

# SUBJECTIVE LISTENING TESTS IN CONCERT HALLS: METHODOLOGY AND RESULTS

Eckhard Kahle & Jean-Pascal Jullien  
Ircam, 31, rue St-Merri, F-75004 Paris, email: kahle@ircam.fr

Published in:  
Proceedings of the Int. Congr. Acoustics, Trondheim, Norway, June 26 - 30, 1995

## Summary

The aim of this paper is to analyze the problems linked to listening tests in real halls *during* concerts and to devise a methodology overcoming these problems as much as possible. It will be shown that in this way it is possible to collect reliable data under real concert conditions.

## Introduction

Adapted measurement programs, mainly using maximum-length sequences (MLS) or sweeps, and fast hardware have made reliable measurements of room acoustical criteria in halls readily available. Due to a recent effort of the Concert Hall Research Group (e.g. [GBS93]) some convergence of the different measurement procedures is under way and reliable data on *occupied* halls is in sight. Availability of good data and the possibility to calculate a large number of different criteria has recently made researchers refocus on the aim of finding sets of *perceptually relevant* criteria ([Bar93], [Gri93],[Ber92]). Further attention will now have to focus on how to get reliable subjective data for identifying the most salient criteria.

Two basic methods are known for obtaining subjective data: laboratory experiments and real hall listening tests. Listening tests in real halls, using structured questionnaires, present a reliable way of collecting subjective data as all sources of unnaturalness are excluded. But the problems in analyzing the data are considerable: non-instantaneous comparisons, simultaneous variation of a large number of parameters, non-identical musical stimuli, etc. Those problems have prevented a widespread use of real-hall tests and the only two major studies directly using the questionnaire technique for the evaluation of different concert halls during live performances are the ones by Hawkes & Douglas, [HD71], and Barron, [Bar88].

After a series of laboratory experiments ([Lav89], [JKWW92]) that allowed to identify a set of criteria directly linked to perceptual factors and that helped to establish a structured questionnaire, the room acoustics laboratory at Ircam undertook a campaign of measurements and listening tests in European concert halls and operas. Subjects (mostly acousticians and/or musicians) responded to a structured questionnaire comprising 25 questions while listening to two concerts in each hall, changing seats between concerts as well as during the interval. Whenever possible, listeners attended two *identical* concerts. But nevertheless a number of problems are left: musical works differed between halls as well as for the different parts of a concert. As the number of seats tested per hall and the number of subjects were, on average, of the order of 8, each subject only listened to four seats – and each seat was only occupied by four subjects, additionally for varying musical works. These limitations may seem practical only, but are more or less unavoidable for this kind of listening tests.

## Theoretical considerations

Each response to a question,  $q_i$ ,<sup>1</sup> is a complex function of the following variables: *subject* ( $s_j$ ), *musical work* ( $o_m$ ) and the *place* ( $p_k$ ), i.e. the exact location within a hall:

$$q = FUN(s_j, o_m, p_k). \quad (1)$$

The different dependencies of the responses on the acoustics,  $q = \mathcal{F}(p_k)$ , the musical work,  $q = \mathcal{G}(o_m)$ , and the subject,  $q = \mathcal{H}(s_j)$  can be identified. Additionally, the acoustics at each place still has to be expressed in terms of a set of criteria  $c_k^i$ ,  $q = \mathcal{F}(p_k) \approx \mathcal{F}'(c_k^i)$ . A first approach might be to *factorise* the equation 1,

$$q = \mathcal{H}(s_j) + \mathcal{G}(o_m) + \mathcal{F}(p_k), \quad (2)$$

assuming linearity and absence of interdependence of the different influences. Neither strict linearity nor independence hold and it could be shown that at least the interdependence between the acoustics and the musical work is non-negligible ([WKJ93]), mainly due to the pronounced directivity of some musical instruments. On the other hand we know that we will have to eliminate the *specific* apprehension of a seat or a musical work by a subject — it is neither feasible nor desirable: what we want to get to is a *global* model valid for *all* listeners, or at least some kind of “average” listener. Introducing a term  $\mathcal{R}$  describing the resulting residual noise and introducing a new term,  $\mathcal{I}(i_{m,k})$ , describing the interaction between the musical work and the specific location in a hall, we get to the following final equation :

$$q = \mathcal{H}(s_j) + \mathcal{G}(o_m) + \mathcal{I}(i_{m,k}) + \mathcal{F}(p_k) + \mathcal{R}. \quad (3)$$

Do all of these terms have to be considered? For studying the correspondence between the subjective responses and the objective measures of the acoustical criteria at the individual locations in the halls,  $q = \mathcal{F}'(c_k^i)$ , all other terms have to be eliminated — or at least their relative importance has to be known.

## Importance of the different influences

A comparison of the standard deviations (square root of the variances) of the individual terms in equation 3 is given in table 1. We cannot go into detail here on how these standard deviations were calculated and just a few comments are given: for the subjects’ linear biases,  $\mathcal{H}(s_j)$ , see next section. For the influence of the musical work,  $\mathcal{G}(o_m)$ , only *in-hall* variations are considered, and only for those halls in which substantial changes in musical work occurred. The influence of the acoustics is excluded as identical sets of places are used. For the influence of the acoustics,  $\mathcal{F}(p_k)$ , two different values are given: the first,  $\mathcal{F}_{w/hall}(p_k)$ , corresponds to *within-hall* variations and the responses per place are averaged over musical works; the same subset of halls as for the term  $\mathcal{G}(o_m)$  is considered. The second,  $\mathcal{F}_{a/halls}(p_k)$ , includes *across-hall* variations and 90 places in nine halls are considered. The term  $\mathcal{R}$  was estimated from a specific test in which the same musical work was listened to four times. The term  $\mathcal{I}(i_{m,k})$  had to be inferred from the value of  $\mathcal{R}$  and the variance of  $\mathcal{I}(i_{m,k}) + \mathcal{R}$ , calculated as the distribution of the responses to the questionnaires around musical work and seat averages. The results are given as a standard deviation in perceptual units, averaged over all question (irrespective of the fact that the scale differed between questions). The units hence have to be considered as arbitrary units. More detail can be found in [Kah95].

influence	$\mathcal{H}(s_j)$	$\mathcal{G}(o_m)$	$\mathcal{I}(i_{m,k})$	$\mathcal{F}_{w/hall}(p_k)$	$\mathcal{F}_{a/halls}(p_k)$	$\mathcal{R}$
standard deviation (arbitrary units)	0.41	0.22	0.17	0.36	0.44	0.28

Table 1: *Standard deviations of the different influences on the responses to the structured questionnaire, averaged over all questions, in arbitrary units. The different influences are subject:  $\mathcal{H}(s_j)$ , musical work:  $\mathcal{G}(o_m)$ , interaction acoustics/musical work:  $\mathcal{I}(i_{m,k})$ , acoustics (within-hall variations):  $\mathcal{F}_{w/halls}(p_k)$ , acoustics (across-hall variations):  $\mathcal{F}_{a/halls}(p_k)$  and residual noise:  $\mathcal{R}$ .*

The values in table 1 indicate that none of the terms in equation 3 is negligible. The individual terms are going to be treated in the next sections. The level of the residual noise in the responses to the

<sup>1</sup>The index  $i$  for the individual question will be omitted in all following equations. It goes without saying that there is one *independent* equation for each question. All equations could equally be written in matrix notation.

questionnaires,  $\mathcal{R}$ , is considerable. On the other hand, averaging over different questionnaires will bring noise levels down and the quantitative knowledge of the noise level allows to calculate standard deviations as well as theoretical upper bounds on obtainable correlations.

## Elimination of the term $\mathcal{H}(s_j)$

The term  $\mathcal{H}(s_j)$  in equation 3 is in fact nothing but a global bias for each listener, reflecting the fact that each subject will respond according to his personal scale, or rather to his personal *anchoring* of the scale (the point where “good” starts is really something extremely personal ...). Or, put differently, subjects are much more reliable with respect to *differentiation* tasks rather than for absolute judgments.<sup>2</sup> How is it possible to eliminate the linear offset of the subjects? For a case where all subjects listen to *all* places (and all musical works), the offset, for each question, is simply equal to the difference of the average of the responses of the subject and the average of the responses of all subjects:  $\mathcal{H}(s_j) = \bar{q}_j - \bar{q}$ . In our case things are getting more complicated, as the average of the responses of a subject is influenced by the personal bias as well as by the subset of places occupied. One hence has to impose that the average of the subjects responses equals the average of all responses for the seats occupied by the subject. This yields a circular equation that has to be solved by the following iterative process:

$$\mathcal{H}^{n+1}(s_{j_i}) = \bar{q}_{j_i} - \overline{(q_{j_i} - \mathcal{H}^n(s_{j_i}))_{k \in \{k_j\}}}, \quad (4)$$

where  $i$  is the running index of all subjects,  $l$  the specific subject  $j_i := j_l$  and  $k \in \{k_j\}$  denotes the subset of places occupied by the subject  $s_j$ . Convergence is usually rapid and the iterative process yields stable values after just a few iterations. A redefinition of all the responses to the questionnaires was then carried out:  $q_j \implies q'_j = q_j - \mathcal{H}(s_j)$ . All further analyses used these “new” responses with the subjects’ biases eliminated.

## The terms $\mathcal{G}(o_m)$ and $\mathcal{I}(i_{m,k})$

The main way the musical work influences the responses to the questionnaires is through the variable sound power level of the orchestral ensemble (see [Kah94]). Both the number of players on stage and the sound power levels of the individual instruments have to be considered. A supplementary influence can be found for strongly directional instruments like brass, percussion, piano or the human voice. For those instruments the impression of reverberance will diminish as directional instruments will excite the later part of the room decay to a lesser degree. For the question of reverberance an influence of the *style* of the musical work could equally be observed, as was to be expected from the work of Kuhl ([Kuh54]). Furthermore, even orchestration details will produce significant effects, e.g. the use of instruments with pronounced attacks (percussion, piano) or rich spectra in the upper and/or lower parts of the register. Finally, as it is well known, some instruments have a tendency to enhance the echo sensitivity.

The interaction effect can be linked to specific musical works. Even more interestingly the importance of the interaction effect, as well as the *size* of the global effect of a given change in the instrumentation, constitute an inherent quality of a hall. Some halls are more stable than others for a given change in the orchestration, and this stability can be linked to certain design features.

Two main influences on the interaction musical work/acoustics could be identified. The first is linked to the principal emission direction of directional instruments. For example, a solo piano at the front of the stage will favour orchestra seats on axis. The second is linked to the placement of the sound sources on stage. Inhomogeneities of the stage environment (like pronounced orchestra shells) favour the coupling of parts of the stage towards certain listening zones (e.g. far-away balconies) and affect the judgment of balance as well as the perception of overall loudness.

## Isolating the dependence of the responses on acoustics

In order to concentrate on the dependence of the responses on acoustical criteria,  $\mathcal{F}'(c_k^l)$ , and hence to eliminate the influence of the terms  $\mathcal{G}(o_m)$  and  $\mathcal{I}(i_{m,k})$  there are now several possibilities:

---

<sup>2</sup>This result could equally be found when looking at other studies: confidence intervals for the judgments tend to be much smaller when differences rather than absolute values were considered.

- eliminate the average influence of the musical work in the subjective responses. The procedure is to separate the responses by musical work and to *center* them individually. For each musical work the average for each question is hence equal to zero, the variances remaining unaffected. An additional advantage is that one is concentrating on the variations *within* halls, i.e. reasonably short-term comparisons as the two concerts normally were on subsequent days. On the other hand, comparisons between halls are no longer possible, as the hall averages, together with the musical work averages, are set to zero;
- study the interaction term  $\mathcal{I}(i_m, k)$  by considering different source locations and source directivities for different musical works. This will yield different sets of adapted objective criteria  $c^l$  for different musical works in the relation  $q = \mathcal{F}'(c_k^l)$ ;
- include an objective model of the musical work and its orchestral ensemble on stage. An outline of this model was given in [Kah94], but it would have to be extended to include further parameters. The directivity of the orchestral ensemble has to be taken into account by considering measurements with a *directive* sound source on stage;

## Conclusion

It was shown that when carrying out listening tests in real halls, responses to a structured questionnaire depend on several influences: the acoustics, the musical work performed, the subject and the interaction between musical work and acoustics. A method was given eliminating the personal biases of subjects even for practical cases where each subject does not listen to all places included in the analysis. Furthermore, the relative importance of the different influences could be established and it was found that the musical work performed and its orchestral ensemble has to be taken into consideration when studying the relationships between subjective responses and objective parameters.

When properly separating the different influences the correspondence between subjective evaluation and objective parameters is good, in some cases even excellent. The obtained correlations can be compared to theoretical upper bounds, calculated from the residual noise level contained in the responses to the questionnaires. Even slight changes in the definition of the objective criteria produce noticeable – and often significant – changes. The data obtained during the listening tests in real halls turn out to be sufficiently detailed to be used to *optimize* the definitions for objective criteria linked to the perceptual evaluation of spaces for music.

## References

- [Bar88] M. Barron. Subjective study of British symphony concert halls. *Acustica*, **66**, pp. 1–14, 1988.
- [Bar93] M. Barron. *Auditorium Acoustics and Architectural Design*. E & FN Spon, London, 1993.
- [Ber92] L.L. Beranek. Concert hall acoustics—1992. *JASA*, **92**(1), pp. 1–39, 1992.
- [GBS93] A.C. Gade, J.S. Bradley, and G.W. Siebein. Effects of measurement procedure and equipment on average room acoustic measurements. *JASA*, **93**(4), p. 2265, 1993.
- [Gri93] D. Griesinger. Quantifying musical acoustics through audibility. Vern O. Knudsen Memorial Lecture, 126th ASA Convention, Denver, Colorado, 1993.
- [HD71] R.J. Hawkes and H. Douglas. Subjective Acoustic Experience in Concert Auditoria. *Acustica*, **24**, pp. 235–250, 1971.
- [JKWW92] J.-P. Jullien, E. Kahle, S. Winsberg, and O. Warusfel. Some results on the objective characterisation of room acoustical quality in both laboratory and real environments. *Proc. Inst. of Acoust.*, **XIV**, Birmingham, 1992.
- [Kah94] E. Kahle. Influence of size and composition of the orchestra on the perception of room acoustical quality. In *Proceedings of the Wallace Clement Sabine Centennial Symposium, Cambridge, Ma., June 5 - 7 1994*, pages 207–210. Ac. Soc. Am., 1994.
- [Kah95] E. Kahle. *Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras*. PhD thesis, Université du Maine, Le Mans, 1995.
- [Kuh54] W. Kuhl. Über Versuche zur Ermittlung der günstigsten Nachhallzeit grosser Musikstudios. *Acustica*, **4**, pp. 618–634, 1954.
- [Lav89] C. Lavandier. *Validation perceptive d'un modèle objectif de caractérisation de la qualité acoustique des salles*. PhD thesis, Université du Maine, Le Mans, 1989.
- [WKJ93] O. Warusfel, E. Kahle, and J.-P. Jullien. Relationships between objective measurements and perceptual interpretation: The need for considering spatial emission of sound sources. *JASA*, **93**(4), p. 2281, 1993.