Ratio-Scaling of Listener Preference of Multichannel Reproduced Sound*

Sylvain Choisel[1,2], Florian Wickelmaier[1]
[1] *Sound Quality Research Unit, Department of Acoustics, Aalborg University, Denmark*
[2] *Bang & Olufsen A/S, Struer, Denmark*

Introduction

Subjective evaluation of sound quality is often conducted using direct scaling methods such as magnitude estimation or category scaling. These methods are attractive at first glance because they are fast and readily yield numerical values. However, the validity of scales thus obtained relies on many implicit assumptions, one of them being the unidimensionality of the underlying sensation continuum. While unidimensionality might hold for simple stimuli, this is a non-trivial assumption in the case of complex spatial sounds. Another assumption is the subjects' ability to map their sensation magnitude onto a scale. Paired comparisons, on the other hand, require nothing but simple comparative judgments, and thereby eliminate response biases due to scale usage. In addition, from paired comparison judgments, the unidimensionality of the underlying sensation (e. g. the overall preference) can be verified. Finally, a ratio scale can be derived by means of probabilistic choice models such as the Bradley-Terry-Luce (BTL) model [1].

Method

Setup The loudspeaker configuration (Figure 1) consisted of a 5-channel surround setup following the ITU-R BS.775-1 recommendation with two additional loudspeakers at ±45°. The speakers were Genelec 1031A monitors, placed in a listening room complying with the ITU-R BS.1116 requirements. It had an area of 60 m^2 and a reverberation time between 0.25 and 0.45 s.

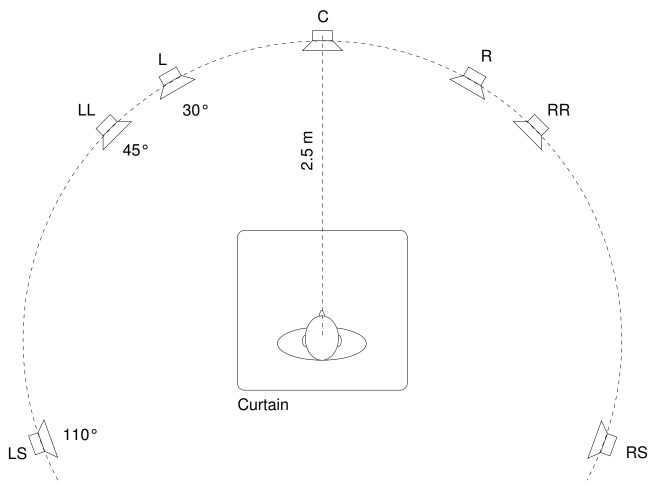

Figure 1: Loudspeaker configuration in the listening room. The setup was hidden to the subject by a curtain.

*This work was supported by the Centercontract on Sound Quality, Aalborg University.

Stimuli Four musical excerpts were selected from commercially available multichannel recordings:

- Beethoven, *Sonata 21, op. 53 (Rondo).* M. Kodama, *Piano Sonatas Nos. 21, 23 & 26.*
- Rachmaninov, *Blazen Muzh.* St. Petersburg Chamber Choir/Korniev, *Vespers.*
- Steely Dan, *Everything Must Go.* In *Everything Must Go.*
- Sting, *Stolen Car.* In *Sacred Love.*

The sounds were about 5 s long, and were carefully cut to include a complete musical phrase. From the original 5-channel recordings (abbreviated *or*), several formats were derived. Stereo (*st*), mono (*mo*) and phantom mono (*ph*) were created by down-mixing, and three up-mixing algorithms were used to recreate a multichannel signal from stereo; a simple matrix decoding (*ma*): $LS = (L-R)/\sqrt{2}$ and $RS = (R-L)/\sqrt{2}$, and two commercially available systems: Dolby Pro-Logic II and DTS Neo:6, referred to as up-mixing 1 and 2 (*u1* and *u2*) in no specific order. Finally, a wide-stereo (*ws*) mode was produced by playing the stereo signal trough loudspeakers LL and RR. Within each type of programme material, all eight reproduction modes were matched in loudness by eight listeners, who did not participate in the main experiment, using an adaptive procedure (2AFC 1-up/1-down).

Subjects 40 subjects participated in this experiment. They were all selected among 78 candidates, based on their hearing abilities and their verbal fluency (for a verbal elicitation task in a later experiment [2]).

Procedure For each of the 28 possible pairs of reproduction modes the subjects indicated which one they preferred. Three replicates of each pair were presented, the order within a pair being balanced across subjects. Every listener repeated this procedure for the four types of programme material. Thereby, no preference judgment was made between samples from different musical excerpts.

BTL model The aggregated choice frequencies were analyzed using the Bradley-Terry-Luce model which relates the probability P of preferring sound x over sound y to scale values u through the equation

$$P_{xy} = \frac{u(x)}{u(x) + u(y)},$$

where u is unique up to multiplication by a constant, and therefore provides a measurement of preference on a ratio-scale level. The BTL model imposes strong restrictions on the choice probabilities, in particular the weak (WST), moderate (MST) and strong (SST) stochastic transitivity, i. e. if $P_{xy} \geq 0.5$ and $P_{yz} \geq 0.5$ then

$$P_{xz} \geq \begin{cases} 0.5 & \text{(WST)} \\ \min\{P_{xy}, P_{yz}\} & \text{(MST)} \\ \max\{P_{xy}, P_{yz}\} & \text{(SST)} \end{cases}$$

Only if the paired-comparison data obey these restrictions, the BTL model can extract a preference ratio scale. It has been demonstrated [3], that especially SST is not likely to hold empirically when the sounds under study are complex. Statistically, the transitivity violations are evaluated by a goodness-of-fit test of the BTL model. If the BTL model holds, the preference judgments are transitive, indicating that the underlying sensation continuum is unidimensional. All data analysis was performed using a software package described in [4].

Results and discussion

Evaluation of the BTL model The number of transitivity violations is shown in Table 1 for the four types of programme material. Also reported in this table is the goodness of fit of the model. In a likelihood ratio test (LRT), the BTL model was compared to a saturated binomial model (see [4] for details). A high p-value indicates that the BTL model cannot be rejected. Only few violations of WST and MST were observed. The LRT indicates a very good fit of the BTL model which consequently classifies the SST violations as *not severe*. Thus, for each programme material the reproduction modes can be represented on a preference ratio scale.

Analysis of the preference scales Point and interval estimates of the BTL parameters (u-scale values) were obtained by maximum likelihood estimation. The resulting ratio scales are displayed in Figure 2. Further analyses were performed based on LRTs. In conclusion, for each musical excerpt the reproduction modes were highly significantly different; this means that the listeners had strong preferences for certain reproduction modes and could clearly distinguish them. The differences between the three repetitions, however, were not significant, which indicates that the preference is relatively stable, at least for the period in which the judgments were made. The types of programme material affect the evaluation of the reproduction modes. Thus, the influence of the reproduction modes cannot easily be generalized. It is, however, justified to conclude that the mono stimuli were generally the least preferred, that *ws* was preferred for the classical recordings but not for the pop music, and that the original 5-channel recording was most preferred only for one of the four musical excerpts.

Table 1: Transitivity violations and goodness-of-fit test. Columns 2 to 4 show the number of weak, moderate and strong stochastic transitivity violations.

Prog. mat.	WST	MST	SST	χ^2	df	p
Beethoven	0	2	14	9.13	7	0.988
Rachmaninov	2	4	19	16.96	7	0.714
Steely Dan	0	0	12	18.13	7	0.640
Sting	0	0	13	10.72	7	0.968

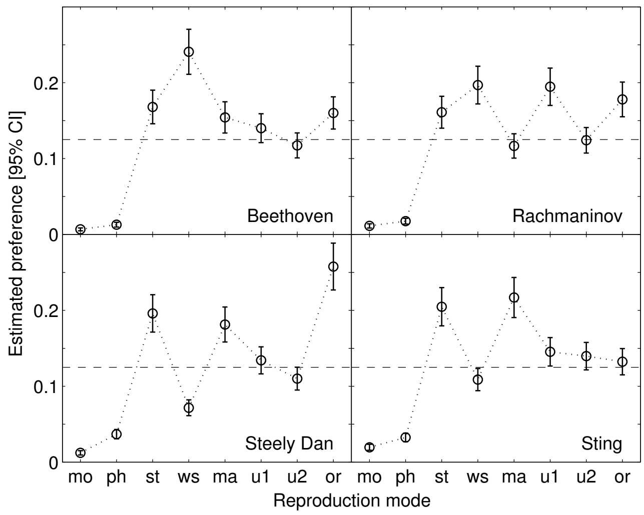

Figure 2: Preference ratio scale of the reproduction modes. Displayed are the maximum likelihood estimates of the BTL parameters and the 95% confidence intervals. The parameters are normalized to sum to unity; the indifference level (1/8) is shown by a dashed line.

Conclusion

The BTL model was found to predict the choice frequencies well. This implies that listeners were able to integrate the complex nature of the sounds into a unidimensional preference judgment. It further implies the existence of a preference scale on which the reproduction modes can not only be ordered, but also their ratios are meaningful. It is concluded that the BTL scaling constitutes a more rigorous approach to the quantification of listeners' preference than methodologically unfounded direct estimation procedures.

References

[1] R. A. Bradley: Paired comparisons: some basic procedures and examples. In: Handbook of Statistics. Vol. 4 P. R. Krishnaiah, P. K. Sen (eds.). Elsevier, Amsterdam, 1984, 299–326.

[2] F. Wickelmaier & S. Choisel: Selecting participants for listening tests of multi-channel reproduced sound. Presented at the *AES 118th Convention*, Barcelona, Spain, May 28-31, 2005.

[3] K. Zimmer, W. Ellermeier & C. Schmid: Using probabilistic choice models to investigate auditory unpleasantness. *Acust. Acta Acust.* **90** (2004) 1019–1028.

[4] F. Wickelmaier & C. Schmid: A Matlab function to estimate choice model parameters from paired-comparison data. *Behav. Res. Meth. Instr. Comp.* **36** (2004) 29–40.