Subjective evaluation of musical instruments on the basis of solo pieces of music

Sebastian Merchel[1], *Rüdiger Hoffmann*[2]

Technische Universität Dresden, 01062 Dresden,
Institut für Akustik und Sprachkommunikation,
[1] E-mail: sebastian.merchel@ias.et.tu-dresden.de,
[2] E-mail: ruediger.hoffmann@ias.et.tu-dresden.de

Abstract

This study examines the human quality perception of musical instruments. It provides the background for future development of objective algorithms ([1],[2]) to distinguish between musical instruments on a quality basis. Previous studies showed that evaluation using single tones is not sufficient ([3],[4],[5],[6]), thus tone sequences will be used. This corresponds also more to the natural situation where a guitar is played. Three listening tests have been developed, using two different methods (serial and block by block presentation of stimuli). They have been realized and evaluated. It was asked to judge the acoustical overall quality of selected classical guitars by listening to recorded tone sequences (scale and melody). The selected binaural technology in combination with headphone compensation seemed to achieve good results. Not only the attributes of the instrument itself affect the perceived quality, also other factors (independent variables) might influence. Parameters like the playing *musician*, the *room* in which the instrument is played, the played *sequence* and the *repetition* of the same sequence by the musician have been included into the experimental design. The *listeners* were divided into two groups, those who play guitar themselves and those who do not.

The complexity of human quality perception can be seen from a multitude of interactions between the mentioned factors. It was concluded, that hierarchical plans, which would reduce the necessary effort in listening experiments, can only be applied very restricted. To present stimuli in blocks gave no benefits, because of difficult comparison between blocks. There was no significant difference between the quality judgment of guitarists and non guitarists.

1. Introduction

First the influencing factors will be specified in order to record the required stimuli. Afterwards the design of the listening tests is described. Two experiments evaluate if the independent variables interact. A third experiment compares the quality perception of ten different instruments, when all other variables are held constant. The results will be summarized and an outlook will be given.

2. Stimuli

The recording of samples for the listening test was done using a binaural recording head. Some attributes of the guitar, like the playability, the optical appearance or the radiation characteristic, have only direct influence during the recording. 3 selected sequences (30 s each) have been played on 10 different classical guitars by 5 professional musicians. An overview of the selected guitars, representing a wide commercial spectrum, can be seen in Table 1. The first 6 instruments have very distinct construction. The last 4 are similar, but differ as a group from the previous ones.

Table 1: Overview of the selected guitars

instrument	description
G1	Takamine C-128 (1979) (industrial instrument from Japan)
G2	Armin Gropp (1977) (master craftsman from Germany)
G4	Marlin MC 315 (industrial instrument from Japan)
G5	Landola SL 3 Nr. 151472 (industrial instrument from Finland)
G22	Session C 425 (layered, low-cost, assumedly Indonesia)
G23	Doppelbodengitarre Eberhard Kreul and IfM (approx. 1975) (prototype with twin corpus)
G24	Höfner HF 12 (2002) (test model with modified top, Germany)
G25	Höfner HF 12 (2002) (test model with modified top, Germany)
G26	Höfner HGL 50 SE (2001) (test model with modified top, Germany)
G27	Höfner HGL 50 (2005) (test model with modified wood, Germany)

The recording took place in two extreme situations, the reverberant conference room and the anechoic room of the Institut für Musikinstrumentenbau (Zwota). Each sequence was repeated once. This resulted in 600 recorded samples, which are too many to examine in a single complete listening experiment.

3. Listening experiments

In the listening experiments participants were asked to judge the overall acoustical quality of the guitars by listening to the recorded stimuli. There was *no guiding* towards possible quality aspects like sharpness, loudness, spectral richness or dynamics of a guitar.

To evaluate the influence of many factors in a hierarchical lis-

tening test design, the independent variables must not interact. The interactions should be evaluated in a reduced **first listening experiment**. Only two levels of each factor were used, resulting in (2 musicians x 2 rooms x 2 instruments x 2 sequences x 2 repetitions) 32 samples. In a first experiment, they have been presented one after the other (serial) after a short introduction sequence. The participant had the possibility to stop each sample after 5 s. The experiment was still long (20 min). The participant was asked to judge only the acoustical overall quality on a discrete 5 point ITU MOS scale [7], which was translated into German [8] (ausgezeichnet, gut, ordentlich, dürftig, schlecht).

To introduce direct comparison, it was tempted to group the samples in blocks. This **second experiment** consisted of 4 blocks with 6 samples each (2 hidden anchor samples, which are repeated in all blocks and 4 random samples) not to exceed the needed time, compared to the first experiment. Thus 16 samples (4 random samples x 4 blocks) can be evaluated. The anchor samples have been selected to align the judgment between blocks. It was found by experience that a quasi continuous MOS scale was more suitable for this task. A screenshot can be seen in Figure 1. The first block was rerun in the end, to see if judgment remained constant.

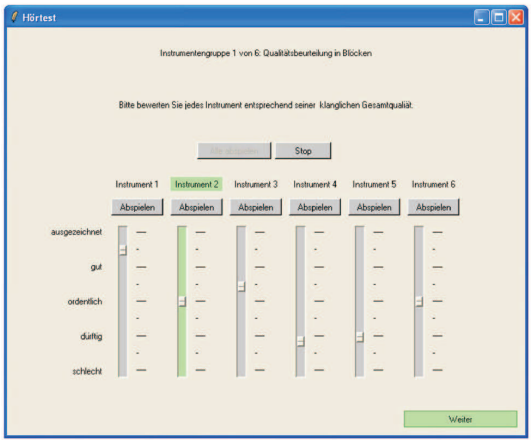

Figure 1: Screenshot of interface for second listening test

In a **third experiment** all 10 guitars were compared in one block keeping all other factors constant (repetition of 1 melody played by 1 musician in the conference room using all 10 guitars).

4. Results

32 participants took part in the experiments, half of them guitarists. No data were removed. A multifactorial analysis of variance with repetition was carried out for statistical analysis. Data were checked for normal distribution with the KS-test. A multitude of significant interactions on a 5% significance level in the **first experiment** shows the complexity of human quality perception. An example can be seen in the interaction diagram Figure 2, where the significant main effect (quality difference between two guitars) is relativized by a significant interaction between *instruments* and played *sequences*. The main effect (better quality valuation for guitar $G1$ than for guitar $G5$) is only true for sequence Se_1, but not for sequence Se_2. The acoustical quality differences are small compared to a high standard deviation of the judgments for each stimuli. This can be seen in Figure 3. Thus it is necessary to ask for practical relevance of the results.

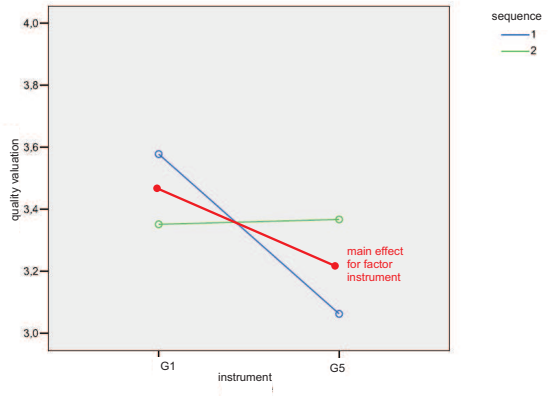

Figure 2: Interaction diagram that shows the mean quality valuation for *instrument* $G1$ and $G5$ in dependence of the factor *sequence*. Main effect for factor *instrument* is significant with average difference of 0.25, but relativized by the significant interaction between *instrument* and played *sequence*. Notice that the quality scale is reduced to the interval 'gut' (3) to 'ausgezeichnet' (4) for better graphical presentation

All significant effects are summed up in Table 2. The case plotted in Figure 2 can be found in line one to three of the table.

Listeners were debriefed after participating. They found it hard to judge the quality without having a direct comparison between different samples. Judgement might have also been difficult, because of the great variability of the boundary conditions. E.g. the two rooms used are very dissimilar and in addition unnatural. Typically we tend to listen to guitars in a concert space and not in a reverberant conference room or an anechoic environment.

Most of the participants (not only non guitarists) had difficulties to name or describe the quality criteria they attended to.

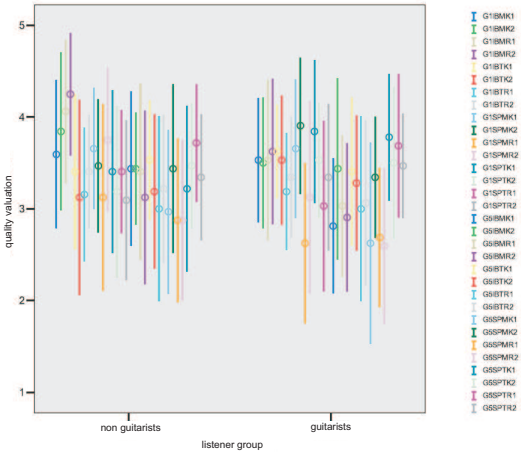

Figure 3: Error bars for results of first experiment with 32 participants, of which 16 guitarists and 16 non guitarists. It shows the mean quality valuation ± one standard deviation of each listener group for each stimuli

Table 2: Summary of the results (main effects and interactions) in the first experiment. The quality influencing factors will be abbreviated as follows: guitar G, musician M, room R, repetition A, sequence Se and listener group HG. Grey lines show significant effects, but are relativized by interactions of higher order. Only first and second order interactions are interpreted. Notice that there are significant interactions of third and fourth order, that complicate the interpretation. Some interactions (e.g. $G*R*M$ and $M*R*G$) occur twice in dependence of the effect studied

interaction	valuated quality of the guitar for	condition
G	$G_1 > G_5$	
$G*Se$	$G_1 \approx G_5$	Se_2
	$G_1 > G_5$	Se_1
$G*R*M$	$G_1 > G_5$	$(R_1 \vee R_2)M_1 \vee R_1M_2$
	$G_1 \approx G_5$	R_2M_2
$G*R*HG$	$G_1 > G_5$	
$G*R*A$	$G_1 > G_5$	
$G*R*A*M$		
$M*R$	$M_1 \approx M_2$	R_1
	$M_1 > M_2$	R_2
$M*R*G$	$M_1 \approx M_2$	$(R_1 \vee R_2)G_5 \vee R_1G_1$
	$M_1 > M_2$	R_2G_1
$M*R*G*A$		
$M*Se$	$M_1 > M_2$	Se_1
	$M_1 < M_2$	Se_2
$M*Se*R$	$M_1 > M_2$	$Se_1(R_1 \vee R_2)$
	$M_1 < M_2$	$Se_2(R_1 \vee R_2)$
$M*Se*A$	$M_1 > M_2$	$Se_1(A_1 \vee A_2)$
	$M_1 < M_2$	$Se_2(A_1 \vee A_2)$
$A*Se$	$A_1 < A_2$	Se_1
	$A_1 > A_2$	Se_2
$A*Se*R$	$A_1 > A_2$	Se_1R_1
	$A_1 \approx A_2$	Se_2R_2
	$A_1 < A_2$	$Se_2R_1 \vee Se_1R_2$
$A*Se*M$	$A_1 > A_2$	Se_2M_2
	$A_1 \approx A_2$	Se_2M_1
	$A_1 < A_2$	$Se_1(M_1 \vee M_2)$
$G*M*A*HG$		
$M*Se*R*A*HG$		

The **second experiment** in blocks gives similar results. Again the main effect for the independent variable *instrument* is significant with an average difference of 0.5. In addition the main effect for *musician* becomes significant. Again many interactions need to be interpreted. The important influence of the factor *sequence* (as seen in Figure 2) can not be evaluated, because it was canceled due to time restrictions. The needed time to evaluate the selected samples was even longer than before. The alignment between blocks with hidden anchor samples gives some crucial interpretation difficulties. Thus there was no apparent advantage of the blockwise method, if samples needed to be segmented in several blocks.

The comparison of all guitars under defined conditions in a **third experiment** showed that only one instrument (instrument $G22$) is judged significantly different from most of the other instruments. This result can not be generalized, considering the findings from the previous experiments. In the previous experiments, instrument $G1$ was preferred to instrument $G5$. This effect is not significant under this conditions.

Surprisingly there was no significant difference between the quality judgment of guitarist and non guitarist under the given conditions. This can be seen in Figure 4. A detailed report can be found in [9].

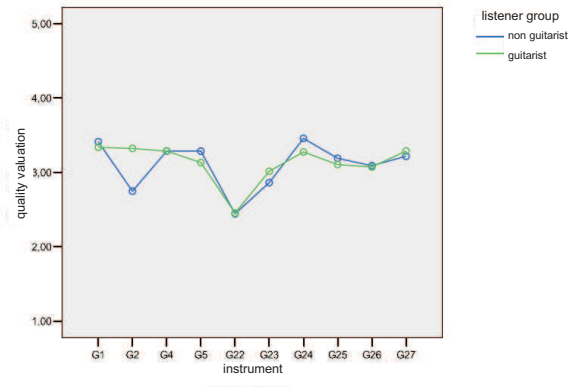

Figure 4: Interaction diagram for not significant interaction between *instrument* and *listener group* in third listening experiment

5. Outlook

In further experiments more instrument groups (e.g. violins) will be studied, using the serial method. The described analysis implies consistent intervals between the verbal steps of the translated MOS scale. To justify this assumption, additional numbers on the MOS scale should be used in further experiments. More unconsidered quality influencing factors might exist (e.g. the used strings), which have been kept constant until now.

There are no significantly tested quality differences between most of the guitars under the described conditions. This doesn't imply that there are no perceivable differences at all between these instruments. This should be proved in future experiments.

The mentioned interactions indicate the complexity of human quality perception. Instrumental methods for estimation of perceived acoustical quality differentiation between musical instruments might become relative extensive. They have to be validated with careful subjective experiments.

6. Acknowledgment

This study was part of the cooperative project "Bewertung und Beurteilung von Musikinstrumenten anhand von Solomusikstücken" (Assessment and evaluation of musical instruments on the basis of solo pieces of music) and resulted in a diploma thesis.

The author wants to thank M. Eichner for supervision, G. Ziegenhals (Institut für Musikinstrumentenbau, TU-Dresden) for informative discussions and R. Jäger (Institut für Allgemeine Psychologie, Biopsychologie und Methoden der Psychologie, TU-Dresden) for competent support. Thanks to all staff members, students and guitarists, who voluntarily participated in the listening experiments.

7. References

[1] Eichner, M., Wolff, M., Hoffmann, R. *Instrument classification using HMMs* (to appear at 7th International Conference on Music Information Retrieval 2006, Canada)

[2] Löschke, H. *Differenzierbarkeit von Musikinstrumenten* (2006, Fortschritte der Akustik - DAGA)

[3] Krishna, A. G., Screenivas, T. V. *Music instrument recognition: From isolated notes to solo phrases* (Proc. ICASSP 2004, Montreal, IV-265–IV-268)

[4] Fasold, W., Kraak, W., Schirmer, W. *Taschenbuch Akustik, Teil 2* (1984, Verlag Technik, Berlin, S. 1106 - 1110)

[5] Meinel, E., Boehm, T., Miklaszewski, K., Blutner, F. *Estimation of guitar sound quality* (1986, Archives of Acoustics 11 3, S. 203 – 229)

[6] Valenzuela, M. N. *Subjektive Beurteilung der Qualität und Ähnlichkeit von Flügelklängen* (1995, Fortschritte der Akustik - DAGA 95, S. 587 – 590)

[7] ITU-R *BS.1284-1: General methods for the subjective assessment of sound quality. International Telecommunications Union - Radiocommunication Assembly* (2003)

[8] ITU-T Rec. P.800.1 *Mean Opinion Score (MOS) terminology* (2003, International Telecommunication Union, CH-Geneva)

[9] Merchel, S. *Diploma Thesis: Untersuchungen zur subjektiven und objektiven Bewertung und Beurteilung von Musikinstrumenten anhand von Solomusikstücken* (2005, TU-Dresden, IAS)