Estimation of Marine Mammals Using Recordings from One Microphone

Xanadu Halkias & Dan Ellis
Laboratory for Recognition and Organization of Speech and Audio
Dept. Electrical Engineering, Columbia University, NY USA
{xanadu,dpwe}@ee.columbia.edu

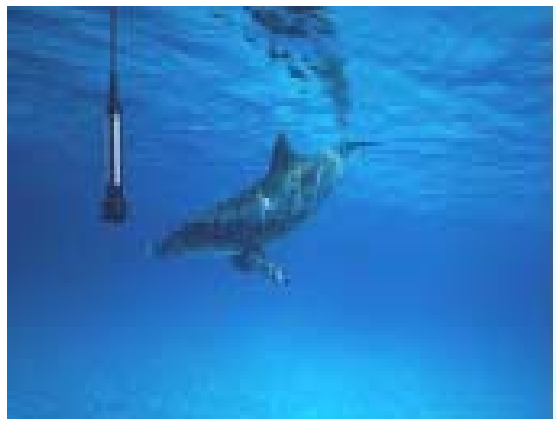
1. System Organization
2. Processing
3. Experiments
4. Conclusions

Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

1. System Organization

- Localizing marine mammals
  - Track their location and population
  - Preserve and protect
- Systems used require an array of hydrophones e.g. M3R
  - Cross-correlation along all hydrophones
- Accessibility to the hobbyist who just wants to have a crude idea



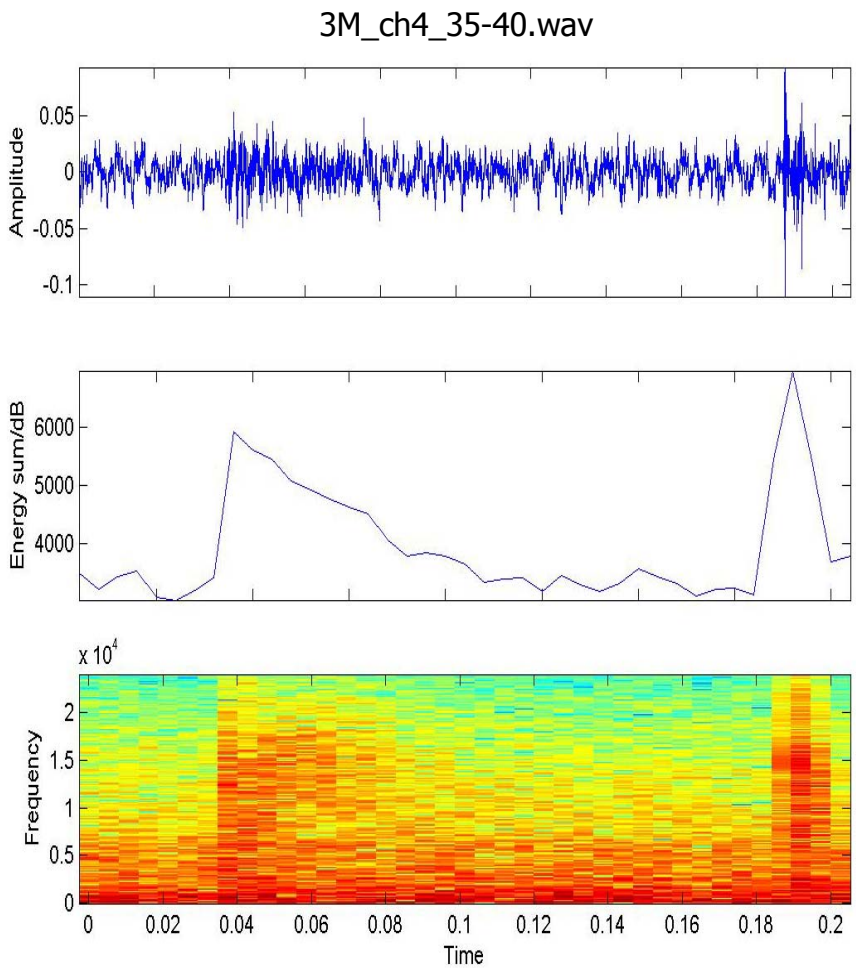
Estimation of the number of marine mammals

- Task:

  Automatically and in real time extract the number of marine mammals in a region given their audio recordings
  - Multiple hydrophones
  - Manual interference for elimination of clicks or whistles in the recordings
- Goal:
  - Provide a good approximation of the number using click recordings from one microphone

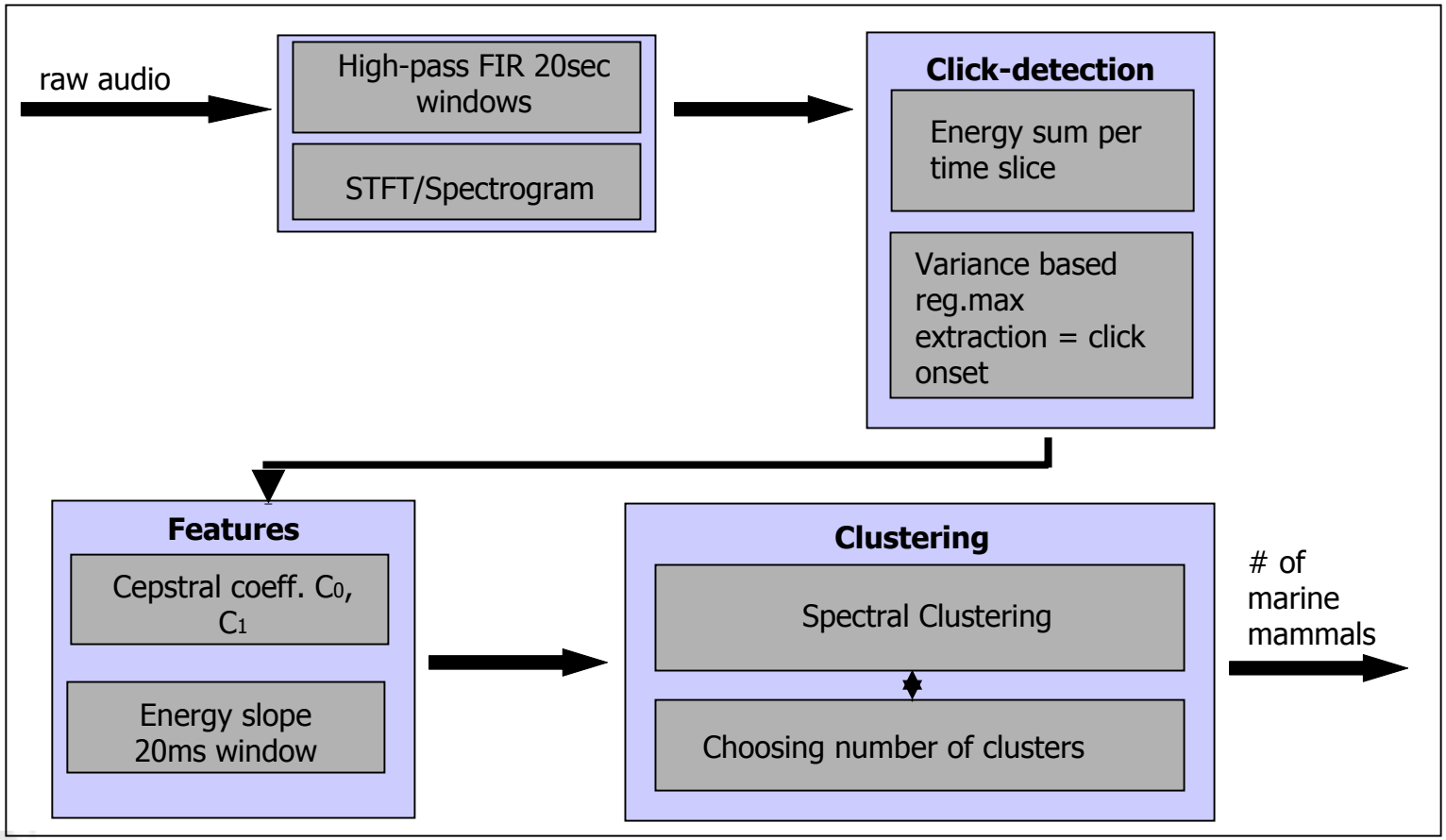
The Data - Click Examples

- Click sounds from sperm and pilot whales
- Several hydrophones
- Ground truth of possible whales available through M3R and visual inspection
- Reverberated and non-reverberated



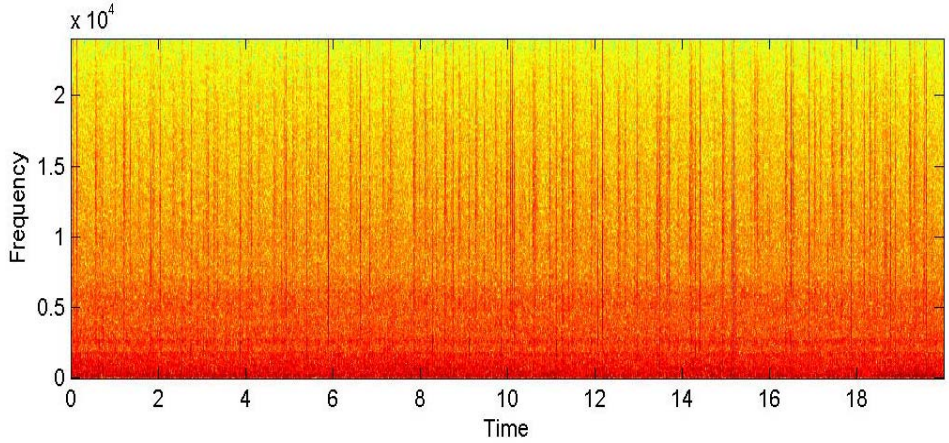
2. Processing

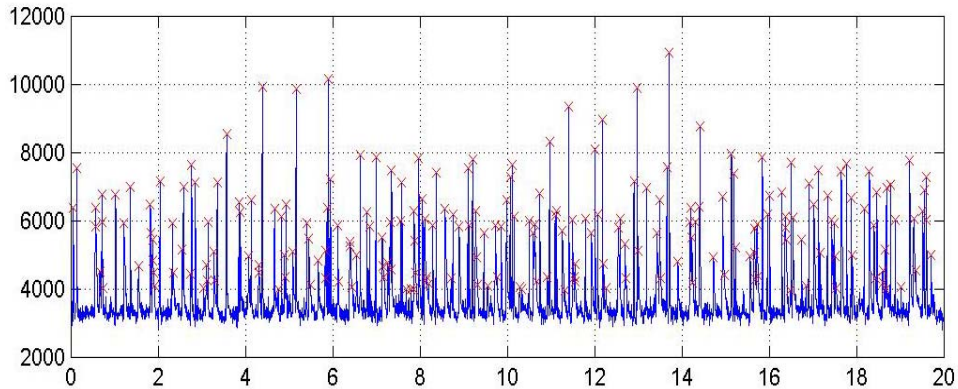
- Detection → Features → Clustering



Click Detection

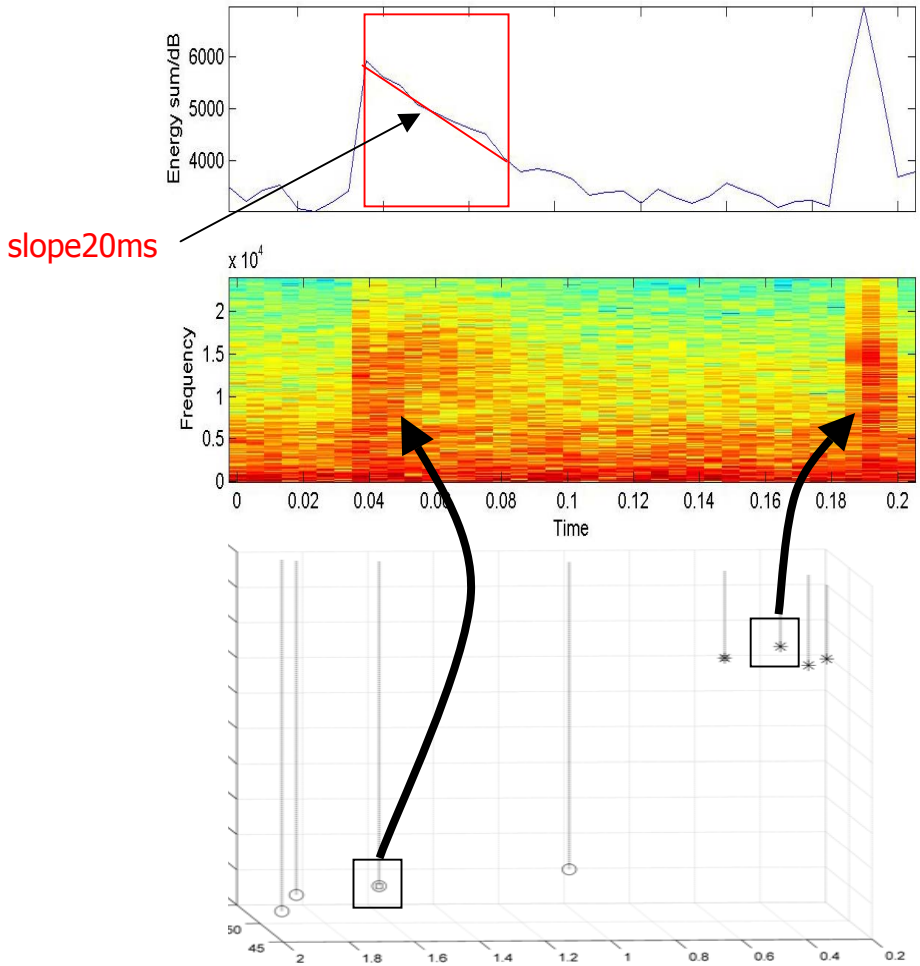
- Click detector based on variance thresholding
- Obtain onset of clicks to extract desired features
- Lack of known number of clicks per frame could lead to false positives/negatives





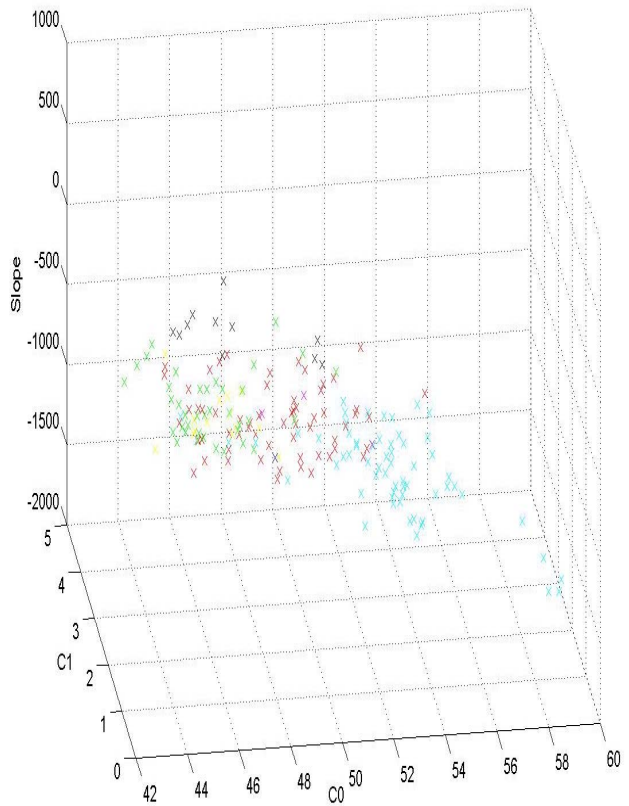
Features

- Cepstral coefficients C_0, C_1 for click time slices as obtained from click detection
  - Energy
- Energy slope at click onset in a 20msec window
  - Discriminate reverb



Spectral Clustering

- Build affinity matrix *A* using reverse euclidean distance and have Aii = 0
- Define D as the diagonal matrix whose elements are the sum of A's rows
- Form $L = D^{-1/2} A D^{-1/2}$
- Choose dominant eigenvectors of L and stack them in columns to form matrix X
  - Normalize X's rows to have unit length
- Perform Kmeans on X assuming each row as a feature vector



K for K means

- Cluster distortion

$$I_j = \sum_{t=1}^{N_j} \left[d\left(x_{jt}, w_j\right) \right]^2$$

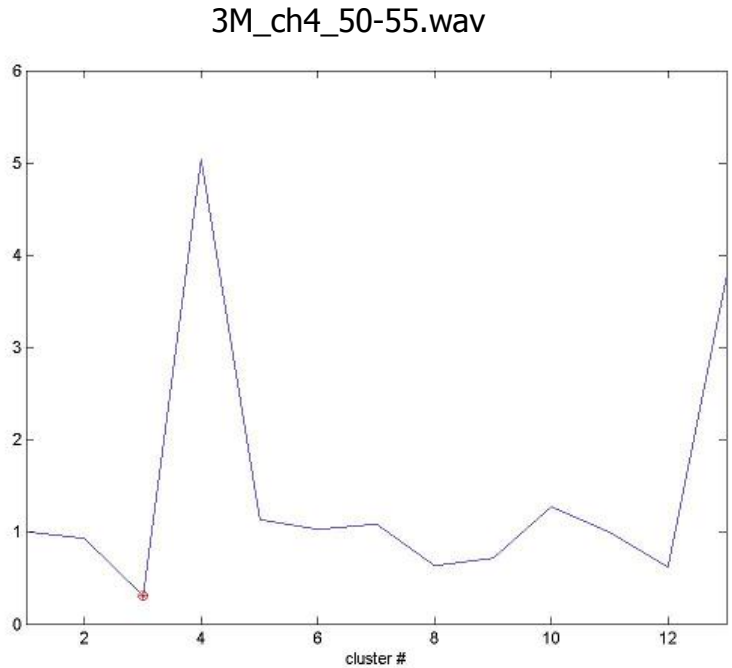
- Total distortion

$$S_K = \sum_{j=1}^{K} I_K$$

- Cluster function

$$f(K) = \begin{cases} 1 & if\ K = 1 \\ \dfrac{S_K}{\alpha_K S_{K-1}} & if\ S_{K-1} \neq 0, \forall K > 1 \\ 1 & if\ S_{K-1} = 0, \forall K > 1 \end{cases}$$

$$\alpha_K = \begin{cases} 1 - \dfrac{3}{4N_d} & if\ K = 2\ and\ N_d > 1 \\ \alpha_{K-1} + \dfrac{1 - \alpha_{K-1}}{6} & if\ K > 2\ and\ N_d > 1 \end{cases}$$



3. Experiments

- Use different hydrophones for cross-validation

Audio file (wav)	# of clusters	Size per chunk (min)	Hydrophone
3M_ch4_35-40	5	1	A
3M_ch5_35-40	7	1	B
3M_ch6_35-40	9	1	C
3M_ch4_40-45	5	1	A
3M_ch5_40-45	7	1	B
3M_ch6_40-45	7	1	C
3M_ch4_45-50	5	1	A
3M_ch5_45-50	4	1	B
3M_ch6_45-50	8	1	C
3M_ch4_50-55	5	1	A
3M_ch5_50-55	4	1	B
3M_ch6_50-55	4	1	C

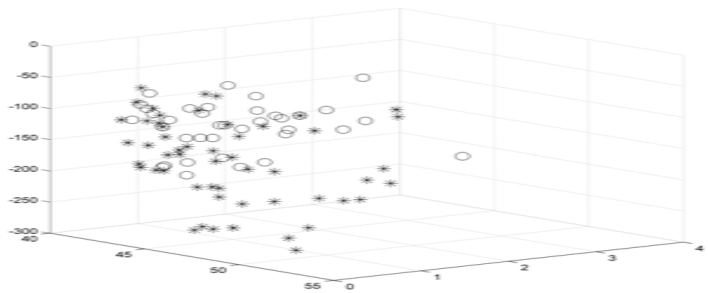
Audio file (wav)	# of clusters	Size per chunk (sec)	Hydrophone
3M_ch4_35-40	4	20	A
3M_ch5_35-40	5	20	B
3M_ch6_35-40	4	20	C
3M_ch4_40-45	5	20	A
3M_ch5_40-45	5	20	B
3M_ch6_40-45	6	20	C
3M_ch4_45-50	4	20	A
3M_ch5_45-50	4	20	B
3M_ch6_45-50	4	20	C
3M_ch4_50-55	3	20	A
3M_ch5_50-55	4	20	B
3M_ch6_50-55	4	20	C

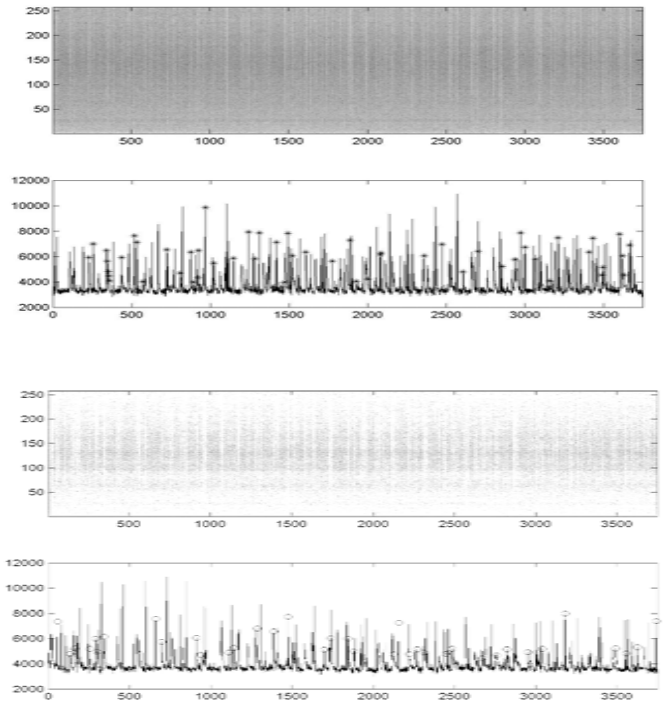
KL Divergence

- Assume single gaussian distributions for each cluster to have closed form

$$KL_N(\mu_p, \Sigma_p; \mu_q, \Sigma_q) =$$

$$\log \frac{|\Sigma_q|}{|\Sigma_p|} + Tr\left(\Sigma_q^{-1}\Sigma_p\right) + \left(\mu_p - \mu_q\right)^{-1}\Sigma_q^{-1}\left(\mu_p - \mu_q\right)$$





4. Conclusions

- Approximate number of marine mammals in area with the use of recordings from one mic
- Take advantage of click inter-timing for localization of sperm whales
- Different features could yield a better result
  - Energy ratios
- Cross-correlation with different hydrophones will identify individual whales
- In order to have a full verification of the above absolute ground truth is needed