EVALUATING VIOLIN QUALITY: A COMPARISON OF PLAYER RELIABILITY IN CONSTRAINED VS UNCONSTRAINED TASKS

Charalampos Saitis[1†] Gary P Scavone[1] Claudia Fritz[2] Bruno L Giordano[3]
[1]Centre for Interdisciplinary Research in Music Media and Technology, McGill University
[2]Lutherie-Acoustique-Musique, Université Pierre et Marie Curie, UMR CNRS 7190
[3]Institute of Neuroscience and Psychology, University of Glasgow
[†]charalampos.saitis@mail.mcgill.ca

ABSTRACT

The overall goal of the research presented here is to better understand how players evaluate different qualities of the violin. To this end, we investigated intra- and inter-individual consistency in preference judgements by experienced violinists. 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, the perception of the same attributes widely varies across individuals, thus likely resulting in large inter-individual differences in the preference for violins. A third study was conducted to further investigate the perceptual evaluation of richness and dynamic range in constrained- vs. unconstrained-playing tasks. Results indicated that specifying the musical material and technique removes a significant amount of inter-individual variability: the more focused the task, the more self-consistent violinists are and the more they agree with each other.

1. INTRODUCTION

For a period spanning more than ten years, Bissinger conducted a wide range of acoustical and structural dynamics measurements on 17 violins [1]. Those instruments were quality-rated from "bad" to "excellent" by a professional player and Bissinger himself. Attempts to quantify the characteristics of "excellent" violins were largely inconclusive, which led Bissinger to remark: "What truly defines violin excellence? If the answer is truly excellent violinists, then the reliability-reproducibility of their psychoacoustic judgements must draw more attention." The research presented here takes this "contrarian viewpoint" (in Bissinger's own wording) as a starting point and aims to quantify the extent to which skilled players are consistent at assessing violins and whether there is agreement between violinists.

We previously carried out two perceptual experiments based on a carefully controlled playing-based procedure for the perceptual evaluation of violins [2]. The first experiment was designed to examine both within-individual consistency and between-individuals agreement across a certain range of violins. In the first session of the experiment, 20 skilled violinists were asked to freely play a set of 8 different violins, evaluate their quality, and order them by preference. Upon completing the task, participants had to comment on the ranking process and provide rationale for their choices. They had to repeat the ranking task 5 times and return for a second, identical session 3–7 days after having completed the first session. In total, players ranked each violin $5 \times 2 = 10$ times. Results showed that violin players are self-consistent when evaluating different violins in terms of overall preference. However, a significant lack of agreement between violinists was observed.

A second experiment was then conducted to investigate the origin of inter-individual differences in the preference for violins and to measure the extent to which different attributes of the instrument influence preference. Thirteen experienced violin players were asked to freely play a set of 10 different instruments and rate them according to *playability* (how easy they are to play), *response, richness, balance* (across all strings), *dynamic range* and preference. The rating attributes-criteria were determined based on the analysis of verbal data collected in the first experiment as well as the potential for the descriptors to be correlated with measured vibrational properties of the violin. Participants had to rate one violin on all scales at a time. The rating task was repeated 3 times. Results showed that the perception of the same violin attributes widely varied between individual players, while confirming the large inter-individual differences in the preference for the violins observed in the first experiment. Importantly, despite the variability in the evaluation of both preference and violin attributes, violinists appeared to strongly agree on their preference for violins with a rich sound and, to a lesser extent, a broad dynamic range. As such, what makes a violin good might, to a certain extent, lie in the ears and hands of the performer 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 vibrational measurements.

From verbal responses collected in the first experiment, a classification scheme emerged that illustrates the complex links between the different player-typical concepts (e.g., *response, timbre*), properties (e.g., *ease, richness*), and underlying themes (*handling, sound* and their *relevance* to the individual) [3]. In particular, *richness* emerged as a key perceptual factor in violin quality, supporting the observations in the second experiment.

One of many hypotheses 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. In the previous two experiments, no playing constraints were imposed on the evaluation process (e.g., specific repertoire). 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 skills 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. The prescribed evaluation materials and techniques were determined based on verbal data collected in an online survey that was conducted prior to the main experiment. 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.

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). They had at least 17 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%)].

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

Table 1. Violins used in Study 3. Violin D was included in Study 1 (highest preference score) and Study 2. Its origin is based on a luthier's informal appraisal, as there is no information regarding the make and age of this violin. The names of living luthiers are not provided for confidentiality purposes.

Figure 1. Richness-constrained task

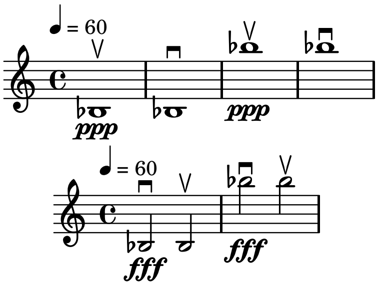

Figure 2. Dynamic range-constrained task

2.2 Violins

Five violins of different make (Europe, North America, China), age (1914–2011) and price ($2.7K–$71K) were used (see Table 2). 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. Participants were given the option to either use a provided shoulder rest (Kun Original model), or use their own, or use no shoulder rest. The experiment took place in a diffuse room with a surface of 46.8 m^2 and reverberation time of about 0.3 s to help minimize the effects of room reflections on the direct sound from the violins [4]. All other experimental conditions (i.e., visual occlusion and choice of a bow) were as in the previous studies [2].

2.3 Tasks

For each one of the perceptual characteristics of richness and dynamic range, a constrained and an unconstrained task were designed. For richness-constrained, participants were asked to play certain notes on the *G*-string (see Fig. 1). They were instructed to play *détaché*, first *without vibrato* followed by a repetition *with vibrato* using the whole bow. For dynamic range-constrained, participants were asked to play the same note first in the lower and then in the upper register (see Fig. 2). They were instructed to play *détaché, without vibrato*, as soft and as loud as possible to obtain a clear sound (i.e., the sound doesn't break).

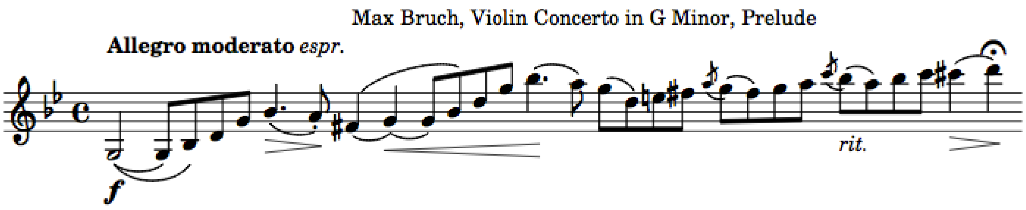

Figure 3. Unconstrained task

The unconstrained task was recurrent across the two criteria and involved playing the opening solo passage from Max Bruch's Violin Concerto No. 1 in G minor, Op. 26 (Movement I: Prelude; see Fig. 3). The particular excerpt was chosen because it incorporates the whole range of the instrument (as opposed to certain registers) as well as a variety of techniques and dynamics. The unconstrained task was also used for the evaluation of 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" is not similar to that of "free." In the latter, which was not used in the current study but only in the previous two, the participants were encouraged to choose both their own materials and techniques—and those would often change from one trial to the next, whereas the musical material was common for all players in the former.

2.4 Procedure

The experimental session lasted two hours and was organized 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 familiarize 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 according to the respective constrained task. Each task involved three repetitions (trials) and all players carried out the two tasks in the same order. 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 randomized (determined by computer calculations). In total, participants ranked-rated all violins $2 \times 3 + 3 \times 3 = 15$ times.

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 on the same unipolar discrete scale using separate, identical on-screen sliders, thus providing a ranking of the five violins at the same time (see Fig. 4). They had to move each slider (i.e., assess each instrument) before being allowed to move to the next trial. 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 maximize evaluation speed and accuracy. They were encouraged to play their own violin whenever they needed a reference point during the experiment. To minimize fatigue, participants were encouraged to take breaks between trials whenever needed.

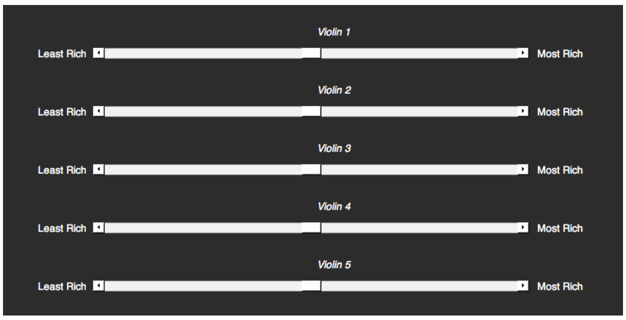

Figure 4. Testing interface

3. RESULTS

Three different analyses were carried out. Firstly, the measures of intra- and inter-individual consistency for each of the evaluation tasks were assessed and compared. Furthermore, a two-way repeated-measures analysis of variance was employed to investigate the effects of condition (i.e., constrained versus unconstrained) and attribute (richness versus dynamic range) on the measures of intra-individual consistency. The measures of intra- and inter-individual consistency recorded during this study were also compared with those recorded during Experiments 1 and 2. Secondly, the effects of participant characteristics (self-reported) on the measures of intra-individual consistency computed for each of the tasks were assessed. Thirdly, an overall score for each of the violins was derived.

3.1 Intra- and inter-player consistency

For each task, intra- and inter-individual consistency were measured and assessed based on Lin's concordance correlation coefficient ρ_c between ratings given on different blocks of trials [5]. 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 $\pm 45°$. As such, ρ_c does not assume linear relationships. 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 minimize 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 (see Fig. 5) or t-tests should be treated with caution.

For the constrained tasks, the average measure of intra-individual consistency was substantially high for richness, average value = .697, but less so for dynamic range, average value = .472. Concerning the unconstrained tasks, the average measure of intra-individual consistency was relatively high for richness and preference, average value = .443 and .442 respectively, but considerably lower for dynamic range, average value = .292. Inter-individual consistency was generally low for both constrained and unconstrained tasks, $.145 \leq$ average value $\leq .189$, except for richness-constrained, average value = .305. Considering the unconstrained tasks, no 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 $t(15) \leq 1.87,\ p \geq .081$].

3.2 Constrained vs. unconstrained evaluation

To examine the effect of constrained versus unconstrained task (condition) in the perceptual evaluation of richness and dynamic range (attribute) on self-consistency, a two-way repeated-measures analysis of variance was conducted on the corresponding measures of intra-individual consistency. 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, the analysis of variance revealed that both condition and attribute had a significant effect on how self-consistent participants were in their judgements [$F(1, 15) = 8.64, p = .01$ and $F(1, 15) = 7.72, p = .014$ respectively]. The interaction between attribute and condition fell short of significance [$F(1, 15) = .25p = 0.628$], hence the two factors do not appear to influence each other.

	Within	Between
Experiment 1		
Preference	.62(.09)	.015(.04)
Experiment 2		
Easy to play	.24(.16)	.064(.04)
Response	.328(.1)	.042(.05)
Richness	.389(.16)	.068(.06)
Balance	.203(.12)	-.005(.03)
Dynamic range	.333(.16)	.071(.03)
Preference	.38(.14)	.089(.05)
Experiment 3		
Richness (c)	.697(.17)	.305(.09)
Dynamic range (c)	.472(.16)	.154(.09)
Richness (u)	.443(.2)	.189(.08)
Dynamic range (u)	.292(.17)	.145(.06)
Preference (u)	.442(.18)	.179(.08)

Table 2. Across-participants average intra-individual consistency and inter-individual agreement measures in each of the experiments (c = constrained, u = unconstrained).

3.3 Comparisons with previous studies

The overall measures of intra- and inter-individual consistency for richness and dynamic range were compared with those measured in the respective attribute-rating scales used in the second study, wherein players were instructed to develop their own strategy. Both intra-individual consistency and inter-individual agreement for the evaluation of richness-constrained were notably higher than in the second experiment, average value = .697 and .305, and .389 and .068 respectively. Indeed, the large increase in both intra- and inter-individual consistency was found to be significant [independent samples $t(27) = 2.81, p = .009$, and $t(27) = 4.59, p < .001$, equal variance]. In the case of dynamic range-constrained, intra- and inter-individual consistency were also higher, albeit to a lesser extent, than in the second study, average value = .472 and .154, and .333 and .071 respectively. Although the relative increase in intra-individual consistency fell short of significance [independent samples $t(27) = 1.32,\ p = .199$, equal variance], the increase in inter-individual agreement was significant [independent samples $t(19.78) = 3.36,\ p = .003$, unequal variance]. Intra- and inter-individual consistency in richness-unconstrained were moderately higher than in the second experiment, average value = .443 and .189, and .389 and .068 respectively. The increase in self-consistency was not significant [independent samples $t(27) = .44,\ p = .665$, equal variance], but the increase in inter-individual agreement was [independent samples $t(27) = 2.33,\ p = .028$, equal variance]. In dynamic range-unconstrained, intra-individual consistency was slightly and not significantly lower than in the second study, average value = .292 and

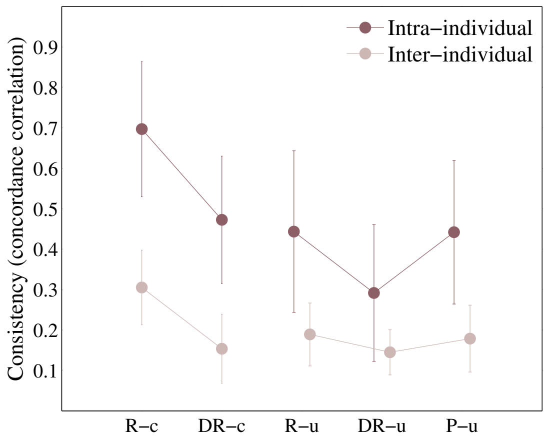

Figure 5. Across-participants average intra- and inter-individual consistency scores for each of the constrained and unconstrained playing tasks (R = richness, DR = dynamic range, P = preference; c = constrained, u = unconstrained; error bar = 95% confidence interval of the mean). See text for details on averaging of concordance correlations.

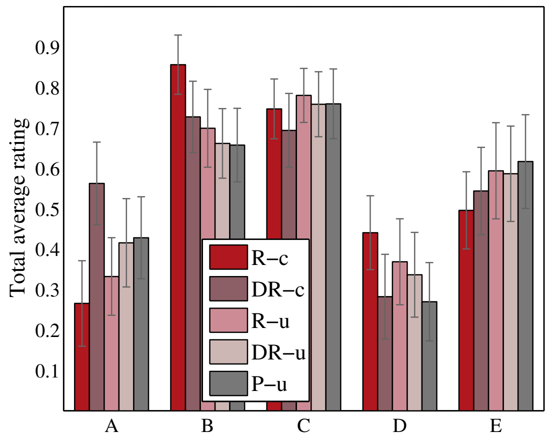

Figure 6. Across-participants across-trials average rating scores for each violin (R = richness, DR = dynamic range, P = preference; c = constrained, u = unconstrained; error bar = 95% confidence interval of the mean).

.333 respectively [independent samples $t(27) = -.38$, $p = .709$, equal variance]; inter-individual agreement was significantly higher than in the second experiment, average value = .145 and .071 respectively [independent samples $t(21.32) = 3.56, p = .002$, unequal variance].

The overall measures of intra- and inter-individual consistency collected during the first experiment (i.e., preference judgements) and those measured during the second study for the preference-rating scale were compared with those measured during the current study for the preference task. Intra-individual consistency for the evaluation of preference was higher in the first than in the third experiment, average value = .62 and .442 respectively, but the decrease was not significant [independent samples $t(23.5) = -1.88$, $p = .072$, unequal variance]. In the second study, intra-individual consistency in preference judgements was lower than in the current one, average value = .38 and .442 respectively, but the increase fell short of significance [independent samples $t(27) = .56$, $p = .577$, equal variance]. Inter-individual consistency in preference judgements gradually increased from the first to the second to the third study, average value = .015, .089 and .179 respectively. Despite the increase from the second to the third experiment not being significant [independent samples $t(23.52) = 1.91$, $p = .068$, unequal variance], the overall increase from the first to the third study was found to be significant [independent samples $t(20.38) = 3.79, p = .001$, unequal variance]

3.4 Violin ratings

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 Fig. 6, where we observe 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 dynamic range-constrained, the grouping was only slightly different as violin E alternated with A). In particular, violin A was consistently chosen as the least rich instrument and violin D was consistently thought to have the narrowest dynamic range. Violin B was characterized 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. The difference in the playing range between constrained and unconstrained tasks should be considered here. It may be possible that violin B has a particularly rich low end compared to violin C but the latter “wins” in the middle and high register. Similarly, violin B may allow better control of dynamics for the low and high $B\flat$ than violin C but the latter’s dynamic range may be better highlighted through playing the Bruch passage.

4. DISCUSSION

The results of this experiment showed that experienced violin players are self-consistent when evaluating different violins by focusing on a specific attribute of the instrument and following prescribed musical material and technique. An analysis of variance revealed that violinists are 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-randomized order of the constrained tasks (i.e., first all richness trials followed by all dynamic range trials) gave participants a better opportunity to stabilize their responses than in the unconstrained tasks (where the order of the three tasks was randomized in each of the 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., randomized) for each participant. Playing a violin concerto passage that involves a wider range of notes and nuances (unconstrained tasks) 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 affected the level of attention in evaluating richness and dynamic range as well as preference.

Participants were considerably more self-consistent in the constrained-playing tasks involved in this experiment than in the respective attribute-rating scales involved in the second study whereby there were no playing constraints. Several methodological differences between the two experimental settings could explain this effect. The rating of richness alongside other attributes (in the second study) did not allow the same level of attention as focusing only on richness. Similarly, the level of attention is increased when the number of violins is reasonably small. Furthermore, being able to compare the various violins to determine ratings is ecologically more valid than rating one violin at a time (as in the second experiment).

Participants were less self-consistent when evaluating preference in this study than in the first experiment. This could be explained by the higher number of repetitions in the first study (10 ranks for each violin across the 2 sessions) than in the current experimental setup (3 ranks for each violin) as well as the presense of two attribute tasks alongside preference. On the other hand, participants appeared slightly more self-consistent in this study than in the second experiment. To a certain extent, these observations seem to suggest that when evaluating a set of violins, comparing all instruments at a time is more meaningful and thus more reliable than assessing each violin individually. Furthermore, no significant differences were observed in this study between the level of intra-individual consistency in the preference ratings and that in the attribute ratings (unconstrained tasks).

More importantly, a higher inter-individual agreement in the playing tasks relative to the previous studies was present. This is further confirmed by the average ratings of the violins, whereby we observe three distinct groups in all tasks but for dynamic range-constrained (though the difference in the respective ordering is relatively 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 because they had to evaluate only five violins, a relatively smaller number than in the previous studies.

5. CONCLUSIONS

We reported a study aimed to investigate the effect of playing constraints on the assessment of violins by experienced musicians 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 as the tasks became more controlled.

While specifying the musical material and technique may improve consensus, there remains the issue of addressing differences in how people play. Different violinists may use different combinations of gestures when playing, each producing a fundamentally different behaviour of the instrument for a certain attribute. For example, player A may use more bow force than player B and thus produce a more *bright* timbre [6]. Further exploration is needed in this direction. That still would not address differences in the semantic interpretation of such verbal tags. To this end, we are currently studying player verbalizations using a linguistic approach to identify the meaning(s) of *richness* [7].

Acknowledgments

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 luthiers Olivier Pérot and Tom Wilder for loaning the violins.

6. REFERENCES

[1] G. Bissinger, "Structural acoustics of good and bad violins," *J. Acoust. Soc. Am.*, vol. 124, no. 3, pp. 1764–1773, 2008.

[2] C. Saitis, B. L. Giordano, C. Fritz, and G. P. Scavone, "Perceptual evaluation of violins: A quantitative analysis of preference judgements by experienced players," *J. Acoust. Soc. Am.*, vol. 123, no. 6, pp. 4002–4012, 2012.

[3] C. Saitis, C. Fritz, C. Guastavino, and G. P. Scavone, "Conceptualizing the "good" violin in preference descriptions by experienced players," in *Proc. Stockholm Music Acoust. Conf.*, Stockholm, Sweden, 2013.

[4] G. Bissinger and F. Gearhart, "A standardized qualitative violin evaluation procedure?" *Catgut Acoust. Soc. J. (Series II)*, vol. 3, no. 6, pp. 44–45, 1998.

[5] L. I. Lin, "A concordance correlation coefficient to evaluate reproducibility." *Biometrics*, vol. 45, no. 1, pp. 255–268, 1989.

[6] E. Schoonderwaldt, K. Guettler, and A. Askenfelt, "An empirical investigation of bow-force limits in the Schelleng diagram," *Acust. Acta Acust.*, vol. 94, pp. 604–622, 2008.

[7] P. Cheminée, "Est-ce bien *clair* ? Stabilité, instabilité et polysémie d'une forme lexicale en contexte," in *Le sentir et le dire: Concepts et méthodes en psychologie et linguistique cognitives*, D. Dubois, Ed. Paris: L'Harmattan, 2009.