INTERNATIONAL
SYMPOSIUM on
MUSICAL and ROOM
ACOUSTICS

ISMRA
2016

September 11-13, 2016

La Plata, Buenos Aires, Argentina

RA-MS: Measurement of concert halls / opera houses: Paper ISMRA2016-39

Measurements of IACC during music performance in concert halls

Magne Skålevik[(a)]

[(a)] AKUTEK and Brekke&Strand, Norway, msk@brekkestrand.no

Abstract

Spatial aspects of the sound field at concert listeners' ears have for many decades been considered important for the experience of concert hall acoustics. Strangely enough, this knowledge has not led to development of direct measurement methods. Acousticians seems to have arrived at the consensus that two distinct aspects are important, namely Source Broadening or Apparent Source Width (ASW), and Listener Envelopment (LEV). The aspects are considered predictable from impulse response measurements, in terms of LF and 1-IACCE for ASW, and LLG and 1-IACCL for LEV. In order to explore cues of ASW and LEV directly at listeners' ears, this author has analyzed binaural recordings in concerts with symphony orchestras. Until otherwise proven, one would expect that any difference between two halls that can be perceived by our ears could also be measured as soon as the technical challenges are overcome. At least, one would like to test the common understanding that there are significant differences in IACC from hall to hall.

This paper presents results from measurements of IACC(t) obtained from binaural recordings during symphony orchestra performances worldwide, including well-known halls and orchestras, with statistics from N=337989 correlation periods measured over 33799 seconds, i.e. >9 hours, from 10 big concert halls in Europe and the US. The hypothesis, "Binaural signals, i.e. signals at listeners' pair of ears, can exhibit statistically significant hall-to-hall differences in cross-correlation", is not rejected by the data. In further work, the issue of predicting ASW(t) and LEV(t) in terms of parallel streams will be pursued.

Keywords: Spatial impression, spaciousness, envelopment, binaural hearing, IACC

INTERNATIONAL
SYMPOSIUM on
MUSICAL and ROOM
ACOUSTICS

ISMRA
2016

September 11-13, 2016
La Plata, Buenos Aires, Argentina

Measurements of IACC during music performance in concert halls

1 Introduction

Spatial aspects of the sound field at concert listeners' ears have for many decades been considered important for the experience of concert hall acoustics. Strangely enough, this knowledge has not led to development of direct measurement methods.

Perception of spatial aspects rely on binaural hearing. Generally, scientific knowledge about the significance of binaural hearing in performance spaces can be traced back to the World Exhibition 1881 in Paris and an installation at the Paris Opera [1].

In 1931, Alan Blumlein lodged the patent for "binaural" sound in a paper which patented stereo records, stereo films and also surround sound. In January 1934, Blumlein took his stereo-cutting equipment to the newly opened Abbey Studios and recorded Sir Thomas Beecham conducting the LPO, as it rehearsed Mozart's Jupiter Symphony [2]. Since the 1960-es, the audible difference between stereo audio and mono audio became evident to the public due to the development in broadcasting and music reproduction technique, and the availability of such equipment. In stereo sound processing, the goniometer and the cross-correlation indicator were commonly used to measure the stereophonic content of a stereo signal, not least to check polarity or phase differences that could cause unwanted cancellations when stereo signals were reproduced in mono equipment (mono compatibility check), Figure 1. Cross-correlation close to 1 would correspond to a frontal or centered sound image, while values close to 0 would correspond to a wide sound image. Sound engineers have learned how to measure the wideness of the sound image by monitoring the cross-correlation between the signals presented at each ear [3].

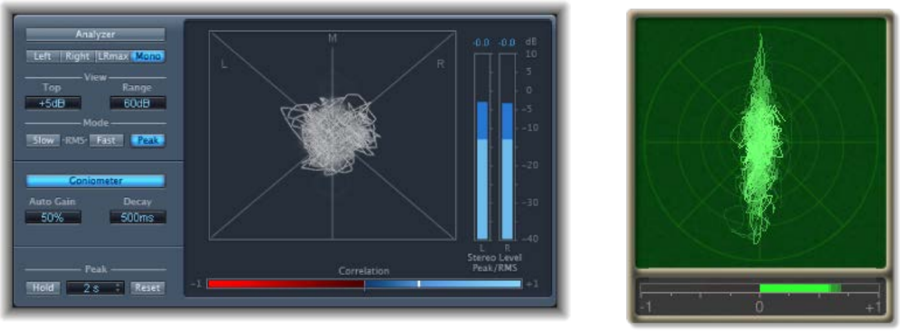

Figure 1 Goniometers and Correlation meters. An example of an apparently wide stereo signal with low correlation (Left), and an example of an apparently narrow and frontal signal and high correlation. -1 means anti-phase or reversed polarity and potentially cancellation in mono.

A coincidence or not, the 60-es was the decade when some researcher started to investigate 'room impression', and Harold Marshall (1967) was the first author to publish a paper suggesting that different concert halls could produce different spatial perception of a sound source, in particular proposing the importance of cross section geometry and early lateral reflections [4][5].

It is unclear why the knowledge about binaural perception developed since 1881 up to present seems to have had little impact in concert hall acoustic research and measurement technique.

The established knowledge about spatial aspects in concert hall acoustics developed between 1960 and 2000, with contributions from several researchers and authors [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], as reviewed by Barron (2001) [4].

Today, there is a common understanding among acousticians that a symphony orchestra can produce a broad sound image in one concert hall and a narrow or frontal sound image in another concert hall, even if playing the same piece. This perceptive effect is referred to as Apparent Source Width (ASW), or Source Broadening. A similar common understanding goes for perceived Listener Envelopment (LEV): One concert hall offer the sensation that musician and listener is in the same space, enveloped by the same physical environment, while another hall fail to do so. However, no available measurement apparatus can measure such differences directly from the sound field at listeners' ears during a symphony orchestra performance.

In contrast, any statement about an orchestra sounding louder in one hall than in another could be tested with a sound level meter during music performances.

Instead of using direct measurement methods, concert hall acoustics researchers use indirect measurement methods based on room acoustical impulse responses (RIRs) and corresponding quantities, namely LF and 1-IACCE for ASW, and LLG and 1-IACCL for LEV. Here, LF is (Early) Lateral Fraction, LLG is Late Lateral G (G is sound strength), and IACCE and IACCL is the inter-aural cross-correlation of early and late parts of RIR, respectively. However, there are some problems with these methods.

While impulse responses are quite straightforward to acquire in an empty auditorium, they are less so with orchestra and audience in place.

Measurements in unoccupied conditions leave some uncertainties as to their relevance for occupied conditions.

The connection between the early part of the impulse response and ASW as well as the connection between the late part of the impulse response and LEV is not straightforward. While ASW and LEV can be quite intuitive during percussive parts of the music or during the decay after so-called stop chords, it is unclear how this is processed to the brain in running music. Several authors have suggested that the brain can process two separate streams of information. Kahle [17] suggested that early sound relates to Source Presence while late sound relates to Room Presence [18].

How ASW and LEV in practice can be predicted separately from a continuous signal like the one arriving at concert listeners ears needs to be addressed in further work.

Moreover, recent research has challenged the convention that ASW is determined by early sound (arriving less than 80ms later than the direct sound) and LEV is determined by late sound (arriving more than 80ms later than the direct sound). Klockgether et al had respondents assessing ASW and LEV while presenting anechoic music convolved with impulse responses where cross-correlation as well as the contribution from early and late parts were controlled [6]. The results of the experiment showed that the manipulation of the reverberant tail of an impulse response also affects the perception of source width, and that late, and maybe even early, reflections have an impact on the perceived listener envelopment. With relevance to this authors current work, Klockgether described a method for extracting ASW-cues and LEV cues separately from running music in a binaural signal.

Cross-correlation in binaural signals is known to be a cue of spatial perception aspects. This paper presents results from measurements of cross-correlation in binaural signals recorded during symphony orchestra performances worldwide, including well-known orchestras and halls.

Among the tasks in the initial stage of this research project is to make the method able to produce repeatable and significant results even while the source, i.e. orchestras and musical content, vary.

An introduction and a report from initial studies have previously been presented by this author. This paper presents results from measurements of IACC obtained from binaural recordings during symphony orchestra performances worldwide, including well-known halls and orchestras, using inter-aural cross-correlation and other common measures.

2 Towards a method for direct measurement of cues for spatial perception in sound at concert listeners' ears

2.1 Motivation for hypothesis

- Hall-to-hall differences between inter-aural cross-correlation (IACC) in binaural impulse responses (BRIRs) have been reported
- From experiments, differences in IACC are known to correspond to differences in ASW and LEV
- There seems to be an established opinion that there are hall-to-hall differences in ASW and in LEV
- Until otherwise proven, one would expect that any difference between two halls that can be perceived by our ears could also be measured as soon as the technical challenges are overcome.

2.2 Hypothesis

Binaural signals, i.e. signals at listeners' pair of ears, can exhibit statistically significant hall-to-hall differences in cross-correlation.

2.3 Cross-correlation, qualitative description

Cross-correlation have several meanings in binaural hearing. Cross-correlation between the sound signals entering a pair of ears, commonly referred to as signal L (left) and signal R (right), is a measure and predictor of the perceived sound image. Incoming sounds in or near the median plane, i.e. the vertical plane perpendicular to the axis through the left and right ears, will enter the two ears simultaneously. The information about the sound will, from each ear, travel along a delay-line and will at some point "meet", or coincide perfectly, at one of the many so-called coincidence cells along the delay-line. The brain has learned from experience that this coincidence cell corresponds to events in the median plane, and a group of neighboring coincidence cells corresponding to events in the near-median plane. Events outside the near-median plane, i.e. sounds arriving from azimuth angles different from zero, will arrive at the two ears at different times, i.e. inter-aural time-difference ITD. The L and R signals will therefore not coincide at the coincidence cells associated with the median plane (but at some other coincidence cell; however, details are superfluous in this context). We can predict to which degree the brain will associate the incoming sound to the median plane by using a well-known mathematical algorithm that mimics the coincidence process, namely the aforementioned cross-correlation function. A perfect point source located somewhere in the median plane will, in perfectly noise-free and an-echoic environment, produce identical L and R signals, exhibiting a perfect cross-correlation, having the maximum value of 1.0. If introducing an increasing amount of sound reflections (or noise), L and R would become more and more different, thus resulting in decreasing cross-correlation. In contrast, increasing reflections from a flat floor or a flat ceiling, e.g. outdoors far from vertical surfaces or indoors far from side walls, would arrive in the median plane and therefore increase the cross-correlation. The latter case explains why reflective ceiling and floor produce an experience of a frontal sound source. Even late reverberant sound will add to the cross-correlation if the environment is dominated by reflective surfaces intersecting with the median plane. This means that reverberance, loudness and clarity can all be independent of the spatial aspects. EDT, G and, C80 can be perfect, even if ASW and LEV are zero. However, the ill-reputed, wide, fan-shaped halls of the post-war era tended to have too low EDT, and to weak G toward the back of the halls, in addition to lack of ASW and LEV.

By combining the coincidence cue with information about high frequency (e.g. 4kHz) content, and slight turning of the head, it is possible for the brain to decide where in the median plane the sound source is, whether the source is located up front, at the back or above the head.

2.4 Cross-correlation, quantitative description

The (Normalized) Correlation (Coefficient) r between the left signal L and the right signal R, assuming sample means equal to zero, can be computed from a stereo sample-pair sequences {Li,Ri} over a correlation period of arbitrary duration T, containing n sample pairs with index $i=1,2,..,n.$ where $n=T \cdot f_s$ *when* f_s is the sampling frequency:

$$r = N/D, \qquad (1)$$

where

$$N = \Sigma L_i \cdot R_i \ \text{ and } \ D = \text{sqrt}\, [\Sigma L_i^2 \cdot \Sigma R_i^2] \qquad (2)$$

Here, Σ means sum of all terms with index $i = 1, 2, \ldots, n$,

For the purpose of the work presented in this paper, the inter-aural cross-correlation function is defined as a function of (discrete) time, integration time T and frequency band

$$IACC(t, T, f) \text{ or } IACC_{T,f}\,(t) \qquad (3)$$

In the measurement period (t_0,t_1), the discrete times are $t=t_0+T$, $t_0+2T,\ldots,t_1$.

Note: The inter-aural cross-correlation function used here differs from the IACC conventionally applied on binaural impulse responses (BRIR) according to ISO-3382. E.g., the latter would also be a function of ITD (τ), returning the maximum value of IACC from the interval $-1\text{ms}<\tau<1\text{ms}$. Inherently, ISO-3382 emphasize strong reverberant components from single azimuth angles rather the average from all azimuth angles like in (3).

While cross-correlation indeed has been used as an indicator of the wideness of stereo signals, there is no standard method for how to use the output of the cross-correlation between signals at listeners' as a function of time in order to find features that effects ASW and LEV, and to distinguish between them.

Klokgether analysed binaural signals in 40ms segments, using the energy ratio between one segment and the preceding segment as a measure of early energy content and late energy content of the segment. This author has explored a similar approach, as will be describe below.

2.5 Problem: to predict ASW and LEV from running music

In this paper results are based on the inter-aural cross-correlation analyzed in periods of T=100ms, i.e. 10 periods per second. This choice is arbitrary, motivated by the fact that the period is long enough to contain the shortest possible musical transient and the following early energy. E.g., orchestral stop chords typically have a rise time of 3-400ms. A level step at time t is defined

$$dL(t) = L(t) - L(t-T) \qquad (4)$$

Periods with $dL(t)>0$ are denoted Up-steps, while periods with $dL(t)<0$ are denoted Down-steps.

This author suggests that an Up-step in the Median Energy, $dL_M>0$, indicates that the actual period is dominated by direct energy and early reflected sound, while a Down-step, $dL_M<0$, indicates that the actual period is dominated by reverberant energy. The bigger the Up-step, the higher the probability that the period contains an ASW cue. The bigger the Down-step, the higher the probability that the period contains an LEV cue.

Assuming C80≈0dB, early energy is approximately equal to late energy, thus a period with dL close to zero would contain an equally strong cue for ASW and LEV. These cases are common (half of the periods have dL between -2dB and +2dB) and force us to make assumptions as to what ASW and LEV is and is not.

2.5.1 Static aspects

If we assume that ASW is a static aspect that can be predicted from the early part of an impulse response, we would choose a selection of periods with high dL-values as a basis for our prediction of ASW from running music. Likewise, with LEV we would choose a selection of periods with low dL_M-values (big negative values) as a basis for our prediction of LEV from running music.

2.5.2 Dynamic aspects

In contrast, if we assume ASW(t) and LEV(t) being dynamic, time-varying aspects, like continuous parallel perceptive streams, we would rather apply a continuous weighing, or filtering, function on every period. An example from the current study is presented in Figure 2 (bottom). A long-time average of such a stream is expected to converge to a static value that characterizes the room's spatial responsiveness.

2.5.3 Synthesis

The two assumptions above may be considered a thesis and an antithesis. A synthesis of the two could be to define a dynamic ASW(t) and a dynamic LEV(t) resulting when a room with static properties ASW and LEV responds to dynamical content, e.g. music, similar to an impulse response being convolved with a music signal. In order to measure the static ASW and static LEV during running music, we would need to select periods with "pure" early energy and "pure" late energy, similar to selecting periods with high signal-to-noise ratio when measuring sound from sources in general. This synthesis allows the spatial aspects to exist when dL is close to zero even if they are not measurable.

In the results reported in this paper, the chosen criterion for selecting periods dominated by early energy has been dL_M>6dB, and dL_M<-3dB for selecting periods dominated by late energy. However, methods for splitting IACC(t) into continuous IACCE(t) and IACCL(t), i.e. cues of ASW and LEV, like the example in Figure 2 (bottom), are currently being explored.

There is a drawback with the thresholds dL_M>6dB and dL_M<-3dB: Only 1.7% of the periods would be qualified as early energy periods, meaning that ASW will not be observed 98.3% of the time. Likewise, LEV will not be observed in 86% of the time. The problem is not restricted to this choice of threshold values, but is a principal one. It would be unsettling, given our assumption: "- if we can perceive it, we can measure it". Instead of observing ASW and LEV as continuous streams, we would be waiting for impulses or transients on whom we can apply established measurement techniques. It would leave us with ASW being something that cannot be perceived and measured, except for in note-onsets in music. In similar manner, LEV would be something that cannot be perceived and measured, except for in the events of a note release. Moreover, the combination of ASW and LEV cannot be simultaneously perceived and measured in any musical event.

On the other hand, the drawback described above is similar to reverberation time T30 being directly observable only in stop chords or end chords in music, while the listener aspect of Reverberance exists continuously.

In our data, the dL_M-distribution (4) turns out to be largely Gaussian, σ =2.4dB, skewness 0.8dB, $dL_M(95\%)$=4.0dB and $dL_M(5\%)$ is -3.7dB.

2.6 Data recordings

This author has since 2011 collected binaural recordings in occupied concert halls during symphony orchestra concerts, with small microphones in the outer ear canal and a wav recorder. Two graphical presentation of SPL(t) and IACC(t) are presented in Figure 2. In initial studies, special attention has been drawn to impulse-like parts in music, since these are to some degree comparable with results from impulse response measurements, available in literature.

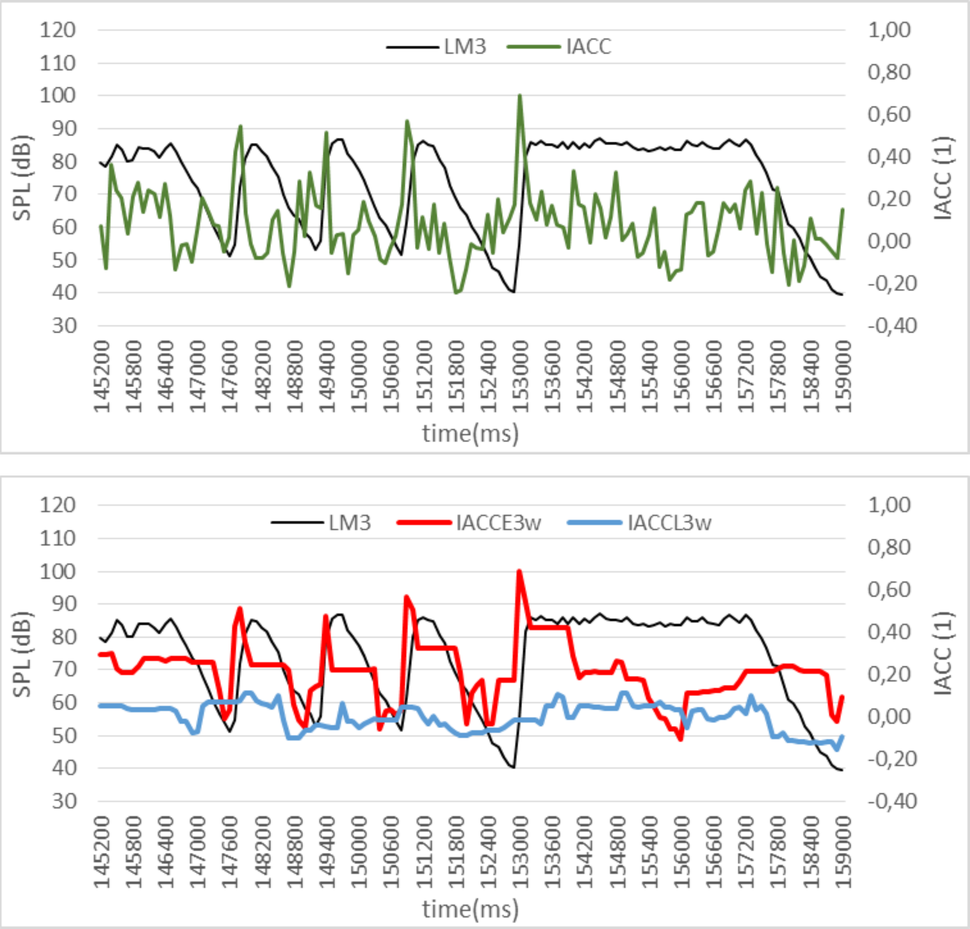

Figure 2: Sound level LM3(t) and IACC(t) in 100ms bins in 500-2k octave bands, 14s Brahms 1st Symphony including stop- and endchords. Bottom: Same part, but example of continuously weighted splitting into early and late IACC, 1s long rectangular weighting window. Boston Symphony Orchestra, in Symphony Hall, 2nd balcony, May 02, 2013

3 Results

In Table 1, results are presented in terms of statistical data for measurements in the following quantities: IACC3(t), IACCE3(t) and IACCL3(t), all of which are time-varying cross-correlation between sound at Left and Right ear, in the 3 octaves 500, 1000 and 2000Hz, and correlation periods T=1/10 second. IACCE3 is measured from periods with dL_M>6dB, and IACCL3 is measured from periods with dL_M<-3dB. A total N=337989 correlation periods measured over 33799 seconds, more than 9 hours, are from over 10 big concert halls in Europe and the US.

In all halls, 95% confidence interval of means are smaller than 0.001 for IACC3, smaller than 0.003 for IACCL3, and smaller than 0.01 for IACCE3. Even if hall-to-hall difference in mean values are small, they are for most halls bigger than the confidence intervals.

Table 1 Results. IACC3, IACC$_E$3 and IACC$_L$3: mean values of respective time-varying functions over N correlation periods; Total measurement duration in seconds is N/10; Phon=equivalent Loudness level; L_M3sd= standard deviation of median energy level; NE% is the percentage of periods with dL_M>6dB; NL% is the percentage of periods with dL_M<-3dB; dL,sd= standard deviation of level steps dL in dB; dL5%= 5%-percentile of dL in dB; dL95%= 95%-percentile of dL in dB; All halls = weighted averages over all halls, except N being the sum of periods from all halls; Bergen=Grieghallen 2012 and 2015; Berlin=Philharmonie 2014; Boston= Symphony Hall 2013; Chicago= Orchestra Hall 2014; Helsinki= Music Centre 2014; Kristiansand= Kilden 2011; New York= Avery Fisher Hall in 2014; Oslo=Konserthus 2011; Paris= Philharmonie 2015; Stavanger= Konserthus 2012; Misc. reference data: AES_off= Norway, 4000cbm hall with RT=0.8s and Acoustic Enhancement System off; AES_on is the same hall with AES system on; "_0"= measurement of typical ambient pre-concert noise in three of the halls. NY_applaus= measurement during applause in New York.

Hall	N	L_M3 Phon	L_M3 sd	IACC3	IACC$_E$3	IACC$_L$3	NE %	NL %	dL sd	dL 5%	dL 95%
Bergen	45480	79	6	0,22	0,49	0,02	2 %	10 %	2,5	-3,7	4,1
Berlin	23164	77	9	0,15	0,42	-0,02	2 %	9 %	2,4	-3,7	4,1
Boston	22301	70	4	0,13	0,37	-0,04	2 %	10 %	2,6	-3,9	4,3
Chicago	54223	75	7	0,15	0,33	0,02	2 %	10 %	2,5	-3,8	4,1
Helsinki	41016	70	7	0,19	0,41	0,04	2 %	10 %	2,7	-4,0	4,4
Kristiansand	7634	-	8	0,06	0,25	-0,08	2 %	9 %	2,4	-3,7	4,1
New York	19153	76	8	0,11	0,25	-0,03	2 %	9 %	2,6	-3,7	4,1
Oslo	4347	-	4	0,14	0,44	-0,05	1 %	7 %	2,8	-3,4	3,1
Paris	31122	71	8	0,08	0,29	-0,09	1 %	6 %	2,1	-3,2	3,4
Stavanger	89549	76	6	0,12	0,31	-0,04	1 %	8 %	2,4	-3,5	3,9
All halls	337989	75	7	0,14	0,37	-0,01	3 %	14 %	2,5	-3,7	4,0
AES_off	494	72	2	0,33	0,56	0,13	6 %	17 %	4,0	-5,5	6,4
AES_on	1290	76	1	0,22	0,52	0,05	4 %	11 %	3,0	-3,8	4,9
Helsinki_0	2394	67	8	0,17	0,41	0,00	2 %	10 %	2,4	-3,8	4,1
Paris_0	1464	92	6	0,06	0,54	-0,07	0 %	3 %	1,5	-2,4	2,4
NY_0	2394	86	2	-0,02	-0,11	-0,08	0 %	2 %	1,5	-2,3	2,4
NY_applaus	1894	92	3	-0,08	-0,13	-0,11	0 %	2 %	1,5	-2,3	2,4

4 Conclusions

Statistical results from a total N=337989 correlation periods measured over 33799 seconds, more than 9 hours, from over 10 big concert halls in Europe and the US are presented. The hypothesis, "Binaural signals, i.e. signals at listeners' pair of ears, can exhibit statistically significant hall-to-hall differences in cross-correlation", is not rejected by the data. Given these results, one should keep in mind that even if binaural cross-correlation is an important cue of spatial aspects, it is not the only one. Sufficient orchestral loudness and Strength (G and G,late) are important prerequisites for the experience of Source Broadening and Listener Envelopment.

In further work, the issue of predicting ASW and LEV in terms of parallel streams will be pursued.

References

[1] Early Radio History. *Scientific American*, December 31, 1881, pages 422–23. *The Telephone at the Paris Opera* Retrieved March 27, 2009.

[2] http://web.archive.org/web/20080807025132/http://news.bbc.co.uk/2/hi/technology/7537782.stm

[3] http://www.rs-met.com/documents/tutorials/StereoProcessing.pdf

[4] Barron M. Late lateral energy fractions and the envelopment question in concert halls. *Applied Acoustics,* 62 (2001), 185-202

[5] Marshall AH. A note on the importance of room cross-section in concert halls. *Journal of Sound and Vibration*, 1967;5:100-12.

[6] S.Klockgether and S. van de Par, A Model for the Prediction of Room Acoustical Perception Based on the Just Noticeable Differences of Spatial Perception, *Acta Acustica United with Acustica*, 100, 964-971.

[7] Bradley JS, Soulodre GA. The influence of late arriving energy on spatial impression. *Journal of the Acoustical Society of America*, 1995,97:2263-71.

[8] Bradley JS, Soulodre GA. Objective measures of listener envelopment. *Journal of the Acoustical Society of America*, 1995, 98, 2590±7.

[9] Barron M. The Gulbenkian Great Hall, Lisbon, II: an acoustic study of a concert hall with variable stage. *Journal of Sound and Vibration*, 1978;59:481-502.

[10] Marshall AH, Barron M. Spatial responsiveness in concert halls and the origins of spatial impression. *Applied Acoustics*, 2000;62(2):91-108.

[11] Barron M, Marshall AH. Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. *Journal of Sound and Vibration*, 1981;77:211-32.

[12] Keet W de V. The inluence of early lateral relections on the spatial impression. *Proc 6th International Congress on Acoustics*, Tokyo, 1968, paper E-2-4.

[13] Barron M. Measured early energy fractions in concert halls and opera houses. *Journal of Sound and Vibration*, 2000; 232:79-100.

[14] Morimoto M, Maekawa Z. Auditory spaciousness and envelopment. *Proc 13th International Congress on Acoustics, Belgrade*, 1989;2:215-8.

[15] Morimoto M, Iida K, Sakagami K. The role of reflections from behind the listener in spatial impression. *Applied Acoustics*, 2000;62(2):109-24.

[16] Reichardt W, Schmidt W. Die hörbaren Stufen des Raumeindruckes bei Musik. *Acustica* 1966;17:175.

[17] E. Kahle: Room acoustical quality of concert halls: perceptual factors and acoustic criteria – return from experience.*International Symposium on Room Acoustics*,Toronto, Canada (2013).

[18] A. Haapaniemi, T. Lokki: Identifying concert halls from source presence vs room presence. *Journal of the Acoustical Society of America* 135 EL311 (2014).

[19] S. Klockgether, J. van Dorp Schuitman, S. van de Par: Perceptual limits for detecting interaural-cue manipulations measured in reverberant settings, *POMA* 19 (2013) 015004.

[20] M. Skålevik: Can source broadening and listener envelopment be measured directly from a music performance in a concert hall? *Proceedings of the Institute of Acoustics*, Vol. 37. Pt.3 2015, paper 34 http://www.akutek.info/Papers/MS_Spaciousness-meter.pdf

[21] International Standard, ISO-3382 Acoustics – Measurement of Room Acoustic Parameters – Part 1, Performance Spaces, 1st edition (2009)