Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Musical Acoustics

Session 2pMU: Musical Preference, Perception, and Processing

2pMU2. Perceptual evaluation of violins: A comparison of intra-individual agreement in playing vs. listening tasks for the case of richness

Charalampos Saitis*, Gary P. Scavone, Claudia Fritz and Bruno L. Giordano

***Corresponding author's address: Computational Acoustic Modeling Laboratory, Centre for Interdisciplinary Research in Music Media and Technology, Schulich School of Music, McGill University, Montreal, H3A 1E3, Quebec, Canada, charalampos.saitis@mail.mcgill.ca**

In previous studies by the authors, it was shown that there is a significant lack of agreement between violinists when evaluating different instruments in terms of perceived richness in free-playing tasks. A new experiment was designed to further investigate the perceptual evaluation of richness using both a 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 vs. listening tasks in order to better understand whether they are based more on auditory feedback or tactile and proprioceptive cues in the wider context of correlating audio features extracted from the recordings with richness judgements. Skilled violinists were asked to rank five different instruments by playing only certain notes on the G-string. Subsequently, the players were asked to listen to their recordings and rank the violins. Results appeared to show a higher inter-individual agreement relative to previous studies. Furthermore, the rankings in the playing task were generally different from those in the listening task, indicating that the evaluation of richness is based on different criteria in the two cases.

Published by the Acoustical Society of America through the American Institute of Physics

©2013 Acoustical Society of America [DOI: 10.1121/1.4800060]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

In previous studies, we have shown that experienced violinists are self-consistent when evaluating different instruments in terms of overall preference and certain perceived qualities in free-playing tasks, though there was a significant lack of agreement between individuals [1]. One of many hypotheses about the origin of the large inter-individual differences in violin preference is that players may take varying approaches to assess different attributes of the instrument. To this end, a new experiment was designed to investigate the perceptual evaluation of richness using a constrained-playing task (i.e. the musical material and technique that players were allowed to use for their evaluations were prescribed), which was recorded, and a subsequent listening task (using the previously recorded sounds). We chose to focus on the perceptual characteristic "richness" as we had previously found it to be highly correlated with violin preference [1].

METHOD

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%)].

Violins Five violins of different make (Europe, North America, China), age (1914–2011) and price ($2.7K–$71K) were used (see Table 1). They were chosen from several 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 ~ 0.3 s to help minimize the effects of room reflections on the direct sound from the violins [2]. All other experimental conditions (i.e., visual occlusion and choice of a bow) were as in previous studies by the authors [1].

TABLE 1: Violins used in the study; also reported are the across-participants across-trials average ratings of richness in the two experimental conditions (95% confidence interval of the mean in parentheses). The most preferred violin (B) is indicated in bold and the least preferred violin (A) in italics. Names of living 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 instrument.

Violin	***A***	**B**	**C**	**D**	**E**
Origin	Italy	Switzerland	Denmark	Germany	China
Luthier	Contino	-	Hjorth	-	-
Year	1916	2003	1914	-	2011
Price	$71K	$30K	$20K	$10K	$2.7K
Playing	*.27(.11)*	**.86(.07)**	.75(.07)	.44(.09)	.50(.1)
Listening	*.45(.11)*	**.72(.09)**	.54(.1)	.49(.1)	.47(.12)

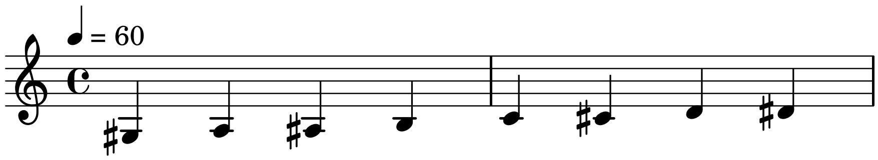

FIGURE 1: Musical material

Procedure The playing session initially involved a short training phase with three violins, which were distinct from the five violins used in the main phase, to help participants familiarize themselves with the constrained-playing task. The main phase consisted of three trials. On each trial, participants were presented with all violins placed on a table in random order (determined by computer calculations) and asked to rank and rate the violins on a unipolar discrete scale using on-screen sliders by playing only certain notes on the G-string (see Fig. 1). They were instructed to play these notes *détaché*, first *without vibrato* and then *with vibrato*. Participants were instructed to always rank-rate their first violin as 1 and their last as 0, and were not allowed to assign the same ranking-rating to two or more instruments. Participants were instructed to maximize evaluation speed and accuracy, and were encouraged to play their own violin whenever they needed a reference point during the experiment. Upon completing the last trial, participants recorded the musical material on each of the five violins. The listening session took place 1-4 days later. Participants were presented with their own audio recordings in random order and were asked to rank and rate the violins as previously (i.e., the interface and instructions were identical to the first session).

RESULTS

For each rating scale, intra- and inter-individual consistency were measured and assessed based on the concordance correlation ρ_c between ratings given on different blocks of trials [3]. The concordance correlation 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.[†] Figure 2 reports the intra- and inter-individual consistency measures for each experimental task averaged across participants.

For both tasks, the average measure of intra-individual consistency was substantially high, average value = .697 and .619 for the playing and listening tasks respectively. We observed no significant difference in the average intra-individual consistency between the two tasks [paired samples $t(15) = 1.29$, $p = .216$]. Further, whereas the average intra-individual consistency was considerably higher than the average inter-individual consistency for the listening task, the same effect was of smaller size for the playing task.

We compared the overall measures of intra- and inter-individual consistency in this experiment with those measured during a previous richness-rating study that involved a free-playing task (see Exp. 2 in [1]). Both intra- and inter-individual consistency for the evaluation of rich-

[†]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. 2) should be treated with caution.

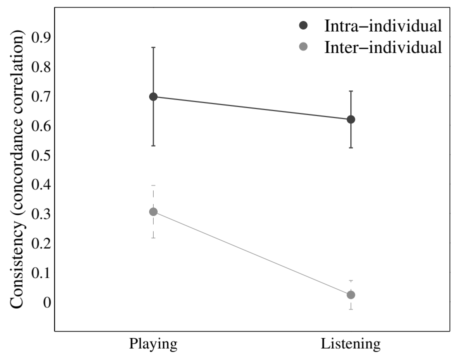

FIGURE 2: Across-participants average intra- and inter-individual consistency scores for each of the richness-rating tasks (error bar = 95% confidence interval of the mean). The confidence intervals for the average inter-individual consistency scores should be treated with caution as the independence criterion is not strictly met.

ness were notably higher in the constrained-playing task than in the free-playing task, average value = .697 and .305, and .236 and .065 respectively. Notably, the increase in intra-individual consistency fell short of significance [independent samples $t(15.18) = -1.73$, $p = .104$, unequal variance].

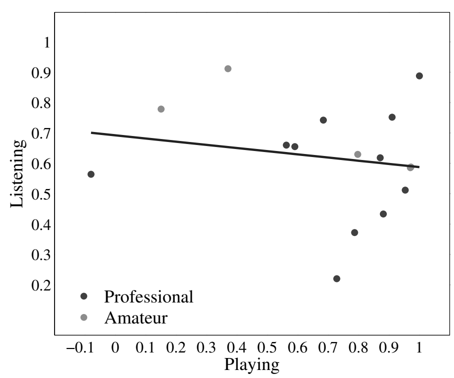

FIGURE 3: Scatter plot of the measures of intra-individual consistency in the playing vs. listening task of each participant.

A two-sample t-test was adopted to assess whether intra-individual consistency significantly differed between professional and amateur violin players (N = 11 and 5, respectively). In the playing task, despite a tendency for professional violin players to be slightly more self-consistent than amateur players, average intra-individual consistency = .717 and .652 respectively, the difference proved to be not significant [independent samples $t(11.31) = 0.87$, $p = .404$, unequal variance]. Professional musicians appeared considerably less self-consistent than amateur players in the listening task, average intra-individual consistency = .583 and .699 respectively, although the difference fell short of significance [independent samples $t(6.96) = -1.1$, $p = .307$, unequal 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 size.

Finally, for each of the violins, we computed a richness score defined as the across-trials and across-participants average rating for each of the playing and listening task. The scores are reported in the lower part of Table 1 and shown in Fig. 4.

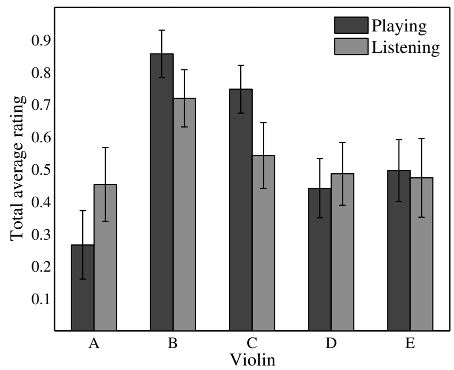

FIGURE 4: Across-participants across-trials average of richness ratings for each violin in each experimental condition (error bar = 95% confidence interval of the mean). The violins are ordered by decreasing price. Violins B and A were evaluated as the most and least rich respectively.

DISCUSSION

The results of this experiment showed that experienced violin players are highly self-consistent when evaluating different violins by focusing on a specific attribute and following prescribed musical material and technique, both in a playing task and a listening task. However, the rankings-ratings of the violins in the playing task were generally different from those in the listening task and we observed no significant effect of the type of task on intra-individual consistency (see also Fig. 3). Players who were more self-consistent in the playing task were not necessarily self-consistent in the listening task and vice versa. This indicates that the evaluation of the richness of a violin is based on different criteria and/or perceptual processes in the two cases.

Perhaps more interestingly, participants were considerably more self-consistent in the constrained-playing task involved in this experiment than in a previous free-playing task (see Exp. 2 in [1]). Several methodological differences between the two experimental settings could explain this effect. The rating of richness alongside other attributes (in the previous 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 previous study).

More importantly, results showed a higher inter-individual agreement in the playing task relative to previous studies. This is further confirmed by the violins' average ratings (see Fig. 4), whereby we observe three distinct groups in both tasks: violins B and C were evaluated as the most and second-most rich respectively; violins D and E alternated between the third and fourth ranks; and violin A was evaluated as the least rich. 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 previous studies. Furthermore, Fig. 4 shows that the violins's average ratings are closer in the listening task than in the playing task. This indicates that there was less differentiation between the violins in the listening task than in the playing task. 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.

Despite the notable increase in inter-individual agreement, a considerable amount of variability in the richness judgments by skilled violinists is still observed. While specifying the musical material and technique may remove a significant portion of inter-individual variability, there remains the issue of addressing differences in the way people play. Different violinists may use different combinations of gestures when playing, each producing a fundamentally different behavior 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 [4]. Further exploration is necessary to tease apart the effects of the playing skills of different individuals.

ACKNOWLEDGMENTS

This project has been partially funded by the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) and the Natural Sciences and Engineering Research Council of Canada (NSERC). All violin shoulder rests were generously provided by Kun Shoulder Rest Inc. (Ottawa, Ontario, Canada). We thank Stephen McAdams, Jean-François Petiot and Jacques Poitevineau for fruitful discussions. We are grateful to luthiers Olivier Pérot and Wilder & Davis for loaning the violins used in the experiment.

REFERENCES

[1] 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. **123**, 4002–4012 (2012).

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

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

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