MUSIC, ROOM, TWO EARS, DESIGN AND PARADIGMS

By Magne Skålevik msk@brekkestrand.no

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.

An introduction to binaural hearing concepts and metrics is presented together with a mathematical demonstration of why IACC can be considered a difference between median and lateral energy fractions 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.

Music is composed for binaural listening in rooms that allows for a wide range of perceptual aspects, among which Source Broadening and Envelopment are frequently mentioned in literature, while a reminder of the following aspects seems to be necessary: Localization –both azimuth and distance, Depth and Relief in the sound image, Auditive Perspective and Presence.

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 experience can be more sensitive to seat position and orientation relative to the room and musicians than to the room geometry itself. E.g., perceived width should be measured along the lateral axis running through the ears.

This paper is an updated and extended version of the paper presented at the IOA Auditorium Acoustics conference in Athen in September 2023, including the latest developments and information.

2 BINAURAL HEARING, PERCEPTION AND METRICS

Definitions of the terms used in this chapter are given in Table 1.

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 Presence, Localization, Source Broadening and Envelopment

Since the late 1960s, when Marshall[1] and other authors drew the attention of concert hall planners toward the significance of room width and the height-to-width ratio, and its implications to the spatial impression of the performance spaces, the emphasised binaural hearing aspects have been source broadening and envelopment, and the metrics used to measure their related acoustical features: Early Lateral Fraction (LEF) and Late Lateral G (LLG), or the pair of metrics based on inter-aural cross-correlation, i.e. IACCE and IACCL.
Strangely enough, the historically earliest publicly known advantages of binaural hearing, today often referred to as *localization* and *presence*, seems to have been forgotten by most authors in the field of performance spaces for classical music: First presented to the public at the 1881 Paris Electrical Exhibition, Théâtrophone ("the theatre phone"), invented by Clement Ader, was a telephonic distribution system available in portions of Europe that allowed the subscribers to listen to opera and theatre performances over the telephone lines. The following two paragraphs is a quotation from an article in Scientific American in 1881[2].

“One of the most popular attractions at the 1881 Paris Electrical Exhibition is the nightly demonstration of the marvelous powers of the Ader telephone, by its transmission of the singing on the stage and the music in the orchestra of the Grand Opera at Paris, to a suite of four rooms reserved for the purpose in one of the galleries of the Palais de l'Industrie. This demonstration is given nightly between eight and eleven o'clock, and the enormous number of people who crowd the entrance to the building before the doors are open to the evening visitors rapidly resolve themselves into patient queues as soon as they can obtain access to the gallery adjoining the telephone rooms….”

> "….We will now consider the new acoustic effect which Mr. Ader has discovered, and applied for the first time in the telephonic transmission at the Electrical Exhibition. Every one who has been fortunate enough to hear the telephones at the Palais de l'Industrie has remarked that, in listening with both ears at the two telephones, the sound takes a special character of relief and localization which a single receiver cannot produce. It is a common experience that, in listening at a telephone, it is practically impossible to have even a vague idea of the distance at which the person at the other end of the line appears to be. To some listeners this distance seems to be only a few yards. To others the voice apparently proceeds out of a great depth of the earth. In this case there is nothing of the kind. As soon as the experiment commences the singers place themselves, in the mind of the listener, at a fixed distance, some to the right and others to the left. It is easy to follow their movements, and to indicate exactly, each time that they change their position, the imaginary distance at which they appear to be. This phenomenon is very curious, it approximates to the theory of binauriclar auduition, and has never been applied, we believe, before to produce this remarkable illusion to which may almost be given the name of auditive perspective….”

Figure 1 Left: Diagram of the Teatrophone and the transmission lines between Grand Opera and the exposition location. Right: an 1896 lithograph by Jules Chéret.

There is an inherent conflict, or at least opposite concerns, between the two groups of binaural perception aspects:

a) Localization – both azimuth and distance, Relief[1] and Depth in the sound image, Auditive Perspective, Presence
b) Source Broadening and Envelopment

a) would require some portion of high values of IACC, while b) would occur when IACC is low. One can only speculate over why the volume of papers in late 20^{th} century on concert hall acoustics have been dominated by b), but the historical background might offer some explanation: Since Boston Symphony Hall was opened in 1901, there was no new concert halls that had acoustics comparable to the reference halls in Vienna and Concertgebouw, and by 2023 maybe there still aren't any. After WW1, acousticians were concerned about directing ceiling reflections. After WW2, for various reasons, halls were wide, and ceilings were low. The Haas effect was a proof that reflections from above in reasonable amount did not create false localizations. Besides, ceiling reflections could mask unwanted late reflections from side walls in wide rooms. It took time for acousticians in general to fully realize that lateral energy and the height-to-width ratio in the old reference halls were important success factors, and not accidents. But once they did realize, no one seem to be concerned that too much early lateral energy could ever be an issue, unless false localization in the side walls occurred. One notable exception is Griesinger (2010)[3] who in a trilogy of papers holds Localization to be a key to Engagement, and in Part 3 commented on Boston Symphony Hall that "Localization and engagement are restored in the front of the first balcony because the primary reflection from the side wall is blocked by the side audience." However, a statement in the Part 1 of the trilogy suggests that the potential problem with the side wall reflection was not excess lateral energy, but its arrival time interval: "excess reflections in the time range of 10ms to 100ms reduce engagement, whether they are lateral or not."

[1] Here, relief is in the geographical meaning, a sense of depth in the sound image.

By the turn of the millennium, the most important listener aspects were Reverberance, Strength, Clarity, Apparent Source Width (or Source Broadening) and Level of Envelopment, as expressed in the Annex of the standard ISO-3283. None of the aspects in a) were mentioned.

This author proposes a revival of the aspects in a), Localization – both azimuth and distance, Depth and Relief in the sound image, Auditive Perspective, Presence. A balance between the somewhat conflicting concerns, a) and b), should be pursued, instead of an either-or approach, e.g., with questions like: How much relative lateral energy is necessary and sufficient, and what at what limits would lateral energy tend to compromise localization, presence, depth or other aspects in a)?

2.3 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.
Perceived width - another implication of the variant coordinate system is discussed below in this paper, in the comments to Figure 20.

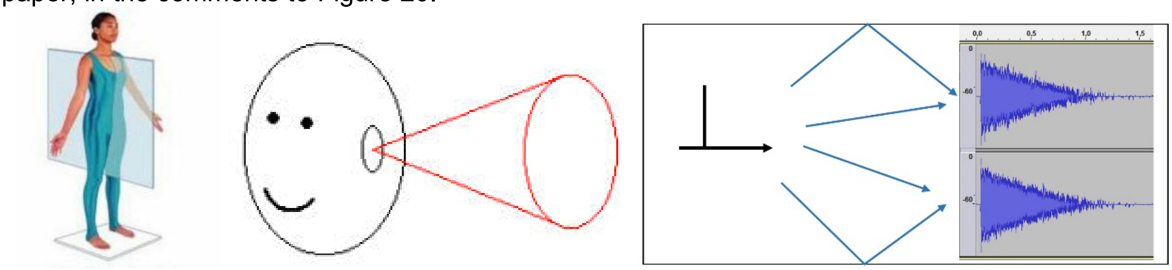


Figure 2, from left, a) median plane, b) cone of confusion, c) Binaural Impulse Response (BRIR)

2.4 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.

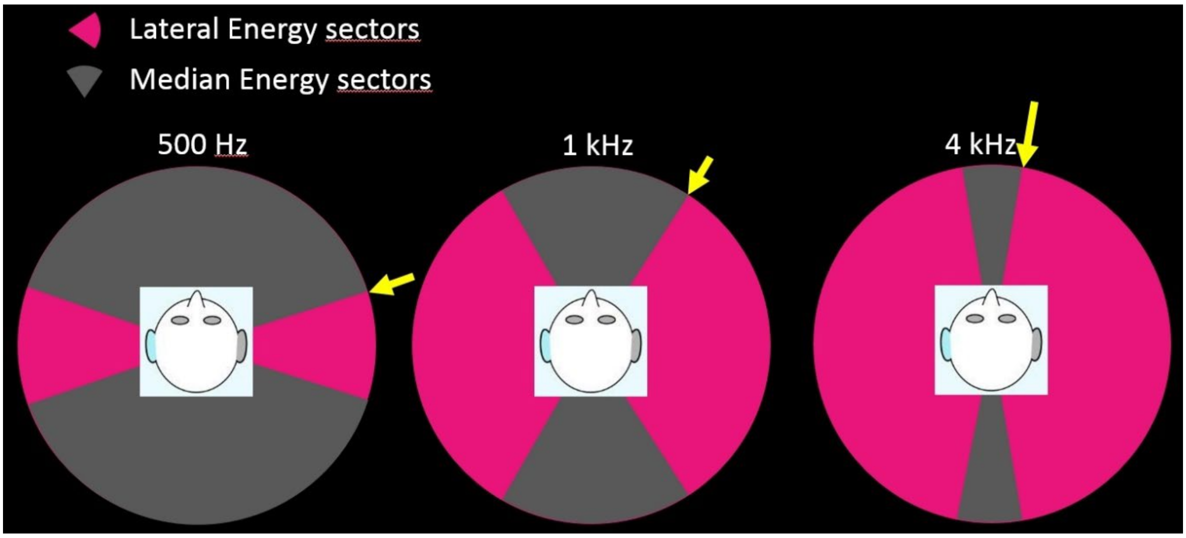


Figure 3 Borders between sectors of median and lateral energy are frequency dependent.

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.

2.5 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].
Since sound pressure signal does not have a pressure bias, like the one from atmospheric pressure, the mean sound pressure is equal to zero, and the correlation formula can be applied to the sound pressure signals (functions of *t*) incident at the left ear and right ear, p_L and p_R, respectively, as follows:

$$IACC = \frac{\sum p_L \cdot p_R}{\sqrt{\sum p_L^2 \cdot \sum p_R^2}} \qquad (1)$$

Assume that the signals are sampled and the summations are carried out over a range of *n* samples. Multiply with *1/*n in both nominator and denominator, and note that the square root of a product equals the product of the square roots,

$$IACC = \frac{\frac{1}{n}\sum p_L \cdot p_R}{\sqrt{\frac{1}{n}\sum p_L^2} \cdot \sqrt{\frac{1}{n}\sum p_R^2}} \qquad (2)$$

Now note that the two roots in the denominator are the root-mean-squares of the left ear signal $RMS(p_L)$ and the right ears signal $RMS(p_R)$, respectively, over the n samples. For each binaural signal sample *1* to n, let $L=p_L/RMS(p_L)$ and $R=p_R/RMS(p_R)$ be the normalized, unitless, binaural sound pressure signal pair, and write

$$IACC = \frac{1}{n}\sum_1^n L \cdot R \qquad (3)$$

In binaural signals with any differences between L and R, the terms in the summation could be separated into two summations, the sum of the products with factors of equal signs and the sum of product with factors of opposite signs:

$$IACC = \frac{1}{n}\sum_1^n (L \cdot R)_{equalsign} + \frac{1}{n}\sum_1^n (L \cdot R)_{oppositesign} \qquad (4)$$

Note that both summations are average of unitless energies from the n sample pairs, i.e. unitless powers, of sample pairs with equal polarity and sample pairs of unequal polarity, respectively. This separation is highly significant because any sample in the first summation contributes to increase *IACC*, while any sample in the second summation contributes to reduce *IACC*. A common interpretation of cross-correlation is that it is a measure of similarity between two signals, and in the case of IACC namely similarity between the left and right signal. Therefore, it would be natural to let the summations represent the energy fractions of similarity and dissimilarity in the binaural signal, denoted E_{sim} and E_{dissim}, respectively, and since all terms in the second summation are negative, write:

$$IACC = E_{sim} - E_{dissim} \qquad (5)$$

Here, the sum of energy fractions is $E_{sim} + E_{dissim} = 1$. All sounds arriving in the median plane, will contribute to the median energy fraction, E_{sim}, and all sounds arriving laterally, will contribute to the lateral energy fraction, E_{dissim}. Refer to Figure 3, and note that there is no sharp geometrical border corresponding to the split between two groups of samples in (4). It is not meaningful to map the individual samples to the two sectors in the diagram.

Now, let $m = L+R$ be the sum signal of the two normalized (unbiased) signals L and R in (3), and $s = L-R$ be the difference signal between them. These signals are the normalized binaural counterparts to the mid- and side-signals used in the M-S or Blumlein stereo microphone coupling of a figure-8 and an omni-directional microphone centered in point with the figure-8 axis pointing to the side. It can readily be shown that the difference between their energies is $m^2 - s^2 = 4 \cdot L \cdot R$. Insertion in (3) yields,

$$IACC = \frac{1}{4n}\sum_1^n m^2 - s^2 \qquad (6)$$

If we borrow the Blumlein terms and replace them with “median” and “lateral”, a transition similar to the one from (4) to (5), we let $E_{median} = m^2/4$ and $E_{lateral} = s^2/4$ be the average median and lateral energy fractions, respectively:

$$IACC = E_{median} - E_{lateral} \qquad (7)$$

Again, the sum of energy fractions is $E_{median} + E_{lateral} = 1$. *IACC* increases as E_{median} increases and *IACC* decreases as $E_{lateral}$ decreases.

Comparing (5) and (7), it is tempting to conclude that E_{median} equals E_{sim} and that $E_{lateral}$ equals E_{dissim}. However, this is an assumption that needs to be tested mathematically, and it is outside the scope of this paper.

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, or displacements, in one signal relative to the other signal in the binaural pair, between lag limits $+\tau$ and $-\tau$, which allows arbitrary azimuth angles to be analyzed, each angle corresponding to unique values of τ. 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. Examples of IACCF as a function of sample lags are given in Figure 4 and Figure 5.
IC	Interaural Coherence is the special case of IACCF where τ = 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*)

An ideal frontal sound, or other sound incident in the median plane, would in anechoic conditions produce $E_{median}=1$, and $E_{lateral}=0$, thus *IACC=1.* The other extreme, a non-median sinusoidal incident that produces the exact opposite phase at the two ears, would yield *IACC= -1*. The in-between case represents a practically interesting reference, i.e., if $E_{median} = E_{lateral} = 0.5$, then *IACC=0*.

Note that the Lateral Energy Fraction (LEF or just LF) metric is sometimes listed among binaural sound metrics and frequently used to describe listener aspects in concert halls. IACC and LF are indeed statistically, and in some sound-fields mathematically, related to one another. But 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.

2.6 IACCF, IC, IACC and the wind vane

In each time window T, the inter-aural-cross-correlation function IACCF returns the IACC-value not only for the raw binaural signal pair (no lag), but for any arbitrary lag between the left and right signal. This is useful for analyzing from which azimuth angle incident sound is the strongest and how strong it is compared to sound incidence from other directions.

Figure 3 is an example of e binaural signal from the note ***a'*** (*440Hz* fundamental) arriving at 30 degrees (left) azimuth angle relative to median plane, in anechoic conditions. The signal at the right ear is approximately equal to the one at the left ear, but it lags by -12 samples (at 44.1kHz sample rate) due to the longer path. The harmonics are (in dB) -6, 0, -6, -12, -18, …

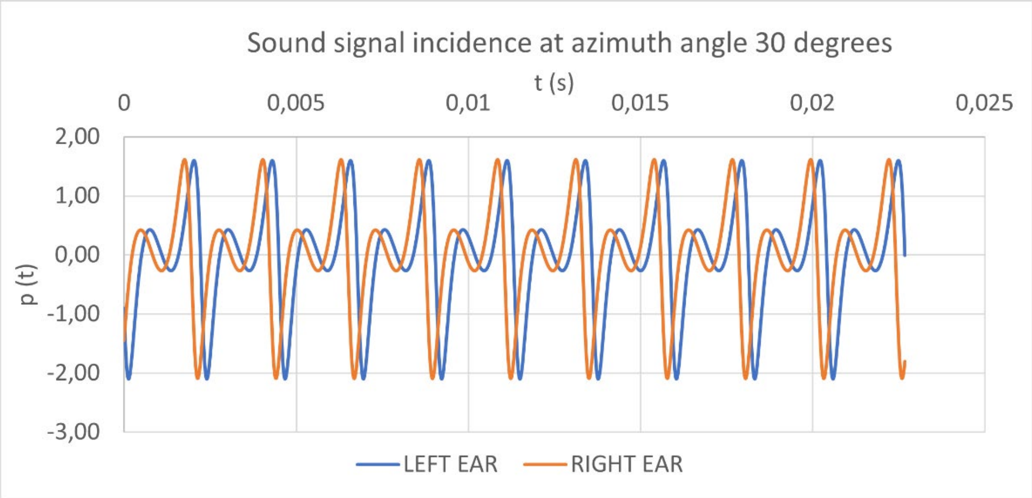


Figure 4 Binaural signal from the note *a' (440Hz)* arriving at 30 degrees azimuth angle relative to median plane.

In Figure 4 we have a plot of the IACCF(lags) in as a function of the number lags between -30 and +30 sample from the binaural signal in Figure 3. As can be seen from the plot, IACCF reaches its maximum at -12 lags. In this case the maximum is $IACC_{max}$ =1.0. If the line spectrum of the tone had been flat, the maximum would be 0.99, and if it also had arrived at 90 degrees, the diffraction around the head (the HRTF) would have distorted the signal at the right ear sufficiently to make it 0.95. The relationship between azimuth angle α and lags L at the far ear for a common binaural model with a spherical head with radius 0.09m is approximately (r^2 =0.9992).

$$\alpha = 0.0257*L^2+2.1563*L$$

Consequently, the IACCF(azimuth) can be expressed graphically as in Figure 5.

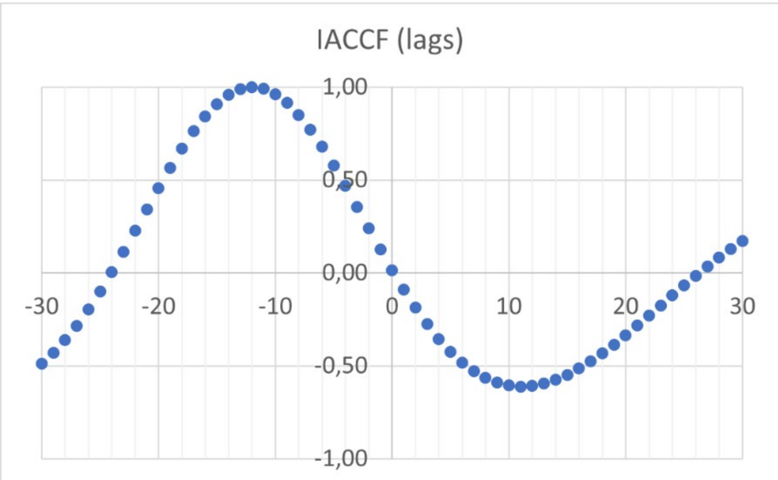


Figure 5 The inter-aural cross-correlation function IACCF (sample-lags) from the binaural signal in Figure 3, see text.

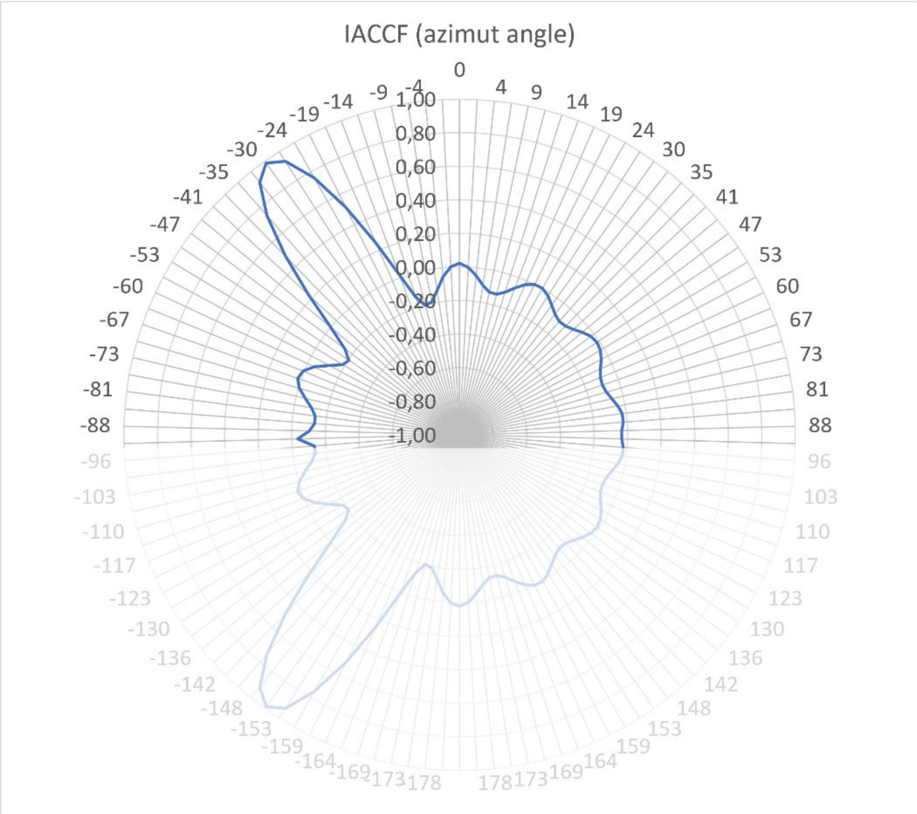


Figure 6 Radar plot of IACCF for the *a'* note arriving at -30 degrees azimuth, see text. 0 and 180 degrees correspond to the median plane. The lower half of the diagram is a mirror image to the upper, and demonstrates the so-called *cone of confusion* in Figure 1b, meaning that the information from the lag within the binaural signal pair is ambiguous, and that the brain in blindfolded perception easily confuses, e.g., front left with rear left. The two lobes in the diagram coincide with a cone centered at -90 and with an opening angle of ±60 degrees.

If the binaural signal in Figure 3 had been received in a concert hall seat, from the tuning note from the oboe, the reverberant sound arriving from all directions would be out of phase with the direct component, causing a decorrelation in the binaural signal pair, and consequently lower $IACC_{max}$ than in the anechoic examples above. Examples of IACCF from such tuning sessions are presented graphically in Figure 6. Apart from reverberant sound causing decorrelation, the usual babble noise from the audience will reduce the maximum value. Here, the relationship between azimuth angles and lags have been used to convert the horizontal axis from lags to azimuth angles. We can see that the angles at which the maximum of IACCF occurs, vary.

Computation-wise, the examples above demonstrates how the algorithm behind IACC(t) in the Binaural Project processes and compresses the raw data in the binaural signal for each 100ms window:
First, the 100ms binaural signal is sampled at 44.1kHz, producing 2x44100x0.1 = 8820 numbers. Then the correlation function algorithm boils it down to 47 IACC numbers, one from each number of lags: -23, -22, …0, …+22, +23. Finally, the maximum value $IACC_{max}$ from the 47 IACC-value is returned and added to the sequence as a sample at time *t*. The sequence is indeed the IACC(t) used to represent the binaural signals in the Binaural Project. The compression is 8820 to 1. Further, the IACC(t) is analyzed in various ways, e.g., statistically by its mean and standard deviation. In the latter case, from a 45-minute symphony, the total compression would be 238 million to 2, or 119 million to 1 if you like.

Conceptually, the IACCF radar plot resembles the *wind rose* diagram (Figure 6 a), only they express the long-term distribution of prominent wind directions, instead of the short-term distribution of prominent incident sound direction. Incident wind direction have been indicated by the traditional *wind vanes*, (b), quite analogous to the direction of maximum IACCF in Figure 5. Finally, the vane-style anemometer (wind-speed meter) would measure the wind speed in the direction from which the wind blows the strongest at any time. The latter would be analogous to the IACC(t), which indeed measures the (relative) strength, or prominence to be precise, of sound in the direction from which it is strongest, or most prominent.

Now, tying together the analogy of a binaural "vane-style anemometer", the radar plot in Figure 6, IACC(t), and the median-lateral concept, the nose in Figure 3 always points in the direction of maximum IACCF(azimuth). But in practice, it is not necessary to turn the head like the wind vane every 100ms. For the concert listener, the maximum IACC is detected by our binaural hearing, and for the investigating acoustician, it is taken care of by measurement technology and mathematics.

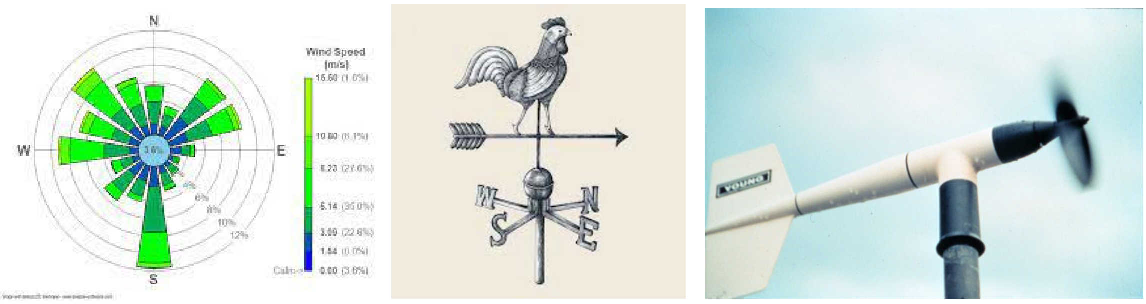


Figure 7 a) the wind rose, b) the wind vane, and c) the vane-style anemometer

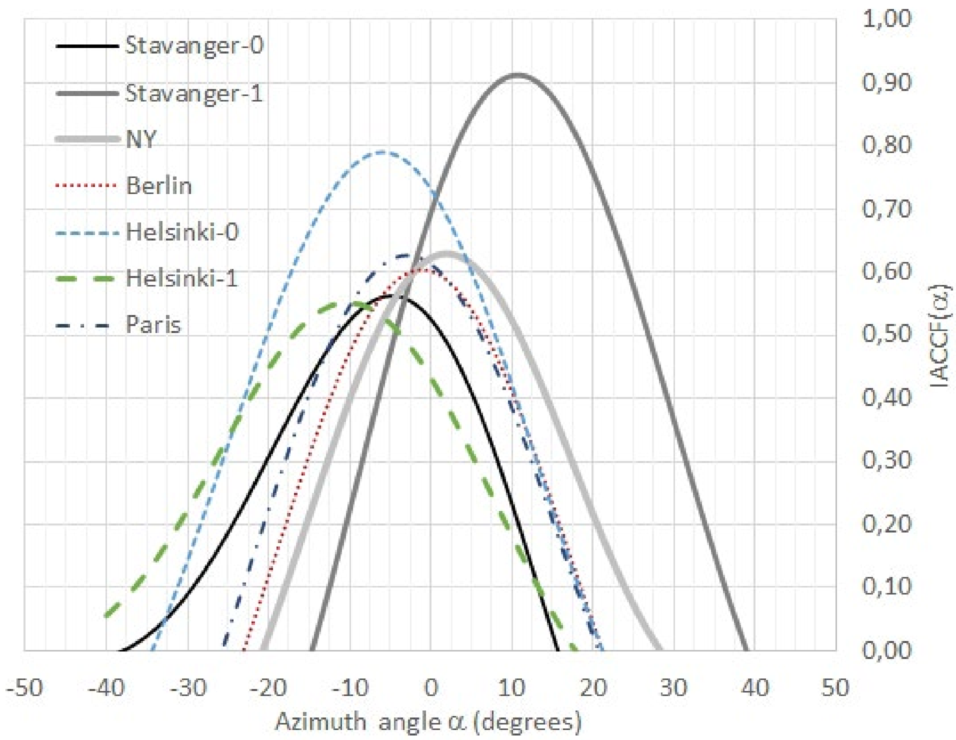


Figure 8 IACCF in binaural signals received in various concert halls, the a' note (440Hz) of an oboe during the tuning of the orchestra. Here, the function graphs are truncated at 0.00.

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 form concerts with symphony orchestras in big halls.

As a part of the Binaural Project, Skalevik[7] 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.

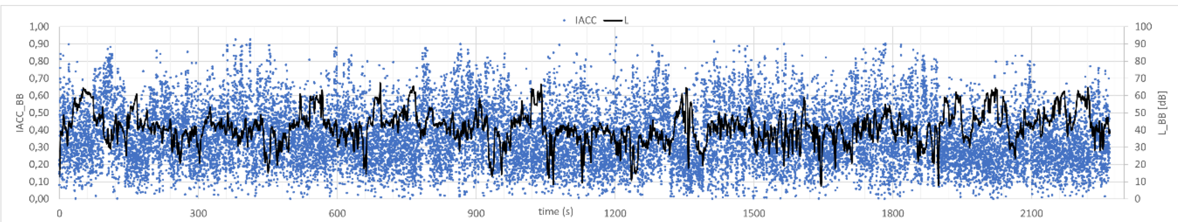


Figure 9 Boston Symphony Orchestra plays Brahms Violin Concerto in D major in Symphony Hall. The cloud of (blue) dots are IACC(t), while the black curve is smoothed (2s) level in dB, in second balcony. The horizontal axis is time in seconds. 100ms IACC-windows. See text for interpretation.

Denser energy spectra decorrelates the sound fields more and reduces IACC. Stronger parts have denser spectra because av 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.

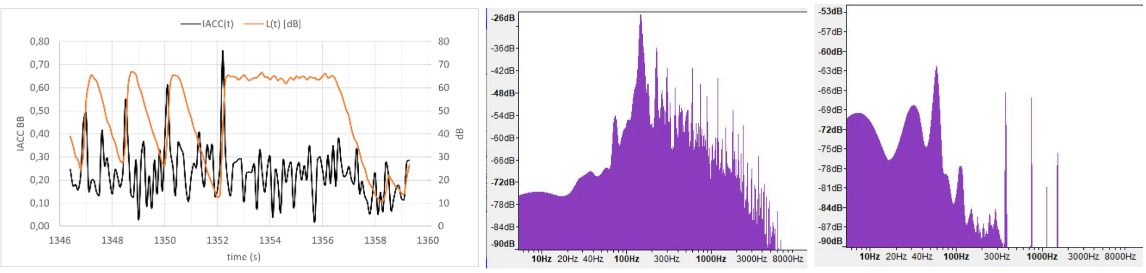


Figure 10 (left) extract 1346-1358 from Figure 9 with stop chords; (middle) spectrum of tutti orchestra in sustained part 1352-1356 s; (right) flute spectrum at 111-113 s in Figure 9;

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±0.11, compared to overall for the whole concert, IACC=0.35±0.17. 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±0.18 from an oboe solo and IACC=0.44±0.04 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.

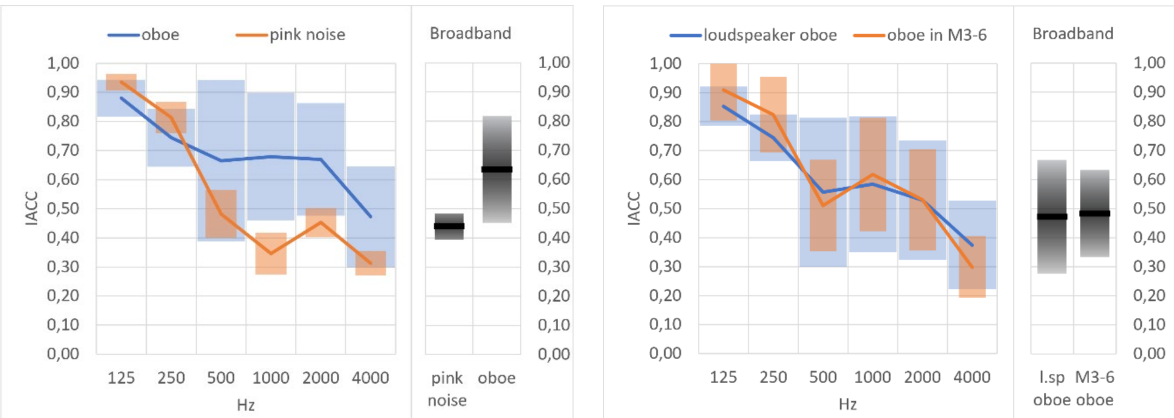


Figure 11 (left), experiments with signals from a loudspeaker, an oboe solo and a pink noise solo, with exact same source receiver positions, at distances 15-19min a 450 seat 6500m^3 hall; (right) same loudspeaker as in the experiment to the left, and a real oboe playing a 9s solo part during a live performance of Mahler's 3rd Symphony, 6th movement, both measured with exact same source and receiver positions, distance 26m, Oslo Concert Hall (19.000m^3), May 2022.

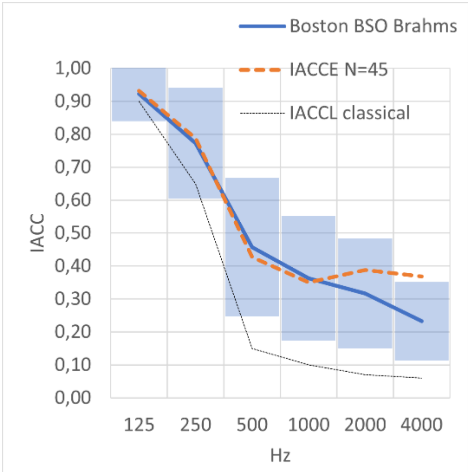


Figure 12, solid curve is IACC average spectrum in octave bands from the whole Brahms Violin Concerto in Figure 3; shaded bars indicate standard deviation, where upper edges measure strength and density of localization cues, while lower edges measure strength and density of envelopment cues. IACCE and IACCL are spectra of early and late parts of binaural impulse responses in Beranek's data collection[7]. This is a typical template for presenting IACC spectra.

4.2 Fluctuation, glimpsing, and parallel streams of binaural ques

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 glimpsing[2]. 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.

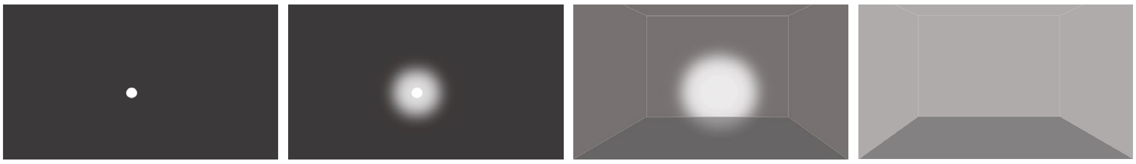


Figure 13 a, b, c and d, left to right: Frames, instantaneous visual input elements, (a) localizable source; (b) localizable source with halo; (c) very broadened source in enveloping environment; (d) enveloping environment. If the frames are displayed serially one-by-one like a video, repeatedly, in random order, at sufficiently high rate, the brain would be able to localize a broadened source in an enveloping environment, all at once.

[2] to see something or someone for a very short time or only partly
https://dictionary.cambridge.org/dictionary/english/glimpsing

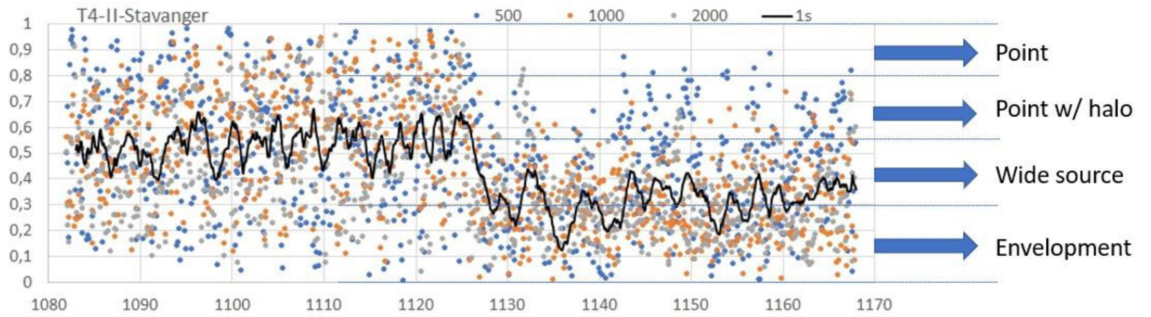


Figure 14 Decoding a serial IACC stream into parallel perceptual streams.

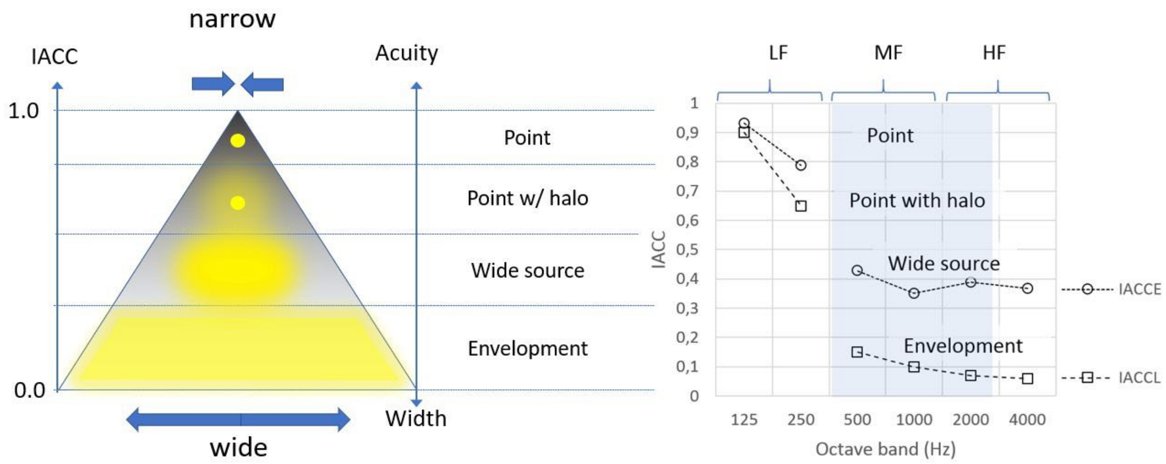


Figure 15 Basic concept of IACC value bins attributed to listening aspects.

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 ques 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.

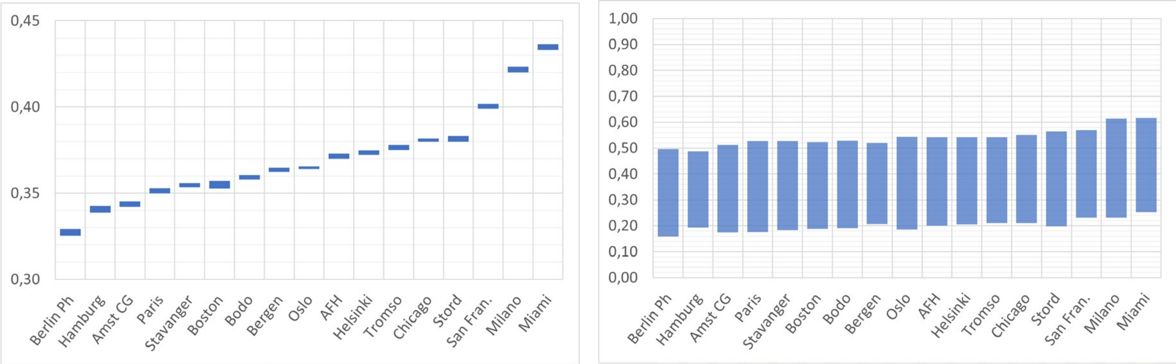


Figure 16 (left) 17 halls, average IACC with 95% confidence bars; (right) Normal range of IACC-fluctuations.

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=4dB. In halls with G=3dB and G=5dB, the same music would be heard with 29-94dB and 31-96dB. Does the great dynamic span of music itself make G less significant?

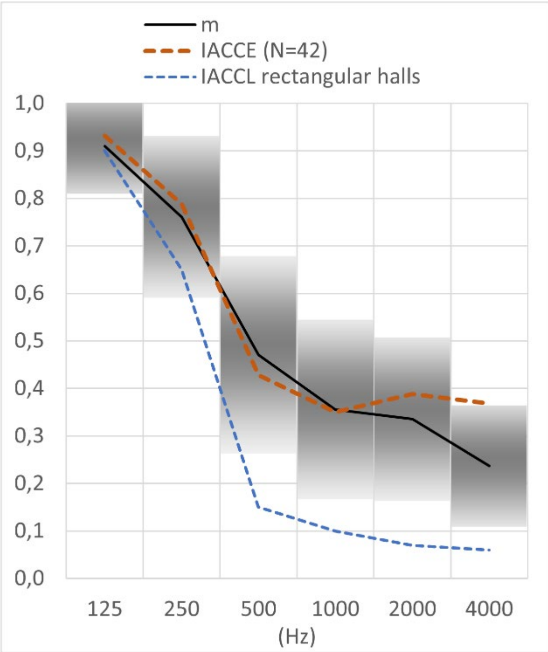


Figure 17 IACC statistics spectrum; *m* is mean value spectrum of all listening data acquired so far, with grey bars indicating standard deviation around *m*.

4.4 IACC and complexity of musical sound

As stated above, higher spectral density, more high frequency energy content and louder music tend to come with lower values of IACC and the impression of more broad sound image. An explanation for this, based on external factors only, is that a more complex and high frequency dense energy spectrum will inevitably cause more decorrelation in the binaural signal. Higher frequencies have smaller wavelengths, causing bigger differences at the two ears with reverberant sound arriving from the sides of the hall. Of course, if the hall also has a higher HF response from side walls and other laterally reflecting surfaces, it will make the decorrelation effect even stronger.

It has been commented in the literature that louder music above some threshold causes more source broadening due to a perceptual mechanism. This cannot be rejected from the data, but data suggests that loud music most often come with high density energy spectra, and that this is sufficient to explain the increased source broadening reported.

A common exception to the coincidence between complexity and loudness of sound, worth noting in this context, is the French Horn. It combines high sound power with a relatively sparse energy spectrum, thus appearing with high IACC even at fortissimo.

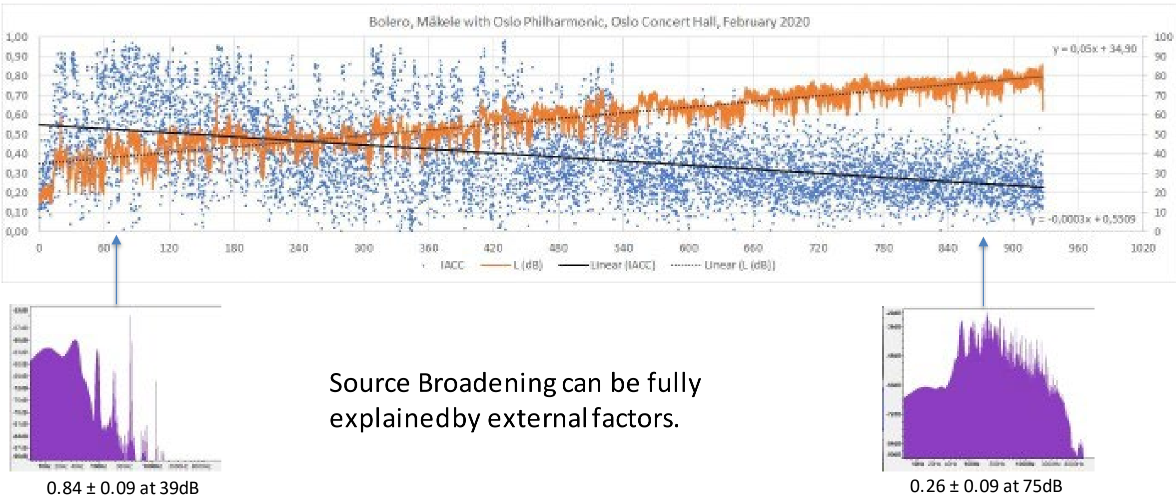


Figure 18 A demonstration of how IACC decreases as loudness increases. Horizontal axis is time in seconds; Blue dots and primary vertical axis is IACC; Red-orange curve with secondary vertical axis is SPL in dB; Energy spectrum in soft part with high IACC to the lower left and energy spectrum in the loud part with low IACC to the lower right. Dark solid curve is the linear IACC trend decreasing with time. Before 540 seconds (left half of the diagram) there are many scattered dots above IACC=0.60, indicating a high density of localization cues, while practically none after 540 s where dots concentrate around the trendline instead, indicating no localization but a lot of source broadening cues.

4.5

4.6 Big data statistics

Figure 17 Presents Statistics from big data; Mean ± standard deviations; 2D surface plot of distribution, broadband IACC (vertical axis) over broadband SPL (dB) on the horizontal axis. From the total volume of broadband IACC data, the grand average is 0.37. In terms of standard deviations, the variations are roughly as follows:

- IACC-fluctuations over single seconds of music listening is within ± 0.11 in 67% of the average second, and more than ± 0.11 in 33% of the average second

- Part-to-part variation: The population of music parts (distinguished by instrumentation) has a standard deviation of ± 0.07
- The population of concert halls ± 0.04

The above reveals that that the hall-to-hall variation is small compared to part-to-part variation, which could suggest that hall average is not important. However, small differences in acoustic parameter hall-to-hall averages compared to the variation in content does not prove that they are insignificant. E.g., a 1-2dB difference in G is not seldom used as argument for one hall being more suitable than another, even though music parts can vary from 30 dB to 90 dB, like in the Bolero example in Figure 18.

Broadband IACC average and standard deviations

Grand average	**0.37**
Hall-to-hall	**± 0.04**
Part-to-part*	**± 0.07**
Fluctuations (1 sec)	**± 0.11**

*Parts distinguished by instrumentation

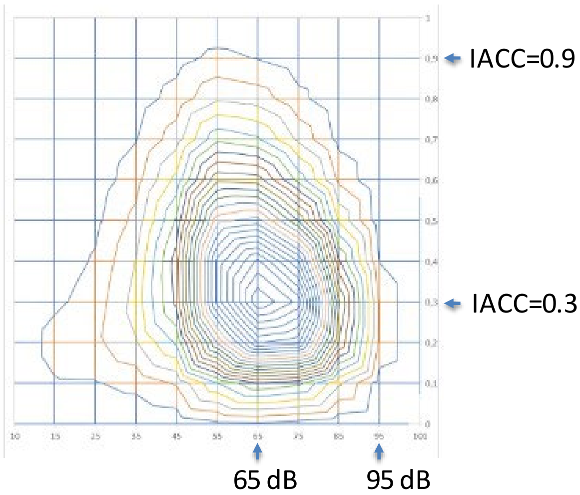


Figure 19 Statistics from big data; Mean ± standard deviations; 2D surface plot of distribution, broadband IACC (vertical axis) over broadband SPL (dB) on the horizontal axis.

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 ± 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 perceptual 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 a wide range of perceptual aspects, among which Source Broadening and Envelopment are frequently mentioned in literature, while a reminder of the following seems to be necessary:
    - Localization – both azimuth and distance, Depth and Relief in the sound image, Auditive Perspective, Presence.
- 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 perceived 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 that 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

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, Figure 20.
Figure 21 is a comparison of simulated binaural metrics in 4 rectangular halls and 3 non-rectangular halls. Red colors are seats with too high values, i.e.>90-percentile in the reference halls Vienna, Amsterdam and Boston, and black colors are seats with too low values, i.e. <10-percentile in the reference halls. Note that there are plenty of seats with too high or too low values in the reference halls themselves. Oslo and Bergen are both fan-shaped, but still very different. There are more black spots in the non-rectangular and more red spots in the rectangular halls. Rectangular shape is no guarantee for good results, but if something, their color plots seem more consistent than those in the non-rectangular halls.

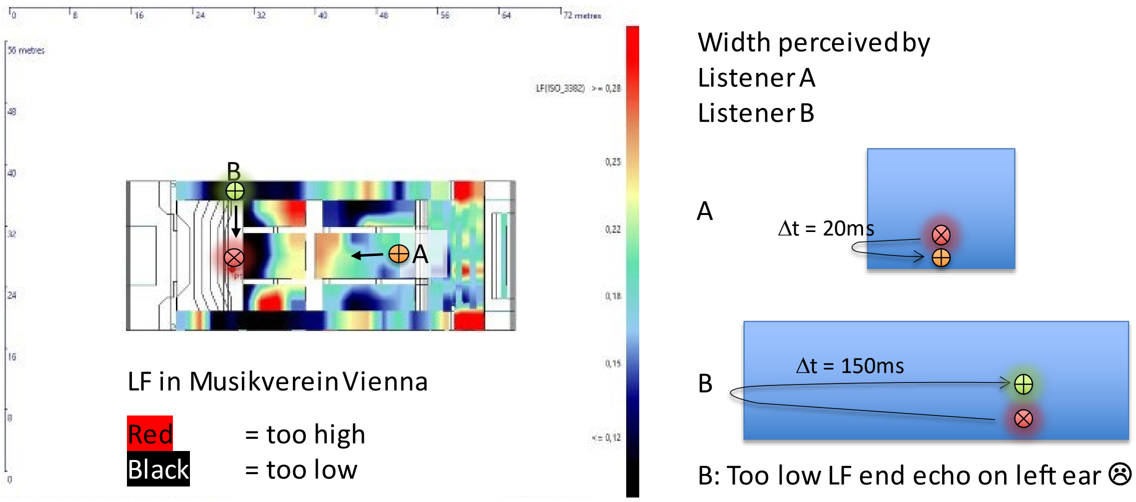


Figure 20 A demonstration of why perceived width is not identical to the architectural width of the performance space.

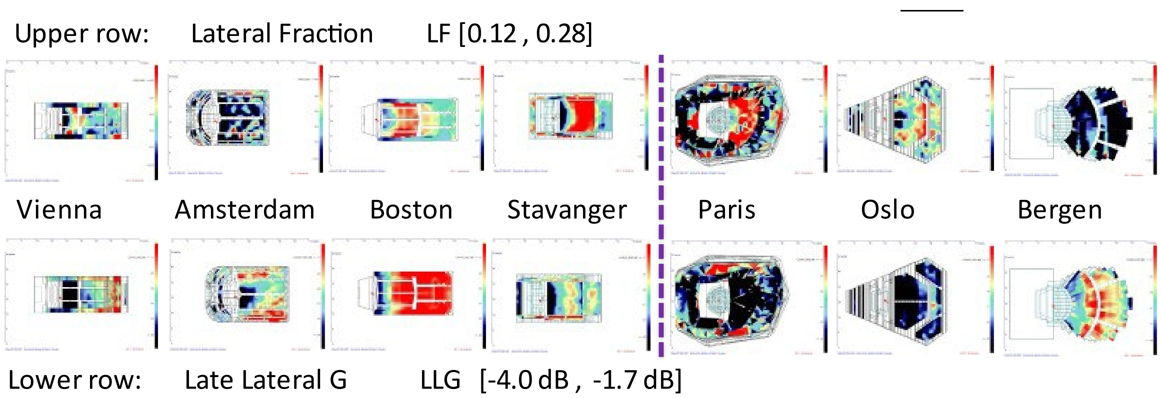
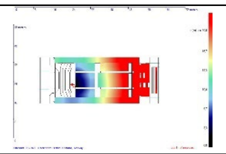
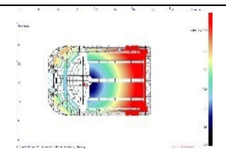
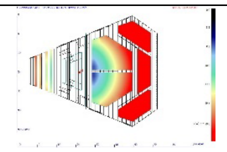
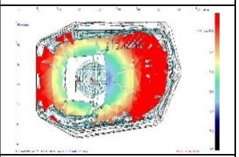


Figure 21 Comparison of simulated binaural metrics in rectangular (leftmost 4 halls) and non-rectangular halls.

Proximity to the musicians is frequently mentioned as a success factor, and it has been claimed that surround halls, like the terraced halls (vineyard halls), bring the audience closer to the stage. This claim is briefly tested with two rectangular, one fan-shaped, and one non-paradigmatic hall in Table 2, without an obvious conclusion as to which geometry or design paradigm is the more effective when it comes to proximity. Among these four halls, the more effective one is a rectangular hall, Amsterdam Concertgebouw, with 90% of the seats closer than 25 meters. The other rectangular hall, Musikverein Vienna, is the less effective one. Paris, obviously aiming for proximity with its surround approach, has 40 more seats than Amsterdam within 25 meters, but also is the one with the highest number of seats (530) outside 25 meters.

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 members, and percentage of total seat count, at less than 25 meters from stage.

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 (68%)	1830 (90%)	1120 (70%)	1870 (78%)

There is no obvious conclusion as to which design paradigm is the better on when it comes to binaural perception criteria. An attempt at conclusion would be:

Rectangular shape is no guarantee for good binaural quality, and Non-rectangular shape is not always bad.

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. Scientific American, December 31, 1881, pages 422-423:
3. Griesinger, D, “Phase Coherence as a Measure of Acoustic Quality” Part 1, 2 and 3, Proc. ICA Sydney 2010 https://www.akutek.info/Papers/DG_Phase_Coherence.pdf
4. Beranek, L., Concert Hall Acoustics 2008, J. Audio Eng. Soc., Vol. 56, No. 7/8, 2008
5. 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
6. Beranek, L.L., Concert Halls and Opera Houses (Springer, New York, 2004).
7. https://en.wikipedia.org/wiki/Decca_tree
8. 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.