Acta Acustica united with Acustica
Vol. 101 (2015) 382 – 393

DOI 10.3813/AAA.918834

Effect of Task Constraints on the Perceptual Evaluation of Violins

Charalampos Saitis[1], Gary P. Scavone[1], Claudia Fritz[2], Bruno L. Giordano[3]
[1] Computational Acoustic Modeling Laboratory, Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montréal, Québec H3A 1E3, Canada. charalampos.saitis@mail.mcgill.ca
[2] Lutheries-Acoustique-Musique, Institut Jean Le Rond d'Alembert, Université Pierre et Marie Curie, UMR CNRS 7190, 4 place Jussieu, 75005 Paris, France
[3] Institute of Neuroscience and Psychology, University of Glasgow, Scotland G12 8QB, United Kingdom

Summary
Results from two previous studies that involved free-playing evaluative tasks showed that players are self-consistent in their preference for violins and tend to agree of what particular qualities they look for in an instrument (in this case, "richness" and "dynamic range"). However, a significant lack of agreement between violinists was observed, likely because different players evaluate the same perceptual attributes in different ways. The present study thus investigated whether there will be more inter-player agreement if musicians evaluate violin richness and dynamic range by playing only certain notes in certain registers. Results showed that the more focused the task, the more self-consistent violinists are and the more they agree with each other. We further examined the evaluation of richness from playing versus listening tasks and observed that players were better able to discriminate between violins in the former than in the latter. Finally, the potential correlation of spectral centroid and tristimulus with violin richness were examined. Results showed that the perception of richness is likely associated with the relative amount of low- and mid-frequency partials in a given sound (i.e., low spectral centroid and high tristimulus 1 and 2), though more exploration would be necessary before drawing any conclusions.

PACS no. 43.66.Jh, 43.75.Cd, 43.75.De

1. Introduction

Since the classical period and the early Cremonese instruments of Amati and Stradivari, the basic lutherie of the violin and its bow has remained largely unchallenged. Its design combines visual charm with ergonomics and a precise acoustical function [1]. Despite a considerable amount of research on the dynamic behavior of the instrument, efforts to understand the link between measurements and the perceived quality of a violin have often been inconclusive [2]. A review of the relevant literature indicated limitations concerning the subjective evaluation process itself as well as the perspective of the musician [3]. Starting from the latter, the research presented here aims to contribute to the understanding of the following two key issues pertinent to the perceptual evaluation of violins: how consistent experienced performers are at assessing violins and whether there is agreement between individuals; and how different experimental situations might affect the consistency of the musicians' psychoacoustical judgements.

We previously carried out two experiments based on a carefully controlled playing-based procedure for the perceptual evaluation of violins [4, 5]. The first experiment examined intra-player consistency and inter-player agreement in violin preference judgements. In a first session, skilled violinists freely played a set of different violins and ordered them by preference. They repeated the ranking task five times and returned for a second, identical session 3–7 days after having completed the first session. We found that violinists were self-consistent in assessing violin quality but a significant lack of agreement between musicians was observed. In the second experiment, experienced performers freely played a set of different violins and rated them according to ease of playing, response, richness, balance (across all strings), dynamic range and preference (one violin on all scales at a time, in three blocks of repetitions). Results showed that whereas violinists tend to agree of what particular qualities they look for in an instrument – preference was strongly associated with richness and, to a lesser extent, dynamic range – the perception of the same attributes widely varied across individuals, thus likely resulting in large inter-individual differences in the preference for violins.

One hypothesis about the origin of the large inter-individual differences in violin preference is that players may take varying playing approaches to assess different attributes of the instrument. For example, player A may

---

Received 14 March 2014,
accepted 09 September 2014.

apply a broader palette of musical material and bowing gestures than player B and thus experience a wider range of the instrument's characteristics. In the previous two experiments, no playing constraints (e.g., specific repertoire) were imposed on the evaluation process. Participants were instead instructed to follow their own strategy with respect to what and how to play. To tease apart the effects of the playing approaches of different individuals, a new experiment was designed to investigate the perceptual evaluation of richness and dynamic range in playing tasks based on prescribed musical material and techniques. The objective was to compare intra-individual consistency and inter-individual agreement in constrained – i.e., playing only certain notes on certain registers – versus unconstrained – i.e., playing a certain excerpt from the violin repertoire – tasks for the cases of richness and dynamic range. We chose to focus on the perceptual characteristics of richness and dynamic range as they had been previously found to be highly correlated with violin preference.

Unlike the free-playing approach adopted in our previous studies, the idea of constrained versus unconstrained playing in this experiment concerned the playing range of the instrument on which violinists were permitted to focus (1 or 2 strings versus all strings) as well as the playing technique they could apply (strict versus loose instructions) during the evaluation procedure. In this respect, the idea of "unconstrained" playing (current study) is not similar to that of "free" playing (previous experiments). In the latter case participants would be encouraged to choose both their own materials and techniques – and those would often change from one trial to the next – whereas in the former case the musical material would be common for all players.

We further investigated the perceptual evaluation of richness using the constrained-playing task, which was recorded, and a subsequent listening task (using the previously recorded sounds). The goal was to compare the evaluation of richness from playing versus listening tasks in order to better understand whether it is based on different criteria and/or perceptual processes in the two settings. From the perspective of the musician, vibrations are capable of providing tactile and proprioceptive cues that contribute to the perception of the radiated sound, so that the player can assess their interaction with the instrument cross-modally [6, 7, 8]. Woodhouse pointed out that what distinguishes one violin from another lies not only on its perceived sound quality, but also on ergonomic considerations, as in how the violinist "feels" the instrument or how easy it is to control a "good" sound [9]. Fritz *et al.* examined the differences between preference judgments made by violinists in playing versus listening situations in conjunction with psycholinguistic analyses of verbal descriptions of the musicians' experience [10]. Results suggested that the overall evaluation of a violin as reflected in the verbal responses of the performers varies between playing and listening settings, the former invoking descriptions influenced not only from the produced sound but also by the interaction between the player and the instrument. More recent results indicate the presence of tactile-only cues in the perception of violin quality by performers [11].

Finally, we investigated potential spectral correlates of violin richness. From verbal descriptions collected in the first of our earlier experiments (violinists were asked to comment on their preference criteria through answering open-ended questions), we previously explored how violin quality is conceptualised by performers [12]. The concept of timbral richness emerged as a key perceptual factor in assessing violin quality, supporting our previous observation that preference is strongly associated with perceived richness in the sound [5]. Descriptions of violin timbre such as full, deep, complex and dark were also found to be conceptually close to richness. As part of a standardised qualitative procedure for evaluating violin quality, Bissinger and Gearhart suggested that a *complex* sound "has many overtones and color," a *deep* sound "includes lower harmonics well" and a *dark* sound contains "lots of lower harmonics" [13]. In a more recent study, 61 common descriptions of violin tone qualities were arranged by violinists on a two-dimensional map, so that words with similar meanings lay close together, and those with different meanings lay far apart [14]. One of three emerging dimensions for the characterization of violin quality was *warm/rich/mellow* versus *metallic/cold/harsh*; the authors suggested it relates to spectral balance, with undesirable qualities associated with excessive high-frequency content or too little low-frequency content.

Starting from the hypothesis that the concept of richness is associated with the perceived amount of harmonics in a given sound, we focused on the features of spectral centroid and the tristimulus timbre model. The former is a well-known measure of energy distribution in the spectrum and has been shown to be highly correlated with perceived timbral brightness [15, 16]. The tristimulus model comprises three ratios that describe timbre in a way analogous to the three primary colors in vision [17]. They measure the relative presence (intensity) of the fundamental or first harmonic (T1), the second, third and fourth harmonics (T2), and all partials above and including the fifth harmonic (T3) in a given sound. A larger T1 is associated with a "strong fundamental," while a larger T2 means "strong mid-frequency partials" and a larger T3 means "strong high-frequency partials." Results from a listening test showed that violin notes described as *sharp* and *narrow* were associated with higher and lower spectral centroid values respectively [18]. Łukasik proposed that a violin sound described as *dark* may be characterized by a spectral centroid of less than 1200–1400 Hz, with higher values indicating a *bright* or *sharp* sound [19]. She further argued that a violin sound with high T1 and T3 values may be described as *deep* versus *empty* in the opposite case. Similarly, a high T1 and a low T3 value may indicate a *full* sound versus *flat* in the reverse configuration. These suggestions were subsequently tested by Łukasik using the recordings of 53 violins (AMATI database [20]) but no distinct trends were observed. For the purposes of this study, we used the recorded sounds from the richness constrained-playing task.

This paper is organized as follows: Section 2 describes the experimental materials and methods. The results of the statistical analyses are reported in section 3, while observations or conclusions related to the results are discussed in section 4. Possible future directions are also considered in section 4. Section 5 summarizes the main findings and concludes the work.

2. Method

2.1. Participants

Sixteen skilled string players took part in this experiment (8 females, 8 males; average age = 32 yrs, SD = 8 yrs, range = 21–55 yrs; 9 native English speakers, 2 native French speakers, 5 other). They had at least 15 years of violin experience (average years of violin training = 25 yrs, SD = 8 yrs, range = 17–48 yrs; average hours of violin practice per week = 15 hrs, SD = 11 hrs, range = 3–35 hrs), owned violins with estimated prices ranging from \$3k to \$70k, and were paid for their participation. Eleven participants described themselves as professional musicians, and 10 had higher-level degrees in music performance (MMus, MA, DMus, DMA). They reported playing a wide range of musical styles [classical (81%), folk (13%), jazz/pop (6%), and contemporary (6%)] and in various types of ensembles [symphonic orchestra (38%), chamber music (31%), folk/jazz band (25%), and solo (19%)].[1]

2.2. Violins

Five violins of different make (Europe, North America, China), year of fabrication (1914–2011) and price (\$2.7k–\$71k) were used (see Table I). They were chosen from two local luthier workshops in order to form, as much as possible, a set of violins with a wide range of characteristics. The violins had not been played on a regular basis as most were from the available sales stock of the workshops. The respective luthiers provided the price estimates and tuned the instruments for optimal playing condition based on their own criteria. Violin D was included in our previous studies (Experiment I, labelled F, highest preference score; and Experiment II, labelled H; see [5]). The fact that some violins may have been less optimally tuned or had strings of varying quality was not a concern, as that should not influence the consistency of the evaluations. Participants' own violins were not included in the set of instruments in order to avoid possible preference biases caused by the mere exposure effect [21] by which familiarity with a stimulus object increases preference toward it. Violinists were given the option to either use a provided shoulder rest (Kun Original model), or use their own, or use no shoulder rest.

Table I. Violins used in the study. Names of currently active luthiers are not provided for confidentiality purposes. (The origin of violin D is based on a luthier's informal appraisal, as there is no information regarding the make and age of this violin.)

Violin	Origin	Luthier	Year	Price
A	Italy	Contino	1916	\$71k
B	Switzerland	-	2003	\$30k
C	Denmark	Hjorth	1914	\$20k
D	(Germany)	Unknown	Unknown	\$10k
E	China	-	2011	\$2.7k

2.3. Controls

Anecdotal evidence strongly suggests that some visual information, such as the color of the varnish, the grain of the wood, or identifying marks of the violin, may influence judgment. More specifically, possible recognition of the instrument's make and origin is likely to produce preference biases (e.g., old Cremonese violins are often considered excellent and hence preferred over modern instruments). To circumvent the potential impact of visual cues on preference while ensuring a certain level of comfort for the musicians, as well as safety for the instruments, low light conditions were used and participants were asked to wear dark sunglasses. As such, violinists could provide unbiased assessments while still retaining some visual contact with the instruments.

A critical issue when conducting violin playing tests is the choice of a bow. In the present study, two options were considered: using a common bow across all participants or asking players to use their own bow. Although neither solution is ideal, by considering the bow as an extension of the player (second option) the potential problems of using a common bow (e.g., participants being uncomfortable with a bow they are not familiar with) were avoided. A common bow would potentially further trigger a similar quality debate [22]. Having the participants use the bow that they are most familiar with was also felt to be more representative of how violinists assess instruments while in the process of purchasing one.

The experiment took place in an acoustically dry room (surface = $46.8\,\mathrm{m}^2$, reverberation time $\approx 0.3\,\mathrm{s}$) to help minimise the effects of room reflections on the direct sound from the violins [13].

2.4. Tasks

For each of the perceptual characteristics of richness and dynamic range, a constrained- and an unconstrained-playing task were designed (see section 1 on how "constrained" and "unconstrained" playing are defined for the purposes of this study). The constrained task was different for each of the attributes (i.e., different musical material and technique) while the unconstrained task was recurrent across the attributes. The unconstrained task was also used for the evaluation of preference.

The richness-constrained task was focused on the lower register of the violin, in particular on the *G* string (see Fig-

---

[1] Subjects were allowed to select as many categories as appropriate, thus the total percentages do not sum to 100%.

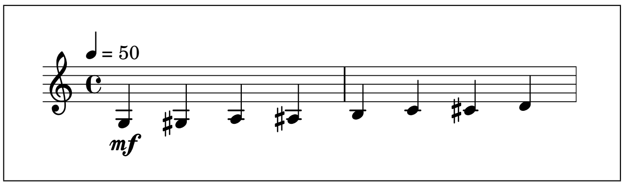

Figure 1. Constrained musical task for the perceptual evaluation of violin richness. See text for details on playing instructions.

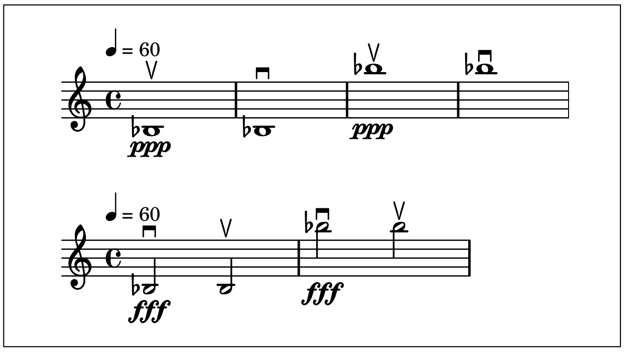

Figure 2. Constrained musical task for the perceptual evaluation of violin dynamic range. See text for details on playing instructions.

ure 1). It involved playing the eight notes of the chromatic scale *G2 → D3 détaché*, first *without vibrato* followed by a repetition *with vibrato*. Participants were instructed to follow a 50 bpm tempo and use the whole bow. The dynamic range-constrained task comprised only the notes *B♭2* on the *G* string and *B♭4* on the *E* string (see Figure 2). Participants were instructed to follow a 60 bpm tempo and play *détaché, without vibrato*, as soft and as loud as possible to obtain a clear sound (i.e., the sound doesn't break).

The unconstrained task used for the evaluation of both richness and dynamic range as well as for preference involved playing the opening solo passage from Max Bruch's Violin Concerto No. 1 in *G* Minor, Op. 26 (Movement I: Prelude; see Figure 3). The particular excerpt was chosen because it incorporates the whole range of the instrument (as opposed to the two constrained tasks) as well as a variety of techniques and dynamics. Participants were instructed to follow the temporal and expressive markings as much as possible (i.e., a certain degree of personal interpretation was expected).

2.5. Recordings

The richness-constrained task was recorded by each participant in order to (a) capture the stereo stimuli for the listening test and (b) extract audio features. For (a), the X-Y stereo microphone positioning technique using a pair of condenser microphones with cardioid patterns (DPA 4011-TL) was followed. The two microphone capsules were mounted on top of each other (i.e., coincident position) at an angle of 90 degrees, their center facing directly at the top side of the played violin from a distance of 2 m. The recorded musical phrases were digitised through a RME Micstasy 8-channel microphone preamplifier and saved in 16-bit, stereo 48 kHz WAV format. For (b), a 1/2-inch free-field microphone (Brüel & Kjær Type 4190-L-001 with Type 2669-L preamplifier) with a sound quality conditioning amplifier (Brüel & Kjær Type 2672) were used. The microphone was positioned 90 cm from the played violin, facing directly at its top side. The gain of the amplifier was set at 20 dB and a high-pass filter with a cut-off frequency of 20 Hz was selected. The recorded notes were saved in 32-bit, stereo 48 kHz WAV format.

2.6. Procedure

The first session (playing test) lasted two hours and was organised in three parts. The first part involved two training rankings with three violins, which were distinct from the five violins used in the actual study, to help participants familiarise themselves with each of the constrained-playing tasks respectively. In the second part, participants were asked to rank-rate (see next paragraph) the violins in terms of richness first and then dynamic range using the respective constrained tasks. This order was based on how much musical input was involved in each task: richness was assessed based only on the *G* string whereas dynamic range was evaluated based on both the *G* and *E* strings. Each task involved three repetitions (trials) and all players carried out the two tasks in the same order. Upon completing the last trial for the richness-constrained task, participants recorded the corresponding musical material on each of the five violins. In the third part, participants were asked to rank-rate the violins in terms of richness, dynamic range and preference according to the unconstrained task. Each of the three criteria was presented once in each of three subsequent blocks of trials. The order of presentation of the criteria within each block of trials was randomised (determined by computer calculations). In total, participants ranked-rated all violins $2 \times 3 + 3 \times 3 = 15$ times. The experimenter was constantly present in the room to facilitate the process.

In each trial, participants were first presented with all violins placed on a table in random order (determined by computer calculations) by the experimenter. Participants were then asked to simultaneously rate each violin using separate, identical on-screen sliders, thus providing a ranking of the five violins at the same time (see Figure 4). In addition to and independently of how they ordered the violins, participants were asked to indicate which of the instruments satisfied their perceived standard for the respective attribute or preference by setting a "limit of acceptability" (i.e., violins rated higher or equal to that threshold were flagged as "acceptable") on a separate on-screen slider. Participants had to move each slider (i.e., assess each instrument and set the acceptability limit) before being allowed to move to the next trial. In order to end a trial and start the succeeding one, participants clicked an on-screen button labelled "Done" that appeared only after all sliders had been moved. Participants were instructed to always rate their top choice as 1 and their lowest as 0. They were not allowed to assign the same rank-rating to two or more instruments. Participants were instructed to max-

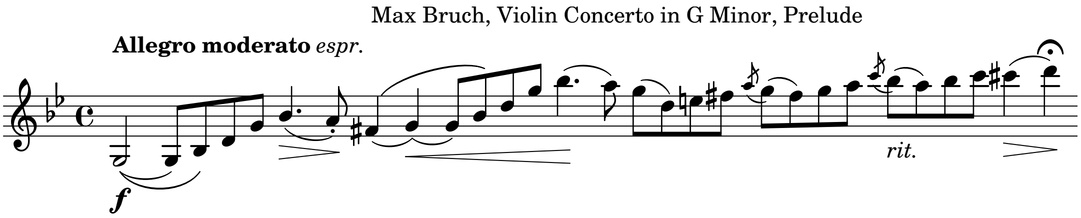

Figure 3. Unconstrained musical task for the perceptual evaluation of violin richness, dynamic range and preference. See text for details on playing instructions.

imise evaluation speed and accuracy. They were encouraged to play their own violin whenever they needed a reference point during the experiment. To minimise fatigue, participants were encouraged to take breaks between trials whenever needed. Upon completing the last trial, participants provided written responses to the questions "Did you have difficulty with any of the tasks?" (QP1) and "To what extent was wearing sunglasses disturbing?" (QP2).

Participants were asked to return for a second session 1–4 days after having completed the first session. It lasted thirty minutes and involved a listening task with three repetitions (trials). On each trial, participants were first presented with their own audio recordings in random order (determined by computer calculations) over closed, dynamic stereo headphones (Sennheiser HD 280). They were then asked to rank and rate the violins following the same procedure as in the first session (i.e., the interface and instructions were identical; recordings could be played as many times as deemed necessary). Upon completing the last trial, participants provided written responses to the questions "In this new condition (listening), did your overall perception of richness change?" (QL1), "On what criteria did you make your richness ranking this time? Did these criteria differ from the ones used in the previous condition (playing)?" (QL2) and "When you evaluate a violin, how important is sound in your overall judgement compared to vibrational characteristics of the instrument?" (QL3).

2.7. Analysis

For each task, consistency was defined as the concordance correlation between evaluations from different trials. The concordance correlation coefficient ρ_c is a special case of the Pearson product-moment correlation coefficient that measures departures from the equality lines with slopes ±45° [23, 24]. As such, ρ_c does not assume linear relationships, thus being a stricter index of agreement than Pearson's coefficient. For a given participant A, intra-individual consistency was estimated as the average of the ρ_c between ratings of A across all trials. Inter-individual consistency was given by averaging the ρ_c between ratings of A and those of all other participants across all trials. Note that according to this definition, the inter-individual consistency measures for participants A and B would be computed by considering the same set of 9 ρ_c measures between the 3 ratings of participant A and those of participant B. In order to minimise one source of dependence between the inter-individual consistency measures for different participants, correlations were distributed among participants at random (e.g., for participant A the inter-individual consistency measure considered 4 or 5 randomly selected $\rho_c\,(A, B)$ measures, whereas for participant B it included the other 5 or 4 respectively). However, there is another source of dependence as all correlations come from the same matrix and are therefore linked to each other. As a result, any statistical inferences on inter-individual consistency such as confidence intervals of the mean or parametric tests of statistical significance should be treated with caution.

Figure 4. User interface (example for the case of richness).

3. Results

3.1. Evaluation of richness and dynamic range in constrained vs. unconstrained tasks

Three different analyses were carried out. First, the measures of intra- and inter-individual consistency for each of the tasks were assessed and compared. A two-way repeated-measures analysis of variance was employed to investigate the effects of condition (i.e., constrained versus unconstrained task) and attribute (richness versus dynamic range) on the measures of intra-individual consistency. Second, we assessed the extent to which certain self-reported characteristics of the participants explained their ability to be consistent across trials (e.g., whether "hours of practice per week" was correlated with self-consistency). A two-sample *t*-test was adopted to assess whether self-consistency significantly differed between professional and amateur violinists. Third, average ratings for each of the attributes and conditions for each violin were derived.

3.1.1. Intra- and inter-player consistency

The histograms in Figure 5 describe the distribution of intra- and inter-individual ρ_c coefficients for each task (including the listening task, the results of which will be discussed in section 3.2). In the same figure, the symbols above the histograms report the across-participants average (circle) and 95% confidence interval (error bar) of the intra- and inter-individual consistency scores. The following were observed:

- For the constrained tasks, the average measure of intra-individual consistency was substantially high for richness, average value = 0.7, but less so for dynamic range, average value = 0.47.
- Concerning the unconstrained tasks, the average measure of intra-individual consistency was moderately high for richness and preference, average value = 0.44 in both cases, but considerably lower for dynamic range, average value = 0.29. Marginally significant differences emerged between the intra-individual consistency measured for the preference task on the one hand, and the richness and dynamic range tasks on the other [paired samples $t(15) \leq 1.87,\ p \geq 0.08$].[2]
- Following the notable decrease in self-consistency from the constrained to the unconstrained tasks for each of the two attributes as well as from richness to dynamic range in both the constrained and unconstrained tasks, a two-way repeated-measures analysis of variance on the respective measures of intra-individual consistency revealed that both condition (constrained versus unconstrained task) and attribute (richness and dynamic range) had a significant effect on how self-consistent participants were in their judgements [$F(1,15) = 8.64$, $p = 0.01$ and $F(1,15) = 7.72$, $p = 0.01$, respectively].
- The interaction between attribute and condition was not significant [$F(1,15) = 0.25,\ p = 0.63$], hence the two factors do not appear to influence each other here (i.e, in the circumstances related to the particular experiment).
- Considering all five tasks, most participants had average self-consistency above 0.4 (13 participants, 81.25%); five participants (31.25%) had average self-consistency of more than 0.5; three violinists (18.75%) had average self-consistency of less than 0.3; and only player 4 was highly, almost perfectly, self-consistent.
- Going from the second to the third trial, average self-consistency dropped noticeably (−0.1) for the dynamic range tasks, while an important increase of about 0.2 was observed for the preference task.

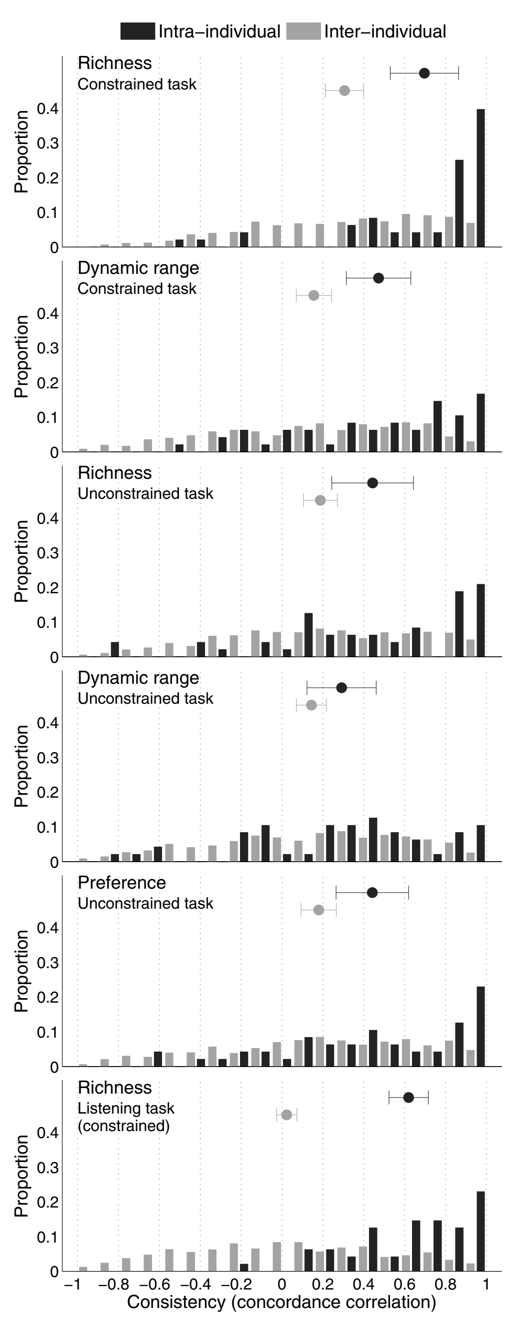

Figure 5. Distribution of intra- and inter-individual ρ_c coefficients: 1 corresponds to perfect consistency, 0 corresponds to no consistency, -1 corresponds to perfect anti-consistency (i.e., exactly opposite rankings given on different trials). The symbols above the histograms report the across-participants average of the intra- and inter-individual consistency scores (error-bar = 95% confidence interval of the mean; the ordinate for the symbols has been chosen arbitrarily for display purposes). See text for details on averaging of concordance correlations.

- Inter-individual consistency was generally low for both constrained and unconstrained tasks, $0.15 \leq$ average value ≤ 0.19, except for richness-constrained, average value = 0.31.

---

[2] By "marginally significant" we denote a value that is close enough to the typical threshold of $p = 0.05$ to be ruled out as not significant.

- In the second trial, there was more agreement between the participants for the unconstrained than constrained tasks, particularly for the cases of preference (wherein inter-player agreement made a +.35 jump) and dynamic range (+0.15). However, participants were much less consistent between themselves in the third trial for these two tasks. Across trials inter-player agreement overall increased except for the richness-constrained condition (−0.1).
- Finally, no significant correlation between the self-consistency of a participant and their level of agreement with the other violinists was observed.

3.1.2. Influence of participant characteristics

For each of the evaluation tasks, the association between the participant-specific measures of intra-individual consistency on the one hand, and the self-reported price of the owned violin, the years of violin training, and the weekly hours of violin practice on the other was assessed. This analysis was carried out by computing the Spearman rank correlation ρ_S between intra-individual consistency scores and participant characteristics. No association was found to be significant [absolute value of $\rho_S \leq 0.42,\ p \geq 0.11,\ df = 14$].

3.1.3. Professional vs. amateur performers

When evaluating richness, professional violinists tended to be slightly more self-consistent than amateur players in the constrained task, average value = 0.72 and 0.65, respectively, but considerably less self-consistent than amateurs in the unconstrained task, average value = .38 and 0.56, respectively. These differences were not significant [independent samples $t(14) \leq 0.37,\ p \geq 0.33$, equal variance]. In the case of dynamic range, professional violinists were more self-consistent than amateur musicians in both the constrained and unconstrained tasks, average value = 0.52 and 0.38, and 0.35 and 0.16, respectively. Again none of the differences was found to be significant [independent samples $t(14) \leq 1.16,\ p \geq 0.27$, equal variance]. It should be noted that due to the small sample size in one of the two groups (amateur players, N = 5), it is not surprising to find effects falling short of significance despite their relatively large magnitude. Finally, professional performers were significantly more self-consistent than amateur violinists in their preference judgements, average value = 0.55 and 0.21, respectively [independent samples $t(12.3) = 3,\ p = 0.01$, unequal variance].

3.1.4. Violin scores

For each of the violins, a task-specific score defined as the across-participants average rating of a violin throughout all trials was computed. The across-participants average violin rating scores for each task are shown in the top graph of Figure 6 (including the listening task, the results of which will be discussed in section 3.2). A different score was also computed for each of the violins by considering the proportion of times a violin was chosen as "acceptable" by the participants in terms of preference/richness/dynamic range. The "acceptability scores" for each violin and task are shown in the bottom graph of Figure 6.

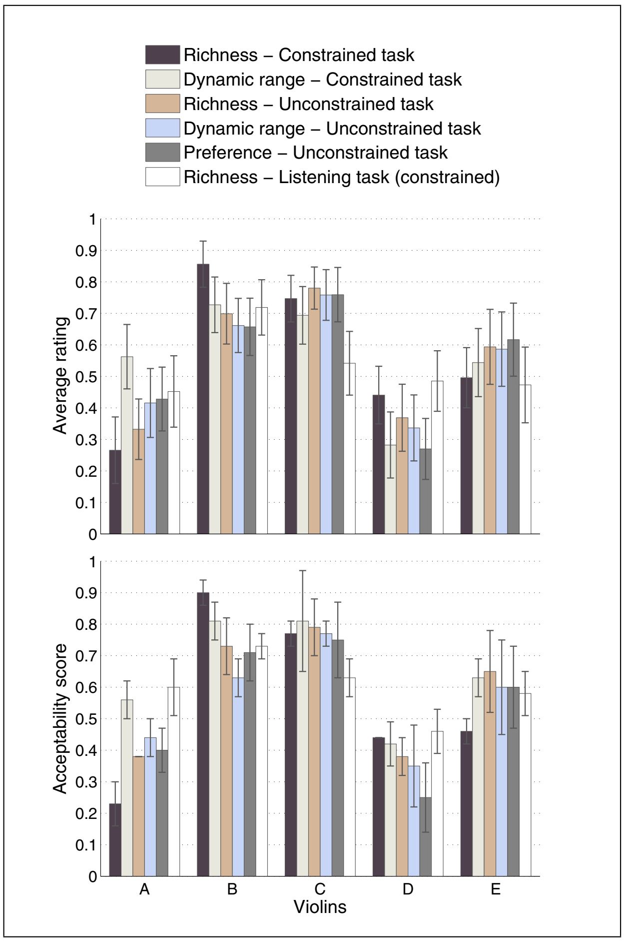

Figure 6. Across-participants and across-trials average rating (top) and acceptability (bottom) scores for each violin (error-bar = 95% confidence interval of the mean). See text for details on computing of scores.

Ordering the violins by any of the two scores revealed the same grouping pattern for all tasks: violins A and D always alternated between the two lower ranks, violin E was always placed in the middle position and violins B and C alternated between the two higher ranks (in the case of the constrained task for dynamic range, the grouping was only slightly different as violin E alternated with A). In particular, violin A was chosen as the least rich instrument and violin D as having the narrowest dynamic range consistently. Violin B was characterised as both the most rich and having the broadest dynamic range when evaluated in the constrained tasks; for the unconstrained tasks participants appeared to prefer violin C over B.

3.2. Evaluation of richness in playing vs. listening tasks

We followed similar analyses as in the previous section. A *t*-test was employed to investigate the effects of type (i.e., playing versus listening) on the measures of intra-individual consistency.

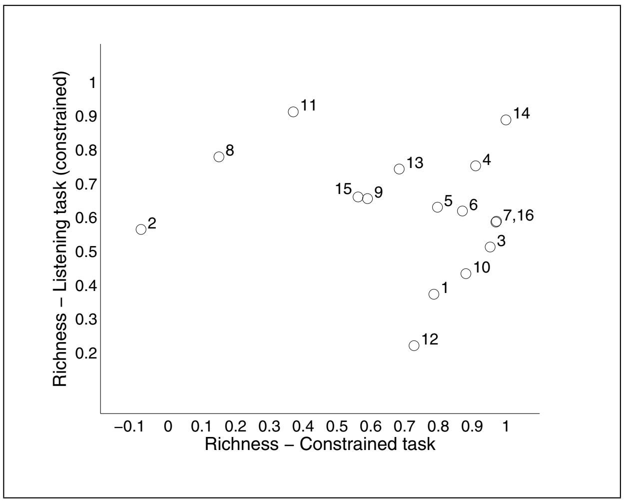

Figure 7. Measures of intra-individual consistency in playing versus listening tasks for the evaluation of richness (constrained task, see text for details; numbers represent participants).

3.2.1. Intra- and inter-player consistency

The distribution of intra- and inter-individual ρ_c coefficients in the listening task as well as the across-participants averages are reported in the bottom graph of Figure 5 (compare top versus bottom graphs for playing versus listening, respectively). For both tasks, the average measure of intra-individual consistency was substantially high, average value = 0.7 and 0.62 for the playing and listening tasks, respectively. No significant difference in the average intra-individual consistency between the two tasks was observed [paired samples $t(15) = -0.8$, $p = 0.45$]. This is further suggested by Figure 7 wherein the individual self-consistency measures for the playing task are plotted against the corresponding intra-individual measures for the listening task. Despite the good level of self-consistency observed in both tasks, agreement between players was considerably lower in the listening task compared to the playing task, average value = 0.02 and 0.31, respectively.

3.2.2. Professional vs. amateur performers

Whereas professional players tended to be slightly more self-consistent than amateur performers in the playing task (see section 3.1.3), professional violinists appeared less self-consistent than amateur players in the listening task, average intra-individual consistency = 0.58 and 0.7, respectively. However, the difference fell short of significance [independent samples $t(14) = -1.2$, $p = 0.25$, equal variance]. As mentioned previously, such inferences should be treated with caution due to the small sample size in one of the two groups (amateur players, N = 5).

3.2.3. Violin scores

Following the same procedures as those described in section 3.1.4, two across-participants average scores (i.e., average rating and acceptability score) were computed for each of the violins in the listening task and compared with those obtained in the playing task. Both scores are shown in Figure 6 (leftmost versus rightmost bars for playing versus listening task, respectively). Despite the notable difference in inter-player agreement between the two conditions, ordering the violins by their across-participants across-trials average rating score resulted in two substantially similar hierarchies: violin B was characterised as most rich followed by violin C; violins D and E alternated between the third and fourth ranks; and violin A was perceived as being the least rich.

3.3. Spectral interpretation of timbral richness

To examine potential spectral correlates of violin richness, the features of spectral centroid and the three tristimulus ratios were extracted from the participant recordings of the constrained task for richness using the Timbre Toolbox [25]. The dataset obtained consists of 640 no-vibrato and 560 with-vibrato notes (the open *G* string cannot be played with vibrato by definition) for each audio descriptor (16 participants × 5 violins × 8 or 7 notes). The average length of each recorded note was 1 second.

3.3.1. Spectral centroid

The spectral centroid measures the relative center of gravity of the spectrum in a given sound,

$$\mu = \frac{\sum_{h=1}^{H} f_h \alpha_h}{\sum_{h=1}^{H} \alpha_h},$$

where f_h and α_h denote the frequency and amplitude of harmonic h, and H the total number of harmonics considered (by default $H = 20$ in the Timbre Toolbox). An initial analysis of all data (i.e., 16 μ values per note per violin) showed a large variability, which was likely a result of variations in how much bow force each violinist used [26]. To reduce the effect of outliers, we chose to compute the median across all participants (see Figure 8).

The violin perceived as most rich in the constrained-playing task (violin B, circle) had the lowest median spectral centroid value in 5 out of 8 notes, more characteristically in the open *G* string (i.e., for the note *G2*); it also had the second lowest median μ value for the note *C3*. Violin A (rhombus), which was perceived as the least rich, often had the highest spectral centroid value, including for the open string. In the listening task, violins B and A were again judged as most and least rich, respectively, though there was less differentiation between the violins in the listening task than in the playing task (see Figure 6 and section 3.2). We observed similar trends in the centroid values obtained from the with-vibrato notes. These findings seem to support the hypothesis that the desirable quality of richness in the sound of a violin, common among violinists as observed in their verbalizations [12], is correlated with increased power at lower frequencies (i.e., a lower value of spectral centroid).

3.3.2. Tristimulus ratios

The tristimulus timbre model considers three groups of harmonics in a given sound, the fundamental or first harmonic (T1), the second, third and fourth harmonics (T2),

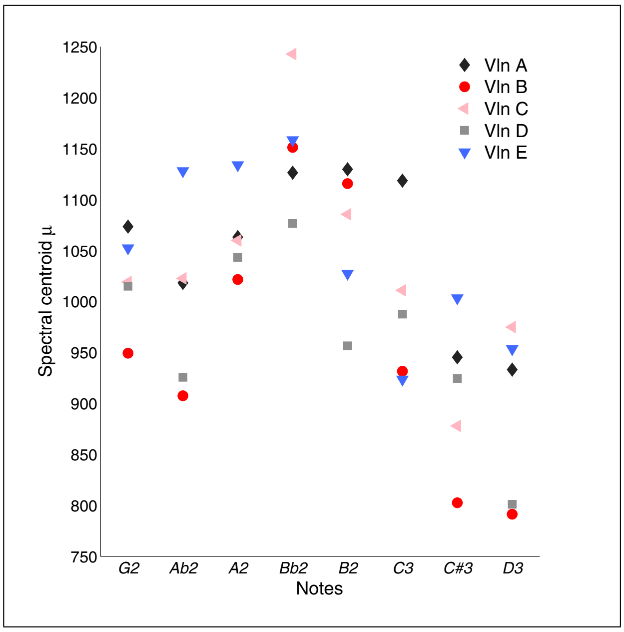

Figure 8. Across-participants median spectral centroid (μ) values for each of the eight notes of the chromatic scale $G2 \rightarrow D3$ and for each of the five violins.

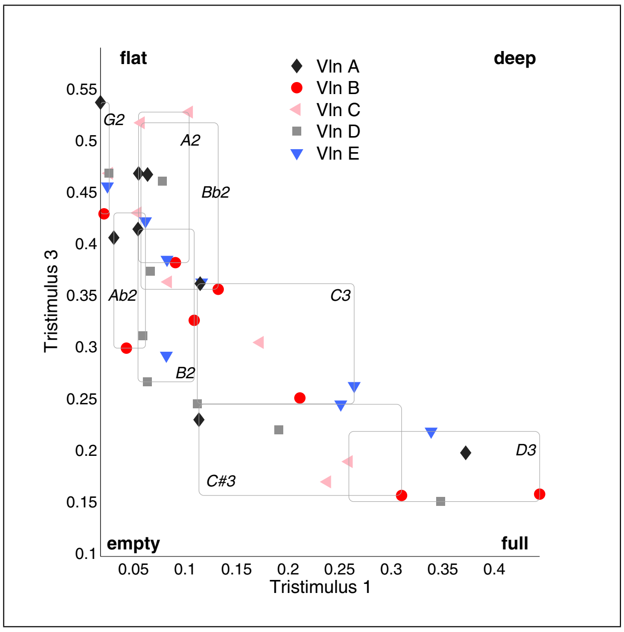

Figure 9. Across-participants median tristimulus 1 versus tristimulus 3 values for each of the eight notes of the chromatic scale $G2 \rightarrow D3$ and for each of the five violins. Gray boxes contain note-specific computed values. In bold are the semantic labels of the formed space as proposed by Łukasik [19].

and all partials above and including the fifth harmonic (T3), and measures the extent to which each group contributes to the timbre of the sound:

$$\mathrm{T1} = \frac{\alpha_1}{\sum_{h=1}^{H} \alpha_h}, \quad \mathrm{T2} = \frac{\sum_{h=2}^{4} \alpha_h}{\sum_{h=1}^{H} \alpha_h}, \quad \mathrm{T3} = \frac{\sum_{h=5}^{H} \alpha_h}{\sum_{h=1}^{H} \alpha_h},$$

where f_h, α_h and H are defined as previously. Similarly to the spectral centroid analysis, we observed a large variability in the tristimulus values for all data and chose to compute the across-participants median per note and per instrument. Figure 9 represents the semantically labelled tristimulus space proposed by Łukasik (i.e, high T1 and T3 values indicate a deep sound, see section 1). We also plotted the T2 against the T3 ratios to examine the role of the former in explaining richness (see Figure 10).

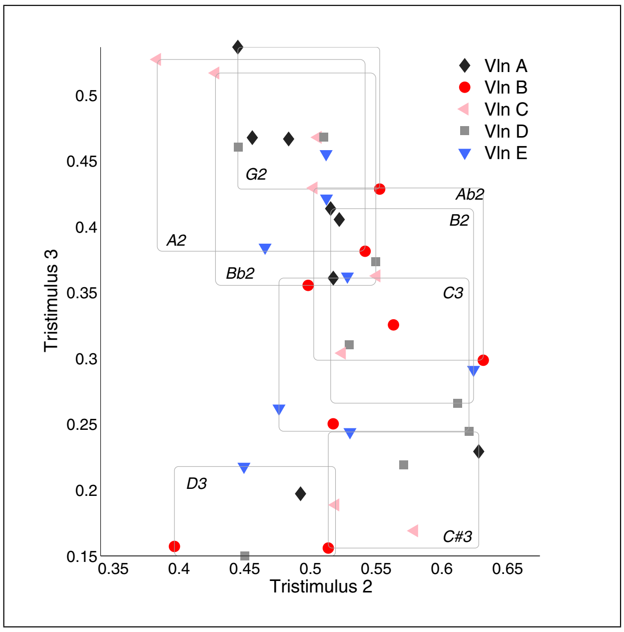

Figure 10. Across-participants median tristimulus 2 versus tristimulus 3 values for each of the eight notes of the chromatic scale $G2 \rightarrow D3$ and for each of the five violins. Gray boxes contain note-specific computed values.

Violin B (red coloured circle), which was rated as the most rich in both the playing and listening tasks, had considerably less present upper partials (i.e., a lower T3 ratio than the other violins) in all notes except for *B2*. For 3 of these notes as well as for *B2*, it also had a stronger fundamental (i.e., a higher T1 ratio than the other violins). However, the violin perceived as least rich (violin A, black coloured rhombus) did not always have the weakest fundamental or the strongest upper harmonics. The “most rich” violin (violin B) finally appeared to have more present mid-frequency partials (i.e., higher T2 ratio than the other violins) in the 3 lower notes (i.e., *G2* to *A♭2*). However, for these notes as well as for two more, it always had a stronger mid-frequency component than the “least rich” instrument (violin A). Similar trends were observed in the tristimulus values obtained from the with-vibrato notes. Based on the above observations, no definite conclusion could be reached about the correlation of the tristimulus ratios with the perception of violin richness (or fullness or depth).

4. Discussion

The results of this study showed that experienced violin players were self-consistent when evaluating different violins by focusing on a specific attribute of the instrument and following prescribed musical material and technique, both in constrained and unconstrained playing tasks. Only 2 players (12.5%) reported being *"a bit"* bothered by the dark sunglasses (QP2). Similarly to our previous studies, attempts to associate self-consistency with known (self-recorded) characteristics of the participants were largely inconclusive.

A comparison of intra-individual consistency in constrained versus unconstrained playing tasks for the assessment of richness and dynamic range revealed that violinists were significantly more self-consistent in well-focused evaluation tasks than in a less restrained setting. Several methodological differences between the two types of tasks could explain this effect. The non-randomised order of the constrained tasks (i.e., first all richness trials followed by all dynamic range trials) might have allowed participants to better stabilise their responses than in the unconstrained tasks (i.e., three tasks presented randomised in three blocks of trials). Moreover, the order of the constrained tasks was recurrent across participants, while the (random) order of the unconstrained tasks was different (i.e., randomised) for each participant. Playing a violin concerto passage that involves a broader range of notes and nuances (unconstrained tasks) likely entailed a more differentiated evaluation strategy than playing certain notes in a certain way (constrained tasks). Furthermore, as the unconstrained tasks were carried out in the second half of the session, fatigue might have affected the level of attention in evaluating richness and dynamic range as well as preference. We also observed that violinists were less self-consistent in assessing dynamic range than richness in the constrained tasks (see Figure 5). When asked "Did you have difficulty with any of the tasks?" (QP1), 5 participants (31.25%) explicitly expressed difficulties in ranking the violins in the constrained task for dynamic range. It may therefore be possible that in this case, the constrained task was not well designed.

Participants were considerably more self-consistent in the constrained-playing tasks involved in the present study than in the respective attribute-rating scales involved in a previous experiment wherein there were no playing constraints (see Experiment II in [5]). Several methodological differences between the two experimental settings could explain this effect. The rating of richness alongside other attributes (in the previous experiment) did not allow the same level of attention as focusing only on the evaluation of richness. Similarly, the level of attention is likely increased when the number of violins is reasonably small (5 in this study vs. 10 in the previous experiment). Furthermore, being able to compare the various violins to determine ratings appears to be more meaningful for the musician than assessing one violin at a time (as in the previous experiment).

More importantly, results showed a higher amount of agreement between individual performers in the playing tasks relative to the previous studies. This is further confirmed by the average ratings of the violins (see Figure 6), whereby we observe three distinct groups in practically all tasks (the difference in the constrained task for dynamic range is rather minor). On the one hand, this observation seems to support the hypothesis that different violin players may take varying approaches to assess different attributes of the instrument and hence designing focused evaluative tasks may trigger more agreement between individuals. On the other hand, it is possible that participants were able to agree more with each other simply because they had to evaluate only five violins, a relatively smaller number than in the previous studies. Despite the notable increase in inter-individual agreement, a considerable amount of variability in the perceptual judgements by skilled violinists is still observed. While specifying the musical material and technique may improve consensus, further exploration is needed to address differences in how people play as well as how verbal descriptions of sound characteristics such as richness are semantically interpreted by musicians.

Concerning the evaluation of richness, violin players appeared highly self-consistent in both the playing and listening tasks. However, the rank-ratings of the violins in the playing task were generally different from those in the listening task. Violinists who were more self-consistent in the playing task were not necessarily self-consistent in the listening task and vice versa (see Figure 7). For example, participant 2 showed no consistency of evaluations during the playing task but was considerably self-consistent in the listening test. This indicates that the evaluation of richness may be based on different criteria and/or perceptual processes in the two settings for some violin players, but perhaps not for others (since there are a number of participants that performed about the same in the two tasks). Indeed, when asked "In this new condition (listening), did your overall perception of richness change?" (QL1), 11 participants (69%) reported that their overall perception of the richness of the violins did change in varying degrees. A player commented: *"I was able to better hear the instrument from an objective point of view. When playing the instrument, the sound is so close to your ear and there are other elements to take in mind (i.e., vibration, feeling of instrument, loudness etc.) that it can become confusing to isolate richness."* However, only 3 of those players further acknowledged that their richness-related criteria for the evaluation of the violins were altered from those used in the playing task (QL2).

Furthermore, the average ratings of the violins appeared closer in the listening task than in the playing task (see Figure 6). This indicates that there was less differentiation between the violins in the listening task than in the playing task. In fact, 4 performers (25%) reported that it was harder to differentiate between the violins in the listening setting (QL1 and QL2). A possible interpretation of this result is that cues that helped players discriminate between the instruments when they are played are absent as a result of the recordings. Such cues might be related to tactile-kinesthetic feedback due to vibration sensation and finger touch. When asked how important are sound versus vibrational characteristics of the instrument, 12 participants (75%) commented that sound attributes are as essential to the overall quality of a violin as its playability. More specifically, many violinists pointed out that the perception of the produced sound is naturally dependent on the *"physical requirements to produce the sound."* As one musician explained: *"I think sound under one's ears is very difficult to judge. Projection can be limited even when it feels like there is ample sound, and likewise, an instrument may have a tone that carries, though it seems meagre under the ears. Ultimately it is variety of tone, and flexibility of tone production, as well as proprioception (feel), which count for as much as the sound one hears under the ears."*

Whereas no concrete conclusions can be drawn about the relationship between all or some of the three tristimulus ratios and perceived richness in a violin sound at this point (see Figure 9 and 10), increased power at lower frequencies (i.e, a lower value of spectral centroid) appears to indicate a rich sound (see Figure 8). Considering the importance of the concept of richness in evaluating violin quality (see [3]), spectral correlates of perceived richness in the sound of a violin need be further explored. The main challenge lies in teasing apart the effects of the playing skills and playing styles of different performers. For the string player, timbre (i.e., harmonic content of the sound) is controlled primarily through bow force, velocity and the distance of the bow-string contact point from the bridge. In the case of the spectral centroid, Schoonderwaldt showed that its magnitude is mostly determined by the applied bow force: increasing the latter results in higher values of the former [26]. Since different violinists may use different configurations of bowing parameters, a violin may exhibit a fundamentally different behaviour to each musician when assessing richness – for example, player A may use less bow force than player B and thus produce a richer sound. Another challenge lies in the potential influence of vibrato on the perception of richness.

5. Conclusions

A long-standing goal of violin acoustics has been to identify which aspects of the dynamic behaviour of the instrument affect its timbre and feel – for example, its perceived richness – thus distinguishing one violin from another. In most previous research, efforts to answer this question have traditionally been based on structural dynamics measurements and/or listening tests. Both approaches seem unsuitable for addressing the critical role of the violinist in determining the quality of an instrument. To this end, recent studies have focused attention on the perceptual processes that take place when musicians assess violins in playing tests [10, 11, 12, 27, 28]. We specifically investigated the perceptual evaluation of violin quality by experienced performers concentrating on the reliability of their psychoacoustical judgements and the verbalization of their perceptions [5, 12, 29, 30].

Expanding on our previous work, this paper reported a study aimed to investigate effects of playing constraints on the assessment of violins by musicians as well as the perception of sound characteristics from playing versus listening tasks within the context of better understanding how players evaluate violin quality. We focused on the preference for violins as well as the perceptual attributes of richness and dynamic range, which had previously been shown to be strongly associated with preference. We observed that the psychoacoustic judgements of violinists became more reliable (i.e., musicians were more self-consistent and there was better agreement between individuals) as the tasks became more controlled. We further observed that violinists were better able to discriminate between instruments in the playing than in the listening task. As such, cues that helped musicians differentiate between violins when in the former condition might have been absent in the latter as a result of the recordings. Finally, we observed that the perception of richness in violins is likely associated with the relative amount of low- and mid-frequency partials in a given sound (i.e., low spectral centroid and high tristimulus 1 and 2), though more exploration would be necessary before drawing any conclusions.

It is hoped that a better understanding of the challenges in the perceptual evaluation of violin quality has been given, even though not all aspects have been possible to take into account. The long-term goal is to better understand what distinguishes one violin from another, what criteria are considered most important to the quality of an instrument, and how these are related to its dynamic behaviour. Such knowledge can be used to refine the design of violins, inform luthiers on ways to fix problems with existing instruments, and potentially help improve sound synthesis models.

Perhaps most importantly, the results of this and our previous work demonstrate very low agreement between players in assessing violin quality, with no relationship to price or age of the instrument. Thus, these findings are important in helping players "come to terms" with a violin purchase and should diminish to some extent the societal expectations that only the old and expensive violins are of great quality. The strong correlation of violin preference with sound richness signifies that what makes a violin good might, to a certain extent, lie in the ears and hands of the musician not because different performers prefer violins with largely different qualities, but because the perceptual evaluation of violin attributes widely considered to be important for a good violin vary across individuals. This important conclusion may explain the limited success of previous studies at quantifying the differences between good and bad violins from physical measurements.

Acknowledgements

This project has been partially funded by the Centre for Interdisciplinary Research in Music Media and Technology and the Natural Sciences and Engineering Research Council of Canada. All violin shoulder rests were generously provided by Kun Shoulder Rest Inc. We thank Stephen McAdams, Jean-François Petiot and Jacques Poitevineau for constructive comments. We are grateful to musician and researcher Alfonso Pérez for his input on the design of the violin-playing tasks and to luthiers Olivier Pérot and Tom Wilder for loaning the violins.

References

[1] M. Campbell, C. Greated, A. Myers: Musical instruments: History, technology, and performance of instruments of western music. Oxford University Press, 2004.

[2] G. Bissinger: Structural acoustics of good and bad violins. J. Acoust. Soc. Am. **124** (2008) 1764–1773.

[3] C. Saitis: Evaluating violin quality: Player reliability and verbalization. Dissertation. Dept. of Music Research, McGill University, Montreal, Quebec, Canada, 2013.

[4] C. Saitis, B. L. Giordano, C. Fritz, G. P. Scavone: Investigating the origin of inter-individual differences in the preference for violins. Proc. 2011 Forum Acusticum, Aalborg, Denmark, 2011, 497–501.

[5] C. Saitis, B. L. Giordano, C. Fritz, G. P. Scavone: Perceptual evaluation of violins: A quantitative analysis of preference judgements by experienced players. J. Acoust. Soc. Am. **132** (2012) 4002–4012.

[6] A. Askenfelt, E. V. Jansson: On vibration sensation and finger touch in stringed instrument playing. Music Percept. **9** (1992) 311–350.

[7] C. Nichols: The vbow: An expressive musical controller haptic human-computer interface. Dissertation. Stanford University, Stanford CA, August 2003.

[8] C. Traube: An interdisciplinary study of the timbre of the classical guitar. Dissertation. Dept. of Music Research, McGill University, Montreal, Quebec, Canada, 2004.

[9] J. Woodhouse: On the playability of violins. part I: Reflection functions. Acustica **78** (1993) 125–136.

[10] C. Fritz, A. Muslewski, D. Dubois: A situated and cognitive approach of violin quality. Proc. 2010 Int. Symp. Music. Acoust., Sydney and Katoomba, Australia, 2010.

[11] I. Wollman: Perception bimodale des violonistes en situation de jeu: Influence des retours auditif et vibrotactile sur l'évaluation du violon. Dissertation. Université Pierre et Marie Curie, Paris, France, 2013.

[12] C. Saitis, C. Fritz, C. Guastavino, G. P. Scavone: Conceptualization of violin quality by experienced performers. Proc. Stockholm Music Acoust. Conf., Stockholm, Sweden, 2013, 123–128.

[13] G. Bissinger, F. Gearhart: A standardized qualitative violin evaluation procedure? Catgut Acoust. Soc. J. (Series II) **3** (1998) 44–45.

[14] C. Fritz, A. F. Blackwell, I. Cross, J. Woodhouse, 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** (2012) 783–794.

[15] J. M. Grey, J. W. Gordon: Perceptual effects of spectral modifications on musical timbres. J. Acoust. Soc. Am. **63** (1978) 1493–1500.

[16] E. Schubert, J. Wolfe: Does timbral brightness scale with frequency and spectral centroid? Acust. Acta Acust. **92** (2006) 820–825.

[17] H. Pollard, E. V. Jansson: A tristimulus method for the specification of musical timbre. Acustica **51** (1982) 162–171.

[18] J. Štěpánek, Z. Otčenášek: Rustle as an attribute of timbre of stationary violin tones. Catgut Acoust. Soc. J. (Series II) **3** (1999) 32–38.

[19] E. Łukasik: Towards timbre-driven semantic retrieval of violins. Proc. 5th Int. Conf. Intellig. Syst. Des. Applicat., 2005, 55–60.

[20] E. Łukasik: AMATI – Multimedia database of violin sounds. Proc. Stockholm Music Acoust. Conf., 2003, 79–82.

[21] R. B. Zajonc: Attitudinal effects of mere exposure. J. Pers. Soc. Psychol. **9** (June 1968) 1–27.

[22] R. Caussé, J. P. Maigret, C. Dichtel, J. Bensoam: Study of violin bow quality. Proc. 2001 Int. Symp. Music. Acoust., Perugia, Italy, September 10-14 2001.

[23] L. I. Lin: A concordance correlation coefficient to evaluate reproducibility. Biometrics **45** (1989) 255–268.

[24] M. M. Shoukri: Measures of interobserver agreement. Chapman & Hall/CRC, Boca Raton, FL, 2004, 152 pp.

[25] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, S. McAdams: The Timbre Toolbox: Extracting audio descriptors from musical signals. J. Acoust. Soc. Am. **130** (2011) 2902–2916.

[26] E. Schoonderwaldt: The violinist's sound palette: Spectral centroid, pitch flattening and anomalous low frequencies. Acust. Acta Acust. **95** (2009) 901–914.

[27] C. Fritz, J. Woodhouse, F. P.-H. Cheng, I. Cross, A. F. Blackwell, B. C. J. Moore: Perceptual studies of violin body damping and vibrato. J. Acoust. Soc. Am. **127** (2010) 513–524.

[28] C. Fritz, J. Curtin, J. Poitevineau, P. Morrel-Samuels, F.-C. Tao: Player preferences among new and old violins. Proc. Nat. Acad. Sci. USA **109** (2012) 760–763.

[29] C. Saitis, C. Fritz, B. L. Giordano, G. P. Scavone: Bridge admittance measurements of 10 preference-rated violins. Proc. Acoustics 2012, Nantes, France, 2012, 3599–3604.

[30] C. Saitis, C. Fritz, C. Guastavino, B. L. Giordano, G. P. Scavone: Investigating consistency in verbal descriptions of violin preference by experienced players. Proc. 12th Int. Conf. Music Percept. Cogn., Thessaloniki, Greece, 2012, 875–876.