

Sound source separation for music remixing

Andrew Nesbit

`andrew.nesbit@elec.qmul.ac.uk`

Centre for Digital Music
Queen Mary, University of London

5 October, 2004



Music remixing

- remastering old recordings
- original piece remixed into completely new track
- examples

What is the problem?

- no access to master tracks
 - physically lost
 - bouncing tracks
- or, perhaps master track is single channel
- ad hoc filtering techniques are common, but typically perform poorly
- more sophisticated techniques are needed

Separation tasks

- significance orientated (SO): preserve as many of the features as possible to provide a description (e.g., score, text, identification, indexation)
- audio quality orientated (AQO): integrity of audio data is important (e.g., extraction of single source, remixing)
- criteria for evaluation
- Vincent et al (2003)

Blind source separation (BSS)

- the simplest model:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}$$

where $n \geq m$ and each of the source signals s_i are statistically independent

- BSS aims to recover s using only the observed x (Cardoso, 1998)

Audio scene modification

- the model:

$$\hat{x}_{remix} = Bf(\hat{s})$$

where B is a (different) mixing matrix, the f_i are audio processing functions applied to each extracted source \hat{s}_i

- critical step is in isolating the s to yield \hat{s}

Independent component analysis (ICA)

- given the BSS model $x = As$, ICA tries to find the linear transformation A and source signals s such that the recovered source signals \hat{s} are as independent as possible
- can learn higher order structure of natural sounds (Bell and Sejnowski, 1996)
- Hyvärinen, Karhunen and Oja (2001)

Sparse decomposition (SD). . .

- assume we can represent a signal by

$$s = \sum_{i=1}^N c_i \phi_i$$

where the c_i are mixing coefficients and ϕ_i are *atoms* or *elements* from a (possibly overcomplete) dictionary of size N

- *sparsity* means that most of the c_i are zero or close to zero

. . . sparse decomposition continued

- adapting the dictionary to the signal is equivalent to BSS (Lewicki and Sejnowski, 1998)
- exploiting sparsity to separate sources (Zibulevsky and Pearlmutter 2001)

Clustering

- residual dependencies between components
- topographic ICA (Hyvärinen, Hoyer and Inki, 2001)
- independent subspace analysis (Casey and Westner, 2001)
- geometric dependency analysis (Abdallah, 2001)

Applications

- remastering
- spatialisation
- creating a new piece
- creating new sounds

Issues

- source separation
 - monophonic or stereophonic observation signal
 - grouping independent components and manual intervention
 - how many sources are present? when does this question make sense?
 - decomposition and filtering of single sources (complex sounds)
- remixing
 - remixing as filtering
 - introducing new sources
 - effect of remixing on timbre and spatial configuration

- tying it together
 - framework not necessarily tied to BSS
 - is explicit extraction necessary?
- evaluation!

Final comments

- please ask me if you want a copy of my research proposal, or email me at Andrew.Nesbit@elec.qmul.ac.uk