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

SPATIAL LISTENING ASPECTS AND THE TIME-VARYING INTER-AURAL CROSS-CORRELATION DURING MUSIC PERFORMANCE IN CONCERT HALLS

Magne Skålevik

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

Spatial listener aspects like source broadening and envelopment have by many researchers been shown to have big impact on concert goers' preference for the acoustics of a concert hall. These two aspects rely on the amount of lateral reflections arriving early and late, respectively. A concert hall's ability to add source broadening and envelopment to the listening experience during a music performance, can be measured by analysing binaural room impulse responses (BRIR), according to ISO-3382. More recently, this author has investigated the time-varying inter-aural cross-correlation, IACC(t), in binaural recordings during symphony orchestra performances in concert halls, including several well-known halls. The objective is to see whether spatial listener aspects can be observed in features in IACC(t) of binaural signal with musical content. One would like to know how big the hall-to-hall differences are compared to the variation due to music content. Another interesting question arisen during this study is whether source broadening and source localization are mutually exclusive, or not. This paper is a status report from these investigations. A scatter plot of discrete IACC(t) calculated in 100ms periods proved to be useful, revealing spatial perception cues related to hall properties as well as to instruments, orchestra and the music itself. A study of two performances of Tchaikovsky's 4th Symphony, by two different orchestras, one from each side of the Atlantic Ocean, is included.

Keywords: Time-varying IACC, source broadening, listener envelopment, source localization

1. Introduction

There is a common understanding among acousticians[1] 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 offers the sensation that musician and listener is in the same space, enveloped by the same physical environment, while another hall fails to do so. A concert hall's ability to add source broadening and envelopment to the listening experience during a music performance, can be measured by analysing room impulse responses, according to ISO-3382[2]. Early lateral fraction (LF) and early inter-aural cross-correlation (IACCE) are indicators for source broadening, while late lateral strength (LLG) is an indicator for listener envelopment. Beranek (2008)[3] suggested that listener envelopment can be predicted from a combination of late strength (G_{late}) and late inter-aural cross-correlation (IACCL). More recently[4], this author has investigated the time-varying inter-aural cross-correlation, IACC(t), in binaural recordings during symphony orchestra performances in concert halls, including several well-known halls. The objective is to see whether spatial listener aspects correspond to any features in IACC(t). One would like to know how big the hall-to-hall differences are compared to the variation due to music content. Another interesting question arisen during this study is whether source broadening and source localization are mutually exclusive, or not. This paper is a status report from these investigations.

2. Terms and Definitions

For the purpose of this paper, the following definitions are given. Note that the terms and abbreviations may occur with varying definitions in literature.

Table 1: Definitions

Term	Description
BRIR	Binaural Room Impulse Response, measured from an impulse with a pair of microphones at the ears of a dummy head or head of a real person
IACC	Inter-aural cross-correlation, a value between -1 and +1, the correlation between two normalized signals in a binaural pair
IACCE	IACC ‘early’, i.e. IACC from the first 80ms of a BRIR
IACCL	IACC ‘late’, i.e. IACC from the BRIR part after 80ms
IACCA	IACC ‘all’, i.e. IACC from the whole BRIR
IACC(t)	IACC as a function of time, i.e., if not otherwise noted, a sequence of IACC calculated for every 100ms of a binaural signal
Median plane	The vertical plane perpendicular to the lateral axis
$L_M(t)$	Median signal level, the sum signal of the binaural signal, and the level of energy arriving from the Median plane and which contributes to increased IACC, ultimately IACC=1 in absence of lateral reflections
dL(t)	Level step at time *t*, i.e. the difference in signal level (dB) between the RMS levels from the periods [*t-0.1*, *t*] and [*t-0.2*, *t-0.1*]
$dL_M(t)$	Level step in $L_M(t)$
IACCE(t)	Statistic of all IACC(t)-values occurring when $dL_M(t)$ >6dB, interpreted as direct sound and early reflections
IACCL(t)	Statistic of all IACC(t)-values occurring when $dL_M(t)$ < -3dB, interpreted as reverberant sound, i.e. late reflections
IACCE3, etc.	Extension ‘3’ indicates average from octave bands 500, 1000 and 2000Hz
MF	Median energy fraction, 0.5·(1+IACC)
BLF	Binaural Lateral Fraction 0.5·(1-IACC)
LF	Early lateral fraction, the ratio of early lateral energy to all early energy, measured with a figure-8 microphone and an omni-directional microphone, from the first 80ms of an impulse response
LLG	Late lateral Strength, i.e. the lateral energy in the part after the first 80ms from an impulse response, normalized to the energy of direct sound from the same source at 10m distance
G_{late}	Late Strength, i.e. the total energy in the part after the first 80ms from an impulse response, normalized to the energy of direct sound from the same source at 10m distance
ASW	Apparent Source Width, or Source Broadening, the perceptual effect that a sound source appears wider than it is, and the it tends to broaden with increased loudness, depending on early lateral reflections, thus measurable with LF or 1-IACCE
LEV	Listener envelopment, the sonic impression that musician and listener is enveloped in the same space, depending on the strength of late lateral reflections, thus measurable with LLG or a combination of 1-IACCL and G_{late}
m, *s*, *CI*	Statistical parameters: average, standard deviation and confidence interval around *m* at 5% significance level, thus for population *N*, *CI* is $m \pm 1.96 \cdot s/N^{0.5}$

3. Measurement setup and data collection

Since 2011, this author has collected binaural recordings from concerts with symphony orchestras in 10 big halls, including famous concert hall in Europe and the US. So far, more than 10 hours, i.e. more than 36.000 seconds with 360.000 periods have been collected. A schematic presentation of the measurement setup is given in Figure 1.

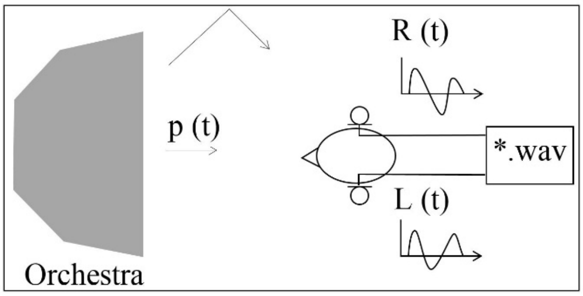


Figure 1: Measurement setup.

4. What can we learn from IACC(t) ?

IACC(t) takes values between –1 and 1, and can be interpreted as an indicator of balance between median energy and lateral energy, i.e. Median-Lateral (M-L) balance. At 1 there is pure median energy, no lateral energy. At 0 they are perfectly balanced, i.e. median energy equals lateral energy. At –1 there is only lateral energy and no median energy. At IACC(t)=0.5, there is 75% median and 25% lateral energy; at –0.5 there is 25% median and 75% lateral energy, as follows from definitions of MF and BLF.

Whenever direct sound dominates, the IACC(t) would take high values. However, the same could occur from vertical reflections, typically from floor and ceiling, and the direct sound path could even be obstructed, e.g. with a music instrument in an orchestra pit. This means that the direct sound component at, instant t, cannot be determined from the value of IACC(t) alone. Our brain combines several cues in order to localize a sound source.

Toward lower frequencies, the binaural signals are generally more correlated. E.g., a pure tone source at lateral direction would, in anechoic environment, produce the values –0.92@1000Hz, –0.04@500Hz, 0.37@250Hz and 0.71@125Hz. The lower the frequency, the less the maximum possible phase difference, and the less the difference in IACC for a given difference in azimuth angle. Inherently, azimuth detection accuracy decreases toward lower frequencies, and ultimately, median sound energy cannot be distinguished from lateral sound energy. Perceptually, low frequency sounds appear to arrive from 'all around', and therefore their sources appear broader than they are. For this reason, low frequency reflections contribute to ASW and LEV even if non-lateral.

In short: IACC(t) indicates M-L balance. The more direct and vertical reflections dominate over lateral reflections, the higher IACC(t), and vice versa. Our brain interprets lower IACC(t) as a broader sound image (more ASW and/or LEV), while high IACC(t) as a narrow or frontal sound image.

4.1 Gaussian behaviour and large fluctuations

One of the first, informal observations when plotting IACC(t) was the great fluctuations and apparently stochastic behaviour of the data. This called for some basic statistical analysis: In the 'average hall', as well as in the sub-populations of each hall, a Gaussian distribution is seen. Statistics, $m\pm s$, of IACC(t) in octave bands 500, 1000 and 2000Hz, are on average 0.24 ± 0.29, 0.10 ± 0.25 and 0.11 ± 0.22 around *0.25* in all halls. If we compress the information in these octave bands to the single number quantity IACC3(t), which is simply the average IACC from the three octaves at each instant, the standard deviation is reduced to 0.16. Mean of IACC(t) is $m=0.14$ in the average hall and *0.06-0.22* in the individual halls, thus the largest hall-to-hall difference observed (*0.16*) equals one standard deviation in the instant-to-instant variation during a concert. Here, 'instant' means a 100ms period.

In order to see any significant hall-to-hall differences, we would need to have confidence intervals (*CI*) around each hall mean that are considerably smaller than 0.16. E.g., in order to divide halls into 8 categories based on IACC(t), we would need *CI* in the order of ± 0.01 or smaller. From our definition of *CI*, given $s=0.16$, the minimum number of correlation periods needed is at least N=1000, as can be verified by $\pm 1.96 \cdot 0.16/1000^{0.5} = \pm 0.01$. With 0.1s periods, we need to measure for a duration of at least 100 seconds to obtain a mean value with this kind of accuracy (significance level 5%).

4.2 Onset, sustain and release of notes

Analysis of transients or impulse-like sounds from brief tutti notes or so-called stopped chords in the orchestra is a common way to measure acoustical parameters in concert halls during concerts. It seemed obvious to try a similar approach when studying IACC(t). One would suspect a tutti onset to contain cues of ASW, and a tutti release to contain cues of LEV.

A graphical presentation of IACC3(t) on the background of the median energy level, LM3(t), in Symphony Hall, 2nd balcony, May 02, 2013, during 14 s with stop-chords in Brahms 1st Symphony, played by Boston Symphony Orchestra.

We clearly identify the four powerful onsets where LM3 steps up 30-40dB, with corresponding peaks in IACC(t) around 0.6, which distinctly localize a source in the median plane. Except from at these peaks, IACC3(t) in the diagram seems to fluctuate between -0.2 and 0.3. There is no obvious difference between the release parts (the five decay slopes in LM3) end the rest. However, a closer look reveals that average IACC3(t) is 0.01 in the 5 decays, 0.08 in the sustained part, and 0.13 over the whole symphony. In all these parts, standard deviation is approximately 0.22. Compared to this, average BRIR values in the hall are IACCL3=0.15, IACCA3=0.21, and IACCE3=0.39. Since IACCA is calculated from a mix of early and late part of the BRIR, it is natural that its average is between the ones of IACCE and IACCL. Similarly, the sustained part as well as the whole symphony is a mix of early sound and late sound, thus it is natural that their average IACC(t) values (0.08-0.13) lie between those of onset (0.6) and decays (0.01).

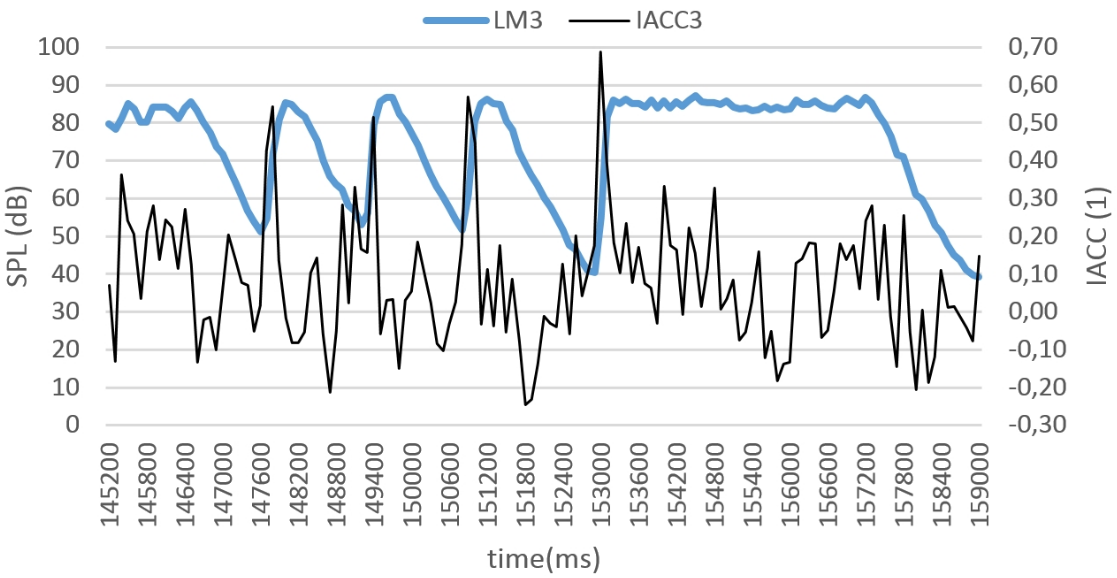


Figure 2: Orchestral tutti notes with onsets, sustained parts and release parts (Boston 2013).

IACC3(t) values at onset is not directly comparable with IACCE3 from BRIR, since the duration of onsets are much longer than that of the impulse in BRIR. This becomes evident by ‘zooming in’ on the onsets in Figure 2, as presented in Figure 3. From the LM3 curve we observe that the levels are rising over a period of 400ms in the first three examples and over 300ms in the fourth (rightmost) one.

The rise phase is very interesting and deserves more attention: A steep rise in LM3 between two periods means that there is much more median energy in the latter period than in the former one, implying that direct sound and early reflections will dominate the latter period.

We suspect that in the steep-rise periods, IACC(t) should be sensitive to early lateral reflections in the same way as IACCE is sensitive to lateral reflections in the early part of BRIR.

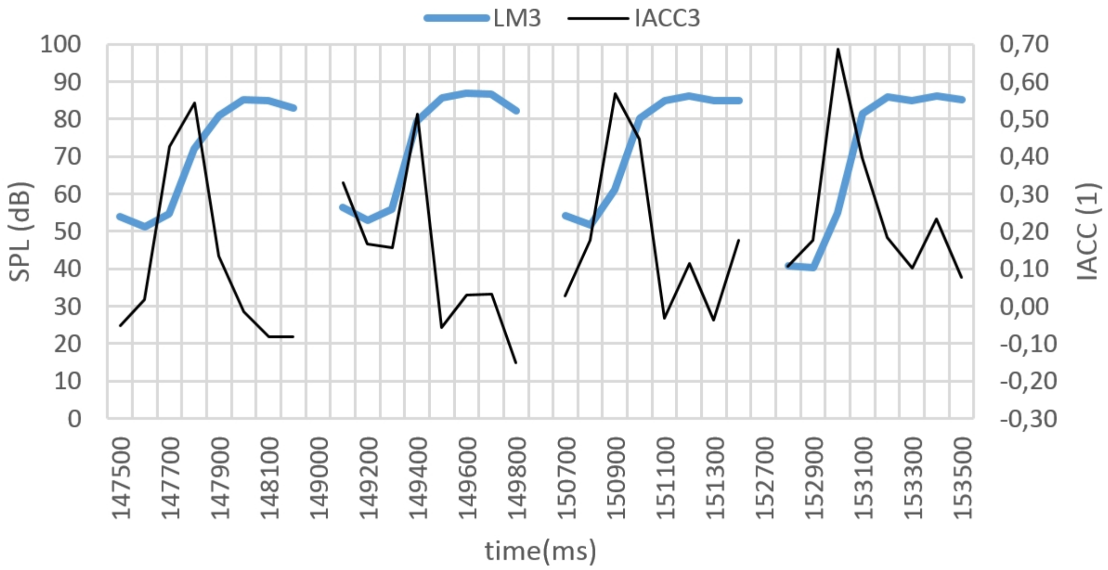


Figure 3: The four onsets in Figure 2.

Given the definitions of MF and BLF, IACC3(t) =0.50 means 75% median energy (MF) and 25% lateral energy (BLF); 0.70 means 85% median and 15% lateral energy. The BRIR value in the empty hall, IACCE3=0.39, means 70% median energy and 30% lateral energy. Bear in mind that the extension ‘3’ means average of the three mid-octave bands 500, 1000 and 2000Hz, and thus the BLF3 is not comparable with Lateral Fraction (LF), who’s single number quantity is calculated from the four octave bands 125-1000Hz.

On this background, we return to Figure 3, observing that only the steepest parts of the rise-phases have particularly high IACC3(t), i.e. high median-to-lateral energy balance. Towards the completion of the onset, the sound energy of the instruments will have filled the auditorium, and direct sound and early reflections are no longer dominating the sound field like they did in the steepest rise-phase of the onset. This can be observed from the IACC(t) behaviour right after its peaks, quickly returning to the fluctuations around zero. Now, how steep must a rise in LM3 be in order to qualify as a cue of so-called early sound? Our first idea was to collect early sound data that is comparable to those acquired with BRIR, meaning we had to collect periods containing pure early sound. Theoretically, dLM>10dB should be sufficient to ensure that at least 90% of the energy is ‘new’, i.e. early median sound energy that was not present in the previous 100ms period. In Figure 3 there are 6 periods having dLM>10dB, and they all have IACC3(t) peak values in the range 0.40-0.69, statistically 0.53±0.10. We do not know the IACCE3 at the seat where the binaural signal was recorded, but the hall average is 0.39. If reducing the criterion to dLM>6dB, 8 periods would be qualified, but these would include two values close to zero, statistically resulting in peak values 0.40±0.25.

To sum it up, the 6dB criterion returns peak averages closer to the hall average IACCE3, while the 10dB criterion returns peak values with higher average and smaller variation. The 10dB criterion seems more trustworthy, but it would come with new problems: Onsets like those in Figure 3 are very rare. We know that the 6dB criterion would qualify less than 2% of all periods in our 10 hours of data from 10 concert halls. With the 10dB criterion there would be even fewer qualified periods, a smaller data base for analysis of early sound, thus harder to achieve significant results. It could provide us with a way of measuring IACCE in occupied conditions, but we would fail to discover cues of ASW as a continuous aspect in running music.

4.3 What does a symphony look like through the eyes of IACC(t) ?

Fortunately, we have two binaural of Tchaikovsky's 4th Symphony (T4), one with State Academy Symphony Orchestra St. Petersburg in Stavanger Concert Hall (Fartein Valen) in 2012, and one with Chicago Symphony Orchestra in Orchestra Hall, Chicago in 2014. Figure 4 offers plots of IACC3(t) from the two performances, consisting of more than 25000 blue dots, together with level curves of LM3(t) for reference. Top to bottom: LM3(t) Stavanger, IACC(t) Stavanger, IACC(t) Chicago, and LM3(t) Chicago. Horizontal axis is time from 0 to 2520 s (0-42 min). Letter indexes are related to comments in the following.

The IACC3(t) plots consist of more than 25000 (blue) dots, forming two elongated clouds with varying vertical scatter around an average of 0.09 in Stavanger and 0.15 in Chicago

Note the different tempi. Diagrams are stretched to align at start and end, and the times 600, 1200, 1800, 2400 seconds, i.e. every 10 minutes, are indicated in each diagram. Statistical analysis of the clouds reveals 0.09±0.16 in Stavanger and 0.15±0.16 in Chicago, indicating a broader sound image in the former than the latter. However, the clouds have equal standard deviation s=0.16.

Solo parts, e.g. flute or oboe, render wide vertical scatter, while big string section parts or homogenous tutti render narrow formations. The lowest values occur in quiet parts with low pitch instrument sections like low strings and low brass, rendering scattered dots underneath the main cloud.

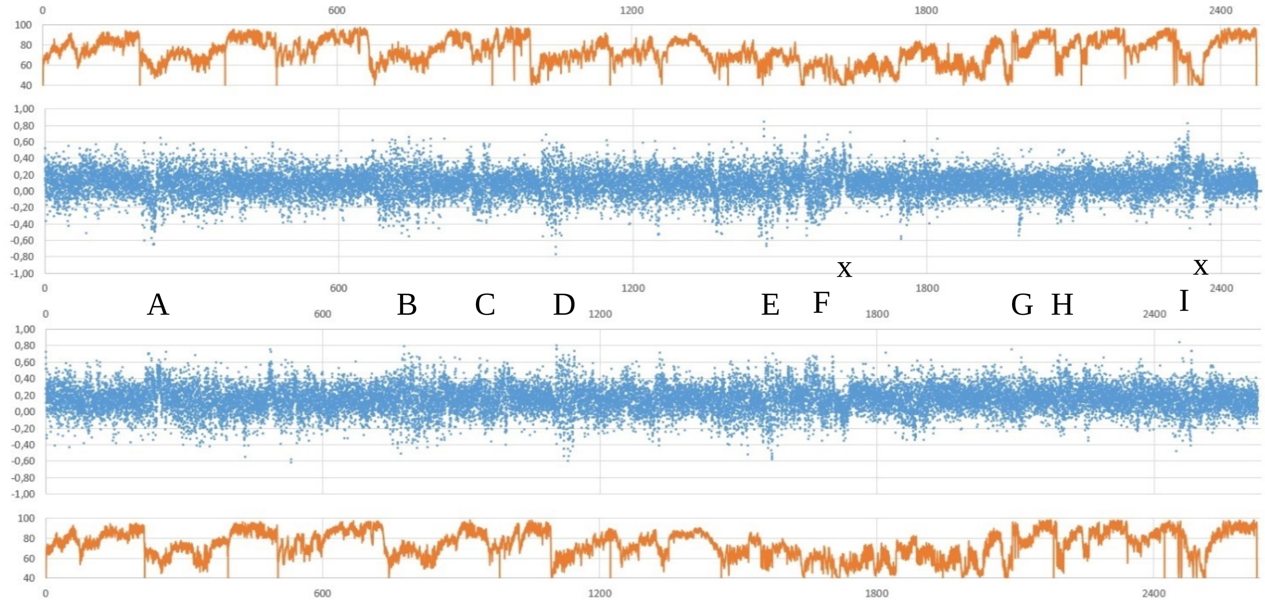


Figure 4: Tchaikovsky's 4th Symphony. Top to bottom: LM3(t) Stavanger, IACC(t) Stavanger, IACC(t) Chicago, and LM3(t) Chicago. Horizontal axis is time from 0 to 2520 s (0-42 min). Letter indexes: see text.

Brief pauses in the performance, where the level curve dips (x), are sensitive to random noises from audience or occasional hiss in the signal chain, thus any details there should be ignored.

Interestingly, the halls exhibit many similarities in vertical scatter patterns (B,C,D,E), and some dissimilarities like the pattern right after the first sudden level drop from *ff* (bar 104 in the score) (A): Both Stavanger and Chicago have large vertical scatter in this clarinet and bassoon solo part, but in Stavanger the cloud shifts downward, while in Chicago it shifts upward. In Stavanger, there is lower M-L balance in clarinet and bassoon, and in Chicago it is higher (more median and/or less lateral sound). Stavanger exhibit lower M-L balance in general, thus lower average and lower values for individual instruments (F,G,H), with one exception: (I) Right after last climax, before the grand finale, the cloud in Stavanger shifts upward and is more scattered than the one in Chicago. This part is dominated by a fanfare in trumpets and horns. Stavanger seems to be generally more sensitive to the directivity of trumpets than Chicago, thus the upward bended formations as trumpets and trombones sets in toward climaxes, e.g. before B, H and I. This may of course vary over the audience area.

A more detailed analysis is given in Figure 5. A pair of diagrams, one for Stavanger – 'S' to the left, and one for Chicago – 'C': A triplet of vertical bars represents the distribution of IACC(t) in 500, 1000 and 200Hz, respectively. Each diagram has four triplets, one for each of the following parts: 'tutti' – the 14s tutti climax before (B), followed by a 17s 'pause', then a 42s 'oboe' solo at (B), and for reference, 'whole' – the entire symphony. Each bar defines the normal interval [*m-s*,*m+s*]. Note that IACC(t) in 500Hz is relatively higher in Chicago. In Stavanger, the triplet spectrum is more 'flat', besides being on average 0.04 lower than in Chicago. Oboe solos have different spectra than the rest. Scattering is bigger (high bars) in solo, and smaller in tutti, than in the Symphony as a whole.

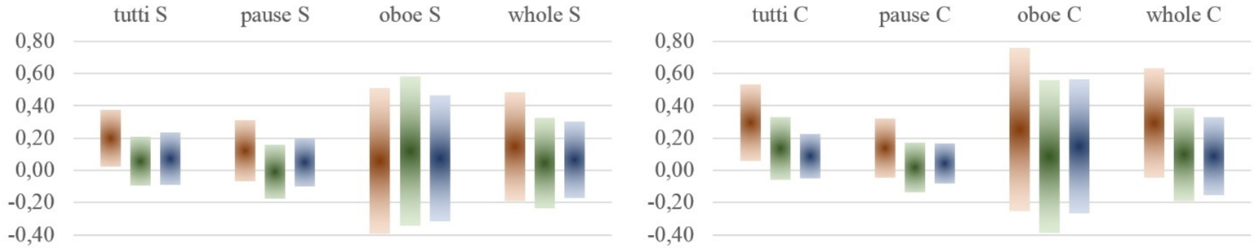
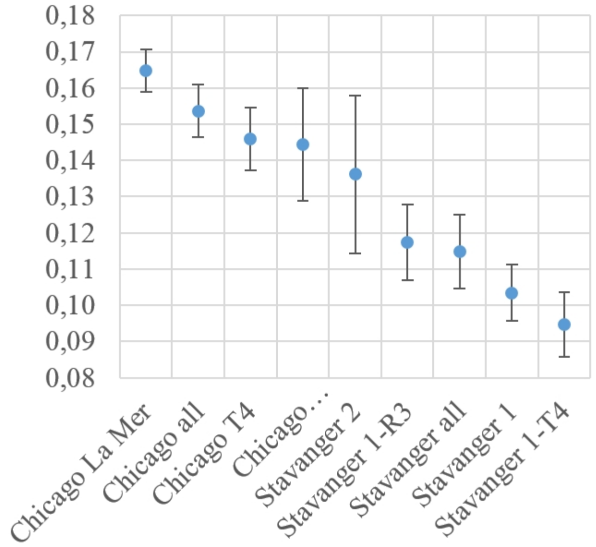


Figure 5: IACC distribution in 500Hz, 1000Hz and 2000Hz, in 4 parts of T4; Stavanger (right) and Chicago.

For a comparison of the vertical scatter throughout the two performances, we divided them into 12 parts, calculated the 2x12 standard deviations in the 1kHz and 2kHz octave bands, and found that they correlated strongly, with r^2 equal to 0.80 and 0.85, respectively. See Figure 6 (right).

Summarized, overall wideness of sound image can be seen in the centroid level of the cloud, i.e. mean IACC(t). High dots in the plot are cues of direct sound and accurate localization, naturally occurring in solo parts or when highly directive instruments peak through in tutti parts. Low dots are ques of lateral reflections and wide sound, and sometimes even false or blurred localization if very low. In solo parts, we see both high dots and low dots, i.e. strong fluctuations in IACC(t), that may explain why perception of solo instruments can combine localization and ASW. A preliminary result from an investigation by the author suggests that clarity and localization in a considerable number of halls correlate more with the standard deviation of IACC(t) in 1kHz and 2kHz octave bands than with any other parameter in the investigation.



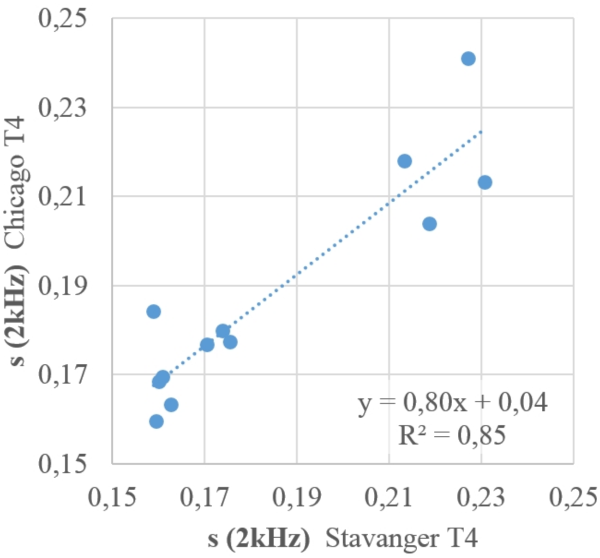


Figure 6: Stavanger vs Chicago, left: Averages and confidence intervals of IACC3(t); Right: Standard deviation of IACC(t) in 2kHz octave band, in 12 parts of Tchaikovsky's 4th, see text.

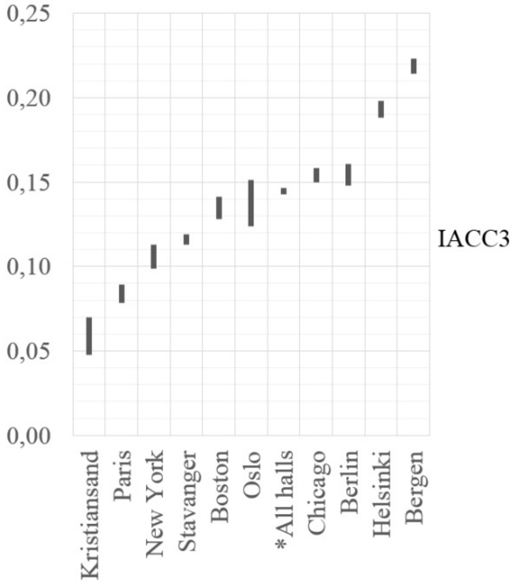
It remains to do a statistical exercise as to the duration and musical content of measurements necessary to get results that characterize the hall, without being too sensitive to the music content of the measurement. Figure 6 (left) illustrates the problem – measurements in the same hall but from different concerts or sessions produce different results. Note that this problem is not necessarily a weakness of the measurement method – people often need more than one performance to make up their minds as to their subjective opinion of a concert hall.

The Chicago concert has three parts – La Mer (Debussy), Tempest (Tchaikovsky) and 'T4' - Tchaikovsky's 4th; Stavanger 1 and Stavanger 2 are two full-length concerts. 'R3', Rachmaninov Piano Concert No 3, was played before the intermission, while T4 was played after the intermission, both in Stavanger 1. 'Chicago all' and 'Stavanger all' are based on the total amount of data from each hall. If we just attended 'Stavanger 2' concert and 'Chicago Tempest', we would not be able to distinguish the halls with certainty (5% significance level). However, after attending both concerts, 'Stavanger all', the total result would distinguish clearly from any individual or total result in Chicago.

4.4 Comparison of 10 big concert halls in Europe and US

In Figure 7 (left), confidence intervals (95%) from measurements of IACC3(t) in big concert halls in Europe and US. Total measurement time is 10 hours, but varying from hall to hall, as can be seen from the size of CI's. We note that at least 8 of 10 halls are distinguishable. From knowledge about lateral reflections, results are to be expected: 4 of the 5 halls with lowest values are rectangular halls, and the one with highest value has a pronounced fan shape. Non-rectangulars are in-between, except Paris, where early lateral reflections had emphasised attention throughout the planning of the hall.

Figure 7 (right) shows results from an attempt to separate cues of early sound (IACCE3) and late sound (IACCL3) from the IACC3(t) data, by selecting periods with dLM>6.0dB and dLM< -3dB, respectively, ref. Section 4.2. The criterion dLM>6dB resulted in less than 2% of the periods being selected, explaining why the CI's are much wider in 'IACCE3' than in 'IACC3'. Similar explanation applies to 'IACCL3'. These two do not alter the ranking provided by IACC3, but may suggest that Chicago has more ASW than Berlin, even if equal in IACC3, and that Paris' low IACC3 may not be due to high ASW, but due to high LEV. True ASW and LEV cues are currently being explored.



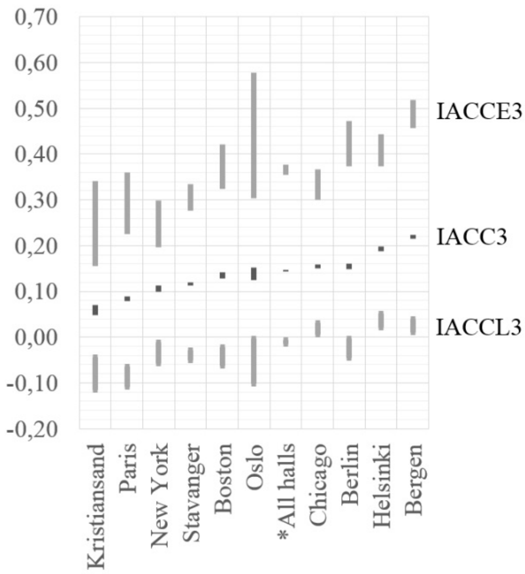


Figure 7: Confidence (95%) intervals around means of IACC3(t) from big concert halls

REFERENCES

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

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

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

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