Room Boundary Estimation from Acoustic Room Impulse Responses

Luca Remaggi, Philip J. B. Jackson, Philip Coleman and Wenwu Wang
Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, UK

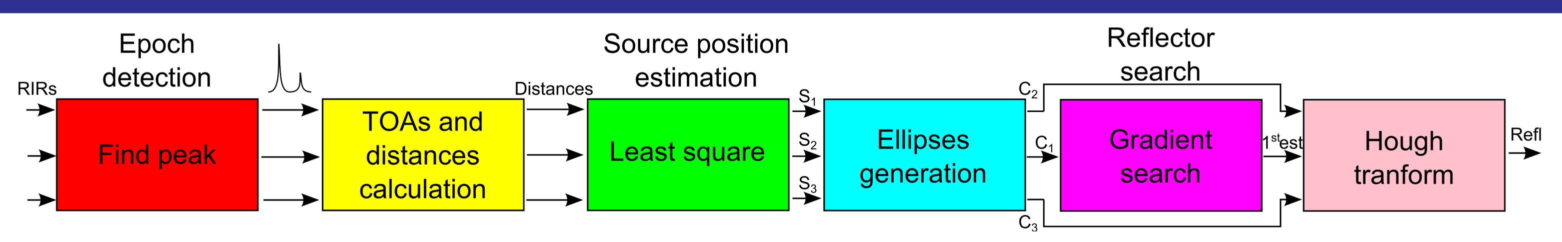
Abstract

Boundary estimation from an acoustic room impulse response (RIR), exploiting known sound propagation behaviour, yields useful information for various applications: e.g., source separation in reverberation, simultaneous localisation and mapping, and spatial audio. The baseline method uses reflection times of arrival (TOAs) to hypothesize reflector ellipses. Here, we modify the algorithm for 3-D environments and for enhanced noise robustness: DYPSA is proposed for epoch detection, direction of arrival (DOA) via MUSIC combined for source localization, and numerical search adopted for reflector estimation. Both methods, and other variants, are tested on RIR data measured; the proposed method performs best.

Introduction

- Purpose: to estimate the room geometry knowing its RIRs
- Results are useful for source localization, source separation and surveillance
- In [1] the authors present a new method based on the creation of ellipses to find the reflector position
- DYPSA [2] is proposed here to improve the robust TOA detection
- MUSIC [3] is exploited to find DOAs for improving the source localization

Baseline method

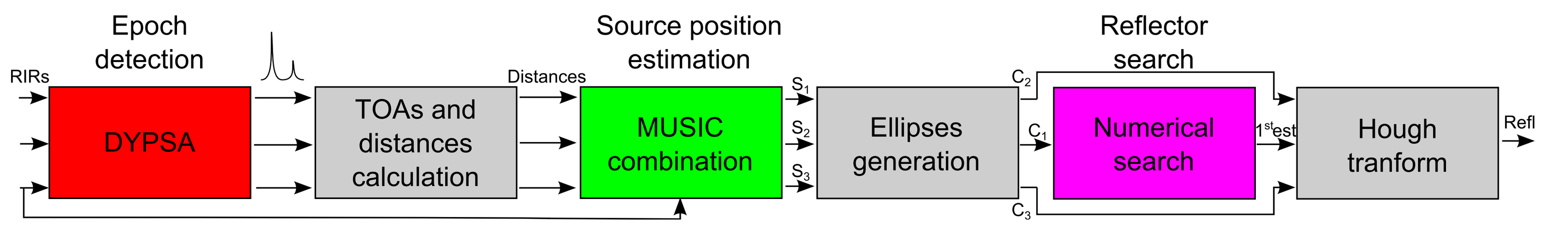


- RIRs peaks detected using the "Find Peak" algorithm
- Distances calculated knowing the sound speed and the TOA relative to the i-th microphone and k-th reflector
- The sources position are estimated using the "Least square" (LS) algorithm
- Ellipses are generated for each microphone-source couple
- Finding the line that minimize the gradient of the cost function

$$J(l) = \sum_{i=1}^{M} |l^T C_i^* l|$$

- Hough transform used to refine the results

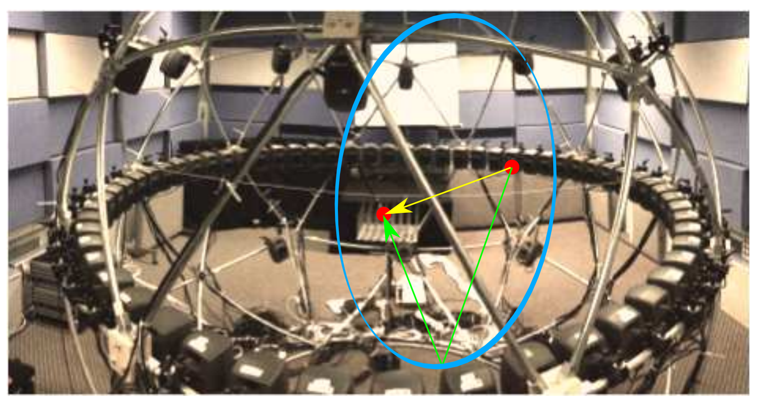
Proposed method



- RIRs peaks detected using DYPSA
- The sources position are estimated combining the distances and DOAs estimated through MUSIC
- Numerical search of the line which minimize the cost function

Surrey Studio-2 RIR Data-set

- Data-set: http://www.cvssp.org/soundzone/resource/Studio2RIR



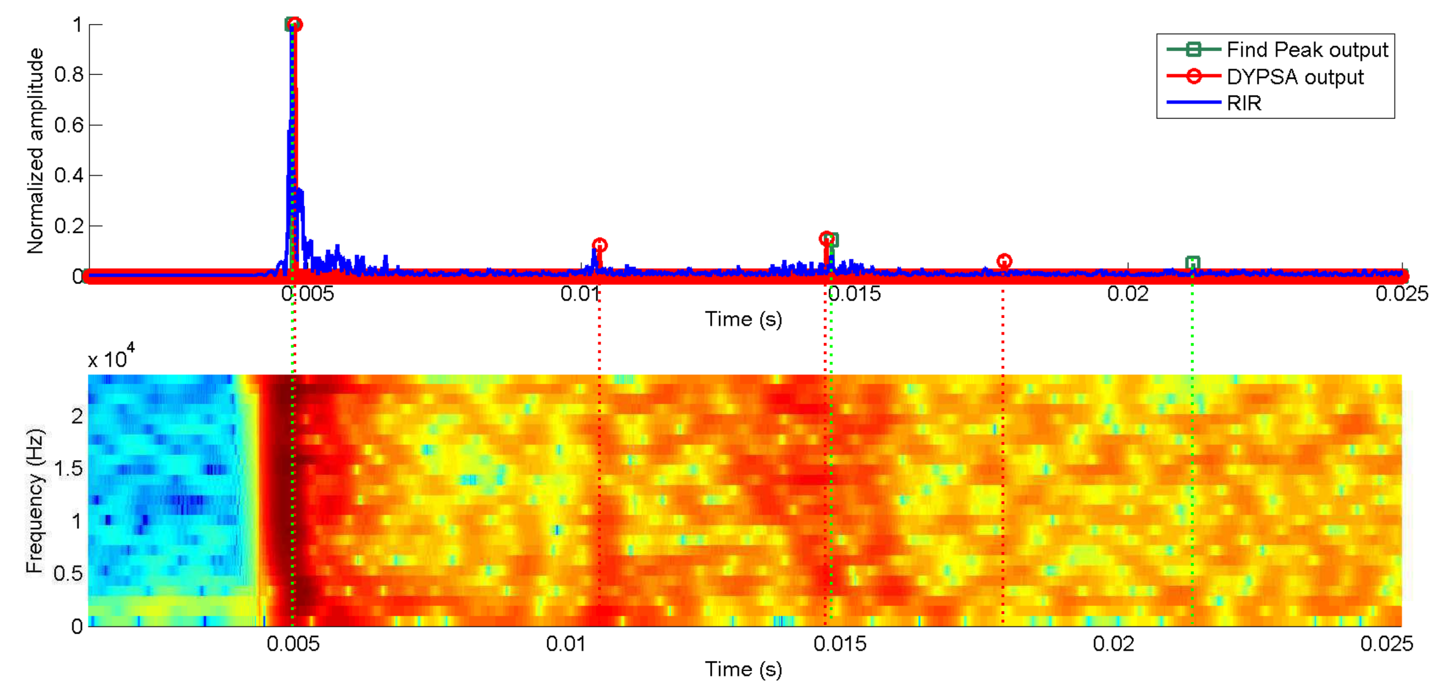
- 336 RIRs evaluated
- Sphere radius: 1.68m
- Studio-2 dimensions: 6.55×8.78×4.02m
- RT60: 235ms

References

[1] F. Antonacci et al., "Inference of room geometry from acoustic impulse response", 2012.
[2] P. A. Naylor et al., "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm", 2007.
[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", 1986

Acknowledgments:
This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/K014307/1 and the MOD University Defence Research Collaboration in Signal Processing. Thanks to Dr. J. Gudnason of Reykjavik University for the help given with DYPSA.

DYPSA outcomes and "Find Peak" comparison

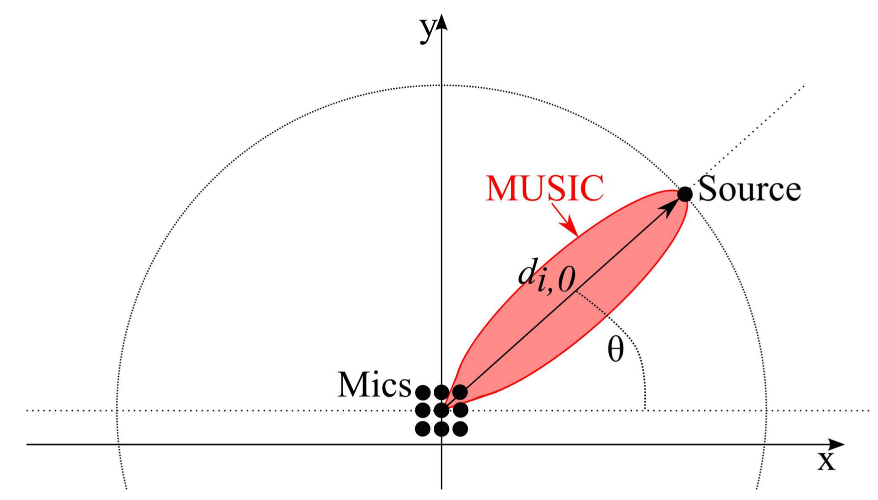


- The first picture compares DYPSA and "Find Peak" outputs in time
- The second picture shows the spectrogram of the recorded RIR
- In table are reported errors for a given RIR used as ground-truth

Table: Errors (mm (ms)) calculated for the "Find peak" and DYPSA.

	Direct sound	1^{st} refl.	2^{nd} refl.
DYPSA	**6.8 (0.020)**	**2.8 (0.008)**	**148.5 (0.433)**
Find peak	10.9 (0.032)	19.6 (0.057)	691.3 (2.000)

MUSIC outcomes and comparison with other algorithms

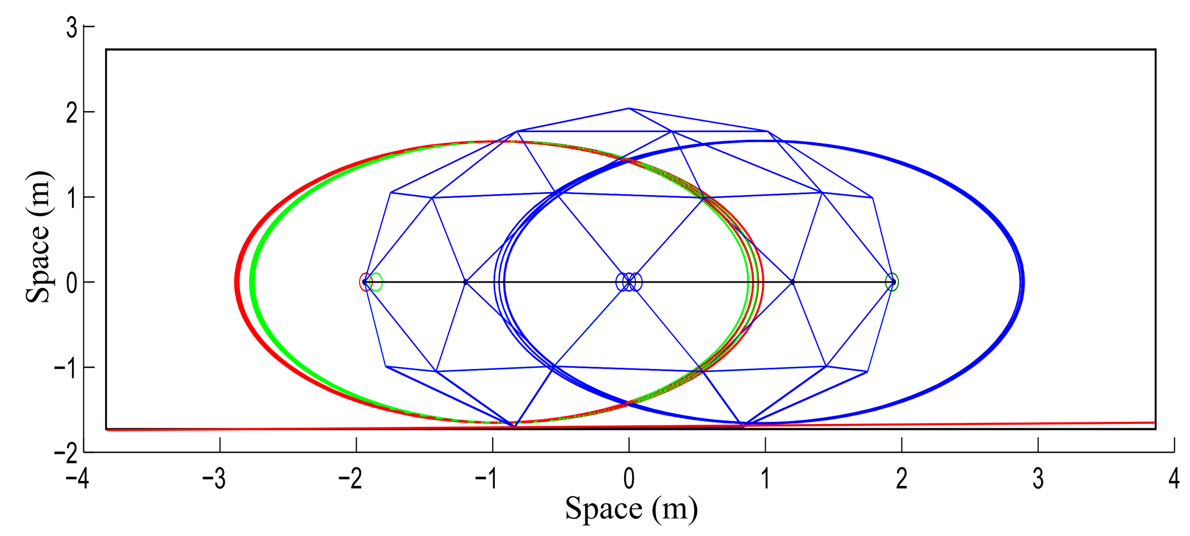


- The figure shows the MUSIC beam-pattern and the distance calculated
- The table shows errors from 3 different algorithms averaged over 8 sources

Table: Errors (Deg) calculated using MUSIC, Capon and Bartlett.

	MUSIC algorithm	Bartlett algorithm	Capon algorithm
Errors	**1.3**	2.6	3.0

RMSEs from the different configurations



- In figure the ellipses generated, the line obtained and a schematic reproduction of the sphere are reported
- For the RMSE in table, 1950 microphone-source combinations are used and the results averaged over 10000 trials.

Table: RMSEs (mm) for the simulated configurations.

	Baseline [1]	Hybrid 1	Hybrid 2	Hybrid 3	Proposed
RMSE	117.0	113.2	67.4	59.9	**18.9**

Conclusion

- RMSE is calculated considering the floor estimation
- RMSE shows that the combination of DYPSA and MUSIC to estimate the source position introduces most of improvements
- The proposed method is strong to noise (can be used with real RIRs)
- The proposed method can be applied to 3-D environments
- Further work to apply a full 3-D implementation
- Other RIRs data-set will be tested