Institute of Sound
and Vibration Research

UNIVERSITY OF Southampton

A blind source separation approach for humpback whale songs separation

Zhenbin Zhang, **Paul White**

Institute of Sound and Vibration Research,
University of Southampton, UK

Acknowledgements: Olivier Adam, Ceta mada

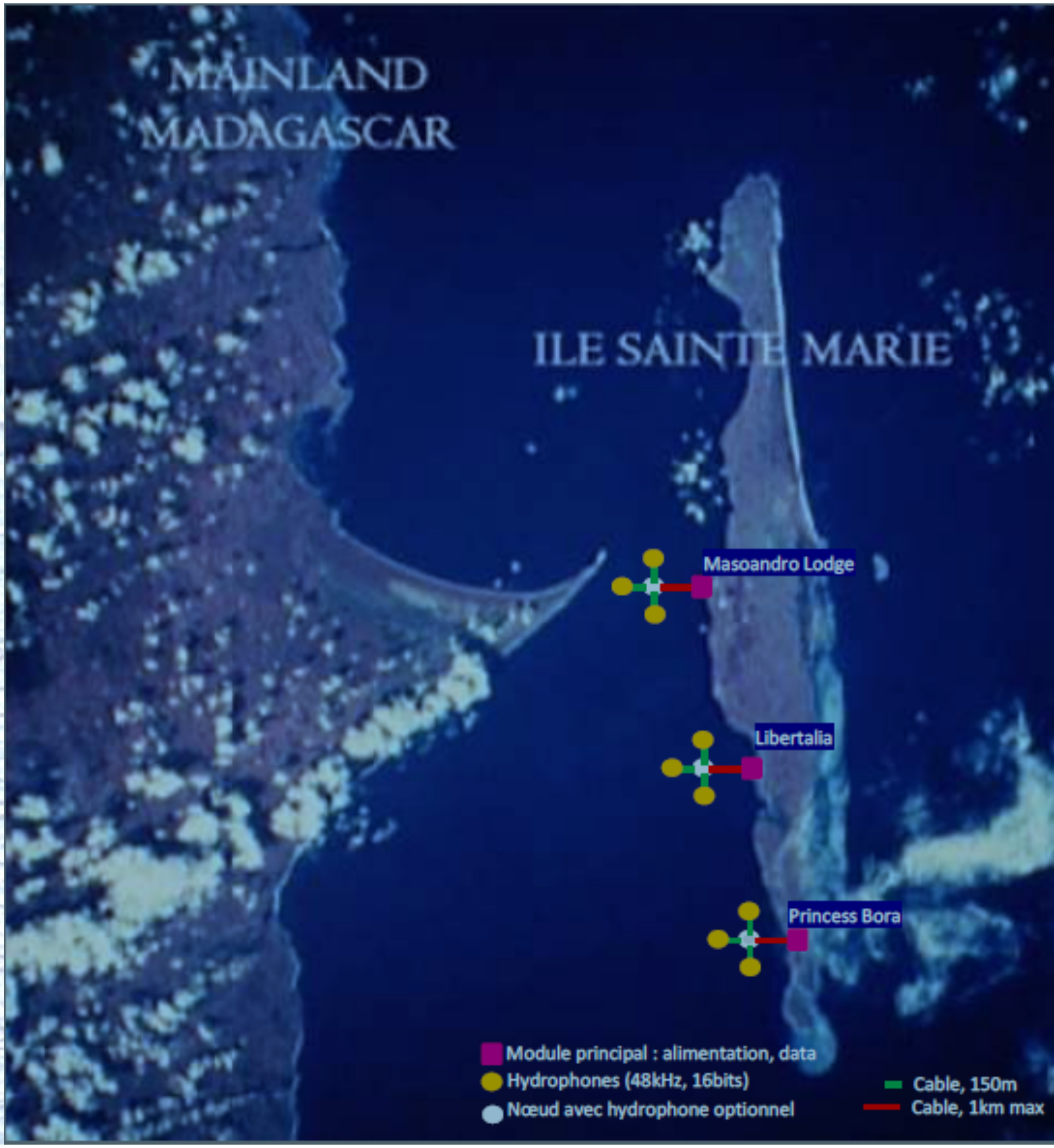
Outline

- Background and objectives
- Method
  - Principles Underlying Blind Source Separations
  - The method we employ
- Results
- Conclusion

Background

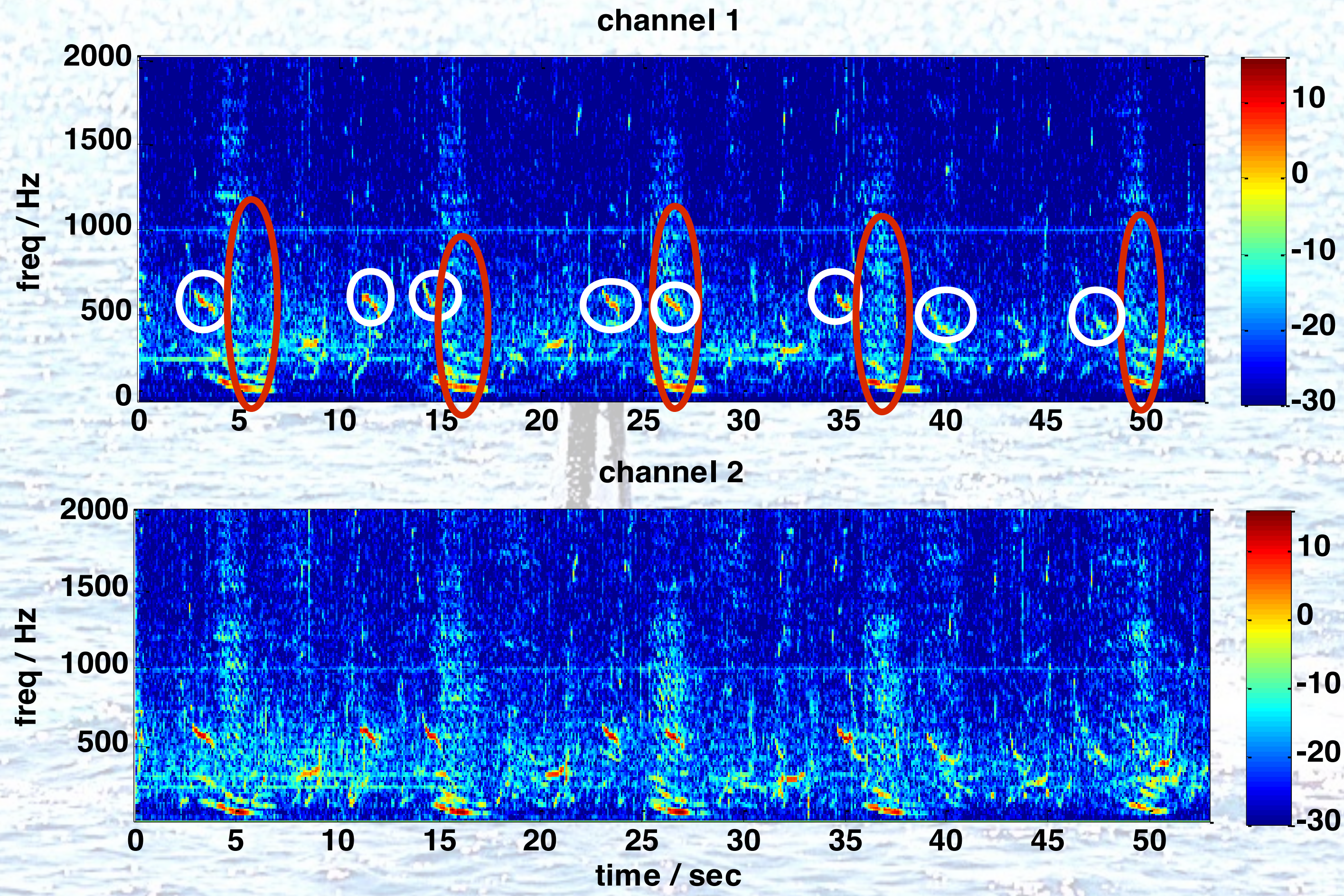
- We are interested in automated methods to assist with the study of the songs of humpbacks in the Ste Marie channel, Madagascar.
- In this area there is a high density of singers.

The Fixed Acoustic Arrays



The Problem

- Recordings always (?) contain the song of many singers – we would like to take such recordings and decomposed them into the songs of individuals.



The Human Equivalent

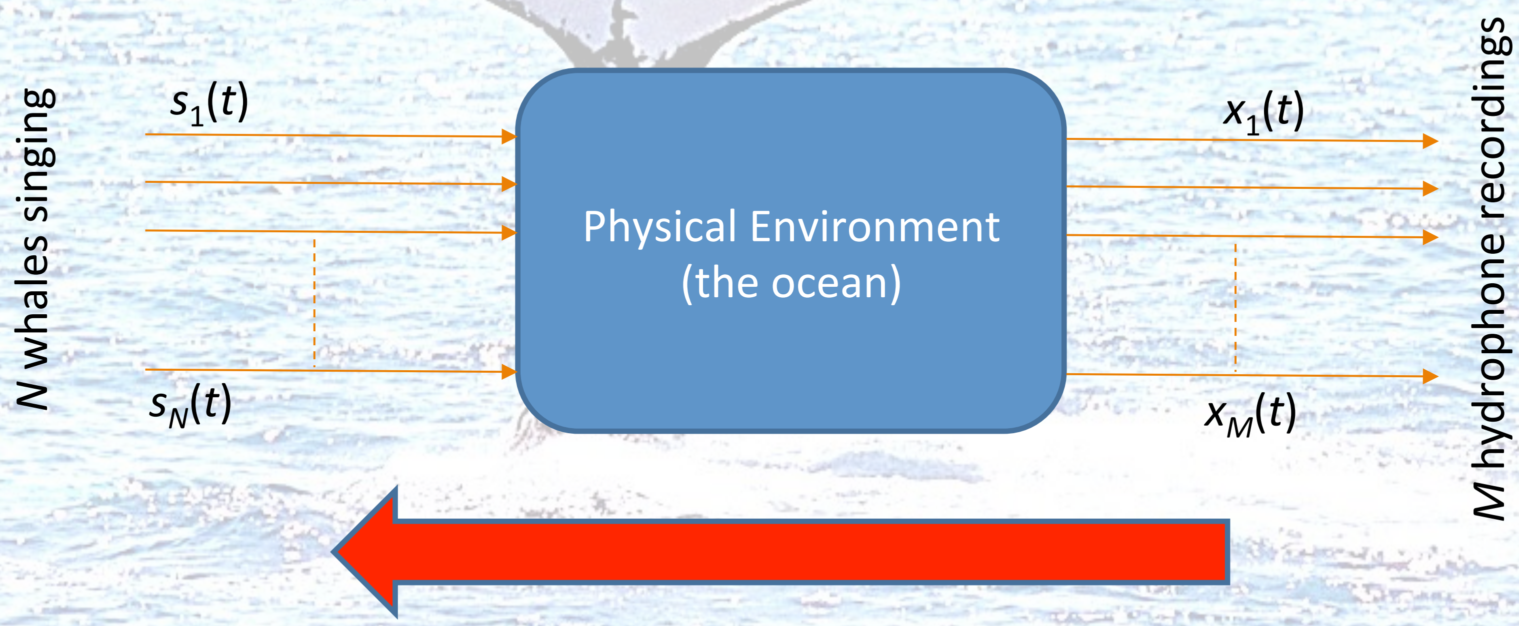
- In a room with many speakers talking at once we are able to focus on one speaker at a time.
- This is "easy" for us to do…. much harder to get a computer to achieve.
- Is generally called the "cocktail party" problem.

- Focus of much recent signal processing interest
  - As you will see it is a long way from being completely solved!

Blind Source Separation

- What is Source Separation?
  - A definition of source separation might be:

    Given a set of N measured signals $x_m(t)$ (m=1,…,M) can one estimate the N source signals, $s_k(t)$ (k=1,….,N) that were combined to form the measurements?



We want to go from the measurements back to the sounds from each whale

Blind Source Separation

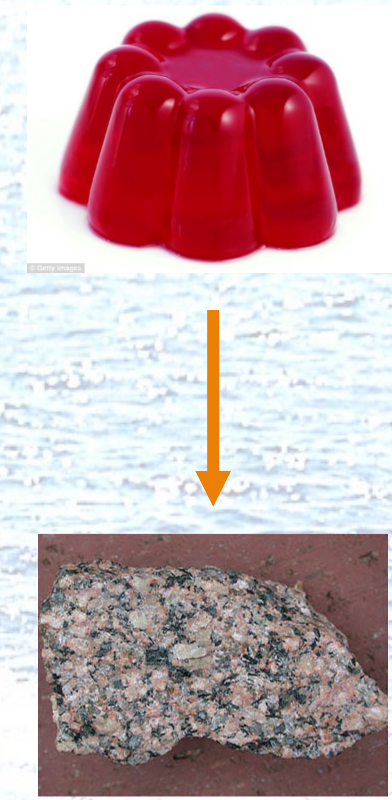
- Why blind?
  - We are assuming that we know neither the properties of the environment nor the properties of the sounds (actually we do rely on some very general properties).
- This blindness means we don't need to assume the hydrophone locations, where the whales are, ocean properties or the frequency content of the whale vocalisations.
- This makes the method potentially powerful, but also challenging!

The Key to Solving a BBS Problem is Statistical Independence

- We assume the whales sing independently.
- The measurements of the song will not be independent (in general).
- If we can process the data to restore statistical independence then we can claim to have unmixed sources.
  - There will remain the amplitude and permutation uncertainties.
- Such methods, based on independence, are called Independent Component Analysis

Grades of BSS Problem

- Various algorithms exist, based on different complexities of the model for the environment:
  - Instantaneous mixtures
    - Assumes **no** delays between measurements
    - Not suitable for acoustics, good for electromag.
    - Easiest problem, largely already solved.
  - Anechoic
    - Assumes **only** delays and gains.
    - Maybe suitable for simple acoustic environments
  - Convolutive
    - Includes reflections, so models realistic acoustic scenarios.
    - Difficult to solve.



The Simplest BSS Problem

- Consider the following problem

Unknown signals

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$
$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

Measured signals

Unknown multipliers

- We only know the x_i, not the a_{ij} or s_i
  - a seemingly intractable problem?
  - Assuming independence of s_i then we can solve this problem to within a couple of ambiguities.

Ambiguities

- If we can find a solution then there are two insurmountable ambiguities.

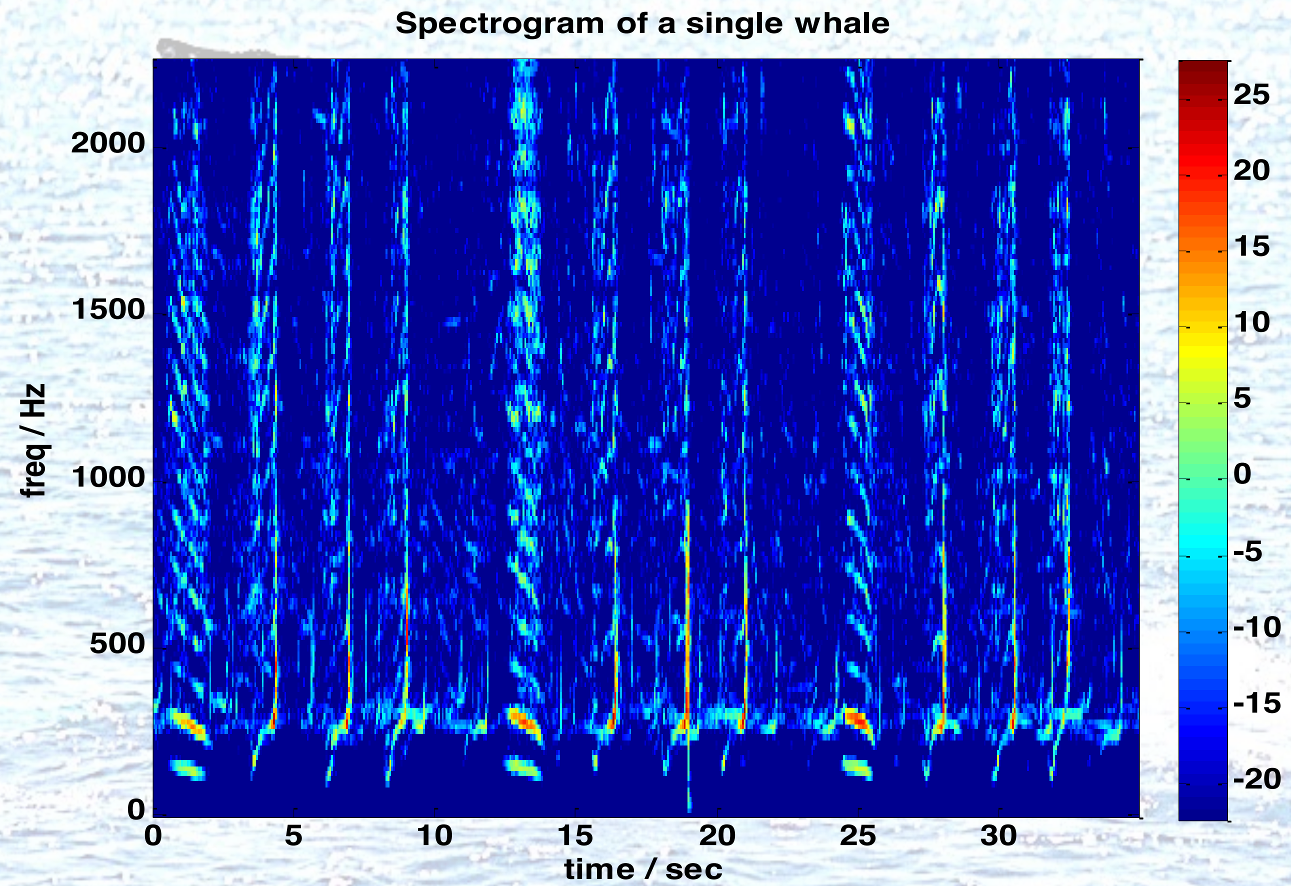
1) Permutation ambiguity
   - The solutions $\hat{s}_1 = s_1, \hat{s}_2 = s_2$ and $\hat{s}_2 = s_1, \hat{s}_1 = s_2$ are equally valid.

2) Scale ambiguity
   - The solutions we can half the *a* coefficients and double the estimated signals and still have a valid solution.

- For our method the permutation ambiguity represents the greater challenge.

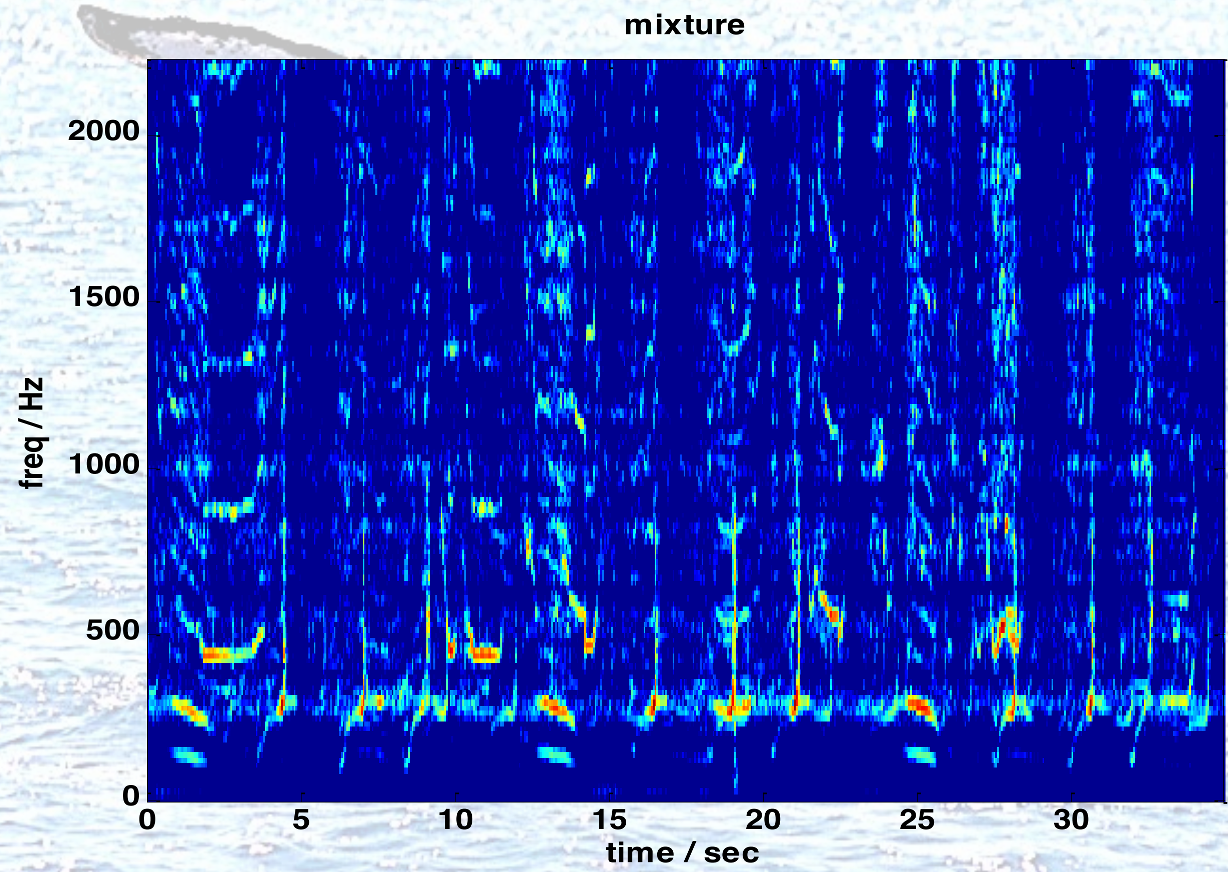
Sparsity in Time and Frequency

- The solution we adopt relies on the idea of sparsity.
- In the time-frequency domain the song of one whale occupies only a small proportion of the TF bins.



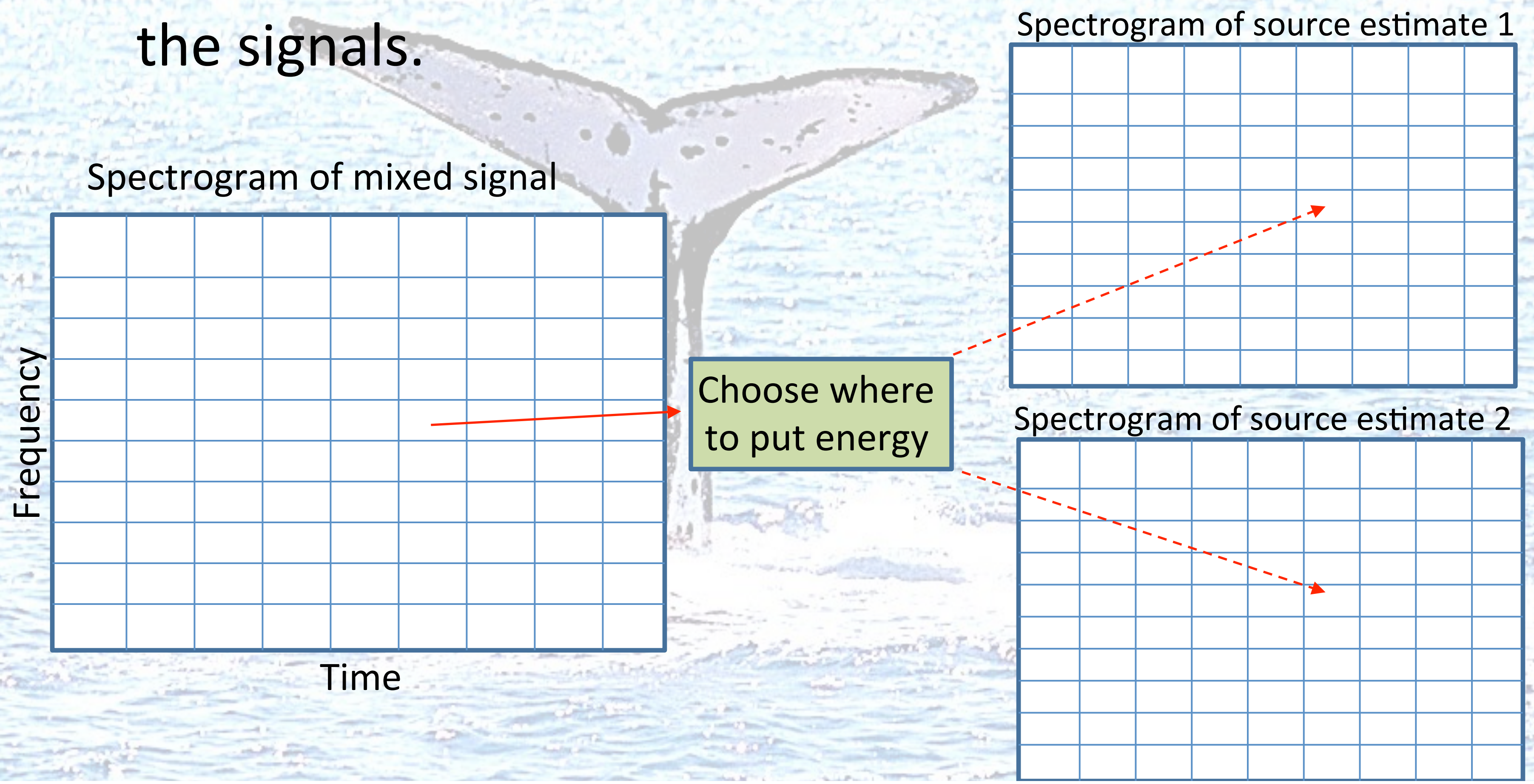
W Disjoint Assumption

- We assume that at each point in time-frequency one energy from one whale dominates over the others.



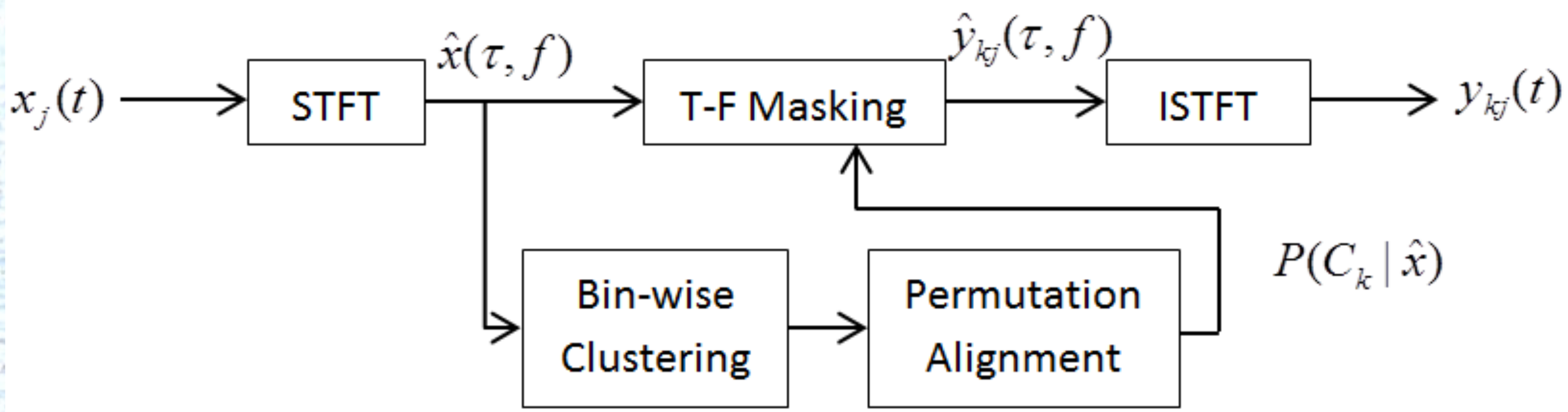
Time Frequency Masking

- Because the signals are W-Disjoint then one can aim to allocate each TF bin of the **mixed** signal to one of the signals.



The **Sawada** algorithm

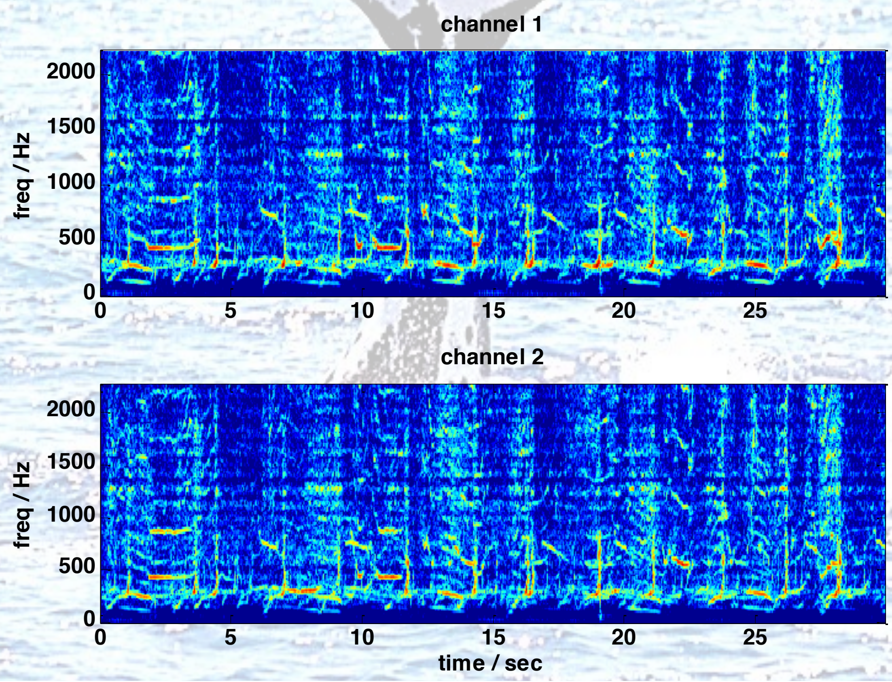
- It performs separation via the bin-wise clustering and permutation alignment.



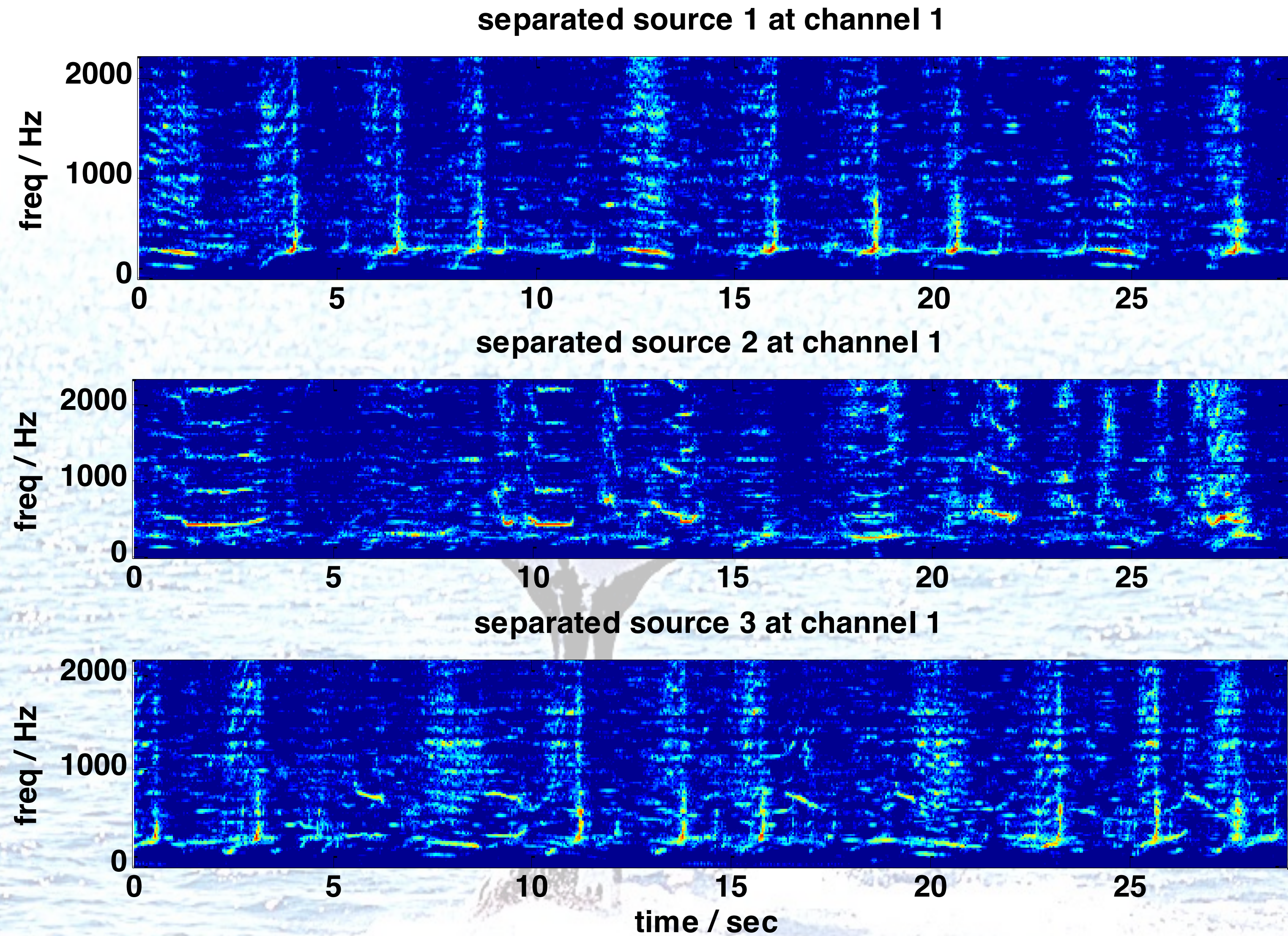
The Sawada algorithm system block diagram

Simulated Data

- We take a pair of high quality HW song recordings.
- Use a standard underwater acoustic propagation model (RAM) to simulated the environment.
  - Assuming: a constant depth, homogeneous sound speed, and a sandy seabed.
- From that we generate impulse responses assuming 3 locations for the sources (whales) and 2 hydrophones.

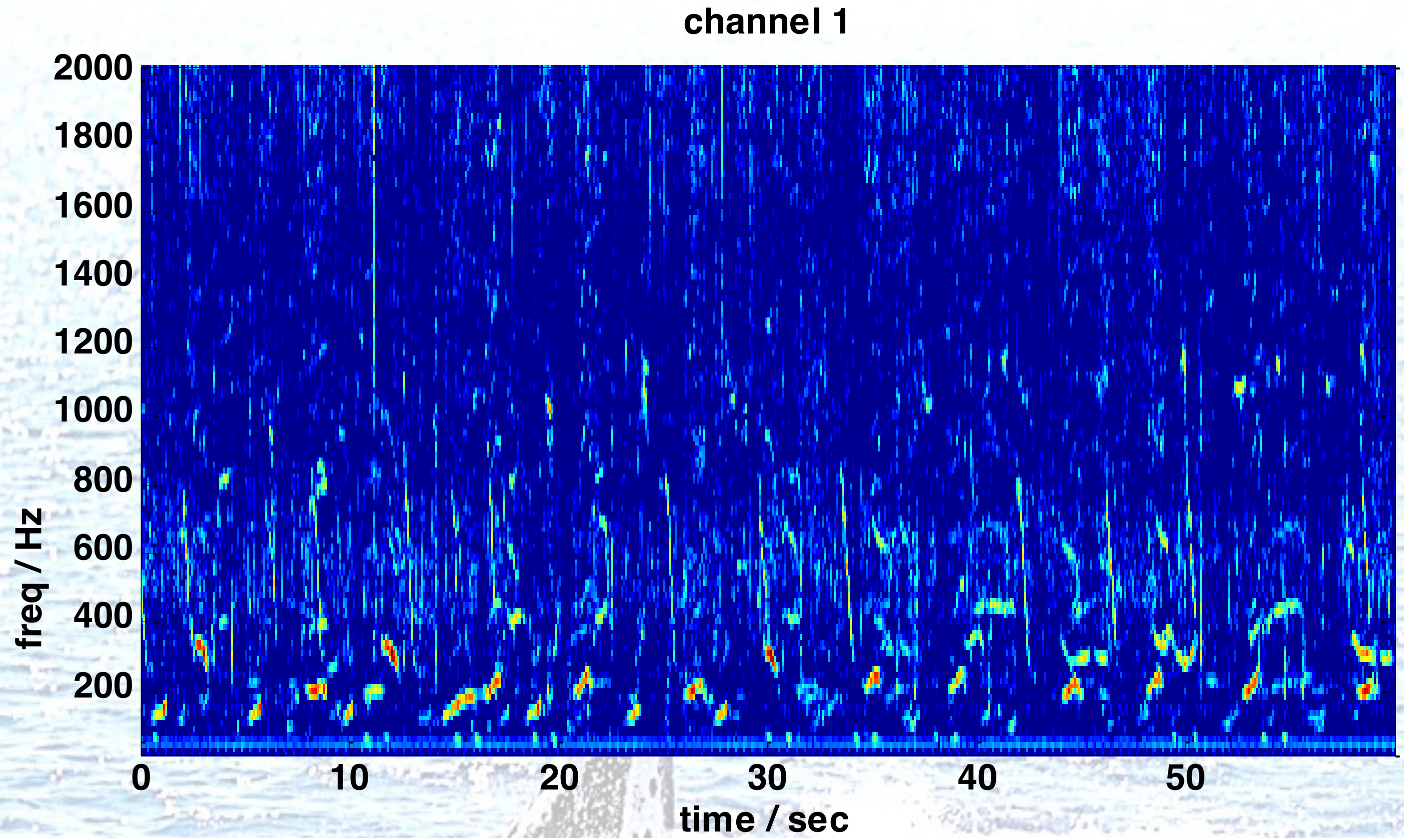


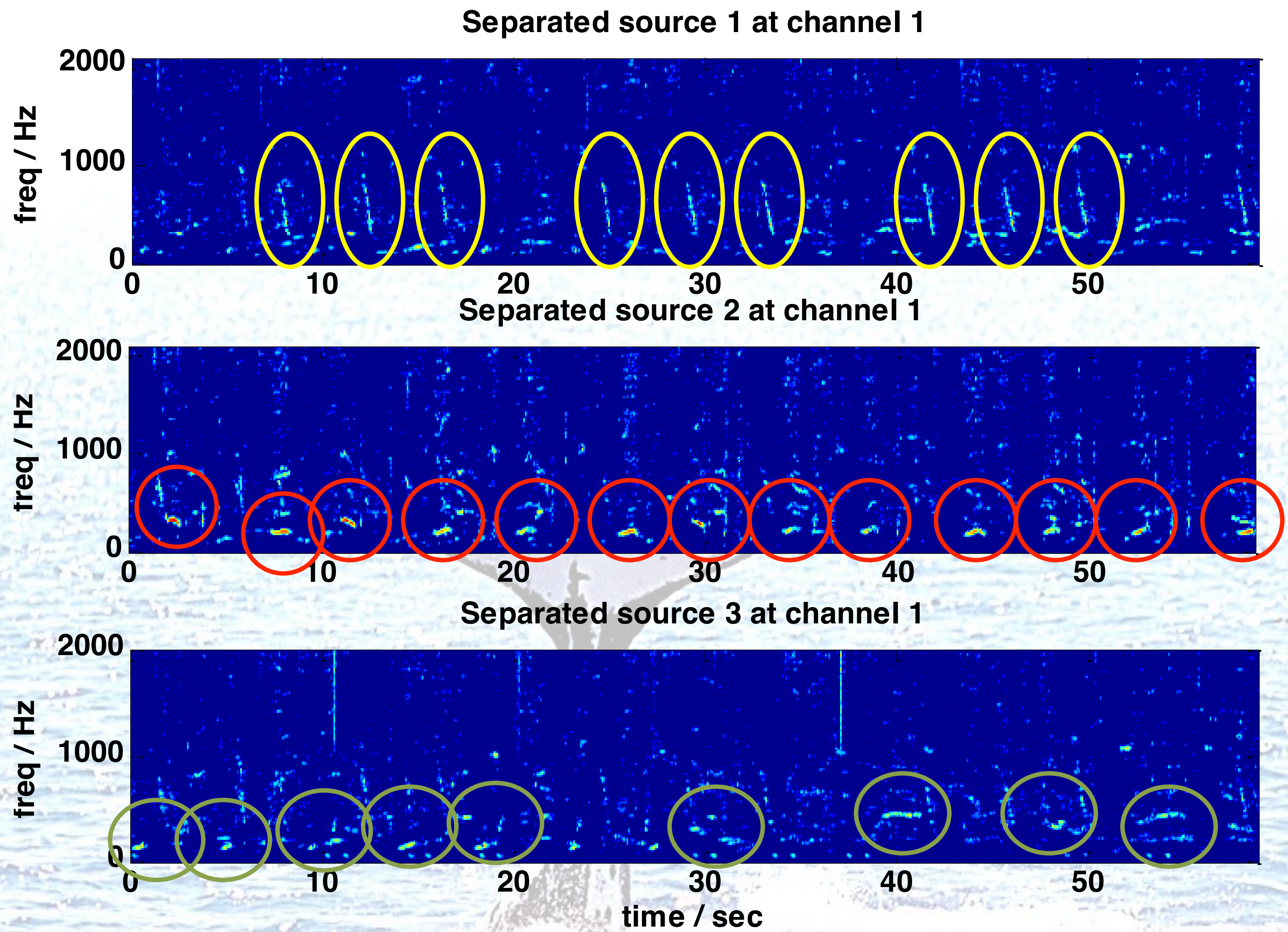
Under-determined



The separation performance is evaluated in terms of the signal-to-distortion ratio (SDR).

Real recording from Madagascar





Due to the fact that the true source information is unavailable, the SDR cannot be evaluated.

Conclusion

- The Sawada algorithm combined with noise reduction significantly improves source separation performance of real recording in severe noise environments.
- We are currently working on automatically estimating the number of sources, and will continue focus on that in the future.

References

- [1]   R. Payne and S. Mc Vay, Songs of humpback whales, Science, pp. 585-597, 1971.
- [2]   H. Sawada, S. Araki, and S. Makino, Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment, Speech Communication, vol. 53, pp. 606-621, May 2011.

Thank you!

Contact email address:
zz1g12@soton.ac.uk