

MUSIC, ROOM, TWO EARS, DESIGN AND PARADIGMS

M Skålevik AKUTEK and Brekke & Strand

1 INTRODUCTION

Concert goers with normal hearing listen to the music with their two ears. Music has been composed and performance spaces have been designed for listeners with binaural hearing.

Preference for music genre, composers, and their different works, varies over the population, and so does preference for certain opera houses and concert halls. While some concert goers are lovers of clarity and details in the music, others prefer to be immersed in well blended orchestral sound, resulting in individual preferences for seating areas and listening positions. These musical and physical differences correspond to differences in the binaural signal.

This paper presents an update on findings in the Binaural Project since it started in 2011, pointing at different perceivable features in the binaural signal, some due to differences in music itself, others due to different rooms or different listening positions.

As to the discussion on design paradigms - justification of any design, be it within a paradigm or not, should take the design's influence on the binaural signal into account. Listening positions and orientation relative to the room and musicians are more important than the geometry of the room itself.

2 BINAURAL HEARING, PERCEPTION AND METRICS

2.1 Basic concepts of binaural hearing

Hearing with two ears is referred to as binaural hearing, and the pair of sound signals that arrives at the two ears as a binaural signal. By nature, in binaural hearing, the brain compares the signal pair for differences and similarities. Binaural hearing allows the brain to learn to interpret the binaural signal to extract information from incident sounds to the two ears. The process has similarities with drawing a sketch or a map of the surroundings from the visual impression obtained with two eyes. So-called binaural mapping can be performed by the brain over a vast span of complexity – from simply localizing a voice in an-echoic environment, via localizing the same voice in a reverberant room, to the so-called cocktail-party effect in which the brain can single out one source among a multitude of other voices. Live orchestra music perception is a particularly complex task to the brain for two reasons – 1) there are many sources and 2) each of them initiates a vast number of reflected sounds from all angles and with various delays, i.e. reverberant sound.

2.2 Binaural hearing related to a variant coordinate system

In binaural hearing, the listener's brain is the origin of a mobile, ever-changing, variant and instant 3D-coordinate system that could be defined with the x-axis pointing straight ahead as the nose is instantly pointing, the y-axis running through the pair of ears, and the z-axis pointing upwards. The x-z plane is termed the median plane (in medicine also known as mid-sagittal plane). Sound incidence in the median plane, conveniently termed *median sound*, would create equal signals at the two ears, thus the Inter-Aural Cross-Correlation (IACC), a metric that takes values in the interval $[-1, +1]$, would approach +1.

Incident sounds with a y-component different from zero, are commonly termed *lateral sound*, and would cause a difference in the binaural signal, partly due to the sound path length difference causing a difference in arrival time, as measured by the inter-aural time difference (ITD), partly due to the head leaving one ear in the sound shadow and the other ear in the highlight zone, as measured by the inter-aural level difference (ILD).

An inherent shortcoming in binaural hearing is the inevitable *cones of confusion*. Imagining a cone symmetrical around the y-axis, sound incidence anywhere along this imaginary cone would theoretically create the same ITD and the same ILD, only slightly modified by the asymmetrical pinnae surrounding the entrances of the ear canals. The set of possible cones with various angles relative to the y-axis constitute the set of cones of confusion. Note that as this angle approaches 90 degrees, the cone approaches the median plane.

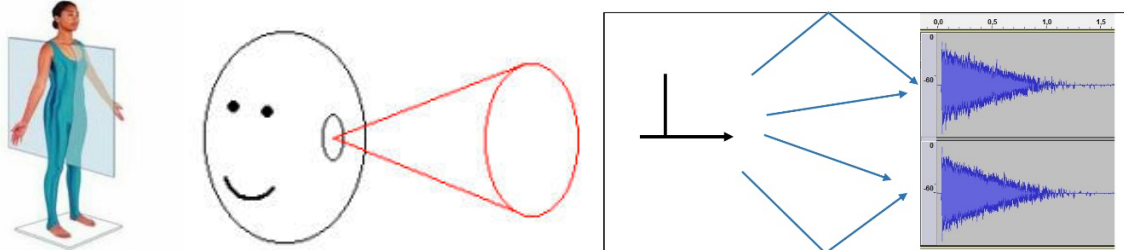


Figure 1, from left, a) median plane, b) cone of confusion, c) Binaural Impulse Response (BRIR)

2.3 Complex sound fields

While ITD and ILD are meaningful in very simple sound fields where sources are easily separated, the various applications of IACC is more useful for measuring complex sound fields like in an auditorium, concert hall or in a cocktail-party analysis task.

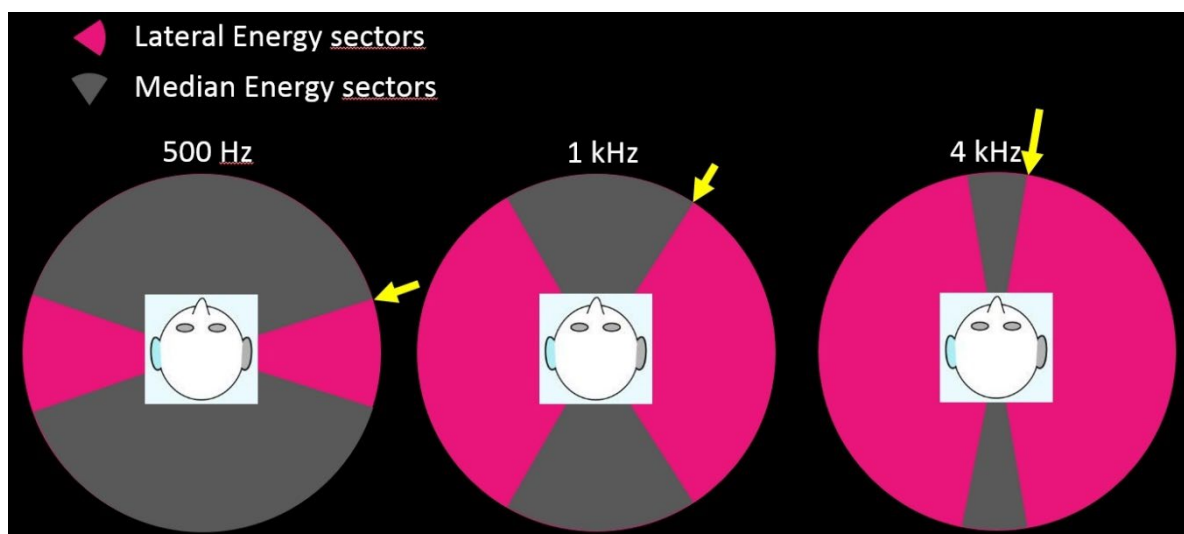


Figure 2 Borders between sectors of median and lateral energy are frequency dependent.

Median sound and lateral sound are theoretically straightforward concepts, but they do in real-life only exist as approximations, limits or references for sound fields consisting of a blend of median sound and lateral sound. Toward high frequencies, i.e. short wavelengths, even sound incidences close to the median plane would cause a difference in the binaural signal, and thus $IACC < 1$. Conversely, towards low frequencies, i.e. long wavelengths, incident sound would cause little difference at the two ears, easily confused with median sound with $IACC$ approaching +1.

A more practical division between median and lateral sound incidences are the sectors shown in Figure 2, where median sound is sound incident in the median sector and lateral sound is sound incident in the lateral sector. Importantly, the division between the sectors is frequency dependent, as indicated by the arrows. As to the $IACC$ metric, sound incidence in the median sector would increase the $IACC$ value, while sound incidence in the lateral sector would decrease the $IACC$ value. An extra sound incidence at the limits indicated with an arrow would leave $IACC$ unchanged.

As a curiosity with little practical relevance, it should be mentioned that IACC can theoretically approach minus one, e.g. if a pure tone arrives with half a wavelength difference at the two ears – leaving the binaural pair in anti-phase.

2.4 Metrics for binaural sound

A collection of binaural sound metrics based on mathematical cross-correlation is presented in Table 1. Cross-correlation is applied in many scientific fields to look for common features in a pair of data sequences. In signal processing, the data set is typically a sequence of samples, i.e. a signal, which means that the data are ordered chronologically. All metrics in Table 1 belong to the important special class of *normalized* cross-correlation, meaning that any bias like signal gain differences or long-term differences in ILD do not affect the result, and that the output is in the interval $[-1, +1]$.

Note that the Lateral Fraction (LF) metric is sometimes listed among binaural sound metrics and frequently used to describe listener aspects in concert halls. IACC and LF are statistically and in some sound-fields mathematically related to one another. Since LF is measured with a Mid-Side (Blumlein) stereo coupling consisting of a pair of omni and figure-8 microphones, they do not record a binaural signal as such and is therefore not included in this table.

Table 1 Metrics for binaural sound

Metric	Description
IACC	Normalized inter-aural cross-correlation, measures the similarity between the normalized Left and Right signals over an arbitrary period of duration T . In the normalization, each signal is divided by its RMS-value from the period T , implying that any bias or long-term differences in ILD do not affect the result, and that the output is in the interval $[-1, +1]$. +1 means that the two signals in the binaural pair are equal, -1 that they are in opposite phase (or polarity), while 0 means that the similarities and dissimilarities during T are weighing one another out. An interpretation of the latter is that incident median sound energy is equal to lateral sound energy in the measured period T .
IACCF(τ)	Inter-aural cross-correlation function, a set of IACCs from one and the same period, calculated with a set of various lags in one signal relative to the other, between lag limits $+\tau$ and $-\tau$, which allows arbitrary azimuth angles (each corresponding to unique values of τ) to be analyzed. Note that technically, to calculate IACCF with lags up to τ on a signal period of netto length T , a brutto data period of length $T+\tau$ is required.
IC	Interaural Coherence is the special case of IACCF where $\tau = 0$, interpreted as the amount of total energy arriving in the median plane.
IACC _{max}	The maximum IACCF returned from all the lags computed in the interval $[\tau, -\tau]$, computed from a binaural signal period of T requiring a practical measurement duration of $T + \tau$
BRIR	Binaural impulse response, a signal pair resulting from a binaural recording of a room acoustic impulse response, Figure 1 c
IACCE	IACC early, i.e. IACC _{max} from the first 80ms of a BRIR
IACCL	IACC late, i.e. IACC _{max} from a BRIR after 80ms
IACC(T, t)	A sequence of IACC _{max} from successive windows of length T , each window beginning at t and ending at $t+T$, meaning that IACC (T, t) is a sequence of samples where T is the sampling period.
IACC (t)	When not otherwise noted, $T=100ms$, and IACC (t) = IACC($0.1, t$)

3 THE BINAURAL PROJECT

The Binaural Project was launched in 2011, with the aim to gain more insight in the relationship between perceptual aspects in concert listeners and features in the binaural signal. A basic activity has been to acquire binaural data for the analysis, in particular to collect binaural recordings from concerts with symphony orchestras in big halls.

As a part of the Binaural Project, Skalevik⁷ 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. Many questions have been raised, including the following:

- What does a plot of $IACC(t)$ look like?
- What are the optimum window T and frequency range settings as to perception relevance?
- Can cues of Localization, ASW and LEV be seen in $IACC(t)$, and if so - how?
- Are there any typical values and statistics of $IACC(t)$ when listening to a performance?
- How big are the hall-to-hall differences compared to the variation due to music content?
- What if the same music is played in two different halls?
- How does $IACC(t)$ respond to changes in listening distance?
- How does $IACC(t)$ respond to changes in surfaces, absorption, G or reverberation time?
- Do differences in IACCE or IACCL correspond to differences in $IACC(t)$?
- How can sources be localized in concert halls where direct sound is often more than 10dB weaker than reverberant sound?
- Are source broadening and source localization mutually exclusive aspects?
- What kind of spatial listener aspects are present in sustained notes or other quasi-stationary sound?
- Can parallel streams of listening aspects be traced back to the single data-stream of $IACC(t)$?
- Do spatial listener aspects correspond to any features in $IACC(t)$ at all?

4 SOME RESULTS FROM THE BINAURAL PROJECT

4.1 Examples of time plots, spectra, and statistics

In Figure 3, the cloud of (blue) dots is a typical example of how $IACC(t)$ fluctuates during music. The black curve is the listening level in dB . Note that the $IACC$ -cloud is a wave-like banner, thinner in top and bottom, and denser in the middle. One typical feature is that the wavy cloud seems to be in anti-phase with the curve level. At the very left we see that during the opening crescendo, the cloud starts reaching up to 0.80-0.90, descending to below 0.60 when the crescendo peaks. Then, as the level drops, the cloud reaches up to 0.90 around 110 seconds. At most of the major level-peaks in the diagram, the cloud is basically found below the level curve, and wherever the cloud reaches the 0.8-0.9 region, the listening levels are moderate. Broadband filter passband is 400-2500Hz.

Figure 6 presents a typical octave band analysis of the $IACC(t)$ from the whole concert.

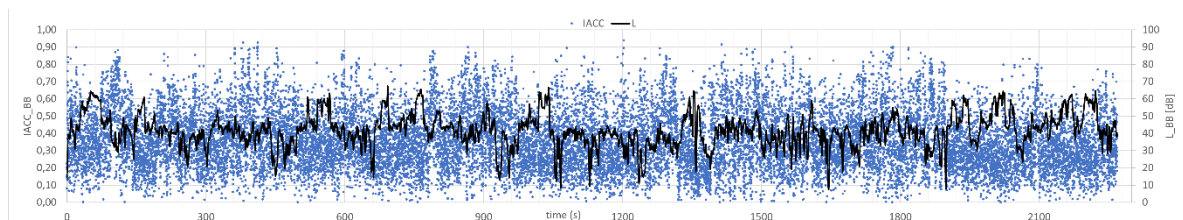


Figure 3 Boston Symphony Orchestra plays Brahms Violin Concerto in D major in Symphony Hall. The cloud of (blue) dots are $IACC(t)$, while the black curve is smoothed (2s) level in dB , in second balcony. The horizontal axis is time in seconds. 100ms $IACC$ -windows. See text for interpretation.

Denser energy spectra decorrelates the sound fields more and reduces $IACC$. Stronger parts have denser spectra because as a larger number of instruments and because they play stronger, see

Figure 4 (middle). On the other hand, soft parts are created by letting 1 or a few woodwind instruments play the leading voice while others play softly in the background, resulting in very few harmonics, and high IACC. Indeed, at 111-113 seconds a flute solo F#4 note Figure 4 (right) has IACC fluctuating up to 0.83.

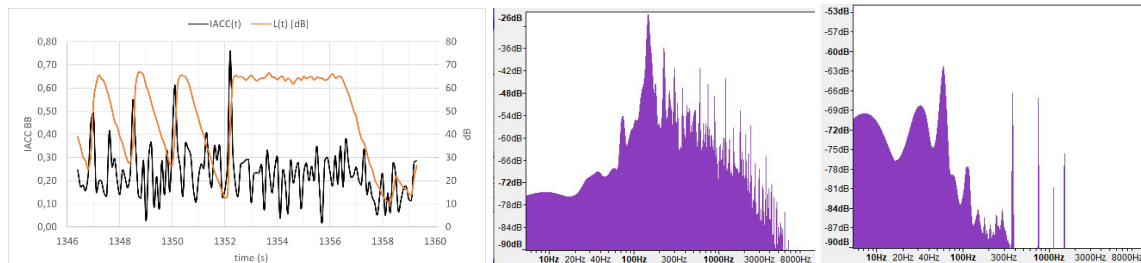


Figure 4 (left) extract 1346-1358 from Figure 3 with stop chords; (middle) spectrum of tutti orchestra in sustained part 1352-1356 s; (right) flute spectrum at 111-113 s in Figure 3;

Stop chords, like those in Figure 4 (left), provides a chance to explore IACC in parts of pure early and late energies. In the four onsets IACC (t) takes peak values 0.49-0.55-0.61-0.76 respectively. Each of the four IACC-peaks all occur in the onset of the chord, 1-2 windows (100ms each) before the level peaks. Except for the brief single window peaks, IACC fluctuates around a relatively low average. In the sustained chord with timpani roll 1352-1356 s, the fluctuations can be expressed as $IACC=0.21\pm0.11$, compared to overall for the whole concert, $IACC=0.35\pm0.17$. Note that both the numbers in this extracted part, 0.21 and ±0.11 , are smaller than the overall ones. This is typical for the stronger parts in concerts and is due to the aforementioned fact – the high spectrum density from all instruments playing *forte* forces the sound field to be decorrelated – thus the lower IACC level, and thus the narrower span of fluctuations.

In experiments with signals from a loudspeaker we have measured $IACC=0.63\pm0.18$ from an oboe solo and $IACC=0.44\pm0.04$ from pink noise, see Figure 5 (left). The diagram to the right compares the “loudspeaker oboe” with a real oboe during a live performance with the exact same source and receiver positions. Solid lines are average values, and the shaded boxes indicate fluctuation range in terms of standard deviations around the average values. A slightly narrower fluctuation range in the real oboe than in the loudspeaker oboe is to be expected due to the soft accompaniment from the full string section making the spectrum denser than with a pure oboe solo.

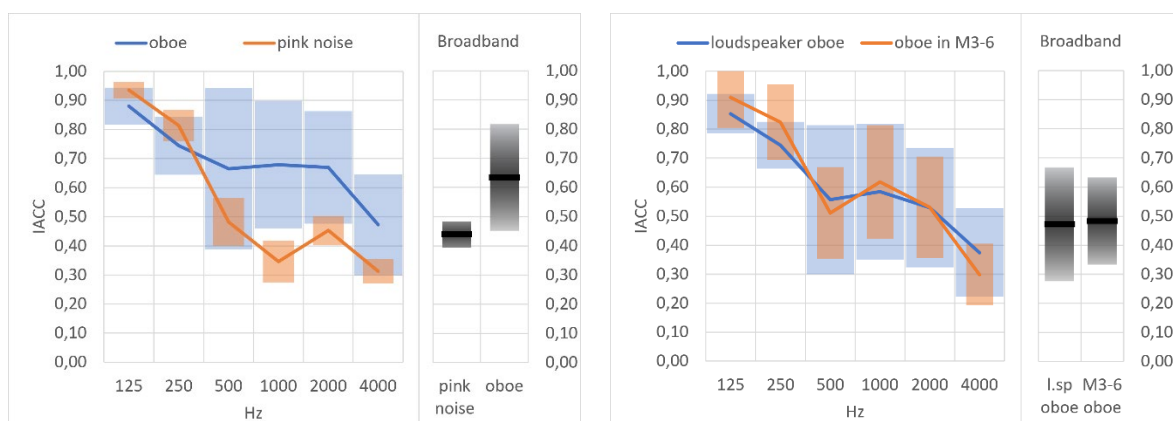


Figure 5 (left), experiments with signals from a loudspeaker, an oboe solo and a pink noise solo, with exact same source receiver positions, at distances 15-19min in a 450 seat 6500m³ hall; (right) same loudspeaker as in the experiment to the left, and a real oboe playing a 9s solo part during a live performance of Mahler's 3rd Symphony, 6th movement, both measured with exact same source and receiver positions, distance 26m, Oslo Concert Hall (19.000m³), May 2022.

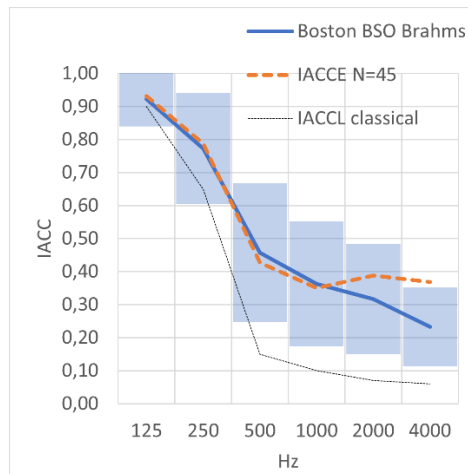


Figure 6, solid curve is IACC average spectrum in octave bands from the whole Brahms Violin Concerto in Figure 3; shaded bars indicate standard deviation, where upper edges measure strength and density of localization cues, while lower edges measure strength and density of envelopment cues. IACCE and IACCL are spectra of early and late parts of binaural impulse responses in Beranek's data collection⁷. This is a typical template for presenting IACC spectra.

4.2 Fluctuation, glimpsing, and parallel streams of binaural cues

We have observed (e.g., in Figure 4) that IACC (t) fluctuates vigorously even in sustained notes of music. On the other hand, localization, source broadening and envelopment are considered and discussed as far more continuous listening aspects, as if they are co-existing like parallel perceptual streams. It is possible to localize a moderately broadened source and sense an enveloping environment all at the same time. While IACC (t) is a serial stream, our brain can learn to continuously analyse and categorize, and assess the density of, e.g., *low IACC*, *medium IACC* and *high IACC*. In this way the density of *low-IACC-events* can be quite invariant over time even if the IACC fluctuates. The same goes for the other categories. In other words, the brain has decoded the serial stream into three parallel streams. The visual analogy to the described process is glimpsing¹. If the reader holds a hand up midway between the eyes and this text, it will obviously be difficult, or even impossible, to read more than a few words here and there. If instead the hand is waved fast from side to side, the brain will integrate the serial stream of brief visual still-images into a continuous complete image. We will not perceive the instantaneous visual inputs, but the stream. Similarly, we do not perceive the instantaneous IACC-values, but the streams. The principle is illustrated and further in Figure 6, Figure 8 and Figure 8. The four perceptual categories and streams are principal suggestions, and any separation limits have not been established.



Figure 7 a, b, c and d, left to right: Frames, instantaneous visual input elements, (a) localizable source; (b) localizable source with halo; (c) very broadened source in enveloping environment; (d) enveloping environment. If the frames are displayed serially one-by-one like a video, repeatedly, in random order, at sufficiently high rate, the brain would be able to localize a broadened source in an enveloping environment, all at once.

¹ to see something or someone for a very short time or only partly
<https://dictionary.cambridge.org/dictionary/english/glimpsing>

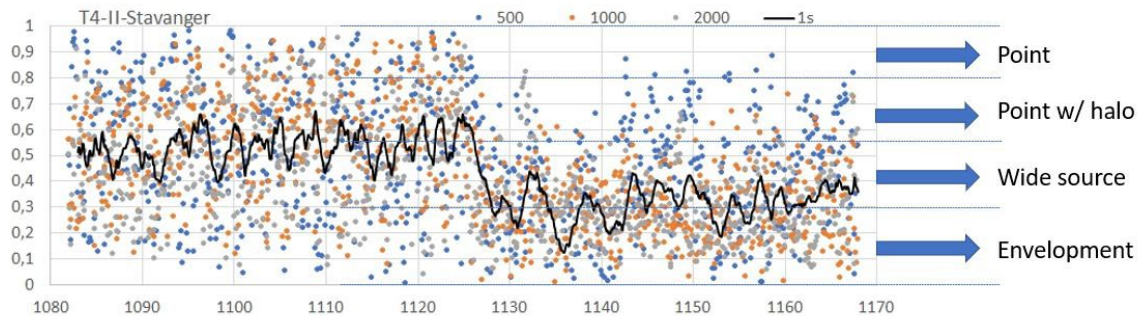


Figure 8 Decoding a serial IACC stream into parallel perceptual streams.

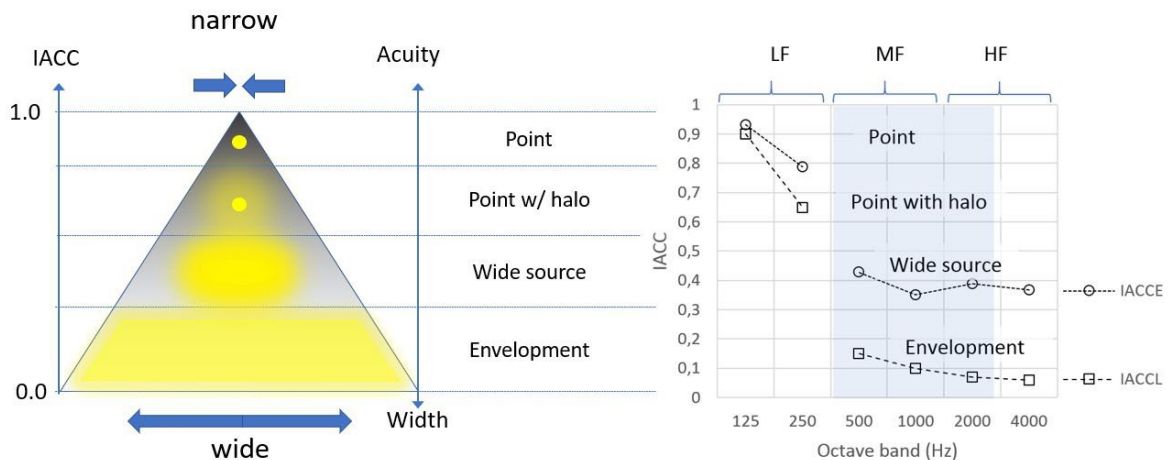


Figure 9 Basic concept of IACC value bins attributed to listening aspects.

4.3 Statistics from many halls

Figure 6 (left) presents statistics for 17 halls based on a total of 1 million samples, where each sample is IACC in broadband (400-2500Hz) from a 100ms window as described in this paper. The average IACC in each hall is with 95% confidence inside the narrow range indicated by the thickness of each bar. As can be read by comparing adjacent bars, some of the neighboring halls cannot be distinguished from one another by the broadband IACC in the data collected so far, e.g. Hamburg and Amsterdam, Paris-Stavanger-Boston, Bergen-Oslo, AFH-Helsinki, and Chicago-Stord. On the other hand, some halls seem to distinguish with big statistical confidence, like Berlin in one end of the scale, and San Fransisco, Milano, and Miami in the other end. However, when looking at the diagram to the right, the fluctuation range in terms of standard deviation around average IACC, all these halls seem to have a lot in common after all. In the diagram to the left, the attention is drawn to the upper and lower edges of the fluctuation range bars in each hall, instead of to the average in the diagram to the left. Upper limits range from 0.49 in Hamburg to 0.61 in Miami, indicating that there were less frequent cues of localization in Hamburg than in Miami. Lower limits range from 0.16 in Berlin to 0.25 in Miami, indicating that there were more frequent cues of envelopment in Berlin than in Miami.

IACC statistics spectrum of all collected data is presented in Figure 11. Note that the IACC mean spectrum is very close to the mean IACCE spectrum except for in 2000 and 4000 Hz.

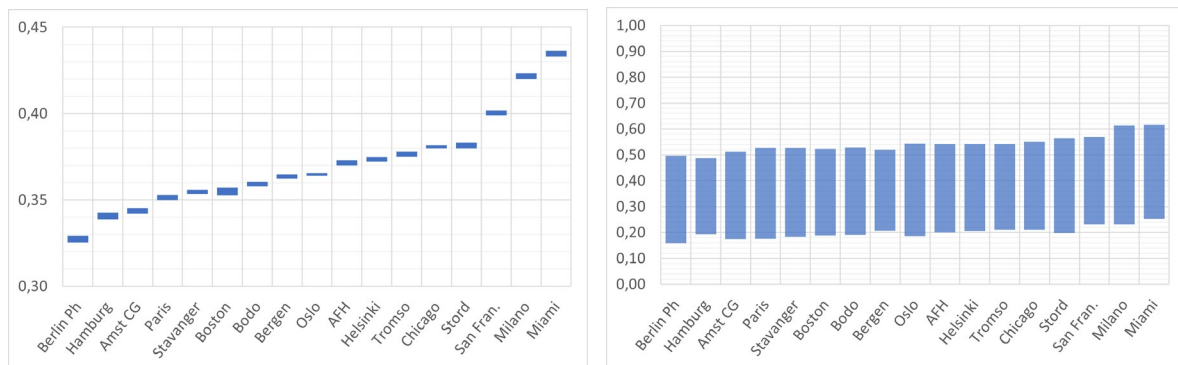


Figure 10 (left) 17 halls, average IACC with 95% confidence bars; (right) Normal range of IACC-fluctuations.

Do these results mean that significant differences in average IACC are not significant after all? That the variation due to music is so big that average differences become un-noticeable? This question is left for discussion, but a similar paradox should be kept in mind: Let's say the listening level in symphonic music spans from some 30dB to 95dB in a hall with $G=4$ dB. In halls with $G=3$ dB and $G=5$ dB, the same music would be heard with 29-94dB and 31-96dB. Does the great dynamic span of music itself make G less significant?

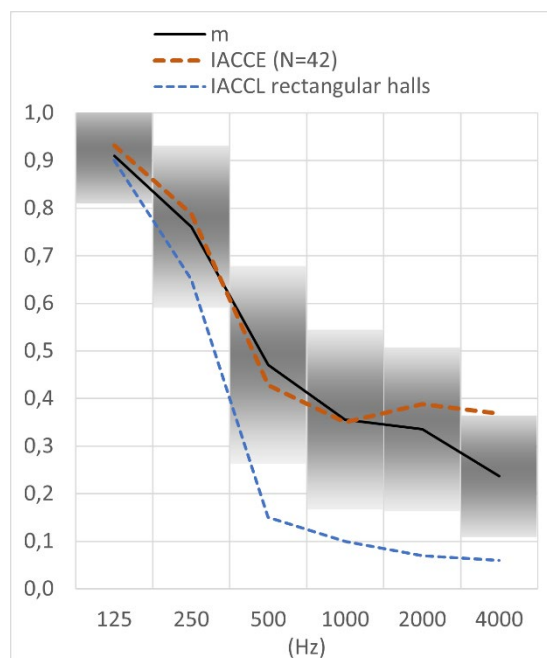


Figure 11 IACC statistics spectrum; m is mean value spectrum of all listening data acquired so far, with grey bars indicating standard deviation around m .

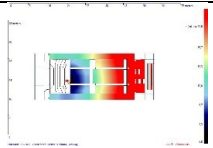
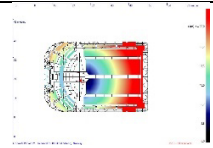
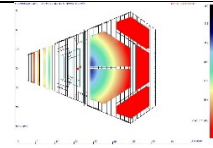
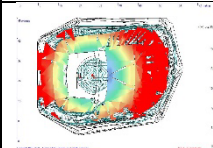
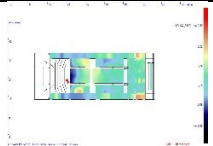
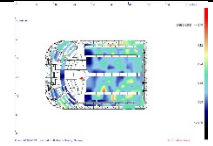
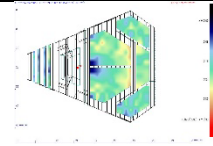
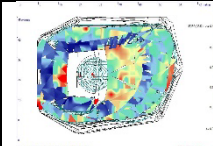
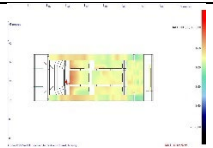
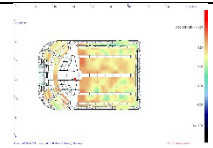
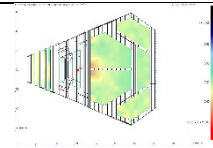
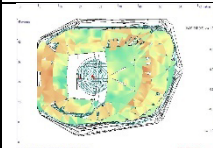
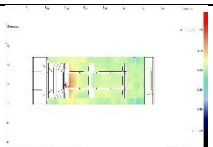
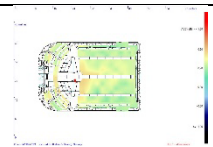
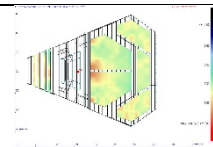
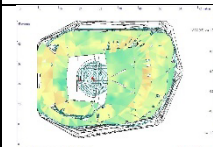
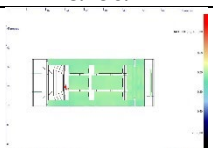
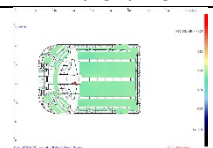
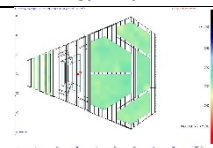
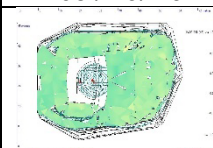
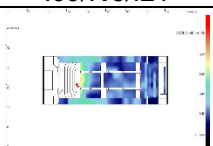
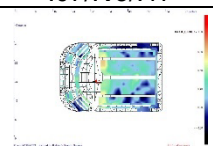
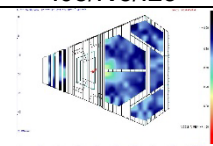
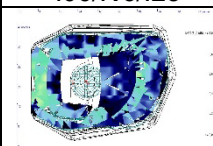
5 LESSONS LEARNED FROM THE BINAURAL PROJECT

- IACC is fluctuating vigorously even during sustained notes, normally ± 0.11 from one 100ms window to the next
- The binaural content in a period of music listening is often well described on the form $IACC = m \pm s$, where m is the mean and s is the standard deviation of the fluctuating IACC in the period
- Significant differences between halls are observed

- IACC varies more from one part of music to another, than from hall to hall
- Oboe and other woodwind instruments have relatively low spectral density and consequently high IACC, which allows the brain to localize the instrument
- String instruments have higher spectral density than woodwind, and the spectral density gets even denser when the whole string section combine in an ensemble, decorrelating the sound field to low levels of IACC, creating a broad sound image
- The brain can decode the serial stream of fluctuating IACC into parallel streams of Localization, Source Broadening and Envelopment, according to the model presented in this paper
- The source can be localized even if direct-to-reverberant ratio is low, because of glimpsing
- Onsets of notes can cause brief peaks in IACC, but this is rare, occurring in less than 2% of the music
- Likewise, decays after offsets of notes typically occur with low levels of IACC and consequently cues of envelopment, but they are seldom prominent except for in stop chords or end chords
- Music is composed for binaural listening in rooms that allows for Localization, Source Broadening and Envelopment
- The orientation of listening positions relative to the musicians is one of the important implications of binaural hearing
 - The relevant coordinate system for a successful performance space is not the one defined by the hall's length-width-height, but the one defined by front-side-up as seen and heard by the listener
 - The height-to-width ratio is still relevant in auditorium design, but the width is defined as the dimension along the axis through the ears of the listener
 - A listener seated to the side of the orchestra will perceive the longest hall dimension as the width
- Seating behind and to the sides of the orchestra is highly problematic when taking binaural hearing into account
 - this problem can hardly be described by monaural listening aspects, or detected by monaural measurements, metrics and parameters
 - to the side of stage, the first reflection will arrive from above and either mask the lateral reflection to the listener, or the lateral reflection will arrive as echo
 - behind the orchestra, music will apparently be an event that happens in front of the orchestra, partly due to directivity, partly because the reverberant centroid is located in the middle of the room, thereby making the listener an observer to a distant event instead of being involved in the music
 - behind the orchestra, trumpets, trombone, singers, and other highly directive sources of music will return as a localizable echo from the back of the hall, particularly disturbing if these voices are soloists
 - music is not composed and conducted for these seats, conductors and composers do not chose these positions when assessing the sound of music
- As to the geometry of the hall as such, some serve music better than others, but so far in the Binaural Project, there is no paradigm that has proven to be the obvious choice

6 DESIGN PATTERNS AND BINAURAL PROPERTIES

Table 2 Four halls of different designs and some of their binaurally relevant properties. Simulations, calculations, and plots over the audience area receiver grids are performed in Odeon 17. N is audience seating capacity, $r(m)$ stage-to-listener distance in meters; in the plot green represents $r=12m$; $N(r<25m)$ is the number of audience at less than 25 meters from stage; LF is early lateral fraction (LF80); IACCE, IACC and IACCL are estimations from directional simulation data processed by algorithms defined by the author, in Odeon. IACCE-IACCL is a measure of the dynamic potential of IACC. The statistical results from the audience grid is, except for $r(m)$, is presented with the following syntax: 5% percentile / Average / 95% percentile, e.g., in Vienna, 5-percentile of LF is 0.10, average LF is 0.18, and 95-percentile of LF is 0.27. Zero before decimal points are omitted.

Hall	Vienna	Amsterdam	Oslo	Paris
N	1680	2037	1600	2400
$r(m)$ plot				
$r(m)$ average	19.4	16.2	19.6	19.8
$N(r < 25 m)$	1140	1830	1120	1870
LF (plot)				
LF80	.10/.18/.27	.04/.14/.30	.09/.20/.30	.06/.19/.38
IACCE (plot)				
IACCE	.22/.40/.62	.20/.48/.76	.13/.34/.60	.07/.37/.47
IACC (plot)				
IACC	.18/.30/.47	.17/.31/.49	.13/.27/.44	.08/.29/.49
IACCL (plot)				
IACCL	.09/.15/.21	.07/.13/.17	.08/.15/.23	.06/.16/.25
IACCE-IACCL				
IACCE-IACCL	.10/.25/.49	.06/.36/.62	-.01/.19/.38	-.08/.21/.52

7 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. Beranek, L., *Concert Hall Acoustics* 2008, J. Audio Eng. Soc., Vol. 56, No. 7/8, 2008
3. 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
4. Beranek, L.L., *Concert Halls and Opera Houses* (Springer, New York, 2004).
5. https://en.wikipedia.org/wiki/Decca_tree
6. Skålevik, M., Spatial listening aspects and the time-varying inter-aural cross-correlation during music performance in concert halls, *Proc. ICSV24*, London, 23-27 July 2017.