JUNE 25 2025

Influences of different string types on the perception of violin qualities[a)]

Lei Fu; Gary Scavone; Claudia Fritz

J. Acoust. Soc. Am. 157, 4604–4612 (2025)
https://doi.org/10.1121/10.0036894

JASA ARTICLE

Influences of different string types on the perception of violin qualities[a)]

Lei Fu,[1,b)] Gary Scavone,[2] and Claudia Fritz[3]
[1]Department of Mechatronic Engineering, Luoyang Institute of Science and Technology, Luoyang, Henan 471023, China
[2]Computational Acoustic Modeling Laboratory, Schulich School of Music, McGill University, Montreal, Quebec H3A 1E3, Canada
[3]Institut Jean le Rond d'Alembert, UMR 7190, CNRS/Sorbonne Université, Paris 75005, France

ABSTRACT:
The brand and model of strings used on violins are considered to play an important role in their playability and sound quality. An experiment was designed to test whether and how violin players perceive changes in violin qualities when strung with different strings. Three models of strings were chosen for this study: Dominant, Kaplan, and Pro-Arté strings. Two violins with similar sound and playing qualities were selected and professional and advanced student violinists were invited to play and compare the violins in two locations: Oberlin, Ohio and Montreal, Canada. Both violins were initially strung with Dominant strings and subjects rated the differences between the two (violin 2 compared to violin 1) according to eight criteria. Then, the strings of violin 2 were changed to a different type and the subjects again rated the differences between the two violins. In Oberlin, subjects compared Dominant and Kaplan strings in two sessions. In Montreal, subjects compared Dominant, Kaplan, and Pro-Arté strings in three trials. No statistically significant results in the perception of the string type differences were found in either location except that violin 2 was found to be significantly brighter with Dominant strings compared to Pro-Arté strings in the Montreal experiment.
https://doi.org/10.1121/10.0036894

(Received 13 December 2024; revised 16 May 2025; accepted 29 May 2025; published online 25 June 2025)

[Editor: Andrew Morrison] Pages: 4604–4612

I. INTRODUCTION

The quality of a violin depends on a number of factors: the physical characteristics[1] and vibrational properties of the plate,[2] the vibrational properties of the strings, the bridge, soundpost,[3] bassbar, and perhaps even the varnish.[4] It has been a long-standing goal of violin research to correlate the changes in these factors with the perceptual qualities of the violin.

Several studies have attempted to relate mechanical characteristics to violin qualities. Moral and Jansson[5] conducted a study employing two professional violinists who rated, from bad to good, the qualities of 24 violins. Based on bridge admittance measurements on the 24 violins, they suggested that the signature modes below 600 Hz and the bridge hill in the 2–3 kHz range are important for violin sound quality. Dünnwald[6] attempted to deduce objective quality parameters from experiments by measuring approximately 700 violins, including 53 old Italian violins, 75 violins of old masters, 300 violins made by masters after 1800, and approximately 180 factory made instruments. He suggested four important frequency bands for the assessment of violin sound quality. Specifically, he associated a large amplitude in the 650–1300 Hz band with "nasality" and a low amplitude in the 4200–6400 Hz band with "clarity," but these associations were made without reporting any perceptual testing. Perceptual tests were conducted later by Fritz *et al.*,[7] using virtual violins, and their results contradict, to a large extent, Dünnwald's perceptual hypotheses. Bissinger[8] 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 five, 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, *A*0. The radiation of this mode was significantly stronger for good than for bad violins. It is uncertain whether the results of this study are reliable or generalizable because the influences of parameters, such as visual condition and choice of bow, were not controlled and the number of subjects making the assessments was very low.

In recent years, scholars have conducted well-controlled perceptual evaluations of violin qualities under playing conditions. Saitis *et al.*[9] and Saitis *et al.*[10] performed a series of experiments to investigate violinists' evaluation process. Twenty skilled players participated in the first experiment in Saitis *et al.*[9] and the results showed that players consistently ranked the violins in terms of preference in different trials and on different days. The lack of agreement between different individuals, however, was significant. They also found that players tend to agree to some extent on

[a)]Portions of this article previously published in Fu *et al.*, Proceedings of the 176th Meeting of Acoustical Society of America and 2018 Acoustics Week in Canada, Victoria, Canada (November 5–9,2018).
[b)]Email: lei.fu2@mail.mcgill.ca

https://doi.org/10.1121/10.0036894

"richness" and "dynamic range" as criteria for determining preference,[9] and players were better able to discriminate between violins in playing tasks than in listening tasks.[10] In 2012 and 2014, Fritz *et al*.[11] and Fritz *et al*.[12] designed two experiments to examine musicians' preference between distinguished old Italian violins and new violins made by professional violin makers. The studies found that the violinists could not tell old violins from new ones at better than chance levels. 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 are of most concern, and how these qualities relate to physical characteristics of the instruments, whether old or new. In 2021, Fu *et al*.[13] explored violinists and luthiers' perception of soundpost tightness through a playing test and a listening test employing a length-adjustable carbon fiber soundpost. The results of the playing test showed that the optimal soundpost length varied among subjects, and they could not discriminate soundpost length variation of $\leq$0.11 mm at above chance levels. In the listening test, subjects could recognize soundpost lengths with a difference of about 0.2 mm at better than chance level.

It is known that with the development of the violin, the traditional gut strings installed on the violin were gradually replaced by metal wound gut or synthetic (mainly nylon) cores or single steel strands. The basic theory of the mechanical and acoustical properties of the string,[14] and the stick-slip Helmholtz motion of the bowed string[15] are well understood today. More recently, Woodhouse and Lynch-Aird[16] discussed the various factors that constrain the choice of musical instrument strings, especially the limit arising from the effect of material damping. The authors also explained the relationship between the bending stiffness of the string and the damping limit, as well as inharmonicity. Detailed results relating to monofilament strings were presented. About measurements with violin strings of different commercial brands, Pickering[17,18] carefully measured the physical properties of some brands of violin strings that were widely used during the 1980s, including steel strings, synthetic core strings, and gut core strings. He compared the elasticity, tension, the time that new-fitted strings take to stabilize, etc. Firth[19] investigated the construction details employing a scanning electron microscope and measured the inharmonicity of different high quality commercial violin strings in 1987. The author indicated that the preference order of the tested strings by players was inversely proportional to inharmonicity. However, no perceptual experiment related to players' preference for strings was reported.

Thus, in this study, a perceptual experiment was designed to investigate the influences of different types of strings on the violin qualities.[20] A playing-based evaluation approach was adopted with controlled experimental conditions. A detailed description of the experimental design is presented in Sec. II. Results and analyses are provided in Sec. III. Concluding remarks and discussions are given in Sec. IV.

II. 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 the detailed procedure.

A. General design

The aim of this experiment was 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.

B. Test violins and strings

A pool of similar student quality violins (all with a value of about $600 Canadian dollars), with the same type of strings, was 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. Detailed information about the three sets of strings, i.e., core, wound material, and tension are displayed in Table I. All strings employed in this experiment were in medium tension. The strings labeled "Dominant" in this study were a set of Thomastik 135 Dominant for the *G*, *D*, and *A* strings and a Pirastro gold for the *E* string (Thomastik-Infeld GmbH, Vienna, Austria). They were installed on both violins initially. According to the luthiers, Dominant strings were very commonly used, especially among student players. The manufacturer claims that the Dominant strings provide a soft, warm, tone with clear, stable intonation, and they are rich in overtones. The Pirastro gold *E* string is declared by the manufacturer to respond easily and sound brilliant and powerful. The other two types of strings were 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 (D'Addario & Company, Farmingdale, NY). The Kaplan strings were a set of model KA310. According to the string maker, the Kaplan strings offer a bold, warm tone and flexible response. The strings allow players to convey the rich tonal palette of their violins with enhanced projection; thus, they suit most advanced players. The Pro-Arté strings were a set of model

https://doi.org/10.1121/10.0036894

TABLE I. Detailed information of three sets of strings.

		G	*D*	*A*	*E*
Dominant set	Core	Synthetic	Synthetic	Synthetic	Steel (Pirastro gold)
	Wound	Silver	Aluminum	Aluminum	Unwound
	Tension (lbs)	9.9	9.1	12.1	17
Kaplan set	Core	Synthetic	Synthetic	Synthetic	Tinned carbon steel
	Wound	Silver	Silver	Aluminum	Unwound
	Tension (lbs)	10.5	10.5	12.5	17.5
Pro-Arté set	Core	Nylon	Nylon	Nylon	Tinned high-carbon steel
	Wound	Silver	Aluminum	Aluminum	Unwound
	Tension (lbs)	10.6	11.3	12	16.8

J56, manufactured for a lower tension to maximize tonal blend, response, and playability, and are suitable for players who seek a warm, dark tone. The cost of the three sets of strings (Dominant, Kaplan, and Pro-Arté) was approximately \$78, \$108, and \$49 American dollars, respectively. All strings were relatively new at the beginning of the experiment.

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. 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 makes and prices can lead to perceptual differences in violin qualities and what the perceptual differences are. On the other hand, synthetic strings are more popular than gut or steel strings among modern violinists; thus, the results should have more practical significance.

C. Venues and controls

This experiment took place at two locations. The first was at Oberlin College, Oberlin, OH. 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, etc.) on judgement, the participants were asked to wear dark sunglasses and the lighting in the room was reduced. 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 an unfamiliar bow. For this experiment, it was decided to allow players to use their own bow to evaluate the violins, as was done in several previous playing tests.[9–12] The violinists were given the option to either use their own shoulder rest, no shoulder rest, or one we provided (Kun Original model).

D. 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 feedback 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 Chinese speakers); average age = 28 years, standard deviation (SD) = 8 years, range = 21–52 years. They had at least 16 years of playing experience (average years of playing = 23 years, SD = 8 years, range = 16–45 years; average years of training = 20 years, SD = 5 years, range = 13–32 years; average hours of practicing per week = 28 h, SD = 10 h, range = 1–45 h). The estimated prices of their own violins ranged from \$6–\$40 K. Eighteen subjects described themselves as professional musicians, five were doctoral candidates in music performance, four had master's degrees in music performance, seven were currently master students in music performance, and one 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%)].

E. Detailed procedure

The first part of this experiment took place during the 16th 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 from Montreal. The experiment consisted of two sessions. Each session lasted approximately 20 min 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 min 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.[9,21] A short definition was provided for each criterion. The list of criteria and their definitions are given in Table II.[9,21] More explanations were provided orally whenever needed by the subjects.

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. 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 in quality for violin 2 with higher quality strings, which would reduce the difference between the two violins. Subjects were given 2 min to rate each criterion.

TABLE II. Definitions of rating criteria.

Responsiveness	*Responsiveness* describes how fast the violin can respond to different bowing techniques by the violinist, and how easy 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.

It took 3 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 subsequent 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 1–1.5 h. 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 in 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 min 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 (PIRASTRO GmbH, Offenbach am Main, Hessen, Germany), 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. Because of the damaged bridge, the condition of violin 2 changed. Consequently, including more subjects was not possible for this specific experiment.

III. 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 Sec. II E, the first subject in Montreal (subject 10) did not evaluate the experimental condition of the Kaplan strings, and subject 19 was not able to evaluate the

Pro-Arté strings. 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.

A. Comparison between $D1$-$D2$ and $D1$-$K2$ experimental conditions based on Oberlin results

During the first session, violins 1 and 2 were both strung with Dominant strings. The across-subjects average criteria difference ratings are shown in Fig. 1. Error bars of two-sided 95% confidence interval (CI) (all CIs are two-sided 95% intervals throughout this article) 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 *resonance*, *brightness*, and overall quality difference ratings remained approximately 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 experiment 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 experiment conditions criteria difference ratings by all subjects were normal, except for clarity. Thus, paired-samples t-tests were carried out on the other seven criteria difference

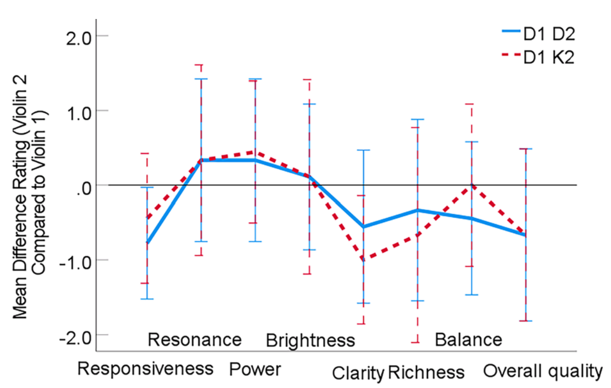

FIG. 1. Across-subjects average of the criteria difference ratings (error bar = 95% CI of the mean) for both sessions in Oberlin.

ratings. The results showed that the differences between the two conditions on these seven criteria were not significant, with the absolute value of paired samples $t(8) \leq 0.936$, $p \geq 0.377$. A 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, $z = 0.707$, $p = 0.48$.

B. Comparison between each pair of experimental conditions based on Montreal results

In this section, a comparison between each pair of experimental conditions was conducted based on the results obtained in Montreal. Three figures in Fig. 2 display the across-subjects average criteria difference ratings for each pair of experimental conditions with error bars of two-sided 95% CI. As mentioned in Sec. II E, the subjects who participated in each of the three pairs of experimental conditions were different. In Fig. 2(a), the analysis results were based on nine subjects (subjects 11–19); in Fig. 2(b), the analysis results were on the basis of eight subjects (subjects 11–18); and in Fig. 2(c), the results were based upon nine subjects (subjects 10–18).

In Fig. 2(a), the comparison was between $D1$-$D2$ and $D1$-$K2$ experimental conditions. For the $D1$-$D2$ condition, 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. For the $D1$-$K2$ condition, i.e., violin 2 was changed to Kaplan strings and the same set of Dominant strings were kept on violin 1; 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 Sec. III A 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 $z \leq 1.192$, $p \geq 0.233$.

In Fig. 2(b), the comparison was between $D1$-$K2$ and $D1$-$P2$ experimental conditions. 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. Statistical analysis revealed that 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 $t(7) \leq 1.59$, $p \geq 0.156$; related-samples Wilcoxon signed rank test $z = 1.622$, $p = 0.105$.

In Fig. 2(c), we compared the $D1$-$D2$ and $D1$-$P2$ experimental conditions. For the $D1$-$D2$ condition, the observed mean difference ratings of all criteria were negative. For the

https://doi.org/10.1121/10.0036894

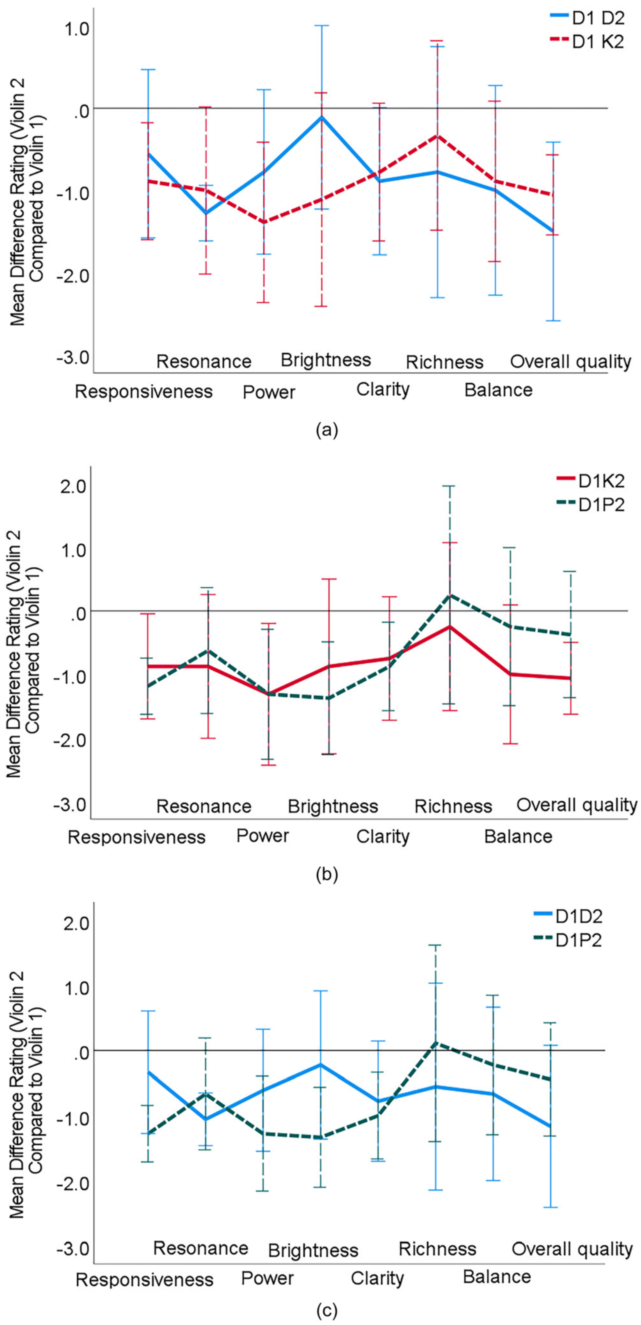

FIG. 2. Across-subjects average of the criteria difference ratings (error bar = 95% CI of the mean) for every two trials in Montreal.

$D1$-$P2$ condition, we can see that *resonance*, *balance*, *richness*, and overall quality of violin 2 were improved while its *responsiveness*, *power*, *brightness*, and *clarity* deteriorated when strung with Pro-Arté strings compared to Dominant strings. Of the improvements observed, *richness* was most noticeable: violin 2 with Pro-Arté strings became richer than violin 1 with Dominant strings. Statistical analysis indicated 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 $z = -2.06$, $p = 0.039$).

C. 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 (subjects 11–18). Figure 3 shows the across-subjects average criteria difference ratings for each experimental condition with error bars of two-sided 95% CI. 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 the *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 experiment 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 experiment 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

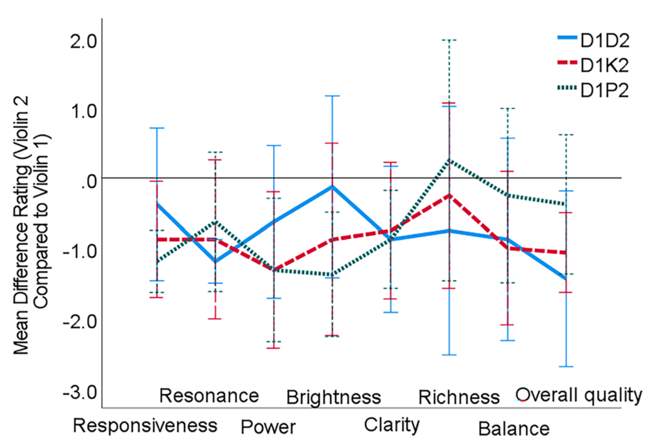

FIG. 3. Across-subjects average of the criteria difference ratings (error bar=95% CI of the mean) for three trials in Montreal.

of the difference ratings for every criterion across the three experimental conditions was the same could not be rejected, $\chi^2(2) \leq 5.7, p \geq 0.58$.

D. 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, we conducted independent-samples *t*-tests or independent-samples Mann–Whitney U tests (when the normal distribution assumption was violated).

When comparing Figs. 1 and 2(a), 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 3 months between these two parts of the experiment and the differences between the Oberlin and Montreal results might be partly attributable to seasonal changes and perhaps effects related to frequent changing of strings (for violin 2) in Montreal. 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 $U \leq 55.5$, $z \leq 1.357$, $p \geq 0.175$.

E. Relationship between overall quality and attribute ratings

We analyzed the relationship between the overall quality difference ratings and attribute difference ratings through a multiple rating-regression model and the computation of partial correlations. 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 are shown in Table III. Therefore, the multiple regression equation can be written as

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

While only the coefficients of richness and resonance were significant at the 0.05 level as shown in the last column of Table III (highlighted in boldface), all attribute difference ratings correlated with the overall quality difference ratings positively except the *brightness* difference ratings. The result of $R^2 = 0.635$ implies that the seven criteria difference ratings can explain 63.5% of the variation of the overall quality difference ratings.

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 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 the 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* was controlled by the calculation of ρ_p [*resonance*, overall quality $\cdot$ (*responsiveness*, *power*, *brightness*, *clarity*, *richness*, *balance*)].

Partial correlation coefficients ρ_p were computed between each of the attribute difference ratings and the

TABLE III. Multiple rating-regression analyzing the attributes that affect the overall quality difference ratings. $R = 0.797$; $R^2 = 0.635$; adjusted $R^2 = 0.568$; $F = 9.438$.

	Unstandardized coefficients		Standardized coefficients		
Independent variable	B	Std. Error	β	t	p
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

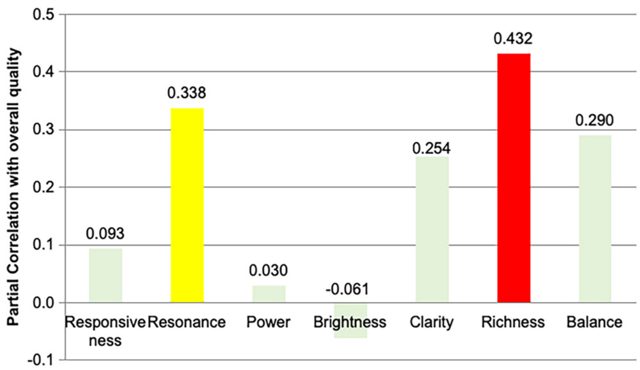

FIG. 4. Partial correlation coefficient ρ_p between difference ratings of each attribute and overall quality.

overall quality difference ratings for all subjects involved in this experiment. The results are shown in Fig. 4. *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 $\rho_p(38) \leq 0.29$ ($p \geq 0.069$).

IV. DISCUSSION AND CONCLUSION

This study investigated how different string types 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 experiment 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 $D1$-$P2$. That is to say, violin 2 was found to be significantly brighter with Dominant strings compared to Pro-Arté strings. The finding was somewhat consistent with the statement about the string characteristics by the manufacturer: Pro-Arté strings are suitable for players who seek a "dark" tone. 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. It was unexpected that the results of the three experimental conditions would lack significant differences. 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. However, having a greater number of subjects could help reduce random error effects.[22] 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. Also, 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. Also, 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 Figs. 1–3. This is similar to what has been observed previously in playing tests (e.g., Saitis *et al.*[9]) 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 similarly affected 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. This is in line with a previous finding[9] that players tend to agree that *richness* is a determinant criterion in preference evaluations.

ACKNOWLEDGMENTS

This project has been funded by the Natural Science Foundation of Henan province (Item No. 232300421356), Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Hector and Ada Ma China Scholarship Council and China Scholarship Council. We would like to thank Tom Wilder for his helpful suggestions at the beginning of the study. We thank Stephen McAdams for fruitful discussions about the analysis of the results. We are grateful to Fan Tao and D'Addario & Company for donating some of the strings and Wilder & Davis for loaning the violins used in this experiment.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The ethics certificate number is McGill REB File #430-0415, titled "Perception and Evaluation of Music Instruments." Informed consent was obtained from all subjects who participated in this experiment.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon request.

[1]J. Woodhouse, "The acoustics of the violin: A review," Rep. Prog. Phys. **77**, 115901–115942 (2014).
[2]C. M. Hutchins, "The acoustics of violin plates," Sci. Am. **245**, 170–187 (1981).
[3]G. Bissinger, "Some mechanical and acoustical consequences of the violin soundpost," J. Acoust. Soc. Am. **97**, 3154–3164 (1995).
[4]J. C. Schelleng, "Acoustical effects of violin varnish," J. Acoust. Soc. Am. **44**, 1175–1183 (1968).
[5]J. Alonso Moral and E. Jansson, "Input admittance, eigenmodes and quality of violins," Report STL-QPSR, 2-3, Department for Speech, Music and Hearing, Royal Institute of Technology, Stockholm, Sweden (1982).
[6]H. Dünnwald, "Deduction of objective quality parameters on old and new violins," Catgut Acoust. Soc. J. **1**, 1–5 (1991), available at https://oac.cdlib.org/findaid/ark:/13030/c8gt5p1r/dsc/?query=catgut%20acoustical%20society#ref1.
[7]C. Fritz, A. F. Blackwell, I. Cross, J. Woodhouse, and B. C. J. Moore, "Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties," J. Acoust. Soc. Am. **131**, 783–794 (2012).
[8]G. Bissinger, "Structural acoustics of good and bad violins," J. Acoust. Soc. Am. **124**, 1764–1773 (2008).
[9]C. Saitis, B. L. Giordano, C. Fritz, and G. P. Scavone, "Perceptual evaluation of violins: A quantitative analysis of preference judgments by experienced players," J. Acoust. Soc. Am. **132**, 4002–4012 (2012).
[10]C. Saitis, G. P. Scavone, C. Fritz, and B. L. Giordano, "Effect of task constraints on the perceptual evaluation of violins," Acta Acust. united Acust. **101**, 382–393 (2015).
[11]C. Fritz, J. Curtin, J. Poitevineau, P. Morrel-Samuels, and F.-C. Tao, "Player preferences among new and old violins," Proc. Natl. Acad. Sci. U.S.A. **109**, 760–763 (2012).
[12]C. Fritz, J. Curtin, J. Poitevineau, H. Borsarello, I. Wollman, F.-C. Tao, and T. Ghasarossian, "Soloist evaluations of six Old Italian and six new violins," Proc. Natl. Acad. Sci. U.S.A. **111**, 7224–7229 (2014).
[13]L. Fu, C. Fritz, and G. P. Scavone, "Perception of violin soundpost tightness through playing and listening tests," J. Acoust. Soc. Am. **150**, 540–550 (2021).
[14]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer, New York, 1998).
[15]J. C. Schelleng, "The bowed string and the player," J. Acoust. Soc. Am. **53**(1), 26–41 (1973).
[16]J. Woodhouse and N. Lynch-Aird, "Choosing strings for plucked musical instruments," Acta Acust. united Acust. **105**, 516–529 (2019).
[17]N. C. Pickering, "Physical properties of violin strings," Catgut Acoust. Soc. J. **44**, 6–8 (1985), available at https://oac.cdlib.org/findaid/ark:/13030/c8gt5p1r/dsc/?query=catgut%20acoustical%20society#ref1.
[18]N. C. Pickering, "Elasticity of violin strings," Catgut Acoust. Soc. J. **46**, 2–3 (1986), available at https://oac.cdlib.org/findaid/ark:/13030/c8gt5p1r/dsc/?query=catgut%20acoustical%20society#ref1.
[19]I. Firth, "Construction and performance of quality commercial violin strings," Catgut Acoust. Soc. J. **47**, 17–20 (1987), available at https://oac.cdlib.org/findaid/ark:/13030/c8gt5p1r/dsc/?query=catgut%20acoustical%20society#ref1.
[20]L. Fu, G. P. Scavone, and C. Fritz, "How different strings affect violin qualities," in *Proceedings of the 176th Meeting of Acoustical Society of America & 2018 Acoustics Week in Canada*, Victoria, Canada (November 5–9, 2018).
[21]C. Saitis, C. Fritz, G. P. Scavone, C. Guastavino, and D. Dubois, "Perceptual evaluation of violins: A psycholinguistic analysis of preference verbal descriptions by experienced musicians," J. Acoust. Soc. Am. **141**, 2746–2757 (2017).
[22]C. Robson, *Experiment, Design and Statistics in Psychology* (Penguin, Harmondsworth, UK, 1973).