A loudspeaker-based room auralisation system for auditory perception research

Jörg M. Buchholz, Sylvain Favrot

Centre for Applied Hearing Research, Technical University of Denmark, Email: jb@elektro.dtu.dk

Introduction

Most research on basic auditory function has been conducted in anechoic or almost anechoic environments. The knowledge derived from these experiments cannot directly be transferred to reverberant environments. In order to investigate the auditory signal processing of reverberant sounds, a loudspeaker-based room auralisation (LoRA) system is proposed here. The LoRA system efficiently combines modern room acoustic modelling techniques with high-order Ambisonic auralisation. Thereby, aspects of the auditory precedence effect are utilized to realise highly authentic room reverberation. This system provides a flexible research platform for conducting auditory experiments with normal-hearing, hearing-impaired, and aided hearing-impaired listeners in a fully controlled and realistic environment. Moreover, the LoRA system can be applied for holistically studying the impact of its individual components on overall perception (or authenticity), considering aspects such as the limitations of Ambisonics auralisation or the details of the employed room impulse response.

In the present manuscript, first the LoRA system is introduced. Afterwards an objective evaluation of the LoRA processing is performed considering different objective monaural and binaural room acoustic measures. The applicability of high-order Ambisonics in the LoRA system for perception research is tested by performing a speech intelligibility experiment.

Loudspeaker-based room auralisation

Within the loudspeaker-based room auralisation (LoRA) system a room acoustic model is provided by the ODEON software [1]. The ODEON output provides a room impulse response (RIR) in eight octave bands, containing detailed information (i.e., intensity, direction of arrival, delay) of the direct sound and the discrete early reflections. The late reverberation is described by energy and vectorial intensity envelopes.

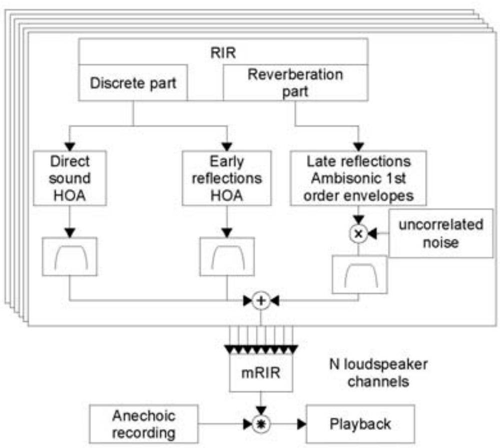

Figure 1: Illustration of the principle signal processing of the LoRA toolbox.

A multi-channel RIR (mRIR) is derived from the ODEON RIR using different auralisation techniques. The discrete RIR elements are either auralised by (high-order) Ambisonics or discrete loudspeakers located in the direction of the elements. The late reverberation envelopes are interpreted as first-order Ambisonics envelopes, which are multiplied with noise that is uncorrelated across loudspeakers. The mRIR is convolved with an anechoic sound signal and played via an array of loudspeakers. The principle LoRA processing is summarized in Figure 1 and further details can be found in [2].

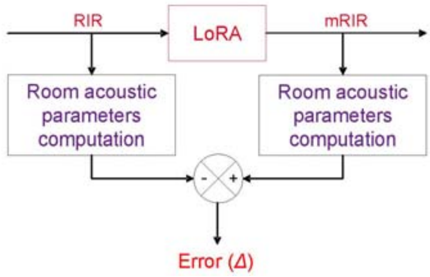

Figure 2: Illustration of the objective evaluation method used to verifying the applicability of the LoRA processing. Different room acoustic parameters are determined at the input and output of the LoRA system, the difference quantifying the introduced error.

Objective evaluation

An objective evaluation is performed to verify that the complex signal processing of the LoRA system does not introduce any significant changes to the original room response provided by ODEON. The principle evaluation method is illustrated in Figure 2.

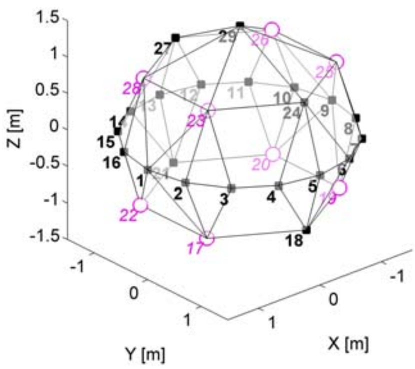

Figure 3: Setup of the 29 loudspeakers exemplarily used in the evaluation of the LoRA system and available in the CAHR laboratories.

Different room acoustic parameters are calculated from: (i) the ODEON room acoustic model at the input of the LoRA system and (ii) the mRIR at the output of the LoRA system. The parameters at the LoRA input are directly calculated by the ODEON software and serve as reference parameter values. The difference between these reference values and the corresponding parameter values at the LoRA output determines the error that may be introduced by the complex LoRA processing. The error is calculated here for eight different source-receiver configurations in a classroom (see Figure 4) and an exemplary auralisation set-up using 29 loudspeakers as shown in Figure 3.

The median and interquartile range of the error introduced by the LoRA processing is shown in Figures 4 and 5. In Figure 4 the errors for the considered monaural parameters are shown: the early decay time (EDT), reverberation time (T30), and clarity (C80). In Figure 5 the errors for the considered binaural parameters are shown: the interaural correlation coefficient for the first 80 ms of the RIR ($IACC_{0,80}$) and the later part of the RIR ($IACC_{80,+}$).

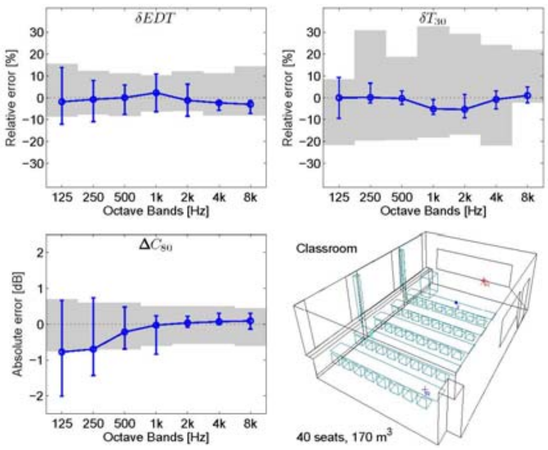

Figure 4: Median and interquartile range of the monaural room parameter errors calculated over eight source-receiver positions in a classroom. The acceptable error range is indicated by the grey shaded areas.

The binaural signals used in the IACC are calculated from the LoRA output mRIR by convolving each of the 29 loudspeaker signals with an ODEON head-related transfer-function that corresponds to the direction of the considered loudspeaker. The final signal is derived by adding all binaural signals from all loudspeakers. A detailed description of the room parameter calculations can be found in the ISO 3382 standard [3]. In order to interpret the error introduced by the LoRA system, an acceptable error range is defined for each room parameter as indicated in Figures 4 and 5 by the grey shaded areas. This acceptable error-range is determined with ODEON by calculating the room parameter variation produced by moving the receiver around an arbitrarily-chosen area of 27x27 cm. Hence, it is assumed that small head movements (within a 27x27 cm area) do not significantly affect the listener's perception of the room.

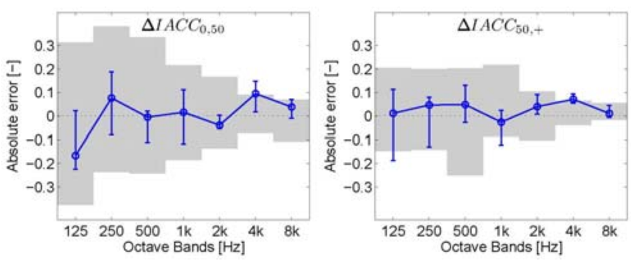

Figure 5: As Figure 4, except that here the binaural room parameter errors are shown.

With reference to Figures 4 and 5, the error introduced by the LoRA processing lies roughly within the acceptable error range for all monaural and binaural room parameters. Similar results were observed for a concert hall in [2], where also a more detailed description of the involved procedures can be found. In synopsis, the objective evaluation encourages the applicability of the LoRA system as a research tool for studying auditory perception in realistic environments.

Subjective evaluation – impact of Ambisonics auralisation on speech intelligibility

The best method of auralising early discrete elements of a RIR is probably by discrete loudspeakers placed in the direction of the individual elements. However, such method requires detailed information on these discrete early elements and moreover, requires a very large number of loudspeakers. Hence, discrete loudspeaker auralisation can not be applied for moving sound sources, complex (arbitrary) acoustic scenes, or when real acoustic scenes are recorded with an array of microphones. In such cases, an alternative auralisation method needs to be applied, such as high-order Ambisonics [4].

In the following subjective evaluation, the limitation of (high-order) Ambisonics auralisation on speech intelligibility measures is investigated with reference to discrete loudspeaker auralisation. Thereby the phenomenon of speech intelligibility enhancement by early reflections [5] is utilized, inspired by the methods proposed by [6].

Methods

Speech intelligibility (word) scores were measured in diffuse background noise as a function of signal-to-noise ratio (SNR). The employed speech signal was either realized by the direct sound alone or the direct sound plus a number of early reflections. The signal level (i.e., the SNR) was varied in two ways: (i) by modifying the direct sound level or (ii) by keeping the direct sound level constant and varying the level of the early reflections. The background noise was realized by diffuse speech-shaped noise with a constant sound pressure level of 60 dB-SPL. The early reflections were calculated for an arbitrary source-receiver (teacher-student) configuration (3.5m distance) in a classroom (see Figure 4). The temporal and directional plots of the employed RIR are shown in Figure 6.

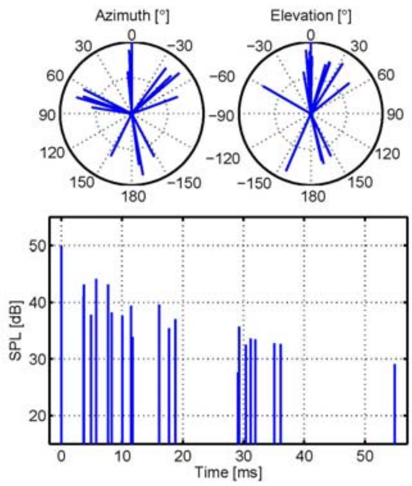

Figure 6: Temporal and directional plots of the RIR considered in the speech intelligibility measurements.

The RIR had an overall duration of 55 ms and the (original) early reflections added 4 dB to the direct sound level. The direct sound was always presented from the front direction. The directions of the discrete RIR elements were modified to fit the discrete loudspeaker directions of the loudspeaker array shown in Figure 3. Three different methods were used for auralising the (reverberant) speech signals: (i) discrete loudspeakers for the individual RIR elements, (ii) 4^{th} order Ambisonics, which is the highest order achievable with the loudspeaker setup shown in Figure 3, and (iii) 1^{st} order Ambisonics, representing the most common Ambisonics realization (in order to minimize coloration artefacts in this case only the eight loudspeakers indicated by the circles in Figure 3 were used).

The experiments took place in an acoustically damped room (T30 < 100 ms for frequencies f > 200 Hz) where a 3D array of 29 loudspeakers was installed (see Figure 3). Nine normal hearing listeners participated in the experiment. The speech corpus consisted of the Dantale II Danish Hagerman sentences [7]. The diffuse background noise started 1000 ms before each sentence with a 600-ms long onset-ramp and stopped 500 ms after the sentence with a 300-ms long offset-ramp. The listener responded via a MATLAB user-interface (see Figure 7) on a hand-held touch-screen, indicating the different words they heard.

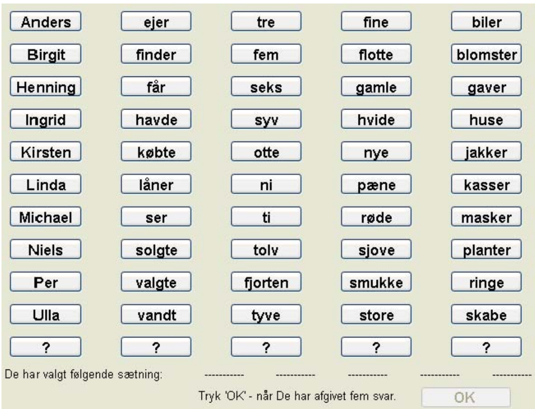

Figure 7: MATLAB user-interface employed in the speech intelligibility experiments and presented via a hand-held touch-screen. Listeners had to indicate the perceived words by pressing the corresponding buttons.

In a first experiment the sensitivity of the individual listeners was determined by measuring the speech reception threshold (SRT). An adaptive procedure according to [8] was applied using the original RIR (see Figure 6) and discrete loudspeaker auralisation. Afterwards, in the main experiment, speech intelligibility scores were measured as a function of SNR relative to the individual SRT, i.e. SNRs were considered of SRT-2, +0, +2, and +4 dB. The listeners had about 30 minutes of training, lists of 20 sentences were used in the SRT measurements, and lists of 10 sentences were used to determine the main intelligibility scores. For further details on the applied methods [9] should be consulted.

Results

The mean intelligibility scores for the 9 subjects are shown in Figure 8 for the case that only the direct speech sound is considered. The standard deviation is presented by the error bars. Discrete loudspeaker auralisation is indicated by blue squares (condition 0), 4^{th} order Ambisonics by red circles (condition 4), and 1^{st} order Ambisonics by black triangles (condition 1). All the data points are significantly different according to a paired t-test. With increasing SNR the intelligibility scores increase, exhibiting psychometric functions that, in Figure 8, are approximated by sigmoid functions $P(L)$ given by:

$$P(L) = \frac{1-\alpha}{1+\exp(4 \cdot s_{55} \cdot (L_{55}+L))} + \alpha \qquad (1)$$

With L the signal level, α the false alarm rate which was here 10 %, L_{55} the 55 % correct threshold relative to the SRT, and s_{55} the slope at $L = L_{55}$. The sigmoid function parameters fitted to the experimental speech data is summarized in table 1.

Conditions	0	4	1	00	44	11
L_{55} (dB)	-2.4	-0.7	1.0	-1.5	0.5	2.5
s_{55} (%/dB)	13.7	12.2	11.3	8.9	10.1	10.45

Table 1: Summary of sigmoid function parameters fitted to the speech data shown in Figures 8 and 9.

According to Figure 8, highest intelligibility scores can be observed for the discrete loudspeaker technique and lowest scores for 1^{st} order Ambisonics. Hence, the spatial comb-filtering effects as well as the limited directionality inherent in the (high-order) Ambisonics technique results in a deterioration of speech intelligibility. This deterioration is strongest for 1^{st} order Ambisonics, which provides the poorest sound field reconstruction. Comparing the different sigmoid function approximations (Table 1), it is observed that mainly the L_{55} threshold is changed by the auralisation technique but not really the shape of the psychometric function. Hence, the auralisation technique seems to mainly affect the effective level of the speech signal (or effective SNR). The observed deterioration of speech intelligibility by imperfect sound field reconstructions is in principle agreement with the deterioration observed for a stereo "phantom" source [10].

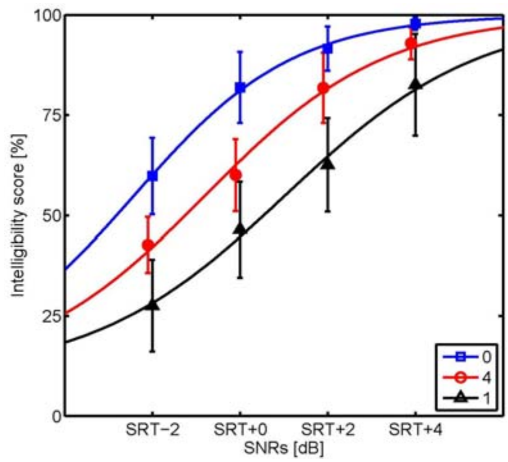

Figure 8: Mean speech intelligibility scores and standard deviations measured for the direct speech sound only. Discrete loudspeaker auralisation is indicated by blue squares (condition 0), 4^{th} order Ambisonics by red circles (condition 4), and 1^{st} order Ambisonics by black triangles (condition 1). Solid lines present approximations by sigmoid functions with parameters summarized in Table 1.

The mean intelligibility scores for the case when early reflections are added to the speech signal are shown in Figure 9. The results mainly indicate a L_{55} threshold shift of similar size as observed in the direct sound only case (Figure 8). Comparing the speech intelligibility scores for the case where the SNR was varied by modifying the direct speech sound level (Figure 8) and the cased where the SNR was changed by modifying the early reflection level (Figure 9), significant differences can be observed. In the latter case the s_{55} slopes are significantly shallower and the L_{55} threshold is increased by about 1 dB (see Table 1). However, this difference is roughly independent of the auralisation technique. Hence, the auralisation technique seems to influence the absolute intelligibility scores, but does not affect the relative change introduced by adding reflections.

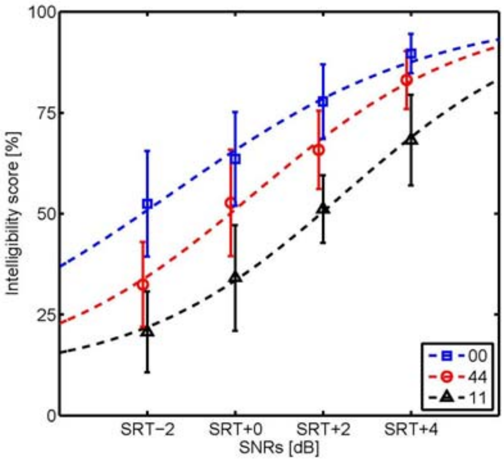

Figure 9: As Figure 8, except that early reflections are added to change the SNR and the direct sound level is held constant.

Although early reflections do significantly contribute to overall speech intelligibility, it is observed here that adding energy to early reflections is significantly less effective than adding the energy to the direct sound. This observation is in conflict with [6], who concluded that adding energy to early reflections produces an equivalent speech intelligibility benefit as adding energy to the direct sound. This difference might be explained by the different speech material and methods used in the two studies.

Summary and conclusions

A loudspeaker-based room auralisation (LoRA) system was introduced, which aims at providing a research tool for investigating auditory (speech) perception in realistic (i.e., reverberant multi sound source) environments. The general applicability of the LoRA processing was verified by an objective evaluation considering different monaural and binaural room acoustic parameters. In order to test if the LoRA inherent Ambisonic auralisation would have an impact on perception research, a speech intelligibility experiment was conducted. In this experiment speech intelligibility scores for Ambisonics auralisation were compared to discrete loudspeaker auralisation. The experiment revealed that (high-order) Ambisonics deteriorates absolute speech intelligibility, but does not affect the shape of the overall psychometric function and does maintain the sensitivity to changes introduced by early reflections. Hence, high-order Ambisonics might be used for speech intelligibility research when overall sensitivity is compensated. However, additional perception experiments are necessary to further support this conclusion. In the case where discrete loudspeaker auralisation can be applied this should be the preferred auralisation method.

References

[1] Rindel, J. H. (2000). "The use of computer modeling in room acoustics." Journal of Vibroengineering **3**, 41-72.

[2] Favrot, S. and Buchholz, J. B. (2009). "LoRA – A loudspeaker-based room auralisation system." Acta Acustica united with Acustica (under revision).

[3] ISO (1997). "Acoustics – measurement of the reverberation time of rooms with reference to other acoustical parameters," ISO3382 – second edition.

[4] Daniel, J., Nicol, R., and Moreau, S. (2003). "Further investigations of high order Ambisonics and wavefield synthesis for holophonic sound imaging," presented at the 114th Audio Eng. Soc. Convention, preprint 5788.

[5] Haas, H. (1951). "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache (On the influence of a single echo on the intelligibility of speech)." Acustica **1**, 49-58.

[6] Bradley, J. S. and Sato, H. (2003). "On the importance of early reflections for speech in rooms." J. Acoust. Soc. Am. **113**, 3233-3244.

[7] Wagener, K., Josvassen, J. L., and Ardenkjaer, R. (2003). "Design, optimization and evaluation of a Danish sentence test in noise." Int. J. of Audiology **42**., 10-17.

[8] Brand, T. and Kollmeier, B. (2002). "Efficient adaptive procedures for threshold and concurrent lope estimates for psychophysics and speech intelligibility tests." J. Acoust. Soc. Am. **111**, 2801-2810.

[9] Favrot, S. and Buchholz, J. B. (2009). "Validation of a loudspeaker-based room auralisation system using speech intelligibility measures." 126th convention of the Audio Eng. Soc., Munich.

[10] Shirley, B., Kendrick, P., and Churchill, C. (2007). "The Effect of Stereo Crosstalk on Intelligibility: Comparison of a Phantom Stereo Image and a Central Loudspeaker Source." J. Audio Eng. Soc. **55**, 852-863.