A study of the changes in musical dynamics in concert hall acoustics as measured by perceptual changes in loudness and envelopment

Laura McLeod

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 5.7.2019

Supervisor and advisor

Prof. Tapio Lokki

Aalto University
School of Electrical Engineering

Aalto University
School of Electrical Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Laura McLeod

Title A study of the changes in musical dynamics in concert hall acoustics as measured by perceptual changes in loudness and envelopment

Degree programme Computer, Communication and Information Sciences

Major Acoustics and Audio Technology **Code of major** ELEC3030

Supervisor and advisor Prof. Tapio Lokki

Date 5.7.2019 **Number of pages** 59+4 **Language** English

Abstract

In this thesis, the perception of musical dynamics in concert hall acoustics is studied by using two of the most influential and more intuitive perceptual factors: loudness and envelopment/width. This study is carried out by using a listening test and interview to assess the loudness and envelopment/width of seven different European concert halls at three different listening positions. The accompanying literature review explores the pschyoacoustics and objective measures of loudness and envelopment as well as the seminal works of perceptual studies in concert hall acoustics. The experiment makes use of a state-of-the-art listening room and a pair-wise comparison listening test to evaluate samples with four different dynamic levels which have approximately constant SPL for the same dynamic levels. The main findings align with previous studies of dynamic responsiveness by Pätynen and Lokki, concluding that shoebox-style halls exhibit better musical dynamics than modern vineyard-style halls. The findings also show that differing seat position has a greater effect on the perception of musical dynamics, where little change is found for differing musical dynamics at the same seating position. Finally, objective measures of strength and SPL are found not to follow subjective loudness suggesting a new objective measure is required.

Keywords Concert hall acoustics, musical dynamics, envelopment, loudness

Preface

I would like to thank the Acoustics and Audio Technology department at Aalto University who participated in the interview and listening test and provided an interesting insight into their perception. Special thanks to my supervisor, Tapio Lokki, who helped me throughout the masters thesis process and to Antti Kuusinen who was always available to answer my questions. Finally, to my friends both in the Acoustics masters programme and to those outside of the department and to my family for supporting me throughout this process and especially my Mum for a last minute read through.

Otaniemi, 5.7.2019

Laura A. McLeod

Contents

Symbols and abbreviations

Symbols

RT	Reverberation Time [s]
EDT	Early Decay Time [s]
$ITDG$	Initial-Time-Delay Gap [s]
C_{80}	Clarity [dB]
G	Strength [dB]
L_J	Late Lateral Energy [dB]
$IACC$	Interaural Cross-Correlation Coefficient
ASW	Apparent Source Width
T_{60}	Reverberation Time [s]
LEV	Listener Envelopment
J_{LF}	Lateral Energy Faction [%]
JND	Just-Noticeable Difference
SPL	Sound Pressure Level [dB]

1 Introduction

Vienna's Musikverein, Amsterdam's Concertgebouw and Boston Symphony Hall are considered to be some of the best sounding concert halls in the world [1]. This is mainly due to their architectural elements that transform a simple anechoic orchestral piece into an overwhelming symphony of sound. The acoustic spaces enhance the musical dynamics, creating a louder and more enveloping space, creating a more moving and emotional experience for the listener.

While musical dynamics have long been an effective musical tool used by conductors and classical musicians to enhance the musical experience by invoking emotion and passion to the music, research with respect to room acoustics has been little to none. Previously, research has shown that changes in musical dynamics are dependent on the sound-level and spectrum of the source and the frequency dependence of spatial hearing affected by the directionality of sound and the room‘s geometry [2] [1].

More recent research has shown that dynamic changes in music are affected by a number of perceptual factors such as loudness, width/envelopment, brightness and clarity among other subjective attributes [3]. It has even been shown that musical dynamics may have a huge influence on measuring the acoustical quality of concert halls, where halls that are more dynamically responsive mirror Beranek's widely accepted acoustic quality ratings [3]. However, concrete subjective measures of musical dynamics and their associated objective measures are lacking in literature and missing from international standards. Hence we look to two of their dominating associated perceptual factors which are more well-understood and have clearer objective parameters to draw conclusions.

In 2016, Pätynen et al. [3] showed that of 57 unique attributes defined by 14 expert listeners from the department of Acoustics and Audio Technology at Aalto University, the two most related attributes with describing musical dynamic changes were loudness and width/envelopment. While 'loudness' is a relatively well understood term, mainly due to its direct association with objective measures of sound level or strength of the sound, changes in loudness level are less understood and have been shown to be affected by a number of different attributes such as reverberance, clarity and envelopment.

'Envelopment' is left open to interpretation throughout literature, mostly found to be associated with width, openness and spatial impression. However, there is some evidence to show that envelopment varies with increasing lateral energy or an increase in perceived bass frequencies, which is in turn attributed by a change in loudness [4]. Mostly envelopment is associated with 'listener envelopment' i.e. how much of the sound feels as though it is around the listener and 'apparent source width' which is the measure of how wide the source is perceived to be. Both subjective attributes have relatively well accepted objective parameters of late lateral energy and lateral energy fraction, respectively.

So how can we use these more accepted and intuitive subjective parameters to draw conclusions for how we perceive musical dynamics in concert hall acoustics? This is the question this thesis aims to explore.

1.1 Research question

The main research question for this thesis is: **How do we perceive changes in musical dynamics in concert hall acoustics?**. This was studied by exploring the relationship between two more intuitive subjective attributes with internationally accepted associated objective measures: loudness and envelopment/width. These were chosen due to the findings of Pätynen et al. where loudness and envelopment/width were found to have the most influence on changes in musical dynamics [3]. These perceptual factors were then assessed at different dynamic levels: *pp*, *mp*, mf and ff rendered for seven different concert halls in three listening positions using a pair-wise comparison listening test. These were rendered using advanced auralisation and signal processing methods and were projected in a state-of-the art listening room for subjective testing.

1.2 Overview of the thesis

The thesis begins with an introduction to the basic concepts of concert hall acoustics and an overview of the subjective and objective measures that are currently used to describe concert hall acoustics.

While the research question focuses on dynamic changes, the questions asked to participants in the test focus on perceived loudness and envelopment. Hence, the psychoacoustics of both subjective attributes are explored in order to understand what may have affected listener‘s answers to the questions asked. Further to this, a literature review is carried out beginning with an overview of previous perceptual studies of concert hall acoustics and alternative methods to the one used in this experiment, as well as experiments more focused and related to dynamic changes in concert hall acoustics.

In the final section, the results of the listening test on the perceptual loudness and envelopment in seven different concert halls with three different seating positions are presented. The listeners were asked to evaluate which of the halls were ‘louder’ or more ‘enveloping/wide’ for the music reproduced at four different sound pressure levels. A supplementary interview then followed to see how participants responded to the subjective listening test. The aim was then to draw conclusions from the perception of changes in dynamic level as a result of listener‘s response to the two questions in order to answer the main research question: **How do we perceive dynamic changes in concert hall acoustics?**

2 Concert hall acoustics: where are we now?

2.1 Introduction of basic concepts

Consider a sound source within an enclosed space and a listener some distance away from the sound source. The first sound to reach the listener is the direct sound which arrives at a time, t. This is the distance from the listener to the sound source divided by the speed of sound, c which is around 334 ms^{-1}. After a short amount of time known as the *initial time delay gap* (ITDG), the first set of reflective sounds reach the listener. According to Beranek [5], an initial time delay gap should be less than 25 ms for good quality halls and no more than 35 ms. Above this, the hall starts to sound more like an arena.

The first reflective sounds usually arrive typically within around 20-30 ms and are known as the early reflections. These have a lower amplitude than the direct sound and arrive as multiple impulses. These early reflections usually result as a reflection from the walls, floor or ceiling of the acoustic space. If the reflections arrive from the lateral directions, the source is perceived subjectively as broadened giving it an *apparent source width* (ASW). This gives the source a fuller sound and is synonymous with the perception of spaciousness.

After the early reflections at around 80-100 ms, the reflections become indistinguishable as individual impulses and contribute the reverberant sound or late reflections. These are lower still in amplitude and are much closer together in time. The *listener envelopment* (LEV) is the degree to which the reverberant sound seems to surround the listener from all directions, where in good halls the sound is free to travel around the spaces to give a full immersion.

If the source emits sound continuously, the acoustic space reaches an equilibrium sound level until the source stops and the sound begins to decay. If the source is an impulsive sound, then the result in Figure 1 is created. The Figure shows a unique fingerprint for a concert hall known as the impulse response, which contains all of the information needed in order to explain a concert hall.

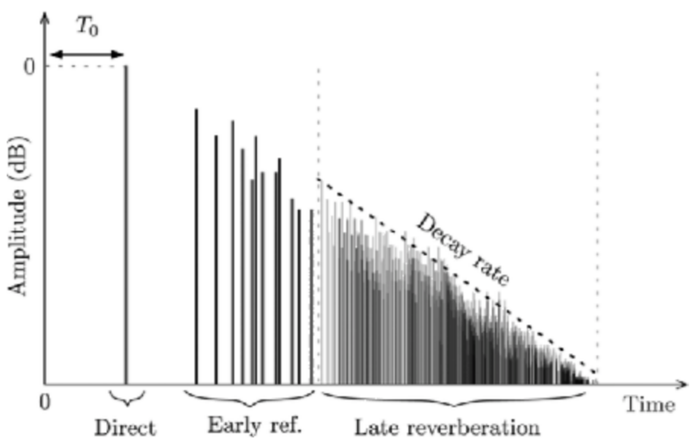


Figure 1: Impulse response [6]

The impulse response of a concert hall can be affected by a number of different factors. Size, geometry, air temperature, humidity, absorption, are all factors which vary in concert halls and help to create their unique impulse response. One of the fundamental measures of concert hall acoustics was proposed by physics Harvard University professor Wallace C. Sabine [7] after he was asked to improve the acoustics of a lecture hall as part of the Fogg Art Museum. Sabine heavily researched architectural acoustics and is credited with building the acoustics of one of the best concert halls in the world: Boston Symphony Hall. His main equation for calculation of the reverberation time, T_{60} is a measure that is still used to determine the acoustic quality of concert halls today:

$$T_{60} = 0.16\frac{V}{A} \tag{1}$$

where T_{60} is the time taken for the sound to decay 60 dB after the source is stopped, V is the volume of the room and A is the total absorption which is sum of all of the absorption coefficients multiplied with the surface areas of the room they cover. While the most renowned of his work was his attribution of reverberation time, Sabine also identified three main perceptual attributes of concert hall acoustics: loudness, distortion of complex sounds (interference and resonance) and confusion (reverberation, echo and extraneous sounds).

2.2 Objective measures of concert hall acoustics and how to calculate them

Since the work of Sabine [7], there has been a lot more research into other objective measures of concert hall acoustics to determine the acoustic quality in addition to the reverberation time. As a result, there have also been many disputes over the subjective and objective measures required to define the acoustics of a concert hall. However, as it stands, the following section presents the objective parameters from ISO3382-1 [8] the current international standard which is widely accepted and used in practice by acousticians. Table 1 lists the subjective measures and their associated objective measures as found in the standard. The following section explains in more detail these objective measures and what they represent as explained in Pulkki and Karjalainen [9].

Table 1: The objective measures defined in concert hall acosutics in accordance with ISO 3382-1, 2009 [8]

Subjective measure	Objective measure
Subjective level of sound	Sound strength G in decibels
Perceived reverberance	Early decay time (EDT)
Perceived clarity of sound	Clarity C_{80} in decibels
Apparent source width (ASW)	Early lateral energy fraction, J_{LF}
Listener envelopment	Late lateral sound level, L_J in decibels

2.3 Calculation of objective measures

2.3.1 Subjective Level

The strength (G) is the ratio of energy of the listening position over the energy that is measured at 10 m in free field. $p(t)$ is the sound pressure measured at the listener's position and $p_A(t)$ is the sound pressure measured in the free field, usually using an omnidirectional source.

$$G = 10log_{10}\frac{\int_0^\infty p^2(t)dt}{\int_0^\infty p_A^2(t)dt} \tag{2}$$

2.3.2 Apparent Source Width

Lateral Energy Fraction (J_{LF}) was proposed by Barron and Marshall [10] as a measure for the apparent source width. J_{LF} is calculated using the sound pressure recorded from a figure-of-eight microphone, $p_8(t)$ with the sound pressure recorded from an omni-directional microphone $p(t)$.

$$J_{LF} = \frac{\int_{5_{ms}}^{80_{ms}} p_8^2(t)dt}{\int_0^{80_{ms}} p^2(t)dt} \tag{3}$$

2.3.3 Subjective Clarity

Clarity (C_{80}) is a measure of the ratio of the energy from the early and late reverberation responses. As presented earlier, since the accepted value for the early reflections is to be at around 80 ms, this results in the energy being calculated in this range.

$$C_{80} = 10log_{10}\left(\frac{\int_0^{80_{ms}} p^2(t)dt}{\int_{80_{ms}}^\infty p^2(t)dt}\right) \tag{4}$$

2.3.4 Listener Envelopment

The late lateral energy (L_J) is the ratio of the sound energy after 80 ms to the energy of the of the measurement source at 10 m away in a free field.

$$L_J = 10log_{10}\left(\frac{\int_{80_{ms}}^\infty p_8^2(t)dt}{\int_0^\infty p_{10}^2(t)dt}\right) \tag{5}$$

2.3.5 Subjective Reverberance

The early decay time was proposed initially by Jordan [11] from the early decay curve proposed by Schroeder [12], as it was found that the reverberation time measured for the first -15 dB drop was similar in two concert halls. The decay time is therefore calculated from the gradient of the curve usually from 0 dB to -10 dB multiplied by 6.

$$EDC(\tau) = \int_\tau^\infty h^2(\tau)d\tau \tag{6}$$

Reverberation time as proposed by Sabine [7] is also used as a measure of subjective reverberance in this thesis, since it is a widely recognised term of describing the acoustic quality of concert halls in past literature.

2.4 Summary

In this section, the basic concepts of concert hall acoustics were introduced as well as the objective measures used by the ISO 3382-1, 2009 [8]. The terminologies used in this section as well as the objective parameters presented are often referred to throughout the background of this thesis. In the following sections, alternative measures brought forward by other researchers due to limitations of the objective measures from the international standard are discussed for describing concert hall acoustics. However, the study presented in section 6 of this thesis uses these definitions in order to calculate the objective measures for each of the seven different concert halls.

3 The psychoacoustics of loudness

Loudness has long been a widely accepted fundamental perceptual acoustic phenomenon, not only in concert hall acoustics. This section aims to explain loudness psychoacoustically. The section begins with the definition of loudness and different scales. The next part discusses loudness thresholds such as the minimum audible pressure and minimum audible frequency and how these can be measured. The following section looks at how loudness depends on level and frequency, mainly exploring loudness curves. The loudness of complex sounds is also discussed with reference to orchestral music, since this is the main source used in the listening test in section 6. Finally, other important phenomena regarding loudness is explored. Most of the knowledge from this section comes from the works of Pulkki and Karjalainen [9], Blauert [13], Moore [14], Meyer [2], Zacharov et al. [15] and Fastl and Zwicker [16].

3.1 Definition of loudness

In the American National Standard Institute [9], loudness is defined as an 'attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud'. Humans can detect ranges of between 20 Hz - 20 kHz in frequency at a 'normal' level, but can even detect lower or high frequency ranges when the level is increased. Sounds that are heard and exist above or below these frequencies are known as infrasound ($<$ 20 Hz) and ultrasound ($>$ 20 kHz) respectively. There are three main measures of loudness that are used in literature: the sone, the phon and the decibel.

The decibel, dB was proposed for loudness as a logarithmic ratio and is usually called the *Sound Pressure Level* (SPL) when applied as a logarithmic ratio of the change in pressure. The sone was proposed as a unit of measurement for subjective loudness by Stanley Smith Stevens [17] in 1936. The measure works such that doubling the value of the sone, doubles the perceived loudness. The reference is such that 1 sone is equivalent to the loudness of 1 kHz tone at 50 dB SPL is judged to be twice as loud as a 40dB SPL with a loudness of 2 sones. This method of assigning an objective measure to a subjective concept is known as magnitude estimation.

The phon was also proposed by Stevens as a measure of the loudness level of pure tones and a way to relate the perceived loudness in sones to the widely accepted objective measure of loudness level, the decibel. The phon is measured using the reference point at 1 kHz so that sound pressure level in dB and loudness level in phons have the same magnitude. This method of assigning an objective measure to a subjective concept is an example of magnitude tuning or production.

There are several shortcomings to the definitions used above. While the definition of perceptual loudness should follow that of perception, the simple relationship proposed by Stevens does not hold for sounds below 40 dB [14]. Subjective measures are also prone to bias from experience, motivation, training and attention [18]. Similarly it has been shown that our perceptual evaluation of loudness is biased by the apparent source distance [19]. While the decibel is considered the SI unit for loudness level, it also has some problems as different signal types at the same SPL do not exhibit the same perceived loudness [9]. This phenomena is explored further in the following sections.

3.2 Loudness thresholds

According to Moore [14], the loudest sound we can hear without damaging our ears is at about 120 dB over the faintest sound which we can audibly hear. This implies that we have some loudness threshold, where the minimum absolute loudness we can hear is called the absolute threshold. There are two different ways to measure this threshold. The first involves measuring the sound pressure at some point close the the entrance of the ear using a small probe. The threshold obtained is called the minimum audible pressure (MAP). An alternative way to calculate the threshold is to use a loudspeaker in an anechoic chamber. The measurement is then made for the sound level once the listener is removed at the point of which the listener's head had occupied the space. This threshold is called the minimum audible frequency (MAF).

There are deviating ways to calculate these thresholds in the literature and each give rise to different values. MAP has been proposed by Killion in 1978 [20], whereas Robinson and Dadson [21] and ISO-389-7 proposed a measured MAF. Since MAP is calculated mainly using monoaural hearing and MAF is calculated using binaural hearing there is usually an average of 2 dB difference between the two thresholds due the threshold being lower when two ears are used. However, they follow a similar pattern.

The thresholds for MAP and MAF increase rapidly at high and low frequencies depending on the characteristics of the middle ear and are best in the mid-frequency range [22]. There are also notable changes in audibility in age ranges. The phenomena presbyacusis is known as the loss of sensitivity with increasing age and this is highly emphasised at low and high frequencies. Measurements from Whittle [23] showed that results from 50 Hz to 3.15 kHz mirror those in the high frequency range, but for 3.15 kHz the threshold was 120 dB. Johnson and Gierke [24] suggested that 16 Hz was around the lowest frequency humans were able to detect audible sounds, however under 16 Hz, we could still sense sound as distortion because of the way the sound still passes through our middle ears.

3.3 Loudness and its dependence on frequency and intensity

The first introduction of equal-loudness curves came from Fletcher and Munson in 1933 [25]. In 1956, Robinson and Dadson [21] revised these curves which led to the international standards, ISO-226 in 1987, then revised in 2003 [26]. The standard gives the equal loudness curves from 20 phons to 120 phons equivalent to the range of sound pressure level -10-130 dB in the frequency range 20 Hz to 20 kHz represented on a logarithmic scale. Figure 2 shows the equal-loudness curves from Pulkki and Karjalainen [9] which were adapted from Fastl and Zwicker [16].

Moore [14] drew two main implications from the loudness curves. First that the loudness of the different frequency components in a sound change as a function of the overall level. This means that if the sound is not reproduced at the same level, the 'tonal balance' can be altered. The second implication is that humans are more sensitive at high levels to low and high frequencies. Other dips can be attributed to resonances due to the shape of the ear.

Moore [14] noted that the accuracy of these curves and values should not be taken literally as in some curves a dip appears at 500 Hz and in some there is no such dip. These curves can clearly be seen to be biased by different stimuli where for example, some of the curves created by Fletcher and Munson in 1933 were found using pure tones with headphones, whereas the curves created by Robinson and Dadson were found using frontal sound incidence of tones using a loudspeaker in an anechoic room.

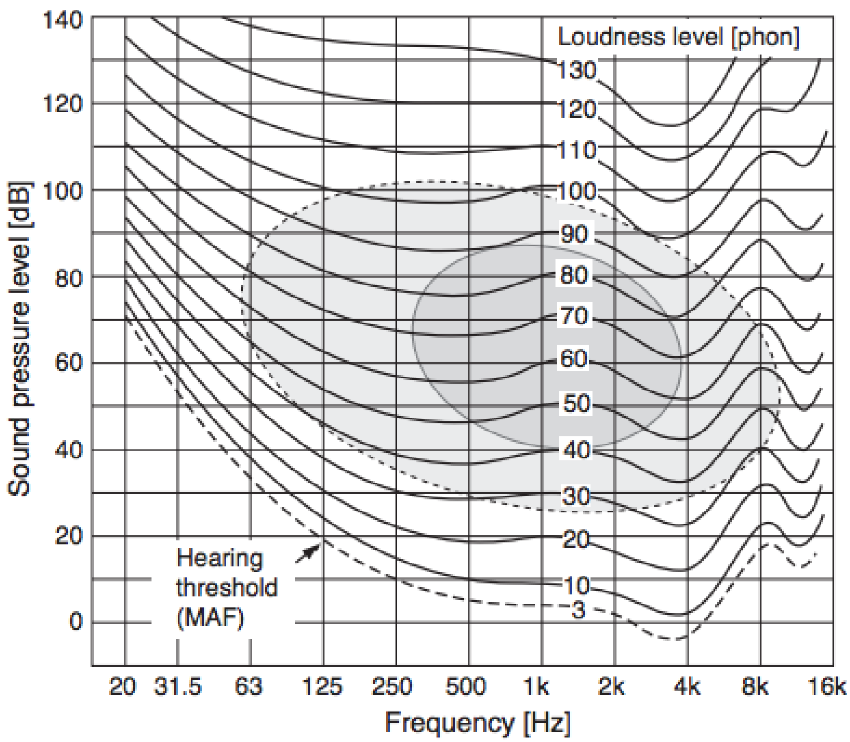


Figure 2: Figure showing the equal-loudness curves for different phon levels and their equivalent sound pressure level for different frequencies. The dashed line shows the hearing threshold (MAF). The range of audible acoustic music is shown in light grey. [9]

3.4 Weighting curves for loudness levels

Due to the frequency dependence of the perceptual nature of loudness, standards have introduced different weighting curves which aim to 'correct' the measured sound pressure level. There are many types of loudness weightings, but the most commonly used are A, B, C and D weightings which each account for different perceptual factors.

The A-weighting originated from the work by Fletcher and Munson [25] which aimed to account for the loudness sensitivity in high and low frequencies by adding more weighting to the mid-range. The weighting is based around the 40 phon loudness curve, which means that it has deviations from perceptual loudness at higher loudness levels. The 'B' weighting is similar, being based on the 70 phon loudness curve. The 'C' weighting is more linear since at high levels the frequency contribution is more or less equal for the loudness level. Finally, the 'D' weighting is similar to the 'B' weighting, only accounting for resonances in frequencies 2-5 kHz.

While the A-weighting curve is most commonly used as it is the most similar to human perception, the weighting is only valid at lower loudness levels. At higher levels, the C-weighting would be more accurate to use. Other weightings are less used in literature or in measures.

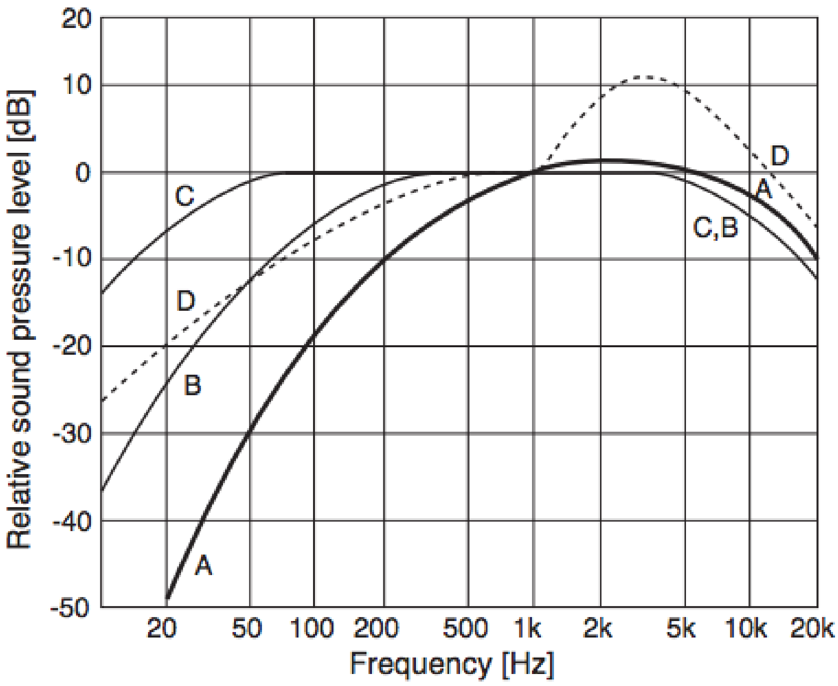


Figure 3: Graph showing the weighting curves as a function of the sound pressure level and frequency. [9]

3.5 Loudness of complex sounds and orchestral music

For complex wide-band signals, the absolute SPL can be derived from the sum of the individual intensities from the signals.

$$L_{\Sigma} = 10 \log_{10}(10^{\frac{X_1}{10}} + 10^{\frac{X_2}{10}} + ...10^{\frac{X_N}{10}}) \tag{7}$$

Where $X_1...X_N$ is the SPL of each of the signals up to the N^{th} signal.

In orchestral music, complexities arise due to dynamic ranges of the instruments and their production of harmonic overtones. Figure 4 shows the dynamic ranges as sound power levels for several different instruments in the orchestra as an average of their overall range. For example, at higher sound power levels, woodwind instruments sound 3 dB louder than the string instruments, whereas brass instruments sound 10 dB louder.

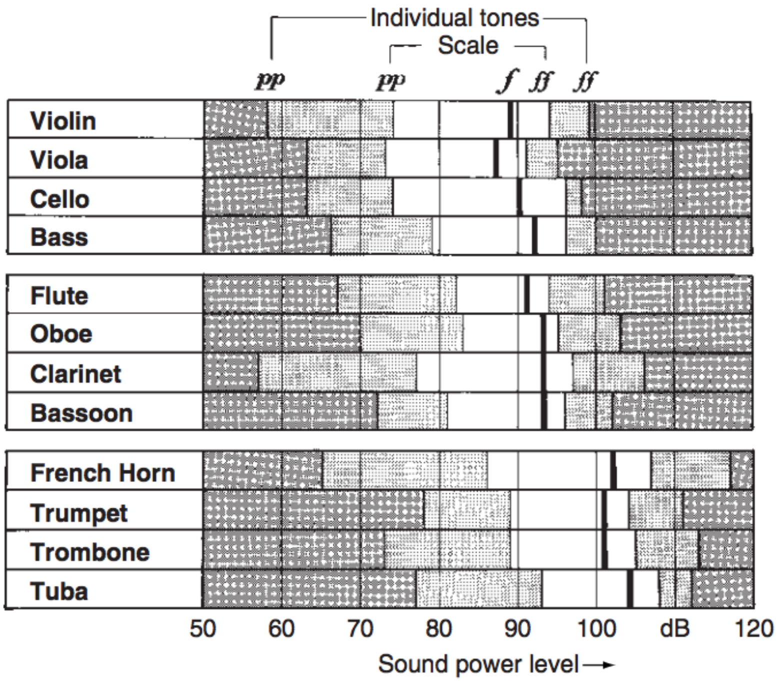


Figure 4: Figure showing the relative sound power level for different instruments of the orchestra and their relative dynamic expressions. [2]

Figure 5 shows the effect of the production of different harmonic overtones for different musical instruments. Overall, the effect is rather minor, but is most prominent for brass instruments and the clarinet. When the orchestra plays in a room, the effect is more enhanced since room acoustical properties affect the sound level and spectrum.

3.6 Other interesting acoustic phenomena in loudness

3.6.1 Measuring the intensity of loudness using Weber's Law

Another way to measure the intensity of sound as proposed by Moore [14] is using Weber's Law.

$$\Delta L = 10log_{10}\Big(\frac{I + \Delta I}{I}\Big) \tag{8}$$

Where ΔI is the minimum audible intensity that can be heard usually at around 1-2 dB [14]. This is more formally known as the just-noticeable difference (JND). For wide-band frequency noises, it has been shown that Weber's law holds since $\frac{\Delta I}{I}$

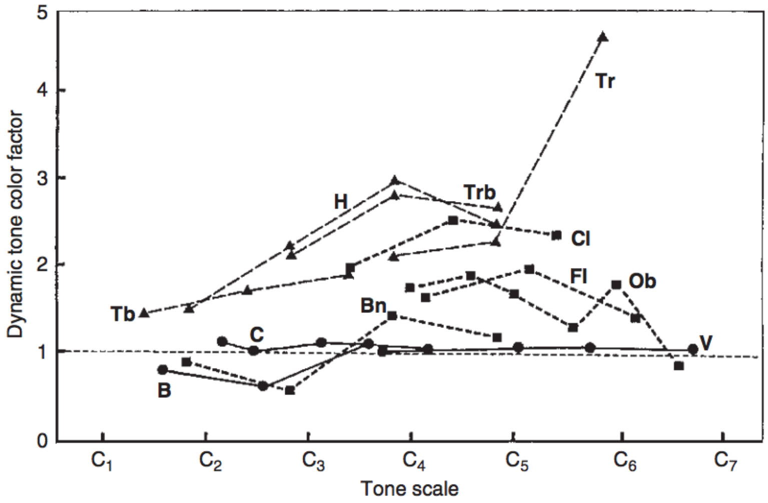


Figure 5: Figure showing level change of the strongest partial components for orchestral instruments depending on the pitch of the instrument. [2]

is approximately constant, however for pure tones, there is a deviation known as Weber's near miss [27]. This is where the constant ratio of Weber's law deviates to approximately around 0.9.

3.6.2 Spectral masking

Simple calculations of the sound pressure level explained in section 3.5 do not hold when the signals lie within certain frequency ranges (critical bands [9]), since they give rise to a phenomena known as masking. The idea of this is where two sounds are played such that one sound can be distinguished whereas the other sound is inaudible. According to Pulkki [9], spectral masking is where sound containing a specific spectral content makes the detection of another sound with a different spectral content harder to identify. This phenomena can even occur when the spectra don't necessarily overlap.

An example of this includes masking by noise, where Fastl and Zwicker [16] showed the effect of broadband noise on the masking threshold of different frequency tones. They showed that the thresholds for the pure tones were constant up to around 500Hz, after which they had an increase of around 3 dB per octave.

Further to this, spectral masking can occur between both pure tones and complex tones. In pure tones, if the frequency of the first tone is within a certain range of the frequency of the second tone, a phenomena called beating can occur. This results in changes in amplitude of the signal which are perceived as periodic fluctuations or roughness [16] [9]. The same applies for complex tones with multiple partials.

Pulkki [9] also identifies further different types of spectral masking such as co-modulation masking release [28] and information masking [29].

3.6.3 Temporal masking

While many of the discussions previously have applied to continuous sounds, it is also important to note that sound is masked in time also both before and after the sound, known as backward/pre- masking or forward/post- masking respectively. Backward masking has an effect usually around 5-10 ms before the onset for relatively low level sounds [9]. However, forward masking has an effect around 150-200 ms after the offset of sound for relatively high-level sounds and the masking is effective for a longer period of time. Explorations by Fastl and Zwicker [16] showed that the masking is dependent on length of the sound and the level, however the main shape shows a linear decrease for the first 5-10 ms, then tends towards silence after around 200 ms. It has also been observed that impulses arriving between a time window of 1-2 ms will be perceived as the same auditory event [9], which Pulkki identifies to be the 'best time resolution of hearing'.

3.7 Summary

In this section the pscyhoacoustic phenomena of loudness were explored. First the definitions of loudness and the scales used to represent both subjective loudness as well as the objective measure for the accepted international system of units were presented. Their advantages and shortcomings were also discussed and presented.

Loudness thresholds were also presented with consideration to possible reasons for the variations between the minimum audible pressure and the minimum audible frequency curves. The frequency and intensity dependence of loudness was also explored with reference to the equal-loudness curves proposed by Fletcher and Munson, Robinson and Dadson and their basis for forming the international standard, ISO-226:2003.

Weighting curves based on the equal-loudness curves that are commonly used in acoustic measurements practice were also discussed including A, B, C and D-weighting curves. Finally other interesting acoustic phenomena for loudness was discussed such as using Weber's Law to measure the intensity of loudness and the limitations of spectral and temporal masking.

In the section, it is clear to see that loudness is a complex phenomena with multiple dependencies such as frequency, intensity and dynamic variations. It is also difficult to find any research relating to the loudness of broadband or wide-band signals, as most of the previous research has been done on pure tones. However, it is mentioned by both Moore [14] and Pulkki and Karjalainen [9] that the hearing mechanisms are more complex and less well understood, and many of the phenomena mentioned is relevant for wide-band signals also. Complexities for wide-band signals also appear due to frequency masking if signals lie within a certain frequency range, known as the critical band [9]. In orchestral music, there is also a clear difference in the musical dynamic ranges for different sound power levels. Similarly, room acoustics emphasises harmonic overtones, especially at higher dynamic levels. The following section presents a similar analysis for the psychoacoustics of envelopment.

4 The psychoacoustics of spatial impression and spaciousness

On the contrary to the previous section on loudness, envelopment is a less understood term and appears less frequently in literature where it is more associated with other terms such as width, spaciousness and openness. In most literature, this comes under the umbrella term of 'spatial impression'.

Spatial impression is heavily influenced by our ability to localise sound. We do this by utilising different localization cues such as binaural and monoaural cues. In the following section, the key concepts of spatial impression are presented. Our ability to perceive geometry and source distance is also explored. Most of the work from this section is an accumulation of research from Pulkki and Karjalainen [9], Moore [14] and Blauert [30].

4.1 Basic concepts of spatial impression and localization

In the same way a concert hall impulse response is unique to the type of concert hall, we inhibit our own unique impulse response which means that perceived sound is different for each person. These depend on our own morphology, especially the torso, shoulders, head and external ear which all give rise to added reflections affecting the resulting sound. These discrepancies all contribute to what is known as the *head-related transfer function* (HRTF) or the *head-related impulse response* (HRIR). The study of HRTFs gives important conlusions into phenomena such as the bright spot and cone of confusion. As shown by Pulkki [9] in a study of a large number of HRTFs, a bright spot occurs around 1-4 kHz which results in an amplified response due to the sound arriving not only via the shortest route but also by diffraction around the head, causing multiple peaks in the response. Similarly, the head has been observed to show a cone of confusion protruding outwards from the ear canal. The phenomena is such that any sound source located on a cross-section of the cone renders binaural localization cues useless due to the same phase delays and transients.

Another interesting contribution comes from the asymmetry of the head which gives rise to important localization cues. These were first proposed by Grantham [31] and can be used for locating a sound source (localization) in both binaural listening (when the ears hear different signals) and in monoaural listening (when the ears hear the same signal).

The interaural level difference (ILD) is one of the main binaural cues. A difference in level occurs between the two ears due to the shadowing and reflection of soundwaves from the opposite side of the head. The phenomena is called scattering which is frequency-dependent and is stronger for higher frequencies. However, ILD is still audible across all frequencies most likely due to incoherent signals in the case of diffuse sound and the use of ILD at low frequencies as a distance cue.

Another of the main binaural cues is the interaural time difference (ITD), which occurs due to the finite velocity of sound due to the differing distances between the sound source and each of the ears. The ITD has a low frequency dependence [9] where below 700 Hz is the ITD slightly higher. Also, some irregularities are present above 2 kHz due to a lower coherence of binaural signals on the contralateral side of the head.

Interaural coherence, IC is the measure of the similarity of the signals arriving to the different ears. According to Pulkki and Karjalainen [9], it is not known as to whether this is perceived as an independent cue or as a result of the two cues presented above. Interaural coherence has an affect on localization, for example, when the IC is high such as in an anechoic chamber, the perceived source is point-like. If the IC is low such as in pink noise, then the perception is that the sound comes from many or all directions. In rooms, the IC is higher for the direct sound and lower for the early and later reflections if the source is in front [32].

For monoaural cues, there are two main types: the analysis of spectral cues and dynamic cues. For spectral cues, this is due to the dependence of the magnitude spectrum of the sound entering the ears on the direction of arrival, which occurs due to scattering and reflection caused by the head and pinna. Learning also has a strong effect on spectral cues, as shown in an experiment by Hofman et al. [33]. Another monoaural cue is the use of movement via rotation or tilting of the head. These provide a good understanding as to whether the source is in front, behind, up, down or inside the head.

While this section is mostly applicable for simple sound sources, for complex sound sources such as orchestral music our localization worsens. This is worse still in halls that have stronger reflections or are more reverberant. An interesting experiment was carried out by Pulkki and Santala [34]. The experiment consisted of asking subjects to identify which loudspeakers pink noise was coming from. For wide, dense sources the distribution was perceived almost correctly although biased towards the centre. However, for perceived distribution in the central area of the source did not match at all with the actual distribution of the sound source. For spatially complex scenarios, the subjects had no idea which of the loudspeakers the sound was coming from.

4.2 Perception of geometry and source distance

The perception of geometry in an enclosed space is quite limited. However, the listener has some mechanisms that can compensate for this. One of these mechanisms was first proposed by Wallach in 1949 [35] called the 'precedence effect'. The idea of this mechanism is that the ear is able to filter the effect of early reflections in an acoustic space to locate the direction of the source. Since in enclosed spaces the direct sound is the most dominant, the precedence effect utilises this. The effect is synonymous with the 'Haas effect' originated by Hermut Haas the same year [36]. According to Pulkki and Karjalainen [9], there are further mechanisms that help localization in room acoustics that have been proved in experiments such as the improvement of localization in rooms with an increasing number of trials, however their exact mechanisms are unknown.

For the perception of distance, several cues have been proved to be in work by Kolarik et al. [37]. Sound level, reverberation and frequency are all considered primary cues. Further to this, binaural cues such as the ILD are considered important when the source is closer than 1 m to the listener [9]. However, Kolarik [37] claims that the area is still relatively under-researched and not well-understood.

4.3 Defining envelopment

One of the biggest challenges presented for this thesis is defining envelopment. In 1967 [38], Marshall gave an interesting quote that relates spatial impression to envelopment from the manager of the Concertgebouw Orchestra of Amsterdam at the time. He described spatial impression as the difference of feeling inside the music rather than being outside looking at it through a window. Barron [39] describes this as different to reverberation where instead of removing the anechoic music, the spatial impression provides a certain degree of envelopment in the sound giving an impression of the distance of the source. Furthermore, in Long's 2009 published article entitled: 'What is so special about Shoebox Halls? Envelopment, Envelopment, Envelopment' [40], he describes how the most important factors for hall quality measures are listener envelopment, reverberation, diffusion (contributing to envelopment), sound strength, clarity and warmth. However, while appearing in the title, there is little explanation as to the definition of envelopment in his paper.

In literature, there are many conflicting subjective measures for envelopment. Mason [41] blamed the confusion of envelopment definitions upon a lack of standard vocabulary which hence triggered different terminologies. Similarly, he claimed that the definition of envelopment depended on the motivations of what the concert hall acoustics studies were trying to find out.

The main accepted subjective measures are the apparent source width (ASW) and listener envelopment (LEV). These concepts were introduced in simpler terms in section 2.1, however in the following sections they are explored in more detail with how to calculate them and their limitations in describing envelopment.

4.3.1 Apparent source width (ASW)

Barron and Marshall [10], defined ASW as the 'subjective sensation associated with the early lateral reflections'. Hence, as the lateral energy increases, the source appears widened or stretched. There are two possible ways to measure the ASW as discussed by Hidaka et al. [42]. The first is to measure the Lateral Energy Fraction, J_{LF} as discussed earlier. Another method is to measure the Interaural Cross Correlation Function which is the binaural measure of the difference in sound at the two ears. The *Interaural Cross Correlation Coefficient* (IACC) is then the normalized maximum of the function. Keet [43] was the first to relate this to the ASW with his coefficient being (1 - IACC). He showed that ASW widened approximately 1.5° for each increase in dB(A) sound level. Another finding was presented by Bradley et al. [44]. They found that at 100 Hz, a 10% change in the IACC was approximately equivalent to a 10 dB change in level. In Figure 6, Hidaka et al. attempted to make 'equal ASW' curves, analogous of the equal loudness curves presented in the previous section. They did this by presenting two synthesized cases: Case A consisting of direct sound with two lateral reflections and Case B consisting of direct sound and twelve lateral reflections. At loudness level 67 dB(A), the ASW of Case A and B were determined to be 30° and 65° respectively. By changing the levels of the reflections, the IACC values were also changed to match a reference signal of an equal ASW with 1 kHz. The biggest deviations between ASW were apparent at 500 Hz above, and hence Hidaka et al. concluded that the 125 Hz and 250 Hz octave bands were less important for determining the differences in ASW.

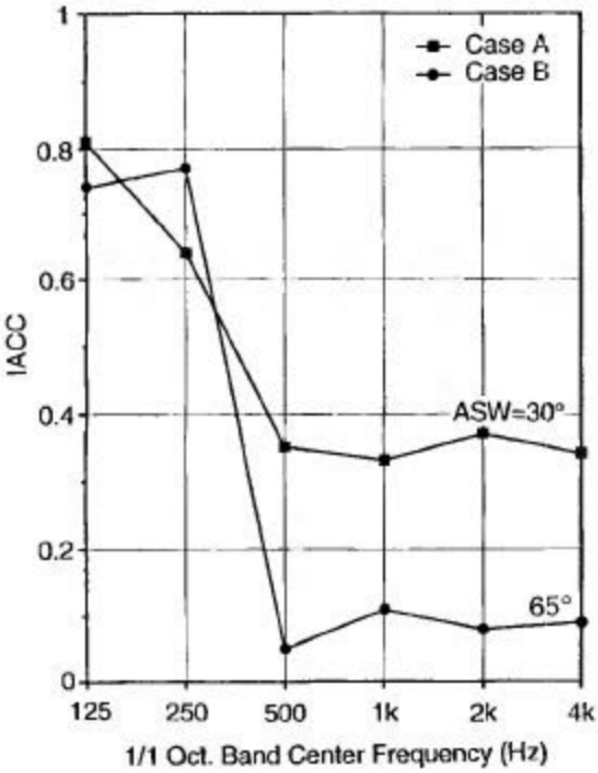


Figure 6: Subjectively determined equal ASW contours for octave band mid-frequencies [42]

4.3.2 Listener envelopment (LEV)

Listener envelopment describes the sense of being surrounded or enveloped by a sound [45] as a result of the late-arriving lateral reflections [46]. The accepted time for the reflections is at 80 ms. However, experiments from Souldore et al. [45] showed that the sum of the lateral energy after 105 ms has a better correlation of the subjective measure than objective calculations with after 80 ms. Sound arriving after around 120 ms are considered to give more of an echo effect. Also, the relation of clarity when high after 80 ms was shown to give a low late energy after 80 ms, showing that LEV decreases with clarity as well as the overall sound level.

Other interesting studies were performed by Griesinger [47], [48], [49] who described the listener envelopment as being split into three main components: the early spatial impression, the background spatial impression (BSI) and the continuous spatial impression (CSI). BSI is considered to be based on the amount of spatially diffused reverberant energy and CSI depends more on the ratio of direct sound to reverberant sound. Griesinger also showed that solo/light music was perceived more reverberant than thickly scored music due to its contributions to the background and continuous spatial impressions respectively.

Griesinger showed that envelopment was created after 150 ms when the late arriving reflections were spatially diffused. This caused an inhibition of the ear starting after 50 ms and ending around 150 ms. After this, the perception becomes the background where envelopment is considered.

Furthermore, as discussed in Kleiner [50], LEV has a characteristic of the reverberation arriving from multiple random angles, hence is related to the IACC. A lower IACC results in a more evenly distributed LEV. Due to this correlation of (1 - IACC) with J_{LF}, Beranek [5] derived calculations for the LEV using a direct substitution of the IACC after 80 ms with the J_{LF}. Cabrera [51] claimed that calculating LEV from the strength of late lateral energy is more effective than calculating from the IACC but he does not provide evidence to back this claim.

4.3.3 The relationship between LEV and ASW

Clearly relations between the LEV and ASW can be drawn due to their dependence on the IACC. As discussed in Kleiner [50] for example, Damaske [52] suggests that the IACC value should be smaller than about 0.1 for a good spaciousness. Whereas, if the sound directions become too many, then the sound becomes uncorrelated and results in a larger ASW. Bradley et al. [53] showed that the early and late arriving energy affect each other in their perception. For example, if one component is 10 dB or 12 dB higher then highest leveled component is the one that is predominantly perceived. Hence if ASW is more dominant then LEV is reduced. It has also been shown by Morimoto [54] that late reflections also contribute to ASW and also that early reflections contribute to the LEV.

4.3.4 Limitations of LEV and ASW in defining envelopment

Mason [41] also outlined some main limitations with the definitions of ASW and LEV. The first was that ASW doesn't refer clearly to a single source, section of instruments or entire orchestra and hence is difficult to compare across studies. Similarly, LEV is not solely perceived reverberation but also contains the level and spatial component as discussed by Bradley and Soulodre [55].

Further to this, Cabrera [51] stated that the ASW's main effect was on the auditorium environment and on the auditory image size of the performance. This means that spaces with strong lateral reflections have a larger ASW, however the measure disregards any expansions other than width.

Other spatial attributes are sometimes recognised in literature such as intimacy [51]. Intimacy is described as a sense of closeness or involvement in the performance where Cabrera [51] suggests this comes in a form of perceived distance. However, intimacy is heavily influenced by a number of other factors such as sound pressure level, frequency content, familiarity with the source, ratio of direct to reverberant sound energy and often as a result, the auditory distance is underestimated.

One of the main things not discussed in this section is the use of visual cues and its involvement with spatial impression. Many researchers have identified a clear interplay of difference and bias between audible and visual results of the three attributes. It has been stated that in perceived visual envelopment, differences arise between different concert halls but not within the same [56] and also that peripheral vision makes a contribution to spatial impression [52].

4.4 Summary

In this section we have explored the fundamentals of spatial impression such as the localization cues, geometry and source distance. The subjective measures of envelopment were presented such as listener envelopment and apparent source width. It was also noted that other subjective measures, while not as widely accepted, such as intimacy and clarity are often used to help define envelopment. Several other subjective concepts and ways to define envelopment were also introduced, although they are not universally used among acousticians.

From the psychoacoustic analysis, it is now possible to draw conclusions of perception of the two different attributes and see where influences come from. For loudness, this would be affected by thresholds and frequency dependence. For envelopment, it is clear to see that localisation and geometry have the biggest affect on perceptual accuracy.

The following section now explores related works in the field of concert hall acoustics related to perceptual testing of these attributes and their main findings.

5 Previous studies of perceptual factors in concert hall acoustics for loudness, envelopment and musical dynamics

In concert hall acoustics perceptual testing there are two main types of listening test: in-situ and methods involving the reproduction of the sound field. Initially used by Sabine [7], the most natural method of evaluating concert hall acoustics is in-situ, where the subject listens and assesses the hall in a real life situation. Questionnaires and interviews of the subjects were then often used to analyse the data. One of the main studies using this method was made by Beranek [1], who in 1962 made a comprehensive review of 100 different concert halls thereby, ranking the best 50. His rankings are still widely accepted by acousticians today. Researchers such as Hawkes and Douglas [57], Barron [58], Marshall [59] and Kahle [60] also used a combination of in-situ listening with subjective questionnaires in order to assess different perceptual attributes of concert hall acoustics. While the method gives the most 'real' acoustics, direct comparisons become problematic due to the limit of acoustic memory which has been shown to be around 10 s [61]. This shows that unless the review is taken almost immediately after the performance the data can unreliable. Further problems arise from ease of transportation and bias due to the visual or customer experience. Hence a newer method was needed to compare multiple concert halls.

As technologies in audio reproduction advanced, emerging methods in binaural and multi-loudspeaker reproduction led to an increased accuracy in recreating the acoustic environment of concert halls. This paved the way for new and innovative ways to test differences in concert hall acoustics, including the direct paired comparisons of particular attributes. Pioneering work by German research groups in Göttingen [62] and Berlin [63] used binaural technology for perceptual evaluations. Lavandier [64] mapped and identified subjective and objective measures but focused more on temporal variations. While their research was profound, neither groups studied auditory spatial impressions. Pätynen, Kuusinen and Lokki [65], [66], [67], [3] carried out a series of subjective listening tests using a state-of-the-art listening room. They used a number of different descriptive analysis methods [15], commonly used in consumer analysis for food and drink, in order to analyse results from subjective listening tests. Like the German research groups, they used pairwise comparison tests which allowed fast assessments, giving a better accuracy for their perception. However, many researchers have argued that this is an unnatural method since perceptual factors are rarely made between concert halls within seconds [68].

The following sections review the extensive literature on perceptual studies of loudness, envelopment and changes in musical dynamics in concert hall acoustics in order to place this thesis in the context of previous works.

5.1 Marshall 1967

Marshall was the first to introduce the concept of spatial responsiveness, where he was the first to notice the importance of the direction of sound reflections in concert hall acoustics. In 1967 [38], using a previously existing masking analysis, he compared two types of concert halls: rectangular, narrow, high halls and broad, fan-shaped halls to understand why spatial responsiveness was lacking in the latter. In doing this, he proved the subjective importance of early lateral reflections in concert hall acoustics. The most important was found to be at low frequency in lateral reflections for the most optimal sensation [59]. It can therefore be concluded that the direction of the sound in the early lateral reflections is the main cause for broadening the source image.

5.2 Keet 1968

In his 1968 paper, Keet applied stereophonic sound reproduction methods to live concert hall acoustics. In comparison to the single channel reproduction, Keet was able to deduce that the increased 'ambience' from the stereophonic reproduction associated with early reflections rather than the reverberation. Hence, the assumption became that time and amplitude differences between the ears created an apparent source broadening. Further to this, he related the 'ambience', directional distributions and spatial responsiveness to the same term and became the first to use 'spatial impression' and carry out experiments to show that the ASW was associated with early reflections. He also found that an increase in ASW was proportional to an increase in sound level and related to the incoherence between signals provided by two different loudspeakers.

5.3 Kuhl 1978

Barron and Marshall [69] describe Kuhl to be a researcher that is less quoted than he deserves. In 1978 [70], he suggested that the early reflections from the spatial impression were due to the space between the orchestra and the walls and ceiling appearing to be filled with sound. Most prominently in the paper, he shows that the sound level in the hall depends on both the dynamic level played and the sound strength at a certain location within the hall. For example, in a narrow hall that a dynamic forte giving an SPL of 93 dB can be attenuated 20 dB before the spatial effect disappears. Compared to a wide hall with a low ceiling, the spatial impression does not occur no matter the dynamic level.

5.4 Göttingen group 1974

As mentioned previously, in the '50s and '60s, anechoic recordings made in concert halls using dummy heads became increasingly popular to reproduce acoustic spaces for listening tests. These were first implemented in Germany where two research groups studied subjective loudness around the same time. In 1974 in Göttingen, researchers [62] attempted to make preference judgments between two different reproduced sound fields. However, they found that subjects were always biased to choose the louder of the two. Since the anechoic recordings could be re-rendered at the same subjective loudness, the study was carried out again. What was found was that the first 50 ms energy fraction, reverberation time and the short-term cross-correlation measures were the most influential on the subject preference.

5.5 Berlin group 1976

At around the same time, a research group in Berlin [71] carried further experiments on subjective loudness in acoustic spaces. Using Factor Analysis [15], they revealed three different main attributes: loudness, clarity and timbre. Furthermore, they found that the perceived loudness was correlated strongly with the objective measure of sound level and that a loud sound was preferred over a clearer sound.

5.6 Barron 1971-2015

In 1971, Barron explored the relationships of lateral and early reflections with the subjective attribute of ‘spatial impression’. He found that this attribute was produced when reflection delays of around 10-80 ms were present. He also found that this was mainly due to the lateral reflections rather than the reflections from the high ceilings, where the main effects of a ceiling reflection were: level change (similar to that produced by a side reflection), image shift and tone colouration. He concluded that the degree of spatial impression was due to the ratio of lateral reflections to non-lateral sound arriving within 80 ms of direct sound.

During 1971-1973, Marshall joined forces with Barron at the University of Western Australia, Perth in order to carry out further experiments to support the findings in his seminal works [10]. Together with Marshall, Barron published extensive further research on early and late lateral reflections and their results on spatial impression [58], [69]. In conclusion to their research, they incorporated a measure of two microphones that gave rise to the objective measure of lateral reflections: the lateral energy fraction [69].

Barron also studied the perception of loudness in concert hall acoustics. In 1988 [58], after a subjective study of eleven British concert halls, he used a subjective questionnaire to use 10 scales including ‘Loudness’ and ‘Intimacy’ which he found to be inter-correlated. For both attributes, the best objective measures found were the total sound level where for ‘Intimacy’, the total sound level was only relevant for the first 80 ms. In 1996 [68], he gave an extensive review of architectural design for loudness considerations in concert halls where he identified seat capacity and the maximum distance to the further seat to lead to the minimum total sound level. He suggested that while enhancing early reflections increases the sound level and loudness, it can risk false localisation of sources as well as unnatural tone colouration, and suggested that methods of diffusion were preferred. He also confirmed further affects of sound level reduction mainly being due to absorption.

5.7 Griesinger 1970-2018

Griesinger has extensively reviewed the objective and subjective measures of spaciousness and envelopment in his work, where most of his relevant work for this thesis has been discussed already in section 4.3. Griesinger commented mainly on the controversies of previous researchers in objective measures to subjective perceptions and proposed many alternative objective measures to those presented by the previous researchers and accepted ISO-3382 standards [49] [47] [49]. While praised by other researchers in their works for his observations, his objective measures are rarely used in practice by acousticians, other than in his own work.

5.8 Wettschurek 1976

Wettschurek [72] wrote the pioneering thesis determining the perception thresholds for a test reflection and its dependence on the overall listening level. He found that back reflections in acoustic spaces have the highest perception threshold for all of the tested listening levels and for all reflections, and that the perception threshold decreased linearly until around 40 dB. For higher than 40 dB, sensitivity to back reflections evens out however, above, reflections from the side increase linearly with listening level where by 80 dB the side reflections' sensitivity is almost 10 dB greater than that from the front or back reflections. This shows that perception and listening level is more complicated than first exhibited. It also suggests that manipulating reflections can have a big effect on factors such as the size or the spatial perception of the room. In doing this, he confirmed the findings made by Marshall [38] [59].

5.9 Bradley and Soulodre 1995-1997

Bradley and Soulodre [73] confirmed links between the source broadening and early reflections using subjective testing of sound fields simulated in an anechoic space. They introduced the concept of LEV and showed that this was related to the level, direction of arrival, and temporal distribution of late reflections and that an increase in LEV descreases the subjective response to ASW, hence suggesting that late lateral reflections have an impact on LEV. After further experiments, they proposed the late lateral energy as a measure for listener envelopment that is widely accepted in international standards [8].

In 1997 [74], Bradley and Soulodre determined to what extent the reverberation time and strength factor at low frequencies are able to determine the perception of bass in concert halls. In their experiments, they made ten listeners rate their perception of the strength of the bass content in music samples where sound strength G and reverberation times in low frequency varied. They saw that the perceived bass level increases almost linearly with an increase in the strength, G in the lowest frequency band. They also noted that the effect of a large change in reverberation time was negligible. For late sound levels, an increase in bass was perceived. Similarly, the direction of arrival of the low frequency sounds was found to have a small effect on assessment of bass content in the sounds. This study was important in showing the relationship between bass and envelopment.

In later work, Bradley carried out further studies with other researchers on the topic of LEV and ASW in explaining spatial impression. In 2000 [53], Bradley et al. carried out subjective studies to estimate ASW and LEV for 16 different concert halls. They found that LEV was a stronger influence on the spatial impression than ASW.

5.10 Beranek 1962-2010

While Beranek's most prominent review of concert hall acoustics [1] has been mentioned several times already throughout this thesis, he also gave a good insight into envelopment and loudness measures. In 2010, Beranek [5] introduced the way to quantify LEV. He used the work of Soulodre and Bradley [74] to derive calculations for the LEV from sound strength and late lateral energy using IACC. For most halls, the observation was that the LEV was highly correlated with overall strength. It was also shown that an increase in G was highly correlated with an increase in the reverberation time. For non-shoebox halls the correlation was shown to be almost zero due to one or more of upper side walls in non-shoebox being all/nearly all covered by seated audience areas and the direct and early sound being radiated from orchestra before being absorbed. Beranek also noted that the strength in shoebox concert halls was about 3 dB higher than in non-shoebox halls due to orchestra position in relation to audience. Also with same reverberation times, G was found to be higher in shoebox than in non-shoebox halls as bass sounds are further augmented.

Another of Beranek‘s main claims was that ITDG was a measure of acoustic quality of a concert hall. Before 'Spatial Impression', and 'Sound Strength' calculated from early lateral reflections, ITDG was seen a substitution since the terms did not yet have accepted objective measures. Beranek initially drew correlation between ITDG with subjective intimacy [1]. He also related the intimacy to the overall loudness level and that a short ITDG means more reflections occur in the first 80 ms, hence producing a greater subjective intimacy. While he initially proposed ITDG to be an objective measure of intimacy, he later retracted this statement [75].

5.11 Dick and Vigeant 2014-2019

Dick and Vigeant studied mainly the perception and objective measures of LEV and suggested that there could be improvements to objective measures. They studied the differences in microphone arrays in measuring the J_{LF} and L_J for concert halls commenting that the figure-of-eight microphones used during calculations have a non-ideal frequency response and directivity pattern. They found that when using a 32 element spherical microphone array, the results agree at higher frequencies than 500 Hz. However, at low frequencies, deviations occur due to differences in directivity patterns. Further to this, they carried out subjective testings using an ambisonic reproduction of the acoustic space through 30 loudspeakers to determine the objective metrics associated with LEV [76]. They found that the highest correlations were with G, the late sound strength and L_J. They found that metrics associated with spatial aspects such as J_{LF} among other metrics, had no relation to LEV, however they determined that perhaps halls with differing geometry may attribute to better correlation.

5.12 Pätynen and Lokki 2011-2016

Pätynen and Lokki's work on concert hall acoustics has been extensive where their superior audio reproduction methods and listening room has led to a number of subjective testings on a wide variety of perceptual factors in concert hall acoustics. This has allowed for similar direct pairwise comparisons between attributes as shown in earlier works by the German research groups using manipulations of different perceptual attributes.

In their paper in 2014 [66], they showed that rooms which had early lateral reflections are able to enhance the musical dynamics in the concert hall and highlighted that the perception of musical dynamics is often ignored in previous works. This led to their next work where they aimed to explored the musical dynamic responsiveness in concert halls. Pätynen and Lokki [3] used further paired comparisons of three rectangular and non-rectangular halls to assess the concert halls' responses to varying orchestral dynamics. It was the first research where non-continuous music was used with dynamics that were not just natural changes in the piece of music. They found that the dynamic responsiveness of the hall was foremost related to the perception of loudness in a concert hall and closely related to the 'width and envelopment' of the concert hall. Similarly, clarity and brightness were two other subjective attributes that related to the dynamic responsiveness. Pätynen and Lokki also analyzed the typology of concert halls to demonstrate that shoebox-type rooms exhibit more perceptual factors and that these are further increased when the distance to the orchestra is increased.

5.13 Kahle and Green 2018

Kahle and Green [4] carried out similar measurements to Wettschurek [72] in real-life concert halls by applying dynamic spatial response filter to three dimensional impulse responses which were measured in the Nouveau Siècle concert hall in France. Their findings mirrored the work of Wettschurek where they found that perception thresholds for reflections from the front and back of the hall decrease as the listening level increases. They also drew similar conclusions that increasing the number of side reflections makes the hall 'wake up' and becomes more perceivable. Their findings confirmed those of Keet [43] when applied to concert hall acoustics, where the reflected sound from the sides are more audible than the front and back when the overall sound level is higher.

5.14 Summary

As shown in the previous sections, loudness is one of the most important of these attributes and that changes in loudness is one of the attributes that humans are most sensitive to and are good at recognising. Therefore, it is interesting to note the absence of literature for understanding the changes in loudness and envelopment.

Motivation was heightened in the 60s and 70s with the introduction of newer technologies in Germany, however interest dwindled in the area of concert hall acoustics until picking up again by Lokki's team in 2015. Perhaps this is due to developments in the understanding of how humans respond to loudness and envelopment, and our increased understanding in the definitions of these terminologies.

The work carried out in this thesis is a continuation of the work began by Pätynen and Lokki [3] who first explored the dynamic responsiveness of several different concert halls (many are the same as used in this thesis). The work was also inspired by the findings of Kahle and Green [4], who studying listening thresholds for reflections from the front, back and lateral of different concert halls. They deduced that the hall would be more perceivable since the hall 'wakes up' as the listening level increases. It was therefore deduced that for this experiment, the perception of the two attributes studied would be easier to answer at louder dynamic levels than softer dynamic levels.

6 A study of the perception of musical dynamics using loudness and envelopment in different concert halls

The following section describes the listening test carried out to evaluate perceptual differences in loudness and envelopment for changing dynamic level in a number of different concert halls and seat positions. The section first begins by introducing the concert halls tested as well as some of the architectural elements and key features identified by Beranek [1] and Pätynen and Lokki [77]. The next section explores the methodology of the listening test: the auralization and reproduction; the listening room; the procedure and experimental design; the analysis process used to evaluate the data and the objective measurements made. In the results section, the listening test results are presented along with the results from the interviews. Following this, these are discussed in more details with comparisons to the architectural elements, objective parameters and psychoacoustic properties of the elements mentioned. Finally, the key findings of the test are summarized and further improvements and advancements are suggested.

6.1 Concert halls used for the listening test

The concert halls used in the experiments were Amsterdam Concertgebouw (AC), Cologne Philharmonie (CP), Munich Herkulessaal (MH), Berlin Philharmonie (BP), Berlin Konzerthaus (BK), Musiikkitalo, Helsinki (MT) and Musikverein, Vienna (VM). The following contains short descriptions of each of the concert halls and their main features as identified by Beranek [1]. CP and MT were inaugurated after Beranek's review, hence they do not appear. An extensive analysis has been found by Pätynen and Lokki [77] for MT, while for CP no analysis could not be found except standard reviews by conductors, performers and audience members.

6.1.1 Musikverein, Vienna (VM)

Designed and constructed in 1870, Musikverein, Vienna is a concert hall that is rectangular in shape and exists at a relatively small size. However, its high ceiling contributes to the long reverberation time as well as its irregular, plastered interior surfaces. The stage is risen slightly above the floor level at around 1 m.

6.1.2 Amsterdam Concertgebouw (AC)

Beranek's [1] competing favourite acoustics hall, Amsterdam Concertgebouw was designed in 1888 and is also a rectangular hall but is wider than Musikverein, Vienna. It is also unsual in that 20% of the audience are seated on a steep stadium situated behind the orchestra. Its stage lies the highest above floor level, situated at around 1.53 m. The hall also emphasises its acoustics with its flat floor and removable seats, irregular walls and coffered ceiling. The hall is known to be strong in bass and that the articulation is better on the balcony than on the main floor.

6.1.3 Munich Herkulessaal (MH)

Munich Herkulessaal was constructed in 1953 and is also rectangular in shape, however it does not appear in Beranek's [1] superior listings. This is due to lack of upper side wall irregularities, little hard surface areas and large columns which create a lower reverberation by reducing the lateral reflections. The flat floor also results in a loss of the high frequency sounds due to the absorption of the seating positions. However, the hall boasts many of the surperior architectural attributes as the other halls such as a coffered ceiling.

6.1.4 Berlin Philharmonie (BP)

Lokki [78] considered Berlin Philharmonie to be a more worse concert hall. It was constructed in 1963 and was the pioneering 'vineyard' style halls, however its architectural elements result in problems for the acoustics. For example, Beranek [1] describes the hall as being split into different blocks where each block receives a very different type of acoustics. The first receives direct sound where many seats receive early lateral reflections from the side walls and walls behind. This is also added to by the terrace which gives further reflections for the middle-seated audience and musicians. Upper blocks receive early reflections from the convex ceiling. As a result the orchestral sound appears heavily distorted where for example the trumpet sound radiates forward, the french horns radiate backwards. There is also an emphasis from higher register instruments which similarly project forward.

6.1.5 Berlin Konzerthaus (BK)

After Schauspeilhaus was destroyed during world war II after its construction in 1821, Karl Friedrich Schinkel was hired to restore the hall and it was renamed the Konzerthaus, Berlin. The hall is another small, rectangular-shaped hall and again boasts many preferred architectural acoustics elements such as coffered ceiling, high side walls and sound diffusing surfaces.

6.1.6 Cologne Philharmonie (CP)

Cologne Philharmonie has rarely been reviewed in concert hall acoustic literature since its inauguration in 1986. Another rather large hall, the audience seats extend to be also around the back of the orchestra. This makes for a rather arena-style hall and hence it is often referred to in works as 'Fan' rather than Vineyard-style.

6.1.7 Musiikitalo Helsinki (MT)

Helsinki Music Centre was inaugurated in September 2011 as a vineyard-style hall and is the most modern of the concert halls used in the listening test. The hall is larger than Berlin Philharmonie but with far less seating, however it exhibits similar forthcomings to those exhibited by other vineyard-style halls such as a dull reverberation time and a loss of bass at further distances from the orchestra position [77]. The level of bass in the hall is adequate near the orchestra but becomes lost at further distances. The noise floor is very low with a clear sound image leading to a good spatial separation of instruments, especially at quieter dynamics.

6.1.8 Summary of the concert halls used for listening test

Measurements of the impulse responses were recorded at positions from the stage: front (11 m), middle (19 m) and first row of the balcony (23-25 m) for the 7 different concert halls. The positions of the different halls are given in Figure 7, where the balcony position is shown to be elevated by the red star.

Table 2: A table showing each of the concert halls as well as their abbreviations, shape, volume, number of seats, strength and their early decay time [3]

Hall	Abbr.	Shape	V $[m^2]$	N	G $[dB]$	EDT $[s]$
Amsterdam Concertgebouw	AC	Rectangular	18 780	2 040	2.8	2.4
Munich Herkulessaal	MH	Rectangular	13 590	1 300	2.9	2.1
Vienna Musikverein	VM	Rectangular	15 000	1 680	4.1	3.1
Berlin Konzerthaus	BK	Rectangular	15 000	1 575	2.7	2.1
Berlin Philharmonie	BP	Vineyard	21 000	2 220	2.1	2.0
Helsinki Music Centre	MT	Vineyard	24 000	1 700	1.4	2.0
Cologne Philharmonie	CP	Fan	19 000	2 000	1.9	1.6

6.2 Methodology

The subjective listening test was carried out using a pairwise comparison test to answer two questions: "Which is louder?" and "Which is more enveloping/wide?" The samples were presented to the subjects in pairs where the pairs compared each of the halls at different listening positions and different musical dynamics. This resulted in 72 results per participant for 4 x 4 x 3 halls, dynamic levels and positions, respectively. The test was repeated for another 4 halls where BP was used in both tests for reference. The experimental design was such that the subject must always answer one hall, even if they noticed no difference. The participants were free to move quickly between the two musical samples where the position in the sample was continuous. The subjects were a range of expert listeners from the Acoustics and Audio Technology department at Aalto University.

The preparing of the recordings and samples used a method similar to that of Pätynen and Lokki [3]. The reproduction of the recording in the acoustic space was in a 39-loudspeaker array listening space. Recordings were made using a sound level meter to record the LAeq and adjustments were made to the samples in order to make them approximately at these levels. The meter was positioned using a tripod to be approximately at the listening position of the subject. Measured LAeq levels made in the listening room can be found in Appendices A and B. The following subsections explain the process of the listening test and procedure in more detail.

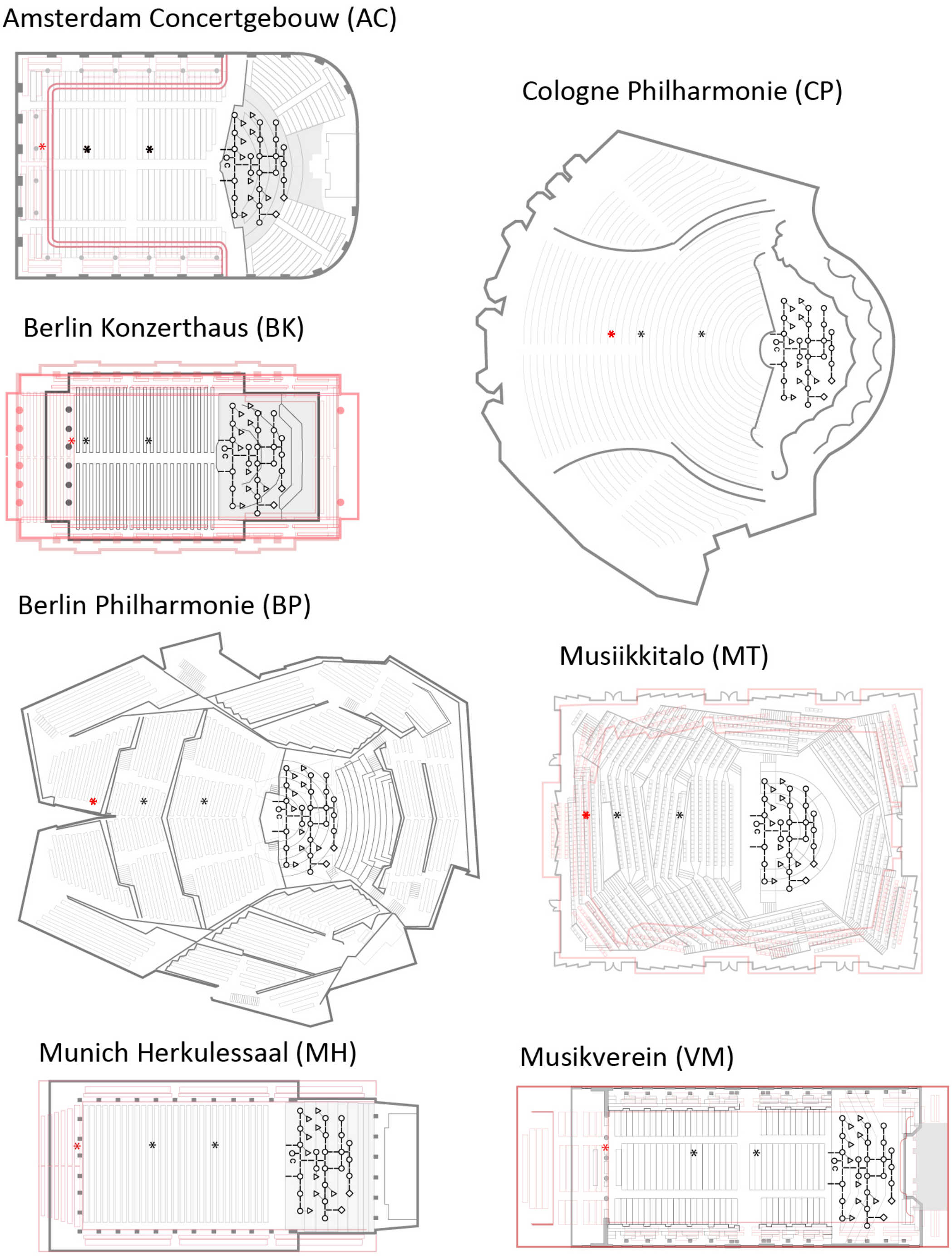


Figure 7: The figure shows the blueprints of the seven different concert halls. The red star shows the position but elevated at the balcony position.

6.2.1 The musical signals

The musical excerpts were taken from the Lemminkäinen Suite, Op. 22, First poem: *Lemminkäinen and the Maidens of the Island* by *J. Sibelius* (1865-1957). The excerpts were as follows:

- *Pianissimo*, bars 553-558, 10 s.
- *Mezzo piano*, bars 257-264, 11 s.
- *Mezzo forte*, bars 172 -185, 20 s.
- *Fortissimo*, bars 514-524, 16 s.

The excerpts were taken where natural musical dynamic steps were exhibited in the musical piece. This corresponded with LAeq levels when reproduced in the acoustic space of around 60 dB (*pp*), 70 dB (*mp*), 80 dB (*mf*) and 85 dB (*ff*) where each were to be approximately within 2 dB per compared hall in accordance with Karjalainen and Pulkki [9]. Adjustments in volume were made only to the anechoic recording and not the reproduction in the acoustic space so the tonal balance was not disturbed.

The instrumentation varied slightly for the musical dynamics. The instrumentation was as follows:

- *Pianissimo*, Flutes, bassoons, horns, trombones, violins, viola, cello and double bass.
- *Mezzo piano*, Flutes, oboes, clarinets, bassoons, trumpet, bass drum (tuned to C), violins, viola, cello and double bass.
- *Mezzo forte*, Flutes, oboes, clarinets, bassoons, horns, violins, viola, cello and double bass.
- *Fortissimo*, Flutes, oboes, clarinets, bassoons, horns, trumpets, trombones, bass drum (tuned to C), violins, viola, cello, double bass.

6.2.2 Auralization process and listening room

Recordings of the concert halls were made using 33 loudspeakers connected to 24 independent channels with a loudspeaker orchestra as the input source [79]. The layout of the loudspeaker configurations for each of the different concert halls is shown in Figure 7. This follows a typical American seating [3]. Care was taken especially to emulate the directivity pattern of the string section, where the nine loudspeakers representing them were connected to nine loudspeakers on the floor which faced upwards [3].

The room response was then measured with a 50-VI 3D intensity probe with 3 co-centric, phase-matched pairs of omni-directional microphones across the x, y and z axes with an opposite distance of 10 cm [3]. The impulse response of the concert hall was then measured using 48 kHz sample rate and a logarithmic sine sweep. These 6 impulse responses were then analysed using the Spatial Decomposition Method [80] which estimates a direction-of-arrival for each of the samples in the impulse response. It does this by analysing the time difference in arrivals of the microphones in short time windows. The method emphasizes the spectral fidelity of high frequencies which leaves a slight reduction in the spatial fidelity [15], however the reduction is negligible for these tests. The anechoic recordings were then convolved with the reproduction channels. The sound is then reproduced by using the direction from the SDM to assign the instant pressure of the impulse response to the nearest loudspeaker in the array. The array consisted of 39 loudspeakers that formed the listening room test space. A summary of the methodology is presented in Figure 8.

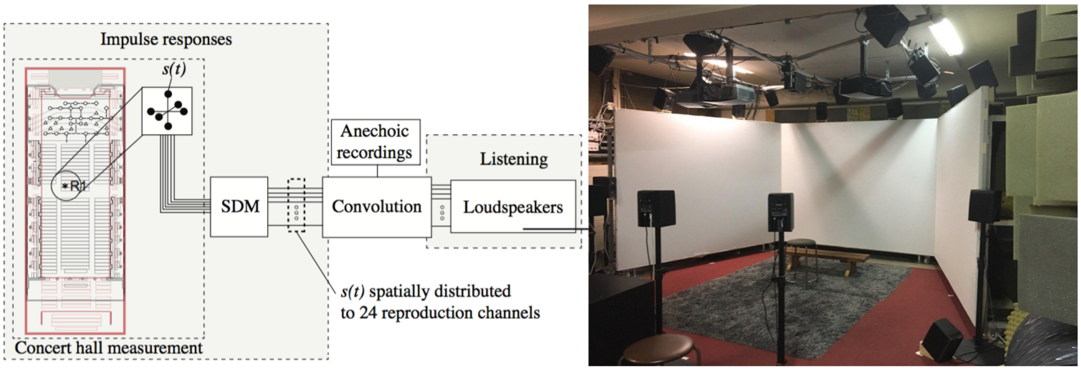


Figure 8: The figure shows the auralization process for the whole orchestra and all samples for testing. Adapted from [3].

The listening room was a rectangular, semi-anechoic treated space. The room was dry and acoustically treated with 80 cm thick foam wedges on the walls. The floor was also covered with carpet which resulted in a short reverberation time of the acoustic space (around 0.11 s at mid-frequencies [3]). This is considered not to bias listening accuracy or differences in stimuli. The frontal loudspeakers existed behind a transparent screen in order to not influence the subjects for their listening position. The room consisted of 40 loudspeakers (Genelec 1029) of which 39 were used, where the configuration was according to the audio control booth for mixing and mastering. The listening room can be seen in Figure 8.

6.2.3 Procedure and experimental design

The test consisted of two main parts: a pairwise-comparison listening test and an interview. First, the participant was seated in the centre of the loudspeaker array and a brief explanation was given on how to use the GUI. The participant then freely listened to the two samples and answered the questions: 'Which is louder?' and 'Which is more enveloping/wide?' Each question was repeated until all 72 of the different samples had been compared: 4 x 4 x 3 of the concert hall, dynamic level and seating position, respectively. The order of the two questions was alternated between participants and the order of appearance of the sample pairs was randomized for each participant.

The listening test was designed in MaxMSP [81] using the MIRA software for iPad. The user interface worked such that the participant had two 'Play' options and two answer pad options. They could then select the play option so that the position in the track was continuous when comparing the signals. Once they were happy with which was louder or more enveloping/wide, they selected the answer pad under the sample and selected 'Next'. This was repeated until all 72 tests had been compared. An identical user interface with the other question was then presented and the test was repeated.

Once the participant had completed the listening test, the subject then participated in a short interview about the listening experience. The interview was only carried out to the 12 subjects who participated in the first study since many of them also took part in the second study. The interview questions were as follows:

- **Question 1** How did you find the user interface?
- **Question 2** What is your definition of envelopment?
- **Question 3** Was the loudest hall always the most enveloping or not always the case?
- **Question 4** Was it easier to answer the questions when the sound level was louder or quieter?

A disclosure document was also signed that explained the process of the test and the handling of data. The disclosure document also meant that the participant could leave the test at any time, although no participants left the test unfinished.

6.2.4 Analysis of subjective listening test results

The data was organised and analysed using both MATLAB [82] and R [83] programming languages. The results from the listening test were analysed using a ranking system evaluated across each of the different subjects. This was a preferred analysis method over other statistical methods in order to better identify changes between less agreed answers among several subjects, as otherwise the less agreed answers would appear as minute differences. The ranking was used to order the concert halls in their perceived loudness and envelopment/width for the different hall positions and dynamic levels.

The objective measurements were computed using MATLAB from the recorded impulse responses of the 7 different concert halls. The L_J, G, C_{80}, T_{20}, $EDT(\tau)$ and J_{LF} were computed from these impulse responses in accordance with the ISO3382 [8]. The measured differences in LAeq values were also calculated for each of the dynamic levels.

6.3 Results and Discussion

The listening test was carried out in two different sessions. The first experiment compared the four concert halls: BK, BP, MT and VM. The second experiment compared the four concert halls: AC, BP, CP and MH. The results presented in the following section are presented for the two different sessions.

Results are presented for the objective parameters: late lateral energy, lateral energy fraction, strength and changes in the LAeq. Also presented, are the results from the listening test from 12 participants in the pairwise comparison of concert halls in determining which hall was considered to be louder or more enveloping. Finally, an overall summary of the answers given during the interviews from the 12 participants about the test itself and the user experience are presented.

6.3.1 Objective parameters

The following objective parameters for the subjective level, subjective clarity, subjective reverberance, apparent source width and listener envelopment were calculated from the measured impulse responses of the 7 different concert halls. These were calculated using MATLAB programming language. Figures 9, 10 and 11 show the objective parameters calculated for the close, middle and balcony positions, respectively. The Figures show the results of the concert halls which are presented using a dashed line for the first sitting of tests and using a continuous line for the second sitting of tests. Since BP was used in both sittings, the results for this hall are represented using a continuous, black line.

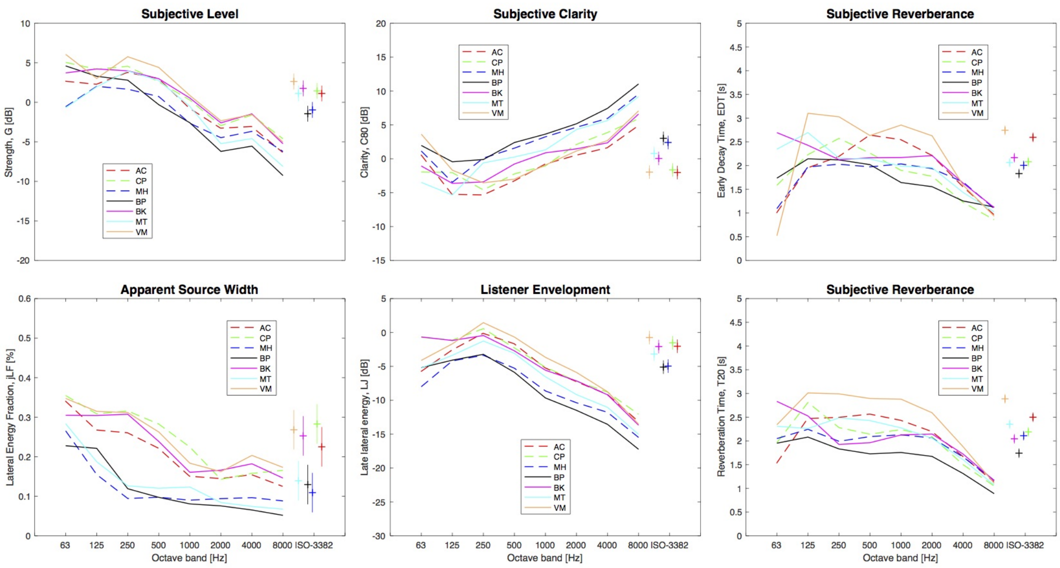


Figure 9: The figure shows the objective parameters calculated for the seven different concert halls at octave frequency bands and the ISO3382 average of the mid-frequencies at close position.

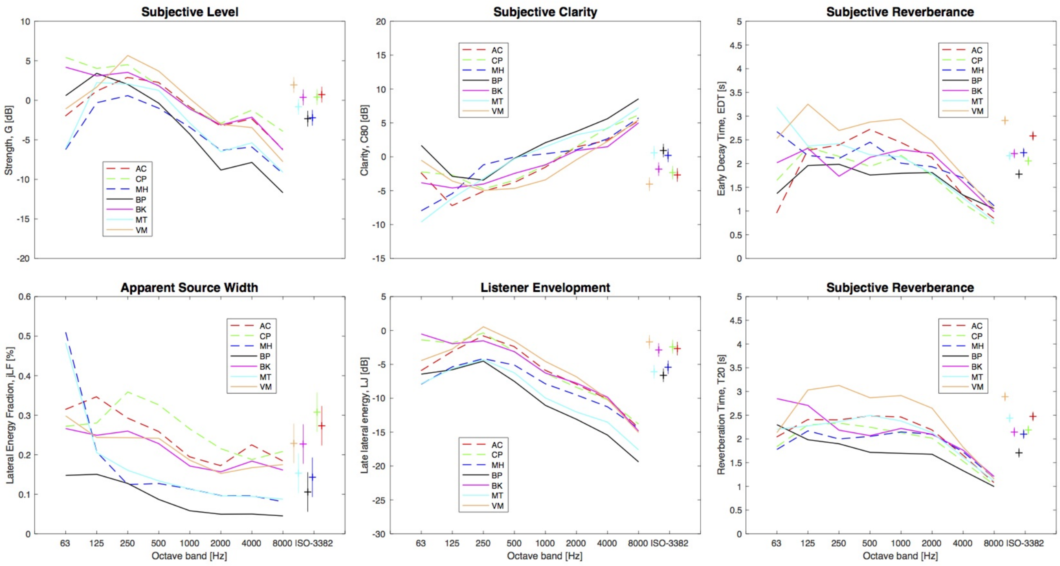


Figure 10: The figure shows the objective parameters calculated for the seven different concert halls at octave frequency bands and the ISO3382 average of the mid-frequencies at mid position

Considering first Figure 9, the graphs show that the resulting subjective level is considerably higher in VM than in the other concert halls. There is around 2 dB difference between 250-500 Hz that should be susceptible to participants. The subjective reverberance is also significantly higher for VM than the other concert halls and BP can be identified as having the lowest at most frequencies. The rest of the halls have close reverberation times at low and high frequencies but are more distinct at the mid-range.

It is also interesting to note that the objective order of level and envelopment is relatively similar according to the ISO3382 values, however considering their dependence with frequency, the order alternates. While BP and BK remain the least, MH, MT, CP and AC vary in their order for different frequency ranges. Apparent source width is perhaps the most clear and separated for the different frequencies where there is two clear groups: VM, MH, CP, AC; and BK, BP, MT. At different frequencies, again the order changes. The clarity works as an almost direct reverse of the listener envelopment.

For Figure 10 at the mid-position, the subjective reverberation separation is more obvious between VM and BP with the other concert halls. Otherwise the pattern is similar. For the subjective level, VM is the loudest at the mid-range frequency whereas the rest of the frequency dependence is more spread vertically. For the

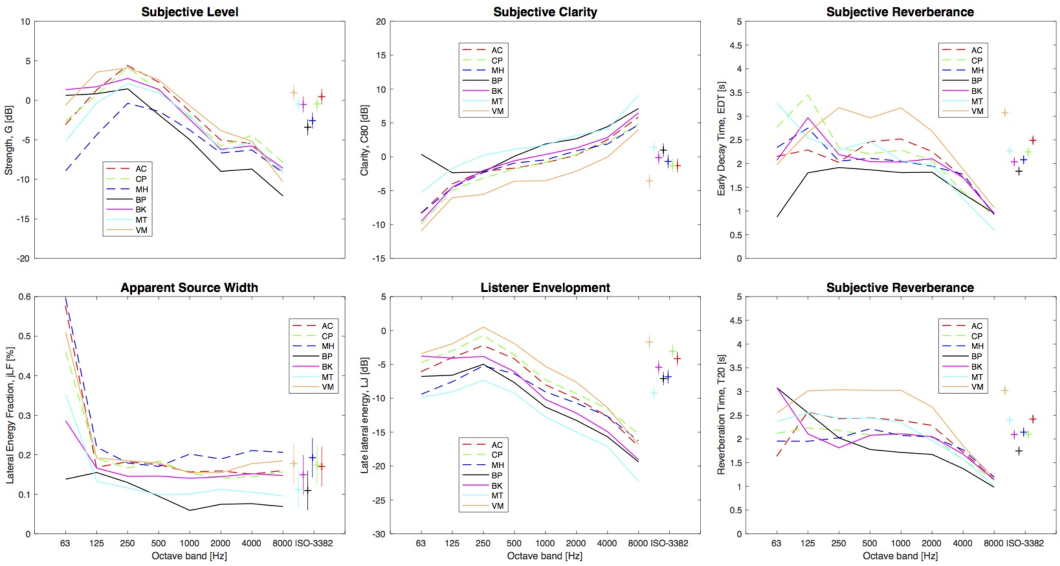


Figure 11: The figure shows the objective parameters calculated for the seven different concert halls at octave frequency bands and the ISO3382 average of the mid-frequencies at the furthest position

apparent source width, at low frequencies the values begin to tend to 1 and the halls become more distinctive with respect to frequency. The groups are identified easier in the frequency range, whereas the ISO3382 values appear as 3 groups: BK, BP; VM, MH, MT and CP, AC.

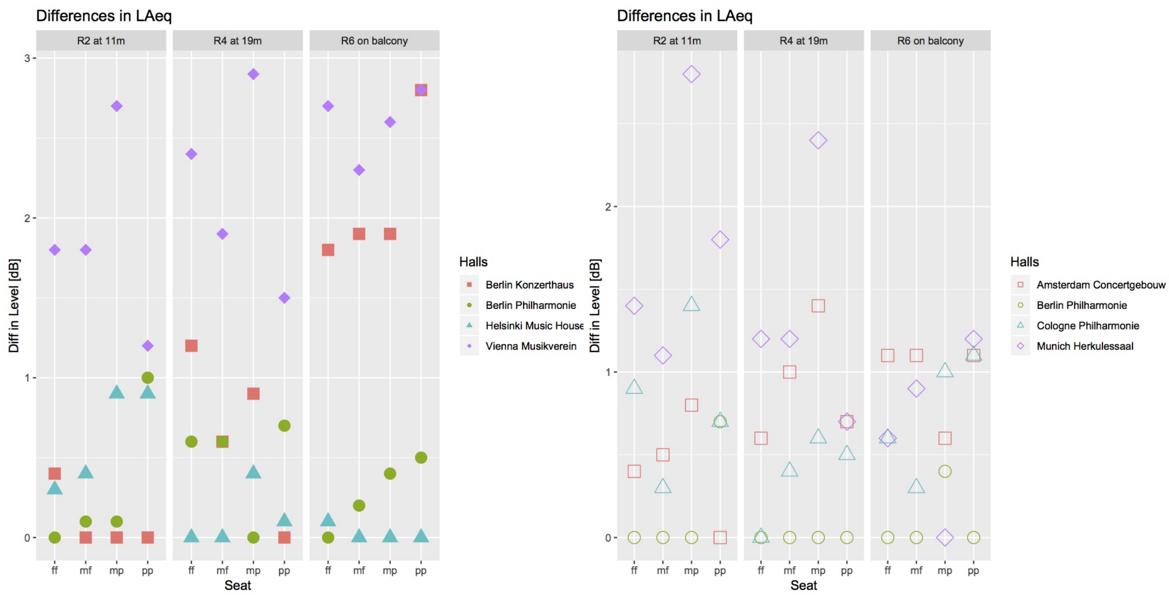
For the mid-position, the subjective level follows less the listener envelopment where now the order for AC:CP and MT:BK:BP is switched according to the ISO3382 values. The clarity again works as an almost direct reverse of the listener envelopment. At mid-range both of these measures‘ order follows similar to the ISO3382, however at the extreme low/high frequencies the order is randomized.

For the final balcony position the results are shown in Figure 11. As expected, the subjective revereberance is much more clear in comparison between the VM and BP where VM is significantly more reverberant than the other halls.

The subjective level and listener envelopment orders are now much more different. The subjective level is similar in order for ISO-3382-2 to that of the mid-position. Whereas, for the listener envelopment, MT now has a significantly lower Late Lateral Energy and is significantly higher for VM. At the mid-frequency range, it is clearer to see the distinctions of the listener envelopment than using the ISO3382 values.

The subjective clarity of VM is significantly worse than the other concert halls across all frequencies and the parameter no longer follows the reverse order of listener envelopment according to the ISO3382. At mid-frequency range, it is also difficult to determine which has the best clarity between the other halls and this order differs significantly again at low frequency. The apparent source width is much less easier to distinguish differences between where nearly all converge to 1 at low frequencies.

For all of the Figures, it is interesting to note that the ISO3382 values clearly differ at the different frequency ranges and are usually representative only in the range of 250-2kHz. This indicates a required objective measure for the 'frequency content' or 'timbre', which would allow for measures at extreme low and high frequencies to be analysed. This is especially important when considering that in the following section, most people noticed considerable differences in the research questions at low frequencies; when the concert hall had more bass.

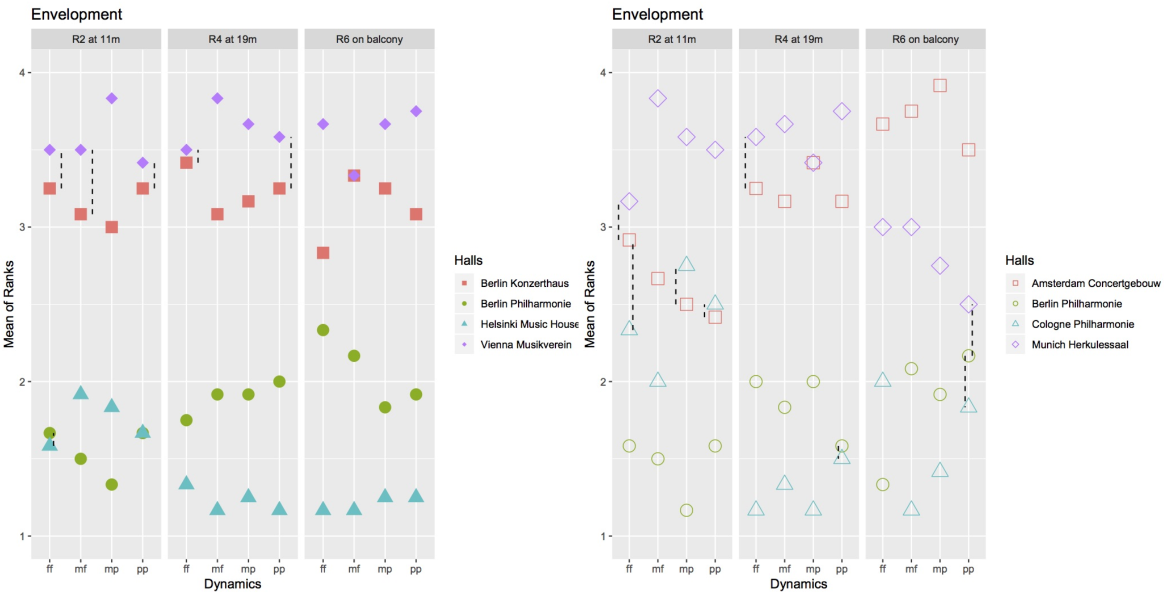


(a) (b)

Figure 12: Figures showing the LAeq values measured for the different equivalent levels for the different concert halls and positions.

The results for the measured LAeq are presented in Figure 12. This Figure shows that the differences in LAeq are within approximately 2 dB difference, the estimated perceivable difference in SPL proposed by Karjalainen and Pulkki [9]. However, for VM, the SPL level is significantly higher (approximately 3dB). In the first test, it is VM that has the higher SPL for each dynamic level (except for *pp* at R6 on the balcony where it is close with BK). The lowest SPL level is interchanging depending on dynamic level and seat position between the other concert halls, where BK increases SPL with increasing dynamic level and further position compared to the other concert halls. BP exhibits the smallest LAeq difference for the highest dynamic level but generally increases as the dynamic level gets quieter. BK is roughly similar for the close seat position for all dynamic levels however increases in general as the seat position moves further away. For the second test, the results are more varied and more within the 2 dB range. The difference grows for AC as the seat position moves further away and generally decreases for MH and BP. CP has the lowest SPL for almost all seat positions and dynamic levels.

6.3.2 Results for perceived envelopment for different concert halls



(a) (b)

Figure 13: Figures showing the envelopment results measured for the different equivalent levels for the different concert halls and positions.

For the first test, the majority agreed on the order that the more enveloping/wide was VM followed by BK. There were some disagreements and closeness especially at extreme dynamic levels at both the mid-position and at close position. The two halls considered less enveloping/wide were BP and MT. It is interesting to note that the volume of VM and BK are relatively the same, however they exhibit definite differences in their perceived envelopment. This seems in-keeping with Beranek‘s narrative on the concert halls as well as the hall‘s measured objective parameters. While VM and BK are similar in size, VM‘s G and EDT are much larger which may have an impact on the width. This is most likely due to the materials and architectural elements of the halls that in VM accentuate the lateral energy in both height and width to create a larger-sounding hall.

It is interesting to compare these results to the objective parameters for envelopment J_{LF} and L_J. There are subtle changes in ASW for VM and BK except in the furthest seating position. For LEV however, the differences are mostly in mid-frequency range and grow to be more different as the seating position moves further. This seems to correspond to the perception since at closer position, VM and BK are less discrepancy between, however at a further position.

It is also interesting to note that VM and BK were found to be the most dynamically responsive by Pätynen and Lokki [3]. Perceptual dynamics are also seen to be worse in MT according to the same reference, which can be confirmed in these results. This correlates the dynamic responsiveness with the perceived envelopment and nicely aligns with the findings of Pätynen and Lokki.

For the second test, concert halls that were seen to exhibit a more bass-heavy impulse response were chosen, to see how this affected the perceived envelopment/width. The spread shows more differences in perception depending on seat position than in the first test. At close position, MH is seen as the more enveloping/wide followed by either CP or AC and then BP. For moving seat position and constant dynamics, MH grows less enveloping/wide and AC is considered more as the position moves further away, which again may be due to its architectural elements and increase in lateral reflections as the volume can be heard more in its entirety. Again, results become more similar at extreme musical dynamics, which again shows that there was not such a clear perception. At close position, CP is considered to have a much more enveloping/wide sound. However, this worsens as the position moves further away where it often ties with BP as the least enveloping/wide hall.

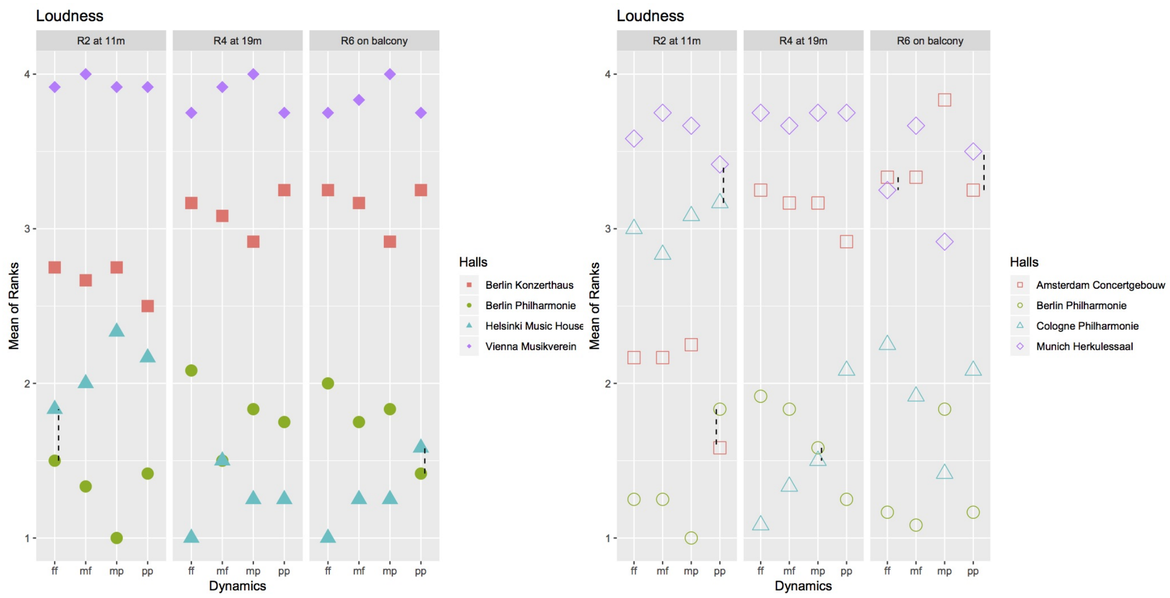
There are clear differences with perceived envelopment and their associated objective parameters in the second experiment. For example, in the second test BP and MH are the worst for exhibiting ASW and LEV at close position, whereas MH grows increasingly better in ASW. This is not reflected in the perceptual results where at MH is considered much more enveloping/wide for all positions and CP is considered the worst at further positions. Perhaps this is due to choosing halls which have a more bass-heavy response which has influenced the perception of the participants.

However, it was predicted that due to findings by Kahle and Green that with increasing dynamic level the hall was seen to 'wake up' [4], that the question would be harder to answer at low dynamic level rather than higher dynamic level, which is not shown in these results. In fact, there is little deviation among dynamic levels in the first experiment where the most confusion appears to arise at extreme dynamic levels. Similarly, in the second experiment, the results are slightly more confused among dynamic levels, however the biggest confusion again arises at the extreme dynamic levels. Perception of which hall is more enveloping/wide also changes slightly more between dynamic levels, especially at close position, however the results for different dynamic levels at the same position mostly agree except for close position.

6.3.3 Results for perceived loudness for different concert halls

The results from the first listening test show that almost all participants agreed that VM was the loudest concert hall across all dynamics. This was followed by BK, then depending on the seat position MT and BP were often competing. Confusion mainly arose at extreme levels. BP was considered less loud at close position than at the other positions. This may be due to a stronger ability to carry the direct sound which is mirrored by its results in the envelopment.

In the second testing, more disparities are apparent. The halls show different orderings usually at lower dynamics such as *pp* and *mp* for all positions. For example, the usual order at R2 position is MH, CP, AC then BP, however for *pp* the order



(a) (b)

Figure 14: Figures showing the loudness results measured for the different equivalent levels for the different concert halls and positions.

switches AC and BP. Similar happens at *pp* for R4. AC and MH compete for the louder hall at R6 position where AC grows increasingly louder. MH remains consistently loud across all positions. CP sounds louder at closer position however moves to be more worse than the two latter mentioned halls, competing with BP to be the quietest hall. Again, some confusion occurs in the extremities of the dynamic levels were participants could not distinguish between certain halls as to which was louder. The results from perceived loudness and envelopment appear to show similar ordering however, the results from the loudness are much clearer and more agreeable among participants.

The results from the listening tests for loudness also correspond with that of Pätynen and Lokki [3]. The loudness of AC is clearly increasing with seat position which corresponds to their findings that the dynamic range increases when moving from close position to a mid position.

Comparing with the objective parameters for subjective level, for the first experiment the order follows such that VM, BK, MT, BP for frequency 500 Hz and over. It is shown in the perceptual loudness testing that the order is mostly like that of lower frequencies, except in close position. For the other positions in the hall, our perceived loudness appears to be more aligned with the J_{LF} or L_J.

For the second experiment, the order follows such that the level is in the order AC, CP, MH, BP for the mid-frequency range. However, at low frequency this order is CP, BP, AC, MH and at high frequencies, CP, AC, MH, BP. This does not correspond at all with the perceived loudness shown in Figure 14, even at close position. Therefore, it again highlights a change in measure for perceptual loudness and that a more suitable parameter is required, perhaps related to the J_{LF} or L_J, for which it corresponds better with.

For the first experiment, the loudness corresponds well with envelopment. The ordering is almost identical. However, for the second experiment, at close position the order is changed. Also, at R6 position, there is more variation between AC and MH and less so with CP and BP in perceived loudness, compared with the perceived envelopment. This may be due to a heavier bass response that influenced more the feeling of envelopment.

6.3.4 Interviews

The following results from the interviews are presented in this section as a summary of the answers from the 12 expert participants.

Summary of answers for Question 1: "How did you find the user interface?"

In the first few listening tests, the user identified a glitch in the MaxMSP patch which meant that clicking the ‘Next’ button before stopping the playing between tracks caused the program to crash. Once the problem was identified, the instructions were modified so that future users were aware of this. The problem however led to no tampering with results.

There were also some limitations identified with the MaxMSP software that meant that using the ‘Next’ button as a toggle meant that the button lit up alternating. This was confusing to some participants. Alternatives such as ‘bang’ buttons were considered, however it was decided that users would be unable to clearly identify if they had completed a test. This was also added to the instructions later as a result from participants‘ feedback.

Otherwise, the feedback from the user was that the UI was very easy-to-use and really fast to switch between the two samples which meant that direct comparisons between the samples could be made easily.

Summary of answers for Question 2: "What is your definition of envelopment?"

A more comprehensive compilation of the answers given to the definition of envelopment is presented in Appendix C. Many participants described envelopment as a feeling of width or wideness, amount of sound coming from all directions and a sense of immersion.

Summary of answers for Question 3: "Was the loudest hall always the most enveloping or not always the case?"

The answer to this question differed among participants. Most of the participants said that it was always the case that the loudest hall was the most enveloping. One participant made an interesting observation that since a lot of the sound felt that it was coming from the front, it created a loudness triangle which in turn made it difficult to hear for the sides. However, it was clear to see among participants that there was a strong correlation with the loudness having an influence on which participants thought was the most enveloping hall.

Summary of answers for Question 4: "Was it easier to answer the questions when the sound level was louder or quieter?"

This question was to compliment the findings from Kahle and Green‘s 2018 paper [4]. However, it incurred split opinions when asked to the 12 participants. Approximately half the participants said that the questions were easier to answer when the sound level was quieter. Few participants said that the questions were easier to answer when the samples were quieter, otherwise participants commented the sound level didn‘t make a difference. Participants also commented that the questions were easier to answer when there was more prominent bass in the halls. One also mentioned that it was easier to answer the question in the middle range and that at the extreme sound levels it was more difficult.

6.3.5 Conclusion of the results

In general, fan-type halls used in the experiments were worst for both the perceived envelopment/width and the perceived loudness, despite their architecturally-designed wider structures. It appears many of the lateral reflections are lost due to their geometry and while visually they are more wide, audibly they are not perceived in such a way. The findings of this experiment contradict with Pätynen and Lokki [3] who claimed that the differences between hall-type are marginal in front position, as this can be seen for almost all seat positions for perceived envelopment/width, and only in the close seat position for the second experiment where AC is considered worse than CP.

There is little deviation among dynamic level where the biggest contrasts come from changing seat position. However, the order is relatively similar. Perceptions appear harder to make when there is more bass in the hall as shown in the second part of both perceptual tests. There is heavy correlation of loudness and envelopment/width in the first experiment, however the correlation seems to disappear in the second experiment. This is most likely due to a more bass-heavy impulse response from the hall, as discussed.

The interviews gave a supporting insight into how the participants carried out the experiments. It appears that the interview results correlate with the perceptual testings where participants said it was harder to answer the questions at extreme dynamics of loud and quiet. For many, perceived loudness and perceived envelopment/width went hand-in-hand, however many listened for different aspects such as frequency content rather than specific loudness. This could be shown more clearly in the differences of the second set of perceptual experiments.

Finally, a lot of the results obtained in these perceptual tests correspond to findings by Pätynen and Lokki [3]. There were that the BK and VM were the most dynamically responsive concert halls; that AC became increasingly dynamically responsive with changing seat position; and that MT and BP were the worst at exhibiting dynamic responsiveness. From this experiment, it can be determined that MH also has a good dynamic responsiveness and CP has a worse responsiveness, similar to other halls of its type.

7 Summary

In light of the research by Kahle and Green in 2018 [4], it was predicted that at higher dynamic levels, the loudness and envelopment/width would be easier to perceive due to the lateral reflections being more paramount, and hence creating a better idea of the size of the space. However, the results show that is not any easier to make comparisons of the concert halls at lower dynamic levels than it is at higher dynamic levels both for perceptual envelopment or for loudness as there are limited differences between results across dynamic ranges. Since these are the two main factors determined by Pätynen and Lokki [3] to be directly influencing our perception of changes in musical dynamics, the conclusion can be applied also for the perception of this attribute.

While this study has been extensive in reviewing two of the main attributes contributing to musical dynamics, there are many other influencing factors that were not considered or tested. The next biggest influence highlighted by Pätynen and Lokki [3] was clarity. This includes attributes such as articulation, definition and sharpness of attacks. While objective measures for clarity are presented, it can be seen that there is not a straightforward relationship between the objective measures of clarity and loudness or envelopment and hence some testing may be required here for the full picture.

Further to this, as shown by Meyer [2] and Pätynen and Lokki [3], tonal balance and brightness are another big perceptual influence on musical dynamics. The musical dynamics are especially emphasized at higher frequencies for increased playing levels as shown by Meyer [2] due to more harmonic overtones becoming apparent. The objective parameters were calculated for the different frequency bands in order to give an estimation of how they depend on the frequency response. As well as this, in the second testing, halls which were known to exhibit a more bass-heavy response were chosen in order to highlight some frequency effects. Despite these considerations, there is still no accepted objective measure for timbre, and this study is one of many that calls for a solution.

Another important consideration is the accuracy of the experiment. Naturally, SPL for each of the tracks were not the same and while changes were roughly within the 2 dB noticeable difference identified by Karjalainen and Pulkki [9], this may have caused some bias to the results. Similarly, in reality, at a higher SPL the C-weighting should be applied since our subjective loudness follows this curve. However, it is interesting to note that even at low SPL, the objective measure does not directly follow the subjective loudness nor envelopment. Further to this, for concert halls with limited bass, it appears that L_J is better suited as on objective measure to subjective 'loudness', whereas with halls with more bass, G is a better objective measure.

In 2019, Lokki released a paper [78] summarized the architectural elements in order to make the sound a concert hall bloom. These 'ideal' concert halls have side walls with small diffusing elements, deep side balconies, high ceiling, vertical wall elements, elevated stage, reduced background noise, reflecting surfaces and flat floor with seats that allow the sound to pass underneath. It can be concluded that since these concert halls produce more enveloping and louder sounds, these architectural elements apply for the perception of musical dynamics.

References

[1] L. Beranek, *Concert hall acoustics: music, acoustics and architecture.* New York, NY, Springer, 1962.

[2] J. Meyer, *Acoustics and the Performance of Music.* Springer, New York, 2009.

[3] J. Pätynen and T. Lokki, "Perception of music dynamics in concert hall acoustics," *The Journal of the Acoustical Society of America*, vol. 140, no. 5, pp. 3787–3798, 2016.

[4] E. Green and E. Kahle, "Dynamic spatial responsiveness in concert halls," vol. 40, 10 2018. Proceeding of the Institute of Acoustics, Hamburg 2018.

[5] L. Beranek, "Listener envelopment lev, strength g and reverberation time rt in concert halls," 8 2010. Proceedings of 20th International Congress on Acoustics, ICA 2010, Sydney, Australia.

[6] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 1421–1448, July 2012.

[7] W. Sabine, *Reverberation: Introduction.* The American Architect, 1900.

[8] I. S. O, "International organization for standardization 3382-1 acoustics – measurement of room acoustic parameters. part 1: Performance spaces," 2009.

[9] V. Pulkki and M. Karjalainen, *Communicaiton Acoustics: Introduction to Speech, Audio and Psychoacoustics.* Wiley, 2015.

[10] M. Barron and A. Marshall, "Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure," *Journal of Sound and Vibration*, vol. 77, no. 2, pp. 211 – 232, 1981.

[11] V. L. Jordan, "Acoustical criteria for auditoriums and their relation to model techniques," *The Journal of the Acoustical Society of America*, vol. 47, pp. 408–412, 1970.

[12] M. R. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, 1965.

[13] J. Blauert, *Communication acoustics.* New York, NY, Springer, 2005.

[14] B. Moore, *An Introduction to the Psychology of Hearing: Sixth Edition.* CRC Press; 1 edition, 2013.

[15] N. Zacharov, *Sensory evaluation of sound.* CRC Press; 1 edition, 2018.

[16] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models.* New York, NY, Springer, 2007.

[17] S. S. Stevens, “A scale for the measurement of a psychological magnitude: loudness.,” *Psychological Review*, vol. 43, no. 5, p. 405–416, 1936.

[18] E. C. Poulton, “Models for biases in judging sensory magnitude.,” *Psychological Bulletin*, vol. 86, no. 4, p. 777–803, 1979.

[19] D. H. Mershon, D. H. Desaulniers, S. A. Kiefer, J. Thomas L Amerson, and J. T. Mills, “Perceived loudness and visually-determined auditory distance,” *Perception*, vol. 10, no. 5, pp. 531–543, 1981.

[20] M. C. Killion, “Revised estimate of minimum audible pressure: Here is the missing 6 decibel?,” *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1501–1508, 1978.

[21] D. W. Robinson and R. S. Dadson, “A re-determination of the equal-loudness relations for pure tones,” *British Journal of Applied Physics*, vol. 7, pp. 166–181, May 1956.

[22] J. J. Rosowksi and A. Graybeal, “What did morganucodon hear?,” *Zoological Journal of the Linnean Society*, vol. 101, no. 2, pp. 131–168, 1991.

[23] “The audibility of low-frequency sounds,” *Journal of Sound and Vibration*, vol. 21, no. 4, pp. 431 – 448, 1972.

[24] D. L. Johnson and H. von Gierke, “Audibility of infrasound,” *The Journal of the Acoustical Society of America*, vol. 56, no. S1, pp. S37–S37, 1974.

[25] H. Fletcher and W. A. Munson, “Loudness, its definition, measurement and calculation,” *The Journal of the Acoustical Society of America*, vol. 5, no. 2, pp. 82–108, 1933.

[26] I. S. O, “International organization for standardization 226:2003 acoustics – normal equal-loudness-level contours,” 2003.

[27] N. F. Viemeister, “Intensity discrimination of pulsed sinusoids: The effects of filtered noise,” *The Journal of the Acoustical Society of America*, vol. 51, no. 4B, pp. 1265–1269, 1972.

[28] G. P. Schooneveldt and B. C. J. Moore, “Comodulation masking release (cmr) as a function of masker bandwidth, modulator bandwidth, and signal duration,” *The Journal of the Acoustical Society of America*, vol. 85, no. 1, pp. 273–281, 1989.

[29] C. S. Watson, “Some comments on informational masking,” *Acta Acustica united with Acustica*, vol. 91, no. 3, pp. 502–512, 2005.

[30] J. Blauert, *Spatial Hearing - Psychophysics of Human Sound Localization*. MIT Press, 1997.

[31] D. W. Grantham, “Chapter 9 - spatial hearing and related phenomena,” in *Hearing* (B. C. Moore, ed.), Handbook of Perception and Cognition, pp. 297 – 345, San Diego: Academic Press, 1995.

[32] C. Faller and J. Merimaa, “Source localization in complex listening situations: Selection of binaural cues based on interaural coherence,” *The Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3075–3089, 2004.

[33] P. M. Hofman, J. G. A. Van Riswick, and A. J. Van Opstal, “Relearning sound localization with new ears,” *Nature Neuroscience*, vol. 1, no. 5, pp. 417–421, 1998.

[34] O. Santala and V. Pulkki, “Directional perception of distributed sound sources,” *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1522–1530, 2011.

[35] H. Wallach, E. B. Newman, and M. R. Rosenzweig, “The precedence effect in sound localization,” *The Journal of the Audio Engineering Society*, vol. 21, no. 10, pp. 817–826, 1973.

[36] H. Haas, *Über den Einfluß eines Einfachechos auf die Hörsamkeit von Sprache*. PhD thesis, Göttingen, 1949.

[37] A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, “Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss,” *Attention, perception psychophysics*, vol. 78, no. 2, pp. 373–395, 2016.

[38] A. Marshall, “A note on the importance of room cross-section in concert halls,” *Journal of Sound and Vibration*, vol. 5, no. 1, pp. 100 – 112, 1967.

[39] M. Barron, “The subjective effects of first reflections in concert halls—the need for lateral reflections,” *Journal of Sound and Vibration*, vol. 15, no. 4, pp. 475 – 494, 1971.

[40] M. Long, “What is so special about shoebox halls? envelopment, envelopment, envelopment,” *Acoustics Today*, vol. 5, pp. 21–25, 2009.

[41] R. Mason, T. Brookes, and F. Rumsey, “Spatial impression: measurement and perception of concert hall acoustics and reproduced sound,” in *International Symposium on Room Acoustics: Design and Science*, 2004.

[42] T. Hidaka, T. Okano, and L. Beranek, “Interaural cross correlation as a measure of spaciousness and envelopment in concert halls,” *Journal of The Acoustical Society of America*, vol. 92, 10 1992.

[43] W. Keet, “The influence of early lateral reflections on the spatial impression,” *Proceedings of the 6th International Congress on Acoustics, Tokyo, Japan*, vol. 3, no. 53, 1968.

[44] J. S. Bradley, G. A. Soulodre, and N. Popplewell, "Pilot study of simulated spaciousness," *Journal of The Acoustical Society of America*, vol. 93, 04 1993.

[45] G. A. Soulodre, M. C. Lavoie, and S. G. Norcross, "Temporal aspects of listener envelopment in multichannel surround systems," Audio Engineering Society: 114th Convention, Amsterdam, The Netherlands (March, 2003).

[46] A. Wakuda, H. Furuya, K. Fujimoto, K. Isogai, and K. Anai, "Effects of arrival direction of late sound on listener envelopment," *Acoustical Science and Technology*, vol. 24, no. 4, pp. 179–185, 2003.

[47] D. Griesinger, "The psychoacoustics of apparent source width, spaciousness and envelopment in performance space," *Acta Acustica*, vol. 83, no. 4, pp. 721–731, 1997.

[48] D. Griesinger, "Objective measures of spaciousness and envelopment," 16th International Conference: Spatial Sound Reproduction (March, 1999).

[49] D. Griesinger, "General overview of spatial impression, envelopment, localization, and externalization," in *Audio Engineering Society Conference: 15th International Conference: Audio, Acoustics Small Spaces*, Oct 1998.

[50] M. Kleiner and J. Tichy, *Acoustics of small rooms.* CRC Press, New York, 2014.

[51] D. Cabrera, A. Nguyen, and Y.-J. Choi, "Auditory versus visual spatial impression: A study of two auditoria," Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia (July, 2004).

[52] J. W. Danahy, "Technology for dynamic viewing and peripheral vision in landscape visualization," *Landscape and Urban Planning*, vol. 54, no. 1, pp. 127 – 138, 2001.

[53] R. R. D. Bradley, J. S. and S. G. Norcross, "On the combined effects of early and late arriving sound on spatial impression in concert halls," *The Journal of the Acoustical Society of America*, vol. 108, no. 2, pp. 651–661, 2000.

[54] M. Morimoto, K. Nakagawa, and K. Iida, "The relation between spatial impression and the law of the first wavefront," *Applied Acoustics*, vol. 69, no. 2, pp. 132 – 140, 2008.

[55] J. S. Bradley, "Contemporary approaches to evaluating auditorium acoustics," in *Audio Engineering Society Conference: 8th International Conference: The Sound of Audio*, May 1990.

[56] A. Nguyen and D. Cabrera, "Visual counterparts to spatial impression in auditorium acoustics," 2004. in Proceedings of International Symposium on Room Acoustics: Design and Science, Hyogo, Japan.

[57] R. J. Hawkes and H. Douglas, "Subjective acoustic experience in concert auditoria," *Architectural Research and Teaching*, vol. 1, no. 2, pp. 34–45, 1970.

[58] M. Barron, “Subjective study of british symphony concert halls,” *Acta Acustica united with Acustica*, vol. 66, no. 1, pp. 1–14, 1988.

[59] A. H. Marshall, “Levels of reflection masking in concert halls,” *Journal of Sound and Vibration*, vol. 7, no. 1, pp. 116–118, 1968.

[60] E. Kahle and M. Bruneau, “Validation d’un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d’opéras / validation of an objective model of the perception of room acoustical quality in an ensemble of concert halls and operas,” 1995. Phd Thesis.

[61] M. Sams, R. Hari, J. Rif, and J. Knuutila, “The human auditory sensory memory trace persists about 10 sec: Neuromagnetic evidence,” *Journal of Cognitive Neuroscience*, vol. 5, no. 3, pp. 363–370, 1993. PMID: 23972223.

[62] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, “Comparative study of european concert halls: correlation of subjective preference with geometric and acoustic parameters,” *The Journal of the Acoustical Society of America*, vol. 56, no. 4, pp. 1195–1201, 1974.

[63] R. Kürer, G. Plenge, and H. Wilkens, “Correct spatial sound perception rendered by a special 2-channel recording method,” in *Audio Engineering Society Convention 37*, Oct 1969.

[64] C. Lavandier, “Validation perceptive d’un modèle objectif de caractérisation de la qualité acoustique des salles,” 1989. Thèse de doctorat dirigée par Bruneau, Michel Acoustique Le Mans 1989.

[65] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, “Disentangling preference ratings of concert hall acoustics using subjective sensory profiles,” *The Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3148–3161, 2012.

[66] 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*, vol. 111, no. 12, pp. 4409–4414, 2014.

[67] A. Kuusinen, “Perception of concert hall acoustics – selection and behaviour of assessors in a descriptive analysis experiment,” *Aalto University School of Electrical Engineering*, 2011. Master’s thesis.

[68] M. Barron, “Loudness in concert halls,” *Acta Acustica*, vol. 82, no. 1, pp. 21 – 29, 1996.

[69] M. Barron and A. H. Marshall, “Spatial responsiveness in concert halls and the origins of spatial impression,” *Applied Acoustics*, vol. 62, pp. 91 – 108, 2001.

[70] W. Kuhl, “Raümlichkeit als komponente des raumeindrucks,” *Acta Acustica United Acustica*, vol. 40, no. 3, pp. 167–181, 1978.

[71] M. H. A. Cremer, L., *Principles and applications of room acoustics.* 2016. translated by Shultz, T. J. in 1982.

[72] R. Wettschurek, "Über die abhängigkeit raumakustischer wahrnehmung von der lautstärke," 1976. Phd Thesis, Technical University of Berlin.

[73] J. S. Bradley and G. A. Soulodre, "Objective measures of listener envelopment," *The Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2590–2597, 1995.

[74] J. S. Bradley, G. A. Soulodre, and S. Norcross, "Factors influencing the perception of bass," *The Journal of the Acoustical Society of America*, vol. 101, no. 5, pp. 3135–3135, 1997.

[75] J. R. Hyde, "Discussion of the relation between intiial time delay gap (itdg) and acoustical intimacy: Leo beranek's final thoughts on the subject, documented," vol. 38, 2018. Proceedings of the Institute of Acoustics.

[76] D. A. Dick and M. C. Vigeant, "An investigation of listener envelopment utilizing a spherical microphone array and third-order ambisonics reproduction," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2795–2809, 2019.

[77] J. Pätynen and T. Lokki, "The acoustics of vineyard halls, is it so great after all?," *Acoustics Australia*, vol. 43, no. 1, p. 33–39, 2015.

[78] T. Lokki and J. Päytnen, "Architectural features that make music bloom in concert halls," *Auditorium Acoustics*, vol. 1, no. 2, pp. 439–449, 2019.

[79] J. Pätynen, "A virtual symphony orchestra for studies on concert hall acoustics," *Aalto University School of Electrical Engineering*, 2011. Ph.D. thesis.

[80] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *The Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, 2013.

[81] C. '74, *Max 7.0.1*, 2018. [computer software].

[82] MATLAB, *version 7.10.0 (R2010a).* Natick, Massachusetts: The MathWorks Inc., 2010.

[83] R Core Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013.

A Table of the A-weighted equivalent levels recorded for the musical extracts for the first listening test

The first test had a background A-weighted equivalent level of 28.0 dB. The main source of background noise was found to be a transformer in close location to the listening room with a constant hum of around 100Hz.

Table A1: A table showing the measured values of the LAeq for the dynamic level in the different positions of each hall for the second listening test.

Hall	Position	Dynamic level	LAeq [dB]
BK	R2	ff	84.9
BP	R2	ff	84.5
MT	R2	ff	85.2
VM	R2	ff	86.5
BK	R4	ff	83.6
BP	R4	ff	82.6
MT	R4	ff	83.1
VM	R4	ff	84.3
BK	R6	ff	83.5
BP	R6	ff	82.2
MT	R5	ff	82.1
VM	R6	ff	83.5
BK	R2	mf	79.4
BP	R2	mf	79.8
MT	R2	mf	80.4
VM	R2	mf	81.1
BK	R4	mf	78.1
BP	R4	mf	78.0
MT	R4	mf	77.5
VM	R4	mf	79.0
BK	R6	mf	78.1
BP	R6	mf	77.0
MT	R5	mf	76.7
VM	R6	mf	78.6
BK	R2	mp	71.5
BP	R2	mp	70.6
MT	R2	mp	72.5
VM	R2	mp	73.9
BK	R4	mp	69.9
BP	R4	mp	69.1
MT	R4	mp	68.7
VM	R4	mp	71.2
BK	R6	mp	70.1
BP	R6	mp	68.9
MT	R5	mp	69.6
VM	R6	mp	71.5
BK	R2	pp	63.2
BP	R2	pp	64.3
MT	R2	pp	63.8
VM	R2	pp	65.2
BK	R4	pp	61.2
BP	R4	pp	61.8
MT	R4	pp	60.8
VM	R4	pp	61.9
BK	R6	pp	62.2
BP	R6	pp	60.1
MT	R5	pp	60.9
VM	R6	pp	62.7

B Table of the A-weighted equivalent levels recorded for the musical extracts for the second listening test

The first test had a background A-weighted equivalent level of 26.8 dB. The main source of background noise was found to be a transformer in close location to the listening room with a constant hum of around 100Hz.

Table B1: A table showing the measured values of the LAeq for the dynamic level in the different positions of each hall for the second listening test.

Hall	Position	Dynamic level	LAeq [dB]
AC	R2	ff	85.3
BP	R2	ff	84.9
CP	R2	ff	85.8
MH	R2	ff	86.3
AC	R4	ff	84.1
BP	R4	ff	83.5
CP	R4	ff	83.5
MH	R4	ff	84.7
AC	R6	ff	83.3
BP	R6	ff	82.2
CP	R5	ff	82.8
MH	R6	ff	82.8
AC	R2	mf	80.6
BP	R2	mf	80.1
CP	R2	mf	80.4
MH	R2	mf	81.3
AC	R4	mf	79.1
BP	R4	mf	78.1
CP	R4	mf	78.5
MH	R4	mf	79.3
AC	R6	mf	78.5
BP	R6	mf	77.4
CP	R5	mf	77.7
MH	R6	mf	78.3
AC	R2	mp	72.6
BP	R2	mp	71.8
CP	R2	mp	73.2
MH	R2	mp	74.6
AC	R4	mp	71.0
BP	R4	mp	69.6
CP	R4	mp	70.2
MH	R4	mp	72.0
AC	R6	mp	70.1
BP	R6	mp	69.9
CP	R5	mp	70.5
MH	R6	mp	69.5
AC	R2	pp	63.5
BP	R2	pp	64.2
CP	R2	pp	64.2
MH	R2	pp	65.3
AC	R4	pp	61.7
BP	R4	pp	61.0
CP	R4	pp	61.5
MH	R4	pp	61.7
AC	R6	pp	61.3
BP	R6	pp	60.2
CP	R5	pp	61.3
MH	R6	pp	61.4

C Definitions of envelopment from the interviewed subjects

Quotes from the 12 expert subjects from the Acoustics and Audio Technology department at Aalto University who participated in the listening test answering the question: **"What is your definition of envelopment?"** The quotes have been anonymized and randomized.

Subject 1

"Envelopment is like the sound is coming from multiple directions. How I felt the difference it would be if coming only from the front then its less enveloping but coming from the sides then more enveloping."

Subject 2

"What's the proportion of the sounds coming from the sides anywhere else but the front."

Subject 3

"Wideness and inner feeling. If something sounds natural then it'll make me feel more enveloped. How wide it is or if I can see something in the back."

Subject 4

"Envelopment is how much I am surrounded by noise or by a sound in my 360 degrees."

Subject 5

"How wide the sound is, is it coming from one direction or multiple."

Subject 6

"Sensation when you feel inside the music or performance. Listen to the band close to me or also all around me. All the sides. But both things. Because dry sound very far away so close to the source so in the middle but sound coming from everywhere."

Subject 7

"More ambience, more diffuse. Ambience something more directional. Both contribute to the envelopment."

Subject 8

"Basically just how does it feel like its coming around the side of you. Does it sound like its coming directly in front of you or from around the sides also. You can just kind of hear like its just coming from the front or its coming from the sides as well."

Subject 9

"It's a balance between sound coming from all directions. So if you hear the same amount of sound coming from all directions."

Subject 10

"Immersed. So basically the sound is all around me and also I associate it with the width of the source so basically if the image of the source is spread its around me but yeah just surrounded by the sound I guess."

Subject 11

"Okay well... um I don't know the word so I went with wider. Listening for the width like how sort of stereo or how wide the sound was because I'm not familiar with the word envelopment."

Subject 12

"How immersed I can feel in the soundscape. Sometimes I felt like the volume was changing so maybe the volume of the audio. I was trying to also listen to the position of the sources but for me it was almost the same and at some point I felt more reverberation but maybe that was related to the volume. That's a confusion of feeling I have there."