Influence of acoustics on emotional impact of music in Konzerthaus and Philharmonie Berlin

Winfried Lachenmayr[1], Jukka Pätynen[2]

[1] *Müller-BBM, 82152 Planegg, Deutschland, Email: winfried.lachenmayr@mbbm.com*

[2] *Dept. of Computer Science, Aalto University School of Science, P.O. Box 15500, FI-00076 AALTO*

Introduction

The impact of music has been defined as powerful, effective, moving, and emotionally impressive as early as the 17th century [1]. Since, the research connecting psychology and music has become an established field. Although the sound experienced by the listeners in live music performances is heavily influenced by the acoustics of the concert venue, only little attention has been brought to the role of room-acoustics with regard to the impact. Instead, a handful of prominent perceptual descriptors, such as strength or spaciousness, have been considered along with the individual subjective preference [2],[3]. In addition, some recent research has explored the emotional component further: A study proposed tentatively that music presented with "good" acoustics activates the reward area of the brain [4]. Another study showed that the reproduced acoustics of certain concert halls elicit a high emotional impact [5] as well as the perceived range of musical dynamics [6]. The present work extends these findings by investigating self-reported impact during reproduced orchestra performance in two concert halls and in different listening distances.

Setup

Stimuli

A professional orchestra, the Staatskapelle Berlin, was accompanied during a pair of concerts in the Konzerthaus am Gendarmenmarkt (BK, Berlin Konzerthaus) and the Philharmonie Berlin (BP). The orchestra performed "Egmont Ouverture" by L. v. Beethoven, a typical Wiener Klassik piece, during the final rehearsal before the first concert as well as the short dress rehearsal of the second concert. The same excerpt, containing a 15 s crescendo, was recorded simultaneously in two receiver positions in both rehearsals with the halls unoccupied. The first position in BK was at 6 m distance from the conductor and 2 m off-center to the right (row 5, seat 15, BK1). The second position was on the first balcony at 22 m distance (seat 9, BK6). In BP, the first position (row 6, seat 19) was approximately at the same physical distance as in BK. As BP does not have a balcony, the corresponding position was at 27 m distance in the rear stalls. The receiver positions are shown in Fig. 1. These are the same halls and fairly similar receiver positions as used in the study investigating musical dynamics [6], but not quite as in the thematically more related study investigating emotional impact [5]. The influence of the different receiver distances will be discussed later. The orchestra was seated along the German arrangement with 1st and 2nd violins on opposite sides, double basses to the far left, and celli and violas to the center and far right, respectively.

Recording and Reproduction

The setup for the orchestra recordings was a four-channel head-like recording system with a pair of DPA 4060 miniature microphones for the front channels, and a Zoom H4N built-in stereo pair for the rear channels. More details on the recording system are given in Ref. [7]. Two calibrated recording systems were used simultaneously in both concert halls, thus allowing a matched A/B-comparison afterwards. The reproduction was accomplished with four loudspeakers (Genelec model 8020B) positioned at $\pm 45°$ angles for the front and $\pm 135°$ for the rear speakers. The radius of the loudspeakers from the listening position was 1.4 m, and the sound level of each loudspeaker was calibrated. To block sound coming from the left loudspeaker to reach the right ear (i.e. cross-talk cancellation) a 4 cm absorptive and isolating panel was placed directly in front of the listener's head. The level for the listening test was adjusted to L_{Aeq} of 78 dB or L_{AFmax} of 86 dB for the loudest stimulus. The presented sound levels were approx. 6 dB higher than in-situ due to calibration issues. However, the difference was constant for all stimuli. The laboratory calibration was done with a class 1 SPL-meter and a human head with two DPA 4060 microphones attached next to the ear for further processing of loudness and levels (calculations in PsySound 3 [8] using the DLM loudness model). A post-hoc analysis of the sound pressure levels showed that BK1 is around 2 dB stronger than BP1 and BK6, which again are another 2 dB stronger than BP6. The loudness normalization for all stimuli led to an average loudness of 26 sone or an L_{Aeq} of 76 dB.

Experiment

Procedure

The listening task was designed for quick comparison between the four stimuli by paired comparison, following the approach in recent studies at the Aalto University [5], [6]. Of the pair of stimuli, the listener was instructed to choose the one which felt to have more impact. The subjects could also indicate a tied response for "no difference". The listening surface was provided on a touch screen (Apple iPad, connected to a MacBook Pro). The

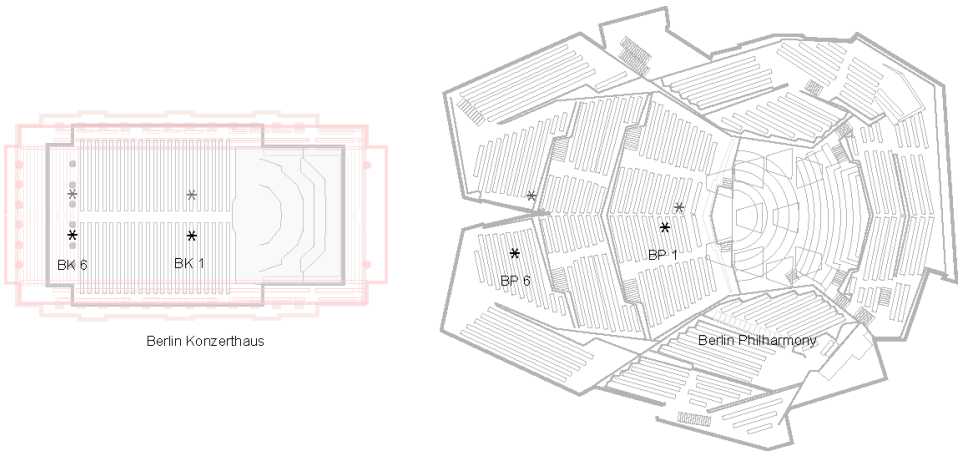

Figure 1: Groundplan of the Berlin Konzerthaus (left) and Philharmonie (right) with each two receivers. Closeby receiver position marked in gray were used in a related study [6].

audio was played back from the computer via Motu 16A audio interface, which feeds the signals to the loudspeakers. The stimuli could be switched seamlessly as the compared recordings were played synchronously. The paired comparison was a full-rank design with two repetitions, i.e. each pair was presented twice in random order.

A brief training session was conducted before the experiment with four stimuli of which two were used in the actual test and the other two were from the same halls but different seats. A maximum of six training comparisons were offered though most participants felt comfortable with the task after four pairs. The verbal test instructions informed the participant about the proper seating position and explained the objective of the study on the experienced impact. The brief description defined the term "impact" as having more influence, being more interesting, or more effective on oneself. The subjects were recommended to initially listen both completely instead of switching quickly back and forth. After the training, the correct understanding of the task was confirmed in a brief discussion. During the test, the subjects were also asked to write on a paper the principal differences driving the decision for certain stimuli for each pair. The test session was concluded with a discussion on the collected criteria in order to resolve ambiguous answers and to narrow down the vocabulary. For instance, a mention of spatial impression was defined further to an indication of envelopment, perceptual distance, or source width. Other general remarks on the test were collected as well. Consent was given by the participant to insure accordance to ethical regulations.

The experiment was conducted in two test periods. In the first test, the stimuli were presented at the original relative sound levels. 18 subjects (2 female, 16 male) participated in the first test. All were Aalto University staff, and half of them could be considered skilled listeners due to their work in room-acoustic or signal processing research groups. The average age of the subjects was 31 years. In addition to the paired comparison between four stimuli with two repetitions, a control pair with two identical stimuli was hidden in the test sequence. Hence, there were 13 comparisons in the test. All subjects reported the pair without difference as a tied comparison, and the control pairs were omitted from the subsequent analysis. The average duration for the paired comparison was 13.6 minutes.

In the second test period, conducted on a different day, the presented stimuli were matched with regard to overall loudness. As this process rendered the task more difficult, relatively more experienced listeners were invited to contribute to the experiment. A total of 10 subjects (1 female, 9 male) participated in this listening test, 7 subjects from the acoustics or signal processing groups and three untrained listeners. The hidden identical pair was removed but three other control pairs included to compare directly between the original and the loudness matched version of each BK1 and BP6 as well as a re-test of the two fairly similar stimuli BK6 vs. BP6. The 50% greater number of comparisons extended the test duration to 18.5 minutes. Most of the subjects were already familiar with the task from the first experiment. However, they were not told how the stimuli differed from the first experiment.

The answers were encoded into a choice matrix and analyzed in Matlab with the Bradley-Terry-Luce (BTL) model for estimating the underlying choice probabilities and variances [9]. For reference, the analysis was reproduced in R environment with the CompR package [1]. The results from Matlab and R analyses were in agreement.

Results

Listening Test

The choice probabilities for impact by the acoustics according to the BTL model is shown in Fig. 2. First, we consider the results of the experiment with original relative sound levels. Position BK1 has a clearly higher perceived impact than the other positions. The error bars indicate ±1 standard error around the mean BTL value.

[1] M. Semenou, available on CRAN ("Comprehensive R Archive Network", cran.r-project.org)

Statistically significant differences ($p<0.05$) appear between all stimuli except for BK6 and BP6 ($p=0.11$). A more detailed inspection of the choice matrix (see Table 1, left) reveals that BK6 is chosen over BP6 33.5 times out of 36. A separate analysis for this isolated pair suggests a highly significant difference ($p<0.01$). It can be concluded that first, front positions produced higher impact than further positions in both halls, and second, positions in Konzerthaus yielded in general a higher impact with regard to the respective positions in Philharmonie. This result follows expectedly the sound pressure levels observed in the compared positions.

The results from the loudness-matched comparison is shown in lighter shade in Fig. 2. Despite the equalized overall loudness, the overall rank order as well as the significant differences are the same as in the first test. In general, the magnitude differences between BTL probabilities are smaller. Also, the choice matrix for the second test shows a tied result between BK6 and BP6 (see Table 1, right), and the significance is further reduced ($p=0.34$). Thus, matched loudness equalized the impact in the further positions but not in the front positions.

For the additional control pairs comparing the original with the loudness matched version of each BK1 and BP6 and original versions of BK6/BP6 the probabilities were in the order of 0.9 for the louder stimulus to 0.1 (not shown here).

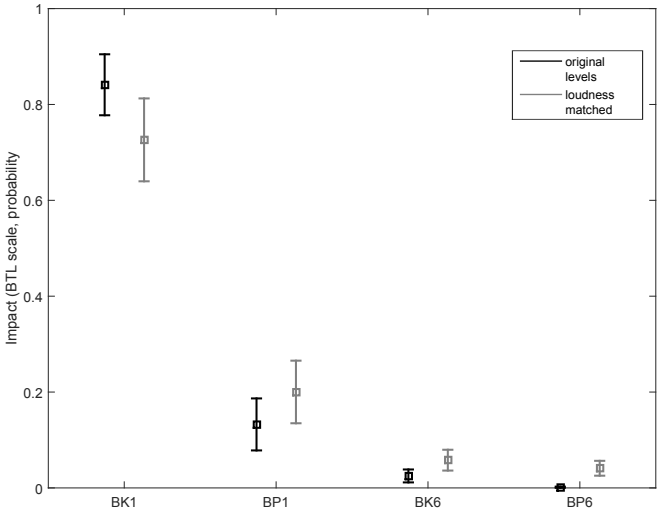

Figure 2: Probability of impact rating for stimuli in original condition (black) and loudness matched (grey). Error bars represent ±1 SE.

Vocabulary Profiling

For each pair comparison, one or more descriptive adjectives suggest the reason for the particular choice. This data also reveals the perceptual dimensions which the subjects used in evaluating the stimuli. The refined adjectives collected after the discussion with each participant were manually categorized into groups of similar attributes. For the first test, almost 90% of the adjectives can be grouped into seven attribute groups. These groups were further combined into two general categories related to strength or dynamics, and spatial properties. This procedure is shown with the respective results in Table 2. For the first test with original relative levels, 25% of the decisions were based on strength attributes. The second most frequent group is proximity, followed by spatial attributes, bass, and clarity. In essence, approximate one third of the decisions were based on dynamics or sound level differences. For the second test with matched loudness, the prominent attributes groups are altered. The strength/dynamic/crescendo group account combined for only 11% of the choices. In contrast, spatial attributes are more frequent, and proximity appeared as the most frequent single attribute.

It can be concluded that when the loudness is normalized, differences in strength (and closely related attributes such as dynamics) are less decisive. Other cues then overweigh the decision, but not sufficiently to alter the order of stimuli regarding impact.

Table 2: Vocabulary Profiling: Adjectives, collected from the participants and grouped to attributes, driving the impact ratings. Percentages greater than 15 are highlighted in bold.

Attribute Groups	Adjective Count Percentage	
	Test1 (original)	Test2 (loudness matched)
dynamic+crescendo	9%	3%
envelopment+spaciousness	10%	9%
clarity	8%	13%
bass	12%	**15%**
width	5%	13%
proximity	**18%**	**22%**
strength	**25%**	8%
timbre		4%
SUM	88%	88%
other	12%	12%
env+spaciousness+width	**15%**	**22%**
strength+dyn.+cresc.	**35%**	11%

Discussion

The physical distances to the receiver positions were not equal between the two halls, and the absolute judgment for these halls would not be entirely justified. However, similar hall areas have been used (e.g. R1 is situated in Row 6 for both halls) and the findings are in line with the study using the artificial orchestra [5], where distances between pairs of receivers were exactly the same, and a significant difference was found between these halls. In other words, the 3 and 5 meters offset in distance between Konzerthaus and Philharmonie receivers here are probably not changing the overall outcome.

An earlier research [5],[6] utilized convolutions of anechoic orchestra recordings and spatial room impulse responses measured from the halls. Although the present

Table 1: Summed choise matrices and row sums for the two tests

Test1 (original stimuli)

	BK1	BP1	BK6	BP6	**SUM**
BK1	0	31.5	36	36	**103.5**
BP1	4.5	0	30.5	36	**71**
BK6	0	5.5	0	33.5	**39**
BP6	0	0	2.5	0	**2.5**

Test2 (avg. loudness matched)

	BK1	BP1	BK6	BP6	**SUM**
BK1	0	15	18.5	20	**53.5**
BP1	5	0	14	18	**37**
BK6	1.5	6	0	10	**17.5**
BP6	0	2	10	0	**12**

study employs recordings from a live orchestra, the results are in agreement with the previous findings. This underlines that both approaches are likely valid for studying subjective impact by concert hall acoustics. The earlier study [5] proposed the early lateral energy as the main cause for the enhanced impact. The present findings give no reason to draw a different conclusion, even though the analysis of objective room acoustic parameters is not included here. The studies also share another similarity, as the loudness-related attributes account up to one third of the comparison decisions [6].

For the more remote seats, the level is more important than for close seats, emphasized in the loudness matching of the second test and also observed in the stimulus-dependent vocabulary profiling of the comparative study [5]. The comparison between differently distanced receivers was included and, as expected, the closer seats have more impact. Also, the loudness matching provides to some extend a hint that overall trends might not change even when the important level/loudness-cue is missing. Unfortunately, the overall listening level was set 6 dB too high due to a mistake in the re-calibration only discovered afterwards, i.e. all stimuli were reproduced louder than in-situ. As louder audio signals are often preferred this could have altered the magnitudes of the rating. However, the overall trend and differences are likely not affected notably. The subjects considered the presentation level acceptable, although one subject preferred to listen to the stimuli at a level reduced by 3 dB.

The findings tackle an interesting question of whether a real orchestra can compensate for less favorable room-acoustic conditions by e.g. playing louder. The present results do not exclude the possibility for an orchestra to adapt to the acoustics, but here the orchestra did not demonstrate a sufficient adaptation to produce equal impact in different halls.

Summary

This study with two high-profile concert halls indicated that positions near to the orchestra, and particularly those in Berlin Konzerthaus, produce prominent subjective impact. When eliminating the level differences between stimuli the order is not affected since other perceptual cues take over: Impact as a measure of the emotional effect on the listener is affected greatly but not solely by level. Using a real orchestra similar results are found as in a study conducted with an artificial orchestra suggesting the validity of both approaches.

Acknowledgment

The research work received funding from the European Commission within the ITN Marie Curie Action project BATWOMAN under the 7th Framework Program [EC grant agreement no. 605867] and the Academy of Finland [project no. 289300].

References

[1] A. Kircher, *Neue Hall- und Thonkunst oder Mechanische Gehaim-Verbindung der Kunst und Natur durch Stimme und Hall-Wissenschafft gestifftet.* Heylen, reprint 1983 ed., 1685.

[2] H. Lehman, P.; Wilkens, "Zusammenhang subjektiver Beurteilung von Konzertsälen mit raumakustischen Kriterien," *Acustica*, vol. 45, pp. 256–268, 1980.

[3] T. Lokki, "Tasting music like wine: Sensory evaluation of concert halls," *Physics Today*, vol. 67, no. 1, pp. 27–32, 2014.

[4] M. S. Lawless and M. C. Vigeant, "Investigating the emotional response to room and acoustics: A and functional magnetic resonance," *Journal of the Acoustical Society of America*, vol. 138, no. 4, pp. EL417–EL423, 2015.

[5] J. Pätynen and T. Lokki, "Concert halls with strong and lateral sound increase the emotional impact of orchestra music," *The Journal of the Acoustical Society of America*, vol. 139, no. 3, pp. 1214–1224, 2016.

[6] J. Pätynen and T. Lokki, "Perception of music dynamics in concert halls," *The Journal of the Acoustical Society of America*, vol. (conditionally accepted), 2015.

[7] W. Lachenmayr, "Höreindruck und raumakustische Kriterien in Konzertsälen," Master's thesis, Institut für Komposition und Elektroakustik. Universität für Musik und darstellende Kunst Wien, 2012.

[8] D. Cabrera, S. Ferguson, and E. Schubert, "PsySound3: An integrated environment for the analys is of sound recordings," in *Acoustics 2008: Proceedings of the Australian Acoustical Society Conference*, 2008.

[9] F. Wickelmaier and C. Schmidt, "A matlab function to estimate choice model and parameters from paired-comparison data," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 1, pp. 29–40, 2004.