

Subject	1	2	3	4	5	6	7	8	9	Mean
Performance violins	3.43	2.67	3. 33	3.77	2.47	4	4	3.67	2.33	3.30
Student violins	1.58	2.33	1. 77	3.88	3.32	1	1	1.33	0.30	1.82
Rating difference	1.86	0. 33	1.67	-0.11	-0.85	3	3	2.33	2.03	1.47

3.3.4 Influence of Participant Characteristics on the Rating Difference between Performance and Student Violins

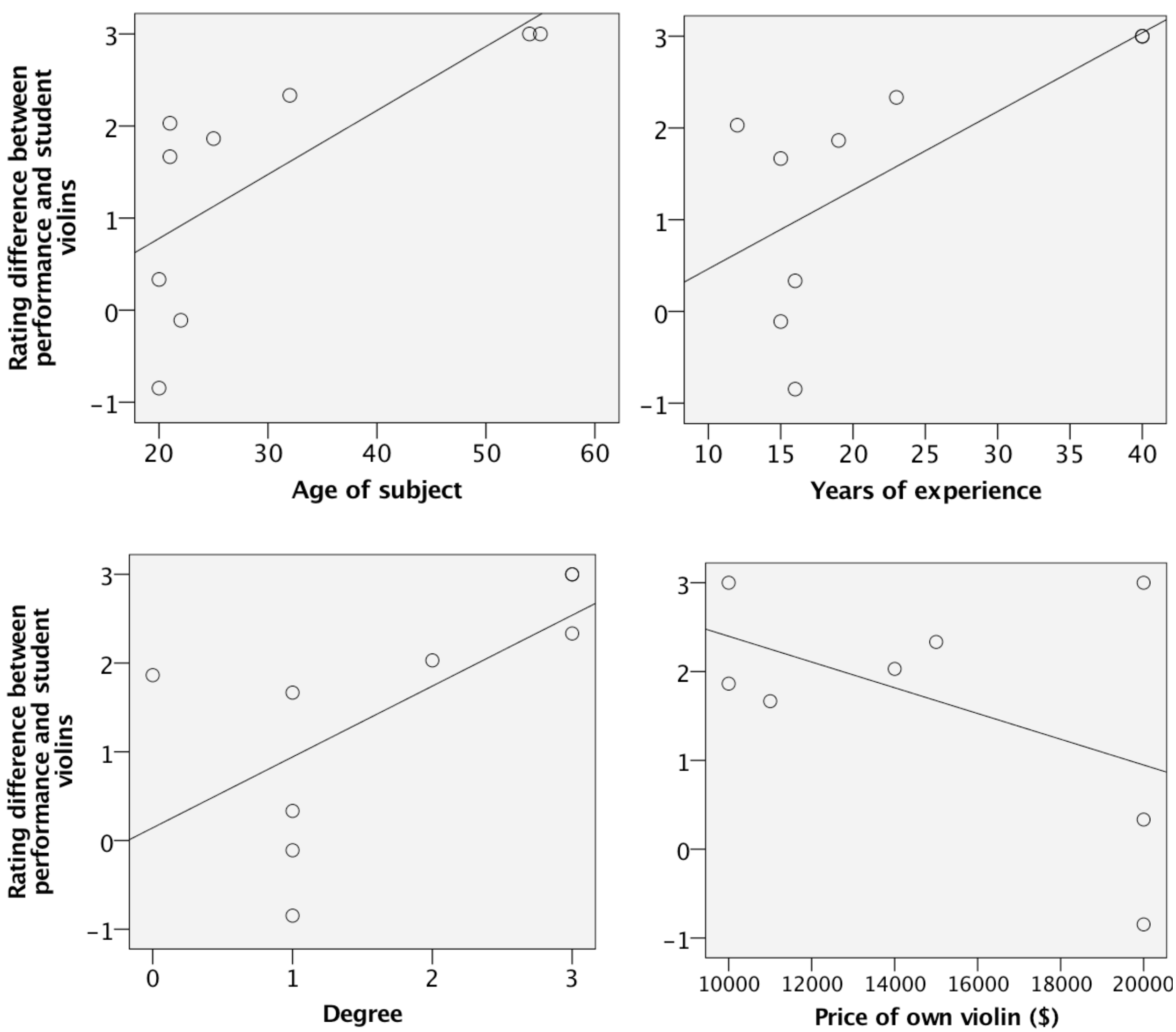
The association between preference rating difference (performance violins and student violins) on the one hand, and the self-reported age, degree in music performance, years of violin experience, weekly hours of violin practice and price of the owned violin, on the other was assessed. This analysis was carried out by calculating the Spearman rank correlation coefficient ρ_S between preference rating difference and participant characteristics, as shown in Table 3.6. To further investigate the significance of the correlations, we also performed bootstrapping analyses, and the resulting 95% confidence intervals of the Spearman rank correlation coefficients ρ_S are reported in Table 3.6 as well. We can see that the correlations between age, degree and rating difference were significant at 0.05 levels. Participants who were older, those who were professional musicians, and/or with a higher educational degree in music performance rated performance violins much higher than student violins. Scatter plots of individual preference rating difference (performance violins and student violins) and self-reported participant characteristics are shown in Figure 3.5. We can see that we lacked participants whose ages were between 35 to 50 and the two participants with ages between 50 and 60 were two of the three professional musicians. The significant correlation between the rating difference and the age could be biased by the small sample. For the correlation between the rating difference and "degree", we have participants for each "degree", thus the significant correlation between the rating difference and "degree" would be more reliable.

Table 3.6 Spearman rank correlation between rating difference (performance violins and student violins) and self-reported participant characteristics, along with the 2-tailed significance value and 95% confidence interval calculated by bootstrapping.

Participant characteristics	Age	Degree[b]	Years of experience	Practice (hours/week)	Price of own violin ($)
ρ_S	0.835	0.766	0.636	-0.553	-0.303
p	**0.005**	**0.016**	0.066	0.123	0.466
95% confidence	[0.279,	[0.092,	[-0.193,	[-0.927,	[-0.904,

interval with bootstrapping	0.996]	0.973]	0.995]	0.107]	0.465]

b: Conservatory degree corresponds to 0; current undergraduate student corresponds to 1, current graduate student corresponds to 2 and professional musician corresponds to 3.



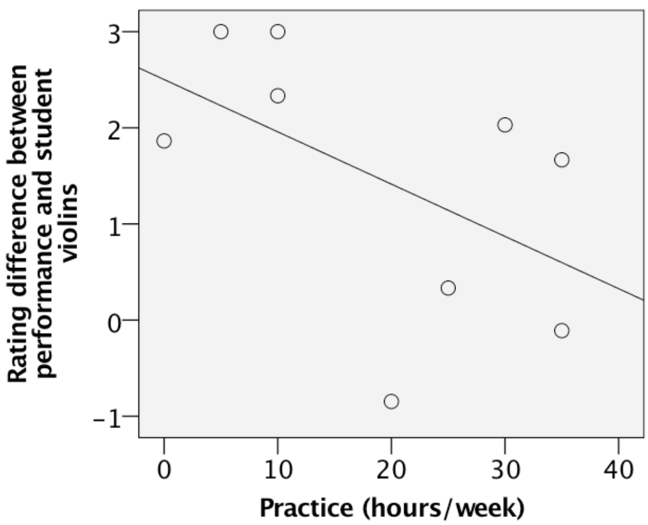


Figure 3.5 Scatter plots of individual preference rating difference (performance violins and student violins) and self-reported participant characteristics. Solid lines show linear fitting to the data.

3.3.5 Verbal Descriptions of Violin Preference

Verbal responses for questionnaire A at the end of Phase 1 collected from the subjects are shown in Table 3.7, 3.8 and 3.9. Those answers were classified into different categories: resonance, *richness*, clarity, etc. and highlighted in different colours, with each category corresponding to one colour. The categorization was performed in a similar manner to [Saitis et al., 2017] though phrases that didn't belong to any of the categories were classified into new classes that were created by the author. As in [Saitis et al., 2017], same phrases can be classified into different categories in the current thesis according to the context of the subjects' verbal response. In the first row of each table, the different categories were listed in the order of number of subjects (denoted in parentheses) who mentioned the phrases that belonged to each category (if one subject mentioned several phrases for each question that belonged to one category, that was counted as one). In response to the question of why they ranked a particular violin as "most-preferred", the subjects indicated the following considerations in order of importance: *richness*, *texture*, *interest*, *response*, *clarity*, *resonance*, *balance*, *projection*, and *craft*. In response to the question of why they ranked a particular violin as "least-preferred", the subjects' answers suggested that they were most concerned about *resonance*, *response,* followed by *clarity*, *interest*, *texture*, *richness*, and *craft*.

Table 3.7 Verbal collections of questions A2 and A3. Subjects' descriptions were classified into different categories, which are highlighted in different colours. Subject number is indicated as "s1, s2, s3…". The colour scheme is shown in the first row.

Violin	Most preferred reasons: Richness (4). Texture (4). Interest (4). Response (4). Clarity (3). Resonance (3). Balance (2). Projection (1). Craft (1).	Least preferred reasons: Resonance (8). Response (3). Clarity (3). Interest (2). Texture (1). Richness (1). Craft (1).
P1	s5: Easiest to play. Required little effort to pull out a rich, pure sound. Had honest sound. Without nasally sound. s9: Weight (usually light). Sound. Neck (comfortable neck)	
P2	s1: Very good bass. E string was bright, too metallic. Well balanced. Rich and warm tone. Other violins were muted or too metallic. s3: Preferred mellow, darker tones. Sound was evenly balanced. Most brilliant sound on the E string. s6: Good at low register and also broad at high register. s7: Rich, good expression.	s5: Difficult to play, hard to play into the string. Muffled. Less concentrated. Less resonance.
P3	s8: Most pleasing/resonant sound. Easier to play.	

S1		s3: Sound was very boxy – no nuance or expression. s6: Too narrow dynamic range. Didn't vibrate. Responsiveness was not bad. s7: Bad resonance. Closed sound. Not thick in lower register, not bright in higher register. s8: Least resonant. 'Tinny' quality. s9: Fat neck. Bad setup. Closed sound.
S2	s2: Projects well. Bright, clean. Ease of playing across the strings. Effective harmonics and bounce worked well.	s4: Difficult to produce a pure sound that rang nicely.
S3	s4: Most direct. G string was easy to play with the exception of a wolf on C.	s1: Cranky sound. Too weak. Annoying, metallic component. A bit muted. s2: Very resonant but the sound quality suffered. Lots of buzzing.

Tables 3.8 and 3.9 compile the answers of A1 "How and based on which criteria did you make your rankings/ratings?" and A4 "In general, what distinguished the less-preferred violins from the more-preferred violins?" The answers to these two questions were similar in many respects, except that the subjects didn't mention *balance* and *projection* in response to question A4. The violinists thought that the less-preferred violins were distinguished from the more-preferred violins in *resonance*, *response*, *interest*, *clarity*, *richness*, *texture* and *craft*. And they

valued *resonance*, *response*, *balance*, *projection*, *richness*, *texture*, *interest*, *clarity* and *craft* when they ranked and rated violins in the preference evaluation.

We asked violinists to evaluate the violins according to five criteria in Phase 2 (*responsiveness*, *resonance*, *clarity*, *richness* and *balance*). Compared to the criteria violinists mentioned, we didn't include *projection*, *texture*, *interest* and *craft*. *Projection* generally needs to be evaluated from a distance (e.g., Fritz et al., 2017), thus it would need other players or audience to help with the evaluation. We felt that *interest* and *texture* were too general to be useful as quality descriptors. *Craft* may affect how the violinist feels while playing, as well as how the violin reacts to the player. Subjects would likely take it into consideration while evaluating criteria like *responsiveness*, *resonance* or *balance*.

At the end of Phase 1, the subjects were asked to provide comments on the evaluation task and their feeling about wearing sunglasses during the process. Two of the nine subjects felt uncomfortable wearing the sunglasses, one of the subjects thought that wearing sunglasses took away from the intimacy of relationship with the violin but concentrated on the sound, whereas the other subjects thought doing so was fun and that it did not bias their evaluation of violins.

Table 3.8 Verbal collections of question A1. Subjects' answers were classified into different categories, which are highlighted in different colours. The colour scheme is shown in the first row.

Most preferred violin	Subject	How and based on which criteria did you make your rankings/ratings? Resonance (6). Response (5). Balance (3). Projection (2). Richness (2). Texture (2). Interest (1). Clarity (1). Craft (1).
P1	5	Resonance. Core of sound/purity/clarity (vs. muffled)
	9	Projection. Sound color (nasally vs. chocolately). Overtones (ringing-well built violin; won't ringing – not well structured to resonant)
P2	1	Balanced sound across strings. Deep bass, tone not too muted nor too metallic. Valued loudness.

	3	Easy to play. Even quality of sound across the strings.
	6	Dynamic range. Responsiveness. Broad range at higher register. Loose sound.
	7	Balance and even sound. Long lasting of the resonance. The flavor of the sound. Wood dried enough.
P3	8	Resonance. Sympathetic vibration. Playability.
S2	2	String crossing. Harmonics. Ease of playing. Projection. Bow bounce test.
S3	4	1st position scale on G and E string. Sound production.

Table 3.9 Verbal collections of question A4. Subjects' answers were classified into different categories, which are highlighted in different colours. The colour scheme is shown in the first row.

Least preferred violin	Subject	In general, what distinguished the less-preferred violins from the more-preferred violins? Resonance (5). Response (3). Interest (3). Clarity (2). Richness (1). Texture (1). Craft (1).
P2	5	Playability. Hardest to draw a simple, clear sound out. Less natural. More forceful. More work with less reward in the sound.
S1	3	Sound not rich. Took more effort to create a good sound. Lack of nuance in the sound-no character.
	6	Loose and broad of the sound.
	7	Bad resonance. Closed sound. Not thick enough in lower register; not bright enough in higher register.

	8	Ring of open strings. Resonance of ringing tones.
	9	Sound color - touch your heart.
S2	4	The sound direction and focus. How easy the violin played.
S3	1	Either weak, metallic (opposed to warm), and/or muted (opposed to rich and bright).
	2	'Tiny' sound. Bridge and string weights are not consistent.

3.3.6 Conclusions of Phase 1

The results of Phase 1 of this experiment showed that the mean preference ratings differed statistically between the six violins, even though post hoc tests revealed no statistically different preference ratings between any two violins. More importantly, it was found that performance violins were on average rated significantly higher than student violins in terms of preference. And it was found that the subjects who were professional musicians, and/or with higher educational degrees in music performance rated performance violins much higher than student violins.

A large amount of variation in the inter-individual consistency of the preference ratings of the violins existed, but three professional musicians highly agreed with each other in this experiment.

From the verbal collections, it was found that the violinists considered *resonance*, *response*, *balance*, *projection*, *richness*, *texture*, *interest*, *clarity* and *craft* when evaluating violins.

3.4 Detailed Analyses and Results of Phase 2

In this section, the results of Phase 2 are analyzed. The analysis was conducted with respect to the following aspects. First, the comparison of the attribute criteria ratings between the six violins and between performance violins and student violins were performed; the corresponding influence of participant characteristics on the rating difference between performance and student violins was also explored; a cluster analysis based on the mean preference and each criterion ratings of performance violins and student violins by each subject was conducted. Second, inter-individual concordance correlation coefficients for each attribute ratings were calculated for each pair of subjects; a cluster analysis was then performed based on the concordance correlations between subjects. Third, the relationship between preference and criteria ratings was examined through regression, partial correlation calculation and criteria ratings comparisons between violins at each rank of preference. Finally, the verbal descriptions of each attribute by subjects were summarized.

3.4.1 Criteria Ratings

Across-subjects average ratings on specific criterion of each violin are shown in Figure 3.6. For each criterion, we tested the statistic difference between the six violins. First, a Shapiro-Wilk test was performed to measure the distribution of each criteria ratings for each violin. The results showed that the ratings of every criterion for violin S1 was not normally distributed. As well, the *responsiveness* ratings for violin S2, *clarity* ratings for violin P1, and *balance* ratings for violin S3 were not normally distributed. We performed a repeated-measures ANOVA testing the differences between the six violins for each criterion rating. The results of the statistical tests are shown in Table 3.10, with only *richness* and *balance* ratings of the six violins having statistically significant differences: *richness*, $F(5, 40) = 4.233$, $p = 0.004$ and post hoc tests using the Bonferroni correction revealed no statistically different *richness* ratings between any two violins; *balance*, $F(5, 40) = 3.31$, $p = 0.014$, and post hoc tests with Bonferroni correction showed that violin P3 was rated significantly more balanced than violin S1 ($p = 0.033$).

Table 3.10 Test of the statistic differences of the ratings of each criterion between the six violins.

Evaluation term	*F* (5, 40 *df*)	*p*
Responsiveness	0.943	0.464
Resonance	1.890	0.118
Clarity	1.709	0.155
Richness	4.233	**0.004**
Balance	3.310	**0.014**

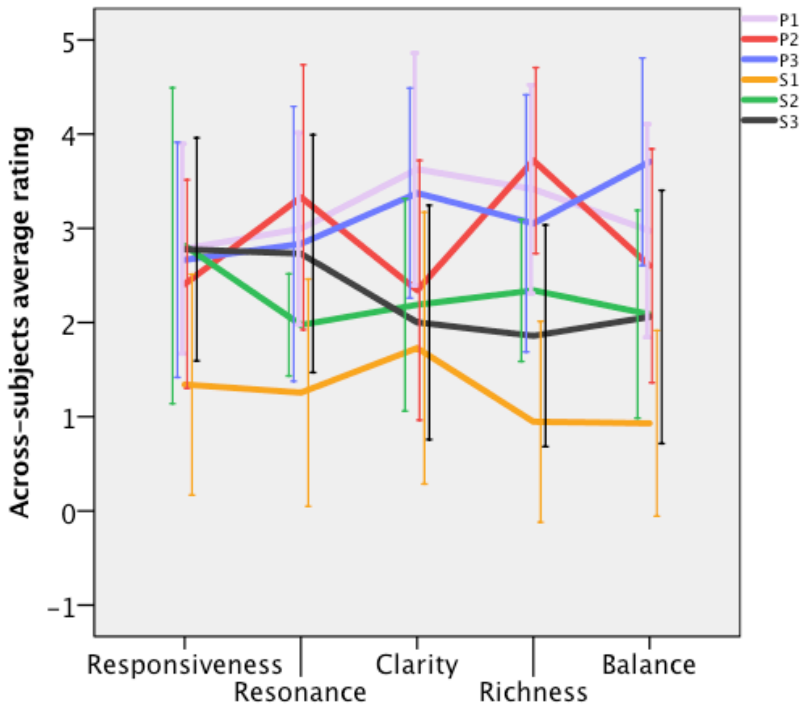


Figure 3.6 Across-subjects average ratings on specific criterion of each violin (error-bar = 95% confidence interval of the mean).

3.4.2 Criteria Rating Differences between Performance Violins and Student Violins

In Section 3.3.3, for each subject, we computed the mean preference ratings of performance violins, student violins and the rating difference between the two types of violins. Similarly, in this section, for each subject, we computed the mean rating for each criterion of performance violins, student violins and the rating difference between the two types of violins. The across-subjects average attributes ratings of performance violins and student violins are displayed in Figure 3.7 with two-tailed 95% confidence intervals, respectively. The confidence intervals were relatively large, especially for student violins.

To test whether the mean criteria ratings between the performance and student violins by each subject were statistically different, we first performed Shapiro-Wilk tests on the distribution of each criterion rating difference between the mean ratings of performance and student violins by each subject. The results showed that for each criterion, the rating differences were normally distributed, therefore we conducted paired-samples t tests. The results are shown in Table 3.11. Based on the paired-samples *t*-tests, the mean ratings of performance and student violins by each subject were significantly different for *richness* [$t(8) = -3.739$, $p = 0.006$], and to a lesser extent for *balance* [$t(8) = -3.035$, $p = 0.016$] and *clarity* [$t(8) = -2.530$, $p = 0.035$], while not significant for *responsiveness* and *resonance*.

Table 3.11 Test of the statistic differences of the ratings of each criterion between the performance and student violins.

Evaluation term	*t* (8 *df*)	*p*
Responsiveness	3.466	0.431
Resonance	6.752	0.092
Clarity	6.778	**0.035**
Richness	14.016	**0.006**
Balance	12.111	**0.016**

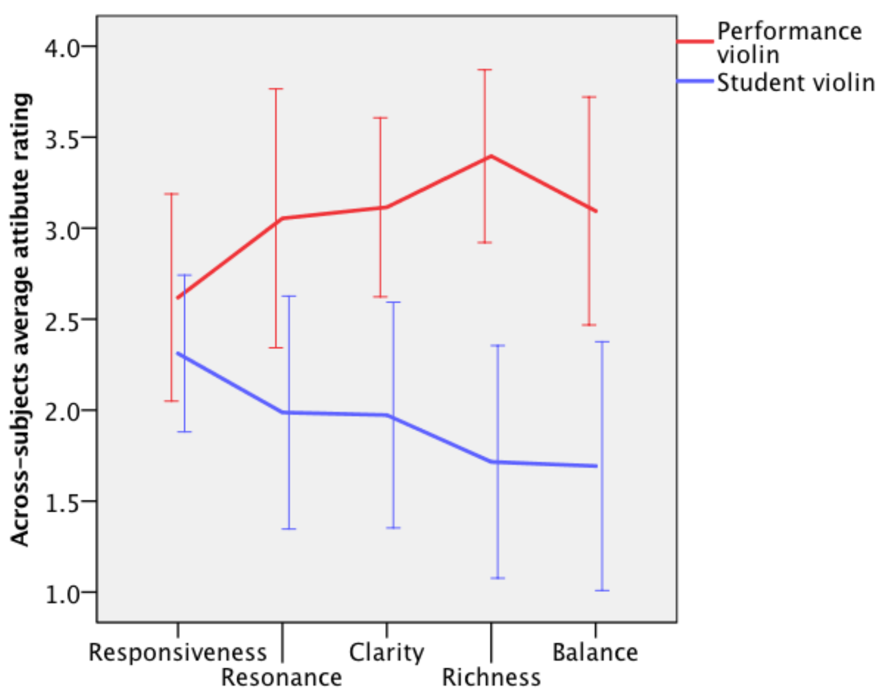


Figure 3.7 Across-subjects average attributes ratings of performance violins and student violins (error-bar = 95% confidence interval of the mean).

3.4.3 Influence of Participant Characteristics on Criteria Rating Differences between Performance and Student Violins

The association between each criterion rating difference (performance violins and student violins) on the one hand, and the self-reported age, degree in music performance, years of violin playing experience, weekly hours of violin practice and price of the owned violin, on the other was assessed. This analysis was carried out by calculating the Spearman rank correlation ρ_S between criterion rating difference and participant characteristics, and the results are reported in Table 3.11. As in Phase 1, we also performed bootstrapping analyses to further examine the significance of the correlations, and the resulting 95% confidence intervals of the Spearman rank correlation coefficients ρ_S are reported in Table 3.11. The correlations between degree on the one hand, and *resonance*, *clarity*, *richness* rating difference on the other were significant at 0.01 levels. The correlations between age and *resonance*, and age and *clarity* rating difference were significant at 0.05 levels. The subjects who were professional musicians, and/or with higher educational degrees in music performance rated performance violins much higher than student violins in *resonance*, *clarity* and *richness*. To a lesser extent, older participants rated performance violins much higher than student violins in *resonance* and *clarity*. As has been noticed in Section 3.3.4, we lacked participants whose ages were between 35 to 50, which makes the correlation between age and the rating differences between performance and student violins in *resonance* and *clarity* not so reliable.

A hierarchical cluster analysis was conducted to detect similarities in different subjects' criteria ratings of performance and student violins. For each subject, the means of preference and each attribute criterion ratings for each type of violins were calculated. Subjects were then grouped according to these averages to reveal concordance among their responses, which were then plotted on the dendrogram (see Figure 3.8). The solid lines that connect subjects and clusters indicate that there are significant positive concordance correlations between every two subjects in the cluster. As the dendrogram reveals, the three professional musicians (subject 6, 7 and 8) were grouped close together. These results indicate that professional musicians rated similarly for the two types of violins in all rating scales, including preference.

Table 3.11 Spearman rank correlation between each criterion rating difference (performance violins and student violins) and self-reported participant characteristics, along with the 2-tailed significance value and 95% confidence interval calculated by bootstrapping.

Participant characteristics		Responsiveness	Resonance	Clarity	Richness	Balance
Degree	ρ_S	-0.168	0.821	0.810	0.866	0.393
	p	0.665	**0.007**	**0.008**	**0.003**	0.295
	95% CI	[-0.853, 0.466]	[0.349, 1]	[0.229, 0.985]	[0.523, 1]	[-0.471, 0.953]
Age	ρ_S	0.067	0.769	0.726	0.165	0.212
	p	0.864	**0.015**	**0.027**	0.078	0.584
	95% CI	[-0.737, 0.895]	[0.158, 0.973]	[0.045, 0.996]	[-0.347, 0.929]	[-0.730, 0.862]

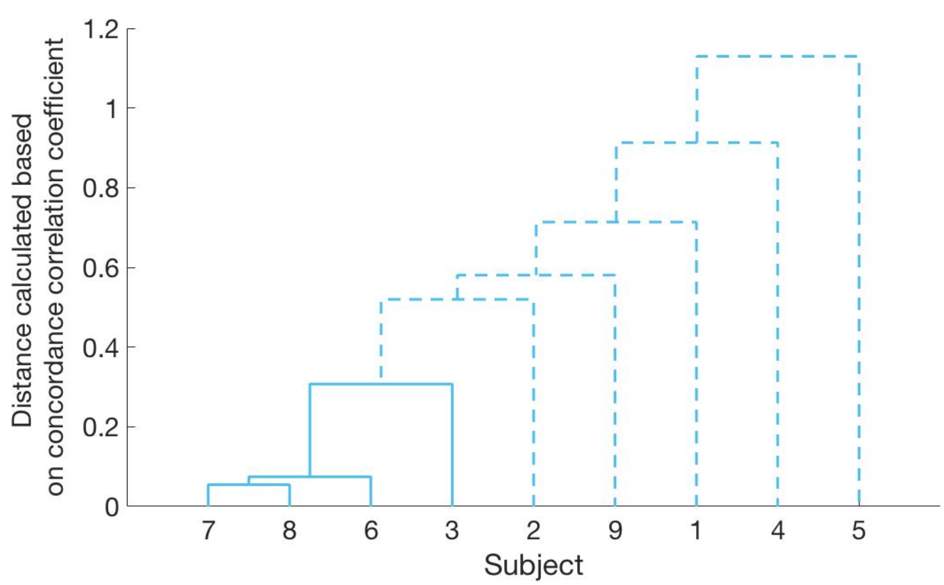


Figure 3.8 Hierarchical cluster analysis of subjects using average ratings of performance and student violins (see text for clarification). The solid and dashed lines that connect subjects and clusters indicate their respective correlations.

3.4.4 Concordance Correlation between Subjects

In Section 3.3.2, we analyzed the concordance correlation between subjects about the preference ratings of the six violins. We calculated the concordance correlation coefficient ρ_c between the preference ratings of every two subjects, and displayed the corresponding histograms in Figure 3.2. On average, the concordance correlation coefficient of the preference ratings between subjects was very low: 0.115 with 95% CI [-0.031, 0.260], which was not significantly higher than 0. Similarly, we calculated the concordance correlation coefficient ρ_c between the ratings of each criterion of every two subjects, and computed the mean concordance correlation coefficient. Figure 3.9 displays the mean concordance correlation coefficient with 95% CI for each criterion. From the figure, we can see that *richness* ratings had the highest mean concordance correlation coefficient: 0.233, 95% CI = [0.075, 0.392], then *balance*: 0.166, 95% CI = [0.047, 0.286], which were significantly higher than 0 [$t(35) \geq 2.835$, $p \leq 0.008$]. While the mean concordance correlation coefficient of the other three criteria ratings (*responsiveness*, *resonance* and *clarity*) were much lower, average ρ_c = -0.021, 0.057 and 0.066 respectively, and were not significantly higher than 0 [$t(35) \leq 0.929$, $p \geq 0.359$]. In [Saitis, 2013], *richness* also had the highest consistency between individuals with an average ρ_c = 0.068. The concordance coefficient of *richness* in this experiment was much higher, but the confidence interval was very broad.

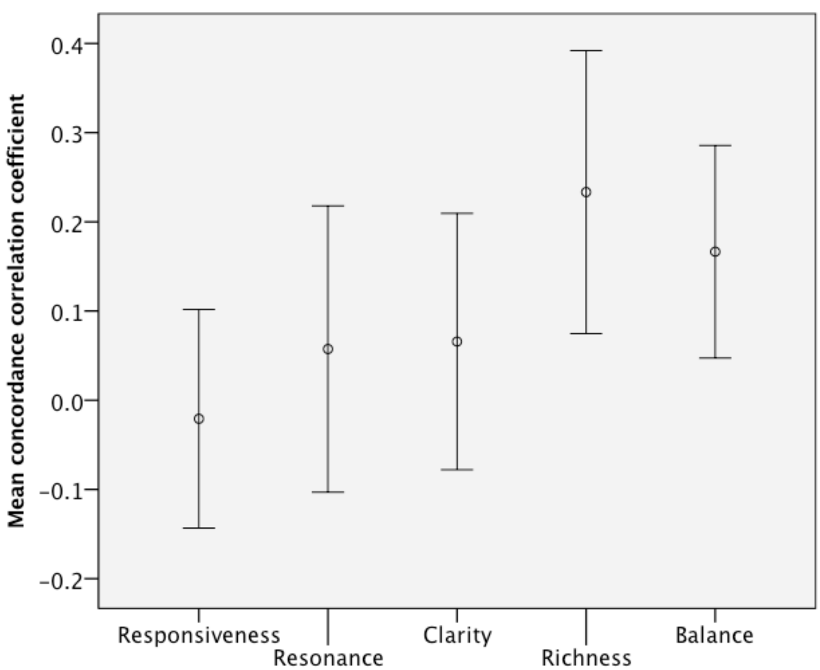
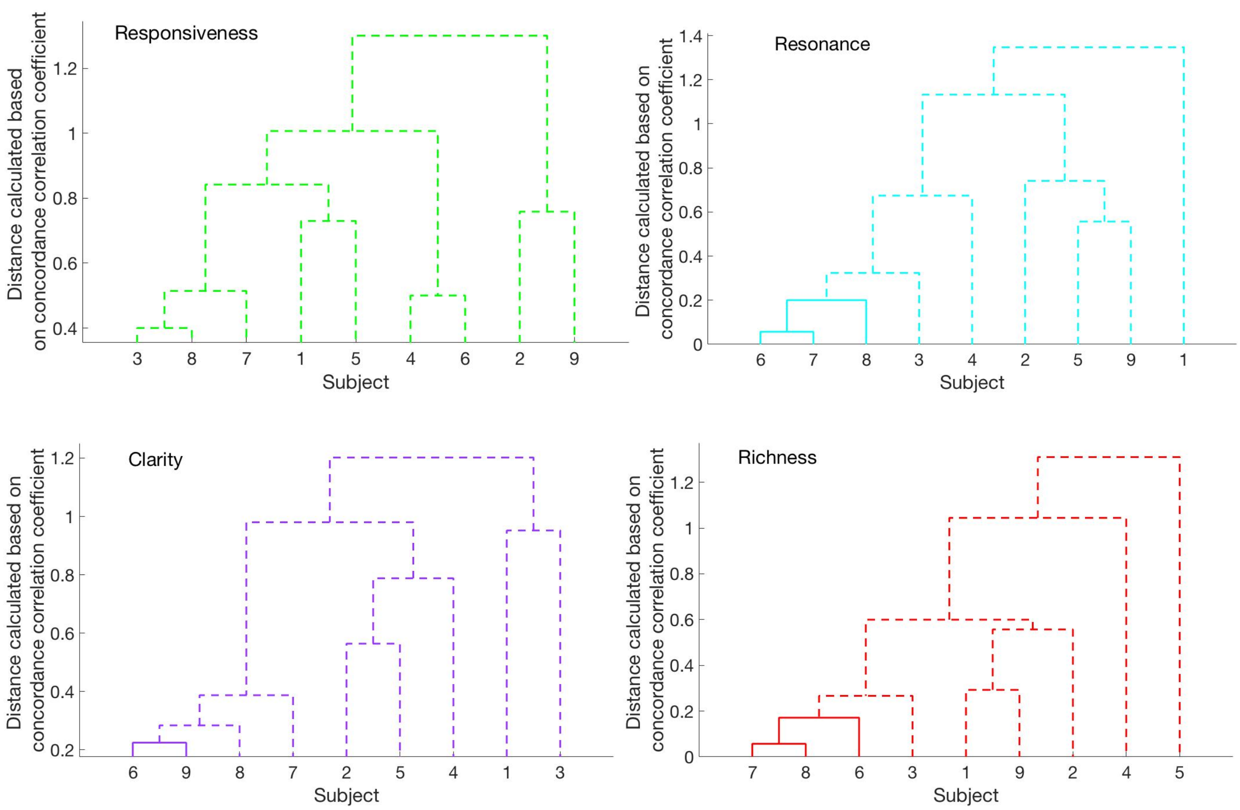


Figure 3.9 Mean concordance correlation coefficient for each attribute criterion rating (error-bar = 95% confidence interval of the mean).

To further investigate inter-individual consistency, a cluster method (hierarchical cluster analysis, average linkage) was employed to detect potential grouping of agreement in each rating scale. Figure 3.10 displays the resulting dendrograms. The y-axis represents the distance calculated based on the concordance correlation coefficient. The solid lines that connect subjects or clusters imply that there was significant positive concordance correlation between every two subjects in the cluster. Cluster {7, 8} was observed in *richness*, *balance*, and the distance between the two subjects was small in other attribute ratings as well as in the preference ratings in Phase 1. This phenomenon implied subjects 7 and 8 had similar opinions in the whole evaluation process. The distance between the three professional musicians (subject 6, 7 and 8) was small in the evaluation of violin *resonance* and *richness*, indicating that high inter-individual consistencies of *resonance* and *richness* among professional musicians in this experiment exist.



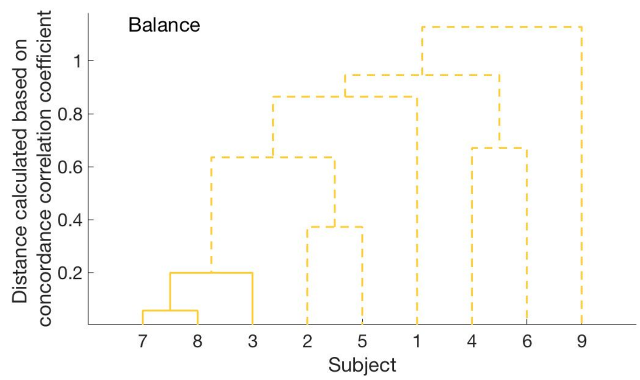


Figure 3.10 Hierarchical cluster analysis on participant-specific attribute profiles. The solid or dashed lines that connect individuals and clusters indicate the respective correlations between them.

3.4.5 Relationship between Preference and Attribute Ratings

The following subsection assessed the relationship between preference and criteria ratings. First, we assessed the relationship employing multiple rating-regression based on all the ratings along each attribute scale and all the preference ratings. A model was obtained to predict the preference ratings from the attribute ratings. The coefficients of the regression model were shown in Table 3.12. Thus, the multiple regression equation predicting the preference ratings can be written as

$$Preference = 0.185 + 0.579\,richness + 0.242\,clarity + 0.128\,balance + 0.056\,resonance - 0.068\,responsiveness$$

only the coefficients of *richness* and *clarity* were significant at 0.05 level as shown in the last column of Table 3.12, and all attribute ratings had positive correlations with preference ratings except the *responsiveness* ratings. The R^2 value of 0.725 indicates that the five criteria ratings accounted for 72.5% of the variation of the preference ratings.

Table 3.12 Multiple rating-regression analyzing the attributes that affect the preference ratings.

Independent variable	**Unstandardized coefficients**		**Standardized coefficients**	***t***	***p***
	B	**Std. Error**	***β***		
Constant	0.185	0.284		0.649	0.519

Richness	0.579	0.142	0.557	4.076	**0.000**
Clarity	0.242	0.119	0.241	2.035	**0.047**
Balance	0.128	0.110	0.124	1.169	0.248
Resonance	0.056	0.132	0.053	0.423	0.674
Responsiveness	-0.068	0.089	-0.065	-0.762	0.450
R = 0.851, R^2 = 0.725, adjusted R^2 = 0.696, F = 25.257					

During the analysis of the multiple rating-regression, we calculated the Pearson correlation between the preference ratings and the ratings of each criterion as well as the correlation between any two criteria ratings. The results showed that all of the correlations were significant at 0.01 levels ($\rho \geqq 0.334$, $p \leqq 0.007$) except that the correlations between the *responsiveness* and the preference ratings were significant at 0.05 level ($\rho = 0.310$, $p = 0.011$). The high R^2 of the regression model predicting the preference ratings from the five attribute ratings and the significantly high Pearson correlations between the preference ratings and the ratings of each criterion, and the correlations between any two criteria ratings may be due to the violinists employing a highly economic strategy in the evaluation process, which might lead to similar ratings in all rating scales. One of the subjects had mentioned at the end of this experiment: "each violin had very different personalities but when broken down into categories it is hard to list them without taking other aspects into account".

To avoid this interpretation when analyzing the relationship between preference ratings of Phase 1 and attribute ratings of Phase 2, we calculated the partial correlation coefficients ρ_p. Partial correlation coefficient $\rho_p(A, B \cdot C)$ measures the relationship between A and B while controlling for the influence of variable C by holding it constant [Gravetter and Wallnau, 2011]. For example, in order to measure the correlation between preference and *resonance*, the effect of *responsiveness*, *clarity*, *richness* and *balance* were controlled by the calculation of $\rho_p(Resonance, Preference \cdot responsiveness, clarity, richness, balance)$.

Partial correlation coefficients ρ_p were computed between each of the attribute scale ratings and the preference ratings of all subjects. The results are shown in Figure 3.11. All criteria ratings correlated to preference ratings positively except the *responsiveness* ratings. *Richness* and *clarity* correlated to preference significantly: ρ_p = 0.507 (p = 0.0002) and ρ_p = 0.282 (p = 0.047), respectively. The results thus indicated that subjects preferred violins with a richer and clearer sound. None of the other partial correlation coefficients between attributes ratings and preference ratings was significant ($p \geq 0.248$).

Figure 3.11 Partial correlation coefficient ρ_p between ratings of each attribute scale and preference.

Table 3.12 shows the most preferred and least preferred violins of each subject along with the criteria in which the corresponding violins were rated as best or worst. *Responsiveness* appears the least number of times in the column "also rated best in" of the most preferred violin, and to a lesser extent, *responsiveness* and *resonance* appear the least number of times in the column "also rated worst in" for the least preferred violin. We might deduce that violinists valued *responsiveness* least compared to the remaining four criteria when considering preference in the evaluation experiment. And the attributes of clarity, *richness* and *balance* were very important criteria during the violin preference assessment. This observation confirmed the analysis results of partial correlation between attributes ratings and preference ratings.

Table 3.12 The most preferred and least preferred violins of each subject along with attributes in which the corresponding violins were rated as best or worst.

Subject	**Most preferred violin**	Also rated best in	Rated worst in	**Least preferred violin**	Rated best in	Also rated worst in
1	P2	Clarity Balance		S3		Responsiveness, Clarity
2	S2		Resonance	S3	Resonance	Richness Balance
3	P2	Resonance Richness		S1		Resonance Clarity Richness Balance
4	S3	Clarity Balance		S2		Responsiveness Clarity Richness Balance
5	P1	Responsiveness Resonance Clarity Richness Balance		P2		Clarity
6	P2	Resonance Richness		S1		Responsiveness Resonance Clarity Richness Balance
7	P2	Resonance Richness Balance		S1		Responsiveness Resonance Clarity Richness Balance
8	P3	Resonance Clarity Richness Balance		S1		Responsiveness Resonance Richness Balance
9	P1	Clarity Richness		S1	Responsiveness	Resonance Clarity Richness

During Phase 1, each subject rated the six violins from 0 to 5. From these results, we can extract implicit rankings of the six violins from least preferred to most preferred. For violins at the same rank of preference, we averaged their ratings of each attribute scale across all subjects. Figure 3.12 displays the across-subjects average ratings of each attribute scale of the violins that were rated at the same rank of preference. Shapiro-Wilk tests showed that for each attribute scale, the ratings of the violins at each rank of preference were not all normally distributed. We conducted repeated-measures ANOVA to test the equality of means of attribute ratings of the violins at each rank of preference. The results showed that all violins having the same population means could be rejected for the five criteria except *responsiveness*: $F\,(5, 40) = 2.216$, $p = 0.072$. Post hoc tests with the Bonferroni correction revealed that the least preferred violin and/or the second least preferred violin was rated significantly lower than the most preferred violin for all five criteria except *responsiveness*.

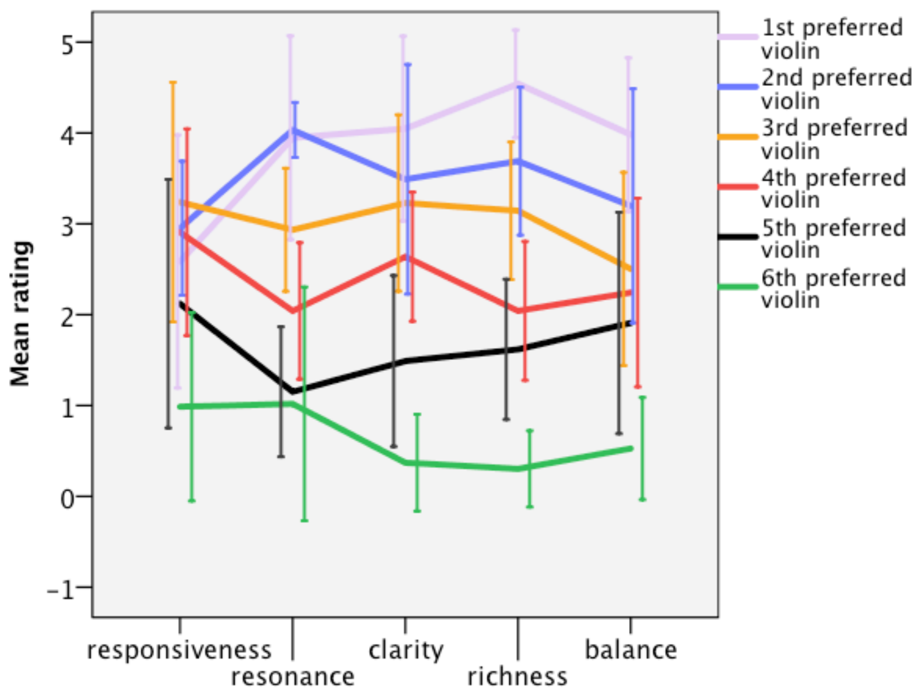


Figure 3.12 Across-subjects average attribute ratings of the violins at each rank of preference (error-bar = 95% confidence interval of the mean).

3.4.6 Verbal Descriptions of Violin Attributes

The subjects were asked to give comments or remarks about the specific criterion that they had evaluated and describe the particular behavior they noticed of the violin that was rated as best or worst for that criterion. The original verbal responses from subjects are provided in the Appendix "Questionnaire B of Study 1 in Chapter 3".

For *responsiveness*, some subjects thought the most responsive violin *had a very clear sound at the attack* and *required less effort to create the sound*. Some subjects thought *responsiveness* meant *ringing* and *sound coming out*. Subjects thought the sound of the least responsive violin was *small*, *noisy*, and *less resonant*. One subject mentioned that the *violin with a bad and dry sound responded quickly*, while the *violin with a thick sound needed more time to respond*. Another subject commented that *the violin with a lower bridge responded fast*, but *the sound wasn't solid*.

For *resonance*, subjects considered that the most resonant violin *had a bright sound* and and *a very open ringing quality with overtones and a powerful quality*; the least resonant violin *had a muted*, *stiff sound with no ringing/brilliant qualities*. Two subjects thought that *resonance was different from responsiveness*, as it was not the case that more resonance was better: good resonance seems to imply ringing very well, but dark sounding instruments usually have less resonance, e.g. muffled, so the challenge is to find *a good "dark"* violin: *ringing but still dark*.

For *clarity*, three of the subjects considered that *clarity* was related to *responsiveness*. They commented that using the clear violin facilitated *easy production* of *brilliant*, *pure*, *concentrated* and *clean* sound, which *didn't change with bow force*. *Each note boomed nicely*, and *ringing well helped with the transition between notes*, *making them connect well with each other*. The least clear violin *had a buzzing*, *muddy sound that lacked purity*, and it became more serious as the bow force increased. There was one subject, however, who thought the clearest violin was *brilliant but lacked flavor*. So, his most preferred violin had good flavor, but was not the clearest violin.

For *richness*, subjects generally thought the richest violin *had lots of colour and undertones in the sound* and *good expression*. The sound was *deep*, *dark*, *sweet*, *thick*, *big* and *fat*. Two subjects considered that richness was *partly related to resonance*. They considered that the least rich violin did *not have many colors* and *lacked depth*, the sound was *hollow*, *open*, *narrow* and *flat*.

For *balance*, the subjects considered the most balanced violin *had an even*, *consistent sound* and *stable*, *balanced playing across strings*. The least balanced violin sound was considered either *not thick in the lower register*, *not bright in the higher register* or *good at one but bad at the other*. Some subjects considered the overall bad sounding violins as balanced. Finally, some subjects thought that *the bridge mattered a lot*. They thought that the most balanced violin *resonated* and *rang well over all strings*; the least balanced violin *had a sloped bridge that did not allow the strings to resonate separately*.

At the end of this experiment, the subjects were asked to provide comments about the test instruments, and whether they would change the preference ranking of Phase 1 after rating the attributes scales. The subjects suggested repairs and playing of these violins. Two subjects indicated that they would consider revising the preference ranking of one or two violins.

3.4.7 Conclusions of Phase 2

Statistic differences existed between the six violins for *richness* and *balance* ratings. For the set of violins we used in this experiment, performance violins were on average rated significantly higher than student violins in all attribute rating scales except *responsiveness* and *resonance*. And subjects who were professional musicians, and/or with higher educational degrees in music performance rated performance violins much higher than student violins in *resonance*, *clarity* and *richness*.

Relatively higher inter-individual consistency of *richness* and *balance* existed among subjects during the violin evaluation. We observed that the three professional musicians had much higher agreement on *resonance* and *richness*.

The analysis of the relationship between preference ratings and attributes ratings showed that violinists preferred violins with rich and to a lesser extent clear sound. The most preferred violin was rated significantly higher than the least preferred violin or the second least preferred violin in all attribute ratings except *responsiveness*.

From verbal collections, the violinists stated that some rating criteria were correlated, e.g., *resonance* and *richness*, *clarity* and *responsiveness*. *Resonance* and *responsiveness* were anti-correlated to some extent. This could be considered as the limitation of the experiment protocol.

We were trying to understand more about what the origin of the violin preference is, therefore, employed an analytical and quantifiable approach similar to previous studies [Saitis et al., 2012; Fritz et al., 2012, 2014]. Future studies may explore different experiment designs. Considering the higher inter-individual consistency among professional musicians, further analysis can be restricted to the results of these subjects.

3.5 Conclusion about the Perceptual Experiment

This experiment explored whether violinists would consistently discriminate entry-level from advanced-level violins. Three student and three performance violins were considered. Nine violinists evaluated the six violins according to their own preference and five attribute criteria. It was found that statistically significant differences existed between the six violins for preference, *richness* and *balance* ratings. The results also showed that performance violins were on average rated significantly higher than student violins in terms of preference, *richness*, *clarity* and *balance*. And subjects who were professional musicians, and/or with higher educational degrees rated performance violins much higher than student violins in preference, *resonance*, *clarity* and *richness*.

Large inter-individual variations in the preference and criteria ratings of the violins existed, except for relatively higher inter-individual consistency on *richness* and *balance* ratings. However, three professional violinists highly agreed with each other on the preference, *resonance* and *richness* ratings in this experiment, which has not been observed in previous experiments (Saitis et al., 2012; Fritz et al., 2012b, 2014). This implies that there were more perceivable differences between the entry-level and performance violins, though this conclusion is limited by the small number of highly-skilled participants.

The analysis of the relationship between preference ratings and attributes ratings showed that violinists preferred violins with rich and to a lesser extent clear sound. The least preferred violin was rated significantly lower than the most preferred violin in all five criteria except for *responsiveness*. From the verbal collections, it was found that the violinists considered *resonance*, *response*, *balance*, *projection*, *richness*, *texture*, *interest*, *clarity* and *craft* when evaluating violins.

3.6 Bridge Admittance Tests

Bridge admittances were measured for the test instruments. A typical measurement procedure was employed. The test violin was tuned, and the strings were damped. The test violin was clamped on a frame with a piece of foam around the violin neck. A bag of sand was placed on the frame to minimize vibrations of the frame structure. This means of support mimicked the way a player holds the violin when playing. The bridge was excited with a miniature force hammer (PCB 086E80) and the resulting velocity was measured by a laser-Doppler vibrometer (Polytec PDV 100) both from the G-string corner. For each violin, we performed 3 to 5 measurements, and the results were averaged. The measurements were conducted in a lab with an area of approximately 30 m^2 and free of strong resonances.

Figures 3.13 and 3.14 display the measurements for performance violins and student violins respectively. The bridge admittances of the most preferred violin P2 and the least preferred violin S1 are shown in Figure 3.15. Calibrated admittance magnitudes in all figures in this thesis are shown in dB relative to 1 $ms^{-1}N^{-1}$. From these figures, we can see that the magnitudes of the B1- (peak between 400 and 500 Hz) signature mode were much smaller for student violins than for performance violins. A cluster of modes forming an apparent "hump" around 1 kHz (transition hill) for performance violins slightly decreased in amplitude between 1 kHz and 1.5 kHz, and then increased to form the bridge hill – another broader hump around 2-3 kHz. The magnitude response of the student violins, however, stayed relatively constant between 1 kHz and 2 kHz and did not show a dip around 1200 Hz. And there was more variation of the bridge hill magnitude for student violins than performance violins. In Bissinger's results [2008], excellent violins showed higher magnitude in three 250-Hz bands, with center frequencies of 875 Hz, 1125 Hz, 2375 Hz, respectively. Similar distinctions are not apparent in the student-performance violins of our experiment.

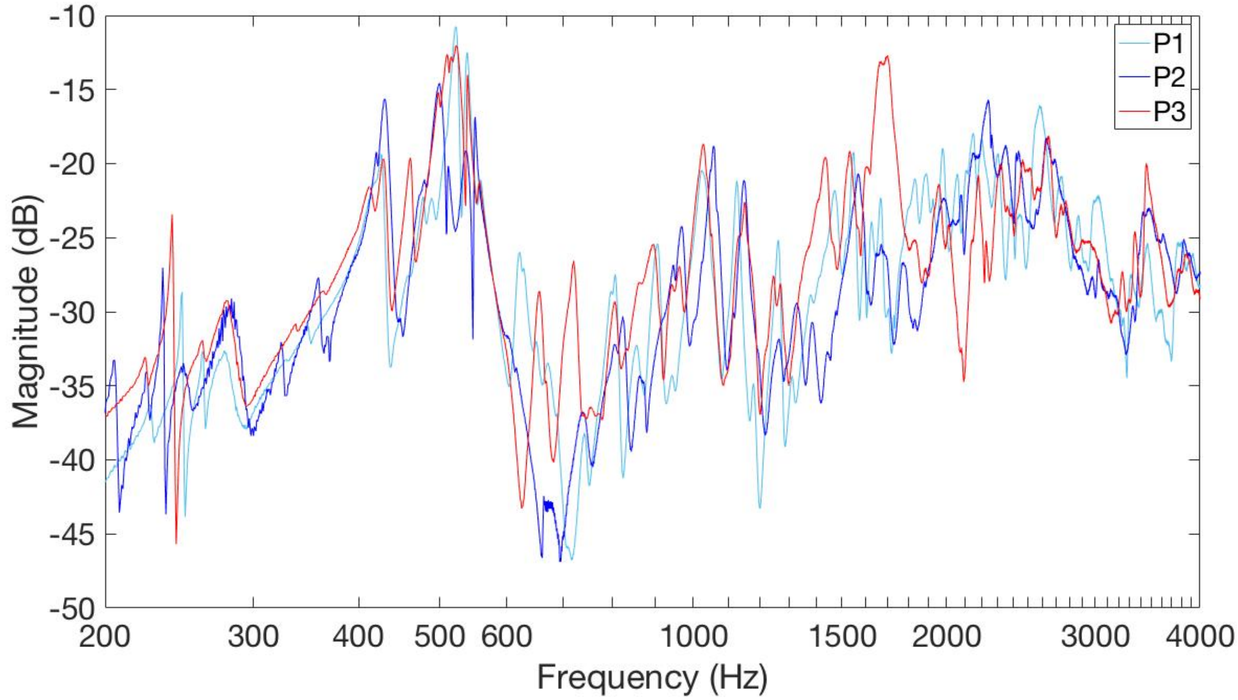


Figure 3.13 Measured bridge admittances of performance violins.

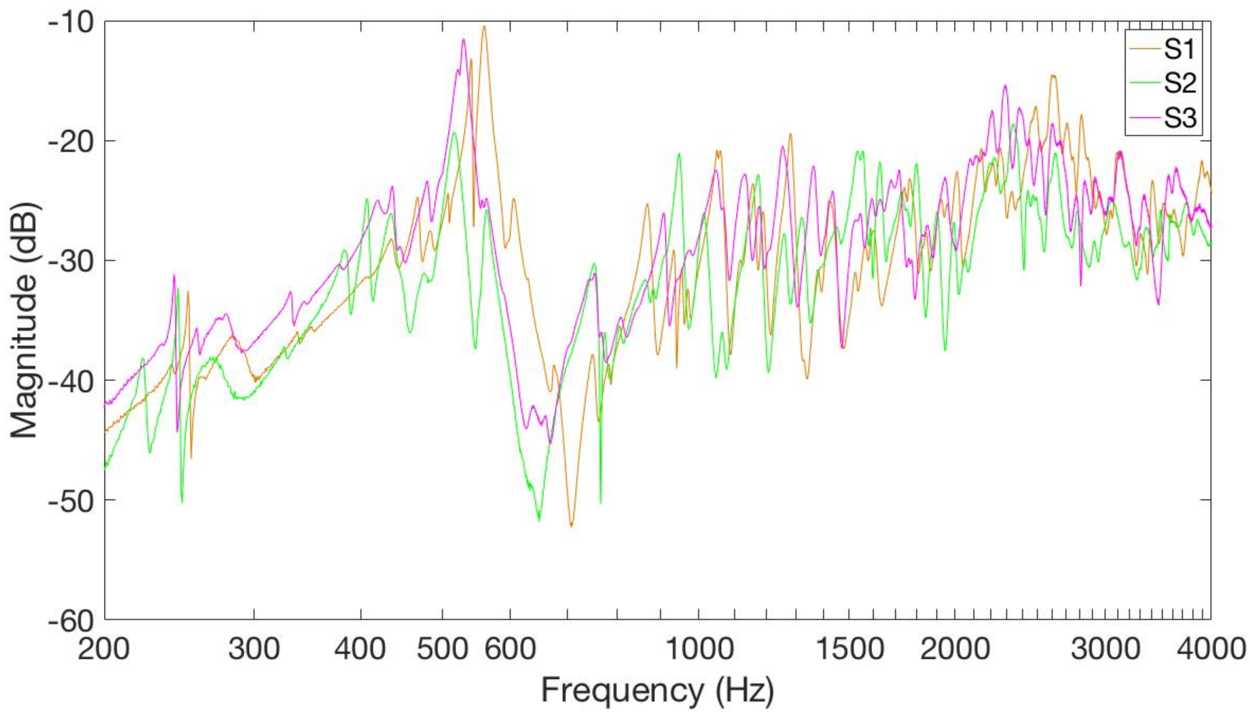


Figure 3.14 Measured bridge admittances of student violins.

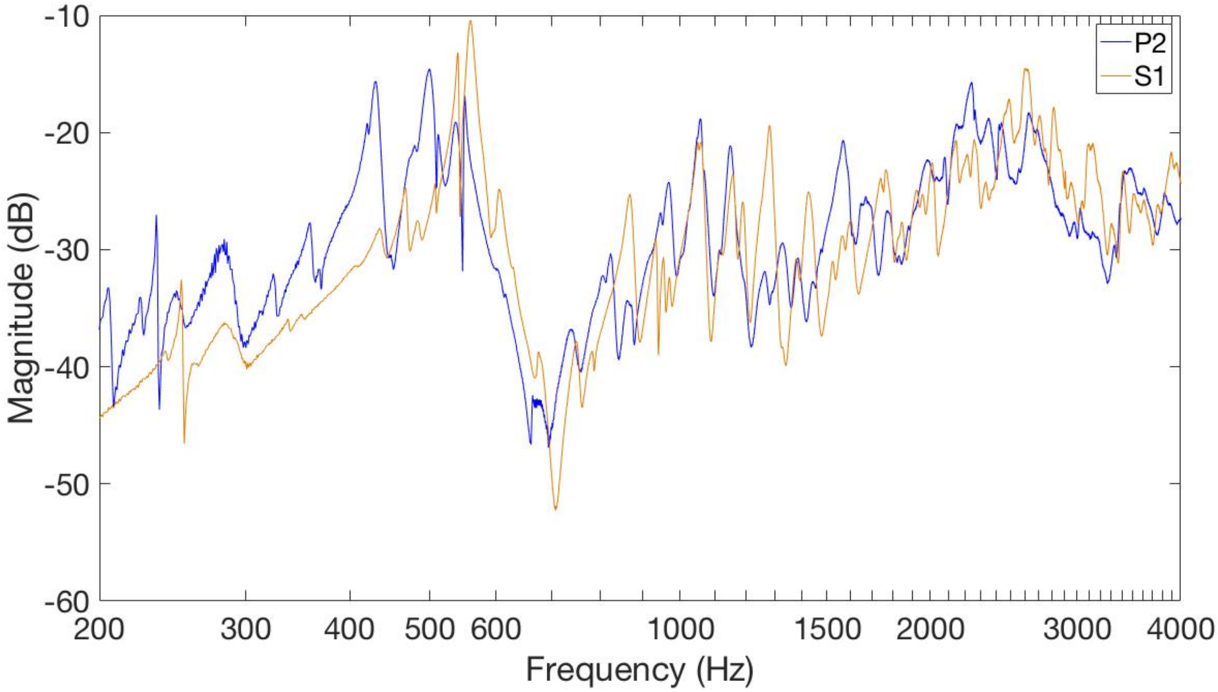


Figure 3.15 Measured bridge admittances of the most preferred violin P2 and the least preferred violin S1.

The values of the signature mode peaks were recognized through a Matlab script employing a mode extraction routine. The mode center frequencies and amplitudes were estimated by parabolic interpolation of the peaks of the frequency-domain admittance data. The Q value of each peak was estimated by fitting lines between each peak and its adjacent notches, picking the slope with the greatest magnitude and using that to compute the 3 dB bandwidth.

The magnitudes of the three signature modes for the six violins are shown in Figure 3.16. The magnitudes of the two signature modes A0 and B1- were higher in performance violins than student violins. We might deduce that the magnitudes of the A0 and B1- modes correlated with violin quality positively in this experiment.

The frequencies of the three signature modes for each of the six violins are shown in Figure 3.17. No significant differences of the mode frequencies between performance violins and student violins were found, which is consistent with Bissinger's finding [Bissinger, 2008]: no obvious quality trend for mode frequencies was found by contrasting the properties of "excellent" and "bad" violins. But the most preferred violin P2 and the least preferred violin S1 differentiated themselves in each group. The frequency of the B1+ mode for violin P2 was lower than the other two performance violins. Frequencies of mode B1- and B1+ for violin S1 were much higher than the other five violins. B1+ and B1- are the first corpus bending modes and the lowest strong corpus radiators [Bissinger, 2005]. Therefore, a lower frequency B1+ mode for violin P2 may affect its tone in the lower register, e.g., sound harmonics in the lower register radiate more strongly, thus making the violin sound "darker", as some subjects' commented about this violin (see Table 3.7).

The Q values of the three signature modes for the six violins are shown in Figure 3.18. The only difference we could find between the two types of violins was that the Q values of the signature mode A0 were slightly higher in performance violins than student violins, implying the damping of the A0 mode was lower for performance violins than student violins.

In summary, the comparison between performance and student violins through bridge admittance measurements showed that the only apparent differences were that the magnitudes of the A0 and B1- modes were higher in performance violins than student violins, and the student violins didn't show the "transition hill" around 1 kHz.

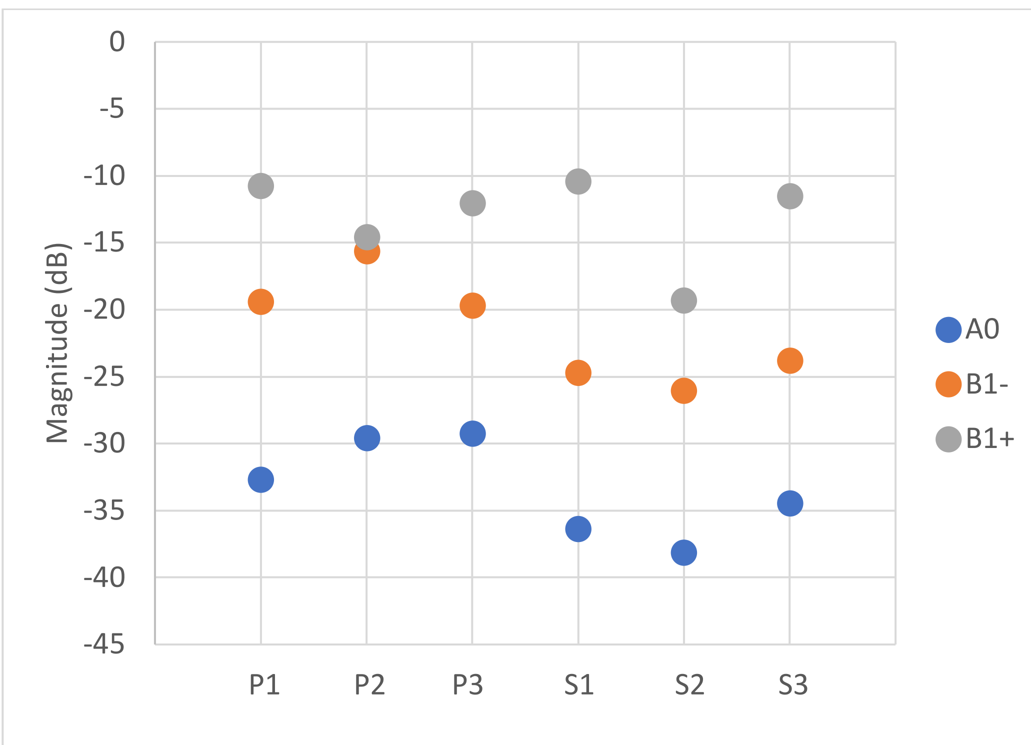


Figure 3.16 Magnitudes of the A0, B1- and B1+ mode for the six violins.

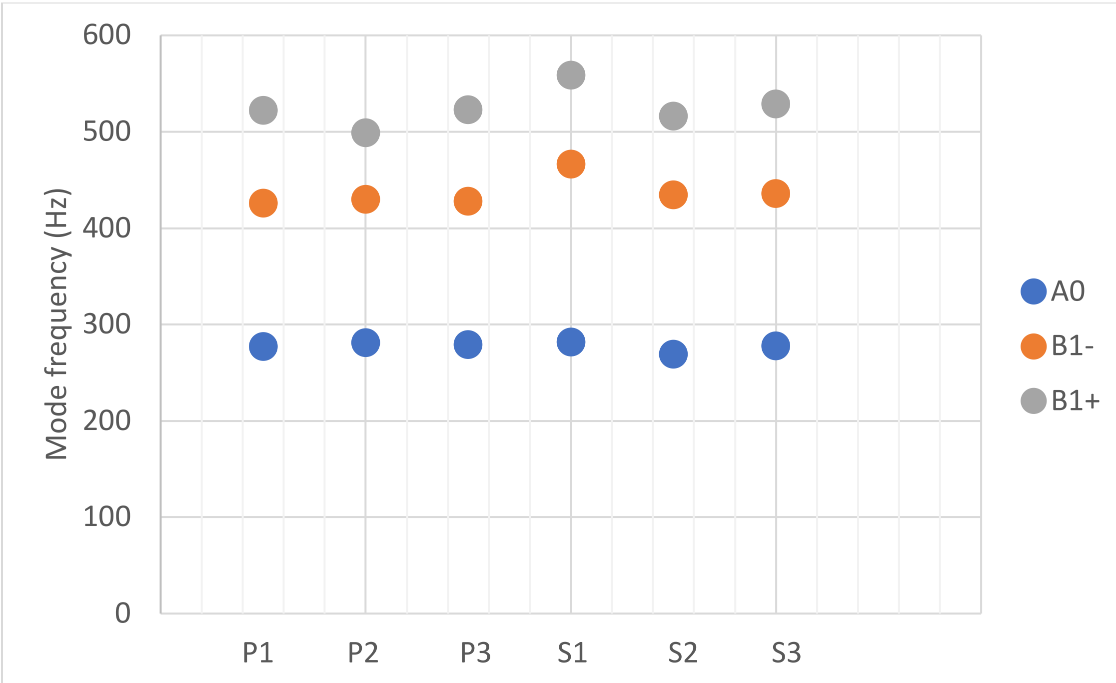


Figure 3.17 Frequencies of the A0, B1- and B1+ mode for the six violins.

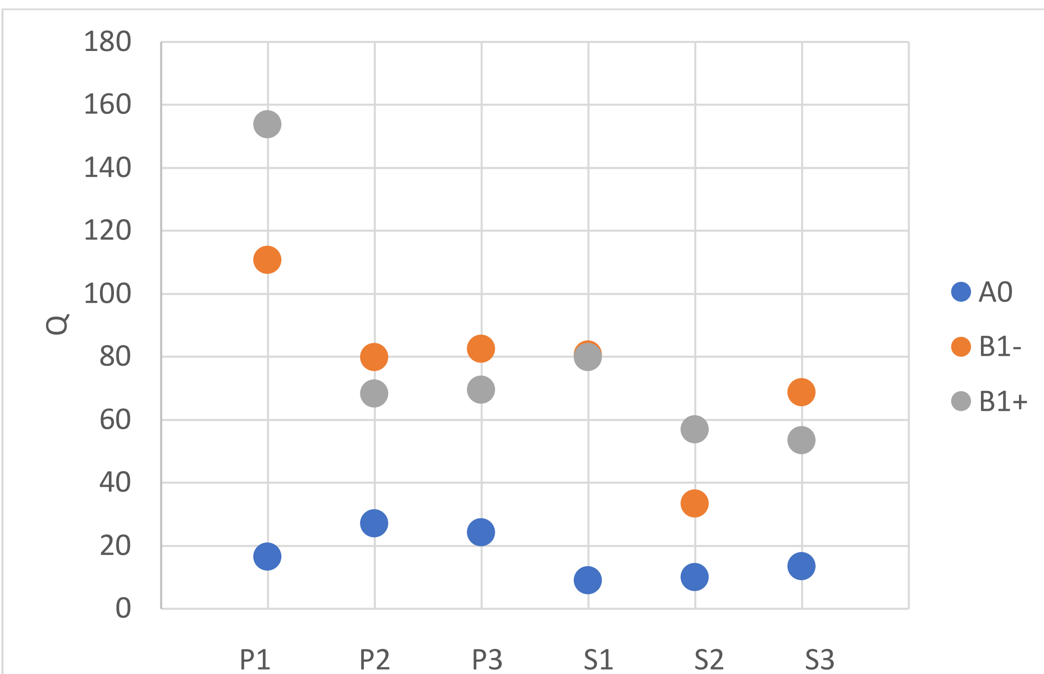


Figure 3.18 Q values of the A0, B1- and B1+ mode for the six violins.

Chapter 4

4 How Different Strings Affect Violin Qualities

4.1 Introduction

The study described in this chapter was designed to investigate the influence of different types of strings on the perception of violin qualities. Both players and makers agree that different string models and brands can make a big difference in how the instruments play or sound. To evaluate the perception of different strings using playing tests, one would ideally have identical violins that could be fit with different string types and players could compare those instruments in quick succession. However, given the nature of wood and the violin making process (with most components carved by hand), every violin is different. Therefore, it was decided that the different string types would be mounted on the same violin. Because the process of changing strings can take several minutes, which might lead to subjects forgetting their impressions of a previous setup, we made use of a reference violin of similar make and quality, such that perceptual evaluations were always made with respect to the unchanged instrument. The experiment was conducted in several sessions. In the first session, the two violins were strung with the same type of strings. During the subsequent sessions, the strings of one specific violin were kept the same and the strings of the other violin were changed to another type. A playing-based evaluation approach was adopted, with controlled experimental conditions. Detailed descriptions of the design of the experiment are presented in Section 4.2. Experiment results and analyses are provided in Section 4.3. Concluding remarks and discussions are given in Section 4.4.

4.2 Materials and Methods

This section describes the details of this experiment, including general design, test violins and strings, venues and controls, characteristics of participants and detailed procedure.

4.2.1 General Design

The aim of this experiment is to investigate whether violinists can tell the difference between different types of strings and how strings affect the perceptual quality of the violin. Two violins with similar sound quality and playability were employed in this experiment. In the first session, the two violins were installed with the same type of strings. Subjects were invited to play the two violins, and then describe and rate the differences between the test instruments according to specific criteria. After this session, the strings of one of the two violins were changed to a different type. This change was unknown to the subjects. Subjects were asked to repeat the evaluation and rating process. By comparing the descriptions and ratings between the two sessions, we could examine whether violinists can differentiate between strings and how strings affect the perceptual quality of the violin.

4.2.2 Test Violins and Strings

A pool of similar student quality violins (all being sold around $600 Canadian dollar) with the same type of strings were assembled at a local luthier shop. An experienced violinist, as well as two violin makers, were invited to select the two most similar violins from the pool. The violins and their strings were relatively new. Because they were coming from the available sales stock of a workshop, they had not been played on a regular basis.

Three types of strings were involved in this experiment. The strings labeled "Dominant" in this study were a set of Thomastik 135 Dominant synthetic core with medium tension for the G (silver wound), D (aluminum wound), and A (aluminum wound) strings; while the E string was a Pirastro gold (steel, medium tension). They were installed on both violins initially. According to the luthiers, Dominant strings were very commonly used, especially among student players. The other two types of strings were generously donated by the string manufacturer d'Addario. We requested two different types of strings of different qualities for the experiment and they provided several new sets of Kaplan and Pro-Arté strings. The Kaplan strings were a set of KA 310 synthetic core with medium tension for the G (silver wound), D (silver wound), and A (aluminum wound) strings; while the E string was made of tinned carbon steel, medium tension as well. The Pro-Arté strings were a set of J 56 nylon core with medium tension for the G (silver wound), D (aluminum wound), and A (aluminum wound) strings; while the E string was made of tinned high-carbon steel, medium tension as well. The cost of the three sets of strings (Dominant, Kaplan and Pro-Arté) were around \$78, \$108 and \$49, respectively.

It would have been interesting to compare strings made of very different materials, such as gut, steel or synthetics, and perhaps look for correlations with string properties (as briefly discussed in Section 2.5.1). However, instead of making an extensive and comprehensive study to correlate string properties with violin qualities, this first study was designed to see whether strings of different prices can lead to perceptual differences in violin qualities and what the perceptual differences are. On the other hand, these synthetic strings are also more popular than gut or steel strings among modern violinists, thus the results could have more practical significance.

4.2.3 Venues and Controls

This experiment took place at two locations. The first was at Oberlin College, Oberlin, Ohio, USA. The second was at McGill University, Montreal, Quebec, Canada. The experiments both took place in rooms free of strong resonances in order to avoid coloring the sounds heard by the subjects. The surface area of the experiment room in Oberlin was approximately 18.4 m^2, while the experiment room in Montreal was approximately 26.7 m^2.

In order to eliminate the possible influence of visual information (colour of varnish, distinctive markings, string wrappings, ...) on judgement, the participants were asked to wear dark sunglasses while the lighting in the room was reduced. In the end, no subject reported feeling uncomfortable wearing dark sunglasses when evaluating the violins.

Players typically use their own bows when testing violins. While the use of a common bow could be considered to reduce variability in the experiment, players might feel uncomfortable using a bow they are not familiar with. As a result, like in all previous playing tests [Saitis et al., 2012, 2015; Fritz et al., 2012b, 2014], players were asked to use their own bow to evaluate the violins. The violinists were given the option to either use their own shoulder rest, no shoulder rest, or one we provided (Kun Original model).

4.2.4 Participants

Nine professional string players (subjects 1-9) took part in this experiment in Oberlin. Among them, there were six violinists, two violists and one cellist. The two violists and the cellist all indicated they had a lot of experience playing the violin. Players in Oberlin were very skilled, and they were good at evaluating instruments. They were invited to the Oberlin violin acoustics workshop to provide luthiers and researchers feedbacks from players' perspective. Ten skilled violinists (subjects 10-19) participated in this experiment in Montreal and were paid for their participation. In total, there were 19 participants (11 males, 8 females; 15 native English speakers, 2 native French speakers, 2 native Mandarin speakers); average age= 28 yrs, SD= 8 yrs, range= 21-52 yrs. They had at least 16 years of playing experience (average years of playing = 23 yrs, SD= 8 yrs, range = 16-45 years, average years of training = 20 years, SD= 5 yrs, range = 13 – 32 yrs; average hours of practicing per week = 28 hrs, SD= 10 hrs, range= 1 – 45 hrs). The estimated prices of their own violins range from $6K to $40K. Eighteen subjects described themselves as professional musicians, 5 were doctoral candidates in music performance, 4 had master's degrees in music performance, 7 were currently master students in music performance, and 1 had an artist diploma. They reported playing a wide range of musical styles [classical (100%), folk (26%), baroque (37%), jazz/pop (26%), and contemporary (26%)] and in various types of ensembles [chamber music (95%), symphonic orchestra (89%), solo (89%), and folk/jazz band (16%)].

4.2.5 Detailed Procedure

The first part of this experiment took place during the sixteenth Oberlin violin acoustics workshop in June 2017 at Oberlin College, which was attended by a mix of professional string players, violin makers and researchers. The two selected similar violins were brought to this workshop. The experiment consisted of two sessions. Each session lasted approximately 20 minutes and there were two phases in each session. Subjects were scheduled individually. The experimenter was always present in the room for instructing and taking notes for the subjects. A small piece of paper was attached on the second violin scroll in order to differentiate the two violins. They were placed on a sofa in random order by the experimenter and the order was switched between subjects. In session 1, both violins were strung with the same type of Dominant strings. During the first phase of session 1, subjects were given 5 minutes to play and compare the two violins and they were told that they would have to describe the differences between the two violins after playing. The experimenter took notes of the subjects' description of the differences. After finishing the first phase of session 1, subjects were given eight criteria to rate that were carefully selected from previous publications [Saitis et al., 2012, 2017]. Compared to the criteria used for the experiment described in Chapter 3, we added two more criteria (*power* and *brightness*) to be rated in this experiment. These additions were made because it was expected that the strings could influence these criteria. A short definition was provided for each criterion. The list of criteria and their definitions are given in Table 4.1 [Saitis et al., 2012, 2017]. More explanations were provided orally whenever needed by the subjects. The definitions in Table 4.1 were changed somewhat in comparison to the previous study based on feedback from violinists and their understanding of the criteria.

Table 4.1 Definitions of rating criteria.

Responsiveness	*Responsiveness* describes how fast the violin can respond to different bowing techniques by the violinist, and how easier the violinist can control the playing process and the played sound.
Power	*Power* describes the intensity of the radiated sound "under the ear".
Resonance	*Resonance* describes sustain time after bowing has stopped.
Brightness	Violinists may use bright, brilliant (trumpet compared to clarinet), lots of high overtones etc. to describe the violin sound in terms of *brightness*.
Clarity	A sound is described as "clear" when perceived as lacking audible artifacts when played, such as wolf notes, "buzzing", or a slow buildup of energy during attacks and transients.
Richness	*Richness* refers to the presence of overtones in the sound, or the perceived number of partial frequencies present in a violin note.
Balance	*Balance* refers to the relative similarity of sound or physical response of the violin across notes and strings of the instrument.
Overall quality	Overall quality includes the sound quality, playability as well as subjects' preference.

Subjects were asked to compare violin 2 to violin 1 according to the given criteria and rate, for each criterion, the difference level between the two violins on a scale from -3 to +3. A criterion difference rating of 0 implies that violin 2 is not different from violin 1 for criterion X. A criterion difference rating of 1 (-1) means that violin 2 is a little more (less) X compared to violin 1. Similarly, criteria difference ratings of 2 (-2) and 3 (-3) signify moderate and significant differences, respectively. The reason for rating the difference between the two violins instead of rating each violin separately was that subjects could be more precise and oriented while rating as they always had a reference in mind during the rating process. If rating each violin separately, subjects would be very free and could not be that precise, possibly making the comparison between the two violins not explicit. While the difference rating might seem more demanding, no subject expressed difficulty during the evaluation. The decision to use violin 1 as the reference was somewhat arbitrary, as the two violins were selected based on their similarity within the pool of available instruments. However, during the selection process, violin 1 was considered to be a bit better according to the violinist and violin makers who participated. We thus chose violin 1 as the reference, hypothesizing an increase of quality for violin 2 with higher quality strings which would reduce the difference between the two violins. Subjects were given 2 minutes to rate each criterion.

It took three days for all nine subjects to attend session 1 of this experiment. Then, a violin maker changed violin 2 to Kaplan strings while the original Dominant strings were maintained on violin 1. The procedure of session 2 was identical to session 1. The same nine subjects were invited back to participate in session 2.

The second part of this experiment was organized in Montreal. Compared to the experiment in Oberlin, subjects completed the whole experiment within one session, as we were concerned about getting subjects to return for a second session. In addition to the Dominant and Kaplan strings, one more string type was added in Montreal, hence there were three trials. Again, the subjects were scheduled individually. The entirety of the experiment lasted 1 hour to 1.5 hours. During the first trial, the two violins were set up with their initial Dominant strings, as in Oberlin. Violin 2 was then changed to Kaplan strings (same set as Oberlin) or Pro-Arté strings during trial 2 or trial 3, respectively. The order of Kaplan strings and Pro-Arté strings was randomized between subjects. The procedure during each trial was identical with each session of the experiment in Oberlin. Between trials, the subject was asked to sit on a chair outside the experiment room without any knowledge about what happened inside but was told that there may or may not be some changes to the violins. The experimenter changed the strings for violin 2 and carefully tuned it, trying as much as possible to avoid any movement of the bridge. It took approximately 8 minutes to change and tune the strings. Once the strings were changed and the violin was tuned, the experimenter asked the subject to continue with the next trial. During the experiments, there were two special cases: 1. The first subject in Montreal (subject 10) evaluated a new set of Evah Pirazzi strings instead of Kaplan strings during trial 2, as the experimenter wanted to try two different types of strings to decide which type of strings to use other than Kaplan strings. 2. During the participation of the last subject in Montreal (subject 19), the bridge was broken during the changing of strings between trial 2 and trial 3. As a result, no data regarding Pro-Arté strings for subject 19 were available for the following analyses. And because of the damaged bridge, the condition of violin 2 changed. Consequently, including more subjects was not possible for this specific experiment.

4.3 Results

In this section, the results of the experiments in Oberlin and Montreal were analyzed separately and a comparison between the Oberlin and Montreal results was performed. As mentioned in the previous section, the first subject in Montreal (subject 10) did not evaluate the Kaplan strings experimental condition, and no data regarding the Pro-Arté strings experimental condition for the last subject in Montreal (subject 19) was available because of the bridge of violin 2 broke. To involve as many subjects' results as possible for the analysis of the Montreal results, we compared each pair of experimental conditions first, then compared three experimental conditions together. We also examined the relationship between attribute difference ratings and overall quality difference ratings. The experimental conditions of violin 1 strung with Dominant strings and violin 2 strung with Dominant, Kaplan or Pro-Arté strings are abbreviated as D1-D2, D1-K2 or D1-P2, respectively.

4.3.1 Comparison between D1-D2 and D1-K2 Experimental Conditions based on Oberlin Results

During the first session, violin 1 and 2 were both strung with Dominant strings. The across-subjects average criteria difference ratings are shown in Figure 4.1. Error bars of two-sided 95%

confidence interval (CI; all CIs are two-sided 95% intervals through this chapter) of the means are also displayed. From the observed means, we can see that violin 2 was rated a little higher than violin 1 for *resonance*, *power* and *brightness*, but lower for *responsiveness*, *clarity*, *richness*, *balance* and overall quality. During the second session, we changed the strings of violin 2 to Kaplan strings and kept the same set of Dominant strings on violin 1. From the observed average criteria difference ratings, we find that the *responsiveness*, *power* and *balance* of violin 2 improved while its *clarity* and *richness* deteriorated. The *resonances*, *brightness*, and overall quality difference ratings stayed about the same. Of the improvements observed, *balance* was most notable: violin 2 with Kaplan strings was as balanced as violin 1 with Dominant strings.

To determine whether the results we observed were statistically significant, we first conducted Shapiro-Wilk tests to measure the distributions of the differences between the two experimental conditions criteria difference ratings by all subjects. Then, depending on the distribution results, we conducted paired-samples t-tests and related-samples Wilcoxon signed rank tests. Shapiro-Wilk tests showed that the distributions of all the differences between the two experimental conditions criteria difference ratings by all subjects were normal except clarity. Thus, paired-samples *t*-tests were carried out on the other seven criteria difference ratings. The results showed that the differences between the two conditions on these seven criteria were not significant, absolute value of paired samples $t\ (8) \leq 0.936$, $p \geq 0.377$. Related-samples Wilcoxon signed rank test was performed to test the differences between the two conditions on the *clarity* difference ratings. The results showed that there was no significant difference between the two conditions, $z = 0.707$, $p = 0.48$.

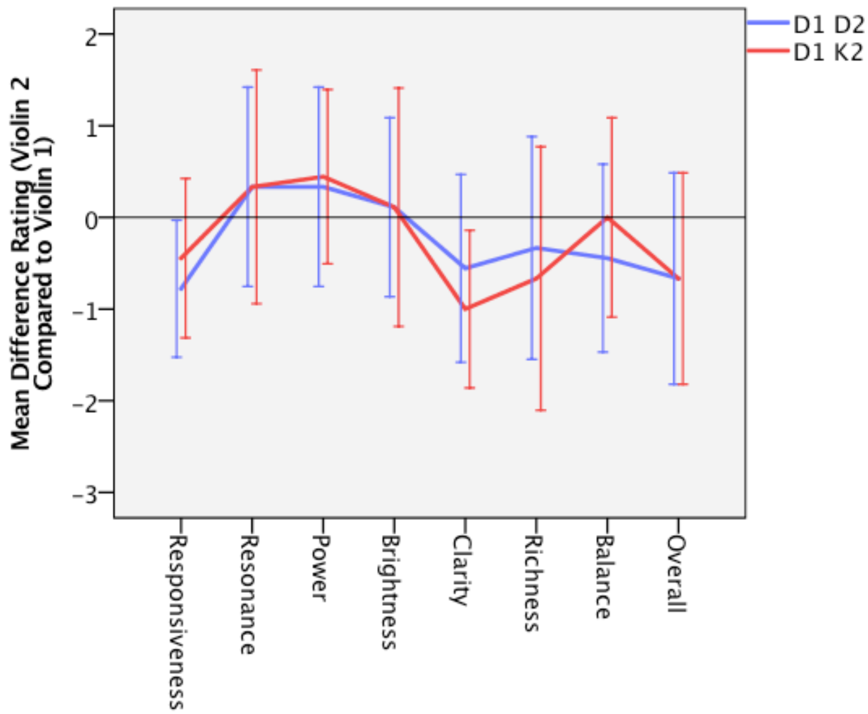


Figure 4.1 Across-subjects average of the criteria difference ratings (error-bar = 95% confidence interval of the mean) for both sessions.

4.3.2 Comparison between Each Pair of Experimental Conditions based on Montreal Results

In this section, comparison between each pair of experimental conditions was conducted based on the results obtained in Montreal.

4.3.2.1 Comparison between D1-D2 and D1-K2 Experimental Conditions

The analysis conducted in this section is based on the results of nine subjects in Montreal (subjects 11-19). The across-subjects average criteria difference ratings for each of the two experimental conditions are shown in Figure 4.2 with error bars of two-sided 95% confidence interval. During the first trial of each subject, violin 1 and 2 were strung with the same type of strings: Dominant strings. The observed mean difference ratings of all criteria were negative, implying that violin 2 was considered worse than violin 1 for all criteria when they were both strung with Dominant strings. During the second or third trial of each subject (different for different subjects), we changed violin 2 to Kaplan strings and kept the same set of Dominant strings on violin 1. From the observed means, we can see that *resonance*, *clarity*, *balance*, *richness* and overall quality of violin 2 were improved while its *responsiveness*, *power*, and especially *brightness* deteriorated.

Similar statistical analysis methods were employed as in the previous section and none of the differences for the criteria difference ratings between the D1-D2 and D1-K2 conditions was found to be significant: absolute value of paired samples t (8) $\leq$ 1.897, $p \geq 0.094$; related-samples Wilcoxon signed rank test $z \leq 1.192$, $p \geq 0.233$.

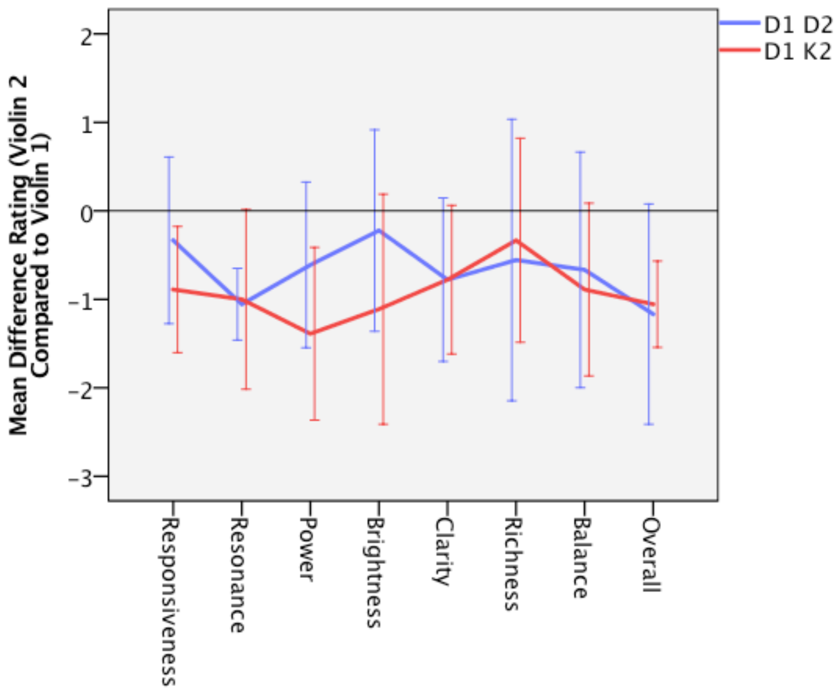


Figure 4.2 Across-subjects average of the criteria difference ratings(error-bar = 95% confidence interval of the mean) for two trials.

4.3.2.2 Comparison between D1-K2 and D1-P2 Experimental Conditions

The analysis to compare D1-K2 and D1-P2 experimental conditions is based on the results of eight subjects in Montreal (subjects 11-18). Figure 4.3 shows the across-subjects average criteria difference ratings for each experimental condition with error bars of two-sided 95% confidence interval. During the second or third trial of each subject, violin 2 was strung with Kaplan strings or Pro-Arté strings (different for different subjects), while violin 1 was maintained with Dominant strings. From the observed means, we can see that the *resonance*, *power*, *balance*, *richness* and overall quality of violin 2 were improved while its *responsiveness*, *brightness* and *clarity* deteriorated when strung with Pro-Arté strings compared to Kaplan strings. Of the improvements observed, *richness* was the most noticeable: violin 2 with Pro-Arté strings was considered richer than violin 1 with Dominant strings.

Similar statistical analysis methods were employed as in the previous sections and none of the differences for the criteria difference ratings between the D1-K2 and D1-P2 conditions was found to be significant: absolute value of paired samples t (7) ≤ 1.59, $p \geq 0.156$; related-samples Wilcoxon signed rank test $z = 1.622$, $p = 0.105$.

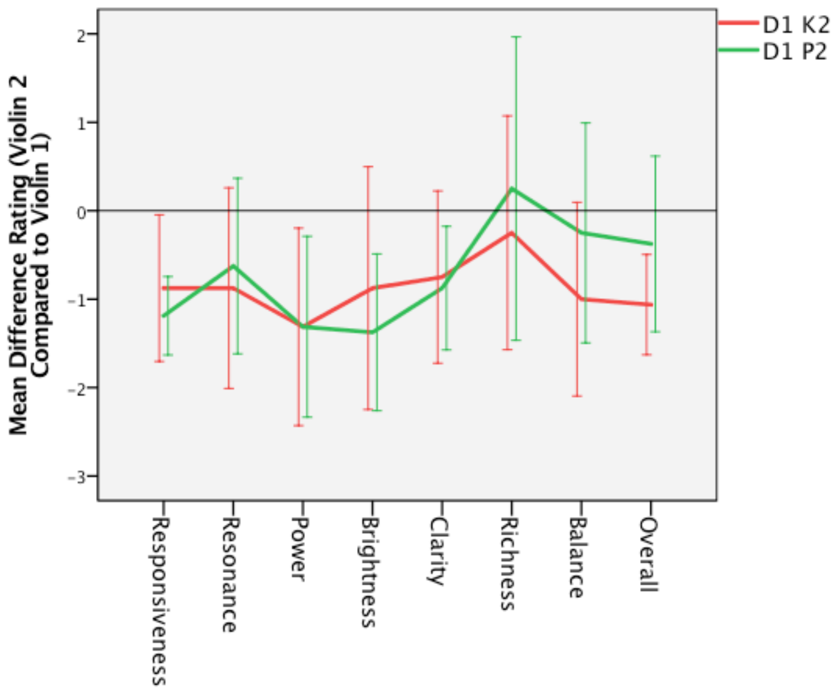


Figure 4.3 Across-subjects average of the criteria differentce ratings (error-bar = 95% confidence interval of the mean) for two trials.

4.3.2.3 Comparison between D1-D2 and D1-P2 Experimental Conditions

The analysis to compare the D1-D2 and D1-P2 experimental conditions is based on the results of nine subjects in Montreal (subjects 10-18). The across-subjects average criteria difference ratings for each experimental condition are shown in Figure 4.4 with error bars of two-sided 95% confidence interval. During the first trial for each subject, violin 1 and 2 were strung with the same type of Dominant strings. The observed mean difference ratings of all criteria were negative, implying that violin 2 was worse than violin 1 for all criteria when they were both strung with Dominant strings. During the second or third trial of each subject (different for different subjects), we changed violin 2 to Pro-Arté strings and kept the same set of Dominant strings on violin 1. From the observed means, we can see that *resonance*, *balance*, *richness* and overall quality of violin 2 were improved while its *responsiveness*, *power*, *brightness* and *clarity* deteriorated. Of the improvements observed, *richness* was most noticeable: violin 2 with Pro-Arté strings became richer than violin 1 with Dominant strings.

Similar statistical analysis methods were employed as in the previous section. The results showed that there were no significant differences between the two experimental conditions on all criteria difference ratings [absolute value of paired samples t (8) $\leq$ 2.054, $p \geq 0.074$; related-samples Wilcoxon signed rank test $z = 0.736$, $p = 0.461$] except for the *brightness* difference ratings (related-samples Wilcoxon signed rank test $z = -2.06$, $p = 0.039$).

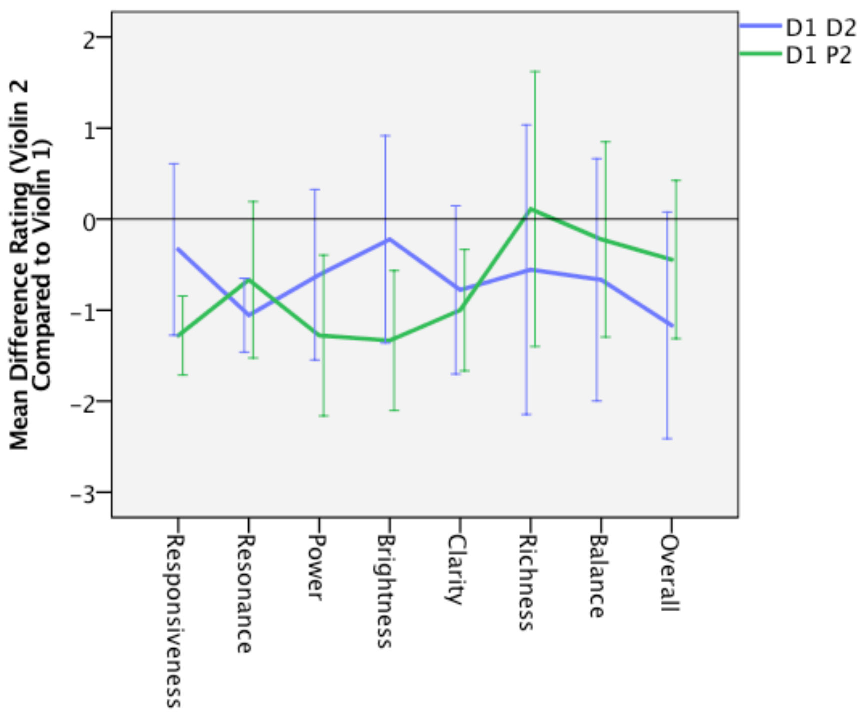


Figure 4.4 Across-subjects of the criteria difference ratings (error-bar = 95% confidence interval of the mean) for two trials.

4.3.3 Comparison among Three Experimental Conditions based on Montreal Results

The analysis leading to the comparison among the three experimental conditions is based on the results of eight subjects in Montreal (subject 11-18). Figure 4.5 shows the across-subjects average criteria difference ratings for each experimental condition with error bars of two-sided 95% confidence interval. Among the three experimental conditions, the observed mean difference ratings of *resonance*, *richness*, *balance* and overall quality were the highest in the D1-P2 condition, *responsiveness*, *power* and *brightness* difference ratings were the highest in the D1-D2 condition, and *clarity* difference rating was the highest in the D1-K2 condition. On the other hand, *responsiveness* and *brightness* difference ratings were the lowest in the D1-P2 condition, *resonance*, *richness* and overall quality difference ratings were the lowest in the D1-D2 condition, *balance* difference rating was the lowest in the D1-K2 condition, *power* difference rating was the lowest in both the D1-K2 and D1-P2 conditions, and *clarity* difference rating was the lowest in both the D1-D2 and D1-P2 conditions.

Statistical analyses were conducted to test whether the observed differences were significant. First, Shapiro-Wilk tests were performed to measure the distributions of the criteria difference ratings, and the results showed that the criteria difference ratings were not simultaneously normally distributed ($p < 0.05$) for the three experimental conditions except *power* and *richness* difference ratings. For that reason, one-way repeated measures ANOVA was only performed for *richness* and *power* difference ratings. The result showed that the *richness* and *power* difference ratings did not change significantly among the three experimental conditions: F (2, 14) = 1.355, $p = 0.29$, partial $\eta^2 = 0.162$; F (2, 14) = 0.797, $p = 0.47$, partial $\eta^2 = 0.102$. For the remaining criteria difference ratings, we conducted related-samples Friedman's two-way analysis of variance by ranks tests. The results showed that the null hypothesis that the distribution of the difference ratings for every criterion across the three experimental conditions was the same could not be rejected, $\chi^2(2) \leq 5.7$, $p \geq 0.58$.

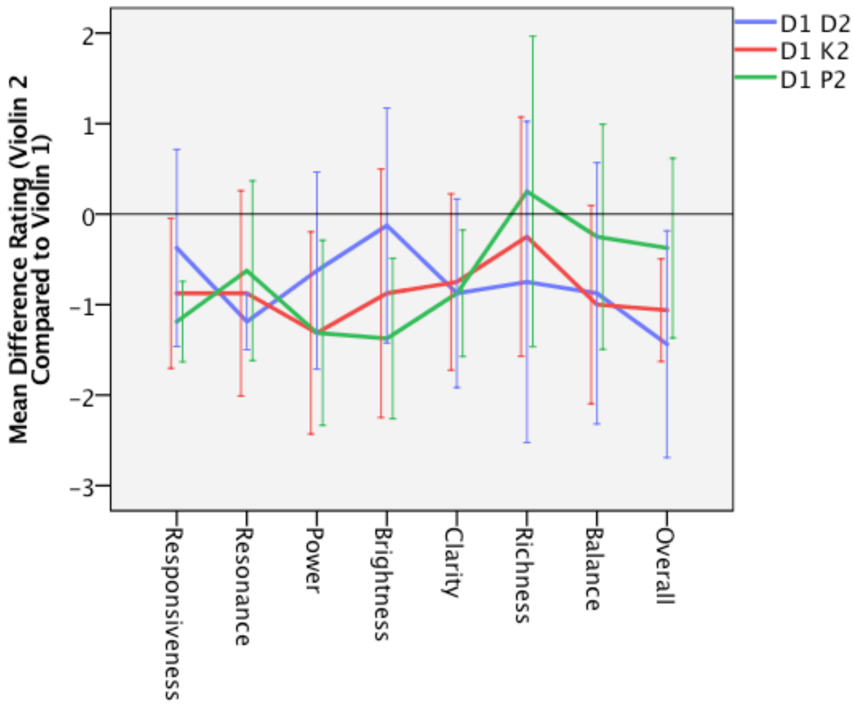


Figure 4.5 Across-subjects average of the criteria difference ratings (error-bar = 95% confidence interval of the mean) for three trials.

4.3.4 Comparison between Oberlin Results and Montreal Results

Comparisons between Oberlin and Montreal were carried out for both conditions D1-D2 and D1-K2 as well as for the differences between these two conditions (despite the variation in the presentation of stimuli). Depending on whether the distributions of the criteria difference ratings were normal or not, we conducted independent-samples t-tests or independent-samples Mann-Whitney U tests (when the normal distribution assumption was violated).

When comparing Figures 4.1 and 4.2, differences can be observed for some criteria. However, the null hypothesis that the distributions of all criteria difference ratings for a given condition were the same across the Oberlin and Montreal results could only be rejected for *resonance* in the D1-D2 condition (independent-samples Mann-Whitney $U = 14.5$, $z = -2.575$, $p = 0.01$) and for *power* in the D1-K2 condition (independent-samples Mann-Whitney $U = 14.5$, $z = -2.357$, $p = 0.018$). There were more than three months between these two parts of the experiment and the differences between the Oberlin and Montreal results might be partly attributable to seasonal changes. As we recruited professional and skilled subjects in both locations, we do not expect there were systematic differences between the subjects.

Despite these few significant differences, the null hypothesis that the distributions of all the differences between D1-D2 and D1-K2 criteria difference ratings were the same across Oberlin and Montreal could not be rejected: absolute value of independent samples $t(16) \leq 1.376$, $p \geq 0.188$; Mann-Whitney $U \leq 55.5$, $z \leq 1.357$, $p \geq 0.175$.

4.3.5 Relationship between Overall Quality and Attribute Ratings

In Chapter 3, we analyzed the relationship between preference and attribute ratings through a multiple rating-regression model and the computation of partial correlations. Similarly, in the following subsection we also employed multiple rating-regression and partial correlation analyzing the relationship between the overall quality difference ratings and attribute difference ratings. The analysis was based on all the difference ratings of different experimental conditions collected from all the subjects in the two experiment locations. A model was obtained to predict the overall quality difference ratings from the seven attribute difference ratings. The coefficients of the regression model were shown in Table 4.2. Therefore, the multiple regression equation can be written as

$$\begin{aligned} Overall\ quality &= -0.238 + 0.334\, richness + 0.246\, resonance + 0.202\, balance \\ &+ 0.242\, clarity + 0.088\, responsiveness + 0.022\, power - 0.045\, brightness \end{aligned}$$

While only the coefficients of *richness* and *resonance* were significant at 0.05 level as shown in the last column of Table 4.2, all attribute difference ratings correlated with the overall quality difference ratings positively except the *brightness* difference ratings. $R^2 = 0.635$, implied that the seven criteria difference ratings can explain 63.5% of the variation of the overall quality difference ratings, which was a bit lower than the R^2 generated in Chapter 3 (0.725).

Table 4.2 Multiple rating-regression analyzing the attributes that affect the overall quality difference ratings.

Independent variable	**Unstandardized coefficients**		**Standardized coefficients**	***t***	***p***
	B	**Std. Error**	***β***		
Constant	-0.238	0.189		-1.259	0.216
Richness	0.334	0.113	0.448	2.953	**0.005**
Resonance	0.246	0.111	0.260	2.211	**0.033**
Balance	0.202	0.108	0.221	1.871	0.069
Clarity	0.242	0.149	0.202	1.619	0.114
Responsiveness	0.088	0.153	0.069	0.576	0.568
Power	0.022	0.118	0.024	0.184	0.855
Brightness	-0.045	0.119	-0.052	-0.375	0.710
R = 0.797, R^2 = 0.635, adjusted R^2 = 0.568, F = 9.438					

As we explained in Chapter 3, the violinists may have employed a highly economic strategy in the evaluation process, which might lead to similar difference ratings for all criteria, as the R^2 of the regression model in this study was still relatively high. To avoid this possibility when analyzing the relationship between overall quality difference ratings and attribute difference ratings, partial correlation coefficients ρ_p were employed. Partial correlation coefficient $\rho_p(A, B \cdot C)$ measures the correlation between A and B while controlling for the effect of variable C by holding it constant. For example, in order to measure the correlation between overall quality and *resonance*, the effect of *responsiveness*, *power*, *brightness*, *clarity*, *richness* and *balance* were controlled by the calculation of ρ_p (*resonance*, overall quality · {*responsiveness*, *power*, *brightness*, *clarity*, *richness*, *balance*}).

Partial correlation coefficients ρ_p were computed between each of the attribute difference ratings and the overall quality difference ratings for all subjects involved in this experiment. The results are shown in Figure 4.6. *Richness* and *resonance* correlated with overall quality significantly: $\rho_p(38) = 0.432$ ($p = 0.005$) and $\rho_p(38) = 0.338$ ($p = 0.033$), respectively. The results indicated that participants rated the overall quality higher for the violin that they considered richer and more resonant. None of the other partial correlation coefficients between attributes difference ratings and overall quality difference ratings was significant, absolute $\rho_p(38) \leq 0.29$ ($p \geq 0.069$).

Compared to the study in Chapter 3, we added two more attributes *power* and *brightness* for rating. From the partial correlation result, these two criteria difference ratings did not seem to have close correlations with the overall quality difference ratings in comparison to other attribute difference ratings. Compared to the partial correlation result in Chapter 3, *richness* also had the highest partial correlation coefficient with overall quality difference ratings; *resonance* correlated significantly with the overall quality difference ratings in this study, while it didn't seem to have a high correlation with the preference ratings in Chapter 3. In both studies, *richness* was most valued by the subjects while evaluating the preference/overall quality of the violins, which was also consistent with the previous finding in [Saitis, 2012].

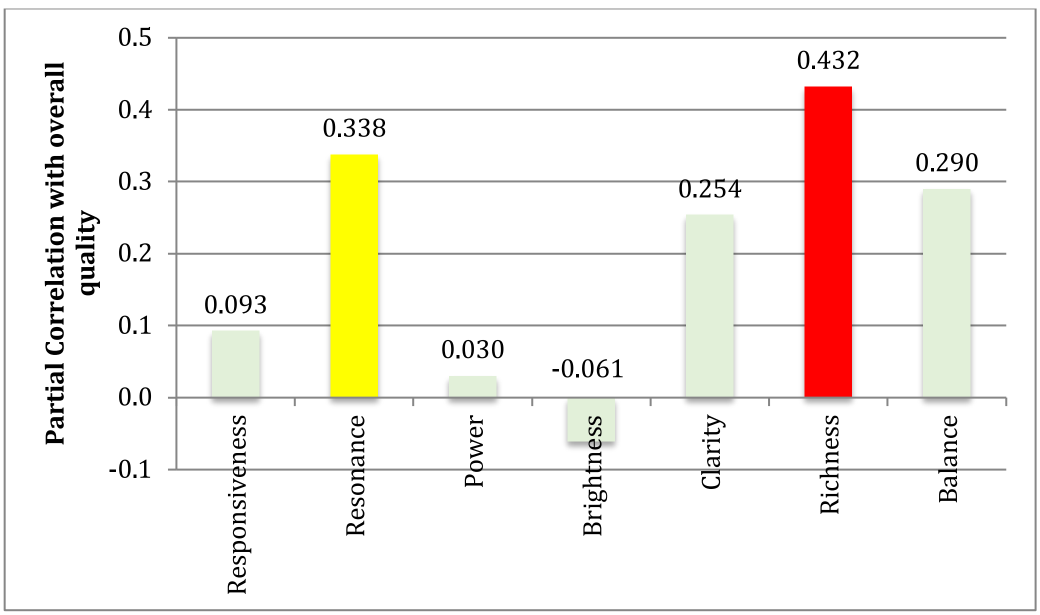


Figure 4.6 Partial correlation coefficient $\boldsymbol{\rho_p}$ between difference ratings of each attribute and overall quality.

4.4 Discussion and Conclusion

This study investigated how different strings affect the perception of violin quality through two carefully designed perceptual playing tests: one in Oberlin and the other in Montreal. In

Oberlin, players compared two types of strings: Dominant strings and Kaplan strings through two experimental conditions D1-D2 and D1-K2. In Montreal, subjects compared three types of strings: Dominant, Kaplan (same sets as in Oberlin) and Pro-Arté strings through three experimental conditions D1-D2, D1-K2 and D1-P2.

The differences between D1-D2 and D1-K2 were not statistically significant based on the Oberlin results. The differences among D1-D2, D1-K2 and D1-P2 were not statistically significant as well based on the Montreal results. If we compare every two experimental conditions based on the Montreal results, differences between D1-D2 and D1-K2, and D1-K2 and D1-P2 were not significant. However, the *brightness* difference ratings were found to be significantly higher in D1-D2 than in D1-P2. There were no significant differences between D1-D2 and D1-K2 even when we combined the results of the two parts of this experiment in Oberlin and Montreal: absolute value of paired samples $t\ (17) \leq 1.342$, $p \geq 0.197$; related-samples Wilcoxon signed rank test $z = -0.288$, $p = 0.773$.

The three types of strings involved in this experiment have different price levels: Dominant strings cost about \$78, Kaplan strings around \$108, and Pro-Arté strings around \$49. The result that the three experimental conditions lacked significant differences was unexpected. There are several possible influences and conclusions. First, the strings we chose for the experiment are widely used on violins and are generally considered to be of good quality. Therefore, the differences between the strings may not be significant enough to be perceptible when presented to players on relatively low-quality violins that are unfamiliar to them (in contrast to installing the strings on their own instrument). Second, the number of subjects that participated was small, though this is inevitable given the nature of this type of experiment due to the need for highly skilled players, scheduling, room availability and subject fee costs. But having a greater number of subjects could help reduce random error effects [Robson, 1994]. Third, for the experiments in Montreal, the strings were changed two times for each subject. Frequent changing accelerates the aging of the strings, which could lead to a variation of the string qualities for different subjects. As well, other violin setup conditions might be inadvertently modified when changing the strings (such as the bridge position). That said, it was decided to design the Montreal experiment as a single session to avoid problems getting subjects to return on subsequent days. As well, the Oberlin experimental design, with sessions separated by several days, has its own set of disadvantages.

Finally, violinists do not share the same interpretation for every rating criterion (despite the definitions we provided) and, there are large inter-individual variations in the criteria ratings, as illustrated by the large error bars in Figures 4.1 - 4.5. This is similar to what has been observed previously in playing tests (e.g. Saitis et al., 2012) and contributes to the lack of significance in our results. Differences in averages were not significant, which again implies that players did not agree with each other and so the differences became small when averaged. Strings may make a difference, but they may highly depend on the player.

We observed a few significant differences between the results of Oberlin and Montreal. D1-D2 *resonance* difference ratings and D1-K2 *power* difference ratings changed significantly from Oberlin to Montreal, respectively. This could be partly attributable to seasonal changes. However, none of the differences between D1-D2 and D1-K2 criteria difference ratings were found to be significantly different from Oberlin to Montreal. The seasonal changes may have affected similarly how the pair of violins was evaluated in the two conditions in Montreal compared to Oberlin, so that when looking at the differences between these two conditions, they are very similar in the two cities. Based on the experience of this experiment, deliberate and compromised choices have to be made during the experiment design, considering the number of professional violinists, the number of different types of strings to be tested, the different service life of different strings, the preservation of the test instruments during the experiment, and the time duration of the whole experiment.

We also examined the relationship between attribute ratings and overall quality ratings. *Richness* ratings and to a lesser extent, *resonance* ratings, were found to significantly correlate to overall quality ratings based on both the Oberlin and Montreal results. The finding that players tend to agree that *richness* is a determinant criterion in preference evaluations is in line with the conclusion of Chapter 3 and the findings of a previous study [Saitis, 2012].

Chapter 5

5 Perception of Violin Soundpost Height Differences: Playing Test

5.1 Introduction

In this and the subsequent chapters, we report the results of an investigation on the influence of the soundpost height on the violin perception. Most of the previous research studying the role of the soundpost is from physics and acoustics aspects. How the soundpost affects the perceptual qualities of the violin, however, has not been properly investigated. As often described by luthiers or players, a very subtle change to the soundpost dimension or position can result in significant variations in the violin quality, especially when the changes are around the optimal soundpost condition. Therefore, the aim of this study was to investigate correlations between a change in height of the soundpost and variations of the quality of the violin, as evaluated by players.

As described in Section 2.5.2, previous reported research on the violin soundpost primarily concerned its role (installed vs. removed) or general trends in its positioning. Our initial interest for a perceptual study was to investigate changes in soundpost position. However, results of a pilot study (conducted at the 2018 Oberlin Acoustics Workshop) demonstrated difficulties in accurate and reasonably fast repeated positioning at specific locations inside the violin soundbox, as well as the need for an experienced luthier to be present for the duration of the experiment. The availability of a commercially available height-adjustable carbon fiber soundpost instead offered the ability to study the perception of soundpost height changes. With a bit of practice, it was found that the height adjustments could be accomplished by one of the experimenters within a minute or less, thus obviating the need (and cost) for a luthier to be present throughout the experiments.

Violinists and luthiers were invited to evaluate the violin with different soundpost heights through playing tests with controlled experimental conditions. The first question to be explored was how big of a change in the soundpost height could result in perceivable variations in the violin qualities by violinists and luthiers. It is not known whether such just noticeable difference in height depends on the reference height. We decided to investigate this question around the subjects' optimal soundpost height, considering this was a more relevant question for makers, who usually make fine adjustments around a height that is already considered as good. Instead of choosing the same reference height for everyone, we decided to look at the just noticeable difference around the optimal height for each person; because we assumed, based on discussions with violin makers, that this is the range of heights where players may be the most sensitive. Indeed, they feel that players are more sensitive between something very good and something "just" good compared to something bad and something slightly less bad. Detailed materials and methods are described in Section 5.2. Section 5.3 presents the results and discussion. Conclusions are given in Section 5.4. We also conducted bridge admittance measurements on the violin for different soundpost heights and the results are then provided in Section 5.5.

5.2 Materials and Methods

5.2.1 General Design

This experiment explores how changes in soundpost height affect the perceptual qualities of the violin and whether there is a threshold of change below which players do not perceive differences. A violin installed with a height-adjustable carbon fibre soundpost was employed. The experiment was designed as a sequence of playing tests. An experimenter was present to change the soundpost height. Violinists and luthiers were invited to participate. The experiment involved two phases. During the first phase, subjects played and described their feelings about the violin with different soundpost settings in order to find their optimal soundpost height. During the second phase, the experimenter randomly increased, decreased or did not change the soundpost height in ten trials around their optimal height. For each trial, subjects were asked to play the violin, comparing it with the previous setting, and to decide whether they were the same or different.

Players were asked to use their own bows to play the violin and evaluate, as they typically use their own bows when testing violins in real life. Luthiers were given the option of either using their own bow if they play violin or to use a bow provided by us. This experiment took place in a room free of strong resonances and a relatively low reverberation time. The area of the experiment room was approximately 26.7 m^2.

5.2.2 Soundpost and Violin

A height adjustable carbon fibre soundpost (Anima Nova) was employed for this study, as shown in Figure 5.1. According to the description by the manufacturer (Anima Nova), the soundpost has flexible ball-and-socket joints at its two ends, which allow the soundpost to adjust automatically to every contour of the violin whilst distributing the pressure evenly over the contact area. The upper cylinder shell possessing a scale on its bottom is sheathed with the lower cylinder through an internal thread, and one can increase or decrease the soundpost height by turning the upper cylinder shell anticlockwise or clockwise. A vertical line indicated on the surface of the lower cylinder acts as the pointer of the scale. A height change is specified by a number of graduations. The minimum graduation value is 0.022 millimeter and the scale employs an octal number system. By turning one complete revolution (0 - 4.4), the soundpost height varies 0.8 mm. There are 5 numbers ranging from 0 to 4 on the scale. Between adjacent numbers, there are 8 minimum graduations. Through the special tools provided by Anima Nova, one can change the soundpost height without taking it out of the violin body. Two adjacent numbers can always be seen simultaneously from the f-hole. We tested the precision of the Anima Nova soundpost. The measurements started from a reading of 3.4 (actual height measured as 53.09 mm) and increased in increments of two graduations, for a total of 26 measurements (i.e., 53 graduations) to a height of 54.23 mm. Subsequently, the process was repeated with 26 decrements of two graduations each, back to the original setting of 3.4 (which was measured as 53.07 mm). By averaging the absolute differences of every soundpost height measured during the increasing and decreasing processes, we obtained an absolute average error of the soundpost height of about 0.007 mm.

Figure 5.1 Anima Nova height-adjustable soundpost.

The violin used in this experiment is a performance-level violin borrowed from Schulich School of Music, McGill University. We asked a local luthier to help replace the original wooden soundpost (around 53.77 mm high) with the Anima Nova height-adjustable soundpost, though subsequent adjustments were made by the experimenters. The height-adjustable soundpost was placed about 3.5-4 mm below the bridge and centered with the treble foot of the bridge according to the soundpost manufacturer's instruction. The soundpost was set initially at a relatively low height, approximately 53 mm.

5.2.3 Participants

Thirteen experienced violinists and six skilled luthiers participated in this experiment. Among the players, there were 8 females, 5 males; 7 native English speakers, 3 native Chinese speakers and 3 other native speakers. Their average age was 30 yrs (SD=9 yrs, range=21-54 yrs). They had at least 16 years of playing experience (mean=23 yrs, SD=7 yrs, range=16-40 yrs), and at least 8 years of training (mean=18 yrs, SD=4 yrs, range=8-26 yrs). They reported to play 23 hours per week on average (SD=10 hrs, range=6-37.5 hrs). Eleven players described themselves as professional violinists. One of the players was a doctoral candidate in music performance, 2 had master's degrees in music performance, 4 were master students in music performance, 3 had bachelor's degrees in music performance, 1 had a bachelor's degree in arts, and 2 were currently undergraduate students in music. They reported to play various types of music [classical (100%), contemporary (69%), jazz/pop (38.5%), baroque (23.1%), and folk (15%)]. 85% of them play in chamber music, symphonic orchestra or solo, respectively. One of the players play in Folk/Jazz band, pop band, chamber orchestra or work as a private music teacher, respectively. Among the luthiers, there were 4 males, 2 females; 3 native English speakers and 3 native French speakers. Their average age was 48.5 yrs (SD=11 yrs, range=36-61 yrs). They had at least 15 years of experience being a violin maker. Five luthiers played violin, among them there were 1 professional violinist, 2 advanced players and 2 beginners. All subjects were paid for their participation.

5.2.4 Detailed Procedure

This experiment consisted of two phases and lasted about 1 hour. Subjects were scheduled individually. Two experimenters were present during the experiment. One experimenter, who made adjustments to the soundpost, sat behind a table, with a screen in front to prevent subjects from observing the adjustments. The other experimenter helped with facilitating the experiment and taking notes for the subjects. During the first phase, the soundpost was initially set at a relatively low height, around 53 mm. Subjects were then asked to play the violin with this initial setting and describe their feelings. Then the experimenter increased the soundpost height by 8 graduations (about 0.176 mm) and the subjects repeated the playing and describing process. Subjects were informed that the experiment was about soundpost height modification using a height-adjustable soundpost before the experiment, however, they were not told which direction the experimenter was adjusting the soundpost height. Subjects were asked whether the modification made the violin better or worse compared to the previous setting and to provide a verbal description of their perception of the change. If the subjects felt the setup was better or the same, the experimenter would continue to increase the soundpost height for a few graduations: 8 graduations or 4 graduations. If the subjects stated that the setup was worse, the experimenter would decrease the soundpost height for a few graduations: a decrease of 2-4 graduations to somewhere in between the two previous soundpost heights. Then the subjects were asked to repeat the playing and evaluation procedure again. This process was repeated several times in order to find their most preferred height, with the number of graduations increased or decreased becoming smaller as the experiment continued. The whole process of searching for the most preferred height usually required from 5 to 9 trials. During the soundpost height adjustment, the experimenter made sure that the soundpost height did not exceed 53.66 mm (30 graduations higher based on the original soundpost height 53 mm) as not to damage the violin. Each soundpost height adjustment took about a minute or less to complete.

There was a 5-minute break between Phase 1 and Phase 2. During the second phase, the experimenter randomly increased, decreased or did not change the soundpost height in ten trials within a range of approximately ± 0.11 mm around their optimal height. Subjects were asked to play the violin during each trial and compare it with the previous setting, to decide whether they were the same or different. At the beginning of Phase 2, subjects were asked to play the violin with their optimal soundpost heights again. Then the experimenter increased, decreased, or did not change the soundpost height by different graduations over ten trials according to a plan determined in advance, which was unknown to subjects. The height variations are ΔH = 0, 0, 2, -2, 3, -3, 4, -4, 5, -5 graduations (actual height of ΔH = 0, 0, 0.044, -0.044, 0.066, -0.066, 0.088, -0.088, 0.11, -0.11 mm). They were randomized differently for each subject, while keeping the variations approximately within ± 0.11 mm around the subjects' optimal height. To minimize subject fatigue, there was a 5-minute break after five trials.

We can see that there are fewer "same" trials than "different" trials. This was decided based on the detection theory model which is used to analyze the data. In our case, this false alarm rate is calculated once and used for all possible height differences. In addition, just noticeable differences are normally obtained by comparing a series of stimuli to the same reference stimulus. Here, this would have been too tedious and tiring to always come back to the optimal soundpost height. The differences being small anyway, we thought it was better to just modify the height while always staying around this optimal height within a small range (+/- 5 graduations) and asked players whether there was a difference between this new height and the previous one. This constrained to some extent the order of the height differences (which was therefore only pseudo-random): for instance, a difference of +4 graduations could not follow a difference of +2 graduations, as this would have led to a height which was too far from the optimal height.

Thresholds were estimated using detection theory [Macmillan and Creelman, 2005]. As shown in Table 1, we have two stimulus classes. Height variations of $\Delta H = 0$ are class S_1, and $\Delta H = 2, -2, 3, -3, 4, -4, 5, -5$ are different cases of the S_2 class. A "Hit" is defined as a correct identification of an S_2 class element (participants recognize a height change); failing to identify it is a "Miss". A "False alarm" is defined as an incorrect identification of an S_1 class element (they thought the height changed when no variation of the soundpost height was made); correctly responding "same" is a "Correct rejection". Table 5.1 summarizes the four possible cases. The hit rate (H) is the proportion of different soundpost heights (S_2) to which the subject responds "different", and the false-alarm rate (F) is the proportion of the same soundpost height (S_1) similarly (but incorrectly) assessed. The hit and false-alarm rates can be written as the following conditional probabilities:

$$H = P\,(\text{“different”} \mid S_2) \tag{5.1}$$

$$F = P\,(\text{“different”} \mid S_1) \tag{5.2}$$

Table 5.1 Different responses for different stimulus classes.

Stimulus Class	Response	
	"Different"	"Same"
Different soundpost heights (S_2)	Hits	Misses
Same soundpost height (S_1)	False alarms	Correct rejections

The perceptual sensitivity is estimated using the *d'* measure: *d'* is defined in terms of the inverse of the normal distribution function *z*:

$$d' = z(H) - z(F) \tag{5.3}$$

The hit or false-alarm rate was thus converted to a z score (i.e., to standard deviation units) by the *z* transformation. The *z* transformation converts a proportion of 0.5 into a *z* score of 0, larger proportions into positive *z* scores and smaller proportions into negative *z* scores. Thus, when $H = F$, $d' = 0$ and the performance is at chance; when $H > F$, $d' > 0$, which means that subjects are able to recognize a difference in height. The sensitivity of detection increases as d' increases. When $H = 0.99$, $F = 0.01$, $d' = 4.65$: this is considered as an effective ceiling by many experimenters. Perfect accuracy implies an infinite d'. A strategy is to convert proportions of 0 or 1 into $1/(2N)$ or $1-1/(2N)$, respectively, where *N* refers to the number of trials that the proportions are based on. By calculating *d'* for each ΔH, we can estimate the sensitivity of the subjects in soundpost height variation.

The standard error of *d'* was calculated according to Gourevitch and Galanter (1967). The variance (square of the standard error) of *d'* is the sum of the variances of the two (independent) transformed proportions: $z(H)$ and $z(F)$. Gourevitch and Galanter showed that observed z scores are approximately normally distributed, with variance

$$var[z(p)] = \frac{p(1-p)}{N[\emptyset(p)]^2} \tag{5.4}$$

where $\emptyset(p)$ is the height of the normal density function at $z(p)$. Consequently,

$$var(d') = \frac{H(1-H)}{N_2[\emptyset(H)]^2} + \frac{F(1-F)}{N_1[\emptyset(F)]^2} \quad (5.5)$$

where N_2 and N_1 are the number of trials in stimulus class S_2 and S_1, respectively.

Values of the function $\emptyset$ can be computed:

$$\emptyset(p) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}z(p)^2} \quad (5.6)$$

By extending 1.96 standard errors above and below observed *d'*, we can obtain a 95% confidence interval around *d'*.

5.3 Results and Discussion

5.3.1 Optimal Soundpost Heights

During the first phase of the experiment, we found an optimal soundpost height for each subject. The optimal soundpost heights were represented relative to the original soundpost height (around 53 mm). Figure 5.2 shows the relative optimal soundpost height of each subject sorted from smallest (0.132 mm) to largest (0.616 mm). The minimum and maximum soundpost height variations that subjects evaluated were 0 mm and 0.66 mm relative to the original soundpost height (53 mm), respectively. Figure 5.3 displays the boxplots of the relative optimal soundpost height for all subjects, players and makers separately. The interquartile ranges of the relative optimal soundpost height for these three groups were 0.33 mm, 0.352 mm and 0.2805 mm, respectively. The interquartile range for makers was smaller than players, meaning that the optimal soundpost heights for makers were more concentrated, which we could also see from Figure 5.2. This could be due to the small number of maker participants. The median relative optimal soundpost height for makers (0.308 mm) was also lower than for players (0.396 mm).

Figure 5.4 displays the mean optimal soundpost height relative to the original soundpost height for all subjects, players and makers separately. Error bars of two-sided 95% confidence interval of the means are also displayed. The mean relative optimal soundpost height and SD for all subjects were 0.371 mm and 0.171 mm. The corresponding mean and SD for players and makers were 0.391 mm, 0.18 mm and 0.326 mm, 0.158 mm, respectively. Players had a higher mean relative optimal soundpost height than makers. The confidence interval error bar of the means for makers was very large, which might be partially due to the small number of maker participants. We compared the relative optimal soundpost height for players and makers by performing the independent-samples Mann-Whitney U test (not employing independent-samples *t*-test for the violation of normal distribution assumption). The results showed that the null hypothesis that the distribution of the relative soundpost height was the same across players and makers could not be rejected, $U = 28$, $z = -0.969$, $p = 0.368$.

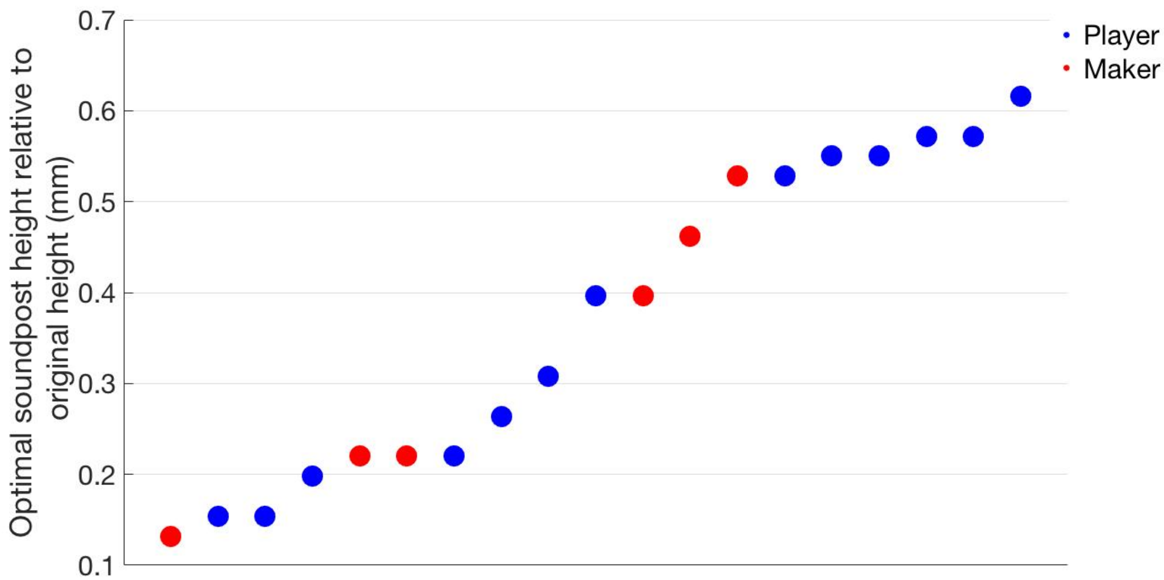


Figure 5.2 Optimal soundpost height relative to original height for every subject sorted from smallest to largest.

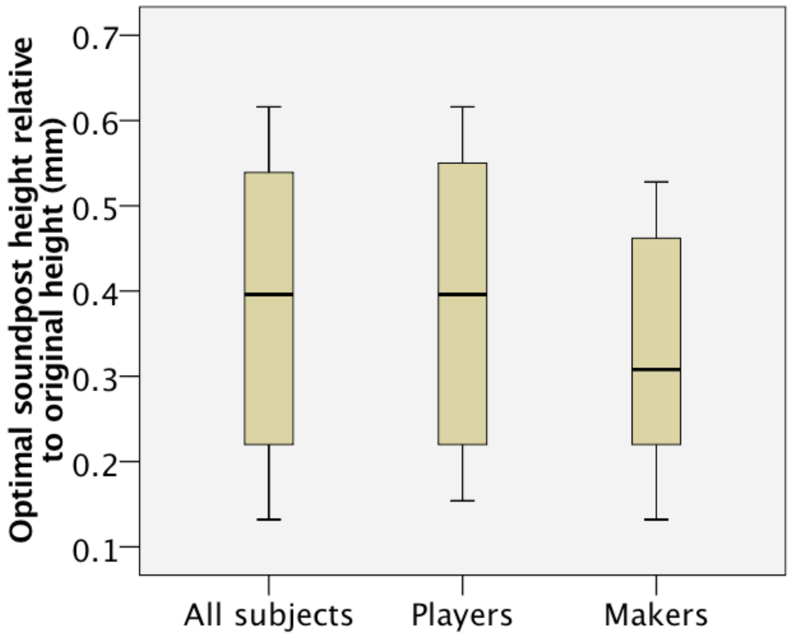


Figure 5.3 Boxplot of the optimal soundpost height relative to original height for all subjects, players and makers.

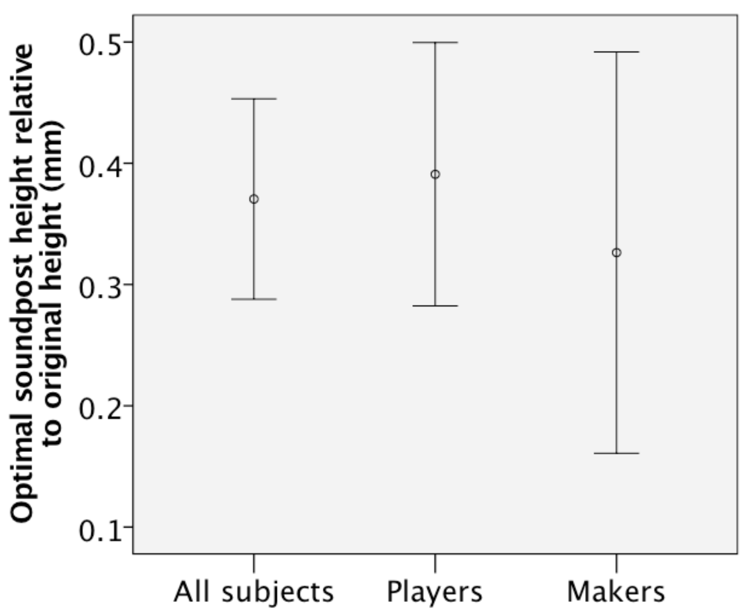
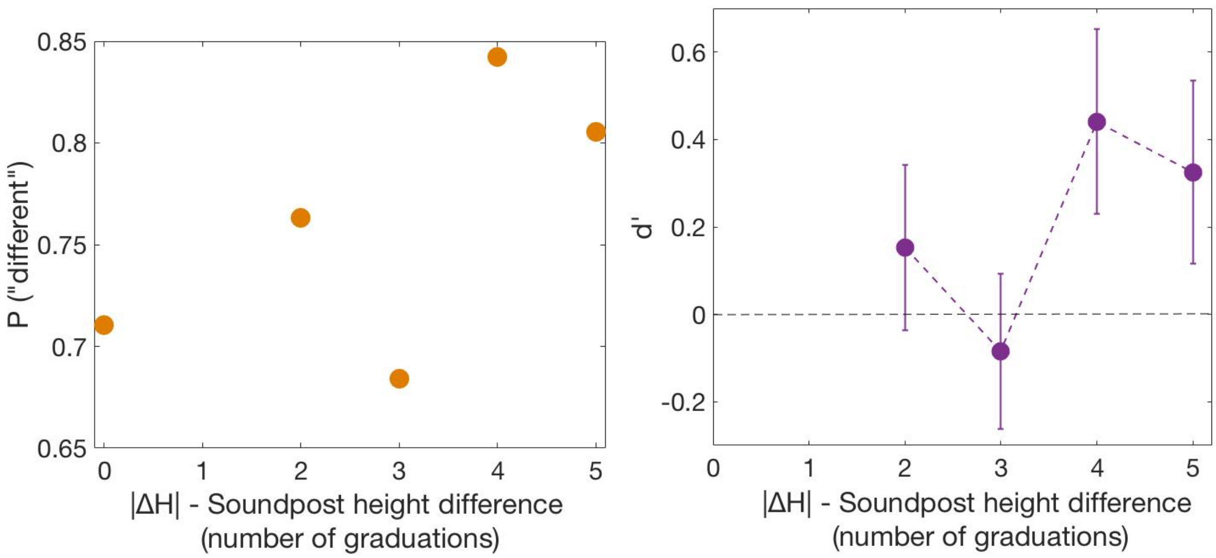


Figure 5.4 Mean optimal soundpost height relative to original height for all subjects, players and makers (error-bar = 95% confidence interval of the mean).

5.3.2 Perceptual Threshold of Soundpost Height Differences

As described in the detailed procedure, we can estimate the threshold of the soundpost height differences by calculating a sensitivity measure *d'* for each ΔH. During Phase 2 of this experiment, positive ΔH and negative ΔH were counterbalanced by randomizing the presentation of positive ΔH and corresponding negative ΔH for subjects. In addition, in order to increase the number of test trials (sample size) and estimate the threshold more precisely, we calculated *d'* for each $|\Delta H|$ instead of each ΔH. Thus, the number of trials for stimulus class S_1 was equal to each case of the stimulus class S_2, as there were two zero height variations among the ten height variations during the Phase 2 for each subject.

Figure 5.5 (a) shows the probabilities that subject considered the two soundpost heights with a height difference of $|\Delta H|$ as "different". We can see that the false alarm rate, which corresponded to *P* ("different") for $\Delta H = 0$ graduation was very high: 0.71. It was even higher than the hit rate for $|\Delta H| = 3$ graduations: 0.68. The highest hit rate was for $|\Delta H| = 4$ graduations: 0.84. The *d'* for each $|\Delta H|$ is shown in Figure 5.5 (b). Error bars of 95% confidence interval around the *d'* are also displayed. We can see that *d'* for $|\Delta H| = 4$ and 5 graduations were greater than 0, implying that subjects could recognize soundpost height changes of 0.088 and 0.11 mm at greater than chance level. It was however surprising that *d'* was much smaller for $|\Delta H| =3$ graduations (i.e., 0.066 mm) as, in this range of $|\Delta H|$, an increase of the sensitivity would have been expected with an increase of $|\Delta H|$.



(a) (b)

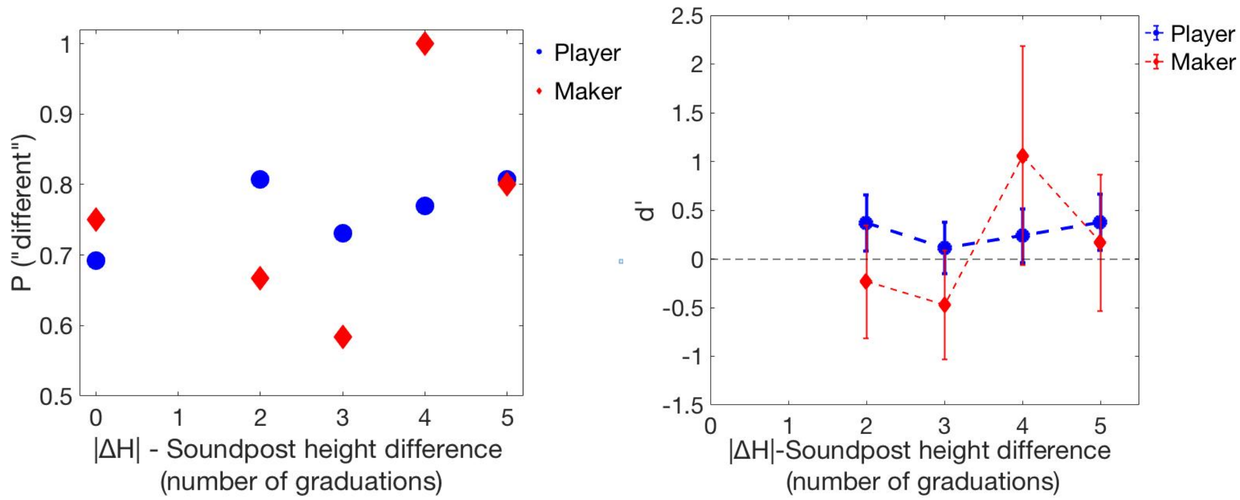
Figure 5.5 (a) Probabilities subjects considered the two soundpost heights with a height difference of $|\Delta H|$ as "different"; (b) perceptual sensitivity *d'* for each $|\Delta H|$ (error-bar = 95% confidence interval around *d'*).

In Figure 5.6, we compare the results for players and makers. Dividing the population into two groups can be problematic because it reduces the amount of data in each group (which was already low), especially in the group of makers. The following results may only be indicative. Figure 5.6 (a) displays the probabilities that players and makers considered the two soundpost heights with a height difference of $|\Delta H|$ as "different". The hit rate for $|\Delta H|$ = 4 graduations for makers was 100% (i.e., perfect accuracy). According to the explanation provided in Section 5.2.4, the proportion was converted to 1-1/(2N), where $N = 2 \times 6$ (number of makers), then the hit rate was converted to 0.96. We can see that the false alarm rate for makers (0.75) was higher than for players (0.69). The hit rates for makers varied more with $|\Delta H|$ than for players. The hit rates for $|\Delta H|$ = 2 and 3 graduations were much lower for makers than players, while the hit rate for $|\Delta H|$ = 4 graduations (0.96) was much higher for makers than players (0.77).

The corresponding *d'* for each $|\Delta H|$ of makers and players are shown in Figure 5.6 (b). Error bars of 95% confidence interval around *d'* are also displayed. We can see that for players, *d'* for $|\Delta H|$ = 2 and 5 graduations were a little bit greater than 0, 95% CI = [0.08, 0.65] and [0.09, 0.66] respectively, which implies that players could recognize height differences of 0.044 mm and 0.11 mm at above chance level. However, they were not able to recognize a height difference of 3 or 4 graduations. (i.e., 0.066 or 0.088 mm). The players' sensitivity to the soundpost height difference seems not to increase with an increase in the soundpost height difference. This could be explained by a few factors. First, the number of trials for each $|\Delta H|$ is relatively small (26, i.e. two by 13 players), and so the results could have been different with a few different answers, which could have easily happened with a different order of trials. Indeed, the task was very difficult (height variations within 0.11 mm), and so the fatigue may have reduced sensibility over the course of the experiment. The sensitivity index *d'* obtained can thus only be considered as estimates and should not be compared too closely. Secondly, though the physical increase in $|\Delta H|$ is linear, the effect on the playing quality may not necessarily be linear and therefore may not lead to an increase of sensitivity.

For makers, all *d'* were not greater than 0. The highest hit rate was for $|\Delta H|$ = 4 graduations: 1, which was converted to 0.96 to avoid infinite *d'*, therefore the corresponding *d'* became 1.06 with the resultant 95% CI = [-0.07, 2.18] having a just below 0 lower limit. The large confidence interval is due to the small number of trials (12, i.e., 2 by 6 makers) and the fact that the CI increases when the hit rate increases (this is explained in [Macmillan and Creelman, 2005]). As we can see that players had a lower hit rate (0.77) for $|\Delta H|$ = 4 graduations than makers and the corresponding *d'* = 0.24 had a less than 0 lower limit as well: 95% CI = [-0.04, 0.51], however, the *d'* for all subjects for $|\Delta H|$ = 4 graduations was greater than 0 with 95% CI = [0.23, 0.65] (see Figure 5.5 (b)). Therefore, we could say that makers were able to differentiate the soundpost height difference of 4 graduations, i.e., 0.088 mm in this experiment. And the first maker participant was not presented with $|\Delta H|$ = 5 graduations, making the number of trials even smaller (i.e., 10). Thus, the results were probably not robust.

These results show that players can detect smaller height differences than makers. This could be explained by the fact that professional players are very experienced at detecting subtle variations while playing. In contrast, makers, who are generally much less skilled in terms of playing than players, are more experienced in listening to players and modifying violins in response to what players say and want.



(a) (b)

Figure 5.6 (a) Probabilities players or makers considered the two soundpost heights with a height difference of $|\Delta H|$ as "different"; (b) perceptual sensitivity *d'* of each $|\Delta H|$ for players and makers, respectively (error-bar = 95% confidence interval around *d'*).

5.4 Conclusion about the Perceptual Experiment

In this experiment, we explored violinists' and luthiers' perception of violin soundpost height differences through a playing test. By employing a height-adjustable soundpost, we were able to find the optimal soundpost height for each subject and investigate the perceptual sensitivity to soundpost height differences around each subject's optimal soundpost height.

The results showed that the subjects' optimal soundpost heights varied from 0.132 mm to 0.616 mm relative to the original soundpost height (53 mm), reasonably well inside the extreme soundpost heights that were tested (0 mm and 0.66 mm). This shows that violin quality increased as the soundpost height increased, and started to decrease after certain height range, which is a phenomenon well known by makers. And the optimal soundpost heights for subjects vary within an interquartile range of 0.33 mm. The variation interquartile range was higher for players (0.352 mm) than for makers (0.2805 mm). The mean optimal soundpost height relative to the original soundpost height (about 53 mm) was also higher for players (0.391 mm) than for makers (0.326 mm). Statistical analysis showed that the differences of the relative optimal soundpost height for players and makers were not significant. Indeed, the fact that there was such a large range of optimal heights definitely showed that subjects did not agree on a "best" setting (even when only one parameter was changed). However, this range has to be interpreted with respect to the detection thresholds of players and makers. During the second phase of the experiment, the perceptual threshold of the soundpost height differences around each subject's optimal soundpost height was estimated through calculating a perceptual sensitivity measure of *d'*. The results for all subjects showed that subjects could recognize height changes of 0.088 mm and 0.11 mm at better than chance levels. Players could recognize height changes of 0.044 mm and 0.11 mm at above chance levels, and makers could recognize the height changes of 0.088 mm at above chance levels. The optimal soundpost heights of different subjects in this experiment varied within a range of 0.484 mm, which was greater than the minimum soundpost height differences that players and makers can detect, therefore confirming the reliability of the optimal soundpost heights that subjects reported.

Overall, the subjects performed at only a little bit greater than chance level in recognizing the differences we presented. And the false alarm rate was very high, i.e., subjects tended to say "different" even though there was no change at all in the soundpost height. That might partly be due to the sequential nature of the trials (they could not compare the different settings at the same time) and thus they might forget what the previous setting was like (though it only took a minute or less to make the soundpost changes). As well, there was a significant amount of variation in their organization approach to violin evaluation. Some subjects used a very consistent set of playing materials for each trial, while others used either very limited or changing materials between trials. Makers were in general significantly less skilled than players and thus may not have been able to "explore" the full range and capabilities of the violin. Additionally, the variation of soundpost height was quite small (within ± 0.11 mm), which made the task very difficult and could have contributed to player fatigue, so perhaps the true perceptual threshold is beyond that range. Finally, there was some imprecision in soundpost adjustments, with an absolute average error of 0.007 mm. All these factors could have had an effect on our results.

5.5 Bridge Admittance Measurements

We measured the bridge admittances of the violin used in this playing test for every soundpost height that had been evaluated by subjects during the first phase of finding optimal soundpost heights. The violin was tuned and damped during the measurements. As we needed to adjust the soundpost height, the measurement setup was different compared to that used for the study reported in Chapter 3. The violin was laid flat on a specially built structure, and the structure was placed on a laboratory table. The violin was clamped where the chin rest was normally located (the chin rest was removed). The neck of the violin was strapped onto a wooden support with a layer of thick foam around it, which resembled the player holding the violin. The shoulder rest supported the violin at the lower bout position. Similarly, as in Chapter 3, the bridge was excited with a miniature force hammer (PCB 086E80) and the resulting velocity was measured by a laser-Doppler vibrometer (Polytec PDV 100), both from the G-string corner. Thus, the experimenter could adjust the soundpost height from the treble-side f hole through particular tools provided by the soundpost manufacturer. For each soundpost height, we performed 3 to 5 measurements, and the results were averaged. The measurements were conducted in a lab with an area of approximately 30 m^2.

The results are displayed in Figure 5.7. The soundpost height was represented relative to the original soundpost height (about 53 mm) by number of graduations, i.e., the original height was 0 graduation. From the original soundpost height to the highest soundpost height (30 graduations), the soundpost height was denoted by gradient colour from blue to magenta. From the figure, we see that the admittances are very similar, with a slight decrease in magnitudes with increasing soundpost height for peaks up to about 1100 Hz. Figure 5.8 shows an enlarged Figure 5.7 particularly between 200 and 700 Hz with three signature modes identified and labeled: A0, B1- and B1+.

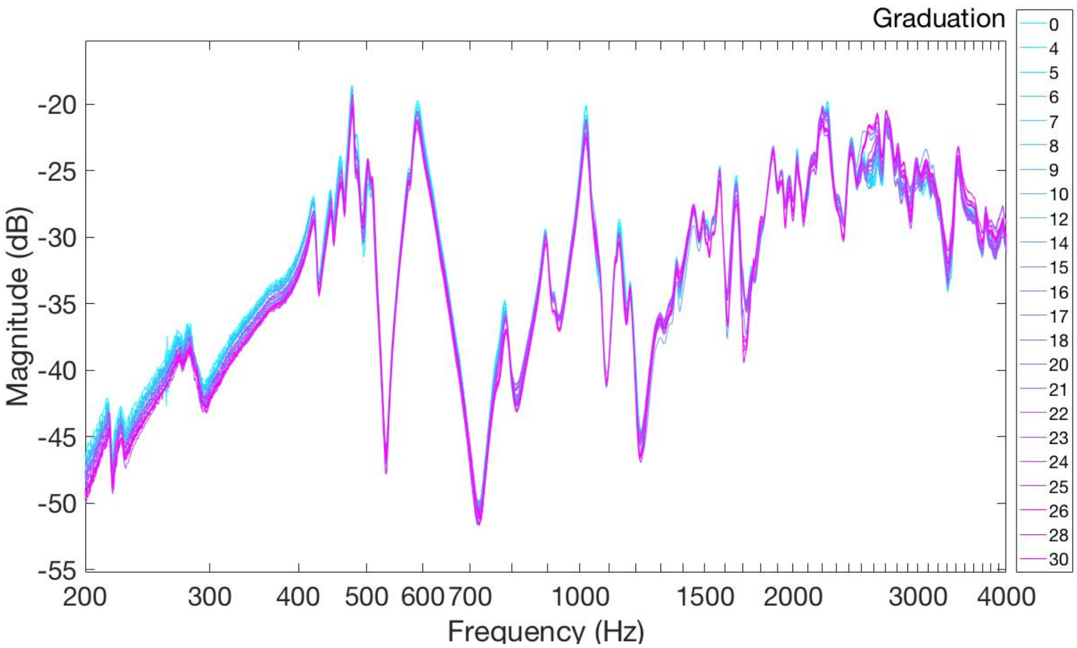


Figure 5.7 Bridge admittance measurements on the violin used in the soundpost height playing test for different soundpost heights. Soundpost height was represented relative to the original soundpost height (about 53 mm) by number of graduations. The colour scheme is shown in the legend: from the original soundpost height (0 graduation) to the highest soundpost height (30 graduations), the soundpost height was denoted by gradient colour from blue to magenta.

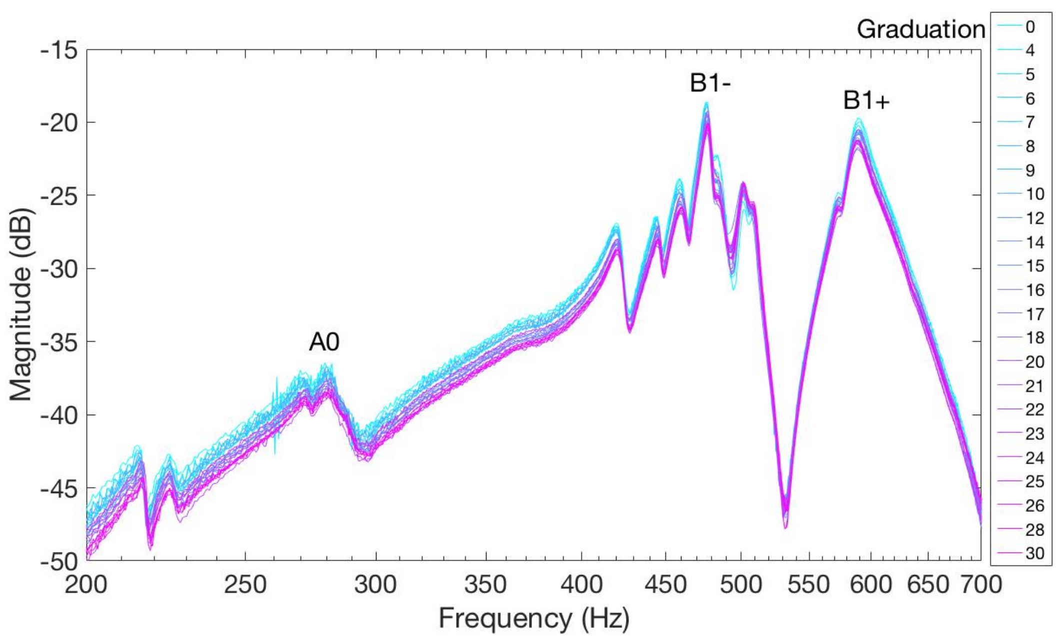


Figure 5.8 Enlarged Figure 5.7 particularly between 200 and 700 Hz with three signature modes identified and labeled. Soundpost height was represented relative to the original soundpost height (about 53 mm) by number of graduations. The colour scheme is shown in the legend: from the original soundpost height (0 graduation) to the highest soundpost height (30 graduations), the soundpost height was denoted by gradient colour from blue to magenta.

For a more detailed examination of the frequency and magnitude variations for different soundpost heights, we used a mode extraction routine [Maestre et al., 2016] applied to the admittance peaks we expect to be the A0, B1- and B1+ signature modes (a full modal analysis would be required to verify these). The mode extraction routine was the same as was used in Chapter 3. The mode center frequencies and amplitudes were estimated by parabolic interpolation of the peaks of the frequency-domain admittance data. The magnitudes of the three modes for all soundpost heights are shown in Figure 5.9. We can clearly see the general tendency that the magnitudes for all three modes decreased as the soundpost height increased and this observation was more obvious for the A0 mode. This makes sense for the structural modes (B1- and B1+), as the overall mobility of the top plate would be decreased as the tension on it increases (as the soundpost height increases). From the original soundpost height to the highest soundpost height, the magnitude decreased for A0, B1- and B1+ mode about 1.96, 1.33 and 1.5 dB, separately. As determined in the perceptual experiment, subjects disliked the extreme soundpost heights they evaluated. Therefore, the violin quality did not associate with the magnitudes of the three violin modes linearly or positively across the full range of soundpost height we tested. The finding in Chapter 3 that the magnitudes of the A0 and B1- modes correlated with violin quality positively should be confined to a certain range.

The variations of the frequency for these three modes were subtle, so we displayed them separately in Figures 5.10 (a) for A0, (b) for B1- and (c) for B1+ mode. The frequencies of the A0, B1- and B1+ modes varied within a range of about 2.4, 1.3 and 2.2 Hz, respectively, for the soundpost heights we measured. These are very small frequency shifts that may not be greater than the measurement error. The frequency for A0 mode stayed consistent generally within about 1 Hz. The unusual low frequency for the highest soundpost height was due to the identification error as there were too many peaks around the A0 mode. The frequency for B1- mode displays a general increase tendency with the increase in the soundpost height except for a few fluctuations. The frequency for B1+ mode seems to show a general decrease tendency with the increase in the soundpost height, though the fluctuations were relatively large. As for the magnitudes, the violin quality did not correlate with the frequencies of the violin modes B1- and B1+ linearly across the full range of soundpost height we tested.

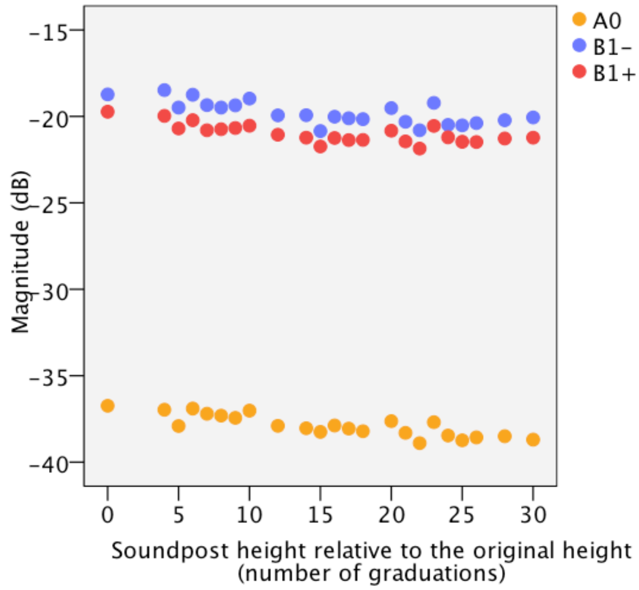
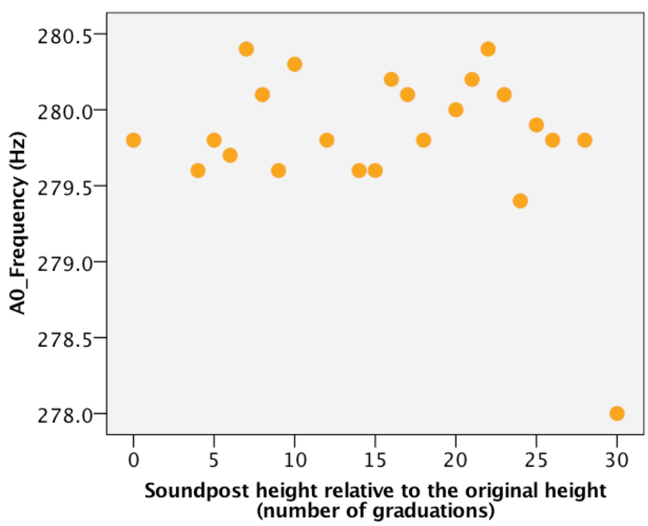
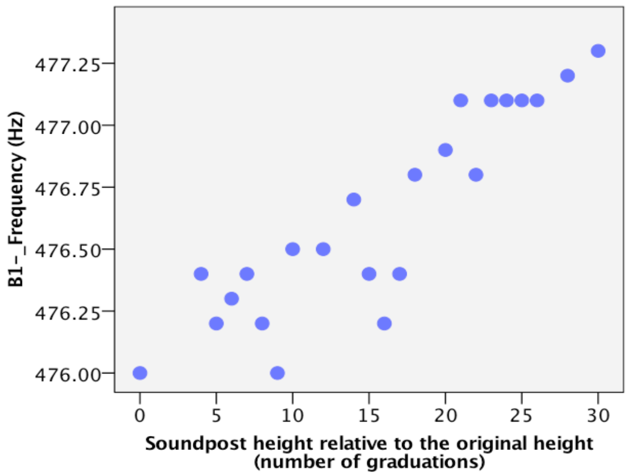


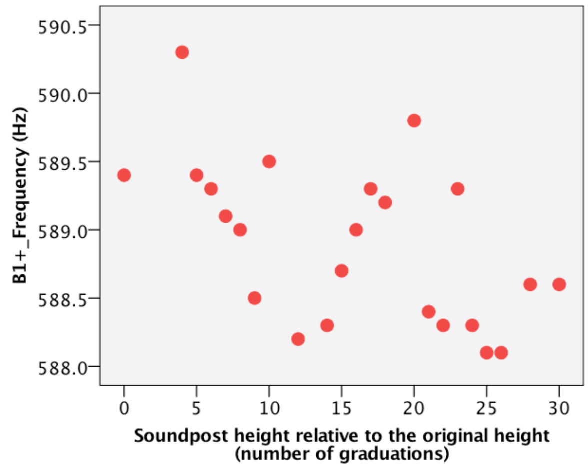
Figure 5.9 Magnitude for the A0, B1- and B1+ modes measured at different soundpost heights, ranging from 0 to 30 graduations (0.66 mm) relative to the original height of about 53 mm.



(a)



(b)



(c)

Figure 5.10 Frequency for the A0, B1- and B1+ modes measured at different soundpost heights, ranging from 0 to 30 graduations (0.66 mm) relative to the original height of about 53 mm.

Chapter 6

6 Perception of Violin Soundpost Height Differences: Listening Test

6.1 Introduction

In Chapter 5, we explored the perceptual threshold of violin soundpost height differences through a playing test. The advantage of the playing test is that subjects can explore the violin with different soundpost heights by themselves, through which they were able to evaluate the violin sound quality as well as playability. They could play freely in different registers of the violin and try various playing techniques to assess the violin's response. The disadvantage of the playing test is that subjects cannot compare two different soundpost height settings simultaneously and thus they might forget what the previous setting was like. Moreover, if they varied the way they played for the different height settings, that could also influence their judgement. For these reasons, we decided to conduct a listening test using recordings made on a violin with different soundpost heights to examine the perceptual threshold of violin soundpost height differences. The detailed materials and methods are presented in Section 6.2. Results and discussion are detailed in Section 6.3. Section 6.4 summarizes the main findings in this experiment and compares them with the playing test in Chapter 5.

6.2 Materials and Methods

6.2.1 General Design

The same height-adjustable carbon fibre soundpost as in the playing test was employed in this experiment. We recorded a soloist playing the same musical excerpt on a violin with different soundpost heights. Violinists and luthiers were invited to listen to the recordings and provide perceptual feedback using a computer interface. Tasks included "same" or "different" pairwise comparisons.

6.2.2 Recordings

Recordings were made in an acoustically dry room to avoid having reverberant effects mask the violin sound. The surface area of the room was approximately 20 m^2. A Sony PCM-D100 was used for the recording. The recorder was placed on a one-meter high table facing the violinist. The table was approximately 1.5 meters away from the violinist. The recordings were saved in 16-bit WAV format. The recording position was chosen to produce a natural quality of sound.

The violin used for the recordings was not the same as the one used in the previously described playing experiment. Before the recording, a violin maker helped replace the original wooden soundpost (around 54.88 mm high) with the Anima Nova height-adjustable soundpost. The height-adjustable soundpost was placed about 3.5 mm below the bridge and 2 mm inside the right end of the treble bridge foot according to the soundpost manufacturer's instruction. It was set initially at a relatively low height (around 54.11 mm, reading on soundpost was 2.0). A professional violinist was invited to help make the recordings. The recorded violin was made and provided by a luthier, who was also present to adjust the height of the soundpost. The recording excerpt was based on the Max Bruch violin concerto and was an "abridged" version specifically for the listening test (see Figure 6.1). The particular excerpt was chosen also by the player because it incorporated the main register of the violin.

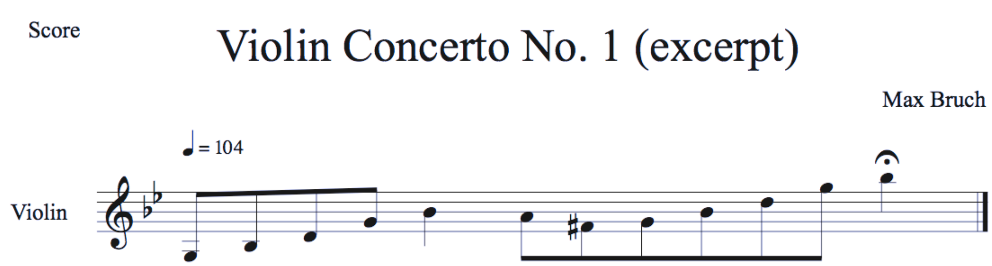


Figure 6.1 Recording excerpt.

During the recordings, the soundpost height was first increased in increments of 3 graduations (0.066 mm) over four iterations, up to a setting (3.4 or 54.374 mm) that was considered "tight" (see details in Table 6.1). Then the soundpost was lowered to a reading of 2.2 (54.154 mm), below the height that the player preferred in the first procedure (2.3 or 54.176 mm), and subsequently increased by increments of 1 graduation (0.022 mm) up to a setting of 2.5 (54.220 mm). Larger increments were used during the first soundpost height increasing procedure to search for the range that was considered "good", and smaller increments were used during the second soundpost height increasing procedure to find the closest "best" soundpost height.

Table 6.1 Recording process and recorded soundpost heights (sorted by soundpost height).

1st soundpost height increasing procedure (reading on soundpost)	2nd soundpost height increasing procedure (reading on soundpost)	Actual height (mm)
2.0		54.110
	2.2	54.154
2.3	2.3	54.176
	2.4	54.198
	2.5	54.220
2.6		54.242
3.1		54.308
3.4		54.374

6.2.3 Participants

Thirteen experienced violinists and eight skilled luthiers participated in this experiment. Twelve of the subjects had participated in the playing test reported in Chapter 5.

Among the players, there were 8 females, 5 males; 4 native English speakers, 4 native Chinese speakers, 3 native French speakers, 1 native English and French speaker and 1 native Italian speaker. Their average age was 32 years (SD=9.7 years, range=22-54 years). They had at least 16 years of playing experience (mean=26 years, SD=8.8 years, range=16-40 years), and at least 8 years of training (mean=18 years, SD=4.6 years, range=8-23 years). They reported playing 27 hours per week on average (SD=12 h, range=5-50 h). Eleven players described themselves as professional violinists. One player had a soloist diploma, one had a doctoral degree, two were doctoral students in music performance, 3 had master's degrees in music performance, 1 was a master student in music performance, 2 had bachelor's degrees in music performance, 1 had a bachelor's degree in arts, and 2 were currently 4th-year undergraduate student in music performance. They reported to play various types of music [classical (100%), contemporary (69.2%), baroque (46.2%), jazz/pop (7.7%)]. 92% of them played in chamber music, 84.6% solo and 77% in symphonic orchestra. Three of the players played in pop band, chamber orchestra or work as a private music teacher, respectively.

Among the luthiers, there were 6 males, 2 females; 4 native English speakers, 4 native French speakers. Their average age was 45.7 years (SD=13, range=29-61). They had at least 10 years of experience being a violin maker (mean=23.7, SD=11, range=10-40). Six luthiers played the violin, among them there were 1 professional violinist, 3 advanced players, and 2 beginners. All subjects were paid for their participation.

6.2.4 Stimuli

The stimuli were created based on each recording made at each soundpost height. The mean stimulus exposure duration was around 4.66 s. They were presented on a laptop in a relatively quiet environment, via Sennheiser HD 280 pro headphones, which was chosen for low distortion and diffuse-field response. We also used an Apogee Duet external audio interface for enhanced sound quality.

6.2.5 Detailed Procedure

We selected 9 recordings as stimuli, eliminating the recordings with playing artifacts or unexpected string ringing or noise in order not to affect the comparison, as shown in Table 6.2. The stimuli are sorted by the soundpost height. For each stimulus, the soundpost reading, actual soundpost height and corresponding recording procedure iteration number are provided.

Table 6.2 Recording chosen as stimuli sorted by the soundpost height.

Stimuli	Reading on soundpost (number in parentheses represents different recordings)	Recording procedure	Actual height (mm)
1	2.0	1	54.110
2	2.2 (1)	2	54.154
3	2.2 (2)	2	54.154
4	2.3 (1)	1	54.176

5	2.3 (2)	2	54.176
6	2.4	2	54.198
7	2.6	1	54.242
8	3.1	1	54.308
9	3.4	1	54.374

Ten pairs were then chosen for pair comparisons as displayed in Table 6.3. Pairs 1(1) and 1(2) were two pairs of identical recordings. Pairs 2(1) and 2(2) were two pairs of different recordings of the same soundpost heights. The two recordings in pair 2(1) were made during the same soundpost increasing procedure, while the two recordings in pair 2(2) were made during the two different soundpost increasing processes (see Section 6.2.2 for details of the recording process). There were two tests for each subject with a 5-minute break in between. Subjects listened to 8 pairs of stimuli in each test. Pairs 3, 4, 5, 6, 7, 8 existed in both tests with stimuli order in each pair switched for the two tests. Pairs 1(1) and 1(2) existed in two different tests separately and randomly, and this was the same case for pairs 2(1) and 2(2). The presentation order of the pairs in each test and the stimuli in each pair were randomized for each subject.

Table 6.3 Pairs of stimuli used in the listening test.

Pair	Height difference (number of graduations)	Height difference (mm)	Stimuli (number indicated in Table 6.2)	
1(1)	0 (identical recordings)		6	6
1(2)	0 (identical recordings)		7	7
2(1)	0 (different recordings of the same soundpost height)		2	3
2(2)	0 (different recordings of the same soundpost height)		4	5
3	2	0.044	3	6
4	3	0.066	1	4
5	4	0.088	1	6
6	6	0.132	4	8

7	9	0.198	4	9
8	12	0.264	1	9

Subjects were given instructions before the listening test as shown in the following:

> *In this listening test, you will be presented with 2 tests, each consists of 8 pairs of recordings.*
>
> *Each pair of recordings will be played twice in the order of* ***A-B-A-B*** *and you will only be able to click each pair a total of 2 times.*
>
> *Please indicate whether you think* ***the violin setup in each of the two recordings is the same setup or a different setup****. If you think it is different, please* ***describe the differences****.*
>
> *These recordings were made by one player on the same violin. Different adjustments were made to the violin between recordings and the player was asked to perform the excerpt as similarly as possible each time, but there still may be slight changes in his playing (tempo, articulation, dampening of the strings…). Please ignore the differences in his playing and decide whether the setup is the same or different.*

Subjects who had participated in the playing test were aware that the "setup" indicated in the instructions meant soundpost height. For consistency, we told subjects who did not take part in the playing test previously that the "different setup" meant "different soundpost height", and showed the Anima Nova soundpost leaflet to them and explained the soundpost working principle.

Before the formal listening test, we played all the stimuli to the subjects so that they could get an initial impression of the recordings. The differences were quite subtle and difficult to hear according to several pilot studies we conducted previously. This was also the reason we put two trials together (A-B-A-B), as two iterations of each pair seemed necessary.

During the formal listening test, subjects performed pairwise comparisons of the stimuli through a computer interface created in Matlab. The initial screenshot of the interface is shown in Figure 6.2. An instruction was displayed on top of the interface “Push the button to listen to a pair of sounds. Please ignore variations in playing technique.” When subjects clicked the “PLAY” button, the pair of recordings were played twice in the order of A-B-A-B (i.e. two trials together). During the playing, the name of the stimulus (A or B) on playback was displayed on the initial “PLAY” button to help the subject track which stimulus was playing, see Figure 6.3. There was a 60 ms duration after the subject pressed the response key and before the 1st stimulus (A) started to play. The interstimulus interval (ISI) was about 0.72 s. The inter-trial interval (ITI) was about 1.28 s. The interstimulus interval and inter-trial interval were determined through testing by the experimenter and pilot studies. Subjects were able to click the “PLAY” button a second time, after which it was disabled and grayed out. They then click the corresponding option shown on the interface “Same Setup” or “Different Setup”. If they clicked “Same Setup”, the button at the bottom of the interface “NEXT” was activated with the colour of the word “NEXT” turned from gray to black. Subsequently, they could click the “NEXT” button to move on to the second pair. The test progress was shown beside the “NEXT” button by displaying the current pair and the total number of pairs in each test (8 pairs). If the subject clicked the “Different Setup”, a text box appeared on the interface which is shown in Figure 6.4 with the instruction line of “please describe how they are different” above. After they entered their answer, the subjects clicked the “NEXT” button to move to the next pair. After finishing all the 8 pairs in test 1, subjects were asked to take a 5-minute break. Afterwards, the experimenter instructed subjects to take test 2, the process of which was identical to test 1.

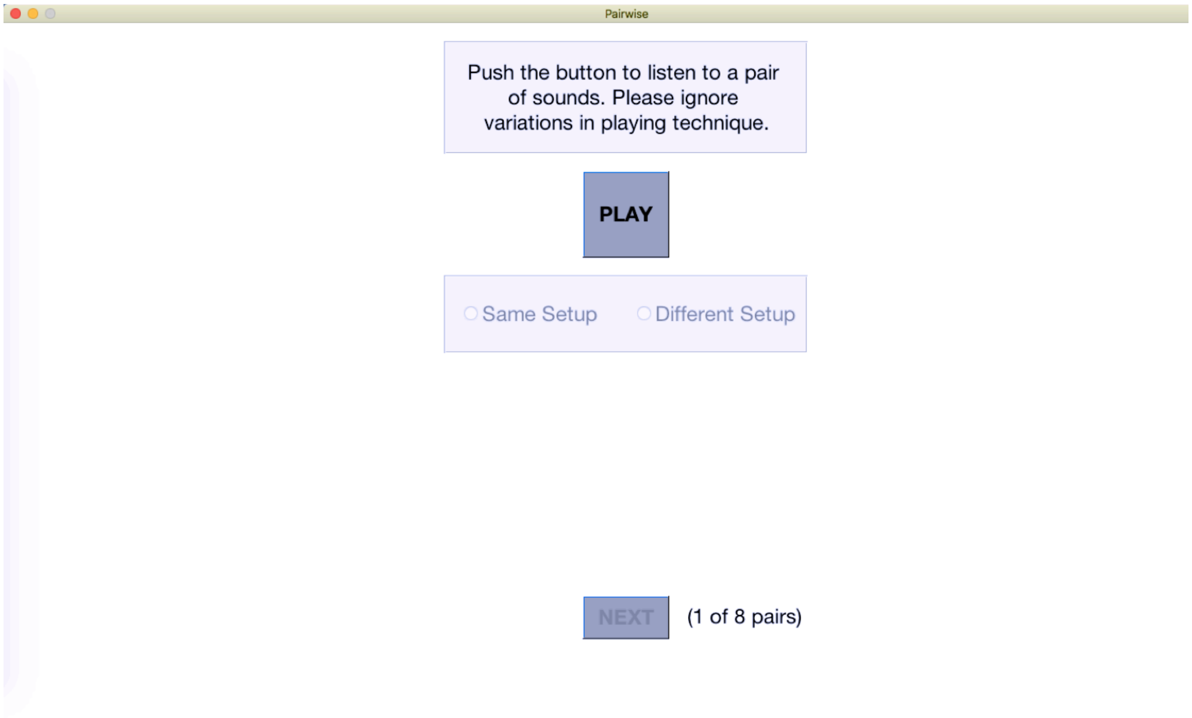


Figure 6.2 Initial screenshot of the listening test interface.

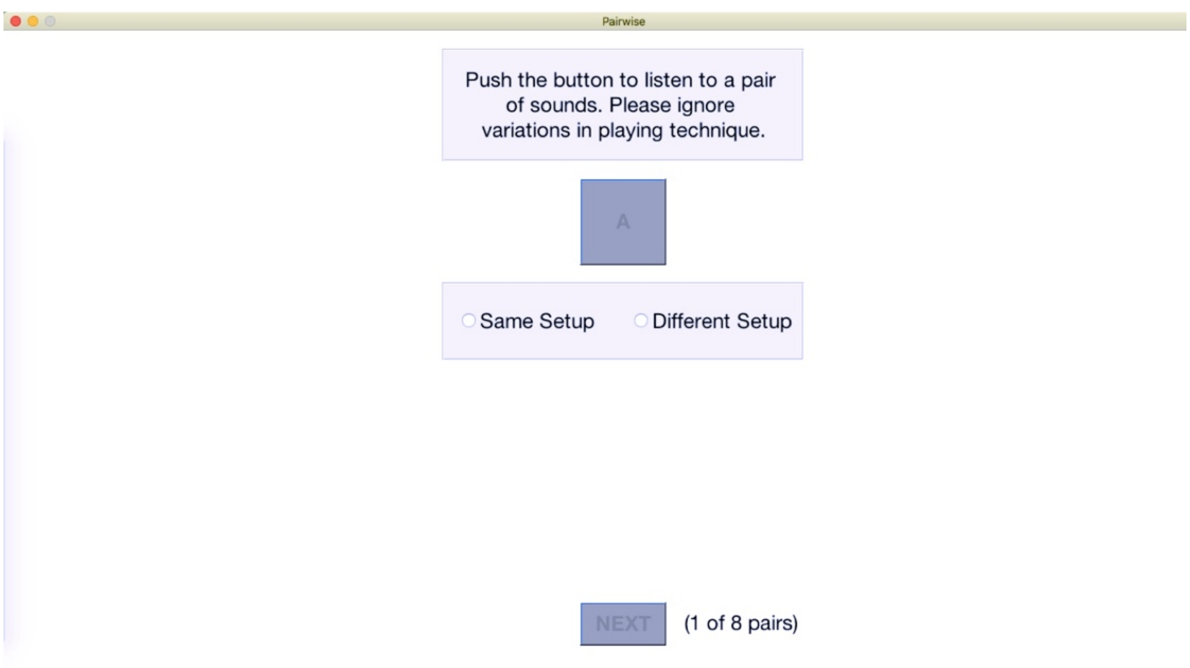


Figure 6.3 Screenshot of the listening test interface during the playing of stimulus A.

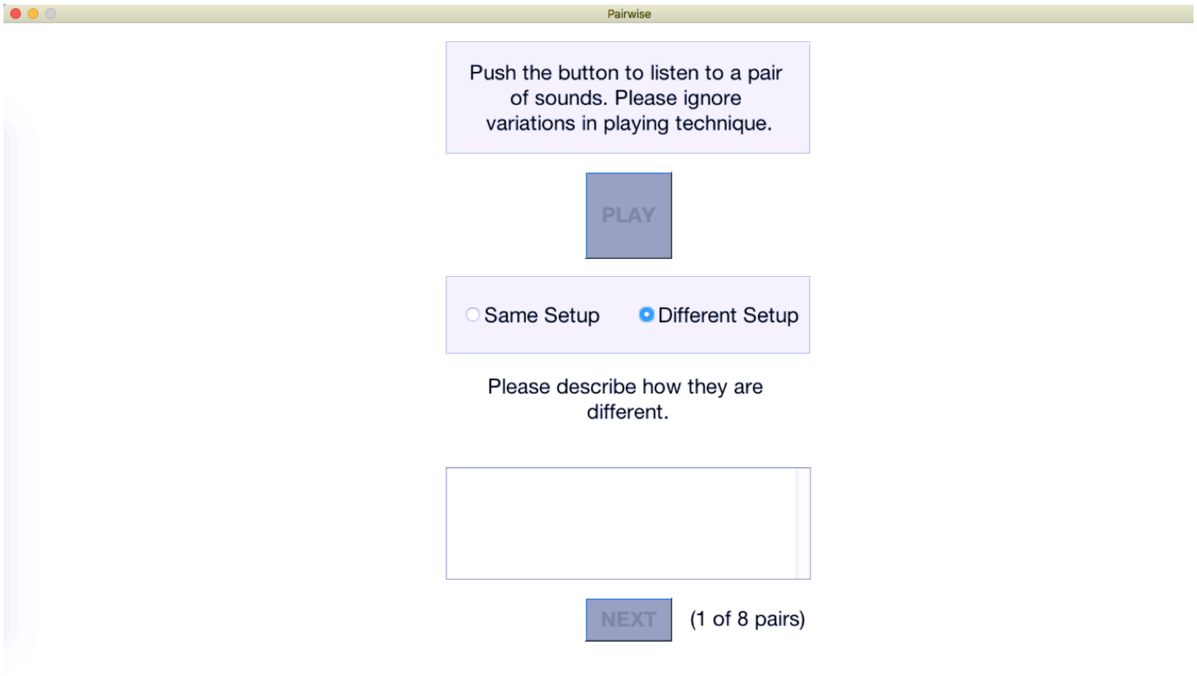


Figure 6.4 Screenshot of the listening test interface when subject clicked "Different Setup".

The perceptual sensitivity was estimated employing the same estimation method as in the playing experiment in Chapter 5. However, in this case there were two different types of reference stimulus classes (S_{1a} and S_{1b}), as shown in Table 6.4; and for each type of the reference stimulus class, we employed two reference pairs, which were shown in Table 6.3. The additional type of reference stimulus class, S_{1b}: "different recordings of the same soundpost height", was added to estimate the effects of possible slight variations in playing technique. That is, even though subjects were told that they should ignore variations in playing technique, there was still the possibility that slight playing variations could be interpreted as different soundpost heights. The reason to have two reference pairs in each reference stimulus class was to include more types of differences that could be caused by playing technique.

Table 6.4 Different responses for different stimulus classes.

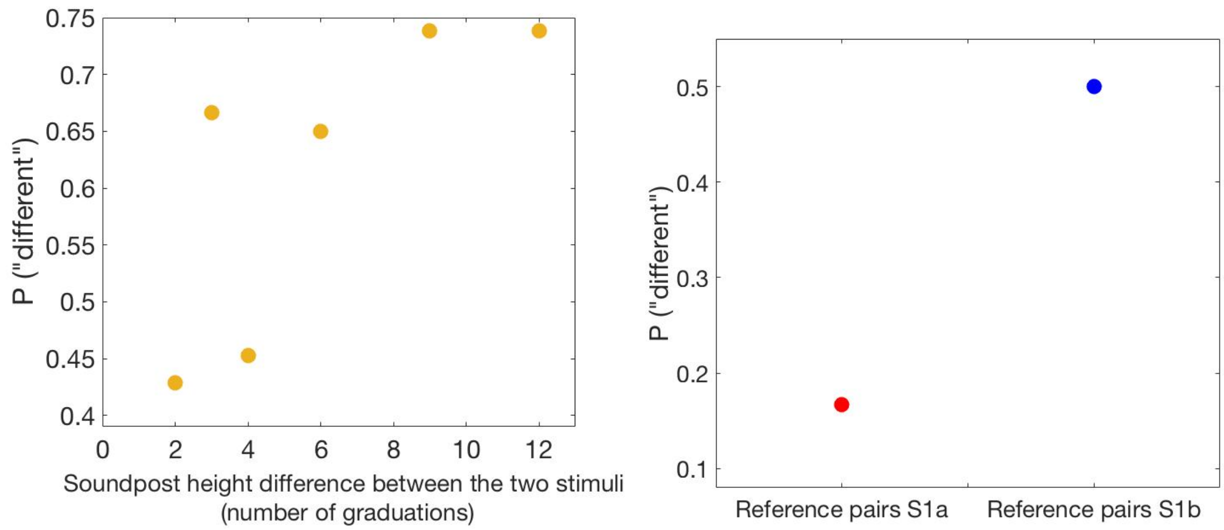
Stimulus Class	Response	
	"Different"	"Same"

Different soundpost height recordings (S_2)	Hits	Misses
Identical recording (S_{1a})	False alarms	Correct rejections
Different recordings of the same soundpost height (S_{1b})	False alarms	Correct rejections

6.3 Results and Discussion

6.3.1 Perceptual Threshold of Soundpost Height Differences

As in Chapter 5, the threshold of the soundpost height differences in this experiment was estimated through the calculation of a sensitivity measure *d'* for each $|\Delta H|$. Figure 6.5(a) shows the probabilities (also known as hit rates) that subjects considered the two soundpost heights with a difference of 2, 3, 4, 6, 9 or 12 graduations as "different", respectively. We can see a trend that the hit rate increased as the soundpost height difference increased except that the hit rate for 3 graduations was extraordinarily higher than its adjacent intervals. Figure 6.5(b) displays the probabilities (also known as false alarm rates) that subjects considered the first type of reference pairs (S_{1a}) that consisted of identical recordings (pair [6, 6] and [7, 7]) and the second type of reference pairs (S_{1b}) that consisted of different recordings of the same soundpost heights (pair [2, 3] and [4, 5]) as "different", respectively. We can see that the false alarm rates of the reference pairs S_{1a} were much lower than the false alarm rates of the reference pairs S_{1b}. Among the two reference pairs of S_{1b}, the false alarm rate of pair [4, 5] ($P = 0.619$) was much higher than pair [2, 3] ($P = 0.381$). Looking back to Table 6.2, we can see that stimuli 2 and 3 were recorded during the same soundpost increasing process, while stimuli 4 and 5 were recorded at different phases of the soundpost height adjustment procedure and thus at significantly different times. Recording in different soundpost height increasing processes may contribute to the inaccuracies of the soundpost height determination, thus enlarging the differences by adding differences due to the setup to the playing differences.



(a) (b)

Figure 6.5 Probabilities that subjects considered each pair of stimuli as different, when the pairs consisted of recordings made at two different heights (a) or at the same height (b).

Figure 6.6 shows the perceptual sensitivity measure *d'* for each $|\Delta H|$ calculated based on the two types of reference pairs, respectively. In addition, we also calculated the perceptual sensitivity of reference pairs S_{1b} (i.e., $|\Delta H|=0$) based on the reference pairs S_{1a}. Error bars of two-sided 95% confidence interval (CI; all CIs are two-sided 95% intervals through this chapter) around *d'* are also displayed. We can see that all *d'* that were calculated based on the reference pairs S_{1a} were greater than 0, including the *d'* for the reference pairs S_{1b}. It implies that subjects could differentiate all different stimuli we presented at above chance levels. However, we couldn't conclude that they could recognize all soundpost height differences because they didn't manage to distinguish the differences caused by soundpost height from those by playing technique. For *d'* that was calculated based on the reference pairs S_{1b}, we can see that all *d'* were greater than 0 except the *d'* for soundpost height differences of 2 and 4 graduations. This means that subjects could differentiate different soundpost heights with a difference of 3, 6, 9 or 12 graduations (i.e., 0.066 mm, 0.132 mm and greater) at above chance levels based on the reference pairs S_{1b}.

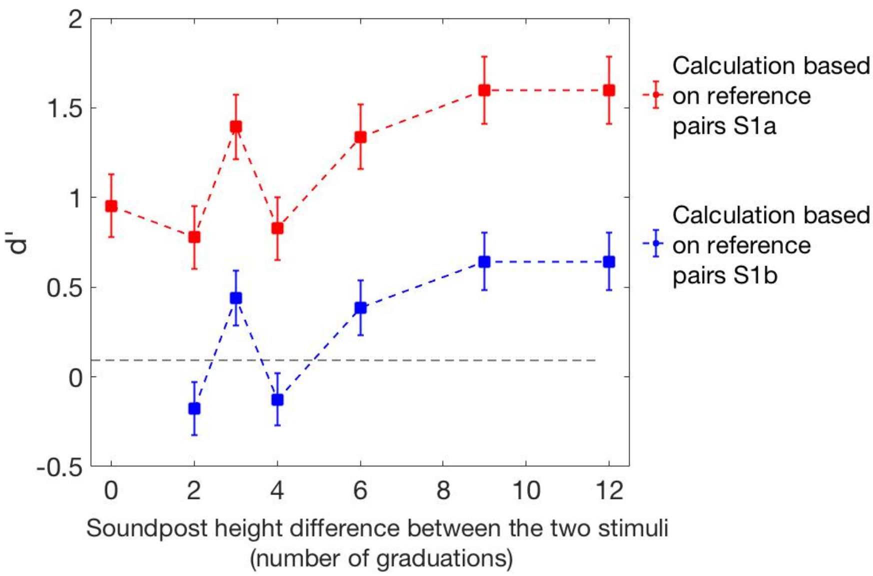
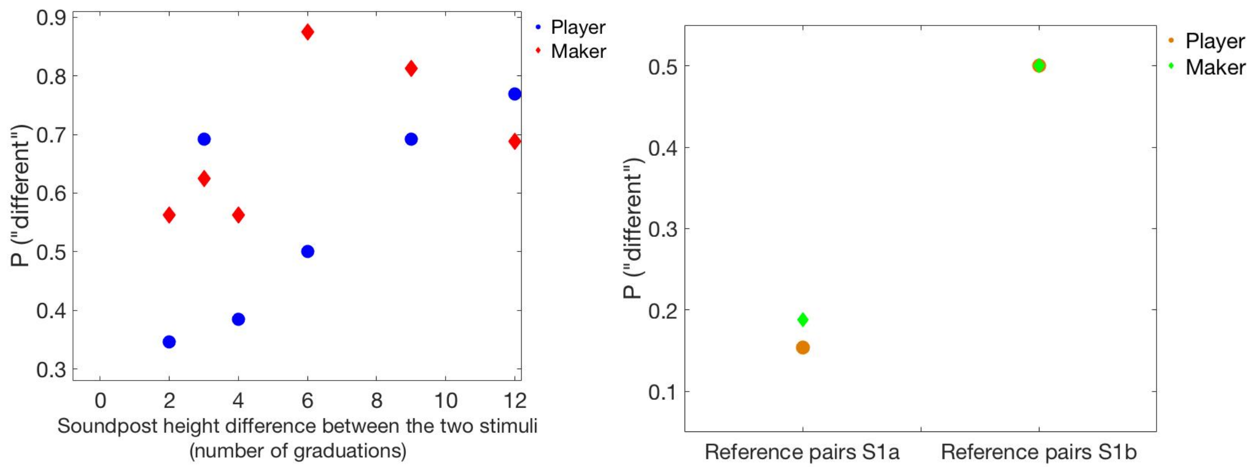


Figure 6.6 Perceptual sensitivity *d'* for each |Δ*H*| calculated using two sets of reference.

6.3.2 Comparison between the Results of Players and Makers

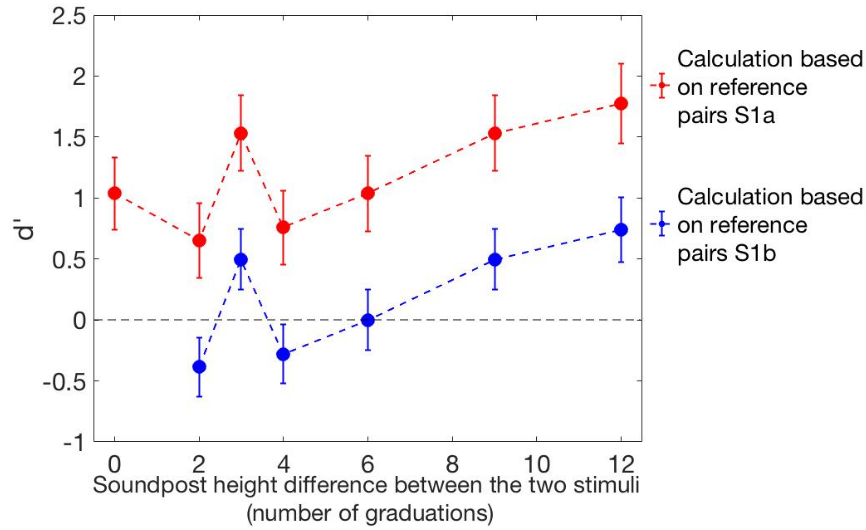
We compared the results for players and makers. Figure 6.7 (a) shows the probabilities (also known as hit rates) that players and makers considered the two soundpost heights with a difference of 2, 3, 4, 6, 9 or 12 graduations as "different", respectively. We can see that the remarkably high hit rate for 3 graduations came more from the players' results. For players, the hit rate generally increased with increases in the soundpost height differences, except for the high hit rate for 3 graduations. For makers, the hit rate for 3 graduations was also a little bit higher than 2 and 4 graduations. 6 graduations had the highest hit rate, then it decreased as the soundpost height difference increased. Figure 6.7 (b) displays the probabilities (also known as false alarm rates) that players and makers considered two different recordings of the same soundpost heights in each of the two reference pairs S_{1b} and the identical recordings in each of the two reference pairs S_{1a} as "different", respectively. We can find that the false alarm rates of the two types of reference pairs were similar for makers and players. Makers had a slightly higher false alarm rate for reference pairs of identical recordings (S_{1a}).



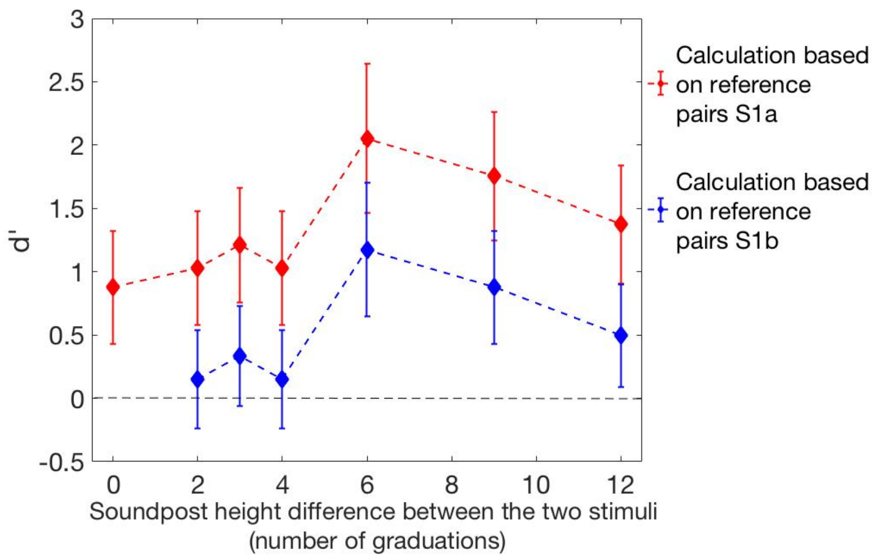
(a) (b)

Figure 6.7 Probabilities that players and makers considered each pair of stimuli as different.

Figure 6.8 (a) and (b) show the perceptual sensitivity measure *d'* for each $|\Delta H|$ calculated based on each of the two types of reference pairs for players and makers, separately. As in Figure 6.6, we also calculated the perceptual sensitivity of the reference pairs S_{1b} based on the reference pairs S_{1a}. Error bars corresponding to the 95% confidence intervals around *d'* are also displayed. We can see that all *d'* calculated based on the first type of reference pairs S_{1a} were greater than 0 for both players and makers. This means that players and makers could perceive all the differences, regardless of whether they were due to changes in soundpost height (($|\Delta H| > 0$) or to playing variations ($|\Delta H| = 0$). For *d'* calculated based on the second type of reference pairs S_{1b}, we can see that only *d'* for soundpost height differences of 3, 9 and 12 graduations for players and of 6, 9 and 12 graduations for makers were greater than 0. It is not very clear why the sensitivity dropped below chance for $|\Delta H| = 4$ and 6 for players, nor why it decreased for makers when $|\Delta H|$ increased over 6. Maybe this could be due to some specificities in some of the recordings and so choosing other stimuli from our dataset to create the pairs may lead to a different sensitivity. In addition, these specificities may have been more obvious to one population than the other, due to their different expertise. More investigation with more recordings and a larger number of subjects would be needed to form more accurate hypotheses.



(a) Players



(b) Makers

Figure 6.8 Perceptual sensitivity *d'* for each $|\Delta H|$ calculated using two sets of reference for players and makers, respectively.

6.3.3 Verbal Description Analysis

When subjects compared two stimuli in each pair, choosing the option "different" led to the appearance of a textbox to describe the differences they heard between the two stimuli. By examining the verbal descriptions provided by the subjects, we found that subjects were generally consistent across the two tests in determining whether the two stimuli in each pair were the same or different as well as in describing the differences between them.

We summarized the number of positive and negative comments for each stimulus in each pair of stimuli across the two tests in Table 6.5. More detailed verbal descriptions are presented in the appendix “Verbal Responses of Study 3 in Chapter 6”. During the listening test, subjects were asked to describe the differences they heard in each pair of stimuli. Some of the comments clearly expressed a positive or negative impression of a particular stimulus in a given pair of stimuli. The table headings of “+” and “-” specify the number of positive and negative comments, respectively, given by the subjects.

Table 6.5 Number of positive (+) and negative (-) comments for each stimulus in every pair of stimuli across the two tests. Stimuli with overall more positive comments and/or less negative comments in one pair were denoted in bold. More detailed verbal descriptions are summarized in the appendix “Verbal Responses of Study 3 in Chapter 6”. The last column of “Different” shows the total number of subjects who thought the two stimuli in that pair as “different” in any of the two tests divided by the total number of trials.

Pair	Stimulus: reading	+	-	Stimulus: reading	+	-	Different
Pair 2(1)	**2: 2.2 (1)**	6	0	3: 2.2 (2)	2	0	8/21
Pair 2(2)	4: 2.3 (1)	3	3	**5: 2.3 (2)**	10	0	13/21
Pair 3	3: 2.2 (2)	4	4	**6: 2.4**	10	2	18/42
Pair 4	1: 2.0	5	7	**4: 2.3 (1)**	10	9	28/42
Pair 5	**1: 2.0**	6	4	6: 2.4	5	8	19/42
Pair 6	4: 2.3 (1)	11	2	**8: 3.1**	14	2	26/40[c]
Pair 7	**4: 2.3 (1)**	11	6	9: 3.4	8	11	31/42
Pair 8	1: 2.0	10	6	9: 3.4	11	12	31/42

c: Subject 1 evaluated a different pair of stimuli with the same soundpost height difference between the two stimuli in pair 6 and thus was not included in this result.

Pair 2(1) was the reference pair consisting of two recordings of one soundpost height that were made during the same recording procedure (see Section 6.2.2 for more details on the recording procedures): stimulus 2 (reading 2.2 (1)) and stimulus 3 (reading 2.2 (2)). Eight subjects thought they were different. More subjects preferred stimulus 2, which was considered to be *warmer*, *had more sustain* and *depth of the tone*, *deeper*, *more open* or *freer*.

Pair 2(2) was the reference pair consisting of two recordings of one soundpost height that were made during two different recording procedures: stimulus 4 (reading 2.3 (1)) and stimulus 5 (reading 2.3 (2)). Thirteen subjects thought them different. Most subjects preferred stimulus 5, made during the second soundpost height increasing procedure, which was considered to be *more focus*, *more resonant*, *rounder* or *had more depth*, *more overtones*, *more balanced sounds across strings*. Stimulus 4, which was the violinist's most preferred soundpost height during the first soundpost height increasing procedure, was considered *a little tight*, *had a more muted quality* and only three subjects found it *more open or* had *clearer articulation* than stimulus 5. Due to possible inaccuracies in the soundpost height adjustment, the two stimuli may not have been recorded at identical soundpost heights.

In pair 3, subjects compared stimulus 3 (reading 2.2 (2)) with stimulus 6 (reading 2.4). Stimulus 6 was preferred by more subjects. Stimulus 6 was recorded at a soundpost height of 2 graduations (i.e., 0.044 mm) higher than stimulus 3. It was the violinist's most preferred soundpost height during the second soundpost height increasing procedure. It was considered *fuller*, especially *the lower strings*, *richer*, *more resonant* by most of the subjects. Stimulus 3 was thought *a bit edgy*, *less responsive*, *slightly muffled* and only four subjects thought it *smoother, fuller, or more balanced*. The result seems consistent with the violinist's opinion during recording.

In pair 4, subjects compared stimulus 1 (reading 2.0) with stimulus 4 (reading 2.3 (1)). Stimulus 4 was recorded at a soundpost height of 3 graduations (i.e., 0.066 mm) higher than stimulus 1 (made at the original soundpost height). It was preferred by more subjects: *more even*, *resonant more*, *better response*, *more open*, *louder*, *brighter* or *projected more*, while stimulus 1 was considered to be *less resonant*, *more stressed*, *tighter*, *more dead*, *had more "inside" sound*. The result is again consistent with the violinist's opinion during recording.

In pair 5, subjects compared stimulus 1 (reading 2.0) with stimulus 6 (reading 2.4). Comparing the results in the two tests, we found that subjects responded "different" more in test 1 (13 subjects) than test 2 (only 6 subjects). The descriptions provided by the subjects showed that stimulus 1, made at the original soundpost height, was preferred by the majority: more subjects found stimulus 1 *smoother*, *rounder*, *more open*, *traveling more* or *more projecting*; while stimulus 6, the violinist's most preferred soundpost height during the second soundpost height increasing procedure, was considered *less focused*, *thinner*, tighter or *more "closed" by more subjects*. This result seems to contradict the violinist' opinion during recording.

In pair 6, subjects compared stimulus 4 (reading 2.3 (1)) with stimulus 8 (reading 3.1). Combining the descriptions in the two tests, the comments about stimulus 8, which was made at a soundpost height of 6 graduations (i.e., 0.132 mm) higher than stimulus 4 (the violinist's most preferred soundpost height during the first soundpost height increasing procedure) seem to be more positive: more subjects described it as *more open*, *fuller*, *more robust* or *rounder*, *more resonant* or with *more projection*; while stimulus 4 was considered *softer*, *more mellow* or *flexible*, *tighter*, *less resonant*, *a little bit muted especially on the lower strings*. The result contradicts the violinist's opinion during recording.

In pair 7, subjects compared stimulus 4 (reading 2.3 (1)) with stimulus 9 (reading 3.4). The soundpost height of stimulus 9 was 9 graduations (i.e., 0.198 mm) higher than stimulus 4. We found that stimulus 4 had a slightly more positive description: more subjects found stimulus 4 *softer*, *clearer*, *smoother*, *more resonant* or *more balanced*. Stimulus 9 was considered as *slightly harsher*, *less responsive*, *with more grain in the sound*, *tighter*, *more noises or was not speaking as clearly or easily as* stimulus 4. The comparison of the results between pair 6 (stimulus 8 preferred over stimulus 4) and pair 7 (stimulus 4 preferred over stimulus 9) seem to show that if the soundpost height increases too much, the sound deteriorates.

In pair 8, subjects compared stimulus 1 (reading 2.0) with stimulus 9 (reading 3.4). They were made at the original soundpost height and the highest soundpost height, respectively. There seemed less description implying preference to any stimuli of the two. Subjects found stimulus 1 *had a more muted tone*, *duller but sounded smoother*, *sweeter* or *velvetier*; while stimulus 9 was *harsher* or *heavier*, *more edgy*, *less precise*, *less rich*, but *more open*, *louder*, *projecting more* or *thicker*. This seems to show that none of the two stimuli was really liked, which is in agreement with an optimal height that is between the two extremes.

To sum up, in general listeners agreed with the player's opinion during the recording, i.e., subjects agreed that the violin sound quality increases as the soundpost height increases up to a point where it then begins to deteriorate. Only pair 5 and pair 6 contradict with the general tendency a little bit. Those two pairs both included one recording that was among the player's most preferred recordings (one in each recording procedure), however, the difference between the number of positive and negative comments about the two recordings in any of the two pairs was not that big. Possible reasons for the differences between the listening test and the player's opinion could include: 1. The listeners were not able to perceive the playing response and interaction with the violin; 2. The player's opinion was only based on one player; 3. The player was conscious of the soundpost increasing procedure, which could affect his opinion. For the reference pairs of pair 2(1) and 2(2), we found that subjects showed preference toward one of the two recordings made at the same soundpost height, which implied that listeners could not differentiate the differences caused by different soundpost heights from those caused by variations of the player's technique. And for pair 2(2), the two recordings were made at the same soundpost height but in different recording procedures, the inaccuracy of the adjustable soundpost could also contribute to the differences between the two recordings.

6.4 Conclusions

In this chapter, we explored players and makers' perception of different soundpost heights through a listening test (using recorded sounds) with a computer interface. A violin installed with a height-adjustable soundpost was used to make the recordings. A concert violinist was invited to repeatedly perform a musical excerpt for different soundpost heights. Subjects then compared six pairs of recordings at different soundpost heights, as well as two reference pairs of identical recordings and two reference pairs of different recordings at same soundpost heights.

The perceptual threshold of the soundpost height differences was estimated through calculating a perceptual sensitivity measure of *d'*. The results showed that based on the two reference pairs of identical recordings, subjects could differentiate all different stimuli we presented at above chance levels, including the stimuli in the reference pairs S_{1b} of different recordings made at the same soundpost heights. Subjects could differentiate soundpost heights with a difference of 3, 6, 9 or 12 graduations (i.e., 0.066 mm, 0.132 mm and greater) at above chance levels based on the reference pairs S_{1b} of different recordings at the same soundpost heights. The perceptual sensitivity seemed not to increase linearly with the increases in the soundpost height differences, as there was an exceptional high perceptual sensitivity for 3 graduations (0.066 mm). That could also be induced by the inaccuracies of the height-adjustable soundpost we used as we have mentioned in Chapter 5. The two stimuli in pair 5 [1, 6] with 4 graduations apart (0.088 mm) were recorded during two different soundpost height increasing procedures, and thus the actual height differences might be smaller than the two stimuli in pair 4 (with 3 graduations apart, 0.066mm).

Comparing the players' and makers' results, we found that players and makers exhibited similar false alarm rates for the different reference pairs. Makers had a slightly higher false alarm rate for the reference pairs of identical recordings than players.

Comparing the results of the listening test and the playing test in Chapter 5, we found that the false alarm rates were much lower in the listening test than the playing test, with even the highest false alarm rate (P = 0.619) of the reference pair [4, 5] in the listening test lower than the false alarm rate (P = 0.711) in the playing test. The reason may be that there were fewer variables in the listening test than in the playing test, which we have discussed in Chapter 5. And subjects can compare two stimuli at the same time. In the listening test, we managed to test bigger soundpost height differences that we didn't involve in the playing test: 6, 9 and 12 graduations.

The verbal response analysis shed some light on the influence of soundpost height on the sound quality: the violin sound quality may increase as the soundpost height increases, though only up to a limit, which is consistent with violin makers' experience. This is to some extent consistent with the finding in the playing experiment. The listeners' opinions were generally in line with the player's opinion (who helped make the recordings). It confirms the reliability of our listening test to some extent. Analyses of the reference pairs S_{1b} of different recordings at the same soundpost heights imply that listeners could not differentiate the differences caused by different soundpost heights or from those caused by inadvertent variations of the player's technique. This is the primary disadvantage of the listening test.

Chapter 7

7 Conclusions and Future Work

Previous perceptual experiments have shown that players do not agree with each other in terms of violin preference. The first study of this thesis tried to explore whether there would be more agreement among players in comparing entry-level Suzuki instruments to more advanced ones. The second and third studies investigated the origin of the disagreement among players through two specific modifications to the violin: the influence of different types of strings on the violin quality and the perception of players and makers on the soundpost height. Section 7.1 summarizes the main findings in the three perceptual studies presented in the previous chapters. Section 7.2 gives some recommendations for possible future studies.

7.1 Main Findings of the Thesis

In Chapter 3, a playing-based violin quality evaluation experiment was performed to examine players' evaluation between performance and student violins. Nine violinists with various degrees in violin performance were invited to evaluate three performance-level and three student-level violins. Subjects rated the six violins according to their own preference and five criteria: *responsiveness*, *resonance*, *clarity*, *richness* and *balance*, in addition to verbal responses to open questionnaire related to their criteria when assessing violin qualities. The results showed:

- It was found that there were significant differences of the preference ratings between the six violins and performance violins were on average rated significantly higher than student violins in terms of preference. And the subjects who were professional musicians, and/or with higher educational degrees in music performance rated performance violins much higher than student violins.
- A large amount of variation in the interindividual consistency of the preference ratings of the violins occurred, but three professional musicians highly agreed with each other in this experiment.

- From the verbal collections of Questionnaire A, it was found that the violinists considered *resonance*, *response*, *balance*, *projection*, *richness*, *texture*, *interest*, *clarity* and *craft* when evaluating violins.
- We found that performance violins were on average rated significantly higher than student violins in all attributes rating scales except *responsiveness* and *resonance*. Subjects who were professional musicians, and/or with higher educational degrees in music performance rated performance violins much higher than student violins in *resonance*, *clarity* and *richness*.
- Relatively higher inter-individual consistency of *richness* and *balanc*e existed among subjects during the violin evaluation. Three professional musicians had much higher agreement on *resonance* and *richness*.
- The analysis of the relationship between preference ratings and attributes ratings showed that violinists preferred violins with rich and to a lesser extent clear sound. The most preferred violin was rated significantly higher than the least preferred violin or the second least preferred violin in all attribute ratings except *responsiveness*.
- In the verbal responses of Questionnaire B, the violinists stated that some rating criteria of the violin were correlated, e.g., *resonance* and *richness*, *clarity* and *responsiveness*. *Resonance* and *responsiveness* were anti-correlated to some extent. Considering the higher inter-individual consistency among professional musicians, further analysis can be restricted to the results of these subjects.
- Bridge admittances were measured for the test instruments. It was found that the magnitudes of the two signature modes A0 and B1- were higher in performance violins than student violins to the extent where it almost appears on the plots as if student violins did not have a B1- mode. And the magnitude response of the student violins did not show an apparent "transition hill" around 1000 Hz. The damping of the A0 mode was slightly lower for performance violins than student violins.

In Chapter 4, we examined the influences of different types of strings on the violin quality through a playing-based perceptual experiment. Two violins of the same make with similar sound quality and playability and three types of strings (Dominant strings, Kaplan strings, and Pro-Arté strings) were employed. The two violins were both strung with Dominant strings initially. Subjects played the violins, described and rated the difference between the two violins (violin 2 compared to violin 1) according to eight criteria (*responsiveness*, *power*, *resonance*, *brightness*, *clarity*, *richness*, *balance* and overall quality) during a session labeled D1-D2. Subsequently, the strings of violin 2 were changed to a different brand (Kaplan or Pro-Arté), unbeknownst to the players, and players had to re-evaluate the differences between the two violins (session D1-K2 or D1-P2). In Oberlin (USA), nine subjects compared violin 2 with Dominant and then Kaplan strings to violin 1 in two sessions (D1-D2 and D1-K2). In Montreal, ten subjects evaluated the differences between the two violins in three sessions (D1-D2, D1-K2 and D1-P2). The results showed:

- The differences between D1-D2 and D1-K2 were not statistically significant based on the Oberlin results.
- The differences among D1-D2, D1-K2 and D1-P2 were not statistically significant as well based on the Montreal results. If we compare every two experimental conditions based on the Montreal results, differences between D1-D2 and D1-K2, and D1-K2 and D1-P2 were not significant. However, the *brightness* difference ratings were found to be significantly higher in D1-D2 than in D1-P2.
- There were no significant differences between D1-D2 and D1-K2 even when we combined the results of the two parts of this experiment in Oberlin and Montreal.
- We also examined the relationship between attribute difference ratings and overall quality difference ratings. *Richness* difference ratings and to a lesser extent, *resonance* difference ratings, were found to significantly correlate to overall quality difference ratings based on both the Oberlin and Montreal results.

In Chapters 5 and 6, we explored the perception of the violin soundpost height differences through a playing test and a listening test, respectively. A height-adjustable carbon fibre soundpost was employed in both tests. In the playing test, we found the optimal soundpost height for each subject and investigated the perceptual sensitivity to soundpost height differences around each subject's optimal soundpost height. In the listening test, we explored players and makers' perception of different soundpost heights using recorded sounds with a computer interface. The results showed:

- From the playing test it was found that the subjects' optimal soundpost heights varied from 0.132 mm to 0.616 mm relative to the original soundpost height (53 mm), reasonably well inside the extreme soundpost heights that were tested (0 mm and 0.66 mm). We found the optimal soundpost heights for subjects vary within an interquartile range of 0.3 mm. The variation interquartile range is higher for players (0.32 mm) than for makers (0.26 mm). The mean relative optimal soundpost height is also higher for players (0.36 mm) than for makers (0.3 mm). Statistical analysis showed that the differences of the relative optimal soundpost height for players and makers were not significant.
- Subjects could recognize soundpost height changes of 0.088 mm and 0.11 mm at better than chance levels. Players could recognize height changes of 0.044 mm and 0.11 mm at above chance levels. Makers could recognize the height changes of 0.088 mm at above chance levels.
- We measured the bridge admittances of the violin used in this playing test for every soundpost height that had been evaluated by subjects during the first phase of finding optimal soundpost heights. We found that the admittances are very similar, with a slight decrease in magnitudes with increasing soundpost height for peaks up to about 1100 Hz.
- Three signature modes were identified (A0, B1- and B1+) in the admittances. There was a general tendency that the magnitudes for all three modes decreased as the soundpost height increased and this observation was more obvious for the A0 mode. From the original soundpost height to the highest soundpost height, the magnitude decreased for A0, B1- and B1+ mode about 1.96, 1.33 and 1.5 dB, separately. The violin quality did not associate with the magnitudes of the three violin modes linearly or positively across the full range of soundpost heights we tested.
- The variations of the frequency of these three modes were subtle, which may not be greater than the measurement error. The frequency of the A0 mode stayed consistent generally within about 1 Hz. The frequency of the B1- mode displays a general increase tendency with the increase in the soundpost height except for a few fluctuations. The frequency of the B1+ mode seems to show a general decrease tendency with the increase in the soundpost height, though the fluctuations were relatively large.
- From the listening test, we found that based on the two reference pairs of identical recordings, subjects could differentiate all different stimuli we presented at above chance

levels, including the stimuli in the reference pairs S_{Ib} of different recordings made at the same soundpost heights, which raises the question whether the participants differentiated the stimuli based on height differences or on playing differences. Subjects could differentiate soundpost heights with a difference of 0.066 mm, 0.132 mm, 0.198 mm or 0.264 mm at above chance levels based on the reference pairs S_{Ib} of different recordings at the same soundpost heights.

- Comparing the players' and makers' results, we found that players and makers exhibited similar false alarm rates for the different reference pairs. Makers had a slightly higher false alarm rate for the reference pairs of identical recordings than players.
- Comparing the results of the listening test and the playing test, we found that the false alarm rates were much lower in the listening test than the playing test.
- Combing the results of the playing test and the listening test, we found what violin makers know very well: the quality increases as the soundpost height increases (starting from a very loose soundpost), but only to a certain height, where it decreases again.

7.2 Suggestions for Future Work

In the first study of the thesis, we investigated the violinists' evaluation of performance and student level violins and found that the three professional musicians highly agree with each other. Subjects who were professional musicians, and/or with higher educational degrees in music performance rated performance violins much higher than student violins. However, the number of the subjects was relatively small. A future study may recruit more subjects with various degrees in music performance, especially more professional musicians. As the professional musicians have more agreement with each other in terms of violin preference, the analysis of the verbal collections can be limited to their answers, as well as the relationship between preference and attribute criteria ratings, the agreement and the correlation between physical measurement and the perceptual evaluation.

In the second study, we studied the influence of different types of strings on the violin quality employing two student level violins and three types of strings that are widely used on violins and are generally considered to be of good quality. There may be several future study directions:

- Strings may have different effects on different violins. It might be interesting to employ two peformance-level violins of similar sound quality and playability to examine the effect of different types of strings on the violin quality.
- The strings we tested are generally considered to be of good quality. More significant differences may exist between strings of low quality and good quality. It would be good to examine the effect that strings of quite different qualities have on the violin perception.

In the third study, we examined players' and luthiers' perception of different soundpost heights through a playing test and a listening test. The following research direction would merit further attention:

- During the playing test, subjects' perceptual sensitivity to the soundpost height differences was quite small, which may be due to the small soundpost height differences we tested (within $\pm$ 0.1 mm). Increasing the soundpost height differences may increase the perceptual sensitivity of subjects. That said, the overall range of soundpost height variations must remain similar to avoid damaging the violins.

Many experiments related to the connection between physical modification of the violin and the changes on the perceived quality can be conducted in the future, some of which were listed in the beginning of Chapter 1 (varnish, violin plate thicknesses, bridge geometry, etc.). Any such experiments should be well-controlled to make sure the modification is made only on the testing factor during each experiment, and all the other factors stay the same.

Appendix: Original Verbal Responses from Subjects

Questionnaire B of Study 1 in Chapter 3

B1: Specific comments or remarks about the "*responsiveness*" of the violins? Was there a particular behavior in the violin rated as least responsive or the one rated as most responsive?

Violin	Most responsive	Least responsive
P1	s1: (comment) I tried playing fast strokes like spicatto and piano, but I couldn't find a very clear pattern.	s3: the least responsive took more effort and had very little nuance in sound.
P2		
P3	s2: very clear sound at attack of string.	s5: the bow skids on the lower ranked violins and even when more weight is added the sound is less resonant, less clear. Less dynamics seem possible as well.
S1	s9: (comment) ringing = sound comes out; setup masks true potential; strings.	s6: I don't like violin S1, the sound is small, and buzzing, stiff, usually new made violin is like this. s7: (comment) usually the bad sounding and very dry violin responds very quick. The violin with low

		fingerboard and low bridge usually respond very quick, but the sound is flat and not solid.
S2	s3: the most responsive violins didn't require much effort to create the sound I wanted. s7: (comment) thick sound often needs more time to respond. s8: (comment) I found responsiveness could best be evaluated among these violins on the G-string.	s2: string wasn't very responsive s4: violin S1 and S2, the neck of the violin felt thicker to me making a clumsier in the left hand.
S3		

B2: Specific comments or remarks about the "*resonance*" of the violins? Was there a particular behavior in the violin rated as least resonant or the one rated as most resonant?

Violin	Most resonant	Least resonant
P1	s5: very easy, simple ringing sound that comes out very naturally. s6: (Comment) P1 is brighter than P2.	

P2	s3: the most resonant violin had an open sound that could be powerful when receiving. s6: violin P2 is muffled, but I like it. The resonance is great with big amplitude. s7: good resonance is that the sound in the higher position of the low register is orotund and thick.	s2: very muted sounding.
P3	s4: violin P2 has a resonant sound but its sound wasn't as defined as violin P3. s6: (Comment) violin P3 is also good. s8: I listened for the ring of the open strings as well as sympathetic vibration. A kinesthetic or tactile sense also came into play – with the resonant violins, the body of the instrument seemed to vibrate more.	s5: the least resonant feels very resistant and stiffed, as if there is a cloth muffling the sound, less satisfying to play.
S1	s1: the most resonant feel a brighter tone	s3: had a muted, boxy sound with no ringing/brilliant qualities. s7: flat, not thick enough in the lower register, and sharp in the higher register.
S2		s4: some like violin S1 and S2 became fuzzy.
S3	s2: very bright sound is ringing very open. s9: overtones come out /ringing; usually dark have less resonance (muffled),	s6: resonance is different from responsiveness. The higher register of violin S3 is bright, but narrow, broader

	challenge good 'dark'; darker one – less resonance (muffled), student model usually bright sounding, not good. So the challenge is to find good 'dark': ringing one still dark.	will be better. It vibrates very quickly, very sensitive, but very narrow, it doesn't vibrate very well.

B3: Specific comments or remarks about the "*clarity*" of the violins? Was there a particular behavior in the violin rated as least clear or the one rated as most clear?

Violin	Most clear	Least clear
P1	s5: the most clear has a very concentrated, narrow sound amidst the ringing tones around it. s6: clarity relates to responsiveness: as long as the bow touches the string, the violin can express clearly. Good clarity means the violin can express clearly under different bow force. s7: the violin with good clarity often sounds bright, but lack flavor, not sweet. Violin with good flavor may not be the clearest. s9: sound itself: nice boom of notes, ringing well help with the transition between notes, making them connect well with each other.	

P2	s1: it was easiest to perceive during the attacks.	s2: violin P2 buzzing, lacked pure and clean sound. s5: the least clear violin sounds overall cloudy. It is difficult to pin point where the exact sound is. It feels like a sketch with pencils and shading used compared to a clean drawing with pen.
P3	s2: violin P3 sound is very pure, no whistles, or buzzing. s4: (Comment) my bow is on the softer side. So this violin while very open wasn't as pure as tone as violin S3.	
S1		s3: the least clear had a muddy sound, even when I went for a brilliant sound s6: some violins become not clear when the bow force increases. s9: strings felt/ bridge buzzing/ tailpiece
S2	s3: the most clear violin had a more brilliant pure sound (goes hand in hand with responsiveness)	
S3	s4: violin S3 general had the clearest sound, with the exception of a G string wolf.	s1: S3 had a very clear buzzing noise. s8: some violins were 'muddier' than others particularly in the lower register.

B4: Specific comments or remarks about the "*richness*" of the violins? Was there a particular behavior in the violin rated as least rich or the one rated as most rich?

Violin	Most rich	Least rich
P1	s2: violin P1 has lots of colour and deep sound. s5: the most rich violin allows me to sink into the string and the resonance and clarity that surrounds the sound creates a warm atmosphere. The sound is clear but also big and fat. s6: (comment) the responsiveness of violin P1 is better than violin P2, the sound is good. s9: (comment) rich (projection/sound color) combination. It's good to have audience. They resonate, may sound sweeter, overtones.	
P2	s3: the most rich violin had dark, rich undertones. The lower register could speak more with time. s6: the sound of violin P2 is thick and can be extended. If you apply more force, it will give you more space to express. s7: (comment) richness means good sound, good expression, and resonance, flavor both relates to richness.	

P3	s8: (comment) richness was partly related to resonance, and I preferred to test richness using double stops.	s5: the least rich lacks depth in the sound even when I use more weight, feels flat and hollow.
S1		s3: the least rich were hollow and shallow sounding. s7: (comment) violin with bad resonance often doesn't vibrate very well, produce very few sounds, then nothing could be expressed.
S2	s1: (comment) I mostly tested it on the G string.	s4: the sound quality was more open.
S3		s2: violin S3 not many colors in the sound, didn't feel as full as the other violins. s6: (comment) the sound of violin S3 is very narrow, it will stay like that no matter how you play.

B5: Specific comments or remarks about the "*balance*" of the violins? Was there a particular behavior in the violin rated as least balanced or the one rated as most balanced?

Violin	Most balanced	Least balanced
P1	s5: strangely enough some of the more muffled violins were well balanced because the sound felt muffled overall. The most	

	balanced was resonant all the way through and although each note was distinct. They were very clear in the same way.	
P2	s1: (comment) here I looked both for balance across strings and also across different notes. s4: (comment) the violin P2, the E string was much brighter than the rest of the violin. s6: (comment) the sound of the lower register of violin P2 is thick, if the higher register can be thicker, and bright as violin S3, that will be great. So overall its balance is not good.	s9: (comment) setup needs. Bridge angles matter a lot, sloped bridge (these six violins) make strings not resonate separately. Less balanced didn't ring every string.
P3	s2: violin P3 very consistent across the strings, even sounding. s3: the most balanced violin had an even, consistent sound across the four strings. s8: (comment) I found balance easiest to test using open strings and arpeggios.	
S1		s3: the least violin had a string that 'stuck out' more to the ear or just no sense of nuance at all. s7: bad balance is that the sound in the lower register is not thick enough; sound in the higher register is not bright, however, dark; or good at one aspect, but bad at the other aspect.

S2	s9: (comment) each string resonates, more balanced rang through at all strings	s5: some less balanced violins had very tangy D string that was hard to work around.
S3	s6: violin S3 is sensitive, the higher register is not bad, comfortable. But the sound is narrow for both higher and lower register, so it's overall balanced.	s2: violin S3 the bridge made the playing uneven, feeling unstable.

Verbal Responses of Study 3 in Chapter 6

"+" (or "-") represents positive (or negative) comments with the number of subjects （one subject in different tests was considered as the number of two） who gave the comments in parentheses.

Subject number and in which test are indicated, e.g., s4/t1, which means subject 4 in test 1.a

Stimulus in bold means it had overall more positive and/or less negative comments than the other one in the pair of stimuli.

	Stimuli 2: 2.2 (1)		Stimuli 3: 2.2 (2)	
	+ (6)	- (0)	+ (2)	- (0)
Pair 2(1)	s4/t1: les graves sont plus chaleureux s9/t1: has more sustain and depth of tone (darker) s11/t2: clear sound quality		s11/t2: better resonance s13/t2: sounds a bit richer	

	s12/t1: sounds deeper s17/t2: more clear and direct s18/t1: the low register sounds more “open and free”			
Pair 2(2)	Stimuli 4: 2.3 (1)		**Stimuli 5: 2.3 (2)**	
	+ (3)	- (3)	+ (10)	- (0)
	s9/t2: clearer articulation s17/t1: Lower strings sound more open s22/t1: sounds a little more brilliant and open	s6/t2: a little tight s12/t2: a flatter sound s16/t1: more muted quality. The sound isn’t sparkling and doesn’t project as clearly.	s2/t2: more focus s3/t1: less tense s6/t2: more resonant s8/t1: darker s12/t2: has more depth s13/t1: sounds more round s15/t2: sounds slightly rounder s16/t1: more overtones in the resonance, more power	

			s18/t2: middle register opens s19/t1: a slightly more balanced string tone	
Pair 3	Stimuli 3: 2.2 (2) s5/t2: something closer		**Stimuli 6: 2.4**	
	+ (4)	- (4)	+ (10)	- (2)
	s4/t1: more smooth sound s8/t1: fuller s12/t2: more balanced s15/t2: sounds smoother	s17/t1: open G string doesn't ring as much s22/t1: sounds a bit thinner s6/t2: a bit edgy s10/t2: slightly muffled attack, less responsive	s3/t1: the lower strings seem advantaged s6/t1: lower strings sounded fuller s6/t2: seemed warmer s8/t2: richer s9/t1: has greater depth of tone s13/t1: sounds a bit richer s13/t2: sounds a bit richer	s12/t2: less equal between lower and higher registers s18/t2: much worse for the high register

			s15/t1: sounds a little bit richer s18/t2: better for the low register s19/t1: sounded slightly more resonant	
Pair 4	Stimuli 1: 2.0		**Stimuli 4: 2.3 (1)**	
	+ (5)	- (7)	+ (10)	- (9)
	s6/t1: rounder, especially lower strings s12/t1: darker s8/t2: fuller s15/t2: sounds smoother and rounder s21/t2: sound appears to be lighter, less difference between each string	s2/t1: more "inside" sound s17/t1: open G string doesn't ring as much s22/t1: sounds a little muffled s1/t2: less resonant s5/t2: more stressed s7/t2: more dead and direct s19/t2: sounded tighter	s5/t1: more even s9/t1: resonates more s10/t1: sound crisper, i.e., better response s12/t1: brighter s14/t1: thicker s3/t2: brighter s6/t2: punchy, louder s7/t2: rounder, richer and more resonant	s1/t1: sounds slightly choked s6/t1: more edgy s7/t1: has a little bit less resonance s13/t1: have more grain in the sound and sounds a bit strangled s21/t1: seems slightly tighter s6/t2: more edgy s10/t2: less overtones

			s12/t2: more brilliant s22/t2: more open, brighter and projecting more	s13/t2: sounds thinner and not free s14/t2: rougher and tighter
	Stimuli 1: 2.0		Stimuli 6: 2.4	
	+ (6)	- (4)	+ (5)	- (8)
Pair 5	s6/t1: smoother s11/t1: project further s12/t1: more round s15/t1: has a more open sound s19/t1: has slightly more open sound s7/t2: travels more and more projecting, has more resonance and more harmonics in the sound	s7/t1: the sound is a little bit more dead s16/t1: more direct sound quality with less ring and more buzz s17/t2: sounds more covered and muted s4/t2: moins définit dans les médiums	s7/t1: more projecting s16/t1: more balanced sound, wider spread, with more natural resonance, the notes "ring" clearer in the higher register at the end of phrase s4/t2: les graves sont plus chaleureux s9/t2: has more depth of tone	s2/t1: less focus s3/t1: higher strings sound shaded s8/t1: thinner s15/t1: sounds tighter s18/t1: the sound is more "closed" in the A string s2/t2: less rich, doesn't resonate as much s7/t2: very direct and the sound

			s22/t2: sounds a little more open	doesn't go very far s22/t2: sounds b less pure
Pair 6	Stimuli 4: 2.3 (1)		**Stimuli 8 : 3.1** s4/t1: le son a plus de grain s7/t2: having more grain in the sound	
	+ (11)	- (2)	+ (14)	- (2)
	s2/t1: more color and more free on the G s3/t1: sounds more soft and subtle s7/t1: rounder s10/t1: has a softer tone s12/t1: brighter s14/t1: thicker and has more core, more depth	s9/t1: sounds tighter (less resonant, sustain and depth of tone) s3/t2: a little bit muted, especially on the lower strings	s5/t1: more open s7/t1: seems to sound louder s8/t1: fuller s16/t1: more robust and round s21/t1: more resonant s22/t1: has more projection and more open s2/t2: softer	s3/t1: maybe the soundpost is closer to the bridge or have more tension s7/t1: the sound is more direct and has less resonance

	s16/t1: more mellow and flexible tone s18/t1: more closed over all the strings s4/t2: les 4 cordes sont plus égales s9/t2: sounds less resonant s11/t2: better timbre, without noise		s4/t2: le son est plus chaleureux s6/t2: very slightly brighter s8/t2: richer s10/t2: sounds more resonant s13/t2: sounds more resonant s18/t2: the sounds open up s22/t2: brighter, more open and has better projection in general (the E string sounds quite free)	
Pair 7	**Stimuli 4: 2.3 (1)**		Stimuli 9: 3.4 s5/t1: close on the high strings s7/t1: has more grain in the sound	
	+ (11)	- (6)	+ (8)	- (11)

	s7/t1: more resonant s8/t1: richer s16/t1: brighter sound s17/t1: clearer, more direct sound s21/t1: sound seems a bit smoother, easier for string-crossing s5/t2: better in the high register s6/t2: slightly smoother s9/t2: sounds more resonant s14/t2: a little soft s15/t2: sounds a little softer/darker, but a nicer and smoother tone s16/t2: more balanced sound production	s10/t1: has less support in the mid range, sounds softer s2/t2: has less focus -s9/t2: sounds thinner s11/t2: has noise in the higher position s18/t2: closed, A string too tensed s22/t2: sound a bit muffled	s4/t1: le son est plus définit et précis, surtout dans les aigues s16/t1: darker fuller low register s22/t1: richer and rounder s3/t2: brighter s5/t2: better in the low register s8/t2: better s12/t2: sounds a bit deeper s19/t2: sounded more open	s1/t1: slightly harsher and less responsive s2/t1: less rich s6/t1: the upper two strings seemed more tight s11/t1: has noise in the higher position s13/t1: less responsive in the attacks s4/t2: moins définit dans les aigues s12/t2: less responsive s15/t2: sounds rougher and louder s16/t2: too direct the sound, power without a lot of

				resonance after the note is played s17/t2: doesn't speak clearly or easily s21/t2: more tight, more noises
Pair 8	Stimuli 1: 2.0		Stimuli 9: 3.4	
	+ (10)	- (6)	+ (11)	- (12)
	s4/t1: le son est clairement plus définit, et les attaques plus précises s5/t1: more even s7/t1: the sound is more resonant, it has a richer sound with a bigger palette of sounds s12/t1: more muted or velvety s14/t1: louder	s1/t1: has a more muted tone with less resonance s6/t1: duller s7/t1: thinner s6/t2: duller s15/t2: sounds harsher s17/t2: sounds a bit more muted and covered	s10/t1: might have more projection s15/t1: sounds bigger/louder but a bit rougher s17/t1: a bit more thicker s22/t1: sounds richer and rounder s3/t2: the lower strings sound more powerful and robust	s7/t1: has a less harmonics in the sound and is more direct. s11/t1: has a bit noise on the third position of D string s14/t1: smaller volume s15/t1: sounds a bit rougher s21/t1: seems a bit harder to grab the string, more friction is needed

	s15/t1: sounds leaner and smoother s17/t1: clearer and more direct s1/t2: has a sweeter sound s3/t2: the higher strings have more depth and richness s12/t2: more velvety		s7/t2: has more resonance and projects more s8/t2: darker, richer s14/t2: louder s15/t2: sounds smoother s19/t2: sounded louder s22/t2: sounds more open and fuller	s1/t2: harsher and heavier s2/t2: less precise, less rich s4/t2: moins définit dans les aigues, généralement le son est plus acide s5/t2: not even s6/t2: more edgy s11/t2: has noise on the B of the D string s21/t2: sound appears to be less more tight, less resonant

Bibliography

Alonso Moral, J., & Jansson, E. V. (1982). Input admittance, eigenmodes and quality of violins (Technical Report STL-QPSR 2-3). Department for Speech, Music and Hearing, Kungliga Tekniska Högskolan, Stockholm.

Bissinger, G. (1995). Some mechanical and acoustical consequences of the violin soundpost. *Journal of the Acoustical Society of America*, *97*(5), 3154-3164.

Bissinger, G. (2005). A unified materials-normal mode approach to violin acoustics. *Acta Acustica united with Acustica*, *91*(2), 214-228.

Bissinger, G. (2008). Structural acoustics of good and bad violins. *Journal of the Acoustical Society of America*, *124*(3), 1764-1773.

Bissinger, G., & Keiffer, J. (2003). Radiation damping, efficiency, and directivity for violin normal modes below 4 kHz. *Acoustics Research Letters Online, 4*(1), 7-12.

Campbell, M., Greated, C. A., & Myers, A. (2004). *Musical instruments: History, technology, and performance of instruments of western music*. Oxford: Oxford University Press.

Cremer, L. (1984). *The physics of the violin.* Cambridge, MA: MIT Press.

Curtin, J., & Rossing, T. D. (2010). Violin. In T. D. Roosing (Ed.), *The science of string instrument*. New York, NY: Springer.

Curtin, J., & Schleske, M. (2003). Can you hear me? *The Strad, 114*, 1103–1110.

Dünnwald, H. (1991). Deduction of objective quality parameters on old and new violins. *Catgut Acoustical Society Journal, (Series II), 1*(7), 1–5.

Firth, I. (1987). Construction and performance of quality commercial violin strings. *Journal of Violin Society America*, *8*(2), 75-88.

Fletcher, N. H., & Rossing, T. D. (1998). *The Physics of Musical Instruments*. New York, NY: Springer.

Fritz, C., Blackwell, A. F., Cross, I., Woodhouse, J., & Moore, B. C. (2012a). Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties. *Journal of the Acoustical Society of America*, *131*(1), 783-794.

Fritz, C., Cross, I., Moore, B. C., & Woodhouse, J. (2007). Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin. *Journal of the Acoustical Society of America*, *122*(6), 3640-3650.

Fritz, C., Curtin, J., Dubois, D., & Verschoore, J. (2016). *The violin: Studying the influence of the model by means of a free sorting task*. Paper presented at the International Symposium on Musical and Room Acoustics, La Plata, Buenos Aires.

Fritz, C., Curtin, J., Poitevineau, J., Borsarello, H., Tao, F.-C., & Ghasarossian, T. (2014). Soloist evaluations of six Old Italian and six new violins. *Proceedings of the National Academy of Sciences, 111*(20), 7224-7229.

Fritz, C., Curtin, J., Poitevineau, J., Morrel-Samuels, P., & Tao, F.-C. (2012b). Player preferences among new and old violins. *Proceedings of the National Academy of Sciences, 109*(3), 760-763.

Fritz, C., Curtin, J., Poitevineau, J., & Tao, F. -C. (2017). Listener evaluations of new and Old Italian violins. *Proceedings of the National Academy of Sciences*, *114*(21), 5395-5400.

Fritz, C., & Dubois, D. (2015). Perceptual evaluation of musical instruments: State of the art and methodology. *Acta Acustica united with Acustica, 101*(2), 369-381.

Fritz, C., Muslewski, A., & Dubois, D. (2010). *A situated and cognitive approach of violin quality*. Paper presented at the International Symposium on Music Acoustics, Sydney and Katoomba, Australia.

Fritz, C., Woodhouse, J., Cheng, F. P. H., Cross, I., Blackwell, A. F., & Moore, B. C. (2010b). Perceptual studies of violin body damping and vibrato. *Journal of the Acoustical Society of America, 127*(1), 513-524.

Gabrielsson, A., & Jansson, E. V. (1979). Long-time-average-spectra and rated qualities of twenty-two violins. *Acta Acustica united with Acustica*, *42*(1), 47-55.

Gough, C. (2007). Musical acoustics. In T. D. Rossing (Ed.), *Springer Handbook of Acoustics*. New York, NY: Springer.

Gough, C. (2015). Violin plate modes. *Journal of the Acoustical Society of America, 137*(1), 139-153.

Gough, C. (2017). *The influence of the bridge, island area, bassbar and soundpost on the acoustic modes of the violin.* Paper presented at the International Symposium on Music Acoustics, Montreal, 67-70.

Gough, C. (2018). The violin bridge-island input filter. *Journal of the Acoustical Society of America, 143*(1), 1-12.

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika*, *32*(1), 25-33.

Gravetter, F. J., & Wallnau, L. B. (2011). *Essentials of statistics for the behavioral sciences*. Belmont, CA: Wadsworth, Cengage Learning.

Guettler, K. (2002a). *The bowed string: On the development of Helmholtz motion and on the creation of anomalous low frequencies*, Doctoral dissertation. Stockholm: Kungliga Tekniska Högskolan.

Guettler, K. (2002b). On the creation of the Helmholtz motion in bowed strings. *Acta Acustica united with Acustica*, *88*(6), 970-985.

Helmholtz, H. L. F. (2013). On the sensations of tone as a physiological basis for the theory of music (Ellis, A. J., Trans.). United States: Dover Publications. (Original work published 1863).

Helmholtz, H. L. F. (1954). *On the sensations of tone: as a physiological basis for the theory of music.* New York, NY: Dover.

Hutchins, C. M. (1981). The acoustics of violin plates. *Scientific American*, *245*(4), 170-187.

Hutchins, C. M. (1983). A history of violin research. *Journal of the Acoustical Society of America*, *73*(5), 1421-1440.

Hutchins, C. M. (1989). A measurable controlling factor in the tone and playing qualities of violins. *Catgut Acoustical Society Journal, (Series II)*, *1*, 10-15.

Hutchins, C. M. (1967). Founding a family of fiddles. *Physics today, 20*(2), 23-27.

Hutchins, C. M., & Benade, V. (1997). *Research papers in violin acoustics, 1975-1993: with an introductory essay, 350 years of violin research*. Woodbury, NY: Acoustical Society of America.

Inta, R., Smith, J., & Wolfe, J. (2005). Measurement of the effect on violins of ageing and playing. *Acoustics Australia, 33*(1), 25-29.

Jansson, E. V. (1997). Admittance measurements of 25 high quality violins. *Acta Acustica united with Acustica*, *83*, 337–341.

Jansson, E. V. (2004). *Acoustics for violin and guitar makers*. Kungliga Tekniska Högskolan, Department of Speech, Music and Hearing. Retrieved June 5, 2019, from http://www.speech.kth.se/music/acviguit4/.

Jansson, E., Molin, N. E., & Sundin, H. (1970). Resonances of a violin body studied by hologram interferometry and acoustical methods. *Physica scripta*, *2*(6), 243.

Jansson, E. V., & Niewczyk, B. K. (1999). On the acoustics of the violin: bridge or body hill. *Catgut Acoustical Society Journal, (Series II), 3*(7), 23–27.

Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food: principles and practices*. New York: Springer.

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45*(1), 255–268.

Loos, U. (1995). Investigation of projection of violin tones. Reviewed by Martin Schleske (2003). *Catgut Acoustical Society Journal, (Series II), 4*(8), 72.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maestre, E., Scavone, G. P., & Smith, J. O. (2016). Design of recursive digital filters in parallel form by linearly constrained pole optimization. *IEEE Signal Processing Letters*, *23*(11), 1547-1550.

Marshall, K. D. (1985). Modal analysis of a violin. *Journal of the Acoustical Society of America*, *77*(2), 695-709.

Meinel, H. (1957). Regarding the sound quality of violins and a scientific basis for violin construction. *Journal of the Acoustical Society of America, 29*(7), 817-822.

Pickering, N. C. (1985). Physical properties of violin strings. *Catgut Acoustical Society Journal*, *44*, 6-8.

Pickering, N. C. (1986). Elasticity of violin strings. *Catgut Acoustical Society Journal*, *46*, 2-3.

Raman, C. V. (1918). On the mechanical theory of the vibration of bowed strings and of musical instruments of the violin family, with experimental verification of results: Part I, *Bulletin of the Indian Association for the Cultivation of Science, 15,* 1–158.

Robson, C. (1994). *Experiment, design and statistics in psychology*. London: Penguin.

Rodgers, O. E., & Anderson, P. (2001). Finite element analysis of violin corpus. *Catgut Acoustical Society Journal*, *4*(4), 12.

Saitis, C. (2013). *Evaluating violin quality: Player reliability and verbalization,* Doctoral Dissertation. Montreal: McGill University.

Saitis, C., Fritz, C., Scavone, G. P., Guastavino, C., & Dubois, D. (2017). Perceptual evaluation of violins: A psycholinguistic analysis of preference verbal descriptions by experienced musicians. *Journal of the Acoustical Society of America*, *141*(4), 2746–2757. doi: 10.1121/1.4980143.

Saitis, C., Giordano, B. L., Fritz, C., & Scavone, G. P. (2012). Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players. *Journal of the Acoustical Society of America*, *132*(6), 4002–4012. doi: 10.1121/1.4765081.

Saitis, C., Scavone, G. P., Fritz, C., & Giordano, B. L. (2015). Effect of task constraints on the perceptual evaluation of violins. *Acta Acustica united with Acustica, 101*(2), 382-393.

Saldner, H. O., Molin, N. E., & Jansson, E. V. (1996). Vibration modes of the violin forced via the bridge and action of the soundpost. *Journal of the Acoustical Society of America*, *100*(2), 1168–1177. doi: 10.1121/1.416301.

Savart, F. (1840). The violin. *L'Institut*, 54-56. [English translation by D. A. Fletcher and printed in Musical Acoustics, Part I, Violin Family Functions, Vol 5, Benchmark Papers in

Acoustics, edited by C. M. Hutchins, Dowden, Hutchinson, and Ross (Stroudsburg, PA, 1975), pp 323 – 327.]

Schelleng, J. C. (1968). Acoustical Effects of Violin Varnish. *The Journal of the Acoustical Society of America*, *44*(5), 1175–1183. doi: 10.1121/1.1911243.

Schelleng, J. C. (1971). The action of the soundpost. *Catgut Acoustical Society Newsletter, 16*, 11–15.

Schelleng, J. C. (1973). The bowed string and the player. *The Journal of the Acoustical Society of America, 53*(1): 26-41, 1973.

Wollman, I., Fritz, C., & Poitevineau, J. (2014). Influence of vibrotactile feedback on some perceptual features of violins. *Journal of the Acoustical Society of America*, *136*(2), 910-921.

Woodhouse, J. (1993a). On the playability of violins. Part I: Reflection functions. *Acta Acustica united with Acustica*, *78*(3), 125-136.

Woodhouse, J. (1993b). On the playability of violins. Part II: Minimum bow force and transients. *Acta Acustica united with Acustica*, *78*(3), 137-153.

Woodhouse, J. (2014). The acoustics of the violin: a review. *Reports on Progress in Physics*, *77*(11), 115901. doi: 10.1088/0034-4885/77/11/115901.

Zhang, A. (2015). *Playability of bowed string instruments,* Doctoral dissertation. Cambridge: University of Cambridge.