PERCEPTION OF MUSICAL DYNAMICS: ORCHESTRA SPECTRA COMBINED WITH AUDITORY MODELING

T Lokki   Aalto Acoustics Lab, Dept of Information and Communications Engineering, Finland
P Llado   Aalto Acoustics Lab, Dept of Information and Communications Engineering, Finland

1   INTRODUCTION AND MOTIVATION

It is generally accepted that the quality of a concert hall cannot be described well enough with objective room acoustical parameters. Even though the parameters would be measured according to the ISO 3382-1:2009 standard, and are inside the recommended limits, the success of a hall cannot be guaranteed. Based on recent research, one feature of the highly renowned concert halls seems to be their responsiveness to musical dynamics[1,2,3,4,5]. Enhanced dynamic variation in music is associated to more expressive performance, which is often the aim of the musicians. A few studies discuss the architectural features that enable large dynamics[4,6] but more research is needed to understand the dynamic responsiveness of a hall and its connection to music and architecture.

The reasons for the dynamic responsiveness of concert halls were revealed a decade ago[7], but no comprehensive objective metrics to predict dynamic responsiveness has been presented. This paper is one in the series of papers[8,9] that propose some objective ways to predict the dynamic responsiveness with measured binaural room impulse responses. Our earlier research led to propose a new metric, binaural dynamic responsiveness (BDR)[8], that was shown to differentiate rectangular halls from non-rectangular halls, but mainly on high frequencies. In addition, it highlights the differences between hall types at larger distances from the stage, result that is also supported by the listening test data of sudden transition in played dynamics[1]. At low frequencies the dynamic responsiveness is most probably related to non-linear sensitivity of human hearing that might be modeled taking into account the equal loudness contours[9].

The measurement of subjective dynamic responsiveness is really hard. Nonetheless, we performed such measurements by measuring the impact of a *crescendo* on subjects, both with skin conductivity measurements and with standard listening tests[2]. Based on that study, it could be confirmed that rectangular halls with strong lateral reflections have larger dynamic responsiveness than other hall types. Moreover, when comparing to objective room acoustical parameters, we found that at low frequencies the best correlation were obtained with strength (G) and late lateral energy (L_j) while at high frequencies with early interaural correlation (as binaural quality index (BQI)).

To summarize the earlier work, it is not clear how to objectively (based on acoustic measurements) predict the dynamic responsiveness of a hall. Therefore, more research is needed and this paper tries to fill in some gaps in the current knowledge. The novelty in this paper is a more accurate analysis of the spectrum changes of the orchestra according to the played levels and first attempts to model non-linear human spatial hearing on different sound pressure levels. Moreover, initial results of a listening test between concert halls rendered at different levels are presented.

1.1   The origins of musical dynamics

Music listened to in-situ in a concert hall is affected by various factors, such as the level of the music, the instrumentation (frequencies excited), the spatial room impulse response, as well as the aspects of human (spatial) hearing. When using the traditional source-medium-receiver model to decribe the
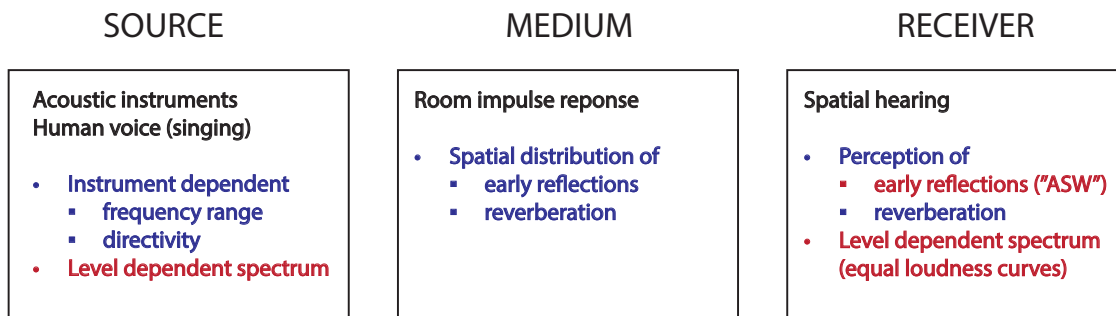
SOURCE

MEDIUM

RECEIVER

Acoustic instruments
Human voice (singing)

- Instrument dependent
  - frequency range
  - directivity
- Level dependent spectrum

Room impulse reponse

- Spatial distribution of
  - early reflections
  - reverberation

Spatial hearing

- Perception of
  - early reflections ("ASW")
  - reverberation
- Level dependent spectrum
  (equal loudness curves)

Figure 1   Source-medium-receiver model in concert hall acoustics. The non-linear aspects are high-lighted with red color. The level dependent perception of early reflections could change the perceived dynamics and auditory source width (ASW)[1,4].

ensemble of all phenomena, the medium, i.e. the spatial room impulse response, is the only linear part of the system. The source is non-linear as the spectrum of the orchestra is level dependent and varies also depending which instruments are in voice. The receiver, i.e. human hearing system, is also non-linear regarding level due to the active process of the cochlear mechanics. Figure 1 illustrates the source-medium-receiver model for concert halls and highlights the non-linear aspects.

2   LEVEL DEPENDENT SPECTRUM OF AN ORCHESTRA

The frequency and dynamic range of an orchestra vary a lot depending on the composition and the size of the orchestra. Moreover, composers tend to write softer passages to instruments that are not usually playing in the extreme ends of the instruments' register. When the full orchestra is playing loud and all insruments are in voice, the frequency range is really wide, even covering from 20 Hz (e.g. bass, drum and other percussions) up to 15-20 kHz (e.g. overtones of brass instruments and cymbals). In other words, the difference in soft and loud passages in classical music is enhanced by the variation in the the frequency range in addition to level differences.

To find out the frequency response of the soft and loud playing of an orchestra, we analysed several anechoic full orchestra recordings. They were 2-4 minutes long passages of Mahler's 1st symphony, Beethoven's 7th symphony, Bruckner's 8th symphony and Mozart's opera aria[10]. Another set of ane-choic recordings analysed contains Beethoven's 8th symphony, entire movements 1, 2 and 4[11]. The analyses were done in the same way as described earlier[9]. First, the full orchestra recording was chopped to one second long frames with 50% overlap. Then, the silent and noise frames were re-moved and the frequency responses of all frames containing signals were computed. The frequency responses were smoothed to 1/3 octave band resolution so that each frequency response had 237 frequency bins on the logarithmic scale on audible frequencies. Finally, at each frequency bin, the levels were ordered to obtain a distribution of spectral magnitudes at different frequencies over the entire recording.

The proposed analysis allows to look at, e.g. the level of first and last percentile at each frequency bin, as illustrated to all analysed music in Figure 2. We refer to them as profiles for *pianissimo* and *fortissimo*. It should be reminded that they are not exact frequency responses of soft and loud playing as the instrumentation might change the frequency content depending on the piece. However, they show statistically the representative magnitude responses in soft and loud playing. Figure 2 plots also the median values, which could be considered statistically the frequency response of *pianissimo* and *fortissimo* playing. The median *fortissimo* response is quite close to all recordings, except Mozart which does not contain full orchestra playing loud. The *pianissimo* responses have more variation. The Mahler extract is a loud full orchestra passage and does not represent well soft playing. In Beethoven 8th symphony 4th movement, the softest passages are played with violins, resulting as high levels be-
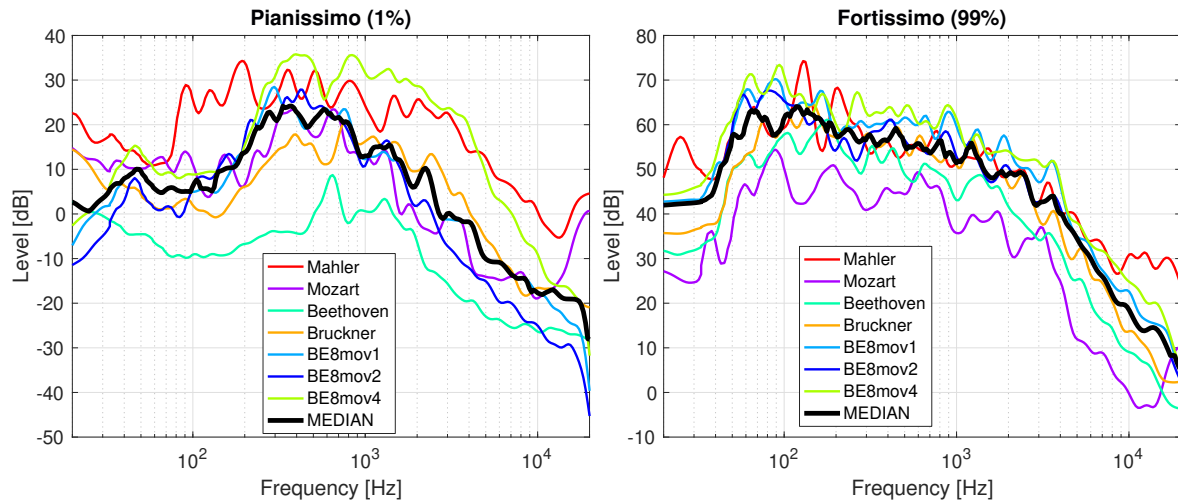
Figure 2   Results of the analysis of the levels from the anechoic symphony orchestra recordings.

tween 350 Hz and 10 kHz. Finally, the beginning of Beethoven's 7th symphony has long notes played by a few woodwind instruments, which most probably affects to much lower *pianissimo* responses. These examples emphasise that the proposed analysis method is not suitable for a single piece, but the median of a handfull of different music passages could statistically give meaningful results.
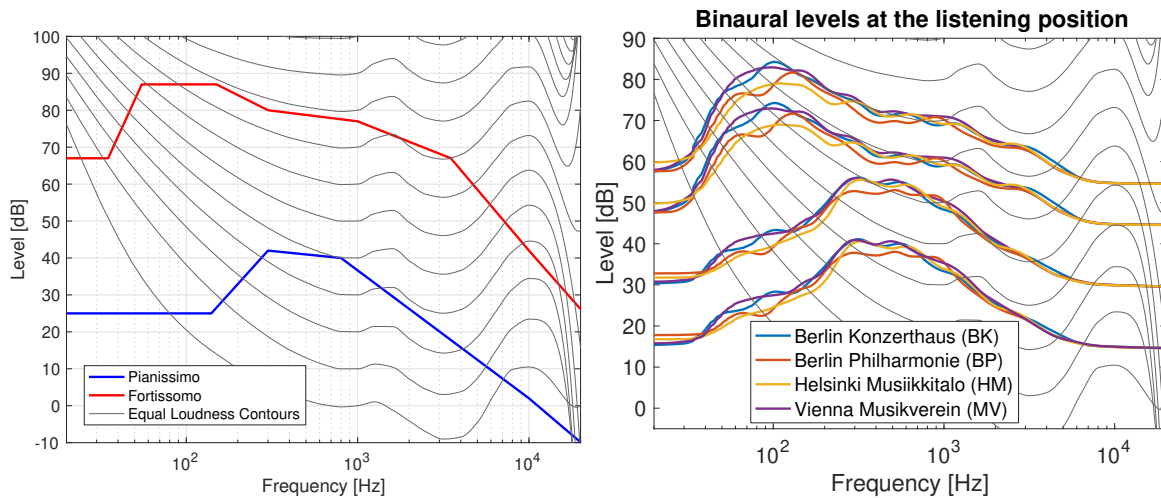

Figure 3   **Left:** Median spectra of *pianissimo* and *fortissimo* plotted on top of equal loudness contours. **Rigth:** Concert hall responses multiplied with *pianissimo* (40 and 55 dB) and *fortissimo* (70 and 80 dB) responses.

The median spectra of the *pianissimo* and *fortissimo* responses could be considered as target spectra for filter design. Such target spectra are plotted in Figure 3 on top of the equal loudness contours (ELC)[12] to highlight the possible perceptual differences. In this work, we used these target curves to design a cascade of shelf filters for processing "*pp*- and *ff*-weighted noise". It should be noted that the ELC are defined using individual tones and it is not well-known if they could be applied to wide band signals. All in all, they highlight the perceptual differences of these two curves, in particular at low frequencies in which the loudness perception is not linear related to sound pressure level changes.

To see better the difference of the concert hall responses at different levels on the ELC curves, the
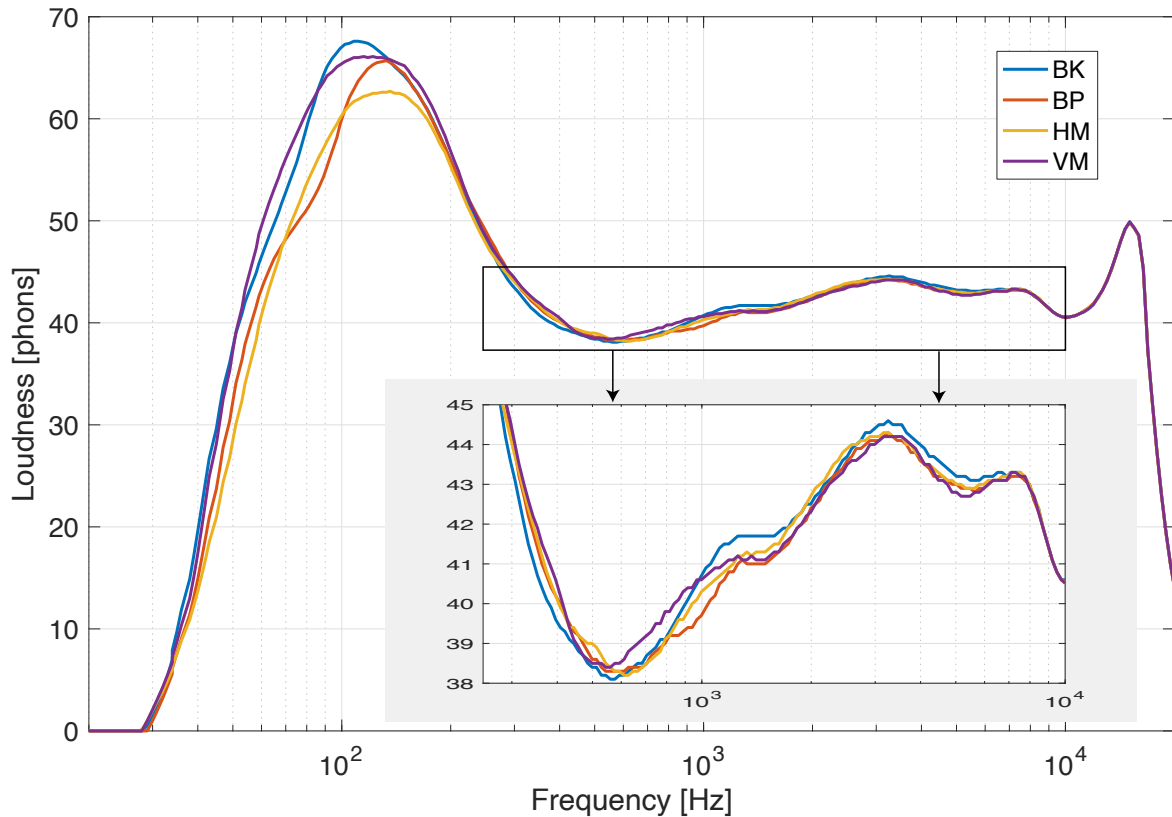
Figure 4 Difference in phons of *fortissimo* (80 dB) and *pianissimo* (40 dB) mapped on the equal loudness contours. The zoomed area is between 250 Hz and 10 kHz.

spectra of 40 dB in Fig. 3 is subtracted from 80 dB so that both of them are first mapped to equal loudness contours. Thus, the levels (dB) are not subtracted. Instead, the computation is done in terms of loudness (phons). The result of this operation is plotted in Fig. 4 which estimates the loudness differences of *fortissimo* and *pianissimo* playing in each hall. The plot reveals that the perceived dynamic range is indeed larger, in particular at low frequencies, and it is even larger in halls with strong bass response. The difference is due to denser ELC at low frequencies, i.e., human hearing expands the perceived dynamic range below 200 Hz.

3 PREPARING SAMPLES FOR OBJECTIVE ANALYSIS AND FOR SUBJECTIVE LISTENING TEST

In order to evaluate the perceptual effects of musical dynamics, two set of samples were prepared. We used the loudspeaker orchestra measurements of four concert halls, i.e. each hall had 24 source positions / channels on stage, and the receiver position used in this study was 11 m from the stage. The measurements, room acoustical parameters and rendering techniques are presented in detail in the recent article[13] and the "fro" position of those halls were used in this study. The auralization method for the listening test was the same as in[5], the 45-channel reproduction system in an anechoic room.

The novelty here is the excitation signals in the concert halls, which are at four different levels, at 40, 55, 70, 80 dB. Moreover, we use two different set of signals:

- ***pp-* and *ff*-weighted white noise**. One-second long white noise burst (24 uncorrelated sequences at different loudspeakers at stage), multiplied with *pianissimo* (for 40 and 55 dB) and *fortissimo* (for 70 and 80 dB) spectra (Fig. 3). In other words, originally flat spectrum is modified with level dependent spectra and convolved with measured concert hall impulse responses. The frequency reponses computed from the binaural samples are plotted in Fig. 3 (left and right ears are merged together with power summation[14]).

- **Music samples**. Two one-second long extract from Beethoven's 7th symphony; bar 13, beat 3 in which only clarinets and bassoons are in voice (*pianissimo*) and bar 15, beat 2, full orchestra playing (*fortissimo*).

As said, both for the binaural auditory modeling and for the listening test the signals were played at levels of $L_{eq} = 40, 55, 70$, and 80 dB. In the listening room the levels were checked with an SPL meter.

4 RESULTS OF THE BINAURAL AUDITORY MODELING

In previous studies, it has been shown that the impact of a crescendo is stronger in halls with strong lateral early reflections[2]. In that study, the impact was assocated strongly to loudness but also to widening of the sound image, at least to some extent. Such widening effect according to musical dynamics has been explained by the threshold of perceived early reflection; the threshold is lower for lateral reflections than for the median plane reflections[4]. Here, we concentrate on the widening effect with a non-linear auditory modelling.

Inter-aural cross-correlation (IACC) has been used as a measure of the spaciousness of concert halls[15]. The IACC is often computed from the signal arriving to the ears of a listener. However, the IACC does not include any level-dependent characteristic that explains the responsiveness of a concert hall, since a louder binaural signal would result in the same IACC result. Level-dependent models have been proposed in the past, by adding a level-dependent gain factor to the IACC. While this step sounds reasonable from an application point of view, it does not help understanding what happens from a perceptual stand point.

In this paper, we assess the effect of the resposiveness with a level-dependent auditory model. The binaural input signal was computed by convolving the noise stimulus (i.e. *pp-* and *ff*-weighted white noises) with the BRIRs of each hall. The left and right signals were processed separately with the model proposed by Verhulst et al.[16]. This model consists of a transmission-line basilar cochlear model followed by an inner hair-cell stage to simulate the neural transduction. This model is sensitive to the signal level and incorporates the non-linearities of the peripheral auditory processing. Thus, the model output for a loud input is not computed by applying a gain factor to a quiet output.

The model outputs a time-domain continuous representation of the neural spikes in the auditory nerve, after the inner-hair-cells neural transduction. These time-domain representations, which are frequency dependent, are then used to compute the IACC between the left and right ears signals for each auditory filter. The estimated apparent source width (ASW) is computed as 1-IACC and the results are plotted in Fig. 5. The results follow an expected behaviour, since the ASW increases over level, representing the responsiveness of the halls.

5 RESULTS OF THE LISTENING TEST

To verify the results of the objective modeling, a focused listening test was organized. Fifteen participants, researchers in the Aalto Acoustics Lab, performed a paired comparison of four halls in four different levels with two signals. The selection within a pair of samples was based on larger perceived SPACIOUSNESS, thus not loudness or preference but the spatial aspect of the sound fields.
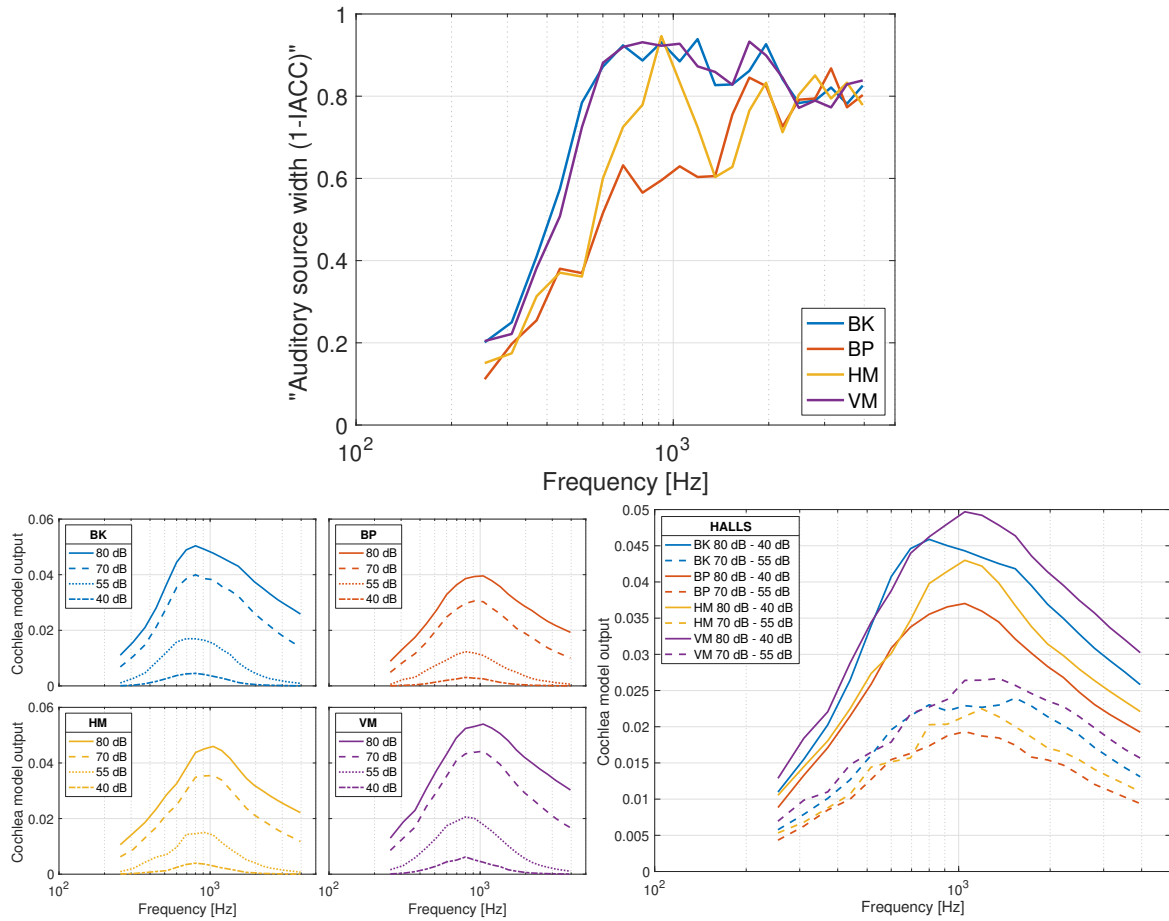
Figure 5 **Top**: Computational auditory source with traditional level independent IACC (plotted as 1-IACC). **Bottom**: Output of the binaural auditory model predicting auditory width of the perceived source. On the left, four curves in each subplot include the source spectrum (40 and 55 dB the *pianissimo* spectrum and 70 and 80 dB the *fortissimo* spectrum). On the right, the differences between 80 and 40 dB as well as 70 and 55 dB are plotted.

The results are plotted in Figures 6 and 7. The number of wins in paired comparison is shown and the maximum number of wins in Fig. 6 for one hall is 45. There are two main results. First, there is hardly any difference between music and shaped noise excitations. Second, regardless of the played level the participants heard the spaciousness of the halls almost always in the same order; Musikverein, Konzerthaus, Musiikkitalo, Philharmonie. We also pooled the data so that the differences between hall types can be compared, i.e. the shoebox halls (MV and BK) against the vineyard halls (HM and BP). Even though the results in Fig. 7 are very clear, some differences are seen between levels. If we combine both signals, we can see that at the lowest levels there are a few more wins of vineyard halls than at higher levels. The numbers are at 10 wins out of 120 (40 dB), 8/120 (55 dB), 2/120 (70 dB), 3/120 (80 dB). This result suggest that the audible defferences between shoebox and vineyard halls are slightly smaller at low levels and larger at high levels.

6   CONCLUSIONS

This paper tries to model objectively the responsiveness of concert halls to dynamical variations in music. The novelties are a more accurate analysis of the full orchestra spectrum, and an attempt to model non-linearities of the peripheral auditory processing. The latter one explains the changes
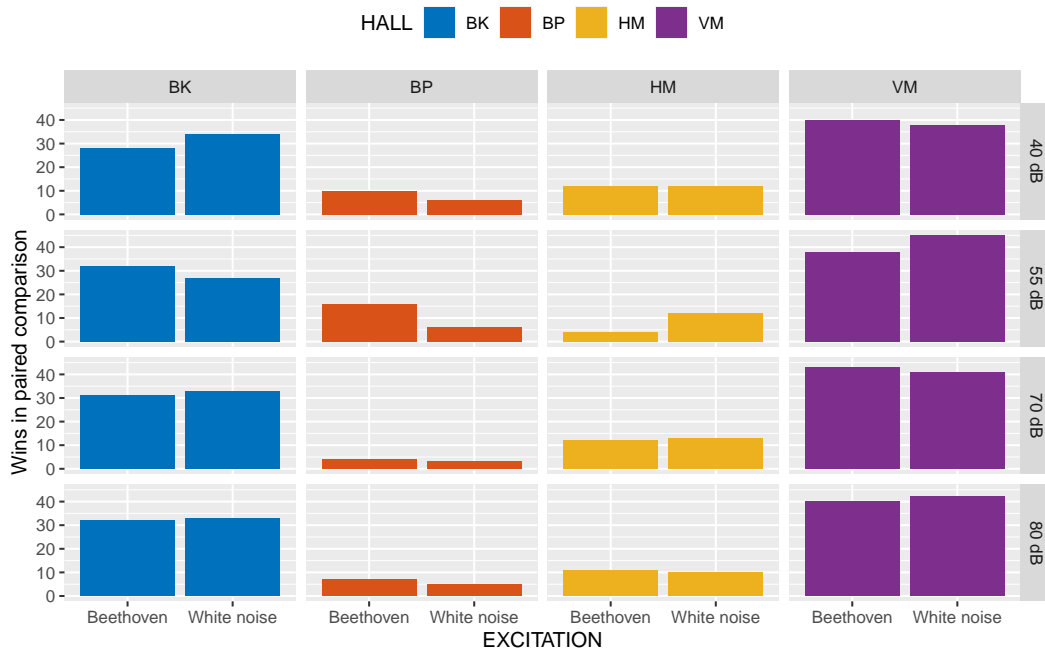
Figure 6   Results of the listening test, wins in the full paired comparison.
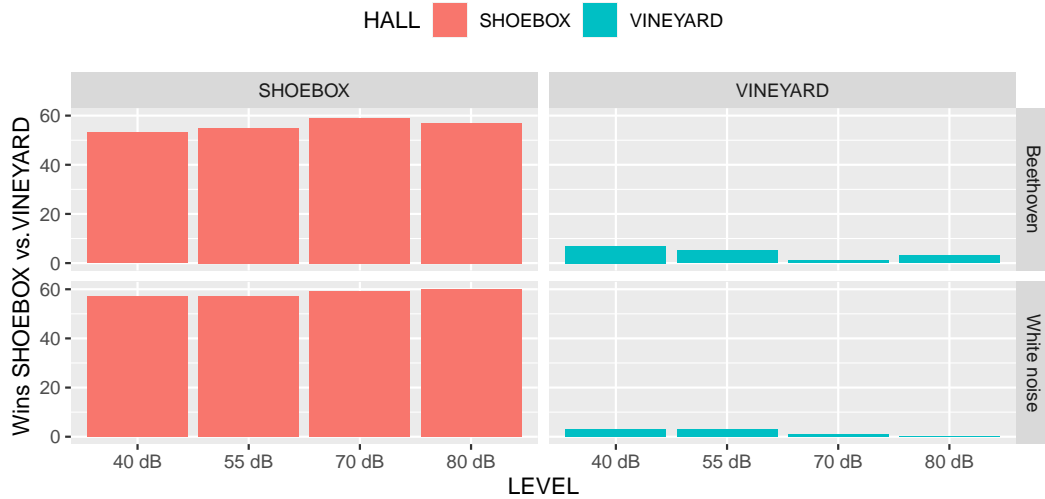

Figure 7   Results of the listening test, wins when shoebox halls are compared with vineyard halls.

in the apparent source width according to sound pressure level and spectral changes, at least to some extend. It is also seen that the differences between halls are much smaller when an orchestra is playing in *pianissimo* than in *fortissimo*. Although, the shoebox halls have already larger predicted ASW in soft dynamics than the vineyard halls, the variation in the cochlear model output is the largest in Musikverein and Konzerthaus, supporting our previous listening test results[2]. The conducted listening test with music and shaped noise samples confirmed this objective modeling result.

7 REFERENCES

1. J. Pätynen and T. Lokki. Perception of music dynamics in concert halls. *Journal of the Acoustical Society of America*, 140(5):3787–3798, November 2016.
2. J. Pätynen and T. Lokki. Concert halls with strong and lateral sound increase the emotional impact of orchestra music. *Journal of the Acoustical Society of America*, 139(3):1214–1224, March 2016.
3. J. Pätynen and T. Lokki. Dynamic responsiveness in concert halls as source of emotional impact. In *the Tenth International Conference On Auditorium Acoustics*, pages 237–244, Hamburg, Germany, October 4-6 2018.
4. E. Green and E. Kahle. Dynamic spatial responsiveness in concert halls. *Acoustics*, 1(3):549–560, 2019.
5. T. Lokki, L. McLeod, and A. Kuusinen. Perception of loudness and envelopment for different orchestral dynamics. *Journal of the Acoustical Society of America*, 148(4):2137–2145, October 2020.
6. T. Lokki and J. Pätynen. Architectural features that make music bloom in concert halls. *Acoustics*, 1(2):439–449, May 2019.
7. J. Pätynen, S. Tervo, and T. Lokki. Binaural dynamic responsiveness in concert halls. In *International Symposium on Room Acoustics (ISRA 2013)*, Toronto, Canada, June 9-11 2013.
8. J. Pätynen, S. Tervo, P. W. Robinson, and T. Lokki. Concert halls with strong lateral reflections enhance musical dynamics. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 111(12):4409–4414, 2014.
9. T. Lokki and J. Pätynen. Objective analysis of the dynamic responsiveness of concert halls. *Acoustical Science and Technology*, 41(1):253–259, January 2020.
10. J. Pätynen, V. Pulkki, and T. Lokki. Anechoic recording system for symphony orchestra. *Acta Acustica united with Acustica*, 94(6):856–865, November/December 2008.
11. C. Böhm, D. Ackermann, and S. Weinzierl. A multi-channel anechoic orchestra recording of beethoven's symphony no. 8 op. 93. *Journal of the Audio Engineering Society*, 68(12):977–984, December 2020.
12. C. Hummersone. ISO 226:2003 Normal equal-loudness-level contours. https://github.com/IoSR-Surrey/MatlabToolbox, 2023. Retrieved August 7, 2023.
13. T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo. Concert hall acoustics: Repertoire, listening position and individual taste of the listeners influence the qualitative attributes and preferences. *Journal of the Acoustical Society of America*, 140(1):551–562, July 2016.
14. V. P. Sivonen and W. Ellermeier. Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. *Journal of the Acoustical Society of America*, 119(5):2965–2980, 2006.
15. L. Beranek. *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. Springer-Verlag (New York), 2nd edition, 2004. 664 pages.
16. S. Verhulst, A. Altoè, V. Vasilkov, and A. Osses. Verhulst et al 2018 auditory model v1.2. https://doi.org/10.5281/zenodo.3717800, 2020. Retrieved August 7, 2023.