Call Detection and Extraction using Sinewave Modeling and Bayesian Inference

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. Whistle Detection
2. Processing
3. Results
4. Conclusions

1. Whistle Detection

- Whistles are considered to be used mostly for interaction
- Signature whistle hypothesis implies that they are of great importance for recognition tasks
- Great amounts of recordings in need of labeling
  - Manual approach is extremely time consuming
  - Automatic, real time detection and extraction would be very helpful for further analysis
- Try to avoid species or other constraints

Detection and extraction of marine mammal whistle vocalizations

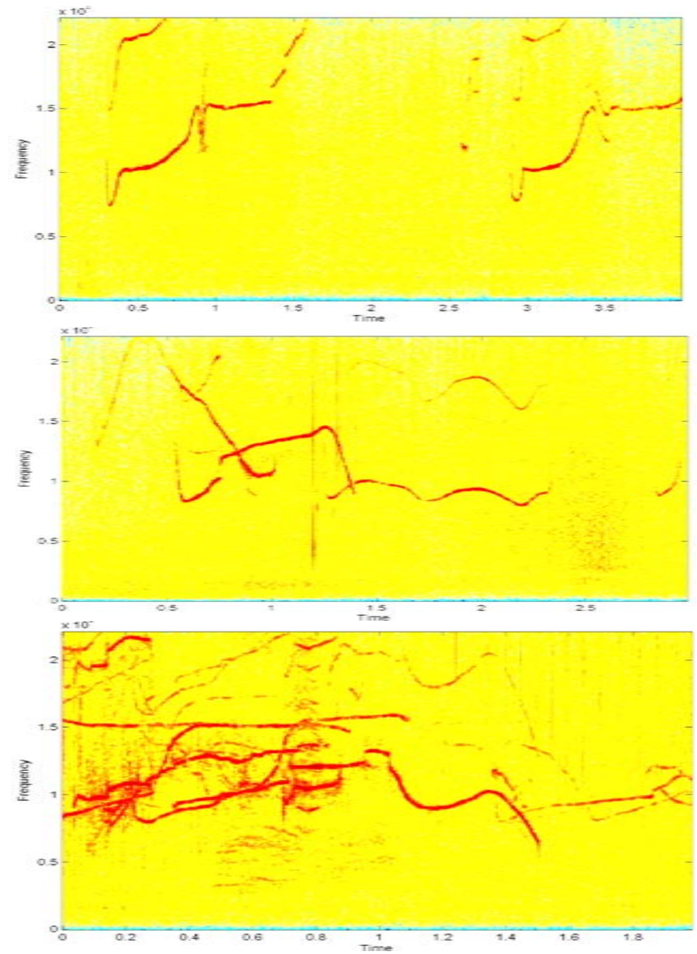
- Task:

  Automatically and in real time extract whistle vocalizations present in marine mammal recordings

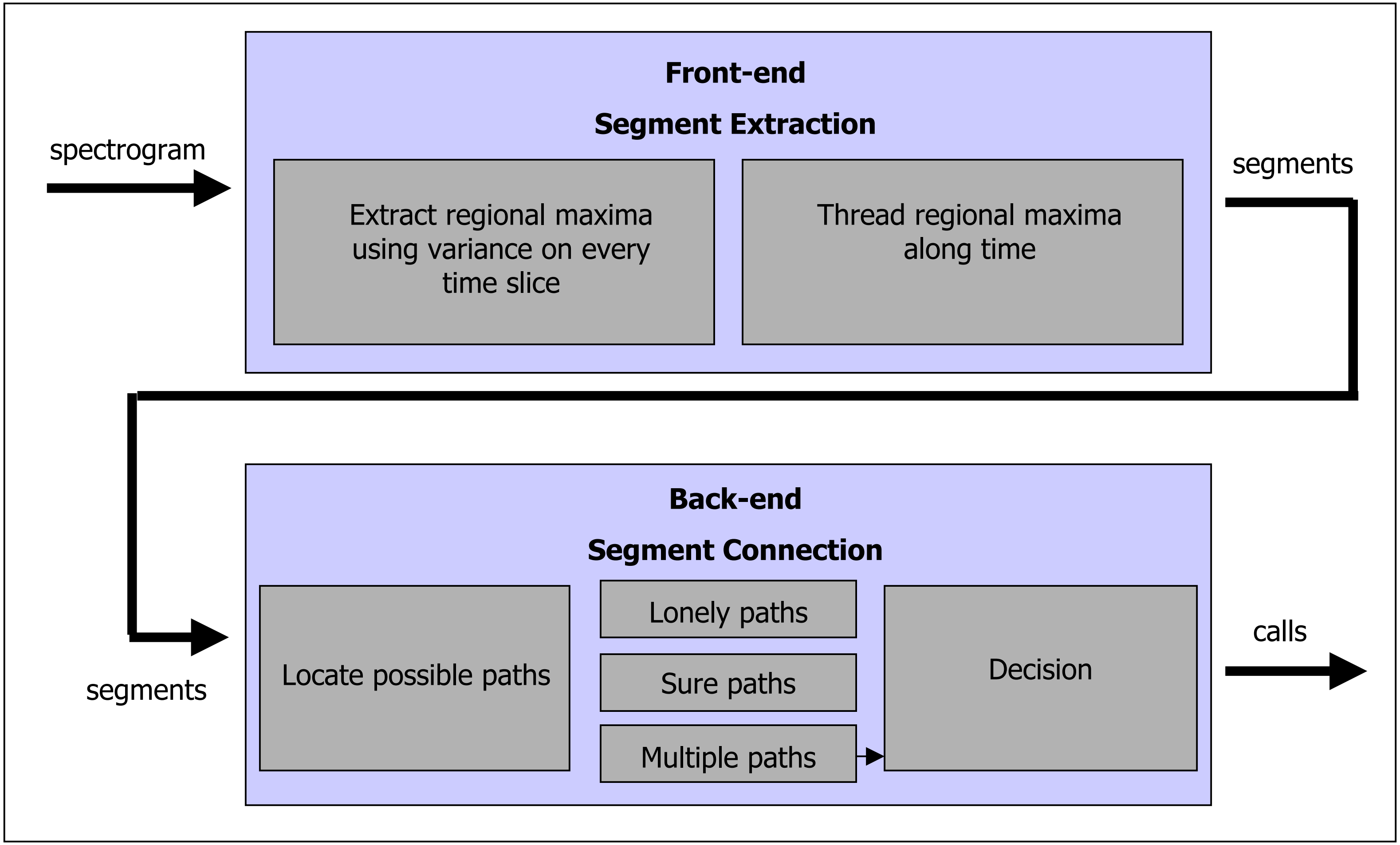
  - Species dependencies inherent in cross-correlation methods
  - Noise and multiple marine mammals cause overlaps of whistles
- Goal:
  - Create a versatile system that can be both species dependent or independent
  - Ability to handle noisy signals
  - Decipher overlaps of whistles

Data-Whistle Examples

- Data of increasing difficulty
  - Easy: simple whistles with low noise
  - Moderate: overlaps and non-uniform whistles
  - Difficult: multiple whistles with mostly overlaps and no distinguishable shape
- Species and technical details are known
- Data obtained from Macaulay Library, Cornell University

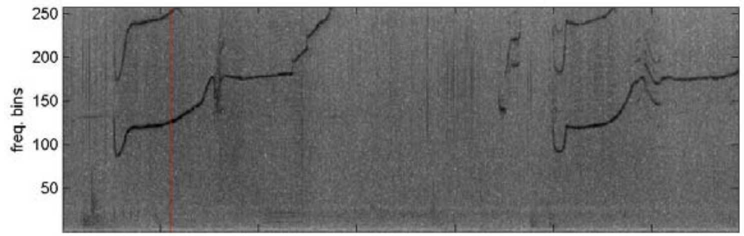


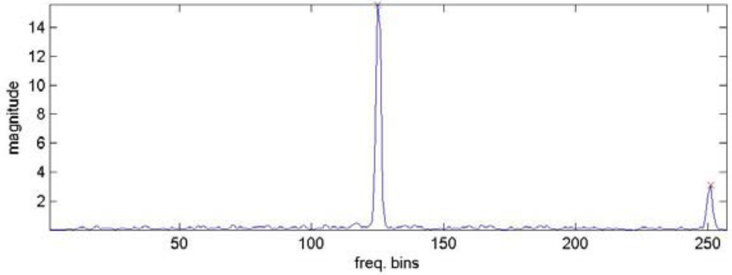
2. Processing

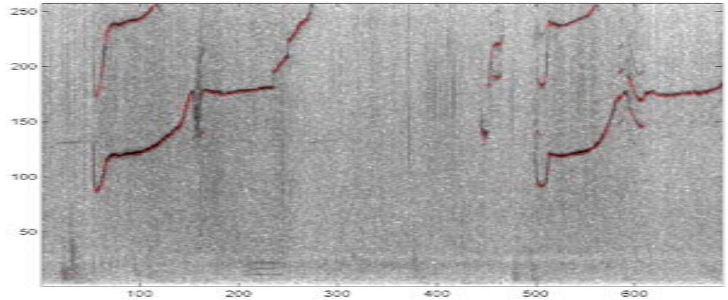


Front-end: Segment extraction

- Extract regional maxima
  - o Variance based thresholding
  - o Minimize false peaks due to noise
- Thread regional maxima in time
  - o Use magnitude threshold
  - o Break the segments according to pre-specified number of dead steps





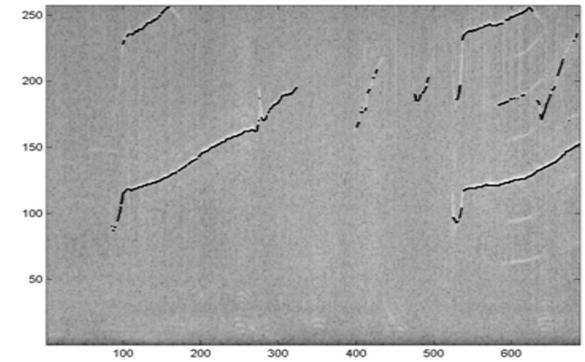


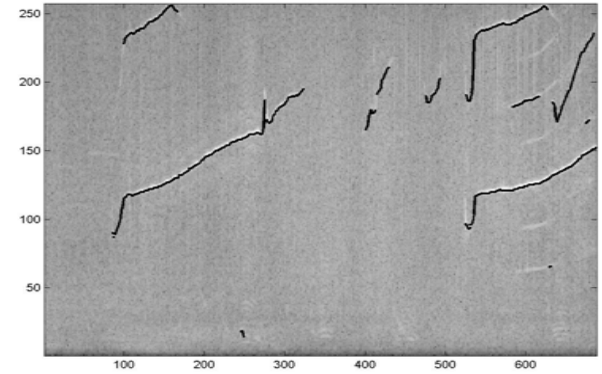
Back-end: Segment connection

- Locate all possible paths for every segment's tip using an adaptive neighborhood
- Soft decision based on the slopes of the tips using ML through training data
  - Directionality of the calls given their short length
- Sort paths:
  - Lonely paths: no connection
  - Sure paths: one possible connection
  - Multiple paths: multiple connections
- Decide the best path using a greedy Viterbi type algorithm according to the likelihoods of global call characteristics such as smoothness in frequency and energy
- Distributions obtained through training data e.g 40 calls

Back-end: Segment connection

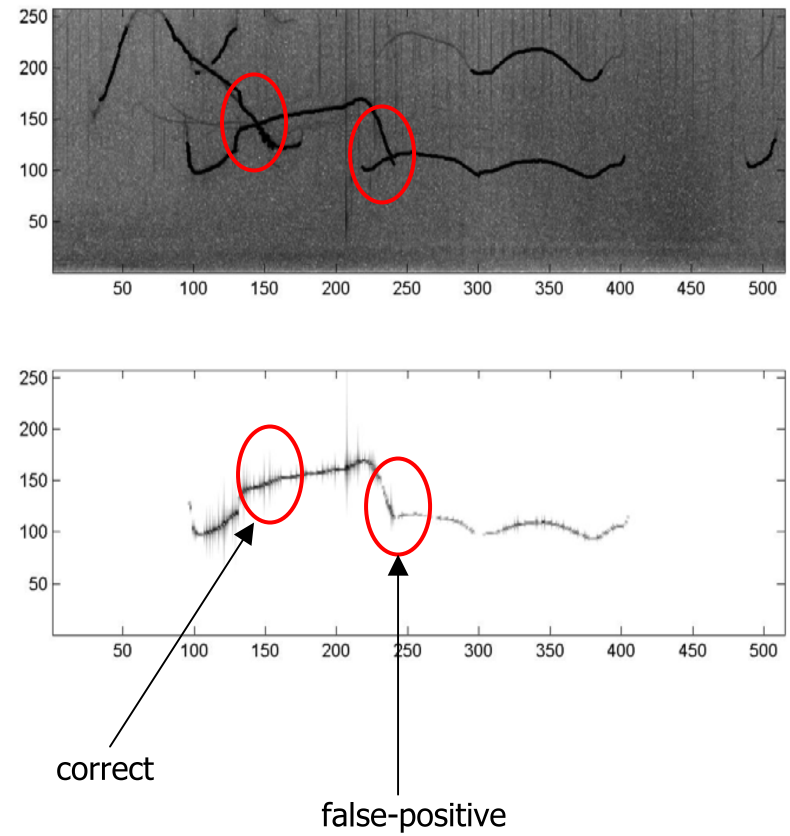
- Criteria works well for simple frames
- Greedy algorithm chooses maximum likelihood at each step
- Ties are resolved by choosing the so far call with the overall highest likelihood





Overlaps

- Novelty:
  - Resolve overlaps between calls
- Global characteristics appear to work adequately for simple overlaps
  - Errors due to resolution fuzziness
- Ability to extract calls that belong to multiple marine mammals



3. Results

- Algorithm applied on 5min of audio approximately 400 whistles
- Overall success rate obtained in the frame level
  - Number of points extracted vs. actual number of points
- False positive and negative rates obtained in the segment level
  - Number of false/correct connections vs. number of all connections
- Errors are based on either bad segment detection or the inability of the characteristics to capture sharp changes in frequency and energy

Rate	Percentage
Success	82%
False-positive	5%
False-negative	3%

4. Conclusions

- Tunable model based on a probabilistic framework for the extraction of whistles in marine mammal vocalizations
- Global characteristics of calls able to deal with moderate complicated frames
- Ability to improve by adding more characeristics
- Improve with multi-resolution approach
- Drawback:
  - Dependency on good segment extraction
  - Two-stage process