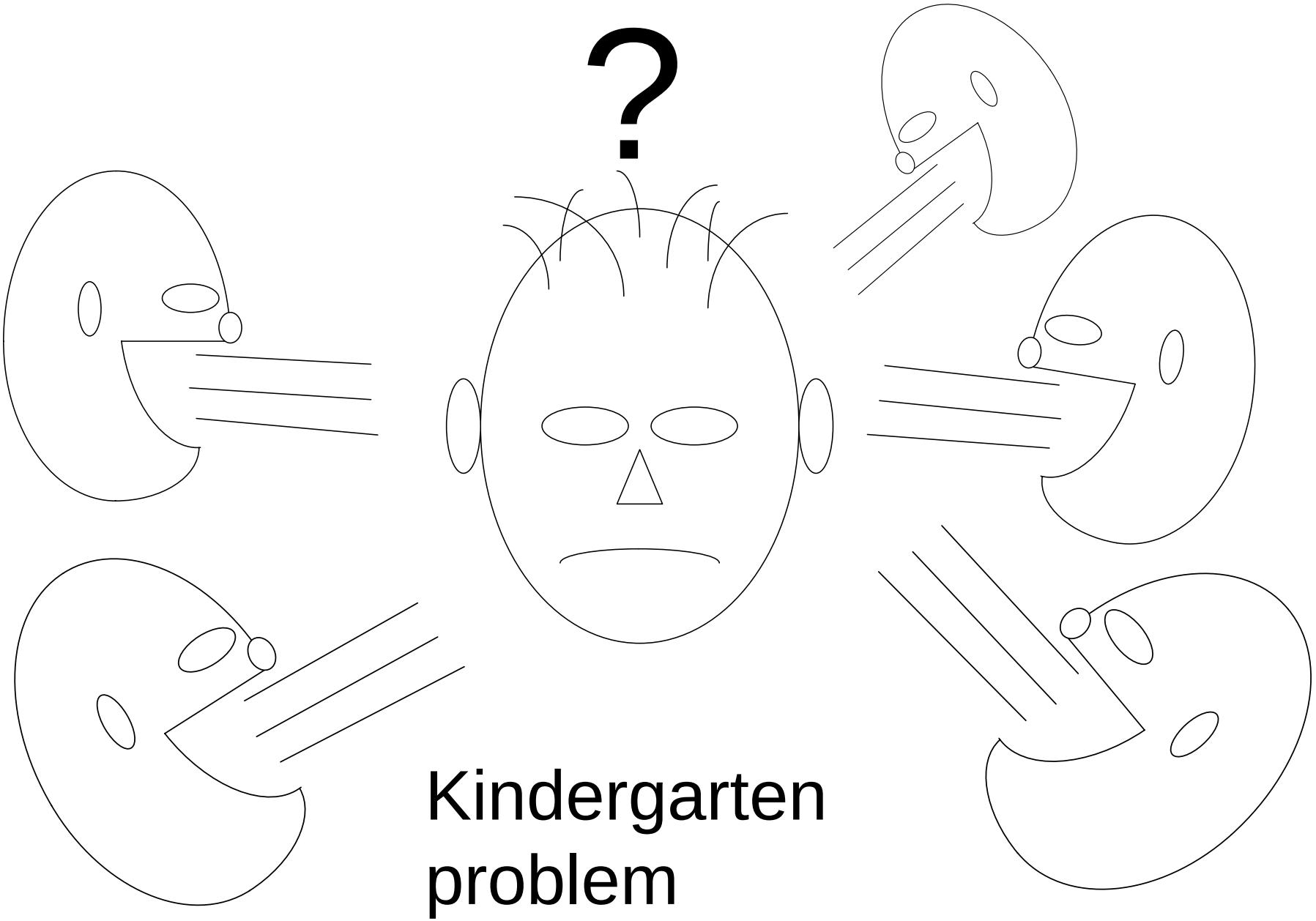
Separating Underdetermined Convolutive Speech Mixtures

Michael Syskind Pedersen
DeLiang Wang
Jan Larsen
Ulrik Kjems
ICA2006, Charleston SC, USA

Outline

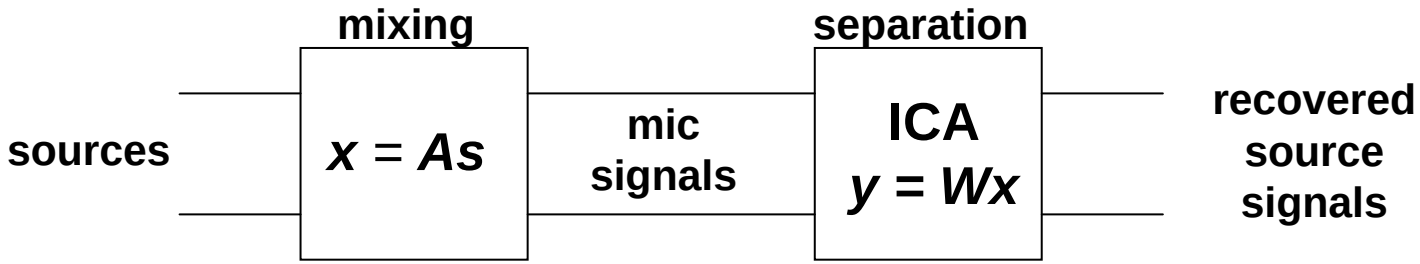
- We propose an unsupervised method for separation of underdetermined convolutive mixtures.
- The method is based on combining two existing methods: ICA and Time-Frequency masking
- The method is an extension of a previously proposed method for separation of instantaneous mixtures.



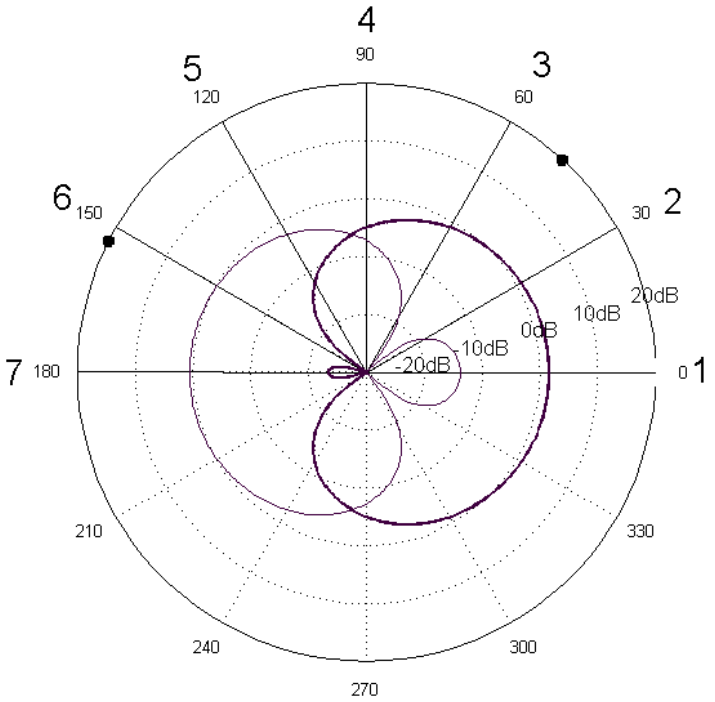
Kindergarten problem

Michael Syskind Pedersen, IMM, Technical University of Denmark

ICA



- We assume that each ICA gain corresponds to a certain direction.
- We apply 2-by-2 ICA to the two recorded mixtures consisting of N sources.
- The parameters in estimated ICA separation matrix is actually an adaptive beamformer.
- Hereby each of the two outputs contain two groups of data, where each group is as independent as possible from the other group.



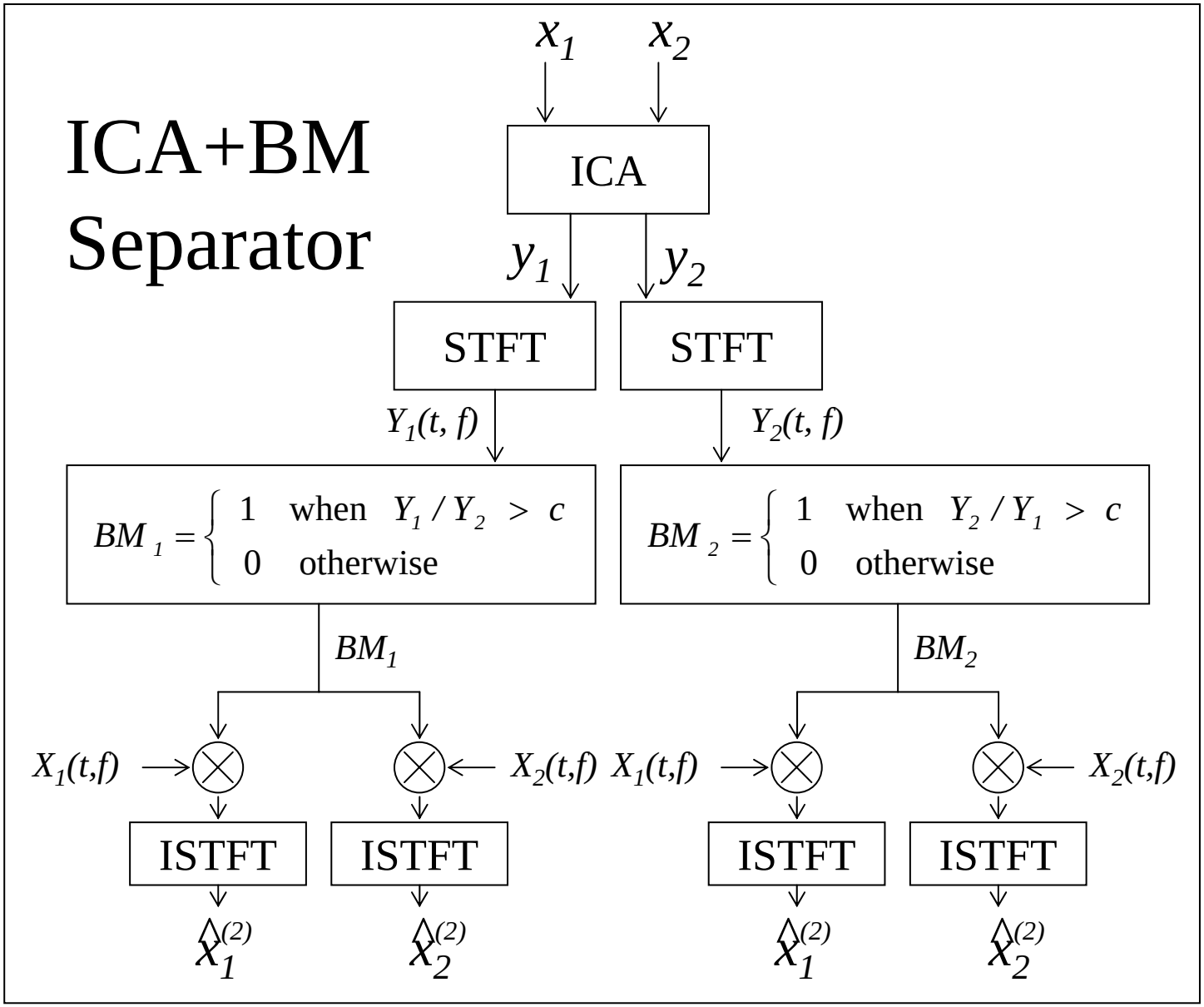
T-F mask

- The T-F mask attenuates T-F units with unwanted sound
- Only T-F units containing the desired sound are attenuated
- Each ICA output amplify a group sources and attenuate another group of sources
- A binary mask that amplifies one group and attenuates another group can be found by simple comparison between the two ICA outputs
- The binary mask is made from a binary decision:

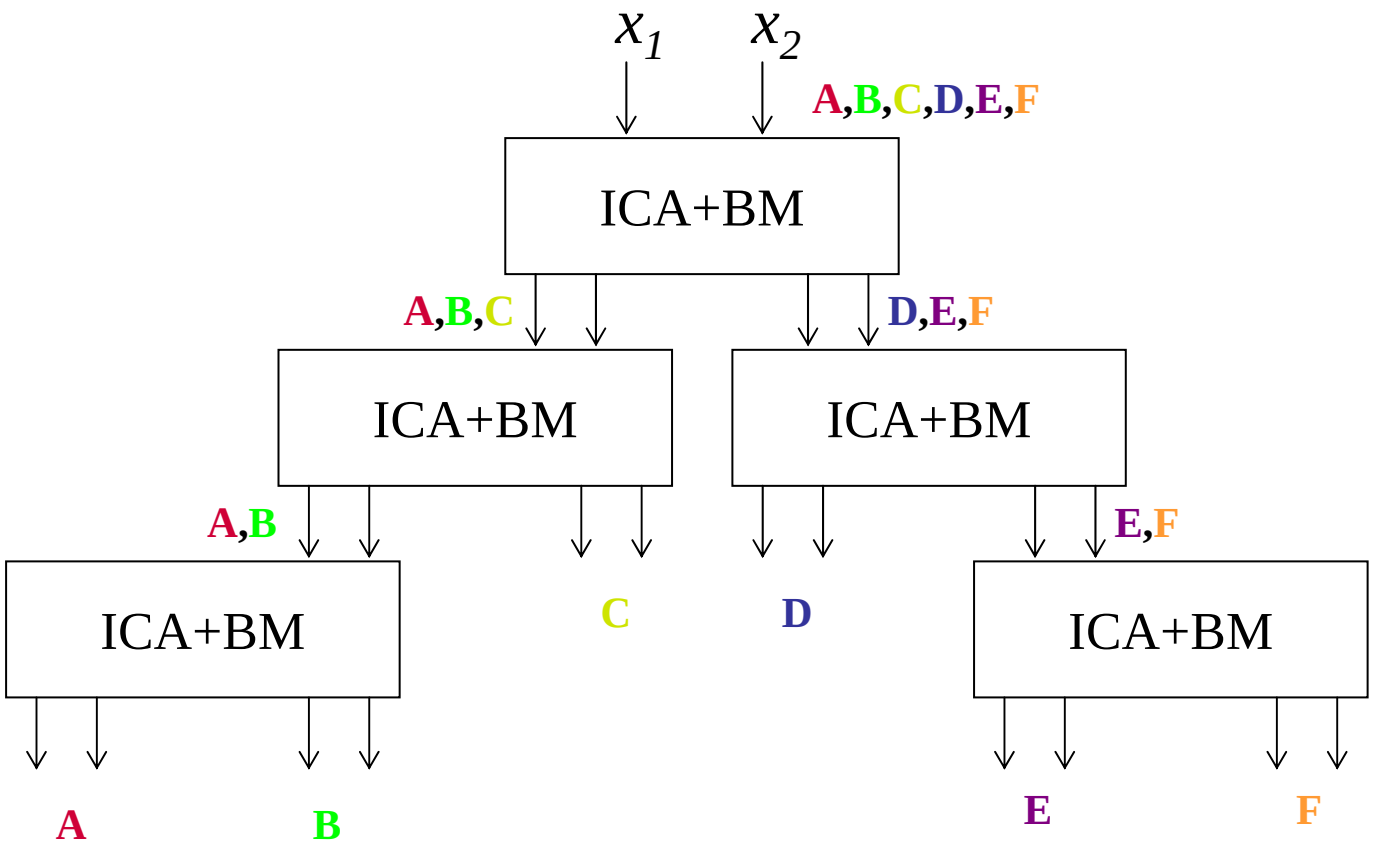
$$BM_1 = \begin{cases} 1 & \text{when } |Y_1(\omega,t)| \ / \ |Y_2(\omega,t)| \ > \ c \\ 0 & \text{otherwise} \end{cases}$$

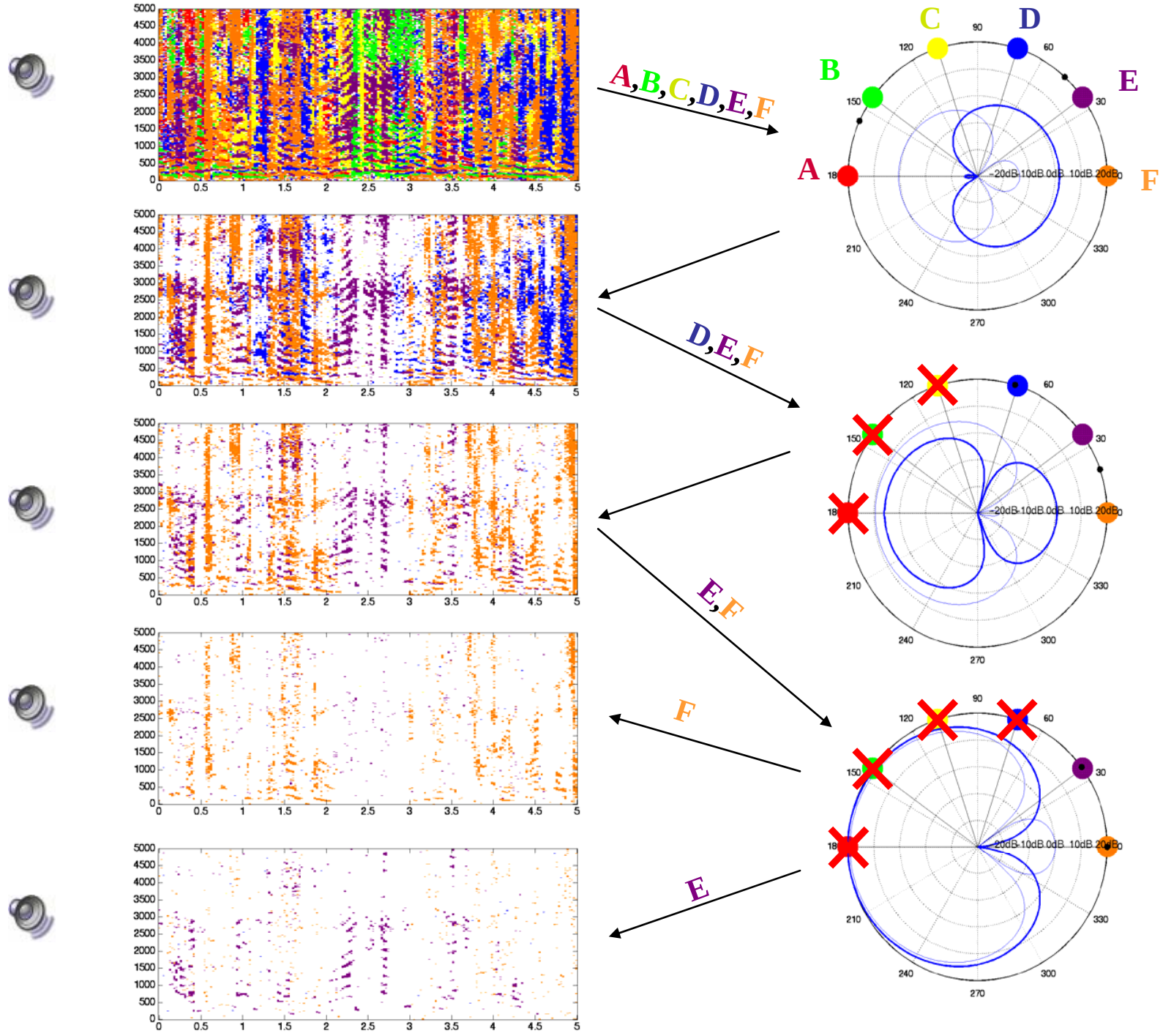
$$BM_2 = \begin{cases} 1 & \text{when } |Y_2(\omega,t)| \ / \ |Y_1(\omega,t)| \ > \ c \\ 0 & \text{otherwise} \end{cases}$$

ICA+BM Separator

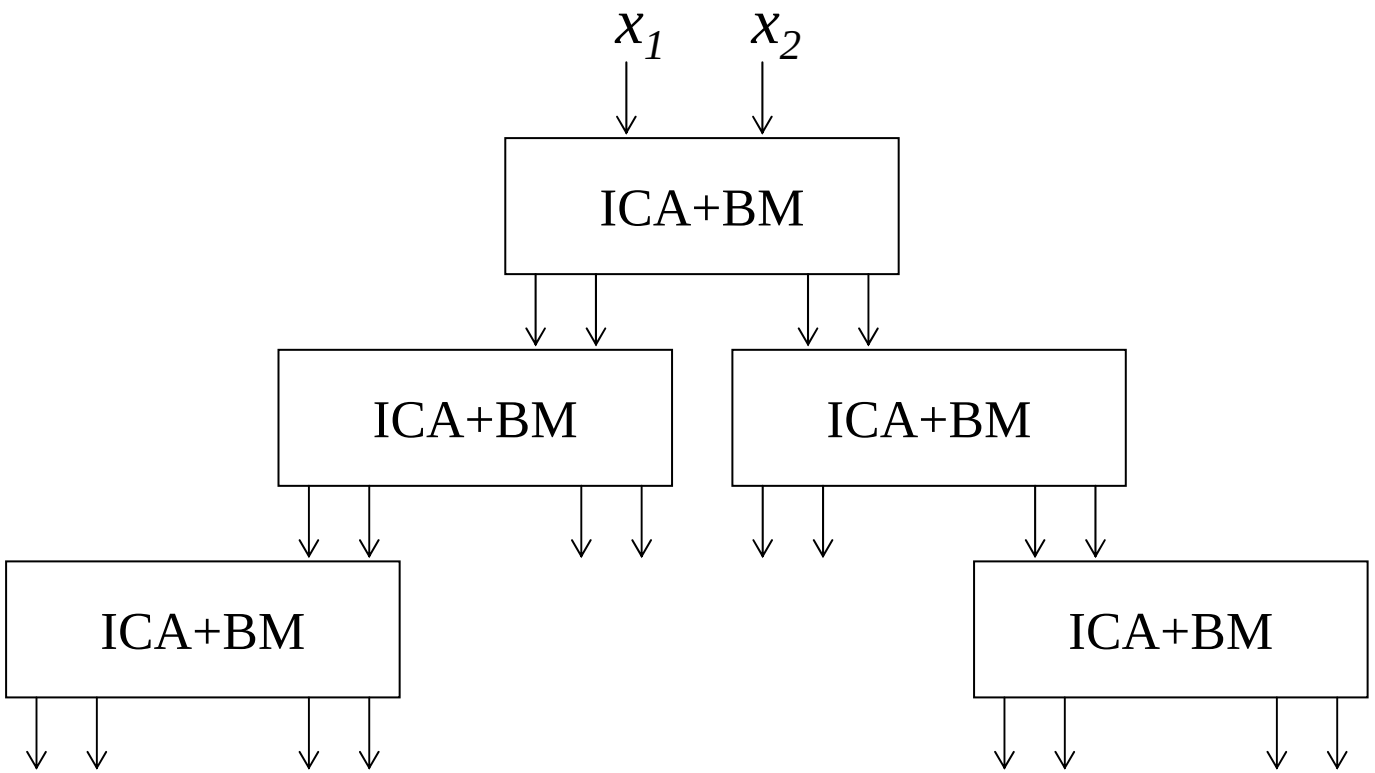


Combined Algorithm

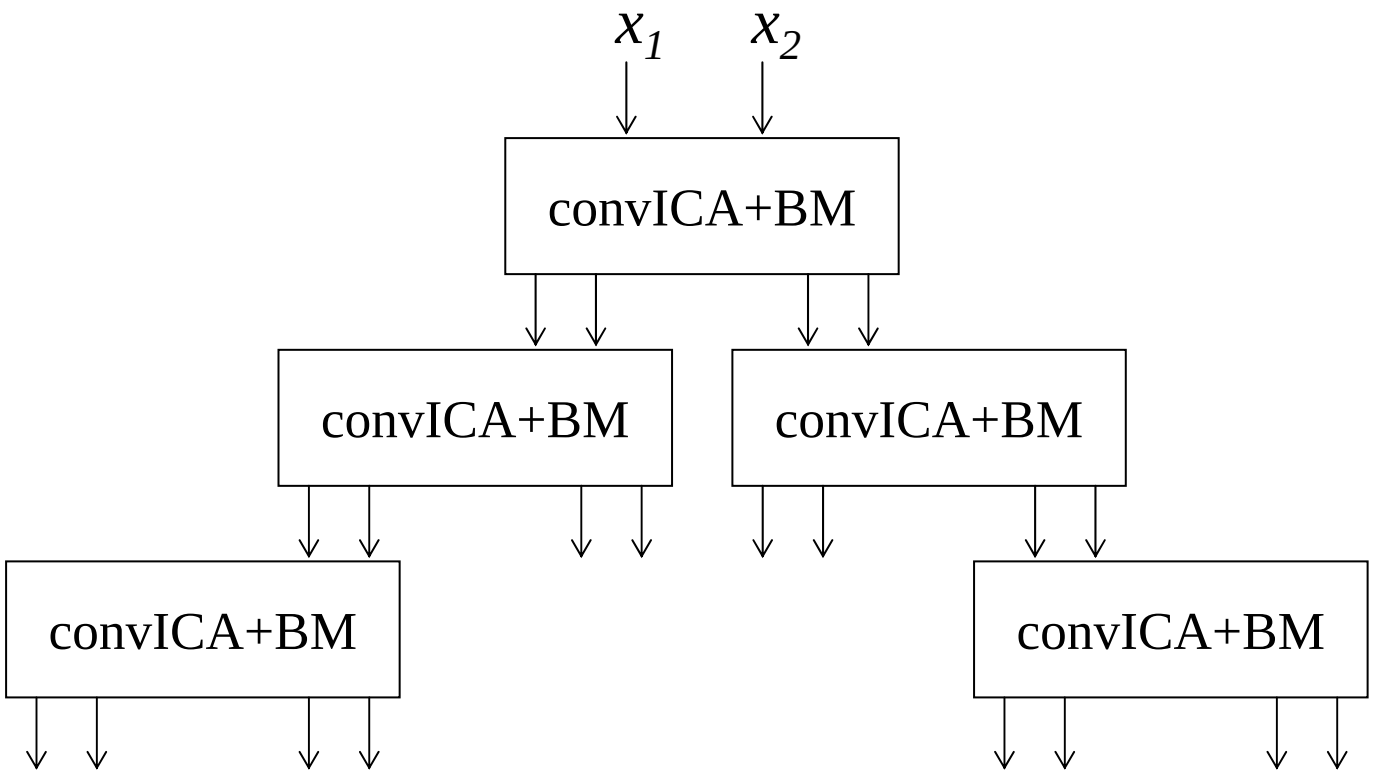



A,B,C,D,E,F
D,E,F
E,F
F
E
A
B
C
D
E
F
-20dB -10dB 0dB 10dB 20dB

Extension to convolutive mixtures



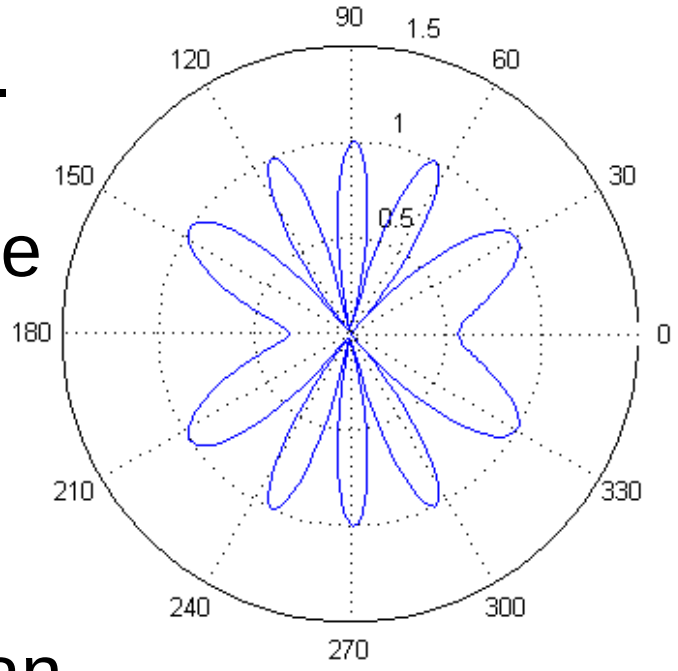
Extension to convolutive mixtures



We used the frequency domain algorithm by [Parra and Spence 2000], because the source code is available

Comments on extension

- It is assumed that the segregated groups mainly contains different sources.
- Spatial aliasing due to a larger microphone distance may deteriorate the mask estimation
- Depending on the chosen convolutive algorithm frequency permutations can occur.

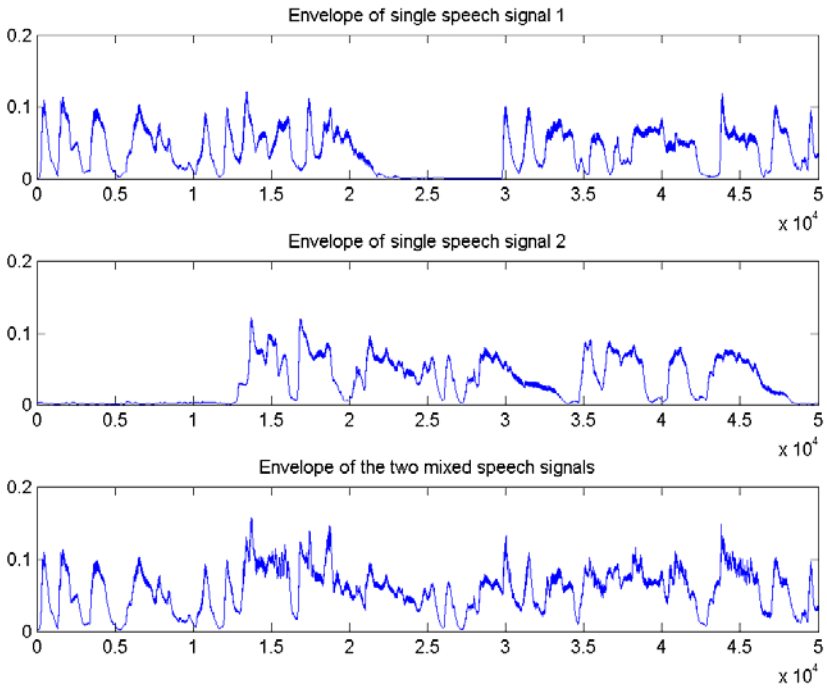


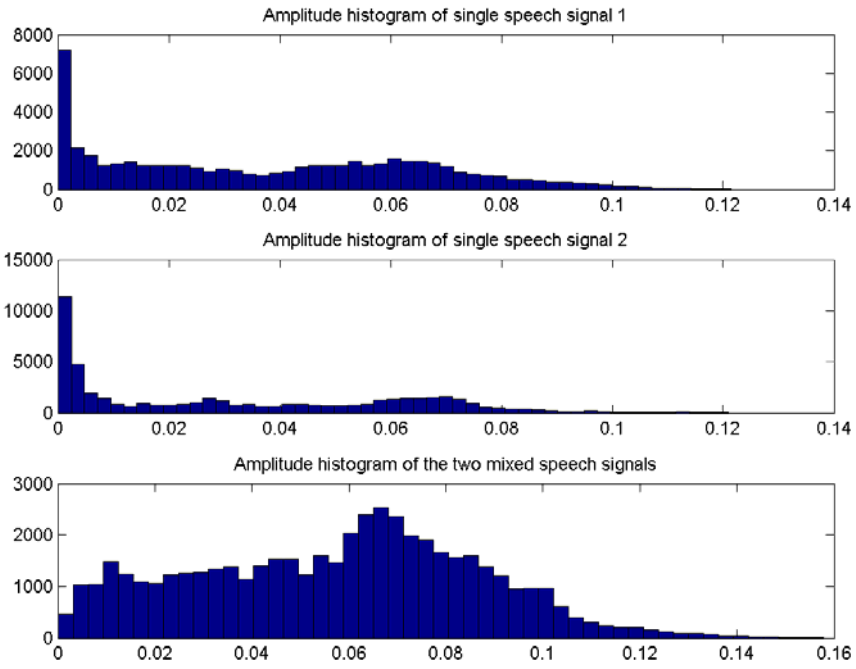
When to stop

For each output signal a decision has to be made:

1. **The mask has removed all sources but one**
2. **The mask has removed too much**
3. **The output contains more than one source**

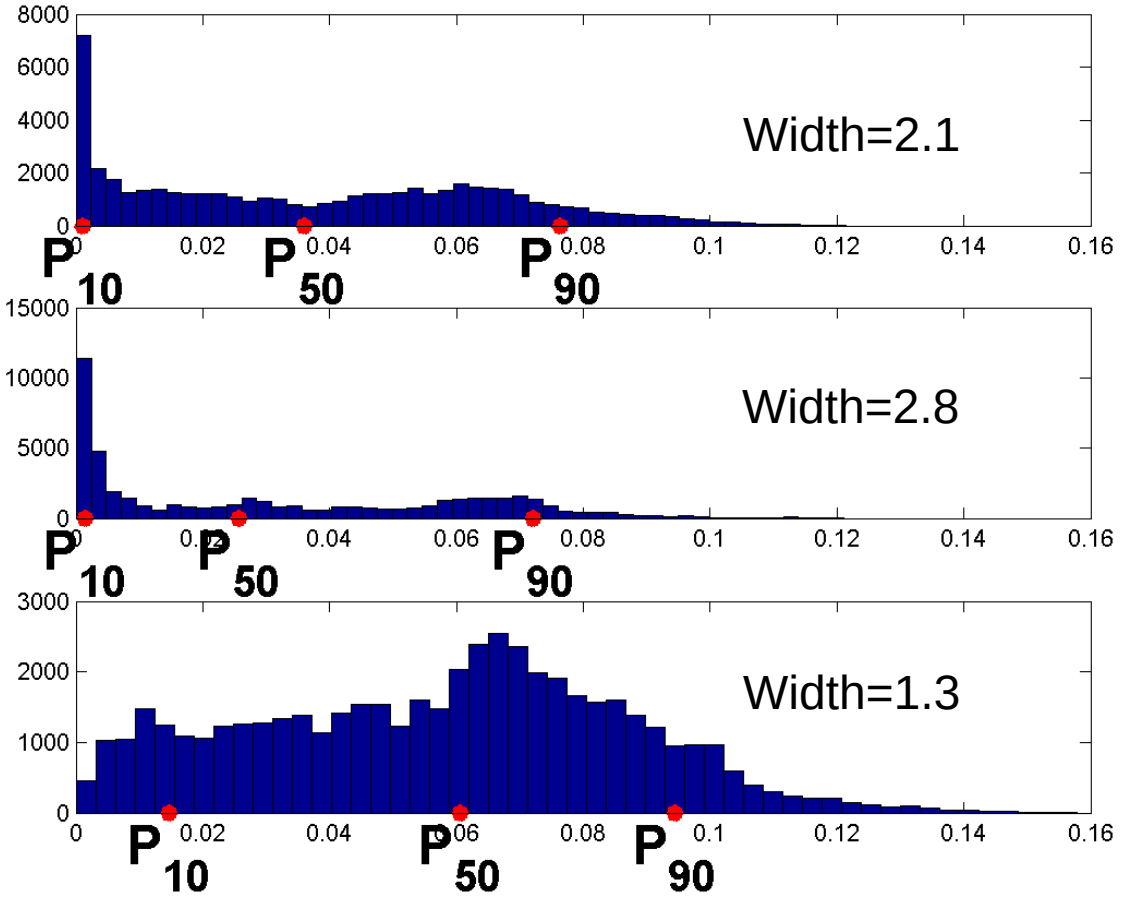
Discrimination between single speaker and more than one speaker





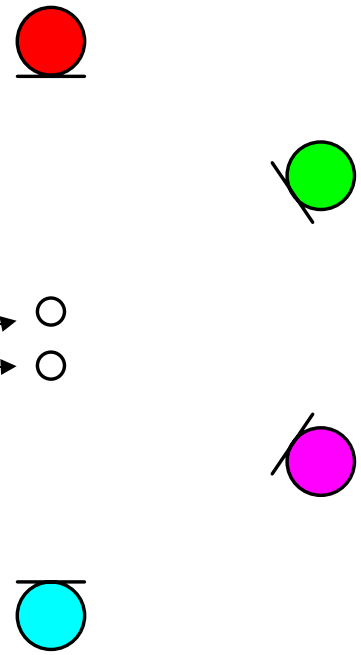
- A single signal has a zero amplitude more frequently than a mixture of two signals.
- The “width” of the histogram can be used to discriminate between one or more than one speaker in the mixture [Büchler 2002].

$$\text{width} = \frac{P_{90} - P_{10}}{P_{50}}$$



Evaluation

- Simulated room recordings
- 6 x 4 x 3 m
- T_{60}=160 ms
- 20 cm



- Separation of four mixtures
- Average SNR gain of **6.4** dB
- Segregated signals are maintained as stereo signals

Conclusion

- We have presented a method for segregation of underdetermined convolutive mixtures.
- The method applies convolutive ICA and T-F masking iteratively, until each mask only contains a single source signal.
- Separation of four speakers from mixtures recorded at two microphones.

Future work

- Merging of masks
- Merging decision based on correlation between the signal envelopes.

Sound example