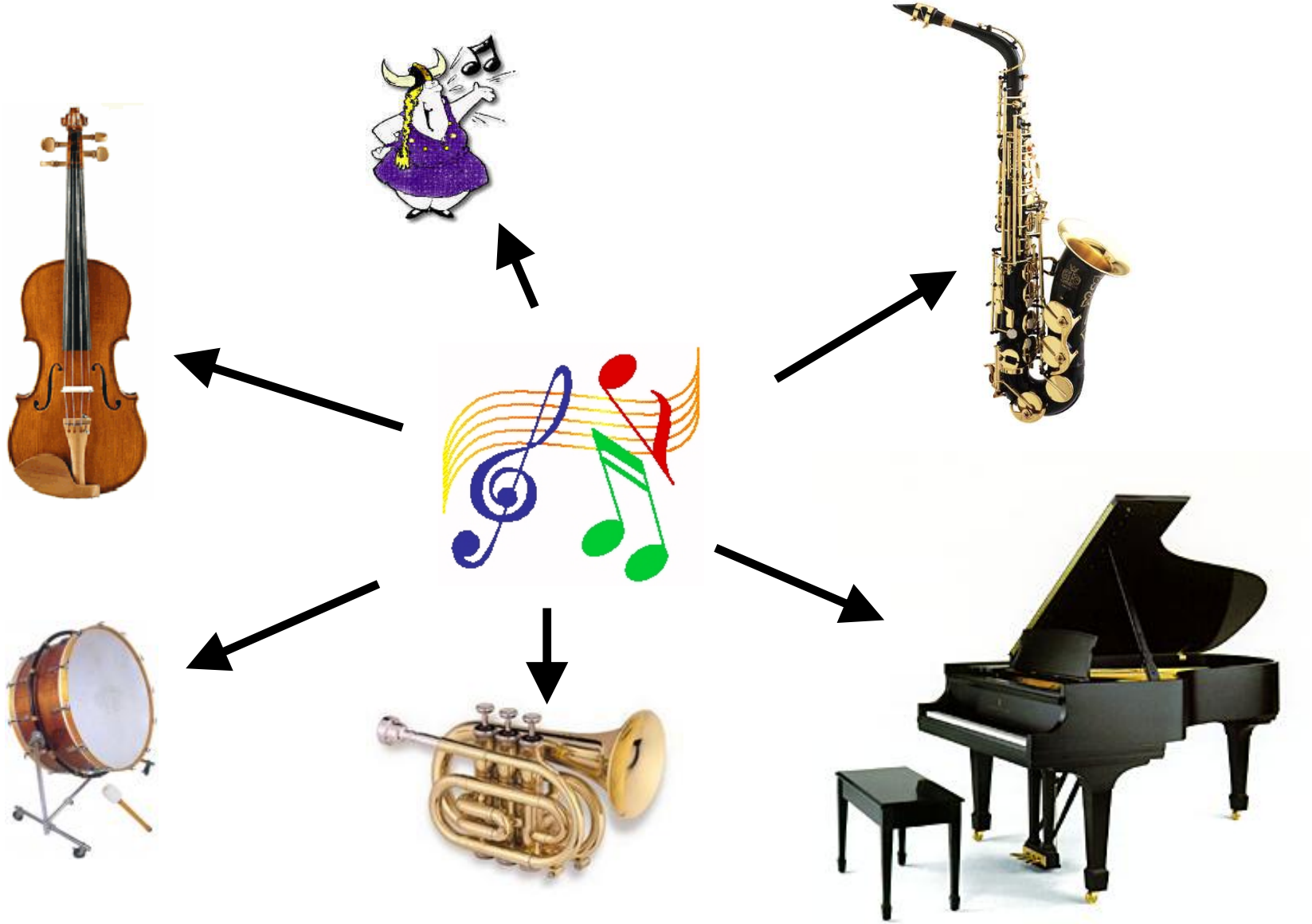
BLUES from Music:
BLind Underdetermined Extraction of Sources from Music

Michael Syskind Pedersen
Tue Lehn-Schiøler
Jan Larsen
IMM, Technical University of Denmark
ICA2006, Charleston, SC, USA

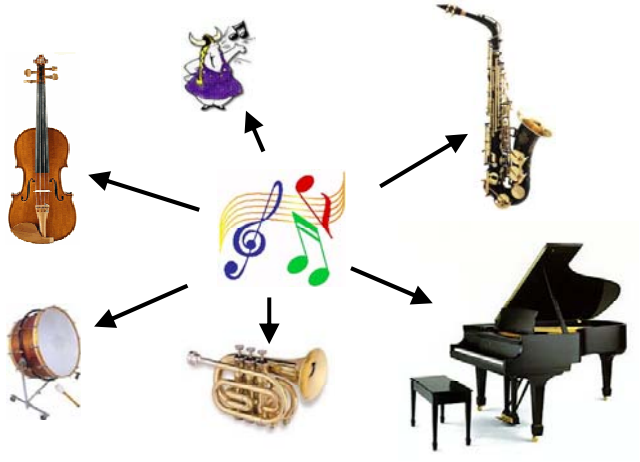
Separating music into basic components



Michael Syskind Pedersen, IMM, Technical University of Denmark

Motivation: Why separating music?

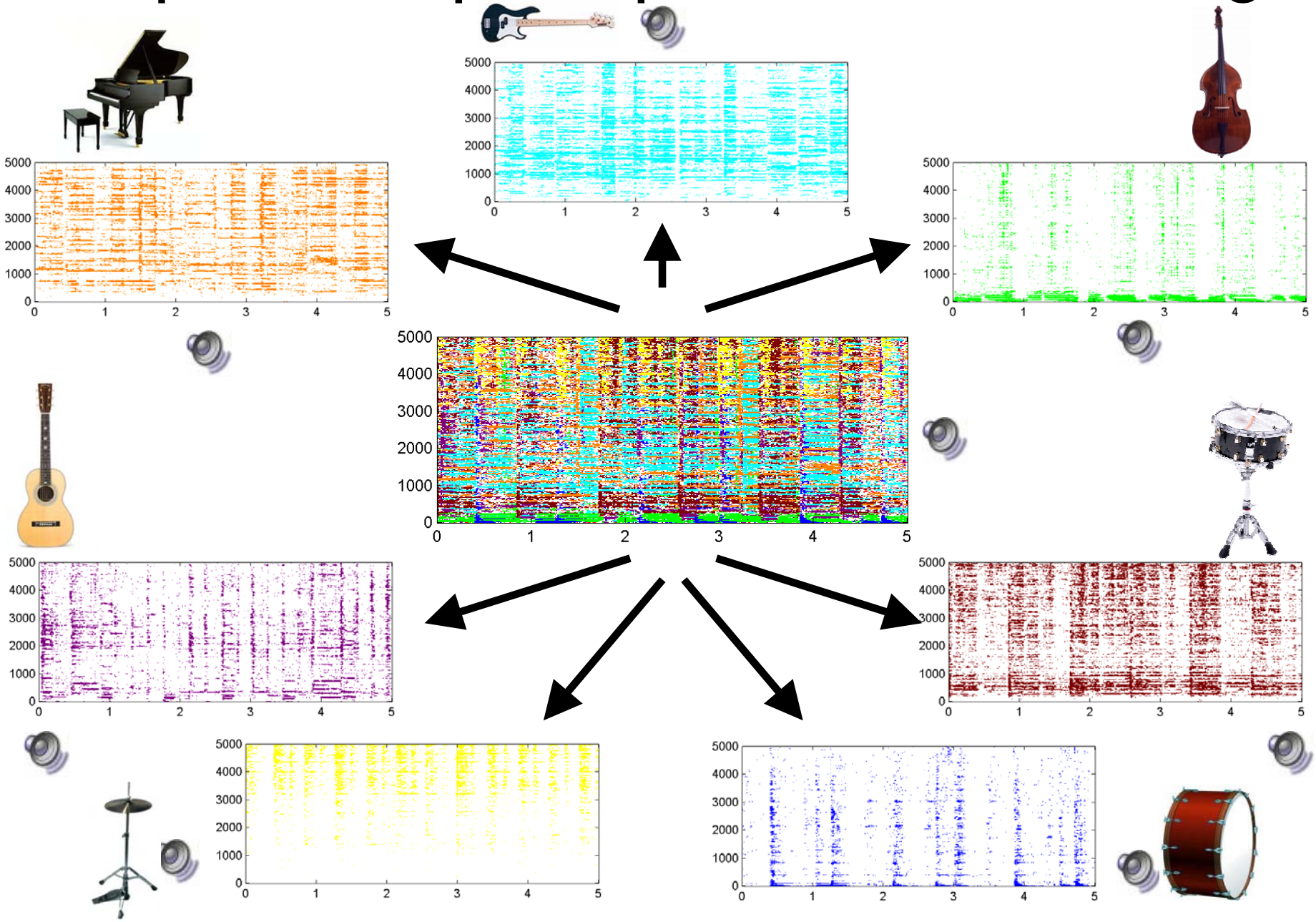
- Music Transcription
- Identifying instruments
- Identify vocalist

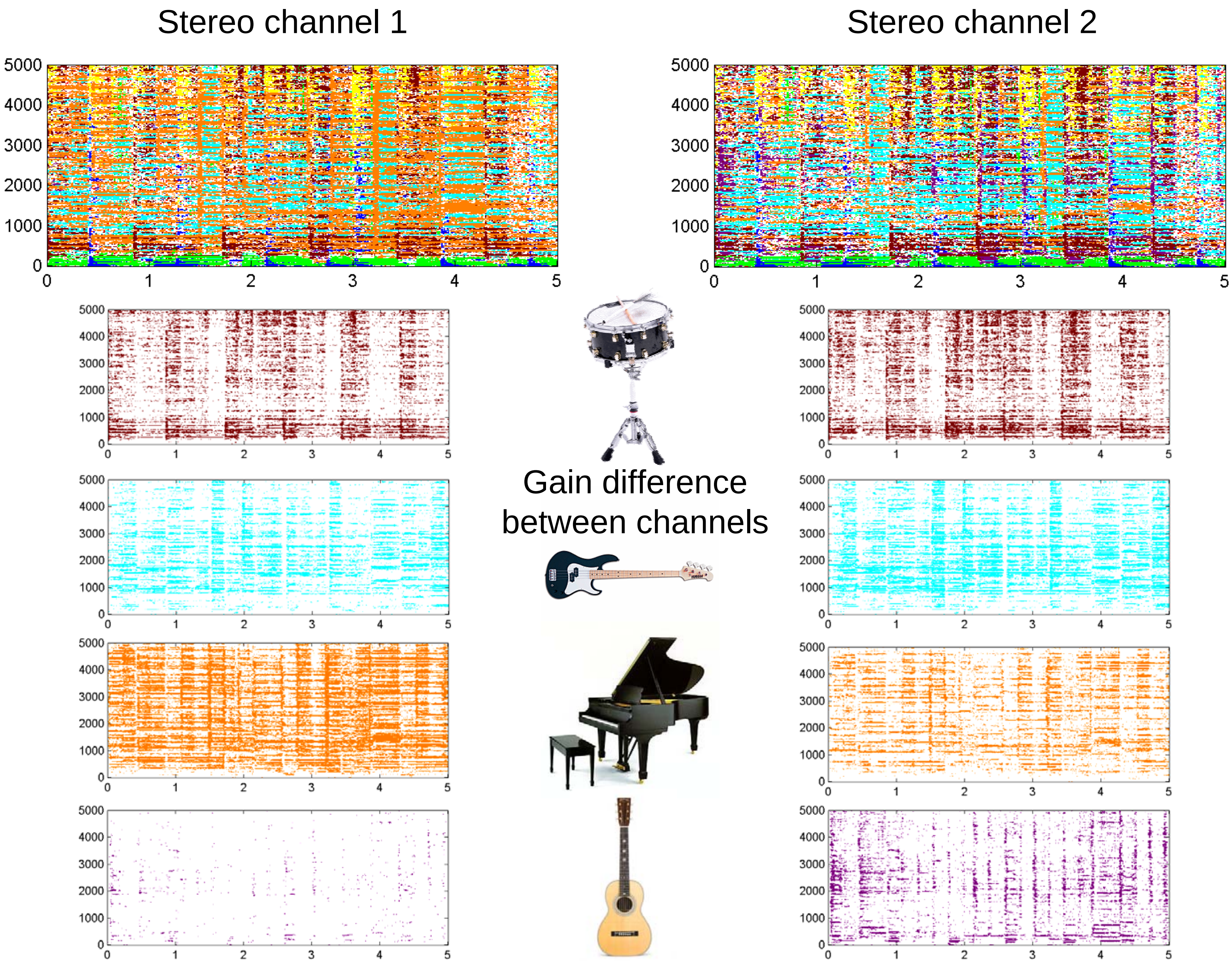


Assumptions

- Stereo recording of the music piece is available.
- The instruments are separated to some extent in time and in frequency, i.e. the instruments are sparse in the time-frequency (T-F) domain.
- The different instruments originate from spatially different directions.

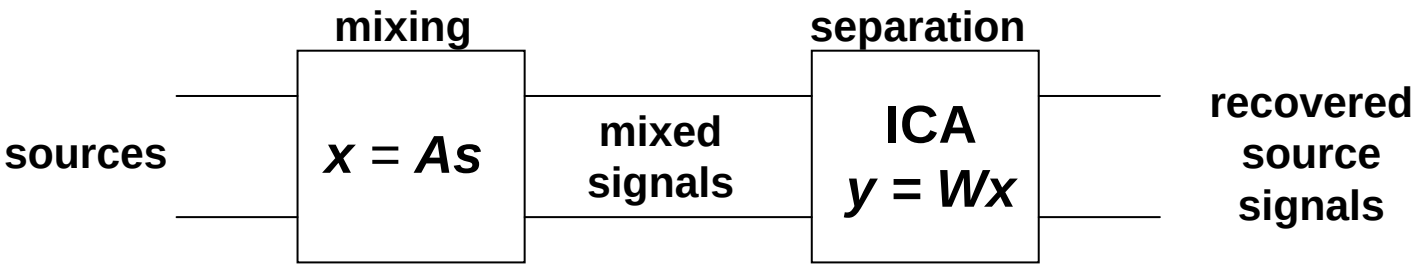
Separation principle 1: T-F masking





Stereo channel 1
Stereo channel 2
Gain difference between channels

Separation principle 2: ICA



What happens if a 2-by-2 separation matrix **W** is applied to a 2-by-N mixing system?

ICA on stereo signals

- We assume that the mixture can be modeled as an instantaneous mixture, i.e.

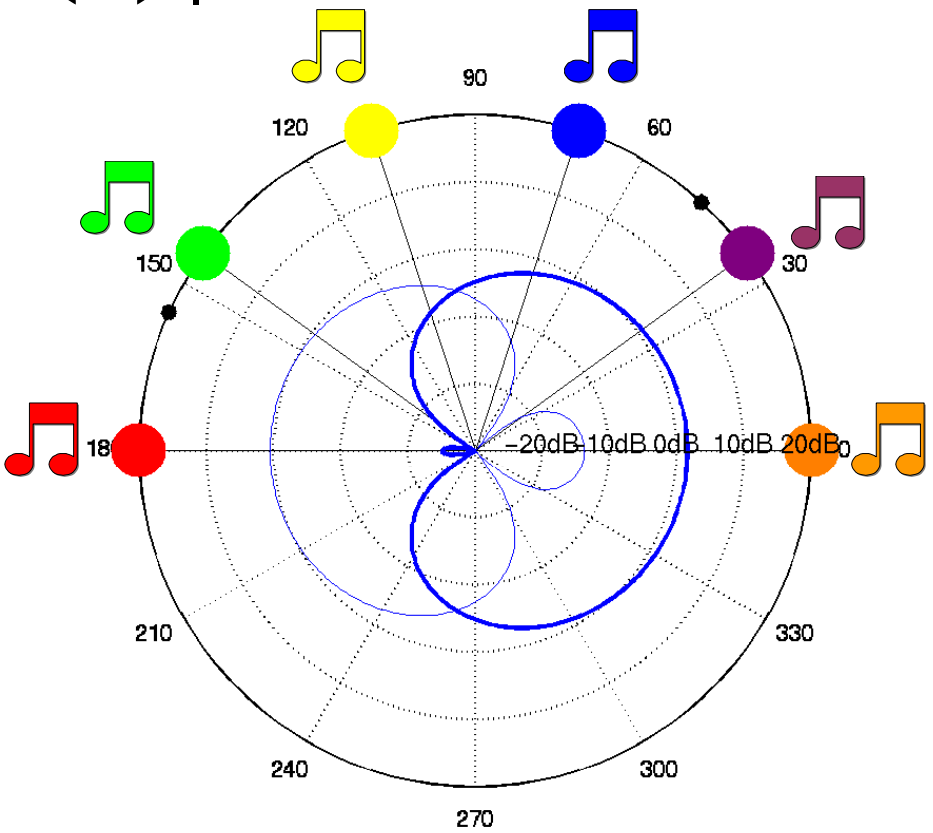
$$x = A(\theta_1, \ldots, \theta_N)s \qquad A(\theta) = \begin{bmatrix} r_1(\theta_1) & \cdots & r_1(\theta_N) \\ r_2(\theta_1) & \cdots & r_2(\theta_N) \end{bmatrix}$$

- The ratio between the gains in each column in the mixing matrix corresponds to a certain direction.

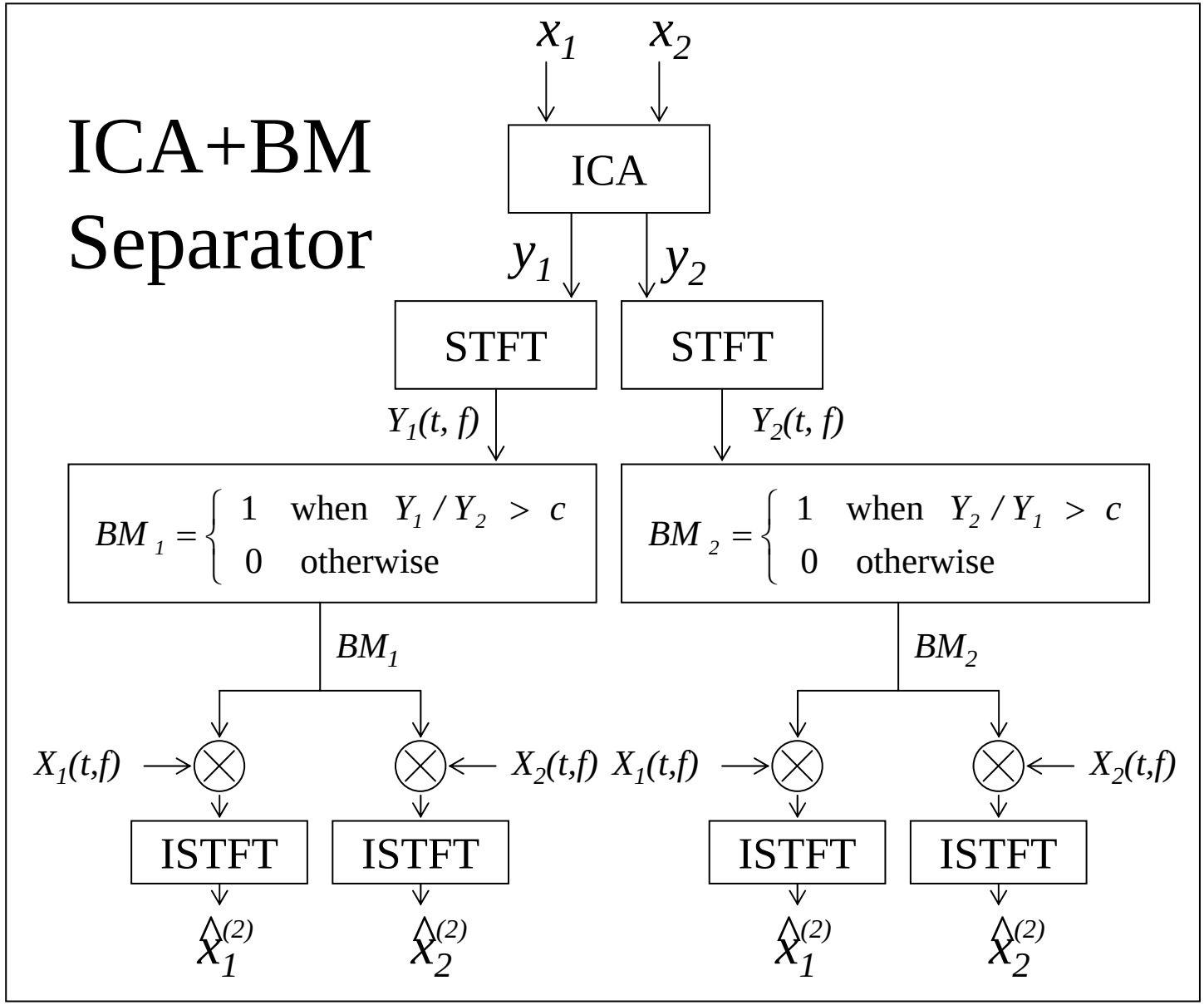
Direction dependent gain

$$\mathbf{r}(\boldsymbol{\theta}) = 20\log|\mathbf{WA}(\boldsymbol{\theta})|$$

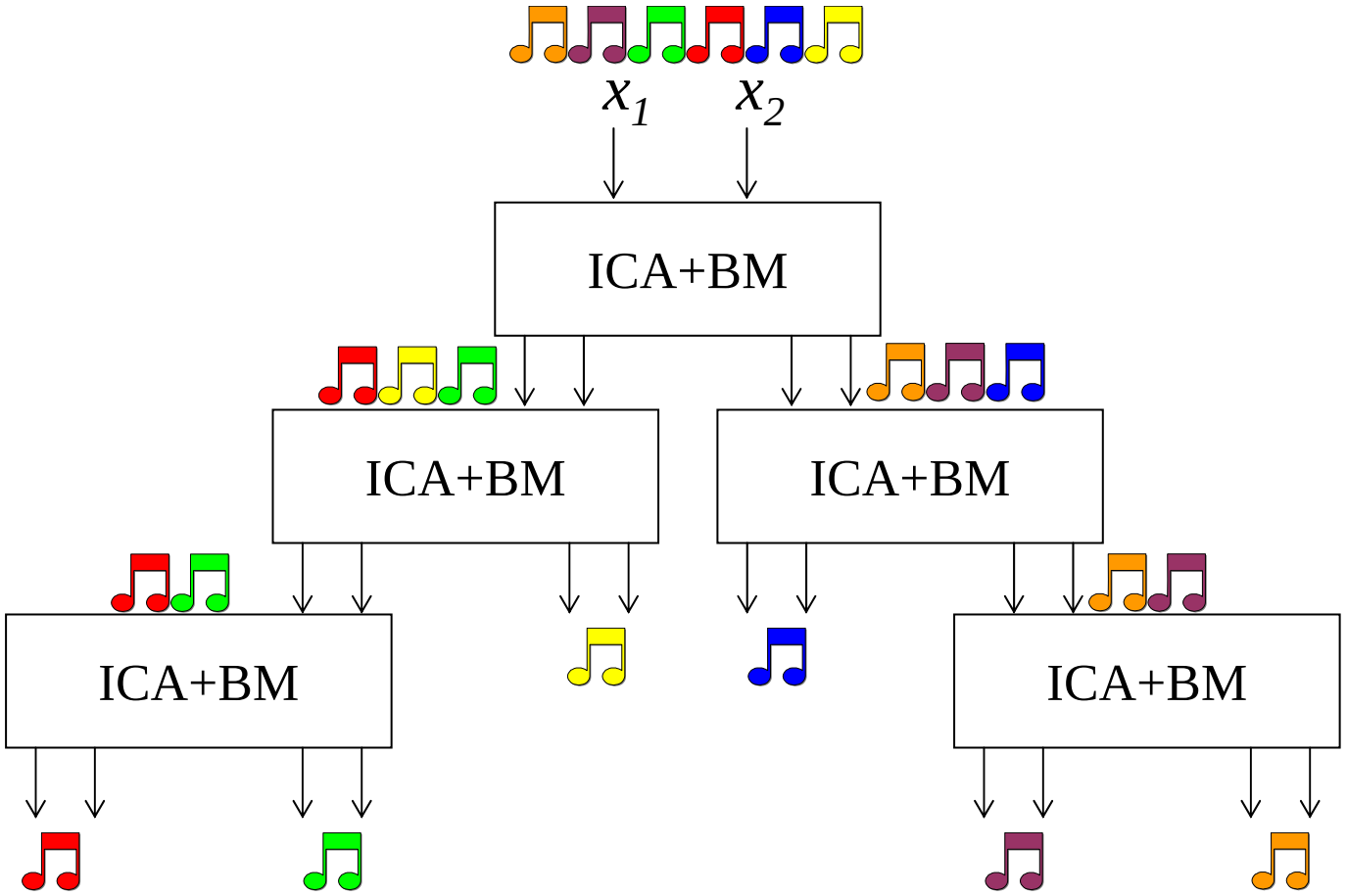
When **W** is applied, the two separated channels each contain a *group* of sources, which is as independent as possible from the other channel.



Combining ICA and T-F masking

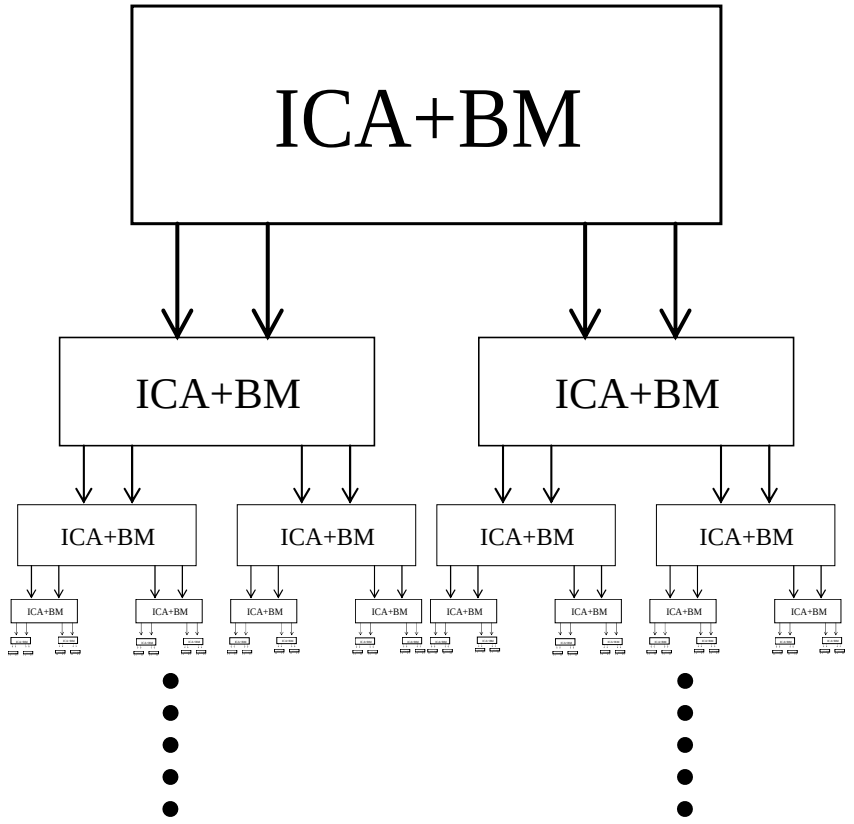


Method applied iteratively

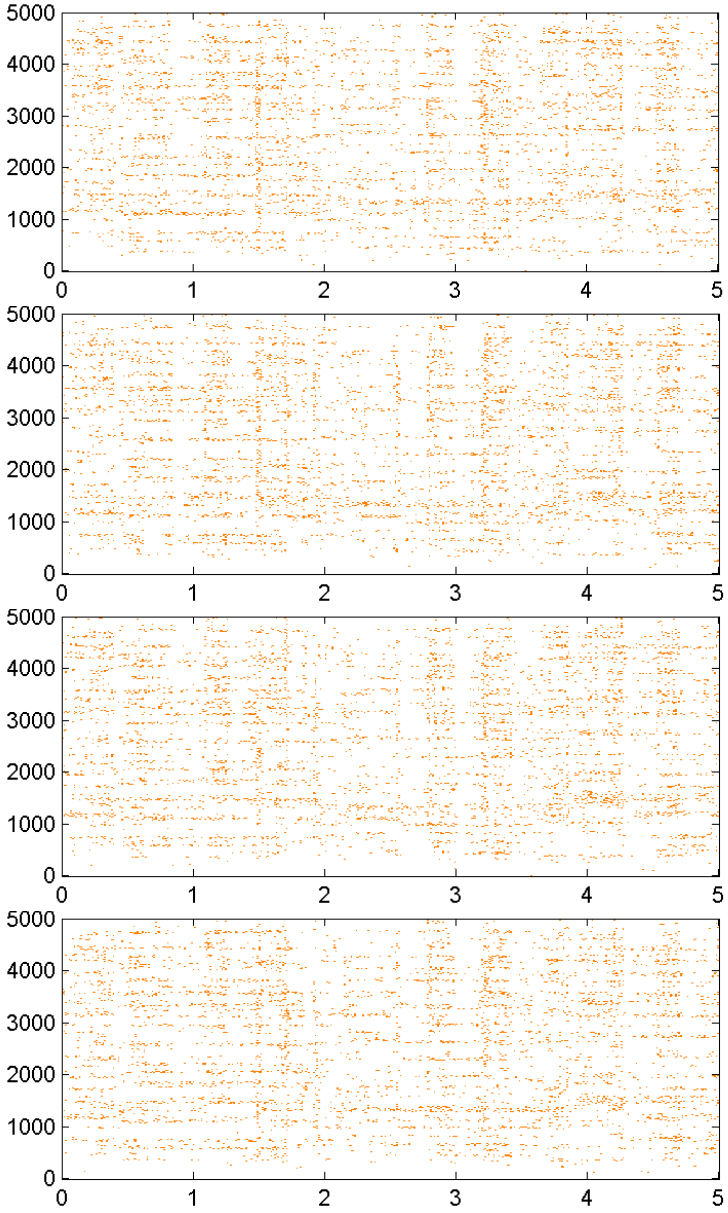


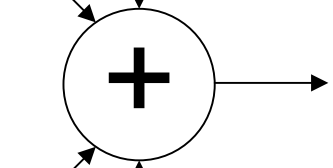
Improving method

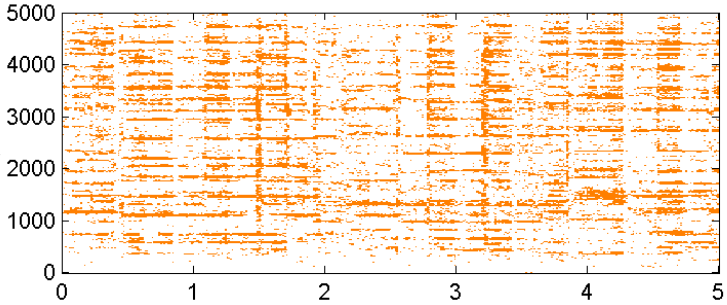
- The assumption of instantaneous mixing may not always hold.
- Assumption can be relaxed.
- Separation procedure is continued until very sparse masks are obtained.
- Masks that mainly contain the same source are afterwards merged.



Mask merging







If the signals in the time domain are correlated, their corresponding masks are merged.

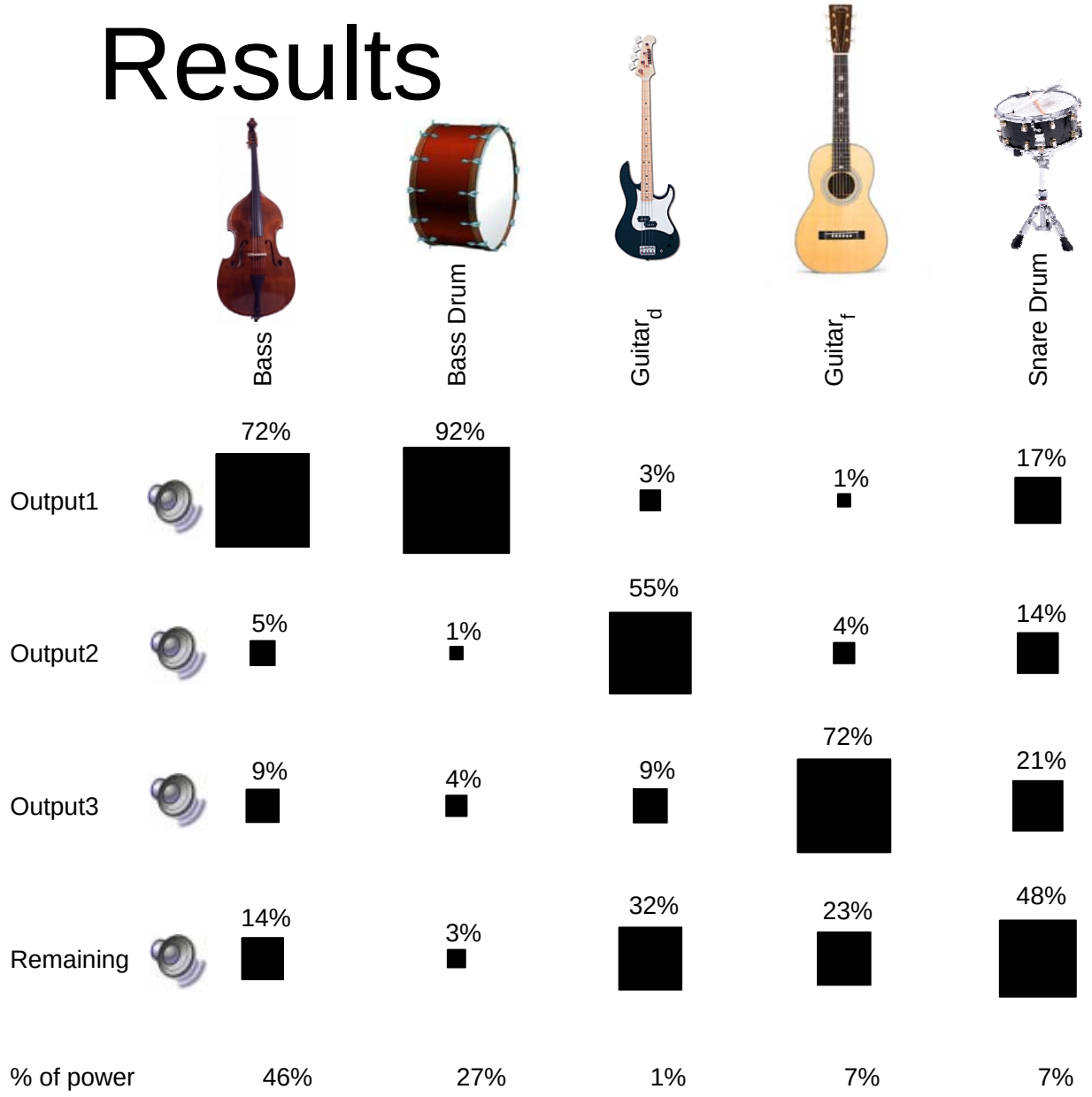
The resulting signal from the merged mask is of higher quality.

Results

- Evaluation on real stereo music recordings, with the stereo recording of each instrument available, before mixing.
- We find the correlation between the obtained sources and the by the ideal binary mask obtained sources.
- Other segregated music examples are available online.

Results

- The segregated outputs are dominated by individual instruments
- Some instruments cannot be segregated by this method, because they are not spatially different.



Michael Syskind Pedersen, IMM, Technical University of Denmark

Conclusion and future work

- We have presented an unsupervised method for segregation of single instruments or vocal sound from stereo music.
- Our method is based on combining ICA and T-F masking.
- The segregated signals are maintained in stereo.
- Only spatially different signals can be segregated from each other.
- The proposed framework may be improved by combining the method with single channel separation methods.