Investigation of acoustic rake receivers in a realistic test scenario

Markus Müller-Trapet[1], Jordan Cheer[1] and Filippo Maria Fazi[1]

[1] *Institute of Sound and Vibration Research (ISVR), University of Southampton Southampton S017 1BJ, United Kingdom*

Email: M.F.Muller-Trapet@soton.ac.uk

Introduction

Acoustic beamforming with microphone arrays is an established tool for source localization and quantification. It is widely used for industrial applications such as automotive [1] or aircraft noise but also for improving speech intelligibility [2]. In the latter case, the microphone array is typically installed in a room. Under such conditions, strong room reflections can make localization and especially quantification very difficult.

An intuitive approach to the problem of room reflections is a matched-filter approach, i. e. using the room transfer function (RTF) from the (estimated) source location to the array microphones in the beamforming processing [3]. However, the RTF has to be either measured, simulated or estimated. Since measurements are not always possible or feasible (especially if many scan positions are needed for the beamformer), the only practical options are simulations or estimations, both of which are likely to be inaccurate and often give a worse result compared to conventional (free-field) beamforming.

Recently, the so-called Acoustic Rake Receiver (ARR) has been introduced by Dokmanić et al. [4]. Instead of estimating the entire RTF, only the specular reflections modeled by image sources (IS) are taken into account. This has the advantage that the complexity can be controlled, as the number of reflections to include can be chosen depending on the actual situation. In [4] the authors have presented 2D simulations of the ARR and found good performance.

In this paper, the performance of an ARR is investigated through both simulations and measurements in a realistic but relatively simple test setup. The performance is evaluated in terms of both source localization as well as quantification, i. e. level estimation.

Theoretical Background

The conventional Delay-and-Sum (DAS) beamformer for a microphone array with sensors at $\mathbf{r}_{\mathrm{mic}}$ and scan points at $\mathbf{r}_{\mathrm{scan}}$ can be expressed by the following equation:

$$\mathbf{B} = 4\pi \frac{\mathbf{w}^H(\mathbf{r}_{\mathrm{mic}}, \mathbf{r}_{\mathrm{scan}})}{\left\|\mathbf{w}(\mathbf{r}_{\mathrm{mic}}, \mathbf{r}_{\mathrm{scan}})\right\|^2} \mathbf{p}(\mathbf{r}_{\mathrm{mic}}) \tag{1}$$

where $\{.\}^H$ denotes the complex conjugate transpose, $\mathbf{p}$ is the vector of sound pressures at the microphones and $\mathbf{w}$ is the matrix of beamforming weights depending on the location of the microphones and scan positions. Typically, the weights in Eq. (1) are calculated with a scaled version of the free-field Green's function, so that a single element of the matrix $\mathbf{w}$ for the m-th microphone and n-th scanning point can be expressed as

$$\begin{aligned} \mathbf{w}(\mathbf{r}_{\mathrm{mic},m}, \mathbf{r}_{\mathrm{scan},n}) &= \mathbf{G}(\mathbf{r}_{\mathrm{mic},m}, \mathbf{r}_{\mathrm{scan},n}) \\ &= \frac{\mathrm{e}^{-\mathrm{j}k\|\mathbf{r}_{\mathrm{scan},n} - \mathbf{r}_{\mathrm{mic},m}\|}}{\|\mathbf{r}_{\mathrm{scan},n} - \mathbf{r}_{\mathrm{mic},m}\|} \end{aligned} \tag{2}$$

In the case of the ARR, the beamforming weights for the scan point at $\mathbf{r}_{\mathrm{scan},n}$ are calculated as a superposition of the free-field weights in Eq. (2) and the contributions by the corresponding image sources at $\mathbf{r}_{\mathrm{scan},n,IS}$:

$$\begin{aligned} \mathbf{w}_{ARR}(\mathbf{r}_{\mathrm{mic},m}, \mathbf{r}_{\mathrm{scan},n}) = \; & \mathbf{G}(\mathbf{r}_{\mathrm{mic},m}, \mathbf{r}_{\mathrm{scan},n}) \\ & + \sum_{i=1}^{I} \mathbf{G}(\mathbf{r}_{\mathrm{mic}.m}, \mathbf{r}_{\mathrm{scan}.n,IS(i)}) \end{aligned} \tag{3}$$

where the position of the image source is obtained by the method described by Borish [5]. In order to improve the level reconstruction capabilities, the image source contributions can be weighted by the reflection factor of the walls. This can be either done with an estimated value or based on the value calculated from an actual measurement of the reverberation time. The impact of the used reflection factor is also investigated in this paper.

Array Simulations and Measurements in the ISVR AudioLab

As a simple and realistic test environment, the ISVR AudioLab was chosen as it is a rectangular room, which is easy to model. The dimensions of the room are $4.85\,\mathrm{m} \times 4\,\mathrm{m} \times 2.6\,\mathrm{m}$. Two source positions were used: position 1 at $(x, y, z) = (0.8\,\mathrm{m}, 1.6\,\mathrm{m}, 0.9\,\mathrm{m})$ and position 2 at $(x, y, z) = (2.5\,\mathrm{m}, 2.6\,\mathrm{m}, 0.9\,\mathrm{m})$. The 36-channel microphone array was placed on one of the walls with its center at $(x, y, z) = (1.83\,\mathrm{m}, 0.0\,\mathrm{m}, 0.9\,\mathrm{m})$. The setup together with the first-order image sources for source position 1 is depicted in Figure 1.

The average measured reverberation time of the laboratory room is 0.15 s with a room volume of $50\,\mathrm{m}^3$, leading to a Schroeder frequency of 110 Hz. The loudspeaker used for all measurements was a Mackie HR824 two-way studio monitor. The frequency response of the loudspeaker and the array microphones was compensated for in the measurements to obtain results comparable to the simulations.

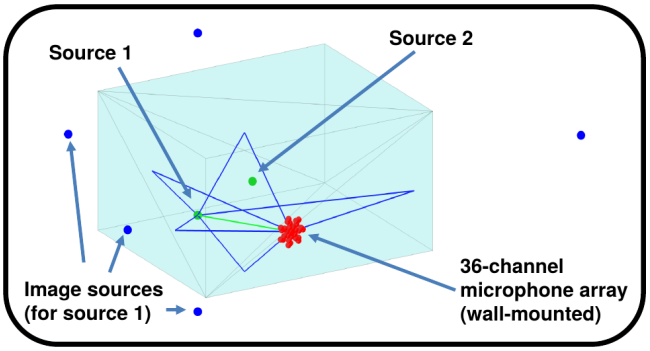


Figure 1: Setup of microphone array (red) and sources (green) in the ISVR AudioLab with first-order image sources (blue)

Simulations of the RTFs were carried out with the image-source method [6] using the reflection factors R determined from the measured reverberation times RT with Eyring's formula:

$$R = 1 - \mathrm{e}^{-\frac{4V}{S}\cdot\frac{6\log(10)}{c\,RT}} \tag{4}$$

where S and V are the room surface area in m^2 and room volume in m^3, respectively, and c is the speed of sound in m/s.

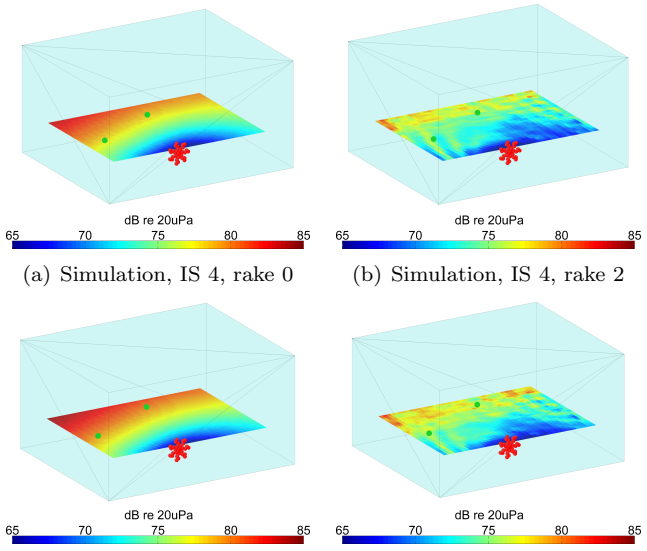
The advantage of using simulations is that the order of reflections can be varied and this was done between order 0 (i. e. free-field) and 4. The source signal used for both simulations and measurements was broadband pink noise with an effective Signal-to-Interference-plus-Noise Ratio (SINR) between source 1 and 2 of 10 dB.

Beamforming was then performed and evaluated at a grid of scan points at the same height of $z = 0.9\,\mathrm{m}$ as the sources. The order of raking was varied between 0 (no raking) and 4, taking into account up to 84 image sources. All results were evaluated in third-octave frequency bands. Both the level maps for certain frequency bands as well as the estimated source level across frequency for source 1 have been evaluated to assess the performance of the ARR.

Results

In order to assess the improvement in localization performance, the beamforming level maps have been evaluated, determining whether "hot spots" could be used to find the two sources.

Figure 2 shows the level maps with a dynamic range of 20 dB for the 500 Hz third-octave band for simulations with a maximum image source order 4 ((a) and (b)) and for the measurements in the corresponding positions ((c) and (d)). The results are presented for conventional beamforming ((a) and (c)) and the raking approach for order 2 ((b) and (d)).



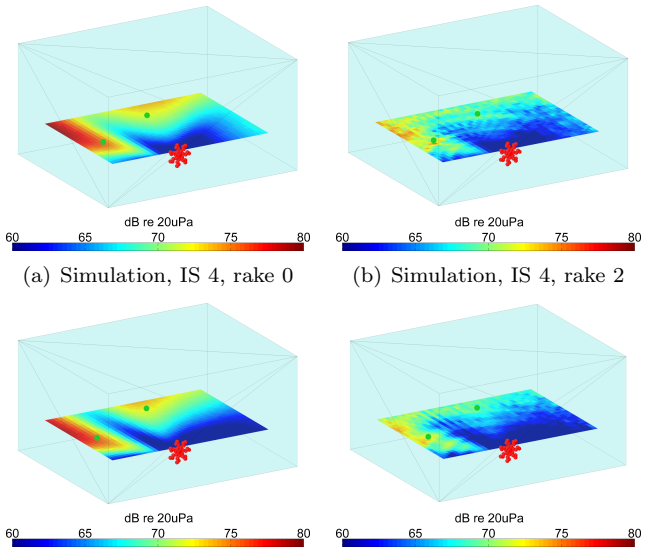
(a) Simulation, IS 4, rake 0 (b) Simulation, IS 4, rake 2

(c) Measurement, rake 0 (d) Measurement, rake 2

Figure 2: Beamforming results (500 Hz third-octave band)

At 500 Hz, the array configuration cannot produce a high spatial selectivity with conventional beamforming (Figure 2(a) and Figure 2(c)), which at first sight can be improved by using the rake beamformer at order 2, taking into account 18 image sources (Figure 2(b) and Figure 2(d)). However, the apparent increased resolution does not necessarily yield better localization, as high levels are also detected at some points closer to the boundaries.

Similar observations can be made for the level maps for the 2000 Hz third-octave band in Figure 3, except that there is a higher source separation in the lateral direction at this frequency even without raking (Figure 3(a) and Figure 3(c)).



(a) Simulation, IS 4, rake 0 (b) Simulation, IS 4, rake 2

(c) Measurement, rake 0 (d) Measurement, rake 2

Figure 3: Beamforming results (2000 Hz third-octave band)

When employing rake beamforming (Figure 3(b) and Figure 3(d)) the depth resolution can be significantly improved as for the data in Figure 2. Nonetheless, large levels are found further away from the sources, especially close to the boundaries.

There is very good agreement between the measured and simulated data, suggesting that the rake approach can be successfully applied in practice. This will be especially true at low frequencies, where an imprecise knowledge of the location of the image sources has a smaller effect than at high frequencies.

Although it was found that the localization accuracy was not improved, the rake approach can be used to achieve a better level estimate than with conventional beamforming after the source positions have been obtained with a different algorithm. To this end, the original level of source 1 has been compared to the reconstructed level of the beamformer depending on the raking order.

In Figure 4, the beamforming output for source 1 is presented for the simulation with a maximum image source order of 4 for raking orders of 0, 1 and 4 together with the original source level. It can be seen that for conventional beamforming (rake order 0) the room reflections have a significant influence on the estimated level up to approximately 1000 Hz, i. e. in the range where the wall reflection factors are higher. Especially between $400 - 800$ Hz, the level difference is up to 3 dB.

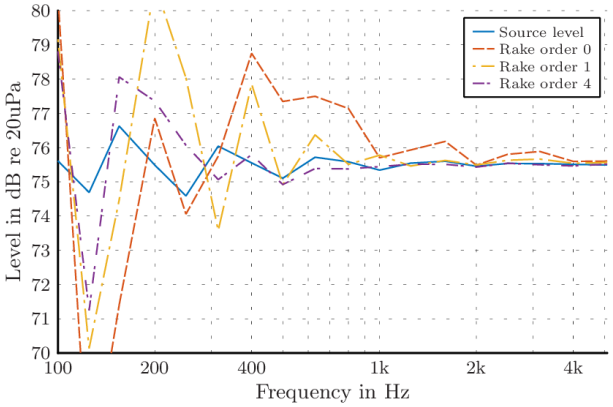
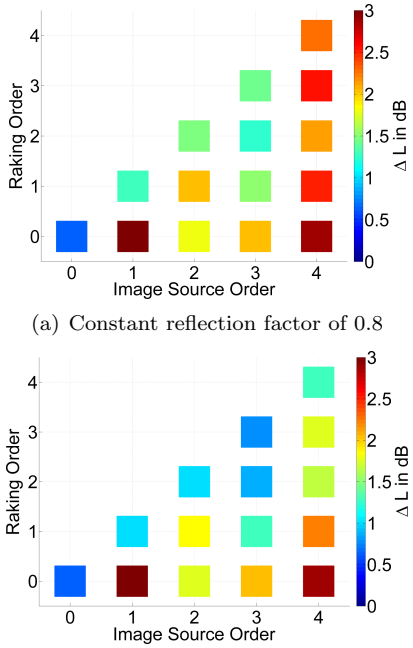


Figure 4: Original and estimated levels for source 1 for IS order 4 and different rake orders

For a rake order of 1 — taking into account 5 image sources — the estimated level is closer to the true result over most of the frequency range, although a large deviation can be found at 200 Hz. With increasing rake order, the beamforming output provides a better estimate of the actual source level.

In order to get a general impression of the performance with respect to the level estimate, the average absolute error between the correct result and the beamforming output has been calculated between $100 - 2000$ Hz as a function of the image source order in the simulation and the rake order of the beamformer. The results are shown in Figure 5.



(a) Constant reflection factor of 0.8

(b) Reflection factor from measured RT

Figure 5: Average absolute error of the level estimation in dB as a function of image source order and raking order

The data in Figure 5(a) has been calculated by weighting the image source contributions in Eq. (3) with an estimated constant reflection factor of 0.8, representing the case that no knowledge about the acoustical properties of the room is available. The results show that especially for higher reflection orders the use of the rake approach does not lead to a significant improvement of the estimated source level, yielding average errors in the range of $1.5 - 2$ dB.

In comparison, using reflection factors calculated from a measurement of the reverberation time of the room gives better results (Figure 5(b)). For almost all combinations of reflection order and raking order, the average level error can be reduced by about 0.5 dB, leaving a maximum error of $1 - 1.5$ dB. This confirms the conclusions from the data in Figure 4.

Discussion

In this paper, the Acoustic Rake Receiver approach has been investigated through a series of simulations and measurements in a simple but realistic environment. A 36-channel microphone array was used to employ beamforming in order to localize and quantify noise sources.

By inspecting beamforming level maps, it was shown that the raking approach does give seemingly better spatial selectivity, especially with respect to depth. However, an actual improvement of the source localization accuracy was not found as large levels could be found at positions away from the actual sources. This suggests that the effect of destructive interference cannot be exploited as robustly as in conventional beamforming when including information about room reflections.

In terms of source level estimation with the raking approach, it was indeed found that on average a better estimate was achieved, especially at low frequencies. Best results were obtained for the case when additional information about the room acoustic properties was included in the beamforming weights, effectively providing a slightly better estimate of the early part of the room impulse response.

It can thus be concluded that a precise knowledge of the environment is crucial when employing Acoustic Rake Receivers. Apart from the geometry of the room needed for estimating the image source locations, information about the reverberation time is helpful and leads to better results. If more precise data for the distribution of reflection factors in the room is available, this could yield an even better estimate.

The measured data agrees very well with the simulations, which seems to indicate that the concept of the rake receiver also works in practice. However, the study presented here has been carried out in an environment with a short reverberation time. For a more thorough evaluation of the performance of the rake approach, an investigation in more reflective rooms should be carried out.

Acknowledgments

The authors would like to thank Fabio Casagrande Hirono for supplying the microphone array. All measurements and part of the post-processing were carried out with the ITA-Toolbox for MATLAB (`http://ita-toolbox.org`).

References

[1] S. Guidati. Source Localisation with Beamforming in Vehicle Acoustics. In *Meeting of the Dutch Acoustical Society (NAG) and 35th German Annual Conference on Acoustics (DAGA), Rotterdam, The Netherlands*, 2009.

[2] Y. Peled and B. Rafaely. Study of speech intelligibility in noisy enclosures using optimal spherical beamforming. In *IEEE 25th Convention of Electrical and Electronics Engineers in Israel.*, pages 285–289, Dec 2008.

[3] E. Jan, P. Svaizer, and J. L. Flanagan. Matched-filter processing of microphone array for spatial volume selectivity. In *IEEE International Symposium on Circuits and Systems*, volume 2, pages 1460–1463 vol.2, Apr 1995.

[4] I. Dokmanić, R. Scheibler, and M. Vetterli. Raking the cocktail party. *IEEE Journal of Selected Topics in Signal Processing*, 9(5):825–836, Aug 2015.

[5] Jeffrey Borish. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827–1836, 1984.

[6] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979.