TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Elektrotechnik und Informationstechnik

Institut für Akustik und Sprachkommunikation, Lehrstuhl für Akustik und Haptik

Perceptual evaluation of audible differences between guitars – Q study

Project partner	Pacific Rim Tonewoods (PRT) 38511 WA-20, Concrete, WA 98237, USA
Contact person	David Olson
Project duration	01.06.2019 – 30.10.2020
Date of report	30.10.2020
Pages of report	21 pages
Contractor	Technische Universität Dresden Fakultät Elektrotechnik und Informationstechnik Lehrstuhl für Akustik und Haptik Helmholtzstraße 18 01069 Dresden
Contact person	Sebastian Merchel

1 Introduction

The aim of this project is to investigate the influence of varying internal damping of tonewood on the perceived acoustic quality of steel string guitars. The internal damping of wood can be measured in terms of the quality factor (Q) and might be useful to evaluate the wood's suitability for a given musical instrument design. Within this project, listening tests were performed to investigate whether the objective testing of this physical parameter can help to make an appreciable and reproducible impact on the sonic quality of the resulting instrument.

2 Selection of wood and fabrication of guitars

In cooperation with PRT and Taylor Co., a unique set of guitars was built for this project: six 814ce V-Class, with the only practical difference being the acoustic characteristics of the soundboard. The guitars were built sequentially by the same team of luthiers, with a factory 'set up' by Taylor's lead technician. The backs/sides, necks, bridges, pin blocks, and fret boards were built from the same density class of seasoned and premeasured wood (EIR, mahogany, ebony and grenadillo). A first group ('Aa') of three guitars was built with low-stiffness and low-density wood for the soundboards and bracewood that was found to be favorable for the 814ce X-Class model in the last study. The tops of the three guitars in the second group ('Cc') were built from high-stiffness and high-density tonewood. The material parameters were measured using the BING software system by Pico Technology, which analyzes standardized rectangular billets of test wood. This system uses the natural resonance frequencies of the test pieces to derive the elastic constants and damping. The pieces were suspended on elastic supports at their natural nodes, at 22:4% of their overall length, and the standard impact of dropping a steel ball (9 grams and 13 mm in diameter) on the pieces was used to induce a transverse wave on one end of the board; a microphone was located on the other end. The dynamic nature of these tests allows for measurements of the logarithmic decrement at the resultant resonance frequencies and a derivation of the internal damping. Figure 1 shows the longitudinal Young's modulus and density for the two groups. Within each group, the quality factor Q was varied across the typical range. Three levels of damping were chosen with Q being approximately 130 (labeled 'Q low' or 'Ql'), approximately 150 (labeled 'Q mid' or 'Qm') and approximately 160 (labeled 'Q high' or 'Qh'). The selected pieces of wood used for building soundboards and bracewood and their material properties are summarized in Table 1.

Table 1 – Measured wood material parameters of the selected boards for the guitar tops.

Label	Q	Density ρ in kg/m³	Longitudinal Young's modulus E_{LONG} in MPa	Shear modulus G in MPa	Ratio of longitudinal Young's modulus and density
AaQl	134	366	10730	701	29.3
AaQm	151	373	11987	620	32.2
AaQh	163	371	11952	638	32.2
CcQl	130	462	16195	932	35.0
CcQm	151	451	15097	847	33.5
CcQh	159	461	15739	687	34.1

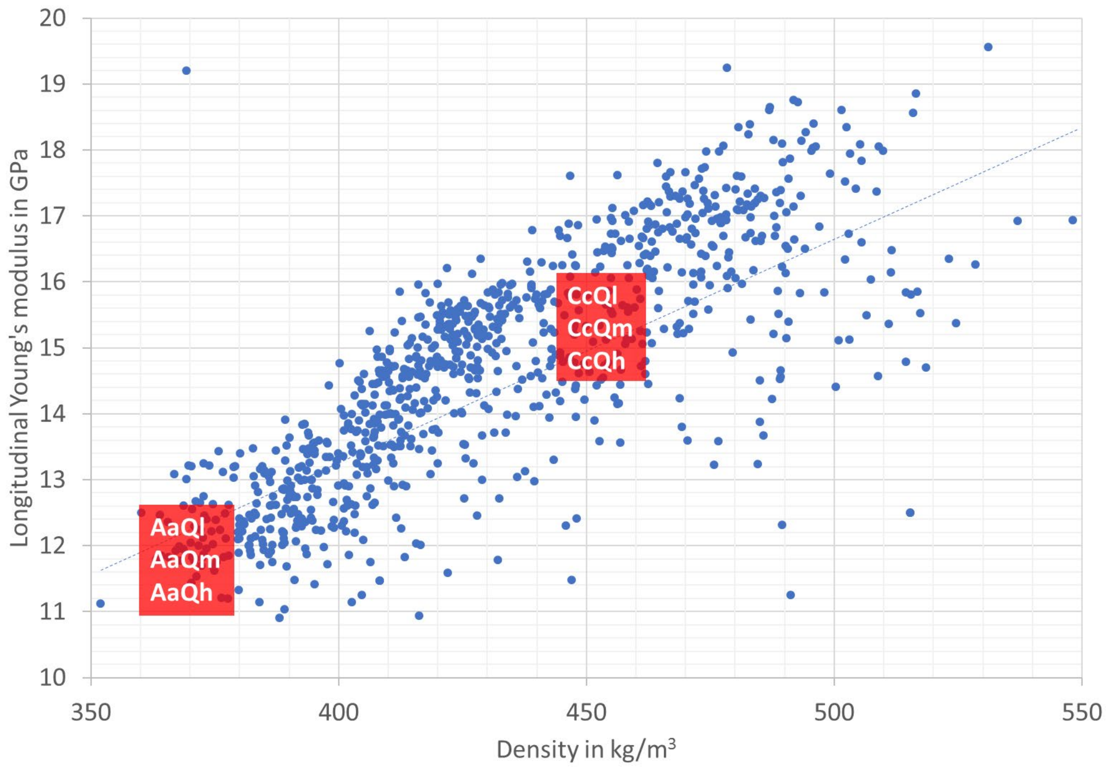

Figure 1 – Longitudinal Young's modulus and density of the tonewood. Marked are the selected parameters that were used for building the tops of groups Aa and Cc.

3 Selection of stimuli

In the last project, no influence of the chosen strumming or melodic sequences on the preference ratings was found. Therefore, only one prototypical strumming sequence, which already proved to be useful, was selected for this project. It consisted of a simple stroking pattern with the following five chords: C, $D^{add9/11}$, e^{add9}, $D^{add9/11}$, and C. The tempo was set to a moderate 70 bpm. The sheet music is shown in Figure 2.

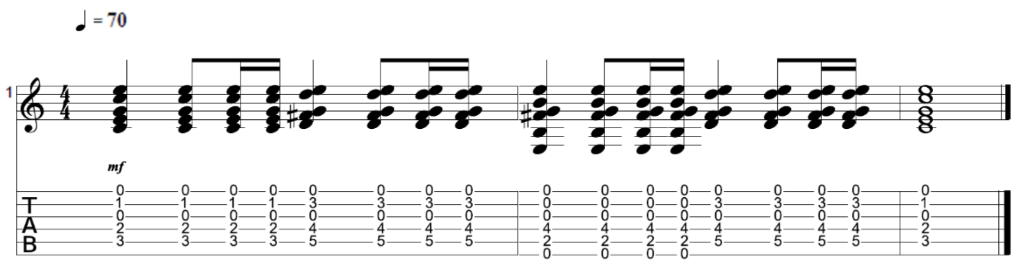

Figure 2 - Sheet music with tabs for the strumming sequence of C, $D^{add9/11}$, e^{add9}, $D^{add9/11}$, and C.

4 Recording of stimuli

In accordance with the last project, the decision was made to record the sound with a single omnidirectional microphone placed 2 m in front of the sound hole of the guitar. This recording position is chosen to be sufficiently far from the guitar to homogeneously integrate sound radiated from different instrument parts. The aim was to have the recording method influence the sound of the guitar as minimally as possible, e.g., summing multiple microphones would always have an artistic component involved. Additionally, the influence of the room was excluded from the recording by using an anechoic chamber. This anechoic recording will be later reproduced using a mono loudspeaker in a normal room. As a result, the reverberation of a real room was included in the reproduction scenario. Using this method, the directivity of the guitar was neglected and replaced by the directional characteristic of a loudspeaker. However, it is assumed that the modifications of the wood in this study mainly influence spectral and temporal cues. Therefore, the recording and reproduction method described above was chosen so that pairwise comparisons can be conducted in a listening test.

A neutral high-quality microphone, the G.R.A.S. 40HL with a flat frequency response (±3 dB from 6 Hz to 20 kHz) and wide dynamic range (6.5 dB(A) to 110 dB), was selected. A HEAD acoustics SQuadriga II front end was applied along with HEAD Artemis software for recording. The microphone was calibrated to correctly reproduce the sound pressure level in the listening test.

The selected music sequence was recorded with all 6 guitars. Additionally, two guitars from the previous study (Aa-114 and Cc-117) were recorded again. To achieve equal conditions, those guitars were equipped with new strings: Elixir Light Gauge Phosphor Bronze NANOWEB (.012-.053). All guitars were evaluated locally by a luthier to ensure that they were in perfect condition. The instruments, including all guitars of the previous study, had been strung and broken in equally beforehand.

It was decided that recordings would be made with a professional guitarist from the Academy of Music ‘Carl Maria von Weber’ in Dresden—Judith Beckedorf. It is self-evident that playing techniques or the individual style significantly influences the sound of a guitar. However, it is assumed that some characteristics of an instrument can be heard independent of the excitation. The guitarist was instructed to perform the sequence naturally, but to maintain a constant playing style. To be able to switch seamlessly between instruments in the pairwise comparison, it was necessary that the tempo between recordings be synchronous. Therefore, a metronome was used to generate a click track, which was reproduced via in-ear headphones in one ear of the guitarist. The level was reduced to a minimum so that the clicks were not audible in the recording. The absolute position of the guitars relative to the microphone was kept constant to avoid changes in the overall recording level. Therefore, the guitarist remained seated throughout the recording session and a second person tuned and exchanged the instruments.

As in the previous study, it was hypothesized that the natural variation in sound between repeated recordings of the same instrument should be perceptually distinguishable from the variation between different instruments. To test this hypothesis, all sequences were recorded twice on each instrument.

Photos of the recording configuration are shown in Figure 3 and Figure 4.

Despite the careful effort to play the instruments homogeneously, there were deviations in sound pressure level for individual strikes within a single recording and between repeated recordings of the same instrument. As in the previous study, the variation was very small, and a decision was made not to adjust the overall level of the recordings.

Figure 3 – Final tuning of the guitars in the anechoic chamber.

Figure 4 – Recording configuration in the anechoic chamber. The main microphone is positioned 2 m in front of the guitar.

5 Listening test

5.1 Setup

To investigate perceptual differences between the guitars, the recordings were reproduced in a quiet laboratory room using a single loudspeaker. A Genelec 8250A speaker was placed 2 m in front of the listener seat. The studio monitor had a flat, free-field frequency response (±1 dB from 38 Hz to 20 kHz). The volume was compensated so that the speaker reproduced the same sound pressure level at a position 2 m in front of the speaker as the original instruments at the main microphone position 2 m in front of the guitar in free-field conditions.

5.2 Subjects

Thirty-five subjects (twenty-six male and nine female) participated in the study. Twenty-two of the participants were professional guitarists either with a completed degree or studying guitar or musicology. The remaining thirteen subjects were passionate amateur guitarists with several years of experience playing the guitar in bands. The mean age of the participants was 36.9 years.

5.3 Experimental design

The goal of the listening test was to identify possible preferences between the guitars. A decision was made to only ask for the overall preference but not for specific quality attributes of the guitars. This strategy was employed as an attempt to avoid influencing the test participants by predefined quality categories, as is usually the case with methods applying a semantic differential. In previous experiments, the authors applied a method adapted from [1] (MUSHRA) to compare musical instruments. However, for small differences between stimuli, the International Telecommunication Union (ITU) recommends applying a paired comparison method [2]. The recommended method was originally designed to evaluate small impairments in audio systems. It was adapted using the open software framework webMUSHRA [3] for preference assessments, as follows. This method should allow a detailed comparison of the subtle, expected differences between the instruments. Each subject was presented with all possible pairs of the $n = 6$ guitars in an individually randomized order. Then, two additional pairs were evaluated to compare the winner of the last study with X-bracing (Aa-114) against two new guitars with V-bracing (AaQl and AaQh). The same was repeated with another X-braced guitar from the last study (Cc-117), which had to be compared against two guitars (CcQl and CcQh) from the current study. This analysis resulted in a series of $n(n-1)/2 + 2 + 2 = 19$ comparisons. In addition to the randomization between comparisons, the order of presentation within each pair was randomized to prevent stereotypical responses.

Before the test, the following background information was provided to the participants: "All of the guitars in this study are the same model (Taylor Grand Auditorium); they were built at the same time, by the same build team, but they differ from one another in one very important respect. The soundboards of these guitars are built from varying spruce wood that has been carefully selected according to its acoustic properties."

Then, the task was introduced: "For each pairing, you will be asked which guitar tone you prefer. We ask that you listen carefully, repeat the recordings as needed, and indicate your preference on each pairing, even if the difference is subtle. You have been invited to participate in this study because of your particular expertise with the acoustic guitar—please bring every bit of that experience to this activity!"

All listening test sessions began with a small training phase. The purpose of the training was to allow the participant to identify and become familiar with subtle differences produced by the guitars that were being tested. During the training phase, the test subjects also became familiar with the test procedure. In each trial, the test subjects heard three recordings of the same sequence. These were labeled A, B and C on the video monitor screen. A was always the *reference* recording, against which both B and C were to be compared. However, either B or C was a repeated recording of guitar A—a so-called *hidden reference*. The other sample was always a different guitar than that played in A. The first task of the participants was to identify the hidden reference. Then, the participants indicated their preference for the remaining pair using a grading scale. It was emphasized during the test instructions that the grading scale had to be considered as a continuous equal interval scale with anchor points (no preference, slightly prefer and prefer) defined at specific values. An example is shown in Figure 5. The ratings were interpreted as numbers, e.g., 'prefer B' corresponding to 100, 'prefer A' corresponding to -100 and 'no preference' corresponding to 0. This number was displayed in a small box below the slider to underline the interval character of the scale.

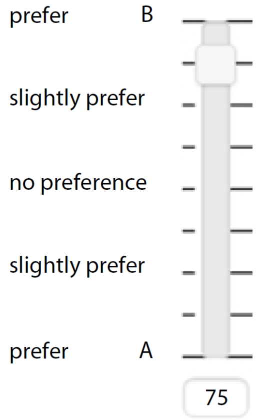

Figure 5 - Preference grading scale for stimuli A and B with verbal anchor points.

The participants were able to switch freely among A, B or C at any time, even during playback. It was also possible and encouraged to adjust the start and end markers of the loop as preferred using the graphical interface shown in Figure 6 to focus on specific aspects of the sound.

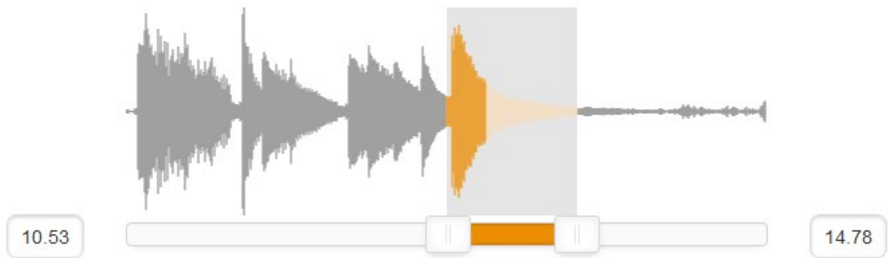

Figure 6 - Graphical representation of the wave form with adjustable start and end markers of the loop. The same wave form image was used for all stimuli.

All audio sequences were played repeatedly until the participants were confident about their evaluations on a given trial. Completing all 19 paired comparisons required approximately 45 minutes.

5.4 Results and discussion – Q

The experimental design assumes that subjects can produce scores on an interval scale level (How much is guitar A preferred over guitar B?). Therefore, the data are analyzed using parametric statistical testing.

Each participant was asked to identify the hidden reference for each paired comparison. This method tests whether the subjects were able to detect the repeated recording of the reference guitar. The mean wrong detection rate of the hidden reference was slightly higher than the last study, at 3.26 of 19 comparisons. A binomial test confirmed with $p < 0.000$ that the mean wrong detection rate is different from detection by chance. *This finding suggests that the participants were able to predominantly distinguish the differences between different guitars against differences between repeated recordings.*

If, in rare cases, the wrong guitar was identified as the hidden reference, the reference guitar was compared with a repeated recording of itself in the preference rating. It is proposed in [4] to exclude subjects who rate the erroneously detected hidden reference in more than 15% of the trials with more than 10% deviation from the expected score. However, the rejection of data is susceptible to shaping the data according to the experimenter's preconception, rather than accepting the empirical evidence of actual outcomes. Therefore, [2] recommends not applying post hoc criteria for the acceptability of data, except for obviously highly deviant single cases. Thus, all data were retained for further statistical analysis. Table 2 lists the absolute preference scores for all subjects and all pairs of the six guitars.

For the following parametric statistical analysis, the software package Statistical Package for the Social Sciences (SPSS) by IBM was used. The presence of a normal distribution was tested with the Shapiro-Wilk test because of the small sample size (35 test participants). Only eight of 19 pairwise ratings showed significance above 0.05, meaning that normal distributions cannot be generally assumed. Therefore, the mean values reported in the following should be carefully interpreted and a visual inspection of the underlying histograms is useful. Since simulation studies have shown that ANOVA with repeated measures is relatively robust to violations of the normal distribution assumption (especially when no other assumption is violated [5]), this parametric evaluation of the data can nevertheless be performed.

To determine whether the ratings of all pairwise scores were in accordance among subjects, the intraclass correlation was tested. As a reliability coefficient, Cronbach's alpha was calculated as 0.73, suggesting sufficient internal reliability. A reliability coefficient of .70 or higher is considered to be acceptable in most social science research situations and comparisons of grouped rating data is meaningful [6]. Mean values were calculated of the individual paired-comparison scores and are shown at the bottom of Table 2.

Table 2 - Absolute preference for all pairs of guitars in the listening test. Positive values mean that the first guitar was preferred over the second.

	Sample														
Number of test participant / vs.	AaQl AaQh	AaQl AaQm	AaQl CcQl	AaQm AaQh	AaQm CcQl	AaQm CcQm	AaQh CcQh	AaQh CcQl	AaQh CcQm	CcQl CcQh	CcQl CcQm	CcQm AaQl	CcQm CcQh	CcQh AaQl	CcQh AaQm
1	40	-68	21	-67	50	39	62	100	60	5	36	-34	0	55	-15
2	-83	-75	-84	42	-38	-42	52	87	-47	20	45	39	0	47	-49
3	53	0	72	74	0	70	2	61	73	44	52	-39	-39	0	-88
4	0	0	-41	32	-75	61	-59	60	63	-56	72	-16	-50	0	0
5	0	-50	26	-24	24	22	-52	55	24	0	100	-23	-3	22	22
6	100	100	0	100	100	100	0	100	0	-100	0	-100	0	-100	-81
7	-100	0	0	0	98	-36	-83	86	0	0	0	-100	-100	-79	0
8	-100	-53	0	-100	-100	0	0	0	0	-48	100	0	-75	47	51
9	-81	-66	54	42	0	-47	0	-68	0	0	0	25	0	-63	-56
10	-75	0	25	-75	0	20	39	-55	69	0	89	-24	0	-7	-67
11	-51	-11	-31	29	19	77	-55	-79	41	36	79	-45	-40	65	0
12	45	-49	51	0	65	78	71	-67	57	67	48	-66	0	-71	-63
13	65	-42	-80	-20	61	24	-81	-62	-73	-74	-47	-75	0	79	77
14	-56	25	-29	-60	0	83	86	61	21	0	57	-38	-59	-40	0
15	-42	0	0	-55	-71	84	68	73	-45	-24	-26	0	-27	-13	20
16	53	-72	42	-48	49	-37	73	100	62	-50	73	85	-75	83	40
17	-63	46	100	-65	-65	0	50	49	0	12	-49	0	58	-51	0
18	-26	-51	0	-51	0	75	0	-47	77	0	45	0	48	-90	-78
19	61	-27	100	49	73	-51	-75	100	50	-75	-79	69	0	-77	-100
20	67	-14	63	-75	65	71	-83	-78	71	0	0	-79	0	0	-82
21	24	-51	24	-51	100	25	-78	100	49	50	0	48	-74	0	0
22	-40	-30	-40	20	-30	30	40	30	30	20	-30	30	-50	0	-30
23	0	0	0	-15	-54	33	-48	89	59	-71	14	21	-51	13	10
24	-75	73	-75	0	-91	100	-87	75	100	75	49	-75	-76	71	47
25	-64	-25	41	-55	70	28	-48	68	0	-57	0	-70	-71	0	27
26	-64	45	57	-42	49	-53	0	57	76	-94	69	72	-100	87	-70
27	0	-50	0	-40	0	0	-35	-20	25	25	50	-50	-40	70	-25
28	-83	-63	-51	-79	53	-57	85	73	77	-71	69	-71	-32	-48	40
29	-79	-83	-75	-79	-43	86	59	-76	-79	79	96	100	93	83	-95
30	-100	-53	0	-100	50	100	49	-100	99	0	100	-100	0	47	51
31	-57	-4	-81	-76	87	17	44	-83	57	-26	61	95	0	0	40
32	100	100	50	100	-100	75	0	-100	100	75	0	0	-25	0	-75
33	0	0	100	-10	100	50	70	100	30	-100	-100	-30	20	-20	-50
34	-19	19	-55	13	-24	25	26	-40	20	-44	50	-9	0	39	0
35	-63	28	35	-30	-30	-40	26	33	33	14	0	33	0	24	43
Mean	22	13	-6	19	-10	-29	-2	-17	-33	11	-29	12	23	-3	16

The mean values in Table 2 can be presented more clearly in the matrix form of Table 3. Positive scores indicate that the guitar in the column was judged to be preferred to the guitar in the row.

Table 3 - Mean preference score matrices for guitars with varying top stiffness and density (model Aa *vs.* Cc) and varying Q (low, mid and high). Positive values mean that the guitar in the column was judged to be preferred to the guitar in the row.

Model		Aa			Cc		
	Q	Low	Mid	High	Low	Mid	High
Aa	Low	-	14	20	-6	-12	5
	Mid	-14	-	20	-11	-29	-16
	High	-20	-20	-	-19	-34	-3
Cc	Low	6	11	19	-	-29	11
	Mid	12	29	34	29	-	22
	High	-5	16	3	-11	-22	-

Some tendencies can already be suspected from the averaged pairwise data in Table 3. However, for statistical analysis, it is meaningful to return to individual ratings but use average scores for each guitar. To obtain an averaged preference score for each guitar (e.g., guitar AaQl), the mean of the pairwise scores (e.g., guitar AaQl compared to all others) was calculated for each guitar as follows: guitar AaQl = (AaQl*vs.*AaQm + AaQl*vs.*AaQh + AaQl*vs.*CcQl + AaQl*vs.*CcQm + AaQl*vs.*CcQh)/5. The resulting individual data are shown in Table 4. The presence of a normal distribution was again tested with the Shapiro-Wilk test because of the small sample size (35 test participants). All averaged ratings show a normal distribution and p > .05. Overall mean values are calculated at the bottom of the table.

Table 4 - Preference score averaged for each guitar.

	Model	Aa			Cc		
	Q	Low	Mid	High	Low	Mid	High
Number of test participant	1	-5.6	21	49.8	-26	-33.8	-5,4
	2	-65.6	17.2	26.6	20	16.6	-14,8
	3	32.8	46.4	1.8	-7.4	-54.6	-19
	4	-5	3.6	6.4	14.4	-52.4	33
	5	-4.6	10	10.2	-1	-34.4	19,8
	6	80	56.2	-20	-60	-40	-16,2
	7	15.8	12.4	20.6	-36.8	-32.8	20,8
	8	-40	-39.6	40	30.4	-35	44,2
	9	-11	23.4	-5.8	2.8	14.4	-23,8
	10	-3.8	2.4	40.6	23.8	-40.4	-22,6
	11	-22.6	27.2	-14.2	41.2	-56.4	24,8
	12	36.8	51	3.2	13.2	-49.8	-54,4
	13	-12.2	6	-52.2	-8	4.2	62.2
	14	3.6	-0.4	56.8	5	-51.6	-13.4
	15	-5.8	-12.4	38.6	-10.4	-8	-2
	16	-29	-0.8	46	-33.6	-17.6	35
	17	26.8	-35.2	45.4	-24.2	21.4	-34.2
	18	2.6	30.6	21.4	18.4	-29.8	-43.2
	19	28.4	39.6	-7	-85.4	29.8	-5.4
	20	39	31.4	-16.4	-10	-44.2	0.2
	21	-10.2	25	19.6	-34.8	-20	20.4
	22	-28	16	24	6	-10	-8
	23	-6.8	-9.2	23	-18.4	-27.2	38.6
	24	-14.6	-22.2	32.6	43	-80	41.2
	25	4.4	8.2	27.8	-47.2	-33.8	40.6
	26	-24.2	-4.2	47.8	-37.6	-24	42.2
	27	-14	7	2	19	-33	19
	28	-15.6	-12	79.4	-15.4	-38.4	2
	29	-84	28.4	12.4	73.8	18	-48.6
	30	-20	10.4	49.6	30	-79.8	9.8

31	-47.4	-1.6	30.2	22.4	-8	4.4
32	50	10	-40	45	-40	-25
33	30	38	42	-100	2	-12
34	-17	-1	2.4	25	-20.8	11.4
35	-11.4	-34.2	37	-4.8	8	5.4
Mean	-4	10	19	-4	-25	4

The individual data shown in Table 4 are plotted again for better illustration in the Appendix, Figures A1 to A4, separately for guitar groups Aa or Cc and professionals or amateurs. A large variation between subjects is evident.

A repeated-measures ANOVA was applied for the statistical analysis of the preference data that are listed in Table 4. *The ANOVA revealed that there is a highly significant influence of Q [$F(2,66) = 8.538$, $p = 0.001$] on the instrument scores. Additionally, the difference between the instrument scores for group Aa vs. group Cc [$F(1,33) = 10.274$, $p = 0.003$] is highly significant. In agreement with the conclusions from the last study, low-density and low-stiffness wood (Aa) is on average preferred, although the revised 814ce V-Class guitars examined differ greatly in construction from the previous ones. However, these main effects have to be interpreted while considering the highly significant interaction effect between group (Aa vs. Cc) and Q [$F(2,66) = 5.362$, $p = 0.007$].* The averaged preference ratings for varying Q are plotted with 95% confidence intervals in Figure 7. The guitars with high Q are judged to be statistically significantly better than guitars with mid and low Q. To explore the significant main effect, post hoc pairwise comparisons were conducted using the Bonferroni correction for multiple comparisons. Guitars with high Q were preferred very significantly to guitars with mid Q (average difference = 21.65, $p = 0.002$) and low Q (average difference = 17.13, $p = 0.017$). To better understand the results, the significant interaction effect needs to be discussed.

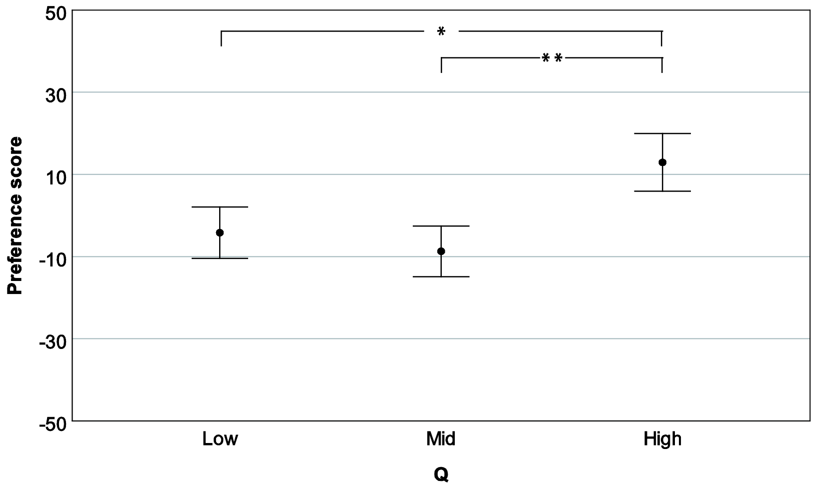

Figure 7 – Mean preference evaluation for guitars with low, mid and high Q plotted with 95% confidence intervals.

Figure 8 visualizes the interaction by showing the grouped preference results for guitar groups Aa and Cc against Q. On average, guitars with tops made from low-density and low-stiffness tonewood (group Aa) are preferred over guitars with heavier and stiffer tops (group Cc). This effect seems to originate mainly from the mid and high Q conditions.

In guitar group Aa, an increase in Q seems to lead to an increase in preference. However, the confidence intervals overlap, and the differences are small. In guitar group Cc, a stronger positive correlation between Q and the preference score is visible for mid and high Q. However, the opposite is observable between low and mid Q.

In summary, the quality factor (Q) of the top wood seems to have an influence on the preference of the examined guitars. The results suggest that there is a tendency of higher Q being preferred over lower Q. However, there seems to be some interaction with the density and stiffness of the tonewood (groups Aa and Cc).

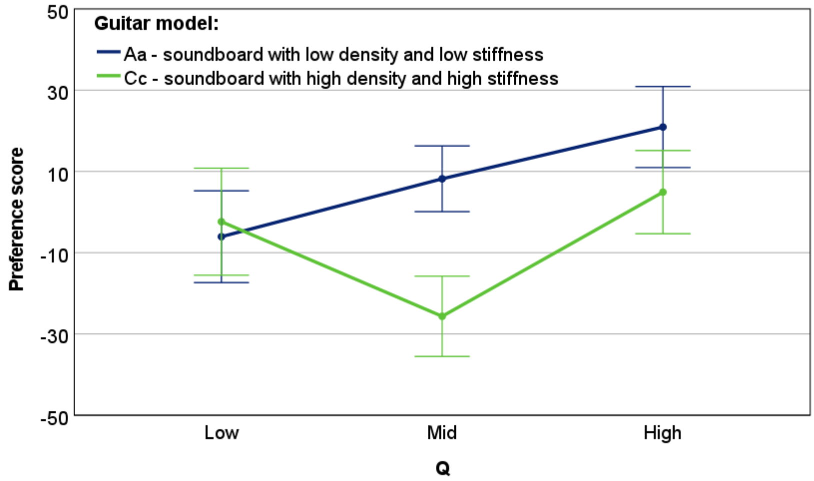

Figure 8 – Profile plot for the significant interaction between guitar group and Q. If there was no interaction effect, the lines would be parallel. Plotted are mean preference scores with 95% confidence intervals.

No between subject effect for expertise (professional vs. amateur guitarist) or any interaction effects including expertise were significant. Although no interaction was found, Figure 9 and Figure 10 show the separate results for professional and amateur guitarists. It should be noted that there were 22 studied musicians in one group compared to 13 amateurs in the other. Good agreement between professional and nonprofessional participants is visible.

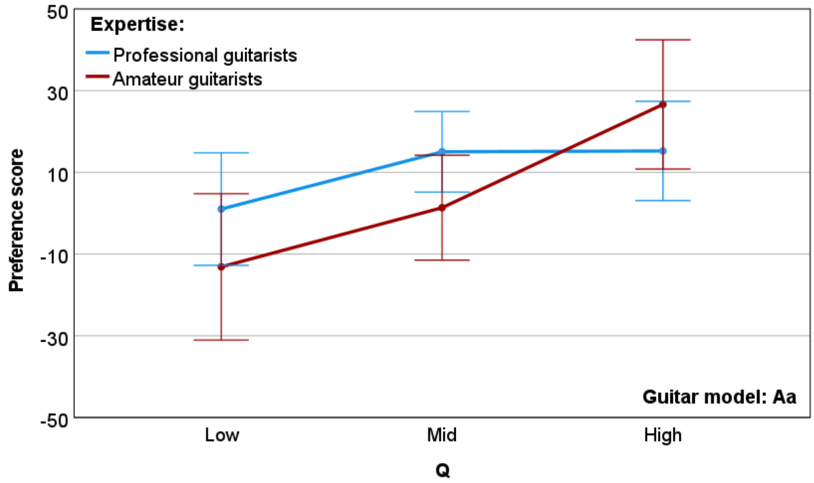

Figure 9 – Preference evaluation for guitar model Aa grouped by professional (blue) and amateur (red) guitarists. Plotted are mean values with 95% confidence intervals.

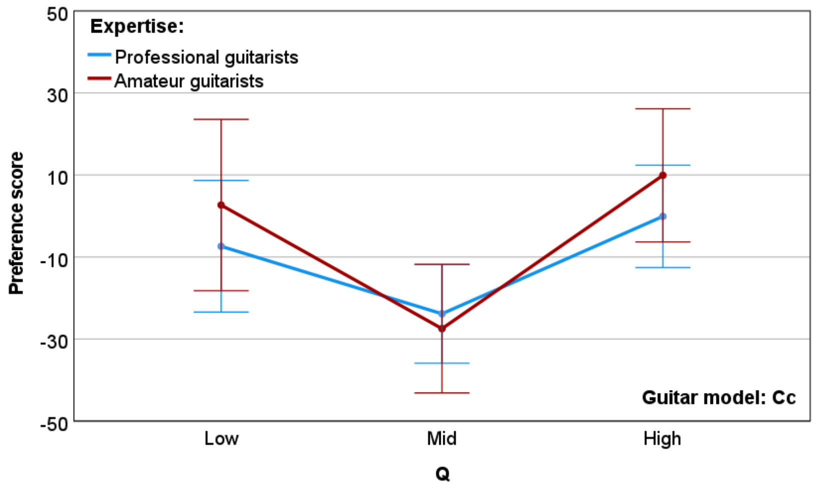

Figure 10 – Preference evaluation for guitar model Cc grouped by professional (blue) and amateur (red) guitarists. Plotted are mean values with 95% confidence intervals.

5.5 Results and discussion – X- *vs.* V-bracing

At the end of the listening test, four additional pairs of guitars were evaluated to compare the winners of the previous study with X-bracing (Aa-114, Cc-117) against the new guitars with V-bracing (AaQl and AaQh, CcQl and CcQh) from the current study. Only four pairs were selected. The resulting preference evaluation is plotted in Figure 11. Positive mean values suggest that the guitars with V-bracing were preferred over those with X-bracing. However, the confidence intervals overlap with 0 (no preference), indicating no statistical significance.

For statistical analyses, the data were averaged individually over all four comparisons. The presence of a normal distribution was tested again with the Shapiro-Wilk test. It was found that normal distribution can be assumed ($p = 0{,}357$). A *t*-test was applied to check if the mean preference score is different from zero. The test revealed that the difference is not significant ($p = 0.055$, 2-sided, $T(34) = 1.992$). **This means that there seems to be a tendency of the selected V-braced guitars being preferred over the guitars with X-bracing; however, there is no statistical significance.**

Figure 12 shows the preference scores averaged over all four pairs but grouped by listener expertise. Again, good agreement between professional and nonprofessional participants is visible.

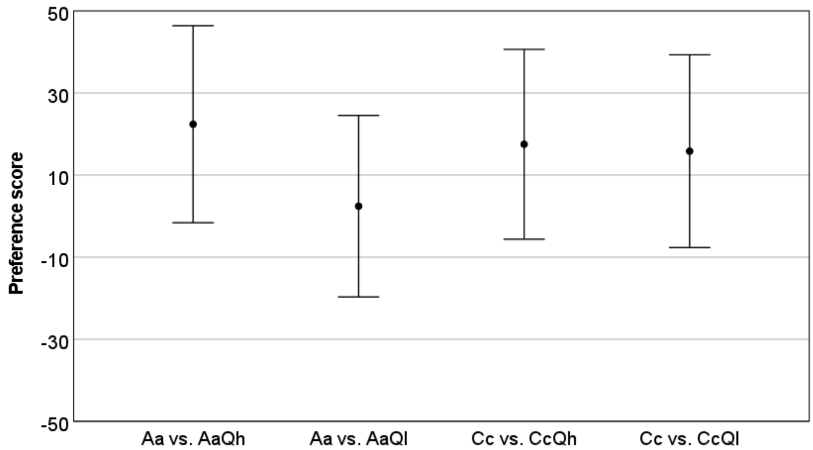

Figure 11 – Preference evaluation for four selected pairs of guitars with X-bracing (Aa-114 and Cc-117) *vs.* V-bracing (AaQl, AaQh, CcQl and CcQh). Plotted are mean values with 95% confidence intervals. Positive values indicate that V-bracing was preferred.

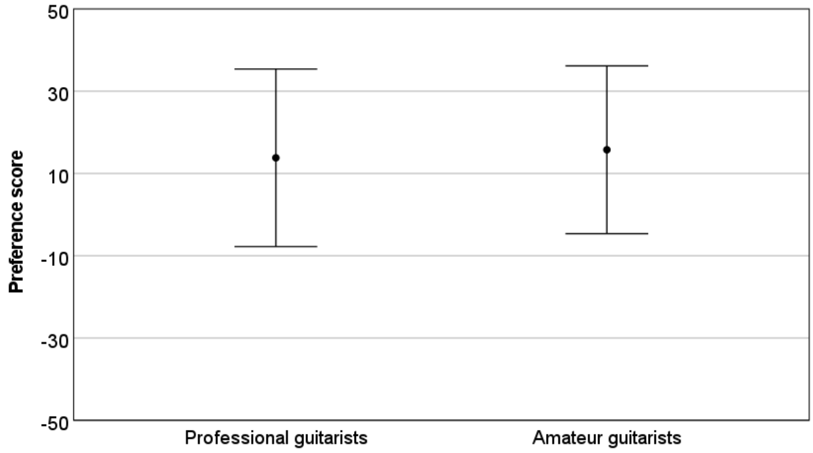

Figure 12 – Averaged preference evaluation of X-braced *vs.* V-braced guitars grouped by expertise. Plotted are mean values with 95% confidence intervals. Positive values indicate that V-bracing was preferred.

6 Playing test

After completion of the listening tests, all test participants were invited to play the six V-Class guitars themselves.

6.1 Setup

A quiet library room was used for the playing test. The guitars were placed beside each other and randomly numbered from 1 to 6. Additional guitar stands were used to be able to quickly change between instruments. The evaluation was done using a PC as shown in Figure 13.

Figure 13 – Configuration for the playing test.

A screenshot of the guided user interface is shown in Figure 14.

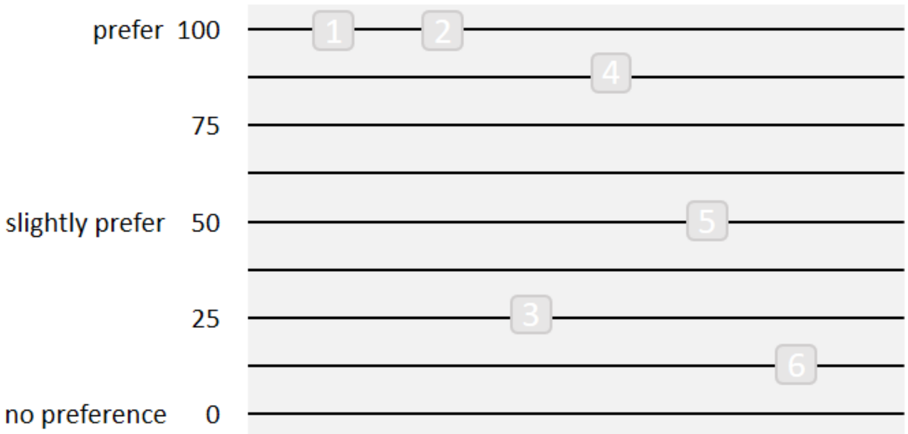

Figure 14 – Preference grading scale for the playing test with six guitars.

6.2 Subjects

The same thirty-five subjects (twenty-six males and nine females) participated in the study. Twenty-two of the participants were professional guitarists either with a completed degree or studying guitar or musicology. The remaining thirteen subjects were passionate amateur guitarists with several years of experience playing the guitar in bands. The mean age of the participants was 36.9 years.

6.3 Experimental design

The test participants were instructed as follows: "In the final playing test, you can try the guitars yourself. You can switch freely between all guitars at any time. Please rate all instruments on the preference scale using the gray boxes numbered 1 to 6. The rating scale is to be regarded as a continuous scale with equal intervals." No time limit was defined. On average, the participants needed 30 minutes for their evaluation.

6.4 Results and discussion

The individual data of the playing test can be found in the Appendix, Figures A5 through A8. The experimental design assumes that subjects can produce scores on an interval scale level. Therefore, the data are analyzed using parametric statistical testing. The presence of a normal distribution was tested with the Shapiro-Wilk test. Only one of 6 ratings showed significance above 0.05, meaning that normal distributions cannot be generally assumed. However, after visual inspection of the data and for the reasons discussed above, a repeated-measures ANOVA was applied. The ANOVA revealed no significant main effects or interactions. On average, no guitar was preferred over another.

This means that the differences found in the listening test could not be proven by the playing test. This can have various causes. First, the method of paired comparisons, implemented in the listening test, is more suitable for detecting small differences. However, the playing test allows for more flexibility (e.g., individual content and varying individual playing style). Therefore, more facets of an instrument, including more tonal aspects, can be evaluated. In addition, other factors in addition to the sound influence the evaluation during a playing test (e.g., interaction with the instrument and optics).

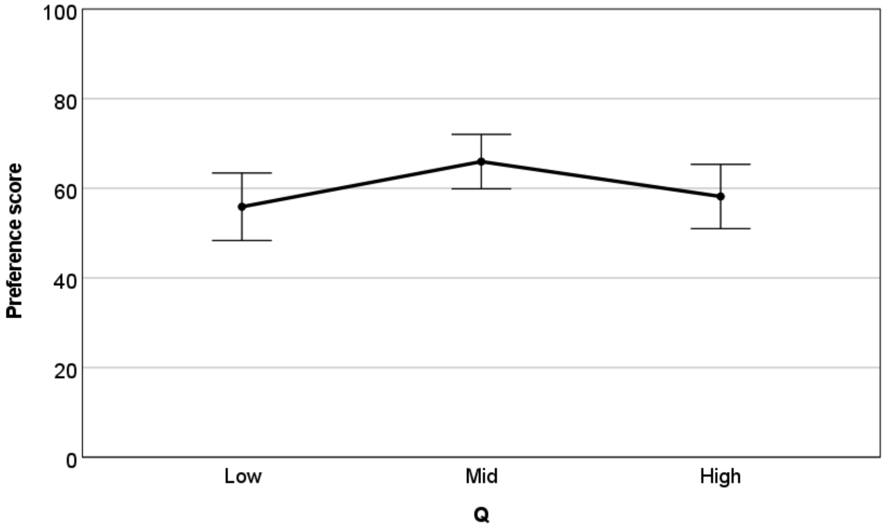

Figure 15 - Mean preference evaluation in the playing test for guitars with low, mid and high Q plotted with 95% confidence intervals. No effects were found to be statistically significant.

7 Appendix

7.1 Individual data of the listening test

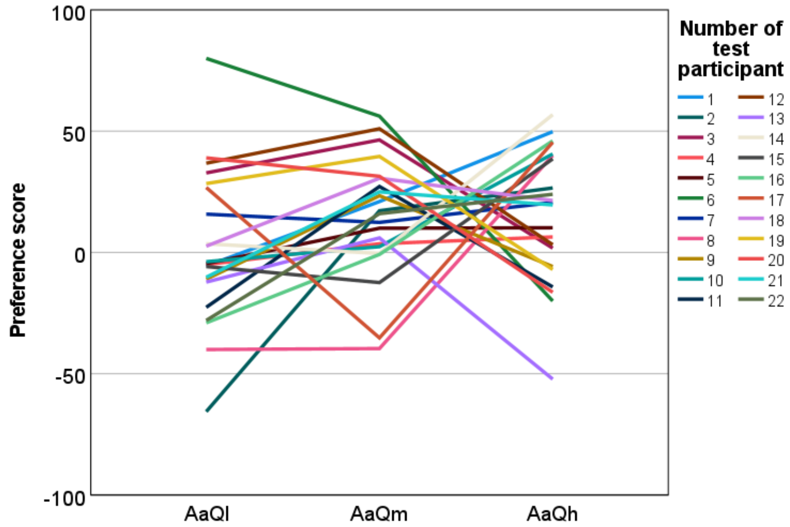

Figure A1 – Individual preference evaluation for guitar models Aa with varying Q of all 22 professional guitarists in the pairwise listening test.

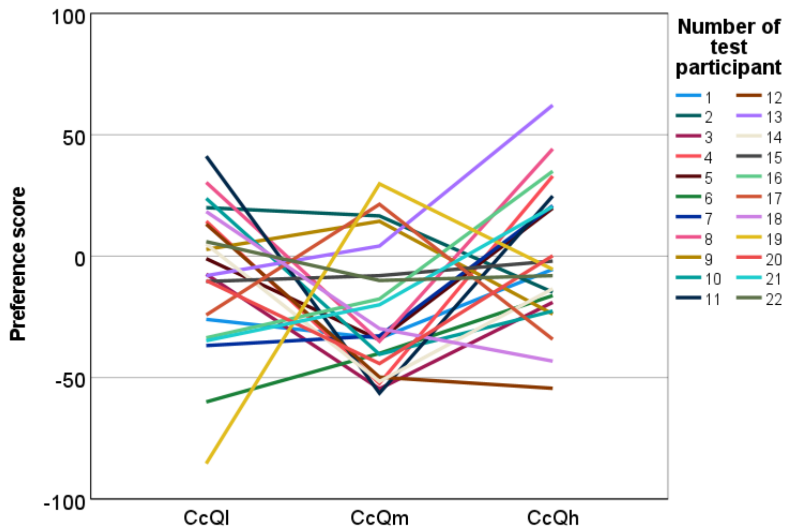

Figure A2 – Individual preference evaluation for guitar models Cc with varying Q of all 22 professional guitarists in the pairwise listening test.

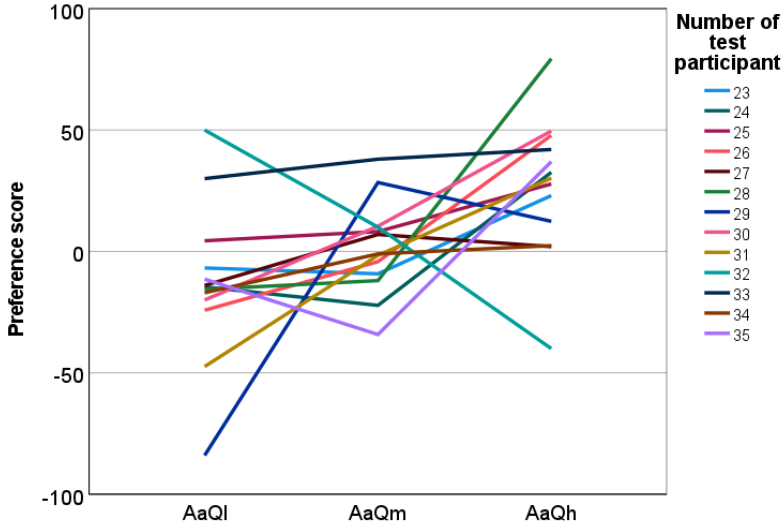

Figure A3 – Individual preference evaluation for guitar models Aa with varying Q of all 13 amateur guitarists in the pairwise listening test.

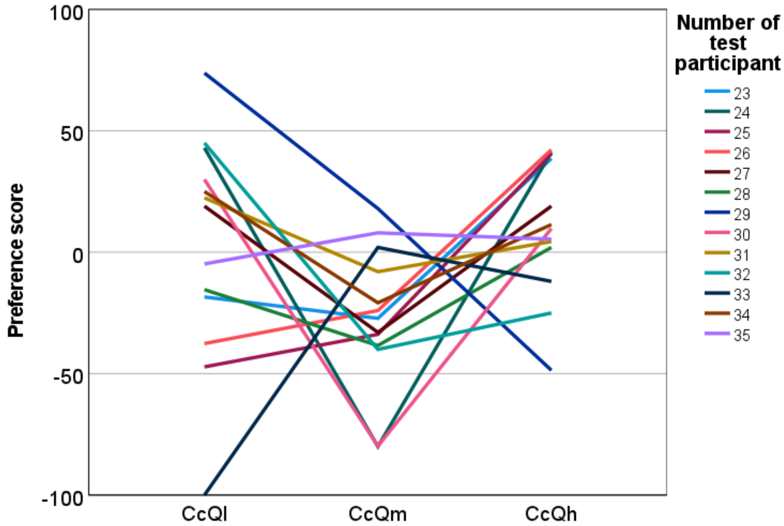

Figure A4 – Individual preference evaluation for guitar models Cc with varying Q of all 13 amateur guitarists in the pairwise listening test.

7.2 Individual data of the playing test

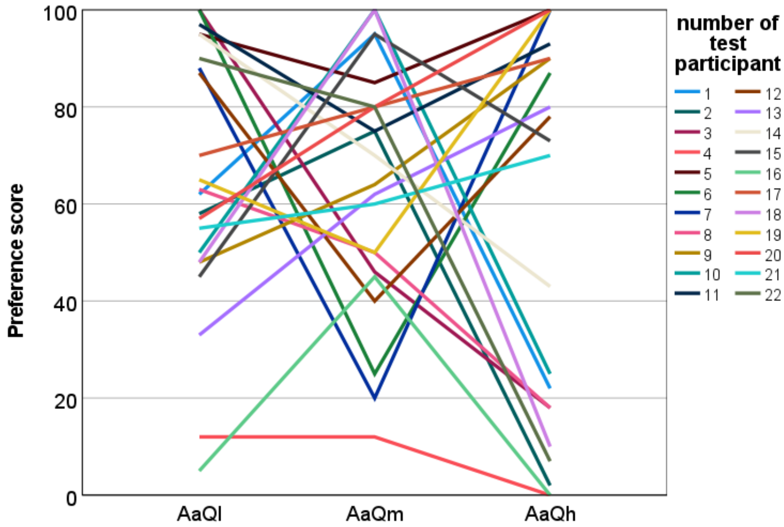

Figure A5 – Individual preference evaluation for guitar models Aa with varying Q of all 22 professional guitarists in the playing test.

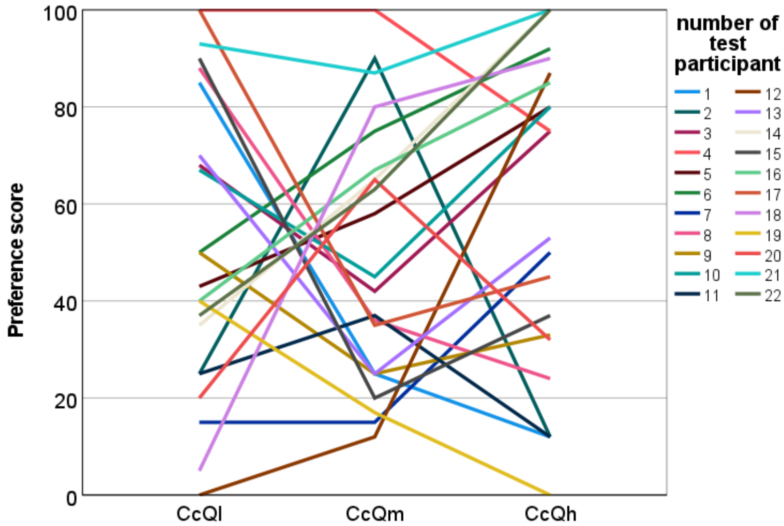

Figure A6 – Individual preference evaluation for guitar models Cc with varying Q of all 22 professional guitarists in the playing test.

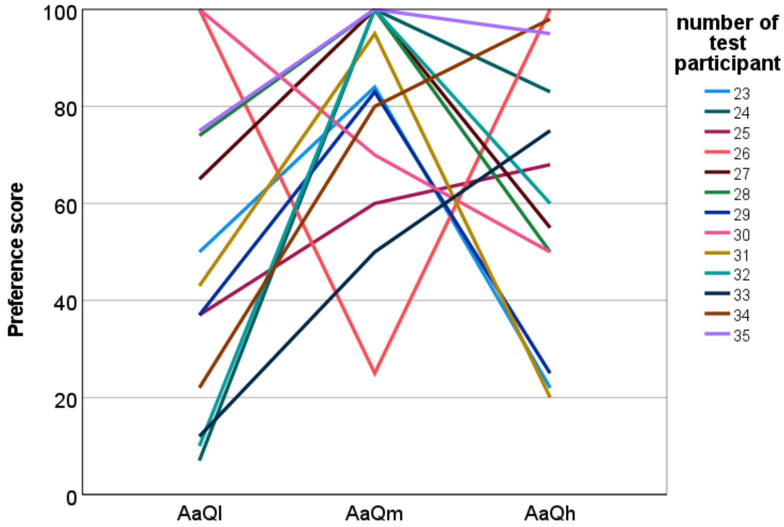

Figure A7 – Individual preference evaluation for guitar models Aa with varying Q of all 13 amateur guitarists in the playing test.

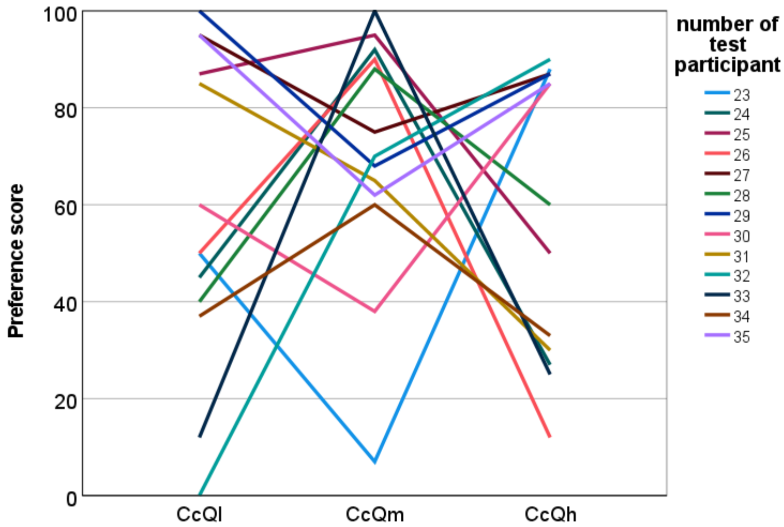

Figure A8 – Individual preference evaluation for guitar models Cc with varying Q of all 13 amateur guitarists in the playing test.

Literature

[1] ITU-R BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems," *Int. Telecommun. Union*, p. 34, 2015.

[2] ITU-R BS.1116-3, "Methods for the subjective assessment of small impairments in audio systems," *Int. Telecommun. Union*, p. 30, 2015.

[3] M. Schoeffler *et al.*, "webMUSHRA — a comprehensive framework for web-based listening tests," *J. Open Res. Softw.*, vol. 6, no. 8, pp. 2–8, Feb. 2018.

[4] J. Herre, F.-R. Stöter, and M. Schöffler, "Lab Course - Statistical Methods for Audio Experiments," 2015.

[5] M. W. Vasey and J. F. Thayer, "The Continuing Problem of False Positives in Repeated Measures ANOVA in Psychophysiology: A Multivariate Solution," *Psychophysiology*, vol. 24, no. 4, pp. 479–486, 1987.

[6] M. Wirtz and F. Caspar, *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen, Germany: Hogrefe, 2002.