akuTEK

www.akutek.info

PRESENTS

CAN SOURCE BROADENING AND LISTENER ENVELOPMENT BE MEASURED DIRECTLY FROM A MUSIC PERFORMANCE IN A CONCERT HALL?

by Magne Skålevik

A paper presented at IOA Auditorium Acoustics, Paris 2015-10-31

INTRODUCTION

Source Broadening, or Apparent Source Width (ASW), and Listener Envelopment (LEV) have for several decades been held to be some of most critical perceptual aspects of a concert hall. There seems to be consensus among acousticians that these aspects relate to quantities that can be analyzed from impulse response measurements, namely LF and 1-IACCE for ASW, and LLG and 1-IACCL for ENV. The first two are calculated from the first 80ms of the impulse response, while the latter two are calculated from the part after the first 80ms of the impulse response. While such data are quite straightforward to acquire in an empty auditorium, they are less straightforward to acquire with orchestra and audience in place. There is uncertainty related to the difference between unoccupied and occupied conditions. Moreover, an orchestra appears, in many seats in the first 20 rows anyway, to produce quite a broad sound image in itself. In such seats it is presently unclear how much difference the room makes to the apparent width, even if recommended values of LF and 1-IACCE are found in the same seat.

Until otherwise proven, one would expect that any difference between two halls that can be perceived by our ears can also be measured as soon as the technical challenges are overcome. It seems obvious to demand a binaural measurement method that in real time can directly detect conditions that are associated with source broadening or envelopment. It seems natural to start with analysis binaural recordings. In broadcasting, sound recording and sound engineering, several devices that in real time are able to detect effects related to spatial perception, e.g. stereo indicators, correlation meters, the Goniometer and the Stereo Performance Meter are being used. In Psychoacoustics and the field of Spatial Hearing, binaural models and binaural activity mapping techniques are under development. It would be interesting to explore whether existing and future techniques could be used in the design of an ASW-meter or an LEV-meter. Since such a device potentially could provide extended insight in perceptual aspects in Auditorium Acoustics, this investigation seems worthwhile. Although the answer is not yet found, this paper is dedicated to the question posed in its title..

akuTEK navigation:

Home
Papers
Articles
Title Index
akuTEK Research
Concert Hall Acoustics

CAN SOURCE BROADENING AND LISTENER ENVELOPMENT BE MEASURED DIRECTLY FROM A MUSIC PERFORMANCE IN A CONCERT HALL?

M Skålevik AKUTEK and Brekke&Strand

1 INTRODUCTION

Spaciousness is the subjective impression created by sound reflections from the surrounding boundaries defining a space or room that contains a sound source and a listener, in particular in an auditorium. Two aspects of spaciousness, Source Broadening, or Apparent Source Width (ASW), and Listener Envelopment (LEV) have for several decades been held to be some of the most critical perceptual aspects of a concert hall. There seems to be consensus among acousticians that these aspects relate to quantities that can be analyzed from impulse response measurements, namely LF and 1-IACCE for ASW, and LLG and 1-IACCL for ENV. The first two can be calculated from the first 80ms of the impulse response, while the latter two are calculated from the part after the first 80ms of the impulse response. While such data are quite straightforward to acquire in an empty auditorium, they are less straightforward to acquire with orchestra and audience in place. There is uncertainty related to the difference between unoccupied and occupied conditions. Moreover, an orchestra appears, in many seats in the first 20 rows anyway, to produce quite a broad sound image in itself. In such seats, it is presently unclear how much difference the room makes to the apparent width, even if recommended values of LF and 1-IACCE are found in the seat.

Until otherwise proven, one would expect that any difference between two halls that can be perceived by our ears could also be measured as soon as the technical challenges are overcome. It seems obvious to demand a binaural measurement method that in real time can directly detect conditions that are associated with source broadening or envelopment. It seems natural to start with analysis of binaural recordings.

In broadcasting, sound recording and sound engineering, several devices that in real time are able to detect effects related to spatial perception, e.g. stereo indicators, correlation meters, the Goniometer and the Stereo Performance Meter, have been used. In Psychoacoustics and the field of Spatial Hearing, binaural models and binaural activity mapping techniques are under development. It would be interesting to explore whether existing and future techniques could be used in the design of an ASW-meter or an LEV-meter. Since such a device potentially could provide extended insight in perceptual aspects in Auditorium Acoustics, this investigation seems worthwhile. Although the answer has not yet been found, this paper is dedicated to the question posed in its title.

Stereophonic sound (hereafter Stereo) in anechoic conditions or with closed headphones can produce the perceptive aspect referred to as Spaciousness in auditorium acoustics.

2 RELEVANT CONCEPTS

The basics of spatial hearing was presented by Blauert (1974)[1]. A more recent quick guide to Binaural Hearing is given online[2]. Our brain has evolved the ability to detect the azimuth angles of incident sound, based on a combination of cues found in the inter-aural time difference ITD and the inter-aural level-difference ILD caused by a source producing sound at both ears. ITD cues divide into phase delay cues and group delay cues. Phase delay cues dominate at low frequencies, while group delay cues and ILD cues dominate at high frequencies, with some crossover region at 0.8-1.6kHz. ITD cues are based on our brains ability to compute correlation between the sound signal entering the left ear and the sound signal entering the right ear. In this paper, only the strictly relevant details will be mentioned.

2.1 Correlation as an indicator of Spaciousness

The (Normalized) Correlation (Coefficient) r between the left signal L and the right signal R, assuming sample means equal to zero, can be computed as follows: First acquire a stereo-sample pair sequences $\{L_i, R_i\}$ from a proper period T, containing n sample pairs with index i=1,2,..,n. Then calculate the denominator D and the nominator N of the coefficient,

$$r = N/D, \text{ where and } N = \Sigma\, L_i \cdot R_i \text{ and } D = \text{sqrt}\, [\Sigma\, L_i^2 \cdot \Sigma\, R_i^2] \qquad (1)$$

Two examples of calculation are presented in Table 1. Note that the “normalizer”, D , is the square of the product of total energy at the two ears during the period T.

In auditorium acoustics, a source to the front of the listener, i.e. azimuth angle equal to zero, is the typical and most interesting case. This case can be investigated by recording signal pairs with two properly spaced microphones on the axis perpendicular to the axis pointing at the source. In order to mimic our binaural hearing, the microphone could be placed in the ear canal of a real person, a mannequin, or a dummy head. Combined with anechoic-like conditions, the signal pairs would exhibit ITDs and ILDs close to zero and correlation close to unity.

Mathematically, correlation is an algorithm which with the input of a pair of number sequences, both having the same length, returns a single number output, denoted r, in the interval of [-1,1]. In audio-signal-processing, the sequence pair would be samples of an audio signal pair. If two signals are equal like in the ideal aforementioned case, the output would be r=1. If adding sounds from other directions than azimuth=0, this would create differences between the left ear signal and the right ear signal, and the value of r would be reduced, depending on how different and how strong the added sounds are. In a reverberant field, with sufficiently low direct-to-reverberant ratio combined with sufficiently high fraction of lateral reflections, values like of r =0.1-0.2 are commonly found. Our brain has learned to associate low values of r with a presence of surrounding walls. If introducing a sound reflecting ceiling or floor, the impression of surround would be reduced, an effect that would correspond to a higher value of r. The reason for this is that sound sources or image sources in the sagittal plane, i.e. the set of points at azimuth angle equal to zero, e.g. straight front, straight above, straight behind, straight below, and all positions in-between, would produce similar signals at both ears, thus contribute to raise the value of r.

Typically, a listener would sub-consciously try to turn or adjust the head toward the source in order to optimize correlation at azimuth equal to zero. The same would be the first choice for a two-microphone measurement, a stereo recording or a binaural recording. For many reasons, it could be practical to use the same measurement setup to measure correlation between binaural signal pairs arriving from sources from other azimuth angles than zero. There may be sources at different positions and many seats in an auditorium are not oriented straight towards the speaker, musician or other source of interest.

Indeed, our brain evaluates correlation in many azimuth directions simultaneously. This can be mimicked with a simple signal processing technique: Every azimuth angle would be corresponding to a time difference (ITD) between the signal paths from the source to the two ears. For a source to the right, simply apply a time delay τ to the R-signal before calculating r, see example in Table 1. By choosing using τ =ITD to cancel the time difference, correlation can be measured in many different azimuth directions from the same recording, in order to detect sources. Given a signal pair sampled at 44.1 kHz, the smallest possible time difference would be 23 microseconds, corresponding to an azimuth angle of approximately 2 degrees. Thus, correlation of binaural signals arriving from this azimuth angle could be evaluated by shifting one of the signal sequences by one sample before applying the algorithm. In general, by computing correlation of the sequence pair over the period T, with a range of varying time delays, τ, we have $r(\tau)$ being a function of time delays.

ISO-3382[3] defines the Inter-Aural Cross-Correlation IACC to be the highest normalized correlation found in all possible delays within ±1ms. Early IACC, IACCE integrates over the first 80ms of the impulse response, while late IACC, IACCL integrates over the 80-1000ms interval. Note that IACC provides no information about which azimuth angle the output value corresponds to, or the correlation at other angles than the one with highest correlation.

Table 1 Two matrices for computation of correlation between two stereo sequences of 8 samples length (L-signal and R-signal), based on the algorithm in (1). The LR-product row is the product of each sample pair. r equals the sum of the LR products divided by the product of the Root in the L-signal row and the Root in the R-signal row. First, calculate the square of each sample, then sum the sample squares to get Square sum in each row. Then Root is the square root of Square sum. In the upper matrix, the L-signal and R-signal are perfectly equal, thus r=1.0. In the lower matrix, the R-signal is shifted (delayed) by one sample, in order to investigate whether there is source at a small azimuth angle to the right. However, the lower correlation (r=0.44) indicate there is not. The Root product is the Denominator in the correlation coefficient, and is not altered by the shift (lead) in the Left-signal.

Sample No *i*	1	2	3	4	5	6	7	8	Square sum	Root	N = Product sum	D = Root product	*r = N/D*
L-signal	0	1	2	1	0	-1	-3	0	16	4		16	16/16
R-signal	0	1	2	1	0	-1	-3	0	16	4			
LR-product	0	1	4	1	0	1	9	0			16		=1,00

Sample No *i*	1	2	3	4	5	6	7	8	Square sum	Root	N = Product sum	D = Root product	*r = N/D*
L-signal	0	1	2	1	0	-1	-3	0	16	4		16	7/16
R-signal	0	0	1	2	1	0	-1	-3	16	4			
LR-product	0	0	2	2	0	0	3	0			7		=0,44

2.2 The relationship between correlation and the difference signal L-R

Note that correlation under typical listening conditions can be interpreted as a measure of similarity or difference between the sounds at the two ears, where close to 1 means similar, close to 0 means different, and close to -1 means similar but with opposite polarity, i.e. in anti-phase. Values in the range 0 to -1 would in sound engineering, production and broadcasting, indicate that a stereo signal includes content that would cause cancellations if converted to a mono-signal. In contrast, *r* in the range 0 to 1 would indicate a so-called mono-compatible stereo signal.
A more direct measure of similarity and difference would be to measure the sum L+R and the difference L-R of the left signal L and the right signal R. For each sample pair *i*, the products in in the series of the Nominator could be obtained by

$$L_i{\cdot}R_i = \tfrac{1}{4}{\cdot}[\,(L_i+R_i)^2 - (L_i - R_i)^2\,]. \qquad (2)$$

Thus, *r* can be interpreted as the difference between the energy of the sum signal L+R and the energy of the difference signal L-R, normalized to the total energy, i.e. the denominator *D,* of the measurement period T.

2.3 Impulse-response based measurements of Spaciousness

The result (2) explains why indicators of Spaciousness can be measured with a combination of a figure-8 microphone and an omni-directional microphone, and why such indicators yield similar evaluation as those utilizing the correlation techniques. A figure-8 microphone is basically two closely spaced omni-microphones with opposite polarity, thus acquiring the aforementioned difference-signal L-R. The omni-microphone would acquire the sum-signal L+R. This measurement technique is used to obtain the Early Lateral Energy Fraction (LF) by integrating $(L\text{-}R)^2$ and $(L\text{+}R)^2$ over the from the first 80ms of an impulse response and calculate LF from the ratios of those integrals. LF correlate with 1-IACCE, and they are both known to be effective indicators of Apparent Source Width, or Source Broadening. Lateral reflections would tend to increase as IACC decreases. However, the different microphone techniques involved in the measurement of LF and IACC may produce significant differences. IACC and binaural recordings applies recording techniques that are more similar to the way our ears pick up sound than the figure-8 and omni technique. Besides, LF is designed to be measured in unoccupied conditions, and would during a concert in occupied auditorium be affected by the audience in a way that is not relevant to the sound perceived by our ears. This is in contrast to IACC and binaural recordings, which would become more relevant to listener aspects and perceived Spaciousness when audience is present than when they are not. Compared to the application of difference signal and sum signal in (2), the sum signal would in binaural recording be less relevant due to the bigger distance between microphones. Instead of the energy of the sum signal $(L_i+R_i)^2$ in (2), we can use the sum of the energies $L_i^2+R_i^2$ and write

$$L_i \cdot R_i = \tfrac{1}{2} \cdot [\, L_i^2+R_i^2 - (L_i - R_i)^2 \,]. \qquad (3)$$

To use the sum of energies, or the average of energies, $\tfrac{1}{2} \cdot (L_i^2+R_i^2)$, could be even more convenient, since it is statistically highly correlated with the Denominator in the correlation. The latter could be interpreted as the logarithmic average of the energy at L and the energy at R over the period T.

2.4 Loudness of late reverberant energy

When evaluating Envelopment (LEV) it is important to include the strength of the late lateral sound, i.e. reflections arriving at least 80ms later than direct sound, and not only the correlation value. This is because even very weak reverberant sound, or background noise, can give low *r*-values and high 1-IACCL values, without being strong enough to create the spatial impression feature known as Envelopment. Late Lateral G is measured with a calibrated figure-8 microphone and thus relates the measure to an absolute reference, namely the level of anechoic sound at 10m from the source. Beranek (2008) suggested a combination of 1-IACCL with G_{Late} in order to emphasize the significance of the strength of the late reverberant sound.

2.5 Tools utilized in music recording and broadcast

While the techniques for measuring ASW and LEV based on impulse responses is well developed, the objective of this paper is to evaluate these perceptual features in running music during a concert. Among Spaciousness, ASW and LEV, the two latter would be the more challenging. Spaciousness in general could be measured with the tools applied in sound engineering in order to evaluate the stereo content in a diotic signal, i.e. a signal presenting sound at both ears simultaneously. The Goniometer would provide an intuitive display of Spaciousness, and the correlation meter would quantify Spaciousness on a scale from -1 to 1, where the range between 0 and 1 would be typical in auditorium acoustics. A goniometer is a development of X-Y displays or Lissajous applied in oscilloscopes, where the display is rotated 45 degrees counterclockwise to obtain the proper left-right orientation as seen by the operator. See Figure 1.

Figure 1 Goniometers and Correlation meters. An example of a spacious stereo signal with low correlation (Left), and an example of poor Spaciousness and high correlation.

3 MEASURING LIVE SPACIOUSNESS

Through decades of research, we have learned to find indicators of ASW and LEV in impulse responses measured in empty auditoria in absence of other sounds. The present task is to learn how to find indicators of Spaciousness in running music, i.e. Live Spaciousness, assuming that since ASW and LEV are perceptible by the brain, it is also possible to measure. The task is a matter of developing the proper technology. In running music, pure impulses followed by silence are rare. However, music is full of pulses, level steps (rise or fall in energy levels) and other dynamic properties.

Based on the similarities between the difference signal and correlation, and the advantages of using a recording technique that mimics our binaural hearing, this author suggests the following binaural measures in order to investigate Live Spaciousness, see Table 2. As a start, we study time-varying Spaciousness in terms of Binaural Lateral Fraction. Ultimately, we wish to study the time-varying Spaciousness as a response to the “perceived input”, i.e. the frontal sound, and try to find techniques to extract the latter from the total energy. However, there are hurdles to overcome. All measures are functions of time t, calculated by integrating the square sound pressures L^2 and R^2 at the left and right ears respectively, and the square difference pressure $(L - R)^2$, over the time interval [t – T, t]. The integration time T needs consideration, and is suggested to be somewhere in the interval 30-160ms. The low limit is related to the infinite separation between events, and the high limit to the buildup of Loudness (compare to time constants 35ms and 125ms in sound level meters). The set of measures assume the total energy at the ears are divided in two parts, namely the sound energy arriving from points close to the sagittal plane, i.e. azimuth close to zero, and lateral energy, i.e. at azimuth not close to zero. Sagittal could be dubbed Non-Lateral. The division between the two parts, in terms of azimuthal limit, would depend on frequency weighting, and needs more consideration. With unfiltered BLL, the sagittal plane would appear to be wider at low frequencies than at high frequencies.

Table 2

Quantity	Notation [unit]	Formula
Binaural Total Energy Level	BTL (T,t) [dB]	10 lg [$L^2 + R^2$]
Binaural Lateral Energy Level	BLL (T,t) [dB]	10 lg [$(L - R)^2$]
Binaural Sagittal Energy Level	BSL (T,t) [dB]	10 lg [$L^2 + R^2 - (L - R)^2$]
Binaural Lateral Energy Fraction	BLF (T,t) [1]	$(L - R)^2 / (L^2 + R^2)$

Returning to the challenge of sub-dividing Spaciousness into ASW and LEV in running music, the problem seems to be the lack of a “trigger” or a que for an onset of direct sound, in order to identify short signal periods that contain early sound similar to those of the first 80ms of an impulse response. Klockgether and van de Par (2014) presented a promising technique. They divided the running binaural signal into energy bins of 30-40ms length. Those bins who exhibited more energy than the former bin was detected as one that contained onset of direct sound. For all these bins, the correlation between the left and right signal was computed. On this basis the ASW of running music signal was evaluated. The results were compared with IACC measurements based on impulse response, and showed quite good agreement. Signal parts dominated by late reverberant sound could be detected in a similar manner, to obtain a basis for evaluating LEV.

Given the discussions above, this author suggests to use correlation to divide the running music signal into two parts, in principle demonstrated on impulse responses in Figure 4. Again, length of energy bins (integration interval T) needs to be settled. Bins with high correlation could form the basis for evaluating ASW, and the parts with low correlation could form the basis for evaluating LEV. As mentioned, low correlation combined with strong sound would be indicating high fraction of lateral sound energy. Again, it would not be sufficient to identify low correlation.

It seems likely that the dynamics of correlation may play an important role. Constant low correlation is not likely to create Spaciousness. The interruptions of direct sound and parts of higher correlation is expected to be, not just inevitable, but actually the contrasting elements that are necessary condition for the Dynamic Spaciousness. These features can be seen in the examples, Figure 2.

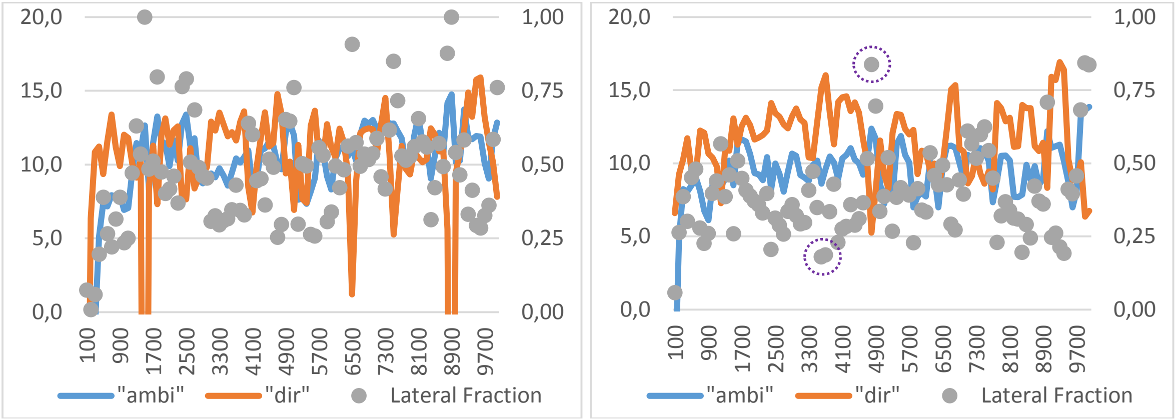

Figure 2 Dynamics in Live Spaciousness. Example of diagrams from binaural recordings, using the energy of the difference signal L-R to measure lateral content of running music in big concert halls in Grieghallen, Bergen (left) and Avery Fisher Hall, New York. The music content is identical, the first 10 seconds of the Dance of the Knights, Prokofiev's Romeo and Juliet, played by the respective residential orchestras. "ambi" is level of lateral sound sound, "dir" is level of frontal (sagittal) sound. Horizontal axis is time in milliseconds. Vertically, left axis is dB, and right axis is dimensionless Binaural Lateral Fraction. Dotted circles correspond to Goniometer displays in Figure 3 in APPENDIX. To the left, Spaciousness appears stronger and with more dynamics.

3.1 Other methods

Lokki (2015) has pointed at the importance of including factors like Lateral Loudness and the spectral and dynamical aspects of Source Broadening: "I would analyze somehow the width of the source from pianissimo and fortissimo passages separately. Marshall calls this 'spatial responsiveness' when source is more wide in fortissimo parts than in pianissimo parts. And I think this is extremely important for 'good acoustics'. This is also related to dynamics, which is enlarged (and compressed) in some halls, but that you cannot analyze dynamics from your data, as your input is not fixed accurately. But the widening of the source according to played dynamics you could analyze…. About HF and LF, I think 63 Hz and 4 kHz octave bands are the most important ones. They are sometimes hard to measure, but they are those that best discriminate the halls."[5]

This author suggests that even other methods could be based on binaural activity models, like those investigated by Blauert[2], Hess[6] and Schuitman[7,8].

4 SUMMARY

In this paper, the possibilities and challenges related to measurements of spaciousness during music performances in auditoria are presented. It is substantiated that Correlation and Difference in the signal pair obtained in binaural recordings contain information that could provide indicators of Spaciousness, and its two listener aspects, Source Broadening (ASW) and Level of Envelopment (LEV). Sound engineers have for a long time monitored Spaciousness during work with stereo recordings, however without any distinction between ASW and LEV.

5 FURTHER WORK

The significance of Integration intervals (T) and frequency weighting of the measures will be investigated in further work. If music was always long sustained notes and legato play at the same level, there would maybe have been only one aspect of Spaciousness, namely envelopment, which could have been measured quantified by the correlation between the left and right signal. However, dynamics is a very important property of music. Statistical Variance in spaciousness and other ways to evaluate the dynamics of Spaciousness will be investigated. It should be investigated whether more advanced binaural hearing models, including binaural activity maps, would provide more effective indicators of Spaciousness, and more insight into the perceptive phenomenon.

6 REFERENCES

1. J.Blauert, Raumliches Hören, Hirtzel Verlag 1974 (Spatial hearing, the psycophysics of sound localization)
2. J.Blauert, Binaural Models and their Technological Application (Key Note lecture) http://www.iiav.org/docs/2_slides_blauert_vilnius.pdf
3. International Standard, ISO-3382 Acoustics – Measurement of Room Acoustic Parameters – Part 1, Performance Spaces, 1st edition (2009)
4. S.Klockgether and S. van de Par, A Model for the Prediction of Room Acoustical Perception Based on the Just Noticeable Differences of Spatial Perception, Acta Acustica United with Acustica, 100, 964-971.
5. T. Lokki, personal communication, 19.04.2015
6. W.Hess, Binaural activity characteristics as indicator of auditory spatial attributes (2006), PhD Thesis
7. J.van Dorp Schuitman, Auditory Modelling for Assessing Room Acoustics, PhD Thesis, TU Delft, 2011.
8. J.van Dorp Schuitman and D de Vries, Developing Content-Specific Parameters of Room Acoustical Quality Using a Binaural, Nonlinear Auditory Model (2010), AES

APPENDIX

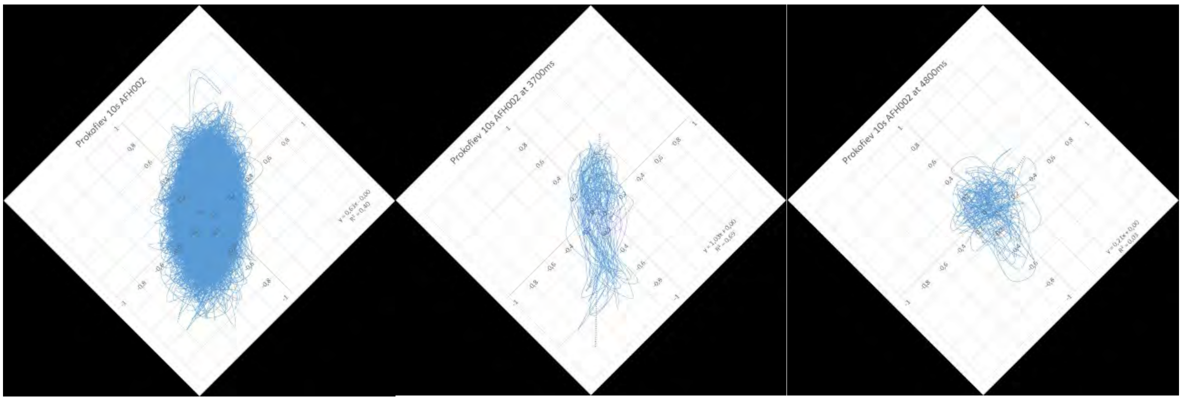

Figure 3 Goniometer displays corresponding to Figure 2: Integrated over 10s (left); Integrated over 100ms at 3700ms (middle), Lateral Fraction =0.18; Integrated over 100ms at 4800ms (middle), Lateral Fraction =0.84.

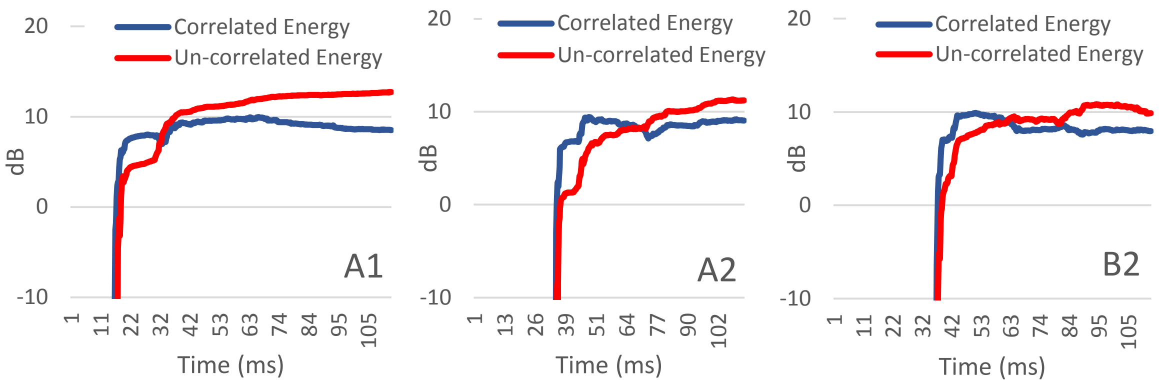

Figure 4 Spaciousness build-up in three Binaural Impulse Responses in the same 1500cbm room. Correlated energy curves are Energy·r^2 , integrated from 0 to t (ms). Un-correlated energy curves are Energy·(1-r^2) , integrated from 0 to t (ms); “A” is central source position, “B” is lateral source position, “1” is central receiver position 6m from “A”, “2” is central receiver position 12m from “A”.

akuTEK navigation:
Home
Papers
Articles
Title Index
akuTEK research
Concert Hall Acoustics