Saurjya Sarkar

Emmanouil Benetos, Mark Sandler

Music Ensemble Separation

Permutation invariant training not only enables separating mixtures of identical instruments in a label agnostic fashion, but also allows separation of unseen instruments with the assumption that each source is monophonic.

How do you define source separation? Ideally you should be able to separate any sonic event in an audio recording. If we assume each monophonic event as a source, could we ideally have universal source separation? How do we train a model with variable number of output nodes that are dynamically activated based on the level of polyphony in the mixture?