Ph.D. in Information Technology
Thesis Defense

May 10th, 2024
at 11:00 a.m.
Sala Seminari Nicola Schiavoni, Building 20

Marco OLIVIERI – XXXVI Cycle

MODEL-AIDED DATA-DRIVEN APPROACHES FOR SOUND FIELD ANALYSIS AND PROCESSING

Supervisor: Prof. Fabio Antonacci

Abstract:

Sound field analysis and processing refers to the study of pressure waves for the characterization and manipulation of acoustic fields. It involves measuring, investigating, and extracting the acoustic properties within a specific environment such as rooms, offices, concert halls, and other enclosed areas. The knowledge of sound generation and propagation and the prediction of wave interactions with multiple objects is a highly active area of research within the signal-processing community. Indeed, the description of sound fields enables the creation of immersive and high-quality sound experiences in a wide range of applications, from the audio production and musical instrument industries to virtual reality and entertainment applications. In this thesis, we propose novel methodologies for both the analysis and processing of sound field taking advantages of the physical knowledge of acoustic models and the extracted information of available recordings. We define compact representations to encode the acoustic spatial information of sound sources and we develop data-efficient solutions for the interpolation and manipulation of the sound scenes. Specifically, we provide novel techniques for the characterization of vibrating sources and for the visualization of acoustic fields to effectively describe the interaction of acoustic entities within the space. Moreover, we face the problem of sound field reconstruction and speech separation with flexible solutions with respect to the specific microphone configuration adopted.

The problem of analyzing and extrapolating the whole pressure field through microphone acquisitions has been widely studied in the literature. The existing solutions can be broadly classified into those based on physical models of the acoustic propagation and on data-driven approaches. In the first case, mathematical formulations of acoustic fields, such as plane-wave decomposition, are employed to describe the interactions of the acoustic system through a set of equations. Conversely, the second case relies on chained mathematical operations, e.g., Deep Neural Networks (DNNs), to learn a representation of the available measured data providing estimates of the desired solution. However, the main challenges of today's audio solutions are represented by the strong modeling assumptions that limit the accuracy and flexibility of model-based approaches and the generalization capabilities of classical data-driven methods, which typically depend on the quality and dimension of the dataset adopted during training. Therefore, new solutions are needed to overcome these limitations promoting the full description of acoustic fields even from a limited set of available measurements.

Here, we present novel methodologies that combine the advantages of Deep Learning (DL) strategies and model-based methods in the context of sound field analysis and processing. The devised model-aided data-driven approaches are able to characterize the pressure radiation of vibrating sources representing acoustic features of the sound field in unique representations that can be investigated and manipulated to extract useful information. The main feature of the proposed solutions is represented by the integration of DNNs to exploit the available data with the prior knowledge of the underlying well-known signal models based on the physics of acoustics.

In the context of sound field analysis, information about the properties of sound sources play a crucial role to describe the acoustics of a given space. The characterization of the vibrating sources, for example, is a key factor to model their pressure radiation and predict the interactions with nearby objects and surfaces. In this sense, Nearfield Acoustic Holography (NAH) enables accurate predictions of the vibrational field on sources in a fully contactless way. This property represents an essential requirement in specific cases where the structure under analysis is particularly fragile, e.g., for musical instruments. Thanks to the collaboration between our research group and Violin Museum settled in Cremona, Italy, which houses several stringed instruments made by the luthier Antonio Stradivari, we gained the experience of simulating and predicting the dynamic properties of musical instruments. In particular, we have the opportunity to study and investigate their acoustic properties and to develop new techniques for providing non-invasive measurements. For this reason, we introduce an innovative data-driven approach based on Convolutional Neural Networks (CNNs) to estimate the velocity field on the surface of different wood-made violin plates with variable outline and shapes starting from the pressure measurements acquired in their proximity. To the best of our knowledge, this is the first time a DL strategy is applied in the context of NAH and we prove its effectiveness with respect to model-based methods present in the literature. Moreover, we take advantage of the prior knowledge of the underlying propagation problem represented by the Kirchhoff-Helmholtz integral equation to increase the accuracy of the CNN estimates, thus providing explainable and physically meaningful solutions.

It is worth noticing that the characterization of vibrating objects and the spatial features of the sound sources contribute to a better design, analysis, and understanding of the properties of sound fields. Such information is typically captured with multiple microphones, usually arranged in clusters or arrays within the environment. Acoustic imaging techniques represent a powerful approach to collect and encode the spatial information of the sound field. Specifically, they enable visual representations of the acoustic fields that can be easily investigated and manipulated. Here, we focus on the analysis of sound scenes where sources are concentrated in a confined space, e.g., for teleconferencing applications. We introduce a novel linear operator that maps the acoustic features captured by a circular microphone array surrounding the region of interest into a compact domain, denoted as "angular space", thus enabling the adoption of pattern analysis techniques to extract source properties. Differently from other acoustic imaging methods, the devised model is able to inherently combine the local information of multiple and distributed microphone arrays into a single and unique representation without complex projection or triangulation operations. We prove the ability of the angular space to efficiently describe the properties of real-world sound fields with examples of source localization application adopting different microphone setups.

Regarding the processing of sound fields, typical application scenarios rely on a small set of available pressure data to extract target signals or identify hidden acoustic features of the sound scene. Nevertheless, the ever-evolving trends in immersive multimedia communication require the estimation of the whole acoustic field within a specific environment. Sound field reconstruction

addresses this problem by recovering the pressure data in target locations starting from a small set of observations, thus requiring specific interpolation procedures. Although classical solutions are based on compressed and intuitive descriptions of the sound field, they are limited by the accuracy and flexibility trade-off between the simplified models and the actual behaviors of the acoustic fields. Recently, data-driven approaches proved their ability for the reconstruction of sound fields thanks to the extracted information from the available data. However, the majority of such DL methods provide accurate estimations considering only the magnitude of the sound fields and for height-invariant cases, i.e., in 2D. Here, we aim at recovering the entire acoustic field in a target 3D region of the environment by combining the feature extraction capabilities of DNNs with the propagation model of sound waves represented by the wave equation. Therefore, we define a time-domain neural network that is able to estimate the desired signals from a small and sparse set of real measurements increasing the reconstruction performance with respect to recent state-of-the-art approaches thanks to the physics-informed prior.

Another audio processing task of particularly interest is represented by speech separation, which involves the isolation and enhancement of a specific speech signal of interest. In this thesis, we focus on the extraction of a target speaker placed in a noisy and reverberant environment from the mixture acquired by a linear microphone array, e.g., for hands-free interaction devices. Classical algorithms rely on beamforming techniques to spatially filter the sound field coming from a specific direction or on recent data-driven approaches to extract the desired signals directly from data. However, such methods lack of generalization with respect to different acoustic conditions and, in general, they are limited by the microphone array configurations adopted in the specific measurement setups. Therefore, we propose a novel real-time model that takes advantage of the combination of classical beamforming operators and data-driven strategies. Thanks to the adoption of a beamspace representation based on a predefined set of steerable directions, we define a lightweight CNN that is agnostic with respect to the measurement setup, i.e., number of microphone and inter-sensors distance of the array. We prove that the devised method is able to estimate the desired speaker even when the sound field is captured by a microphone array configuration not seen during the training phase, thus providing robust generalization capabilities in real-world scenarios.

The promising results presented in this manuscript prove how combining features extracted from data with the prior knowledge given by the underlying physical models represents a powerful technique for audio applications. The devised model-aided data-driven approaches enable data-efficient solutions providing accuracy and flexibility properties required in daily-life acoustic devices. We accurately describe the source spatial properties into visual representations of the sound field to analyzed the acoustic scene. Successively, we provide robust generalization properties with respect to different measuring conditions and a limited number of observations for applications in practical scenarios. Therefore, we believe that this thesis can be considered as the starting point for a new class of algorithms for the characterization and manipulation of sound fields that take advantage form the description of the acoustic models and the learning potentialities of neural networks. Moreover, we envision the integration of this framework also for the rendering of synthesized acoustic environments. With the combination of model-based and data-driven methods, new space-time audio processing solutions will be developed for immersive audio applications that rely on a few increasingly smaller and spatially distributed sensors.

PhD Committee

Alberto Bernardini, **Politecnico di Milano**
Stefania Cecchi, **Università politecnica delle Marche**
Julio Carabias-Orti, **Universidad de Jaén**