Instrument separation in poliphonic recordings using instrument prints

Kristóf Aczél

Decomposing a polyphonic musical recording to separate voice and instrument tracks has always been a challenge. Even extracting one of the many instruments in a complex recording is currently unsolved. The importance of this issue shows when we want to apply different filters to single instruments in a recording that have already been mixed into stereo channels. The paper shows a new way of sound separation that works even on mono-aural digital recordings.

There are already achievements on sound source separation. Although these algorithms may do a good job in separating instruments or groups of instruments from each other, they are not capable of separating one single note from the remaining part of the recording. It means that they are e.g. unable to filter out one and only one single misplayed note from a polyphonic piece.

The paper deals with one approach for separating single notes in the recording. The goal is using this approach later for doing pitch shifting on the single separated notes in a polyphonic recording when needed. Therefore we will concentrate on separating only those parts that determine the base frequency of a musical note. This means that in case of e.g. a piano recording we do not deal with the sound of the hammer hitting the strings - which remains more or less the same regardless of which key we press - but only with the sound of the strings which vary as we play different notes on the instrument.

The separation algorithm is based on the discreet Fourier transform which is used for converting the sound data from time domain to frequency domain. We compute the spectogram of two recordings. One is the original recording, which has to be processed. The second one is a recording that contains sample notes of the instrument of interest.

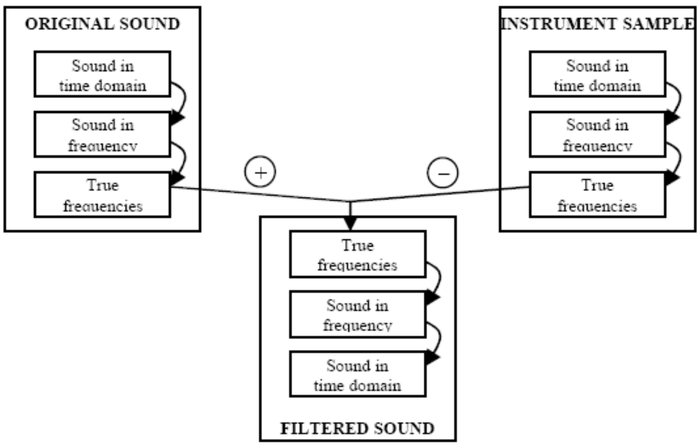

Figure 1: Block diagram of sound separation

We can generate an instrument model of the specific instrument from the second recording, storing its characteristics. We will call this model "instrument print" model. After storing the right instrument print, we can analyse the first recording. We find the spot where we want to separate one or more notes from the remainder of the recording, then subtract the right frequencies from the recording in frequency domain, with the help of the instrument print. The subtracted component can be kept if needed, or thrown away, if not. After this step, the sound data can be converted back to time-domain.