Audio Engineering Society

Convention Paper

Presented at the 131st Convention
2011 October 20–23 New York, NY, USA

The SCENIC Project: Space-Time Audio Processing for Environment-Aware Acoustic Sensing and Rendering

P. Annibale[2], F. Antonacci[1], P. Bestagini[1], A. Brutti[4], A. Canclini[1], L. Cristoforetti[4], E.A.P. Habets[2,3], J. Filos[3], W. Kellermann[2], K. Kowalczyk[2], A. Lombard[2], E. Mabande[2], D. Markovic[1], P.A. Naylor[3], M. Omologo[4], R. Rabenstein[2], A. Sarti[1]*, P. Svaizer[4], M.R.P. Thomas[3]

[1]DEI – Politecnico di Milano, Piazza L. Da Vinci 32, I-20133 Milano, Italy
antonacc/bestagini/canclini/dmarkovic/sarti@elet.polimi.it

[2]LNT – Univ. of Erlangen, Cauerstraße 7, D-91058 Erlangen, Germany [3]
annibale/wk/kowalczyk/lombard/mabande/rabe@lnt.de, e.habets@ieee.org

[3]DEEE - Imperial College London, Exhibition Road, London, UK
jason.filos03/p.naylor/mark.r.thomas02@imperial.ac.uk

[4]Fondazione Bruno Kessler, Povo, Via Sommarive 18, I-38100 Trento, Italy
brutti/cristofo/omologo/svaizer@fbk.eu

ABSTRACT

SCENIC is an EC-funded project aimed at developing a harmonized corpus of methodologies for environment-aware acoustic sensing and rendering. The project focusses on space-time acoustic processing solutions that do not just accommodate the environment in the modeling process but that make the environment help towards achieving the goal at hand. The solutions developed within this project cover a wide range of applications, including acoustic self-calibration, aimed at estimating the parameters of the acoustic system; environment inference, aimed at identifying and characterizing all the relevant acoustic reflectors in the environment. The information gathered through such steps is then used to boost the performance of wavefield rendering methods as well as source localization/characterization/extraction in reverberant environments.

1. INTRODUCTION AND MOTIVATIONS

In any sound analysis and rendering application the environment plays a major role, whether desired or undesired. Before reaching our senses, acoustic wavefronts propagate through the environment in a very complex and hard-to-predict fashion, as they bounce off, pass through, or are scattered by numerous reflectors and obstacles a multitude of times. The sounds that are received do not just depend on the geometry and the physics of the environment but also on the position and the characteristics of the acoustic sources and the receivers. In addition, the environment's impact tends to change over time and becomes more critical as the application requirements become more demanding. This is particularly true with space-time processing of acoustic signals acquired with microphone arrays or rendered by speaker arrays.

Array processing is a hot topic for the audio research community. Microphone arrays allow us to localize acoustic sources; track them as they move; separate them from each other and from reverberation and noise; or even infer their radiation pattern. Similarly, there is a great deal of frenzy around multi-channel and speaker array research, aimed at altering the directivity pattern of sound rendering systems; controlling the perceived spatial location of the source and its radiation pattern; or altering the acoustic response of the environment. Reverberation, however, always makes these goals very hard to achieve. In order to face the challenges posed by the environment, research has mostly focused on developing processing solutions that are robust to reverberation. Looking at the environment as a liability, however, can only go so far. If we are interested in overcoming the main challenges problems that advanced space-time audio processing is facing today, we cannot treat the environment as an enemy to defend against; we need to strike an alliance with it. SCENIC is an EC-funded project whose goal is to do exactly that: bring the environment into the acoustic design of space-time audio project and make it work in our advantage.

Making the best out of the environment's acoustics means understanding propagation in complex enclosures. This is a formidable task that has been the focus of a great deal of research in the past decades, which has produced a plethora of ad-hoc solutions that only seldom are compared with each other, let alone jointly used. The existing approaches to this problem can be roughly divided into

- global approaches, which model the whole wavefield
- local approaches, which model the acoustic channel that links source to receiver (point-to-point approach).

Among the global modelling methods, again we recognize two leading classes of approaches: those that model the acoustic wavefield and its interaction with the enclosure (environment), and those that focus on paths of acoustic propagation (rays). The wavefield approach is based on a space-time discretization of the equations that govern propagation, therefore issues of spatial sampling become relevant. Geometric methods, on the other hand, implement the general solution of the acoustic propagation's equations, as they focus on waves that propagate along acoustic paths and their interaction with the environment. One interesting feature of the geometric approach is that the environment is no longer seen just as a set of complex boundary conditions but it is separately modeled to account for wavefront scattering.

Most space-time processing applications, however, are based on a local approach. Source localization, tracking and separation methods that are designed to withstand a certain degree of reverberation, in fact, are based on channel estimation, where the channel is the point-to-point description that we need to retrieve in order for algorithms to be able to survive the environmental impact. It is in this complex scenario that the SCENIC Project is operating, by developing a corpus of theoretical results that bring together wavefield methodologies, geometric solutions and channel-based methods in a harmonized fashion, with the goal of making the environment part of the global design of space-time audio processing systems.

2. A GEOMETRIC PERSPECTIVE

There are many space-time audio processing tasks that can take advantage of a better understanding of the acoustic propagation in complex enclosures. Let us consider, for example, the task of localizing acoustic sources using microphone arrays. If we consider the "dry sound" as a source of information and the reverberation as a source of disturbance, we can expect the localization accuracy to worsen as the reverberation increases. Knowing the geometry of the environment can be used for reversing this trend. A rather straightforward solution, for example, is to track the locations of multiple virtual (wall-reflected) sources, match them and group them together with the corresponding real source, and use the geometric information on the acoustic paths not just for reducing the localization's uncertainty or for better source separation, but also for doing things that the localization algorithm could not do without the help of the environment, e.g. overcoming the limits of "visibility" (e.g. listening "behind the corner" through "mirror-mediated visibility"). One alternate way to interpret this new potentiality consists of thinking of the surrounding walls as acoustic mirrors onto which each sensor reflects according to the rules of specular reflection. The sensor arrays can thus join forces with the image arrays to form a virtually expanded sensing system. Symmetrically, a rendering system based on an array of speakers can exploit the environment to expand through its mirror images, in order to enable wavefield synthesis in reverberant environments.

2.1. Geometric wavefield analysis

In order for our algorithms to be able to take advantage of the environment, we need to know its geometry and its "radiometric" (reflective) properties. The process of extracting this information is called "acoustic scene reconstruction" and, in the SCENIC project, we show that this can be performed in a fully acoustic fashion, using "unstructured" (natural) acoustic stimuli (e.g. a voice, finger snapping, some abrupt localized noise).

In the area of computer vision, 3D scene reconstruction relies on a corpus of image-based processing solutions based on projective geometry (a nice summary of this approach is presented in [1]). The typical approach consists of extracting image features (corners, edges, etc.) from different views of the same scene, and determining correspondences between them. Each pair of corresponding features determines a projective constraint. Projective invariants are used for stratifying the underlying geometry and the constraints are used for nailing down each geometric layer one by one. Information is progressively incorporated in order to first perform self-calibration (aimed at determining the mutual positioning of the photo/video-cameras that are acquiring the images), and then determining the geometry and the reflectivity of the scene surfaces. A thorough understanding of advanced projective geometry and of the theory of projective invariants has generated a plethora of new applications ranging from shape from motion, to camera ego-motion reconstruction from video, to 3D modelling from unconstrained video footage, which are commonly used today, for example, for the manufacturing of advanced special effects in film production.

Today the SCENIC project is attempting to do in geometrical acoustics what 3D vision did two decades ago in geometry-based image analysis. The SCENIC Consortium, in fact, is developing a harmonic theory of geometrical acoustics that enables scene reconstruction through constraints in the projective space. This theory is not a direct application of the methodologies developed for 3D vision, as such methods cannot be readily applied to audio applications. The underlying geometrical model of clusters of sensors and emitters, in fact, does not generally exhibit a center of projection (acoustic lenses are not practically used), and propagation delays cannot be neglected (we don't just look at distributions of energy like in photo-cameras, but we need to consider phases as well). There are, however, specific points in space where acoustic rays are bound to meet, which are real or virtual (wall-reflected) sources, as well as real or virtual listening points. With respect to such points, the modelling becomes projective and certain rules and constraints of projective geometry can be applied.

Our approach to geometrical acoustics starts from acoustic measurements in the forms of TOAs (Times of Arrival), TDOAs (Time Differences of Arrival) or DOAs (Directions of arrival), which can be swiftly obtained using combinations of channel-based algorithms. Such measurements can refer to direct (free) propagation as well as indirect (wall-reflected) propagation. All such measurements are then converted into constraints that, in a projective space, always take on a quadratic form, regardless of their origin and geometric setup. Direct TOAs will obviously originate projective circles, direct TDOAs will originate hyperbolas, indirect TOAs will originate ellipses [4-6], indirect DOAs will originate parabolas [7], etc. Clever geometrical combinations of quadratic constraints are then used for extracting all sorts of information on the source location, the structure of the environment, and more. Examples of application of this approach to the self-calibration problem are reported in [2] and [3]. Applications of this approach to the case of environment inference are described in [4-7]. This framework is flexible enough to allow us to progressively add information to the environment description repository, as natural sounds are produced and acquired in the environment of interest.

2.2. Geometric wavefield synthesis

As far as geometric rendering is concerned, we propose a new modelling paradigm for performing WaveField Synthesis in a reverberant environment, which exploits the concept of geometric decomposition of the wavefield. Using a real-time beam tracer [9], [10] we can produce a good approximation of the desired wavefield by overlapping acoustic beams generated with an array of speakers, as well with all virtual arrays (wall-reflected versions of the real array) [11]. With this system we can achieve two results that were not possible before:

1. exploiting the environment to perform WaveField Synthesis (WFS) in a reverberating enclosure, using arrays of smaller dimension
2. rendering not just sources, but also a virtual environment (i.e. all virtual sources along with their "visibility" through the virtual walls)

A complete geometrical wavefield synthesis engine requires the following basic functionalities:

- a *beam-tracing engine* for modelling the acoustic propagation in complex environments;
- a *beam-shaping engine*, which allows us to render an acoustic source and its beam pattern by means of a loudspeaker array.

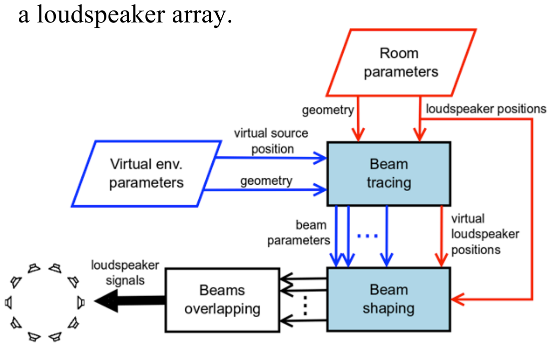

Fig. 1: Block diagram of a geometric WFS engine.

The block diagram in Fig. 1 shows how these functionalities are linked together, in order to realize a completely environment-aware system that uses both real and virtual loudspeaker arrays for synthesizing the acoustics of an arbitrary virtual environment. To this end, all the virtual image sources generated by reflections of a source within the virtual environment have to be rendered. In an arbitrary environment, however, not all the image sources are visible from all points in space. The occlusions that arise greatly influence the wavefield, and determine the sense of presence of a listener in that environment. For an accurate reproduction, therefore, each virtual source must be rendered along with its visibility pattern, which is described as an acoustic beam.

Given the geometry of the virtual environment and the source position, the beam tracer is used for predicting the set of acoustic beams to be rendered by the loudspeakers. The beam tracing algorithm exploits the visibility information in order to speed-up the tracing of acoustic beams. An example is given in Fig. 2, which depicts the prediction of the beams generated by a source located in a densely occluded environment.

The beam-shaping engine is used for rendering arbitrary acoustic beams, which are specified by an origin (i.e. the image source position), the emission direction and the angular aperture. The beam-shaping engine requires knowledge about real and virtual arrays. In particular, the positions of the virtual loudspeakers are predicted using, once again, the beam-tracer technique, provided the geometry of the room in which the beam-shaper operates. The global wavefield is finally synthesized by superposing all the individual beams characterizing the virtual environment, each one properly attenuated and delayed. Fig. 3 shows an example of a geometric rendering system installed into a reverberant room and reproducing the acoustics of a small church. Simulations show that the synthesized soundfield is consistent with the geometry of the virtual environment, as it changes exactly in correspondence with the change of lateral geometry of the environment. We also realized an experiment aimed at reproducing free-field propagation in a highly reverberating L-shaped real environment. The result of geometric room compensation could be perceptively compared with absence of compensation with surprising realism.

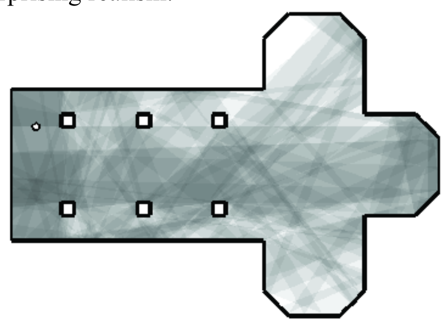

Fig. 2: Superposition of beams in a complex 2D environment (only reflections up to the 2nd order are visualized).

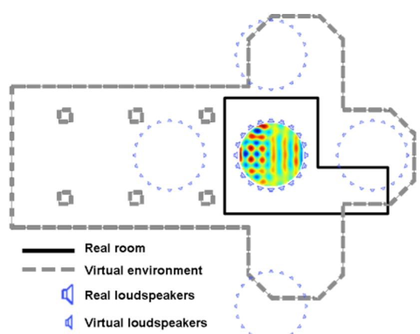

Fig. 3: Spatial audio rendering of the acoustics of a small church in a real reverberant environment. The real environment (solid black line) is used for "augmenting" the array system (real array + image arrays) and constructing the wavefield that we would have in the virtual church (dashed grey line), through beam tracing and beam-shaping.

3. A WAVE-BASED PERSPECTIVE

A wave-based perspective of spatial sound reproduction is adopted by considering a fundamental property of sound fields: The sound pressure of a three-dimensional source free sound field is determined by the sound pressure and the particle velocity on its two-dimensional surface. When sound pressure and particle velocity are created by loudspeakers, almost arbitrary sound fields can be created. In practical setups, the dimensionality is often reduced by one, i.e. the sound pressure in a two-dimensional area is created by a loudspeaker configuration on a one-dimensional boundary like a circular or rectangular loudspeaker array.

The resulting rendering method is called wave-field synthesis and the fundamental acoustical property addressed above is called the Kirchhoff-Helmholtz integral equation. It is found in most classical textbooks on acoustics in various mathematical formulations. The essential statement is that the sound pressure within a volume is determined by the sound pressure and the particle velocity on the surface of this volume. The sound propagation from the surface to the interior is described by the appropriate Green's function.

The form of the Green's function depends on the propagation conditions. Current implementations of wave field synthesis rely on the free-field condition, which is based on two idealizing assumptions: the first of which is that there is no considerable reverberation in the reproduction room; and the second is that the placement of loudspeakers on the surface does not create disturbing reflections. These two assumptions can be met approximately if the reproduction room has a reasonably low reverberation time and if the used array of (usually small) loudspeakers is acoustically transparent. Then the sound propagation from the loudspeakers on the surface to the interior of the enclosed volume can be obtained with the free-field Green's function. However, the validity of the Kirchhoff-Helmholtz integral equation is not restricted to the free-field case. It can be applied also to reproduction rooms with wall reflections if the corresponding Green's function is known. This is where the approach of the SCENIC Consortium sets in. At first the localization of reflecting surfaces provides the required information on the deviation of a reproduction room from the ideal free-field assumptions. Then the modification of the Green's function in the Kirchhoff-Helmholtz integral equation facilitates the sound reproduction under the presence of wall reflections. The practical implementation consists of the use of the modified Green's function for the computation of the loudspeaker's driving functions.

In the simple case of one reflecting surface, the modification of the free-field Green's function consists of adding a second free-field Green's function for the mirror source behind the reflecting wall. In more complicated cases with multiple reflections, mirror source image methods for the approximation of the room impulse response can be used to establish the corresponding Green's functions. But also other methods for the determination of the Green’s function of a reflective room can be used, like the measurement of room modes. In this way, reflections in the reproduction room do not necessarily lead to a deterioration of the free-field based rendering method. When considered appropriately, the room reflections actually contribute to wave field synthesis.

This approach for sound rendering in reflective environments is not to be confused with the reproduction of reverberant virtual spaces with classical model- and data-based wave field synthesis. These methods aim at the reproduction of reverberant spaces in a reflection free reproduction room. On the other hand, SCENIC develops method for the reproduction of (dry or reverberant) sources in reflecting environments.

4. A CHANNEL-BASED PERSPECTIVE

In this section, we review the activities within the SCENIC project that have contributed to the use and understanding of channel-based techniques in acoustic signal processing.

4.1. Channel Equalization

The equalization of an acoustic channel can be divided into two broad modalities: the first is listening room compensation (LRC) that modifies signals prior to being generated in the acoustic environment, the second is dereverberation in which the signals are modified after having been captured from the acoustic environment. Channel equalization techniques can be equally applied in both modalities. Typical Room Impulse Responses (RIRs) are very long, numbering several thousand coefficients, and generally possess a non-minimum phase characteristic. Inevitably, practical systems require causal stable inverse filters whose design has been a focus of the SCENIC project.

Existing inverse filtering systems can be obtained, in the single-channel case, by the method of least-squares (LS) giving an approximate inverse system that is of limited use in acoustic channel equalization [12]. When multiple channels are available, the multiple-input/output inverse theorem (MINT) can provide an exact inverse provided the RIRs do not share any common zeros [13]. Channel shortening (CS) techniques have been extensively developed in the context of digital communications to mitigate inter-symbol and inter-carrier interference by maximizing a generalized Rayleigh quotient. In the case of acoustic channels, channel shortening exploits a useful property of psychoacoustics. It is believed that the early reflections are not perceived as reverberation but as a spectral colouring that is less detrimental to the perceived quality and intelligibility than the late reflections. Channel inversion can therefore be relaxed to suppress only the late reflections while leaving the taps around the early reflections unequalized. This property both simplifies the design of the equalizer [14] and improves its robustness to channel identification error and noise [15]. A contribution within SCENIC was to develop a mathematical link between CS and MINT and to define a criterion for selecting perceptually-advantageous solutions to CS [16]. An alternative approach to CS, called relaxed multichannel least-squares (RMCLS), was also proposed [17]. The choice of an optimum shortening length as a function of noise amplitude and much greater channel errors is an ongoing field of research within the project [15].

4.2. Source Localization and Extraction

Acoustic source localization aims to estimate the localization information of one or several sound sources using the spatial diversity of signals captured by an array of microphones, and it is typically used to steer a beamformer or point a camera in the direction of a source. In general, localization techniques can be divided into three categories: methods that rely on maximizing the output power of steered beamformers, subspace methods based on spectral estimation at high resolution, and approaches based on the estimation of time differences of arrival (TDOAs).

To localize acoustic sources in reverberant environments, blind system identification followed by the TDOA estimation can be used [18-19]. Within the SCENIC project the localization of simultaneously active speakers using only a microphone pair was considered, where independent component analysis based on the TRINICON framework is exploited to blindly learn about the acoustic environment, and next the TDOA information is extracted from the directivity patterns of the BSS outputs [20].

For accurate 3D source localization using a compact spherical microphone array, robust and high-resolution localization algorithms are necessary. Within the SCENIC project the investigation focused on subspace-based and steered beamformer-based localization methods, implemented in both the element space and the spherical harmonics (eigenbeam, EB) domain. In the steered beamformer-based localization approach, an acoustic environment is scanned using, for example, the minimum variance distortionless response beamformer in the eigenbeam domain (EB-MVDR) [21], and the beamformer output power for each look-direction is plotted to form an acoustic map of the environment as exemplified in Fig. 4. The locations of the peaks of the acoustic map determine the estimated directions of arrival (DOAs). The SCENIC contribution includes formulating a novel EB-MVDR beamformer with frequency smoothing and white noise gain control [21], which is robust against array errors and coherent sources. Subspace localization approaches include the EB-MUSIC [22] and EB-ESPRIT [23,24] algorithms, where an eigenvalue decomposition of the covariance matrix constructed from the microphone signals is performed, from which a signal or noise subspace can be extracted and the positions of sound sources can be estimated. In order to improve the robustness and accuracy of the DOA estimation of coherent sources, focusing matrices and frequency smoothing techniques may be employed [21],[25].

Extraction of the localized source signals can be performed using beamforming, in which the desired source signal is extracted from the look-direction and unwanted interference signals are suppressed [26]. In acoustic scenarios with several sources, broadband beamforming techniques characterized by high directivity and arbitrary null placement capabilities are needed. However, such designs are typically highly sensitive to noise [27]. The contribution of SCENIC includes a novel design of the robust broadband beamformer (RBB), proposed in [28], which is optimum in the least squares sense and simultaneously constrains the white noise gain (WNG) to remain above a given lower limit by directly solving a constrained optimization problem, which is convex. This robust data-independent broadband beamforming design enables the placement of nulls in the directions of undesired sources, and it is particularly useful for signal extraction in scenarios where the desired source signals are correlated with the interference signals [29]. Signal extraction in such scenarios using robust statistically optimum beamformers, such as EB-MVDR beamformer, was also investigated [21].

4.3. Applications

The channel-based analysis techniques developed within SCENIC form components of more complex acoustic sensing algorithms.

Of particular interest are a) the extraction of parameters characterizing an acoustic environment and b) exploitation of knowledge of the environment geometry in an acoustic Rake receiver for dereverberation and noise reduction. Block diagrams for the two applications are shown in Fig. 5. In both cases, multichannel observations are used for source localization and extraction that are further processed to extract the desired information.

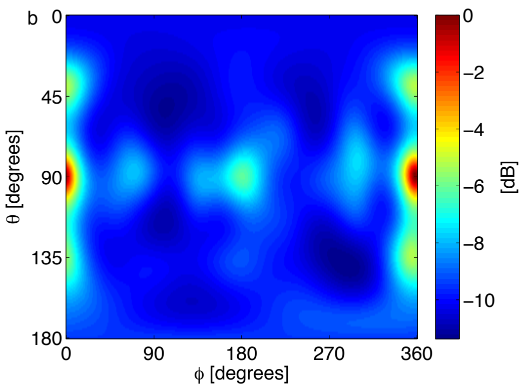

Fig. 4: Example localization result for the EB-MVDR beamformer with frequency smoothing.

Localization and extraction of early room reflections are challenging since their energy is low in comparison to the direct-path signal, they exhibit low SNR, and they are highly coherent with the sound source. In previous reflection localization studies [30-31], non-optimum signal processing methods were typically used, and instead high resolution was achieved by using very large array apertures. During the SCENIC project, we focused our efforts on applying optimum array processing methods such that high accuracy could be obtained using compact off-the-shelf microphone arrays. A comparative study of several subspace-based and steered beamformer-based reflection localization techniques was presented in [32], which shows that EB-MUSIC and robust EB-MVDR with white noise gain control yield very high resolution, and thus lead to an accurate DOA estimation. Having performed the localization, beamformers with high directivity are steered towards the sound sources based on the localizer's DOA information. For this purpose, the EB-RMVDR and the RBB based on convex optimization are used [21][29].

For room geometry inference, we seek the parameters that define the position of the boundary planes. Having performed the localization and extraction steps, statistical analysis of cross-correlations between the extracted signals can then be used to categorize the sources as direct sound and its respective reflections. In addition, the TDOA estimates of room reflections relative to the direct propagation path can be obtained [21], [29]. Next assuming the distance from the source relative to the center of the microphone array is known, the corresponding estimated DOAs and TDOAs of early reflections can be used to calculate the positions of the reflection points (or alternatively image sources) using simple trigonometric relations, as shown e.g. in [21], [29]. The positions of the original and image sources can finally be used to infer the boundary locations. As presented in [21], [29], room inference using the proposed procedure yields highly accurate results, with an error not exceeding 1% relative to the room dimensions.

The acoustic Rake receiver of Fig. 5 (right) aims to add coherently the reflected signals to the direct-path signal in order to improve the output signal-to-noise ratio (SNR). Having extracted the signals from the localized directions, the final step is to linearly combine the extracted beamformer output signals. This can be achieved by aligning the reflections in time with the direct-path arrival using the delay and sum beamformer [33] or by means of a linearly constrained minimum variance (LCMV) filter that combines the reflections under the constraint of distortionless response to the direct sound [34]. The major advantage of the Rake receiver solution developed during the SCENIC project is that an improved performance over previous approaches can be obtained even with a compact spherical array, due to the use of optimum array processing methods in the preprocessing steps.

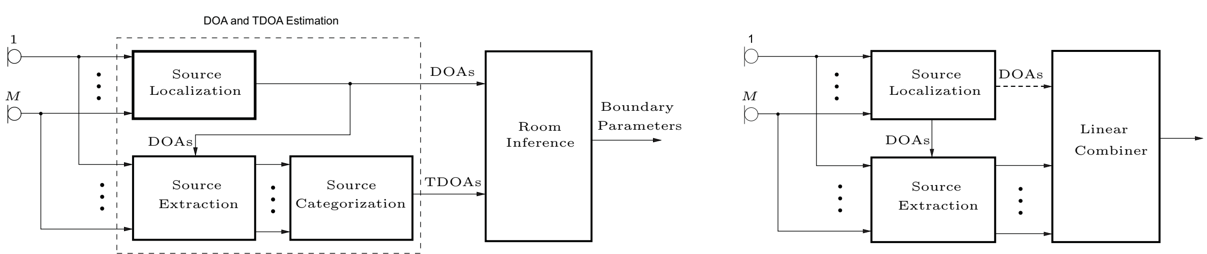

Fig. 5: Block diagram of: geometric inference (left) and the acoustic Rake receiver (right).

5. WAVEFIELD DESCRIPTORS

The SCENIC project also addressed a rather different class of solutions that describe the whole wavefield in a fashion that is only indirectly related to acoustic pressure. Such descriptors, in fact, are related to a measurement of space-time coherence of soundfields, which is estimated locally through the Generalized CrossCorrelation Phase Transform (GCC-PHAT), measured using distributed microphone pairs. This information is then converted into acoustic maps representing the deduced distribution of acoustic sources in space and time from which the location and the orientation of non-omnidirectional sources can be estimated. Maps obtained using microphone pairs distributed all around an area of interest also allow the localization and tracking of multiple overlapping sources (e.g. moving talkers) [35].

In presence of multipath sound propagation the computation of acoustic maps must take into account both real sources and virtual sources generated by reflections. An elementary model of propagation based on the image source method is suitable to this end in environments with simple shapes, while a map-classification approach [36] can be adopted to analyze the acoustic scene in more complex environments.

Specific microphone configurations allow an accurate analysis of the multiple wavefronts bouncing within an enclosure. For example a line array of many closely spaced sensors enables a detailed study of how acoustic impulse responses evolve when the listening point is shifted in space [37]. The related patterns of local coherence along the array, produced by multiple reflected wavefronts, yield acoustic maps (2D or 3D) from which a more accurate localization and characterization of the sources are possible [38].

Since the local coherence is closely related to the ratio between direct-path energy and reverberant energy at the microphones, its measurement can also be exploited to infer the shape of source emission patterns. Provided that a set of microphone pairs with sufficient angular coverage around a source is available, the peaks of GCC-PHAT observed at different azimuth angles, once a model of reverberation in the environment is applied, can be used to characterize the source directivity without the need of anechoic conditions [39].

6. A CONVERGENCE OF METHODOLOGIES

One of the goals of the SCENIC project is to bring together global and local solutions in a harmonic and synergistic fashion. While the relation that exists between global solutions is rather clear, the one that links local and global solutions is a bit more elusive. As a matter of fact, channel-based methodologies are have a long-standing tradition in space-time audio processing, but they tend to play a pivotal role for global methodologies as well. For example, it is through channel-based methodologies that TDOA measurements on unstructured sounds (e.g. finger snapping) are swiftly turned into TOA measurements in [6], thus making geometric inference a great deal more flexible and easier to conduct. It is through channel-based methodologies that a source can be reliably localized and tracked, or that the radiance pattern can be reliably determined. In fact, it is in the diversity and complementarity of the methodologies that the SCENIC project relies for the effectiveness of its solutions. In Fig. 6, for example, we can see how the various categories of solutions concur in creating the necessary environment awareness that is required for challenging space-time processing tasks.

The case of wavefield reproduction can also be addressed in a joint fashion through a synergistic convergence of multiple methodologies, according to the block diagram of Fig. 7. In this case, for example, we see how geometric methods and WFS solutions can follow similar approaches in soundfield reconstruction. The main difference between the two is in the fact that one is based on the computation of green functions, the other on the overlaying of windowed contributions of virtual sources (beams). As the performance of the two approaches are expected to be rather complementary, we are now focusing on their joint use.

7. CONCLUSIONS

The SCENIC project has brought together a wide range of space-time audio processing methodologies in a harmonic fashion, with the purpose of turning the environment into an active functional element of the acoustic sensing or rendering system. We proved that geometric environment inference is a powerful and effective approach for achieving this goal, and we showed that a joint use of the diverse approaches is a promising way to achieve remarkable results in challenging applications such as source characterization, room compensation and virtual acoustics.

8. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's 7th Framework Programme [FP7/2007-2013] under grant agreement n° 226007 (SCENIC - Self-Configuring ENvironment-aware Intelligent aCoustic sensing).

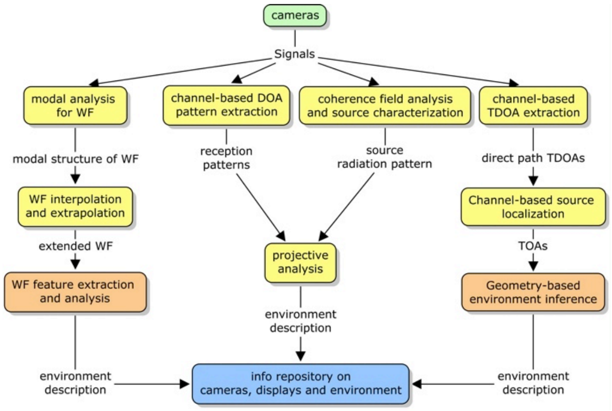

Fig. 6: Gathering information on the environment.

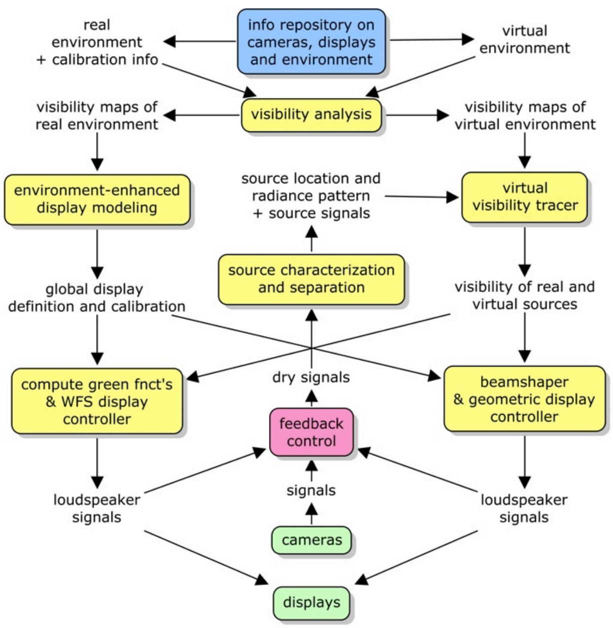

Fig. 7: Wavefield rendering: combining Wavefield Synthesis methods with Geometric Rendering methodologies.

9. REFERENCES

[1] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge Un. Press, June 2000.

[2] A. Redondi, M. Tagliasacchi, F. Antonacci, A. Sarti, “Geometric calibration of distributed microphone arrays”. IEEE Intl. Workshop on Multim. Sig. Proc. MMSP. Oct. 5-7, 2009, pp 1-5. Rio De Janeiro, Brasil.

[3] S.D. Valente, F. Antonacci, M. Tagliasacchi, A. Sarti, S. Tubaro, “Self-calibration of two microphone arrays from volumetric acoustic maps in non-reverberant rooms”. 4th Intl. Symp. on Comm., Control and Sig. Proc. (ISCCSP 2010). Cyprus, Mar. 3-5, 2010.

[4] F. Antonacci, A. Sarti, S. Tubaro, “Geometric reconstruction of the environment from its response to multiple acoustic emissions”, IEEE Intl. Conf. on Acoustics Speech and Signal Processing (ICASSP-10). Dallas, TX, Mar. 14-19, 2010**,** pp. 2822–2825.

[5] A. Canclini, F. Antonacci, M. R. P. Thomas, J. Filos, A. Sarti, P. A. Naylor and S. Tubaro, "Exact Localization of Acoustic Reflectors from Quadratic Constraints," IEEE Workshop on App. of Signal Processing to Audio and Acoust. (WASPAA), New Paltz, New York, Oct. 2011.

[6] J. Filos, A. Canclini, M. R. P. Thomas, F. Antonacci, A. Sarti, P. A. Naylor, "Robust inference of room geometry from acoustic impulse responses," Eur. Sig. Proc. Conf. (EUSIPCO), Barcelona, Spain, Aug. 2011.

[7] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, S. Tubaro, “From Direction of Arrival Estimates to Localization of Planar Reflectors in a Two Dimensional Geometry”. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Rep., May 22-27, 2011, pp. 2620-2623.

[8] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, S. Tubaro, “Soundfield rendering with loudspeaker arrays through multiple beam shaping”. IEEE Intl. Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA '09. New Palz, NY, Oct. 18-21, 2009, pp. 313–316.

[9] F. Antonacci, M. Foco, A. Sarti, S. Tubaro, “Fast Tracing of Acoustic Beams and Paths Through Visibility Lookup”. IEEE Tr. Audio, Speech and Lang. Proc., Vol. 16, No. 4, May 2008, pp. 812-824.

[10] F. Antonacci, A. Sarti, S. Tubaro. “Two-dimensional beam tracing from visibility diagrams for real-time acoustic rendering”. EURASIP J. on Advances in Signal Processing. Vol. 2010, Feb. 2010.

[11] D. Markovic, A. Canclini, F. Antonacci, A. Sarti, S. Tubaro, “Visibility-based beam tracing for soundfield rendering”. IEEE Intl. Workshop on Multimedia Signal Processing (MMSP ‘10), Saint Malo, France, Oct. 4-6, 2010, pp. 40-45.

[12] P. A. Naylor, N. D. Gaubitch, “Speech dere-verberation,” Intl. Workshop Acous. Echo and Noise Control (IWAENC), Eindhoven, The Netherlands, Sept. 2005.

[13] M. Miyoshi and Y. Kandea, “Inverse filtering of room acoustics,” IEEE Trans. Acoust., Speech, Signal Process., vol. 36, no. 2, pp. 145–152, Feb 1988.

[14] M. Kallinger, A. Mertins, “Impulse response shortening for acoustic listening room compen-sation,” Intl. W-shop Ac. Echo and Noise Control (IWAENC), Eindhoven, The Netherlands, 2005.

[15] M. R. P. Thomas, N. D. Gaubitch, P. A. Naylor, “Application of channel shortening to acoustic channel equalization in the presence of noise and estimation error,” in Proc. IEEE Workshop on Applications of

Signal Processing to Audio and Acoustics, New Paltz, New York, Oct. 2011.

[16] W. Zhang, E. A. P. Habets, P. A. Naylor, "A System-identification-error-robust method for equalization of multichannel acoustic systems," IEEE Intl. Conf. on Ac., Speech and Signal Proc. (ICASSP), 2010.

[17] W. Zhang, A. W. H. Khong, P. A. Naylor, "Acoustic system equalization using channel shortening techniques for speech dereverberation", European Sig. Proc. Conf. (EUSIPCO), 1427–1431, 2009.

[18] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," J. Acoust. Soc. Am., vol. 107, pp. 384–391, Jan. 2000.

[19] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, W. Kellermann, "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering", IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Mar 2005, Philadelphia, PA.

[20] A. Lombard, T. Rosenkranz, H. Buchner, W. Kellermann, "Exploiting the self-steering capability of blind source separation to localize two or more sound sources in adverse environments," Proc. ITG Conf. Speech Communication, 2008, Aachen, Germany.

[21] H. Sun, E. Mabande, K. Kowalczyk, W. Kellermann, "Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays," IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 113–116, May 2011.

[22] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, E. Fisher, "Spherical microphone array beamforming," in *Speech Processing in Modern Communication: Challenges and Perspectives*. I. Cohen, J. Benesty and S. Gannot, Eds. Springer, Berlin, 2010.

[23] H. Teutsch, W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," IEEE Intl. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP), pp. 5276–5279, 2008.

[24] H. Sun, H. Teutsch, E. Mabande, W. Kellermann, "Robust localization of multiple sources in reverberent environments using EB-ESPRIT with spherical microphone arrays," IEEE Intl. Conf. on Ac., Speech, and Sig. Proc. (ICASSP), pp. 117–120, May 2011.

[25] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," Proc. IEEE WASPAA, pp. 221–224, Oct. 2009.

[26] J. Bitzer, K. U. Simmer, "Superdirective microphone arrays," in Microphone Arrays: Signal Processing Techniques and Applications. M. S. Brandstein and D. B. Ward, Eds. Springer-Verlag, 2001.

[27] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[28] E. Mabande, A. Schad, W. Kellermann, "Design of robust superdirective beamformers as a convex optimization problem", IEEE Intl. Conf. on Ac., Speech and Sig. Proc. (ICASSP), pp. 77–80, 2009.

[29] E. Mabande, H. Sun, K. Kowalczyk, W. Kellermann, "On 2D localization of reflectors using robust beamforming techniques", IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 153–156, May 2011.

[30] M. Park, B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," J. Acoust. Soc. Am., Vol. 118, No. 5, pp. 3094–3103, Nov. 2005.

[31] B. Rafaely, I. Balmages, L. Eger, "High-resolution plane-wave decomposition in an auditorium using a dual-radius scanning spherical microphone array", J. Ac. Soc. Am., Vol. 122, No. 5, pp. 2661–2668, Nov. 2007.

[32] E. Mabande, H. Sun, K. Kowalczyk, W. Kellermann, "Comparison of subspace-based and steered beamformer-based reflection localization methods," European Sig. Proc. Conf. (EUSIPCO), May 2011.

[33] A. O'Donovan, R. Duraiswami, D. Zotkin, "Automatic matched filter recovery via the audio camera," IEEE Intl. Conf. on Ac., Speech, and Signal Processing (ICASSP), pp. 2826–2829, Mar. 2010.

[34] Y. Peled and B. Rafaely, "Method for dereverberation and noise reduction using spherical microphone arrays," IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pp. 113–116, Mar. 2010.

[35] A. Brutti, M. Omologo, P. Svaizer, "Multiple source localization based on acoustic map de-emphasis," EURASIP, J. on Audio, Speech, and Music Processing, vol. 2010, 2010.

[36] A. Brutti, M. Omologo, P. Svaizer, C. Zieger, "Classification of acoustic maps to determine speaker position and orientation from a distributed microphone network", Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2007.

[37] P. Svaizer, A. Brutti, M. Omologo, "Analysis of reflected wavefronts by means of a line microphone array," in *Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, 2010.

[38] P. Svaizer, A. Brutti, M. Omologo, "Use of reflected wavefronts for acoustic source localization with a line array," in Third Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA), Edinburgh, 2011.

[39] A. Brutti, P. Svaizer, M. Omologo, "Inference of acoustic source directivity using environment awareness," European Sig. Proc. Conf. (*EUSIPCO)*, Barcelona, 2011.