OPTIMISATION OF SPEECH INTELLIGIBILITY IN MULTI-SOURCE ENVIRONMENTS USING DELAY SPREAD MINIMISATION

Evert Start Duran Audio BV, Zaltbommel, The Netherlands

1 INTRODUCTION

The design of public address (PA) or voice alarm (VA) systems in large, complex-shaped spaces can be quite challenging. Particularly in highly reverberant spaces, it is difficult to achieve the desired level of speech intelligibility. Traditionally, numerous spatially distributed loudspeakers are used, causing a strong excitation of the reverberant sound field. Nowadays, the number of sources can be dramatically reduced by applying highly directional loudspeaker arrays, capable of covering large audience areas. Using advanced directivity pattern control techniques[1], the sound can be precisely aimed at the audience. As a result, the ratio between direct and reflected energy increases, leading to a significant improvement of perceived "directness" and speech intelligibility.

Nevertheless, in very large environments or in spaces without clear lines of sight, a (sparsely) distributed loudspeaker array set-up is often inevitable.

In distributed sound systems multiple direct sound paths exist from the loudspeakers to each receiver. Due to differences in path length, the direct sound contributions are dispersed in time. A large time-dispersion has a negative effect on speech intelligibility and may even cause artificial echoes. Therefore, a well-chosen loudspeaker layout and careful time-alignment by electronic delays are essential in any distributed loudspeaker (array) design. Practice shows that time-alignment can be a difficult and a time-consuming job. Moreover, a simple time-of-flight correction scheme only provides a local, and consequently, sub-optimal solution.

In this paper, the Delay Spread, which is a physical measure of time-dispersion in a multi-source sound system, will be introduced. Also, a loudspeaker delay optimisation algorithm is described, which searches for the optimum loudspeaker delays for a given loudspeaker (array) layout. This new tool is implemented in the AXYS Digital Directivity Analysis (DDA) software. By conducting various electro-acoustic simulation experiments, the performance of this new tool is investigated.

2 PERCEPTION OF ECHOES

2.1 Some psycho-acoustic findings

A lot of research has been done into the influence of reflections on the perception of sound. The work done by Haas[2] is probably the most well-known. He found that the disturbance by one single reflection depends on the level of the reflection relative to the direct sound and the delay time from the direct sound. A higher level is required to disturb at a shorter delay time, as can be seen from Figure 1.

As the +10 dB scenario is more disturbing than the –10 dB scenario, it can be concluded that pre-echoes are more disturbing than post-echoes. This corresponds to the generally accepted concept that forward masking effects are stronger than backward masking effects.

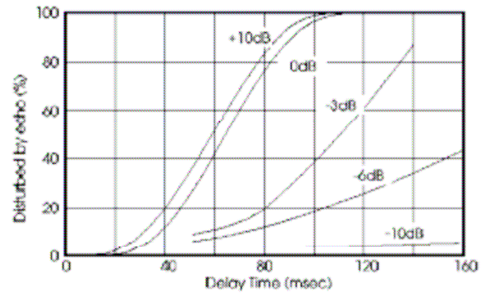


Figure 1: Percentage of listeners disturbed by a delayed speech signal (From Haas[2]). Curve numbers show the level difference between the direct sound and the echo.

Figure 2 gives the results of several measurements of the echo threshold collected by Blauert[3]. Lochner and Burger used the criterion "echo clearly audible", Meyer and Schodder the criterion "echo barely inaudible".

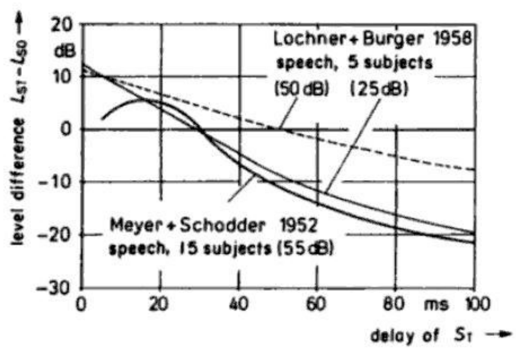


Figure 2: Echo thresholds for continuous speech of average speed (~5 syllables/s) for different detection criteria (From Blauert[3]).

Although the findings of these studies vary, it is widely acknowledged now that early reflections, arriving within approx. 50 ms after the direct sound, support the direct sound, whereas late reflections are detrimental to speech intelligibility.

2.2 STI in single reflection fields

The Speech Transmission Index[4,5] (STI) is widely accepted as a objective measure of speech transmission quality in rooms. Although speech transmission quality is not identical to speech intelligibility, STI is used as an indicator of speech intelligibility.

Based on the results in the previous paragraph, it is expected that, in a single reflection sound field, speech intelligibility is inversely related to echo delay and echo level. A similar inverse relationship would be expected for STI, being an indicator of speech intelligibility. However, it can be easily verified that STI doesn't decrease monotonically with increasing delay time of a single reflection. This behaviour was also reported by Onaga[6], who investigated the relation between STI and speech intelligibility for some synthetic sound fields. Figure 3 shows STI as a function of delay time and relative level of a single reflection.

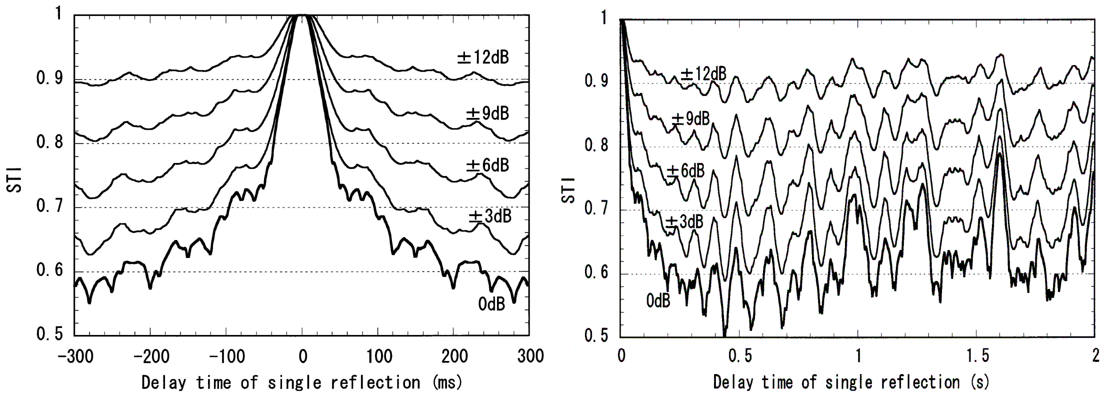


Figure 3: STI as a function of the delay time of a single reflection. Parameter is reflection level relative to direct sound (From Onaga[6]).

Initially, STI rapidly drops with increasing delay times. After approx. 50 ms delay, STI is dropping more slowly and exhibits some minor fluctuations. After 0.3 s, STI hardly drops at all and shows some major fluctuations. From the symmetry of STI curves in the left plot of Figure 3. it is also clear that the STI values for positive and negative delay times are equal. Similarly, STI takes the same value for both positive and negative levels of reflection. This behaviour of STI contradicts the generally accepted concept that pre-echoes are more disturbing than post-echoes (as discussed in §2.1).

3 LOUDSPEAKER DELAY OPTIMISATION

3.1 Method

In order to optimise speech intelligibility for a given distributed loudspeaker (array) set-up, we have to search for the optimum combination of loudspeaker delays. This can be done by defining a cost criterion, which must be a function of the loudspeaker delays. By minimising this cost function, using a numerical search algorithm, the optimum loudspeaker delays can be found. As we are searching for a minimum, the cost criterion should be inversely related to speech intelligibility.

3.2 Initial unsuccessful criteria and considerations

One of the first ideas was to minimise the spatially averaged "speech transmission loss" (1-STI) with respect to the loudspeaker delays. However, pilot tests showed that the iterative calculation of STI is too computationally intensive for large sound systems. Moreover, there is a risk that, due to the irregular relationship between STI and loudspeaker delay, the search algorithm gets stuck in a local minimum. Most likely, this will occur when the loudspeaker delays are still far away from their optimum values, i.e., the when the delay differences are still large.
As an alternative, the use of indirect measures of speech intelligibility, based on early/late energy ratios, such as the Deutlichkeit (D_{50}) and the Clarity Index (C_{50} or C_{80}), was considered. However, due to the sharp split time between early and late energy, a small change in the arrival time of a strong loudspeaker contribution may result a significant change of the energy ratio, which might pose a problem to the search algorithm.

3.3 Delay spread

Multipath propagation is a well known phenomenon in telecommunications. In radio and television broadcast and wireless or AC power line communications, it might cause bad reception or poor transmission rates, respectively. One of the most important indicators of the performance of a communication channel is the RMS Delay Spread[7], which is a measure of the time-dispersion in the transmission channel. In a modified form, the concept of delay spread can also be used in electro-acoustics to quantify time-dispersion in distributed sound systems.

The (mean absolute) delay spread is defined as the first absolute central moment of the squared, anechoic impulse response *d(t)*:

$$\sigma = \frac{\int_0^\infty |t - \bar{\tau}| d^2(t) dt}{\int_0^\infty d^2(t) dt} \tag{1}$$

where

$$\bar{\tau} = \frac{\int_0^\infty t d^2(t) dt}{\int_0^\infty d^2(t) dt} \tag{2}$$

is the mean delay, which is "centre of gravity" of the anechoic impulse response. Note that Eq.(2) is almost identical to the definition of Centre Time (T_s), in room acoustics. The definition used here only takes the direct sound from the loudspeakers into account. The frequency range, for which the delay spread is calculated, covers the 500 to 2k Hz octave bands.

The delay spread σ can be interpreted as the mean, energy-weighted, absolute deviation from the mean delay $\bar{\tau}$. Because only the deviations are taken into account, the delay spread is not dependent of the absolute time of flight. Therefore, no zero-time correction of the impulse response is required.

Figure 4 shows a schematic example of an anechoic impulse response of a distributed loudspeaker system. The impulse response in this example is determined by six loudspeaker contributions. The amplitude of the impulses represents the (relative) intensity of the loudspeaker contributions summed over the 500 to 2k Hz octave bands.

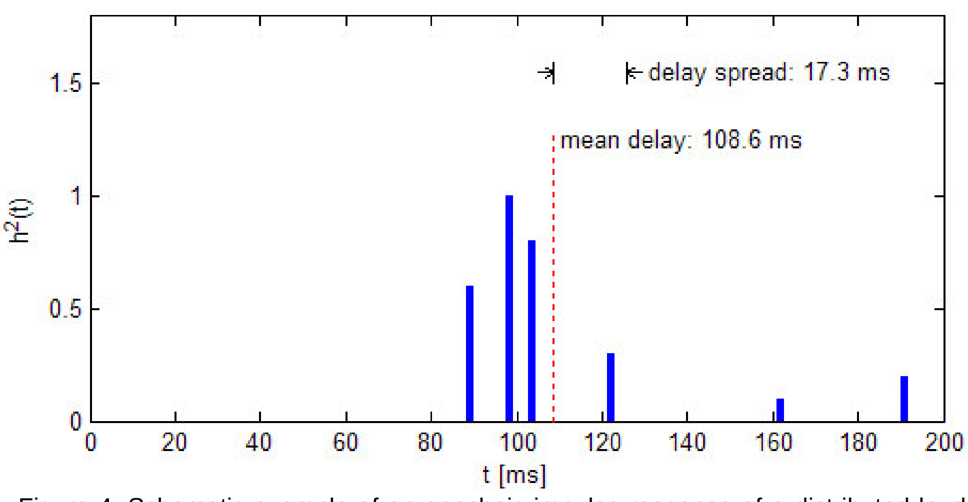


Figure 4: Schematic example of an anechoic impulse response of a distributed loudspeaker system.

Pilot experiments showed that, compared to the RMS delay spread used in communications, the definition of Eq. (1) is a more stable and robust measure of delay spread. Contrary to the RMS method, large delay deviations are not unduly weighted in the proposed method.

3.4 Cost function

The anechoic, squared impulse response of a sound system consisting of L loudspeakers at receiver m is modelled by:

$$d_m^2(t) = \sum_{l=1}^{L} a_{m,l}\,\delta\left(t - \frac{r_{m,l}}{c} - \tau_l\right) \qquad (3)$$

The intensity of the direct sound from loudspeaker l at receiver m is represented by $a_{m,l}$. The first delay term is the time of flight, determined by the distance $r_{m,l}$ between loudspeaker l and receiver m and the speed of sound c. The second delay term represents the electronic delay of loudspeaker l.

The intensity $a_{m,l}$ is determined by the loudspeaker-receiver distance $r_{m,l}$ and the directional response of the loudspeaker. If the line of sight between the loudspeaker and the receiver is interrupted by an obstacle, the intensity $a_{m,l}$ is zero.

The total direct sound energy at receiver m is calculated by integrating $d_m^2(t)$ over time, yielding

$$E_m^{tot} = \int_0^{\infty} d_m^2(t)\,dt = \sum_{l=1}^{L} a_{m,l} \qquad (4)$$

The mean delay for receiver m, as a function of the loudspeaker delays is given by

$$\bar{\tau}_m(\tau_1, \tau_2 \ldots \tau_L) = \frac{\int_0^{\infty} t\,d_m^2(t)\,dt}{E_m^{tot}}. \qquad (5)$$

Now, the delay spread for receiver m, as a function of the loudspeaker delays can be written as

$$\sigma_m(\tau_1, \tau_2 \ldots \tau_L) = \frac{\int_0^{\infty} |t - \bar{\tau}_m|\,d_m^2(t)\,dt}{E_m^{tot}}. \qquad (6)$$

By energy-weighted averaging of the delay spread σ_m over all receivers, the total cost function is obtained as a function of the loudspeaker delays

$$\sigma_{tot}(\tau_1, \tau_2 \ldots \tau_L) = \frac{\sum_{m=1}^{M} E_m^{tot}\,\sigma_m(\tau_1, \tau_2 \ldots \tau_L)}{\sum_{m=1}^{M} E_m^{tot}} \qquad (7)$$

The optimum loudspeaker delays are obtained by minimising the total delay spread σ_{tot} with respect to the loudspeaker delays τ_l.

3.5 Minimisation by simulated annealing

Simulated annealing[8] is a generic mathematical method to find the (global) minimum of a cost function in the presence of many local minima. Random steps in the parameter space are performed. A step is always accepted if the objective function is lowered, and it is sometimes accepted with a certain, decreasing probability, if an uphill step is taken. This scheme allows the algorithm of escaping from local minima.
The name and inspiration come from annealing in metallurgy, a technique involving heating and slow cooling of a material to increase the size of its crystals and reduce their internal defects.

3.6 Software implementation

The algorithm, described above, has been implemented as a loudspeaker delay optimisation tool in the AXYS Digital Directivity Analysis (DDA) software. This tool has the following features:

- Loudspeaker (arrays) can be grouped into zones. All loudspeakers in a zone will keep the same delay.
- The time delay of the loudspeakers or the loudspeaker zones can be locked by the user. Locked delays are kept constant during delay optimisation. Note that, this might prevent the optimiser from finding the global minimum for the delay spread, but will find the minimum delay spread, given the boundary conditions.
- In case none of the delays is locked, the minimum loudspeaker delay is always set to zero.
- If line-of-sight checking is enabled, potential direct sound blocking by obstacles in the model is taken into account.

4 SIMULATION EXPERIMENTS

In order to test the performance of the proposed loudspeaker delay optimisation algorithm, several simulations experiments have been conducted using the AXYS Digital Directivity Analysis (DDA) software.

4.1 Longitudinal loudspeaker array set-up

Consider a sound system consisting of three Intellivox-DS500 loudspeaker arrays aimed in the longitudinal direction, as shown in Figure 5a. The distance between the arrays is 40 m. The relative output levels of the arrays are: 0, -1 and –2 dB, for A1, B1 and C1 respectively. In Figure 5b the direct A-weighted SPL (Male Speech conf. IEC60268-16:2003) is plotted.
In this situation, the loudspeaker delays can be easily calculated by using a simple time-of-flight correction scheme (0, 116.5, 233.0ms, respectively, using c=343.4 m/s). Nevertheless, it's interesting to see which solution is found by the delay optimiser. After optimisation, the delays for array A1, B1 and C1 are 0, 107.3 and 217.4 ms, respectively. The optimised delays are a bit shorter than the "manual" delays.
The calculated STI, before and after delay optimisation, is shown in Figure 5c and d. The spatially averaged STI values are 0.91 and 0.92, respectively. The standard deviations are 0.06 and 0.05 respectively. This indicates that the spatial variation slightly decreases, using the optimised loudspeaker delays.

4.2 Skewed loudspeaker array set-up

Consider a sound system consisting of nine Intellivox-DS500 loudspeaker arrays covering a parallelogram-shaped area, as shown in Figure 6a. The distance between the arrays is 20 m. In Figure 6b the direct A-weighted SPL (Male Speech conf. IEC60268-16:2003) is plotted.
The "manual" calculation of the loudspeaker delays is less straightforward in this situation. So, the loudspeaker delays are initially set to zero . The delay optimiser is used to find the optimum delays. After optimisation, the delays for arrays A1-A9 are ranging from 0 ms for A1 to 297.8 ms for A9. (c=343.4 m/s).

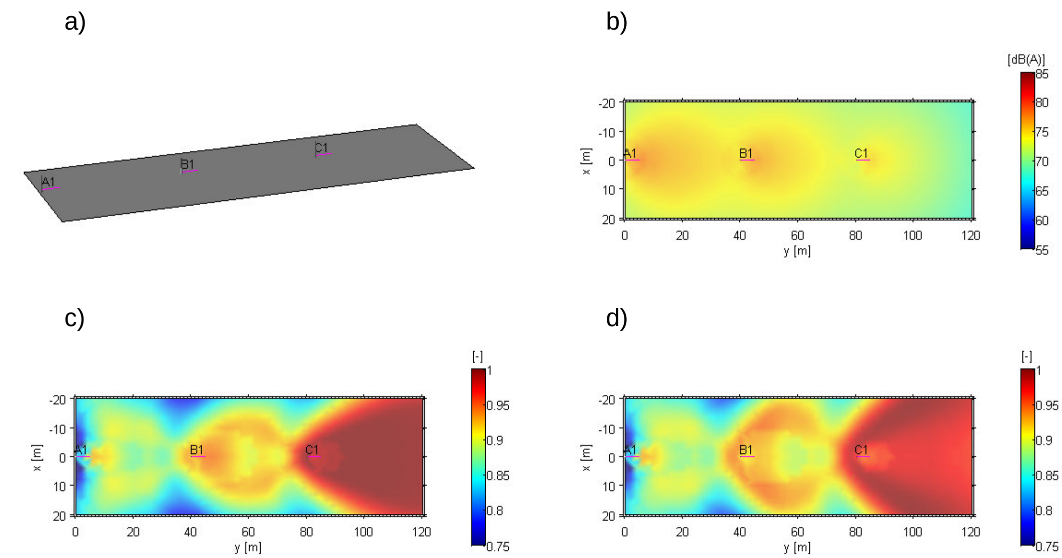


Figure 5: Simulation results for the "longitudinal array set-up".
a: 3D geometry
b: Direct SPL (Male speech, IEC 60268:16)
c: STI, "manual" delays; mean=0.91, std=0.06
d: STI, optimised loudspeaker delays; mean=0.92, std=0.05

The calculated STI, before and after delay optimisation, is shown in Figure 6c and d The spatially averaged STI values are 0.64 and 0.78 with standard deviations of 0.08 and 0.05, respectively. This example shows that by optimising the loudspeaker delays, the average STI can be significantly increased, while less spatial variation is involved.

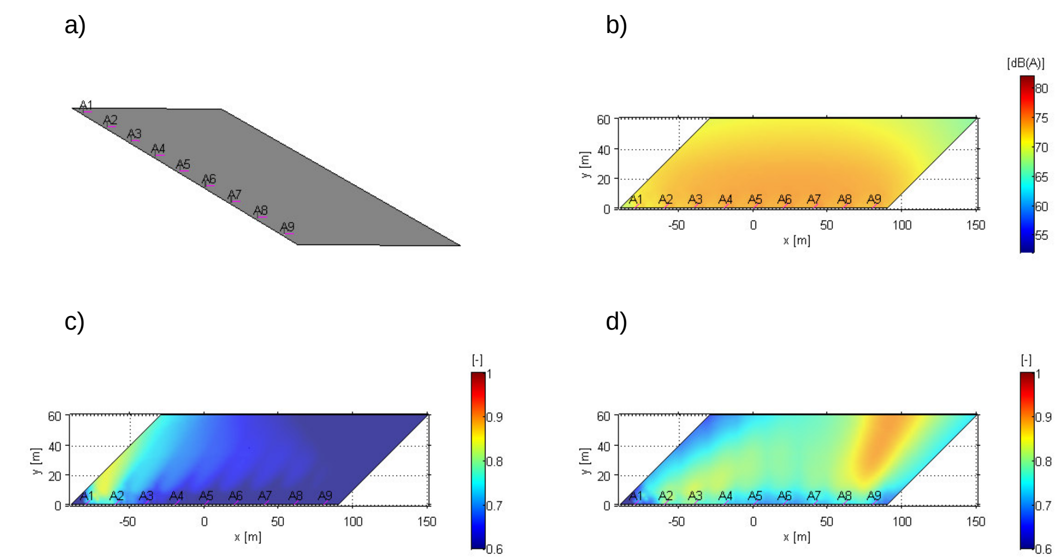


Figure 6: Simulation results for the "skewed array set-up".
a: 3D geometry
b: Direct SPL (Male speech, IEC 60268:16)
c: STI, zero delay; mean=0.64, std=0.08
d: STI, optimised loudspeaker delays; mean=0.78, std=0.05

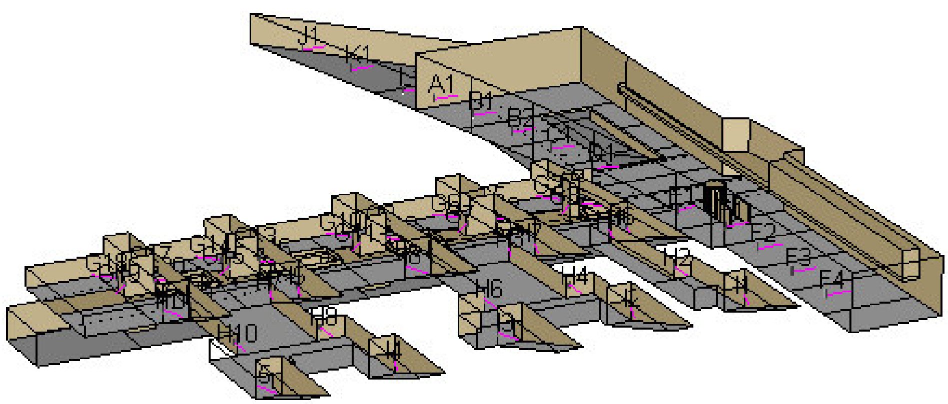
4.3 Complex array set-up

Consider a complex sound system for a train station, consisting of 79 Intellivox arrays of various length as shown in Figure 7a. The building has two underground levels, which are acoustically coupled. The platforms at ground level (not modelled here) are also acoustically coupled via the staircases and escalators. Figure 7b shows the direct SPL (500-2k octaves).

Due to the complex shape of the building and the acoustic coupling between the various floor levels, a well-chosen placement and aiming of the loudspeaker arrays are essential to avoid "unsolvable" timing problems between the loudspeakers. It should be realised that without a sensible loudspeaker set-up, the delay optimiser cannot solve the intrinsic timing problems.

To illustrate the capabilities of the delay optimiser, the course of the delay optimisation process is shown in Figure 8. The total delay spread, due to 79 loudspeaker arrays, was evaluated iteratively for more than 4000 receiver positions. The entire optimisation process (≈15,000 iterations) took less than 15 minutes, which is probably a lot less than the "manual" approach would take.

a)



b)

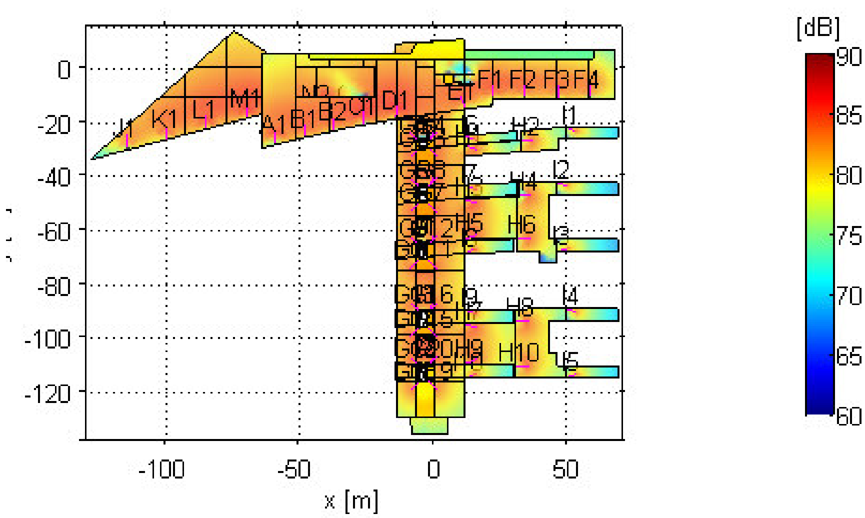


Figure 7: Simulation results for the "complex array set-up".

a: 3D geometry

b: Direct SPL (Male speech, IEC 60268:16)

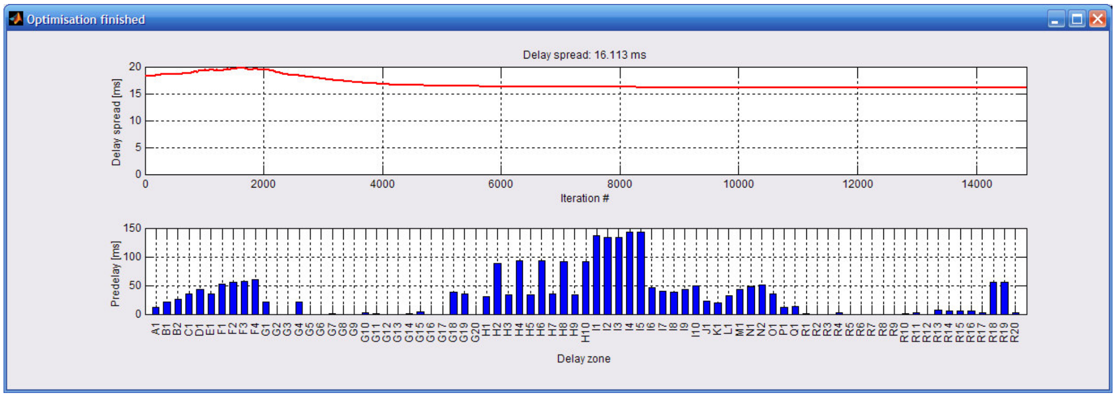


Figure 8: Illustration of the delay optimisation process for the "complex array set-up". Upper graph: course of delay spread for increasing number of iterations. Lower graph: final loudspeaker delays after optimisation.

5 SUMMARY AND CONCLUSIONS

In distributed sound systems multiple direct sound paths exist from the loudspeakers to each receiver. Due to differences in path length, the direct sound contributions are dispersed in time.

In order to quantify time-dispersion in sound systems the Delay Spread was introduced. A loudspeaker delay optimisation algorithm was developed, which searches for the optimum loudspeaker delays for a given loudspeaker (array) layout. This new tool was implemented in the AXYS Digital Directivity Analysis (DDA) software. Using this new tool, the direct sound arrivals of multiple loudspeakers (arrays) can now be automatically aligned in outdoor as well as in indoor applications.

Simulation experiments showed that, even for complex distributed loudspeaker (array) systems, an optimum set of loudspeaker delays can be found, which results in improved speech intelligibility and less spatial variation.

Future work will focus on the optimisation and measurement of STI in real distributed loudspeaker systems.

6 REFERENCES

1. Beuningen van G.W.J. & Start E.W. Optimizing directivity properties of DSP-controlled loudspeaker arrays. Proceedings of the Institute of Acoustics (Reproduced sound) Vol 22 part 6 (2000), p 17 - 37.
2. Haas, H., "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache", Acustica **1** (1951), p. 49-58.
3. Blauert J., Allen J., "Spatial Hearing: The Psychophysics of Human Sound Localization", MIT Press, 1997
4. Houtgast T. and Steeneken H. J. M., ‘‘The modulation transfer function in room acoustics as a predictor of speech intelligibility’’, Acustica 28, 66–73 (1973).
5. Steeneken H. J. M. and Houtgast T., ‘‘A physical method for measuring speech-transmission quality’’, J. Acoust. Soc. Am. 67, 318–326 (1980).
6. Onaga H. et al., `The disagreement between speech transmission index (STI) and speech intelligibility`, Acoust. Sci. & Tech. 22, 4 (2001)
7. Lee, W.C.Y., Mobile Communications Design Fundamentals, 2nd Ed., John Wiley & Soons, 1993
8. Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. 1983, *Science*, vol. 220, pp. 671–680.