Elektrotechnik-Elektronik-Informationstechnik

EEI KOLLOQUIUM

Multi-Microphone Speaker Localization on Manifolds: Achievements and Challenges

Prof. Sharon Gannot
Bar-Ilan University, Faculty of Engineering

Mittwoch, der 19.07.2017, 17^{00} Uhr
Wetterkreuz 15, Raum 00.071

Diskussionsleitung: Prof. Dr.-Ing. W. Kellermann

Speech enhancement is a core problem in audio signal processing, with commercial applications in devices as diverse as mobile phones, conference call systems, hands-free systems, or hearing aids. An essential component in the design of speech enhancement algorithms is acoustic source localization. Speaker localization is also directly applicable to many other audio related tasks, e.g. automated camera steering, teleconferencing systems and robot audition.

From a signal processing perspective, speaker localization is the task of mapping multichannel speech signals to 3-D source coordinates. To accomplish viable solutions to this mapping, an accurate description of the source wave propagation, captured by the respective acoustic channel, is required. The acoustic channels in reverberant environments represent a complex reflection pattern stemming from the surfaces and objects characterizing the enclosure. Hence, they are usually modelled by a very large number of coefficients, resulting in an intricate high-dimensional representation.

We start our talk, by analyzing these acoustic responses with nonlinear dimensionality reduction techniques (diffusion maps). We claim that in static acoustic environments, despite the high dimensional representation, the difference between acoustic channels is mainly attributed to the changes in the source position. Thus, the true intrinsic dimensions of the variations of the acoustic channels are significantly fewer than the number of variables commonly used for their representation, namely, they pertain to a low-dimensional manifold that can be inferred from data collected in a training stage. This claim is validated by a comprehensive experimental study in actual acoustic environments.

Motivated by this result, we present a data-driven and semi-supervised source localization algorithm based on two-microphone measurements, which accurately recovers the inverse mapping between the acoustic samples and their corresponding locations. The gist of the algorithm is based on the concept of manifold regularization in a reproducing kernel Hilbert space (RKHS), which extends the standard supervised estimation framework by adding an extra regularization term, imposing a smoothness constraint on possible solutions with respect to a manifold learned in a data-driven manner.

We then show that an analogue mapping operator between the acoustic channel and the source location can be inferred from the Bayesian inference perspective. This Bayesian framework serves as a corner stone for extending the single node (microphone pair) setup to an ad hoc network of microphone pairs. Each node represents a different viewpoint that may be associated with a specific manifold. Merging the different manifolds is shown to increase the spatial separation and to improve the ability to accurately localize the source.