Measurement-based fuzzy interpolation of room impulse responses

Csaba Huszty, Bottyán Németh, Péter Baranyi and Fülöp Augusztinovicz

Budapest University of Technology and Economics, BME Dept. of Telecommunications,
Magyar tudósok körútja 2, H-1117 Budapest, Hungary
huszty@hit.bme.hu

Application of room impulse responses (RIRs) to acoustic evaluation and auralization often requires many measurements to get enough information about the hall, or to provide enough flexibility for virtual sound source placements in convolution reverberation. In this paper we propose a measurement-based fuzzy modeling method to approximate the RIR function at an arbitrary location between available measured points, without apriori information on the hall geometry or wall reflection parameters. For the fuzzy model identification we define an accuracy indicator of the spatial density of the source positions and predict the required number of them in a selected hall. This indicator quantifies the relationship of the early reflections, determined for various measured positions. This paper also proposes a method that treats non-uniform spatial sampling of the measurement positions, and its implementation for 2D cases is shown. Non-uniform spatial sampling can be useful when RIRs at some source positions - e.g. positions of musicians on a stage of a concert hall - are known or have to be measured precisely, but RIRs at locations in between require an approximation only. The proposed fuzzy model of RIRs actually transforms the measured information into a uniform and tensor product form, enabling the analyst to use further matrix and tensor algebra based numerical methods.

1 Introduction

The room is treated as an infinite-input infinite-output MIMO (or IIIO) bounded-input bounded-output (BIBO) stable causal linear time-invariant (LTI) system. This model is quite difficult to handle and such systems are difficult to identify because of its number of inputs and outputs and lengthy room impulse response (RIR) for each pair of I/O. To measure with least effort, only a few number of inputs and outputs are selected instead of identifying the whole system, and several room acoustic parameters are extracted from the available RIRs. The density of spatial sampling – and other conditions such as time invariance – for an acceptable reliability is given in the [1] standard: it proposes to measure RIRs between 'as many source positions as possible' (> 2) and a given number of listener positions (typically between 6-12 depending on the room size). For applications such as auralization-based on convolution reverberation, users often seek for the availability of moving their sound source positions not only to discrete measured positions, but continuously, therefore some kind of modeling is essential. However, the geometry and surface parameters (such as the complex impedance) is often not known to provide an accurate geometric model. The purpose of this work is to present a method that is entirely based on measurements, capable of approximating the RIR function between measured source positions. We further hope that this method will be later capable of simplifying the identification of the room in a way finite discrete measurement positions are sufficient for building a convex hull on estimated room acoustic parameters. We first present the basic assumptions in Section 2, then discuss the RIR interpolation model in Section 3. Based on the need of local variability, we choose a fuzzy model that we present in Section 4, and a simplified version of the proposed fuzzy model is implemented and evaluated in Section 5. The proposed model can later be extended with non-linear properties as well.

2 Basic assumptions

2.1 Synchronized time support of RIRs

Measurements of the RIRs are based on a Finite Impulse Response (FIR) model. This model is valid, because real-world reverberation is feed-forward and the decay is of finite length (energy conservation). The present assumption of synchronized time support of RIRs states that the end-point of the RIRs are the same within the same room, regardless of the physical locations of the measurement positions, if the room has a perfectly diffuse sound field. This is not difficult to see, because the steepness of the decay slope should be the same everywhere, otherwise the reverberation time would not be the same within the same room. Assuming that this is true, this means that wherever the sources are placed, the reverberation tail will be synchronized for all source positions regardless of the listening position, and the only differences we would see will be at the preceding part, namely the propagation delay, the direct arrival and the early reflection pattern. The differences will be temporal and energetic as well.

2.2 Separability of RIRs

We further assume that RIRs have two separable parts: a deterministic 'early' part and a stochastic 'late' or tail part. These two blocks are following each other respectively with a smooth transition. Finding an exact time limit between the deterministic part and the stochastic part therefore is not possible. However, Hidaka et al [2] proposes a very useful definition for a 'transition time' t_e that can be determined by utilizing short-term correlation analysis on the RIR. They used a fix source position and showed that the transition time is dependent on the receiving position in actual rooms. They found that the t_e is around 225 ms in average in a shoebox-shaped concert hall, varying between 70 and 300 ms. We assume that using a fixed value of 300 ms will be correct for all positions at a concert hall, while in smaller halls we will select smaller, but fixed values for all positions. We propose using the interpolation only at the early part. The tail part will not be interpolated but will be added to the interpolated response later.

2.3 Continuity of RIRs

The assumption of continuity of a RIR states that infinitely small movements of sound sources having a fixed listener will cause a continuous change in the reflection pattern both in the temporal and the energetic properties. A small source movement will result in a different

angle of incidence on a reflective surface, and this difference is growing and accumulating in multiple reflection paths according to the mean free path, but still, the wall impedance of the reflective surface is continuous except for the case of a perfectly rigid wall. Although the surface impedance is not only incidence but frequency-dependent as well and is most likely different at each point of a real surface, we assume that the continuity is likely to be correct at least in the beginning of the early part. Fortunately our main scope of interest is in this part because this is one the most significant part of the reverberation process for a human subject.

2.4 Continuity of excitation positions

We assume that regardless of the source position within a room, all frequencies will be excited in the same manner, so the frequency content of the diffuse field will not depend on the physical location of the sound source. Certain room modes however cannot be excited from certain discrete positions called nodal surfaces. At lower frequencies, standing waves at natural frequencies are dominant in a steady-state sound-field in the room. There is certain frequency limit up to the modal representation of a room is used, because the modes in a real room have a finite non-zero bandwidth due to damping. This bandwidth can be defined as:

$$B_{f,mode} = \frac{6 \cdot \ln 10}{2\pi \cdot T_{60,f}} \qquad (1)$$

The modal density is increasing asymptotically with the square of the frequency for any shape of room [3] – analogously to the density increase of arriving reflections with time –, therefore distance of the maxima of the modes become smaller. This results in the fact that a single-frequency excitation might excite more than a single room mode. The crossover frequency for modal overlap (aka Schröder-frequency) is the frequency limit of the modal approach which is as follows for rectangular rooms:

$$f_{Sch} = \sqrt{\frac{c^3}{4 \ln 10} \frac{T_{60}}{V}} \qquad (2)$$

This concludes that rooms of big size and small reverberation time (such as modern concert halls) have a quite low, while very large rooms with long reverberation times (e.g. cathedrals) may have a higher Schröder-frequency.

Room name	V $[m^3]$	T_{60} $[s]$	f_{Sch} $[Hz]$
Musikvereinsal, Wien	15000	2.0	23.9
KKL Hall, Luzern (closed chambers)	17823	2.05	22.2
Avery Fisher Hall, New York	20400	1.76	19.2
Carnegie Hall, New York	24270	1.79	17.7

Table 1: Modal overlap crossover (Schröder) frequencies of different rooms of shoebox shapes at occupied state. Raw data from [4].

We can conclude from Table 1 that our frequency of interest – that is around 20 Hz to 8 kHz – is very likely to lie in the region of a non-modal representation, therefore all frequencies will be excited at all times, regardless of the source position in the room.

2.5 Detectability of reflections

We assume that we can detect reflections in the RIR. Localizing significant reflections in time-domain is very easy if the propagation medium and the reflections are non-dispersive. A non-disperse response keeps the source spike well localized in time. For a non-disperse propagation, the speed of sound should be frequency-independent and there should be no propagation attenuation. When low-pass filtering occurs, such as in the case of air attenuation, the bandwidth of the reflection will decrease, therefore its temporal support will increase. On the other hand, when a reflective surface has an impedance of such that it can be modelled as a non-linear-phase filter, different frequencies will have different group delays causing a further increase of the time support. Present reflection detection methods all use some kind of localized energy detection methods such as maxima detection or short-term correlation, either concentrating on temporal, frequency or scale-domain properties. One of the simplest method is a windowed local maxima searching method of finding and marking portions of pressure or energy maxima in the interval of interest. Another method might be the short-term correlation based technique we propose using here, but there are other methods as well, such as the cross-wavelet transform (XWT) utilizing the Paul-wavelet, as described in [6].

3 Interpolation of RIRs

We can conclude from our previous assumptions that it is only the early (deterministic) part of the RIR that we can interpolate. We present our approach using a fixed receiver and a large number of different source positions within the same room.

3.1 Spatial interpolation

In common room acoustic measurement situations, source points of interest is measured which are often not on an equidistant grid. The first problem we should overcome is to select appropriate RIRs and their level of contribution that will take part in the interpolation process, given this arbitrary measurement position setup of non-uniform (irregular) spatial sampling. However, interpolation and matrix representations both require a grid of such for simplicity, therefore, a spatial mapping of an irregular grid to a regular grid is required.

3.1.1 2D mapping by using linear interpolation

Constructing a measurement-lattice by connecting real measurement points of an irregular grid will result in having quadrangles and triangles. During the approximation the neighbor points for interpolation will be the corner points of the polygon that is containing the arbitrarily selected point we want to map onto an equidis-

tant grid. The algorithm therefore can be divided into two main parts: we select the appropriate neighbors for position mapping first and then we determine the interpolation factors. To determine the neighbors we take the points of each polygon in the lattice in a given order and for each point we create the vector to our approximation point and to the next point of the polygon and we calculate the vectorial product of them. The rule of selection follows the fact that the sign of the vectorial products shall be the same for all the corner points of the appropriate polygon that contains our arbitrary point.
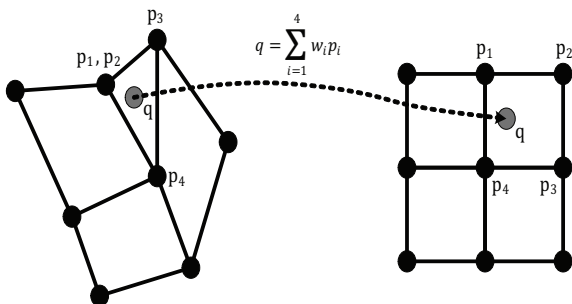

Figure 1: Mapping of an irregular measurement lattice to an equidistant grid.

To form an equidistant grid, the measured points are aligned to an m x n matrix by including a single measurement position for more than one time, if needed. Original neighbors should be neighbors in the matrix as well (See Fig. 1). The equation system for the weights for a quadrangle is:

$$\mathbf{p_0} = w_1 w_2 \mathbf{p_1} + (1 - w_1)w_2 \mathbf{p_2} + (1 - w_1)(1 - w_2)\mathbf{p_3} + w_1(1 - w_2)\mathbf{p_4} \tag{3}$$

There are two unknown weight parameters and two equations for each coordinates, so the solution can be found.

$$w_1 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{4}$$

$$w_2 = \frac{p_{01} - w_1 p_{21} - (1 - w_1)p_{31}}{w_1 p_{11} + (1 - w_1)p_{41} - w_1 p_{21} - (1 - w_1)p_{31}} \tag{5}$$

$$a = (p_{32} - p_{22})(p_{11} - p_{21} + p_{31} - p_{41}) - (p_{31} - p_{21})(p_{12} - p_{21} + p_{31} - p_{41}) \tag{6}$$

$$b = (p_{32} - p_{22})(p_{41} - p_{31}) + (p_{11} - p_{21} + p_{31} - p_{41})(p_{02} - p_{32}) + (p_{12} - p_{22} + p_{32} - p_{42}) \tag{7}$$
$$(p_{31} - p_{01}) - (p_{31} - p_{21})(p_{42} - p_{32})$$

$$c = (p_{02} - p_{32})(p_{41} - p_{31}) - (p_{01} - p_{31})(p_{42} - p_{32}) \tag{8}$$

The same equations can be formed for the case of triangles, when two corner points ($\mathbf{p_3}$ and $\mathbf{p_4}$) are the same on the grid:

$$\mathbf{p_0} = (w_1 w_2 + (1 - w_1)w_2)\mathbf{p_1} + (1 - w_1)(1 - w_2)\mathbf{p_2} + w_1(1 - w_2) \cdot \mathbf{p_3} \tag{9}$$

3.2 Temporal mapping of reflections

Following our previous assumptions, we map significant reflections of the RIRs taking part in the interpolation process by forming pairs. Reflections can be paired if they have a high cross-correlation value, and a similar energy. In this present approach we pair the reflections manually to be able to test other parts of the proposed methods without being affected by the accuracy of temporal mapping. The spatial measurement density can be quantified by examining the change in the early reflection pattern of two neighboring RIRs in 1D. We define the accuracy indicator by measuring the time differences between significant reflections by fitting a 1st-order regression polynomial F_T to the curve $T = T_1(T_2)$ where T_1 contains the arrival times of RIR_1 and T_2 contains RIR_2's. If these are the same, the regression line and the T line are both the same line. To quantify the difference, we evaluate the goodness-of-fit of the regression line to the original data with the R^2 coefficient of determination:

$$R^2 = \frac{\sum_i (F_{T_i} - \overline{T})^2}{\sum_i (T_i - \overline{T})^2} \tag{10}$$

where F_{T_i} is the i-th value of the regression line, $\overline{T}$ is the average of the reflection arrival time values T_i. An R^2 of 1.0 indicates a perfect fit, showing the unrealistic situation of only a constant time-shift between the significant reflections of the RIRs. Therefore there should be a threshold defined in order to find a proper indication. Fig. 2 shows an example in a measured room. An indication of the N_{tot} total required measurements
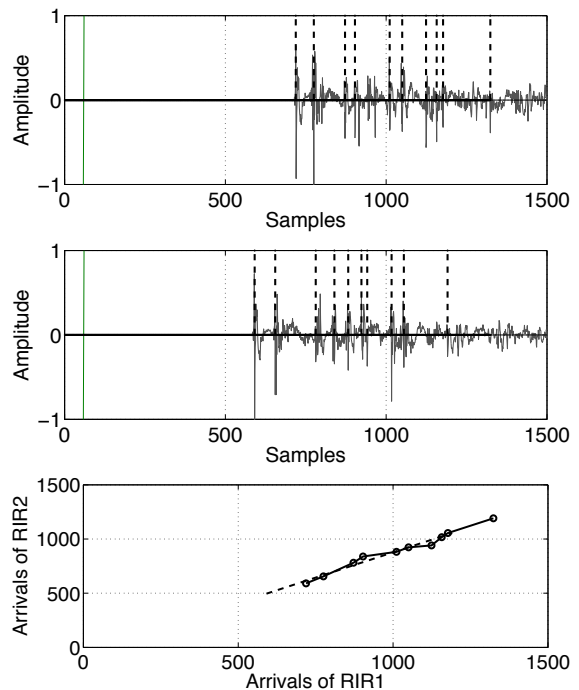

Figure 2: Reflections of RIR1 (up) and RIR2 (middle) showing the RIR function (solid grey line) and the arrival positions (dashed black line). Paired reflections (solid line) and regression (dashed line). Significant reflections were selected manually in this example.

positions for an accurate spatial sampling is possible assuming measuring on a rectangular grid by using the

S surface-of-interest of the room (e.g. the stage) and the d_{min} minimum distance of two measured RIRs that have an lower R^2 value than a given threshold level of desired accuracy:

$$N_{tot} = \frac{4S}{d_{min}^2} \tag{11}$$

3.3 Interpolation of time support

Interpolation of time support practically means time stretching and compression according to a factor that is determined by the fuzzy algorithm. Time stretching can be implemented most easily with resampling, but this affects the frequency response by introduces frequency shifts. This error is low if the time supports of the different RIRs that are taking part in the interpolation process are similar. If they are not, other more complex interpolation methods may be used. Such methods can be frequency preserving time stretching algorithms using either time-domain approaches, analysis and re-synthesis or physical modeling techniques; but to our best knowledge, their applicability on transient signals such as RIRs are not yet discussed. In this present approach, we use resampling instead of other more complex techniques, for simplicity.

3.4 Interpolation of reflections

Following our previous assumptions, we interpolate the reflections as-is without changing its time support at a given time window of the direct sound. Results will be presented in the Evaluation section.

4 Fuzzy model for RIR interpolation

The early part of the RIR has a granular layout in terms of having significant reflections. When these reflections are arriving more frequently as time advances, we reach the stochastic part of the RIR that are not handled in our proposed interpolation approach. Granular layouts are handled more conveniently with soft computing techniques and representations such as neural networks or fuzzy modeling. Granularity also appears because of the fact that we measure at discrete measurement positions. The spatial measurement sampling grid, which is a geometrical grid can be matched well with the fuzzy modeling approach, more conveniently than with neural networks. Therefore we propose using fuzzy modeling.

4.1 Takagi-Sugeno fuzzy system based modelling

A fuzzy model defines a mapping from x to y. This mapping is defined by the linguistic fuzzy rules, such as

IF x_1 is A_{1,i_1} AND x_2 is A_{2,i_2} … AND x_N is A_{N,i_N}

THEN y is $B_{i_1,i_2,\dots,i_N}$

where $A_{n,i}$ is the i-th antecedent set of the n-th input dimension ($i = 1, \dots, I_n$; $n = 1, \dots, N$) given by $\mu_{n,i}(x_n)$ membership function and $B_{i_1,i_2,\dots,i_N}$ is the consequent

set. The fuzzy inference consists of three main steps. First the given value is fuzzified (termed as observation) and then the degree of fitting of the observation and the given rules are computed. The degree of fitting of the rules is used to modified the corresponding consequent fuzzy sets. Finally the modified consequent fuzzy sets are combined and defuzzyificated to yield a scalar value y.

4.2 Simplification for the present approach

In the present approach the fuzzification generates singleton fuzzy sets from the given x value. The antecedent fuzzy sets are defined in Ruspini partition. The consequent sets $B_{i_1,i_2,\dots,i_N}$ are also given by singleton fuzzy sets defined by the location of the corresponding antecedent sets. For defuzzyfication we apply center of gravity (COG) method. The linguistic fuzzy rules are fully specified that means that all combination of the antecedent fuzzy sets are covered by a rule. Having the above simplification the transfer functions of this special Takagi-Sugeno fuzzy model becomes:

$$y = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} \mu_{1,i_1}(x_1)\mu_{2,i_2}(x_2)\dots$$
$$\dots \mu_{N,i_N}(x_N)B_{i_1,i_2,\dots,i_N}. \tag{12}$$

5 Implementation and evaluation

5.1 Correlation-based reflection detection

We detect significant reflections by correlating the received direct sound in each RIR with its forthcoming early reflections (ER). This method has the advantage of using a distant loudspeaker response (the direct sound) with similar travel path lengths and it also includes imperfections of the speaker response. A possible implementation is as follows: first match the spatial measurement matrix to the equidistant grid, then find the RIRs to be used for interpolation and take the early (ER) part of them. For each RIR find the significant reflections by cross-correlating the RIR's direct sound with the ER part by using different time lags yielding the SXCF matrix. Reduce its dimensions by finding the maxima across the time lags, then locate the significant reflections and match the reflections of different RIRs.

5.2 Unchanged support of the direct sound

To visualize the direct sound change with different source-receiver distances, we present an example on Fig. 3 measured by one of the authors at a large concert hall in Budapest, Hungary. It can be seen that there is a smoothing effect with distance and only a few samples of time-domain smearing, if there is any at all. Therefore we will not compensate for the time support extension in the correlation analysis for reflection detections.

5.3 Interpolation in a rectangular office room

We measured a simple rectangular office room (with furniture) at a fixed listener position for 14 closely spaced
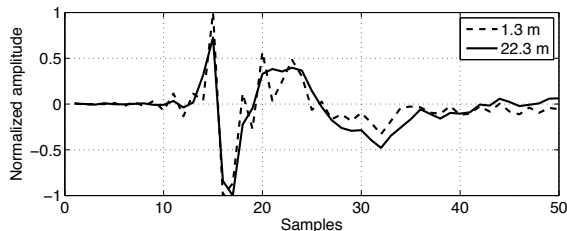
Figure 3: Temporal view of the direct sound of two measured RIRs at a large concert hall (source-receiver distance is 1.3 m and at 22.3 m with a fixed source position). Measurement speaker was Genelec 8050A.

sources. We believe that a small room is a very strict test environment as the reflections are coming very dense. We have found that the length of the direct sound of the speaker we were using was 14 samples at $f_s = 48$ kHz, so the sample length of non-overlapping reflections is equivalent to a 0.1 m distance assuming $c = 343 m/s$. If reflections are coming from two paths closer than this length, we cannot distinguish them. To visualize the accuracy of a linear interpolation we selected significant reflections manually and then interpolated manually matched reflections on the length of the direct sound. We have found that the $k = 0.5$ parameter did not provide an acceptable time support interpolation for the direct arrival (the difference was 10 samples $< 0.1m$) so we had to choose $k = 0.656$ instead, which might be either due to measurement positioning inaccuracy or due to the fact that linear interpolation is not sufficiently complex enough. Anyway, it can be seen that most of the reflections produce an acceptable interpolated result. However, further testing is needed for more positions and halls.
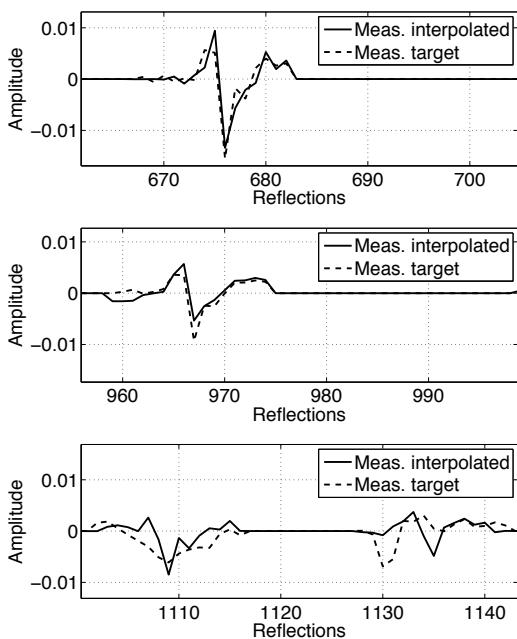

Figure 4: Linear interpolation of reflections of two measured RIRs compared to the measured target RIR of a rectangular office room. Interpolation factor was $k = 0.656$. Top: direct sound, middle: 4th reflection, bottom: 6th and 7th reflection (see also Fig. 2).
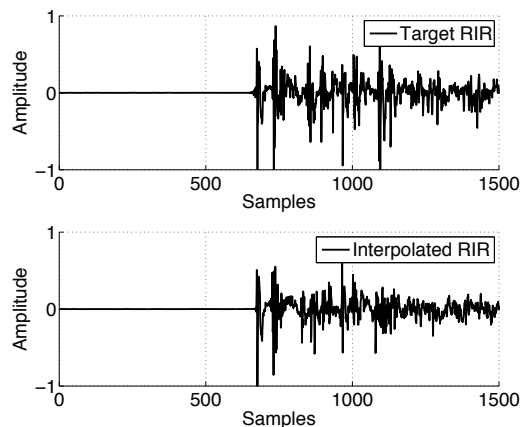

Figure 5: Linear interpolation of two measured RIRs compared to the measured target.

6 Conclusion

We proposed a method of interpolating room impulse responses with a fuzzy model following assumptions for measurement-based modeling. From the modeling we proposed new parameters for measuring the density of spatial sampling in a room and presented a 2-dimensional mapping method for transforming irregular measurement grids to a regular lattice. We presented and tested a simplified implementation of the proposed method by using linear interpolation with a linearized Takagi-Sugeno fuzzy machine. The presented approach relies on significant reflection detection so it is essential to have an accurate detection and mapping algorithm. We also found that the linearized simplification might not be accurate enough so further work should focus on these topics.

References

[1] ISO/FDIS 3382, "Acoustics: Measurement of the Reverberation Time of rooms with reference to other acoustical parameters", *International Organization for Standardization*, (1997)

[2] T. Hidaka, Y. Yamada, and T. Nakagawa, "A new definition of boundary point between early reflections and late reverberation in room impulse responses", *J. Acoust. Soc. Am., Vol. 122, No. 1, pp. 326–332*, (2007)

[3] H. Kuttruff, "Room Acoustics", *4th ed., E & FN Spon, London*, (2000)

[4] L. Beranek, "Concert Halls and Opera Houses: Music, Acoustics, and Architecture", *Second Edition. Springer Verlag, ISBN 0-387-95524-0*, (2004)

[5] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control", *IEEE Transactions on Systems, Man and Cybernetics, 15:116-132*, (1985)

[6] S. Vesa, T. Lokki, "Detection of room reflections from a binaural room impulse response", *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada*, (2006)