Model-based denoising: Examples from speech and hearing science

Yi-Wen Liu, EE, NTHU

Abstract. Speech denoising has been an old topic which surprisingly still has room for exploration. A working principle is to acknowledge that speech is not any ordinary audio signal, but a special kind of signals produced vocally by humans for humans to listen. Hence, a good way for denoising would be to search for the best approximation within a parametrically constrained space, given noisy recordings of the speech. In this talk, I'd like to review two methods, namely subtractive synthesis and additive synthesis. The subtractive method seeks to model speech as filtered wideband source, and the source-filter model can also serve as a preprocessing step for "listening the shape of the mouth". The subtractive approach is actually the corner stone for speech communication over the cellphone. The additive method, in contrast, seeks to model speech as a sparse sum of time-varying sinusoidal functions, and it is most useful for representing sustained vowels. To make it work, one must consider psychoacoustic models to decide whether spurious spectral peaks are worth encoding. For the additive method, I will point out the connection to re-assignment and synchrosqueezing approaches, as well as the model's re-branding as neural networks in recent years. Sound examples will be given throughout the talk.