24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27 JULY 2017
LONDON CALLING

CONCERT HALL ACOUSTICS, ONLINE RATING AND BERANEK’S DATA COLLECTION

Magne Skålevik

AKUTEK and Brekke & Strand
email: msk@brekkestrand.no

One of Leo Beranek’s special contributions in concert hall acoustics was his work on rank ordering of concert halls, based on qualified assessors' preference for their acoustics. In his papers as well as in his book Concert Halls and Opera Houses, Beranek presented the rank orderings together with objective data from the halls included in the ranking. By making this collection of subjective and objective data available for the scientific society, it was possible to investigate the degree to which preference could be explained by the physical and acoustical properties of the halls. This author has tried to take up this legacy by extending the collection of halls, in particular by including halls built after the publication of Beranek’s rank ordering. For this purpose, the Online Concert Hall Acoustics Rating Survey has been launched, inviting all concert hall goers to submit their rating at https://no.surveymonkey.com/r/MMFMZ5W. Indeed, in his last paper in JASA, April 2016, Beranek referred to the preliminary results. In this paper, an updated report from the survey will be presented, together with an analysis of how the current ratings correlate with Beranek’s rank orderings. An example of how to combine data from Ranking, Rating and objective data in Beranek’s data collection, to predict rating of future halls, is presented.

Keywords: Concert hall acoustics, rank orderings, rating, parameters, data collection

1. Introduction

One of Leo Beranek’s special contributions in concert hall acoustics was his work on rank ordering of concert halls[1], based on qualified assessors' preference for their acoustics. In his papers as well as in his book Concert Halls and Opera Houses[2], Beranek presented the rank orderings together with objective data from the halls included in the ranking. By making this collection of subjective and objective data available for the scientific society, it was possible to investigate the degree to which preference could be explained by the physical and acoustical properties of the halls. This author has tried to take up this legacy by extending the collection of halls, in particular by including halls built after the publication of Beranek’s rank ordering. For this purpose, the Online Concert Hall Acoustics Rating Survey was launched[3], inviting all concert hall goers to submit their rating at https://no.surveymonkey.com/r/MMFMZ5W. Indeed, in his last paper[4] (April 2016), Beranek used the preliminary results (N=59 voters had given their votes), to revise and include more recent halls in his rank orderings. In the present paper, an updated report from the survey, at N=84, will be presented, together with an analysis of how the current ratings correlate with Beranek’s rank orderings. Moreover, a discussion of the potential for predicting preference of a planned hall from objective hall properties, is included. In particular, attention is drawn to what happens when comparing a combination several parameters with preference, instead of one parameter at the time.

2. Online Concert Hall Acoustics Rating Survey[3]

In the Online Concert Hall Acoustics Rating Survey, concert hall visitors were presented with a list of concert halls, and asked the following question:

"In those halls in the list where you have attended a concert with a symphony orchestra once or more, how do you rate the acoustics there?"

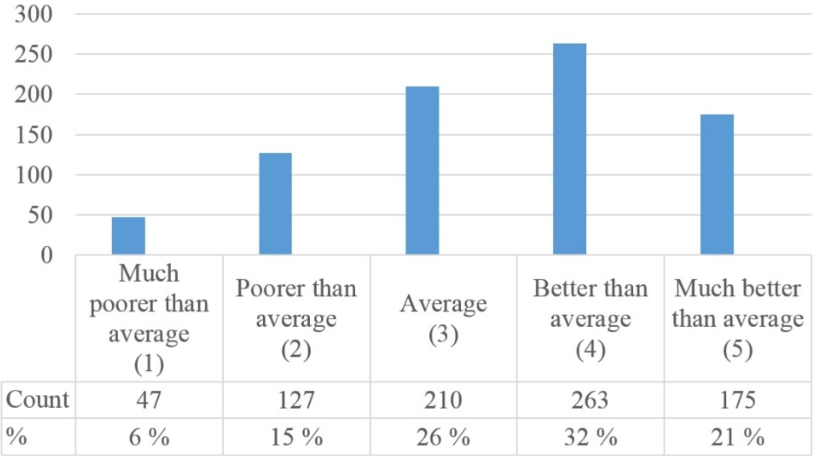
The rating was given by checking one of five alternative judgements

1. Much poorer than average
2. Poorer than average
3. Average
4. Better than average
5. Much better than average

In the analysis of the survey, a vote would be an integer number on a scale from 1 to 5 according to the list, e.g. the vote "Average" would have a value equal to 3. A voter can give a vote to any of the halls on the list, leave the rest blank, and finally click "complete".

In our analysis of results, the rating of a hall is calculated by the sum of values the hall has received from voters, divided by the number of votes. This would yield a decimal value on a rating scale from 1.0 to 5.0. By number of votes, we mean number of ratings submitted. For the ratings to be useful for research purposes, i.e. to identify significant objective differences between good halls and bad halls, it is important to have many ratings of halls in both categories.

As of 2017-03-28, a total of 84 voters had visited the survey, submitting 822 votes, distributed as shown numerically and graphically in Figure 1. The average rating value was 3.48, corresponding to the border between 'Average' and 'Better than average'. We note that 68% of the votes were 'Average' or 'Better than average'.



	Much poorer than average (1)	Poorer than average (2)	Average (3)	Better than average (4)	Much better than average (5)
Count	47	127	210	263	175
%	6 %	15 %	26 %	32 %	21 %

Figure 1: Distribution of the 822 votes.

The fact that the concerts hall in the survey are judged better than what the voters judge as 'Average' may not be surprising. After all, many of the halls are included in the survey because they are being visited by many concert goers, or they are famous for being exceptionally good. In contrast, any existing concert hall with bad acoustics would have a lower probability of being visited, implying a lower probability of receiving a vote in this survey. Have high-ranked halls actually been rated more frequently than low-ranked halls in this survey? The answer is yes, but only a weak tendency is seen. Figure 2 shows that as a trend, the number of votes decreases by 0.12 per rank order step, i.e. 1.2 votes per 10 ranking steps. If we look at those halls that have received many votes, e.g. more than 15 votes, the tendency is somewhat stronger, 1.9 votes per 10 ranking steps.

As a conclusion, we note that there is a slight tendency toward more good halls rated than bad halls rated, but not more than natural from probability as discussed above.

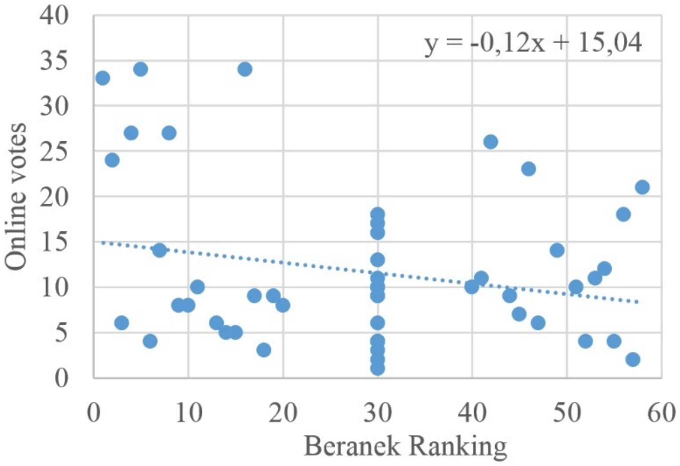


Figure 2: Number of ratings vs ranking.

3. Comparing Rating results with Beranek Ranking

It seemed natural to compare the results from the online rating survey with Beranek's ranking. A total of 52 halls among the 58 halls in Beranek's ranking received at least one rating vote in the online survey. In Figure 3, online rating results are plotted against ranking for these 52 halls. Pearson squared correlation[1], a value between 0 and 1, between the two data sets is r^2=0.58. A common interpretation of r^2 would be: "58% of the variation in rating of halls is explained by their ranking". This value should not be confused with the correlation coefficient, a value between -1 and 1, commonly used in regression analysis, which in this case is r=0.76.

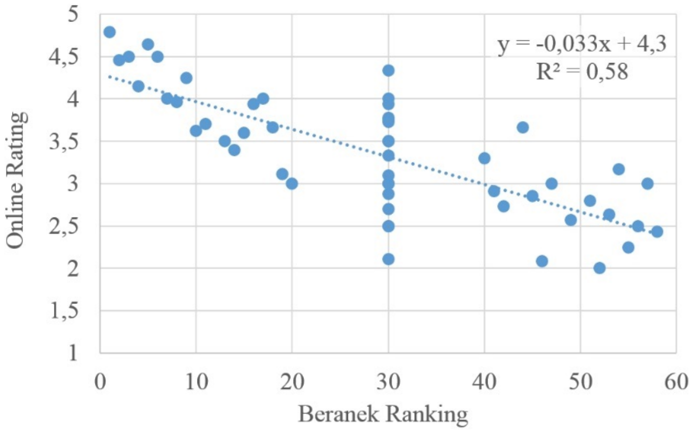


Figure 3: Online rating results of 52 halls plotted against Beranek's ranking

From the plot in Figure 3, we clearly see the three groups of halls in the ranking, as commented by Beranek. To the left is the top 20 group, ranked 1-20. To the right the bottom ranked group, ranked 40-58. The halls in the middle group, 21-39 in the rank ordering, were just listed in alphabetic order, because Beranek concluded they were not distinguishable. However, Beranek made clear that the halls in the middle group was judged poorer than the top 20, but better than the bottom ranked group. Instead of using the alphabetic order to rank the middle group from 21 to 39, this author have chosen to assign all halls in this group with rank order 30, since this is the mean value of the group of integers 21 through 39. From Figure, it is evident that online voters judge some of the halls in the middle group to be better than some of the halls in top 20. Likewise, online voters judge some of the halls in the middle group to be poorer than some of the halls in the bottom group.

---

[1] In this paper denoted r^2 or RSQ

In Beranek's ranking, reliability was strengthened by the fact that no halls had less than six qualified raters[4]. In similar manner, more reliable results, i.e. less uncertainty, can be achieved in the online rating by requiring a sufficient number of voters. Figure 4 is a plot of the 16 halls having received more than 12 votes. Correlation between rating and ranking of the same 16 halls is r^2=0.79, which is considerably more than the correlation seen in Figure 3. By increasing the requirement to more than 18 votes, the number of halls meeting the requirement is reduced to 9, and correlation increases to 0.93. However, a higher correlation between rating and ranking comes with the price of fewer rated halls, and is not always adequate. These are priorities depend on purpose.

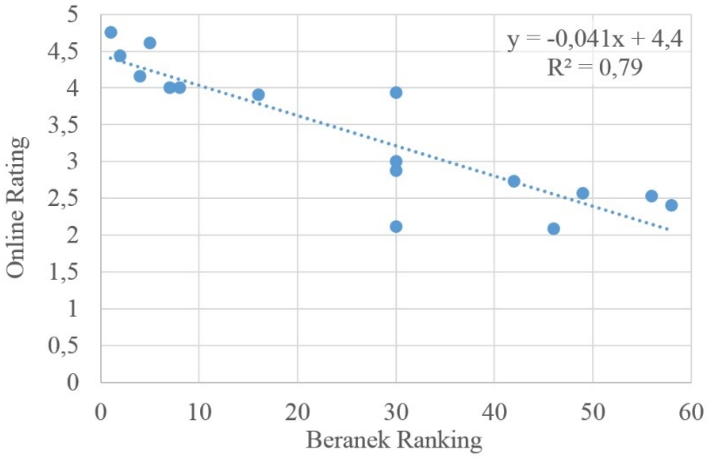


Figure 4: A plot of the halls in Figure 3 that have received more than 12 votes

4. Using Beranek ranking to estimate optimum geometrical and acoustical properties

Beranek combined the subjective data of the ranking with objective data in order to investigate the relationship between the geometrical and acoustical properties of halls, and their ranking. Mathematically, such relationship can be expressed in terms of correlation between subjective and objective data. We shall refer to the objective data, such as geometrical properties, reverberation time, and so on, as parameters. If high correlation is found between one parameter or a combination of parameters, we can estimate the optimum values of the parameters. These optimum values can be used as design criteria in planning of new concert halls, or corrections of existing halls.

There are several parameters measured in existing halls exhibiting high correlation with ranking and rating. However, most of these parameters are unpredictable in the planning phase. That is, existing prediction tools are not reliable, i.e. their predictions of the parameters do not correlate well with ranking and rating. One important exception comes to rescue – RT can be predicted with good accuracy from absorption coefficients, and actual absorption coefficients are quite predictable from knowledge about materials and laboratory testing of material properties. This author has investigated different methods for prediction of concert hall parameters, concluding that the least uncertain predictions come from so-called *TVr*-predictors, based on Barron Revised Theory (BRT) [8].

With input from reverberation time *T*, room volume *V*, and average source-receiver distance *r*, *TVr* predictors can predict the *Strength* parameters *G*, G_{early}, G_{late}, and the early-to-late balance, i.e. *Clarity*, *C* or C_{80}. 'r' is estimated from *Length* and *Width*. Binaural parameters like lateral fraction *LF* and *IACC* cannot be estimated from BRT, and do not come as reliable outputs from state-of-the-art computer-simulation tools. However, there are geometrical parameters that largely determines the probability for spatial impression is likely to occur. In his pioneer paper on spatial impression, Marshall identified the height-to-width ratio *H/W* of the hall as the main critical parameter. E.g., if *H/W*>0.5 a listener in the middle of the stalls in a rectangular hall, would receive lateral reflections before reflections from the ceiling arrived. As we know, lateral reflections increase *LF* and reduce *IACC*, while ceiling reflections do the opposite. Width W has been pointed at as an important parameter in itself. The influence on loudness from reflections via side wall surfaces that are unoccupied by sound absorbing audience, in a hall that is not too wide, was noted already by Sabine in 1900[10]. As a hall gets wider, above a critical limit, the risk of hearing echoes increases. Further it has been suggested that so-called cue-ball reflections from the underside of side balconies serve to fight the seat dip effect, but then W must be within limits. All these arguments substantiate that W can be a critical parameter.

Table 1 presents statistical key figures for 7 parameters, chosen by three criteria: 1) They can be predicted in concert hall planning, 2) They are relevant, ref. discussion above, 3) their values correlate with rating values, and RSQ is >>0. A description of the rows follows.

The header row is the parameters, T and G in mid-frequency octaves 500 and 1000Hz, G, G_{125} is *Strength* in the 125Hz octave, C is C_{80}, the early-to-late balance or *Clarity*, G_L is G_{late}, the *Strength* of late sound, $G_{late},= G\text{-}10\cdot lg(1+10^{C/10})$, where G and C are calculated from *TVr*-predictors, with input from T, T_{125}, and V in Beranek's data collection[2], together with H and W as defined by Beranek. W is hall width, H is hall height above stalls, and H/W is height-to-width ratio, i.e. the cross-section parameter and the parameter that governs lateral fraction in early sound, LF.

Second row, 'm' is the average value of each parameter in the 52 halls among Beranek's 58 ranked halls that have received votes in the online rating. This group of halls is for the purpose of this paper denoted B52. Third row, 's' is the standard deviation around m. Fourth row, 'Po' is the optimum value of each parameter, i.e. the parameter value that from Beranek's ranking is estimated to contribute as much as possible to the highest ranking. E.g., in case of T, Po is calculated from

$$Po(T) = (T_1\cdot e^{-r1}+ T_2\cdot e^{-r2}+\ldots+ T_n\cdot e^{-r52})/(e^{-r1}+ e^{-r2}+\ldots+ e^{-r52}), \tag{1}$$

where T_1, $T_2,\ldots,T_{52}$ are the T-values of the 52 halls, and $r1,r2,\ldots,r52$, are the corresponding Rank orders by Beranek.

Table 1: Statistical key figures for 8 parameters

Parameter	*T*	T_{125}	*C*	*G*	G_{125}	G_L	*H/W*	*W*
m = average in B52	1,82	2,21	-0,1	1,6	2,4	-1,3	0,64	30
s = standard deviation in B52	0,26	0,40	0,8	1,7	1,7	1,8	0,23	8,1
Po = optimum value in B52	1,97	2,89	-0,7	3,4	5,1	0,8	0,86	21

5. An attempt to use deviations from optimum to predict rating of future halls

It would be useful to predict ratings of future halls. In the following, we describe an attempt to do so by considering the deviation from the optimum parameter values in Table 1. We shall consider a large deviation from optimum to forecast a low rating, and a small, or zero, deviation from optimum to forecast a high rating. For this purpose, *w* shall use the normalized deviation from optimum, calculated by $D = | P – Po |/s$, where P is the parameter value for a given hall, *Po* and *s* is given in Table 1.

Not all parameters are equally important, or critical, and we shall put more weight on deviations in the parameters that are found to be more critical, and less weight on deviations in parameters that are found to be less critical. For this purpose, we define the weighting factor $w = RSQ(D ; Rating)$, i.e. the square of correlation between deviations D of a selection of halls and the rating values of the same halls. By using a selection of halls that have received many votes, we ensure a reliable weighting of each parameter, and calculate the weighted deviation for a combination of n parameters $P_1, P_2,\ldots P_n$,

$$D_w = (D_1\cdot w_1+D_2\cdot w_2+\ldots+D_n\cdot w_n)/(w_1+w_2+\ldots+w_n). \tag{2}$$

Finally, to optimize the method, we use an iteration process, i.e. trial and error, to establish exactly which selection of parameters, and how many votes should be required for the method to be as reliable as possible with our available data. In this process we aim for the optimum correlation between weighted deviation and rating, $RSQ(D_w ; Rating)$. The optimum was reached when using the 7 parameters T, T_{125}, C, G_{125}, G_L , H/W and W, while requiring more than 12 votes per hall in the selection of halls. At the present, a total of 16 halls have received more than 12 votes, and we denote this group R16.

Table 2 presents the average deviations D of the halls in R16, and weighting factors w for the 7 parameters. D=100% would mean that average deviation is s, i.e. 1 standard deviation from optimum. Since D is calculated from the absolute deviation, positive and negative deviation contributes equally to D. Note that *RSQ* in T is 0.70, which can be interpreted as ‘deviations in T is able to explain 70% of the deviations in rating of these 16 halls’. 30% remains to be explained.

Table 2: Po from Table 1, average deviations D in R16, and weighting factors w, 7 parameters

Parameter	T	T_{125}	C	G_{125}	G_L	H/W	W
Po = optimum value in B52	1,97	2,89	-0,7	5,1	0,8	0,86	21
$D = \lvert P\text{-}Po\rvert/s$, average in R16	78 %	162 %	90 %	170 %	138 %	107 %	117 %
$w = RSQ(D ; Rating)$ in R16	0,70	0,37	0,51	0,60	0,69	0,23	0,30

Regression analysis of weighted deviations and rating is used to derive a trend formula that can work as a predictor, or a forecast, of rating of future halls. The trend formula is given in (3) and in Figure 5,

$$Trend\ (Online\ Rating) = 4.9 - 1.2 \cdot D_w \tag{3}$$

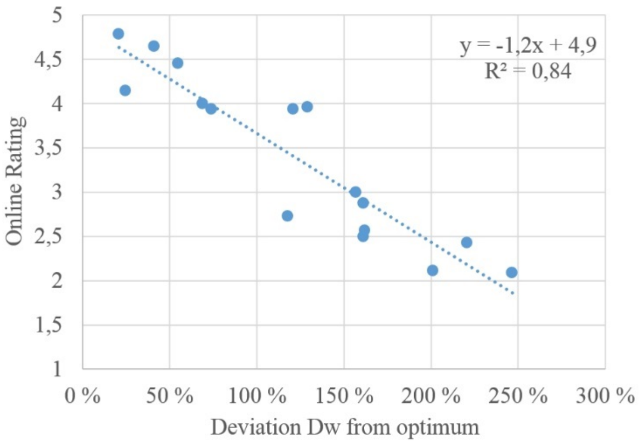


Figure 5: Regression analysis of the halls in group R16

Comments to the result above: Pearson correlation, i.e. squared correlation in the regression analysis is high, $r^2 = 0.84$. This can be interpreted as ‘weighted deviation in 5 parameters is able to explain 84% of the deviations in rating of these halls’. Thus 16% remains unexplained by parameters, which is considerably less than the 30% with parameter T alone. The remaining explanation may lie in factors like resident orchestra, aesthetics, overall experience, visitor satisfaction, etc. When used as a predictor or forecast, these 16% could be interpreted as the uncertainty in the prediction. This uncertainty becomes immediately evident by observing the deviation from the trendline in the interval D_w=100%-150%. Further results from the testing of the predictor function is presented in the next section.

Note that the *Trend* function in (3) is depending on the following:

- Beranek’s ranking of concert halls
- Online rating data
- Online rating submitted for 52 of the halls in Beranek’s ranking, defining the group B52

- Objective data for the halls B52 in Beranek's data collection
- Iteration process based on regression analysis for optimizing the set of parameters, eventually 7 parameters, and optimizing the requirement for number of votes per hall
- Sub-selection of halls having received more than 12 votes, i.e. currently 16 halls, and their rating values
- Beranek's data for the R16 halls: *Volume, Height, Width, Length, T* and T_{125}

As can be seen by comparing Table 2 with Table 1, G has been excluded, but not without doubt. However, the exclusion of G caused r^2 in the regression analysis, Figure 2, to increase slightly, and theoretically G is very much represented by C and G_{125} due to the mathematical dependency $G = G_{late} + 10 \cdot lg(1+10^{C/10})$. Such inter-dependencies are suspected to still exist within the set of 7 parameters. Thus, in future work, one should try to arrive at a number of parameters that equals the size of the input vector, in this case 6, i.e. *Volume, Height, Width, Length, T* and T_{125}. This concern is, basically, the usual concern about the use of orthogonal input variables in linear analysis. In this respect, keep in mind that the method described above distinguishes from multiple linear regression in a crucial aspect: It does not assume a linear relationship between each parameter and the subjective judgement. Instead, it assumes a linear relationship between the absolute deviation from optimum and the subjective judgement.

It should be mentioned that a reduction to only 5 parameters T, C, G_L, H/W and W, would cause an even higher r^2 (=0.87) in the regression analysis. However, this depends very much on the group of halls (R16) having currently received the required number of votes (>12). Since this will change as more votes are submitted, the observation should not affect the selection of parameters.

The inclusion of more than one 125Hz octave parameter deserves a comment: While G_{125} accounts for warmth, T_{125} influence on bass attenuation through the auditorium by $-0.176/ T_{125}$ *(dB/m)*[11].

6. Testing the predictor

In order to test the ability of the *Trend* function above to be a predictor that can forecast ratings of future halls, we compare the *Trend* values with actual rating values of existing halls. Results of this comparison, including calculated by the formula *Error = Trend – Rating*, is given in Table 3. Positive values in Error indicates that the *Trend* would forecast a higher rating that the actual rating, and negative values in *Error* indicate that *Trend* would forecast a lower rating than the actual rating of the hall, given the hall's weighted deviation from optimum parameter values, *Po* in Table 2.

Table 3: Data for the 16 halls in group R16, ordered by their weighted deviation D_w from optimum in the selected 7 parameters; Beranek's Rank order (*B Rank*), *Rating* and number of *Votes* from online survey, *Trend* (forecast), and *Error=Trend–Rating*. *Rating* and *Dw* is plotted in Figure 5.

	B Rank	*Rating*	*Votes*	*Dw*	*Trend*	*Error*
Vienna Grosser Musikverinsaal	1	4,8	33	21 %	4,6	-0,2
Berlin Konzerthaus	4	4,1	27	25 %	4,5	0,4
Amsterdam Concertgebouw	5	4,6	34	41 %	4,4	-0,3
Boston Symphony Hall	2	4,5	24	55 %	4,2	-0,3
Zurich Grosser Tonhalsaal	7	4,0	14	69 %	4,0	0,0
Vienna Konzerthaus	30	3,9	17	74 %	4,0	0,0
New York, Avery Fisher Hall	42	2,7	26	118 %	3,5	0,7
Berlin Philharmonie	16	3,9	34	121 %	3,4	-0,5
New York Carnegie Hall	8	4,0	27	129 %	3,3	-0,6
Washington, DC, JFK Conc. Hall	30	3,0	13	157 %	3,0	0,0
London, Barbican Concert Hall	56	2,5	18	161 %	3,0	0,5
Chicago, Orchestra Hall	30	2,9	16	161 %	3,0	0,1

Paris, Salle Pleyel	49	2,6	14	162 %	2,9	0,4
Munich, Philharmonie Am Gasteig	30	2,1	18	201 %	2,5	0,4
London Royal, Albert Hall	58	2,4	21	221 %	2,3	-0,2
London Royal Festival Hall	46	2,1	23	247 %	2,0	-0,1

The root of mean of squares of errors is *RMS(Error)=0.30*. We note that Avery Fisher Hall would have been overestimated by 0.7 if *Trend* was being used as a forecast of its rating. In contrast, Berlin Philharmonie and Carnegie Hall are being underestimated by -0.5 and -0.6, respectively. In further work, all the errors should be analysed to investigate whether they are random errors or systematic errors due to factors that have been overlooked, and if the latter is the case, whether these factors can be accounted for in predictions.

Note 1: *Trend* accounts solely for variations in physical and acoustical conditions, in particular those that influence on the 7 parameters in Table 2, i.e. T, T_{125}, C, G_{125}, G_L, H/W and W.

Note 2: Ranking, rating and predicted rating of concert hall acoustics presented in this paper are all meant for scientific purpose only, and are particularly not intended to be used for advertising or in any other way influence on concert goers' choices.

REFERENCES

1 Beranek, L.L., *Subjective rank-orderings and acoustical measurements for fifty-eight concert halls*, Acta Acust. Acust. 89, 494–508 (2003).

2 Beranek, L.L., *Concert Halls and Opera Houses (Springer, New York, 2004).*

3 M. Skålevik, *Certainties and uncertainties from using a selection of data to predict concert hall preference*, Build. Acoust. 20, 335–350 (2013).

4 Beranek, L.L., *Concert Hall Acoustics – Recent Findings*, J. Acoust. Soc. Am. 139 (4), April 2016.

5 Marshall, A.H., Barron, M., *"Spatial responsiveness in concert halls and the origins of spatial impression"*, App. Acoustics, 2000;62(2):91-108.

6 International Standard, *"ISO-3382 Acoustics – Measurement of Room Acoustic Parameters – Part 1, Performance Spaces"*, 1st edition (2009)

7 Beranek, L., *Concert Hall Acoustics 2008*, J. Audio Eng. Soc., Vol. 56, No. 7/8, 2008

8 M. Skålevik, M., *Can source broadening and listener envelopment be measured directly from a music performance in a concert hall?* Proc. Inst. of Ac. (IOA), Vol. 37. Pt.3 2015, paper 34 http://www.akutek.info/Papers/MS_Spaciousness-meter.pdf

9 Marshall, A.H, *"A note on the importance of room cross-section in concert halls"*, J. Sound and Vib., 1967;5:100-12.

10 Sabine, W.C. *Collected papers on acoustics*, Harvard University Press, Cambridge, Mass., 1922.

11 Barron,.M., *When is a concert hall too quiet?*, Proc., ICA-2007, http://www.akutek.info/Papers/MB_too_quiet_ICA2007.pdf