



THE WEIZMANN INSTITUTE OF SCIENCE
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Vision and Robotics Seminar

Room 1 ,Ziskind Building
on Thursday, Nov 23, 2017
at 12:15

Aviv Gabbay
Hebrew University

Seeing Through Noise: Visually Driven Speaker Separation and Enhancement

Abstract:

Isolating the voice of a specific person while filtering out other voices or background noises is challenging when video is shot in noisy environments, using a single microphone. For example, video conferences from home or office are disturbed by other voices, TV reporting from city streets is mixed with traffic noise, etc. We propose audio-visual methods to isolate the voice of a single speaker and eliminate unrelated sounds. Face motions captured in the video are used to estimate the speaker's voice, which is applied as a filter on the input audio. This approach avoids using mixtures of sounds in the learning process, as the number of such possible mixtures is huge, and would inevitably bias the trained model.

In the first part of this talk, I will describe a few techniques to predict speech signals by a silent video of a speaking person. In the second part of the talk, I will propose a method to separate overlapping speech of several people speaking simultaneously (known as the cocktail-party problem), based on the speech predictions generated by video-to-speech system.