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
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

**Detection**
- Features
  - High-pass FIR 20sec windows
  - STFT/Spectrogram

**Click-detection**
- Energy sum per time slice
- Variance based reg.max extraction = click onset

**Features**
- Cepstral coeff. $C_0$, $C_1$
- Energy slope 20ms window

**Clustering**
- Spectral Clustering
- Choosing number of clusters

# of marine mammals
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 20 msec window
  - Discriminate reverb
Spectral Clustering

- Build affinity matrix $A$ using reverse euclidean distance and have $A_{ii} = 0$
- 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

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

- Total distortion

\[ S_K = \sum_{j=1}^{K} I_K \]

- Cluster function

\[ f(K) = \begin{cases} 1 & \text{if } K = 1 \\ \frac{S_K}{\alpha_K S_{K-1}} & \text{if } S_{K-1} \neq 0, \forall K > 1 \\ 1 & \text{if } S_{K-1} = 0, \forall K > 1 \end{cases} \]

\[ \alpha_K = \begin{cases} 1 - \frac{3}{4N_d} & \text{if } K = 2 \text{ and } N_d > 1 \\ \alpha_{K-1} + \frac{1 - \alpha_{K-1}}{6} & \text{if } K > 2 \text{ and } N_d > 1 \end{cases} \]
3. Experiments

- Use different hydrophones for cross-validation

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

<table>
<thead>
<tr>
<th>Audio file (wav)</th>
<th># of clusters</th>
<th>Size per chunk (sec)</th>
<th>Hydrophone</th>
</tr>
</thead>
<tbody>
<tr>
<td>3M_ch4_35-40</td>
<td>4</td>
<td>20</td>
<td>A</td>
</tr>
<tr>
<td>3M_ch5_35-40</td>
<td>5</td>
<td>20</td>
<td>B</td>
</tr>
<tr>
<td>3M_ch6_35-40</td>
<td>4</td>
<td>20</td>
<td>C</td>
</tr>
<tr>
<td>3M_ch4_40-45</td>
<td>5</td>
<td>20</td>
<td>A</td>
</tr>
<tr>
<td>3M_ch5_40-45</td>
<td>5</td>
<td>20</td>
<td>B</td>
</tr>
<tr>
<td>3M_ch6_40-45</td>
<td>6</td>
<td>20</td>
<td>C</td>
</tr>
<tr>
<td>3M_ch4_45-50</td>
<td>4</td>
<td>20</td>
<td>A</td>
</tr>
<tr>
<td>3M_ch5_45-50</td>
<td>4</td>
<td>20</td>
<td>B</td>
</tr>
<tr>
<td>3M_ch6_45-50</td>
<td>4</td>
<td>20</td>
<td>C</td>
</tr>
<tr>
<td>3M_ch4_50-55</td>
<td>3</td>
<td>20</td>
<td>A</td>
</tr>
<tr>
<td>3M_ch5_50-55</td>
<td>4</td>
<td>20</td>
<td>B</td>
</tr>
<tr>
<td>3M_ch6_50-55</td>
<td>4</td>
<td>20</td>
<td>C</td>
</tr>
</tbody>
</table>
KL Divergence

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

\[
KL_N(\mu_p, \Sigma_p ; \mu_q, \Sigma_q) = \log \left| \frac{\Sigma_q}{\Sigma_p} \right| + Tr(\Sigma_q^{-1} \Sigma_p) + (\mu_p - \mu_q)^{-1} \Sigma_q^{-1} (\mu_p - \mu_q)
\]
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