Audio processing methods on marine mammal vocalizations

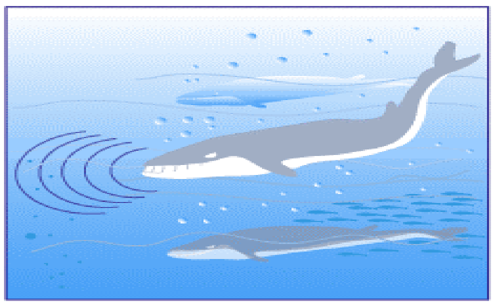
Xanadu Halkias

Laboratory for the Recognition and Organization of Speech and Audio

http://labrosa.ee.columbia.edu

Sound to Signal

- sound is pressure variation of the medium (e.g. speech air pressure, marine mammals water pressure)

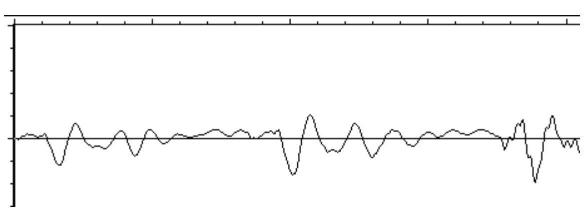


Pressure waves in water

↓

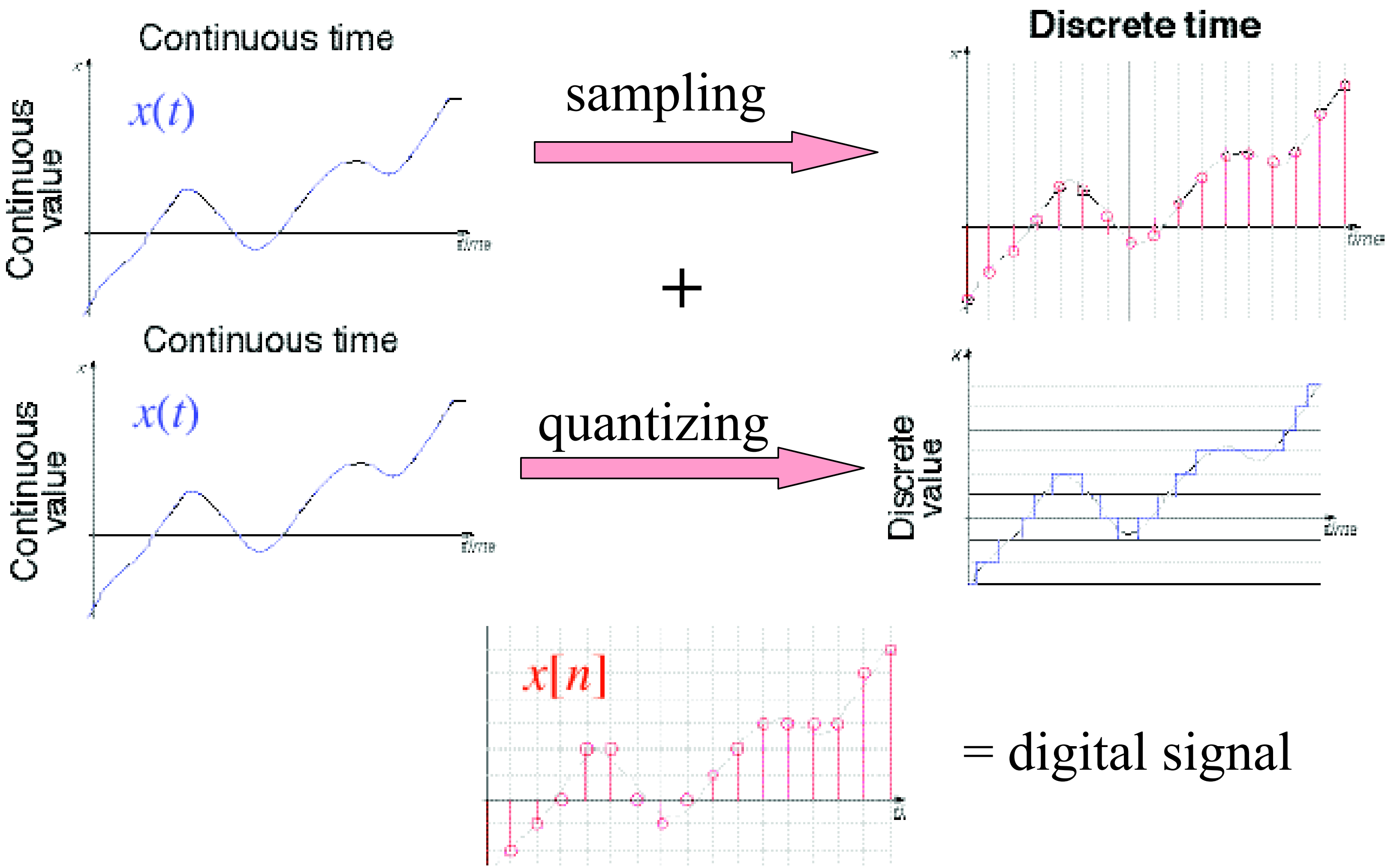
Converting waves to voltage through a microphone

↓



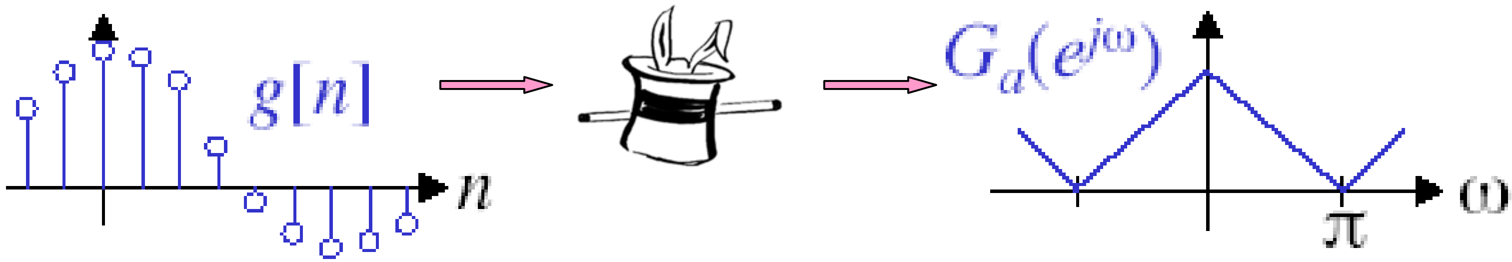
Time varying voltage

Analog to digital



Time to frequency and back…

•Fourier transform=decompose a signal as a sum of sinusoids and cosines

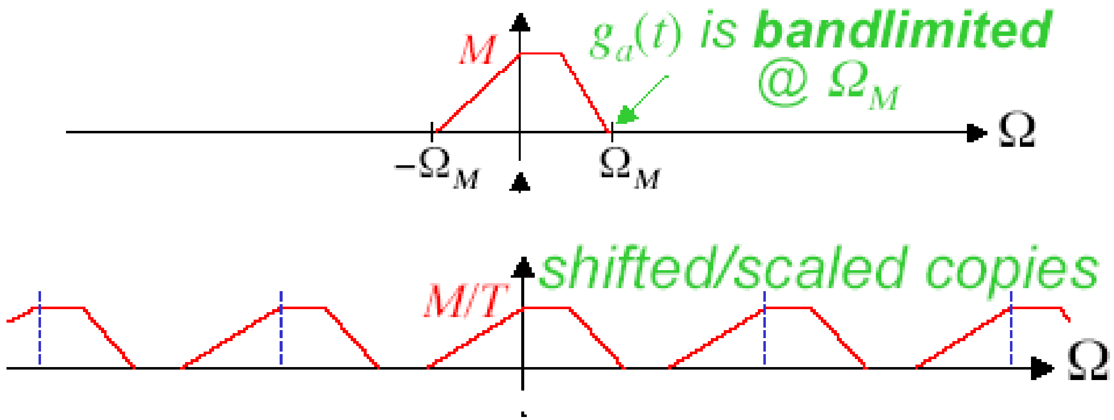


Digital signal Fourier spectrum

Spectrum = the frequency content of the signal (energy/frequency band)

Back to sampling…

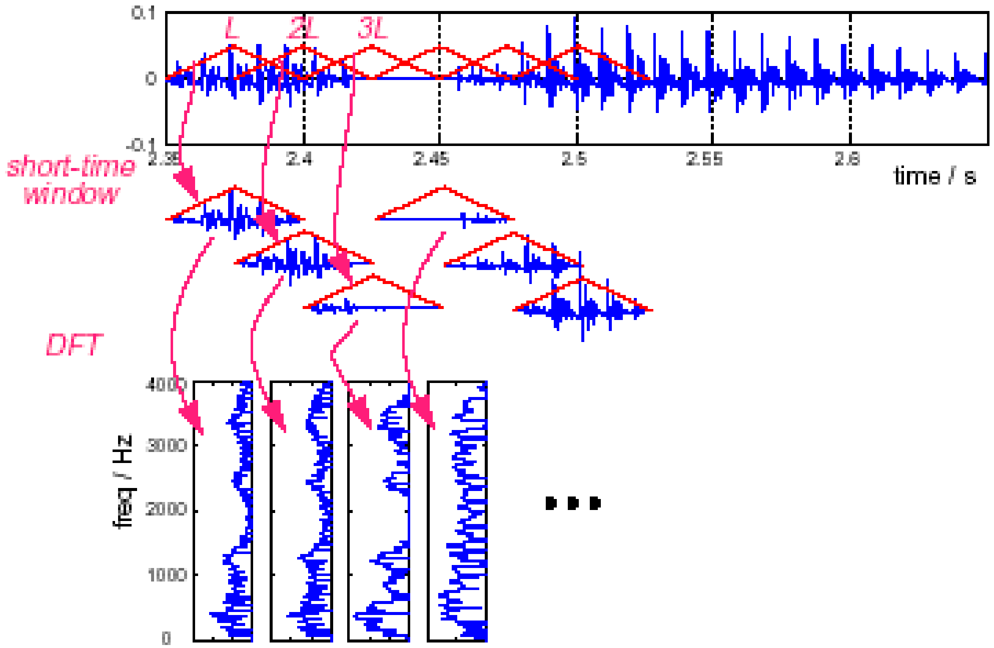
- Signal has to be bandlimited eg. energy up to some frequency Ω_M
- Sampling needs to obey the Nyquist limit: $\Omega_T \geq 2\Omega_M$



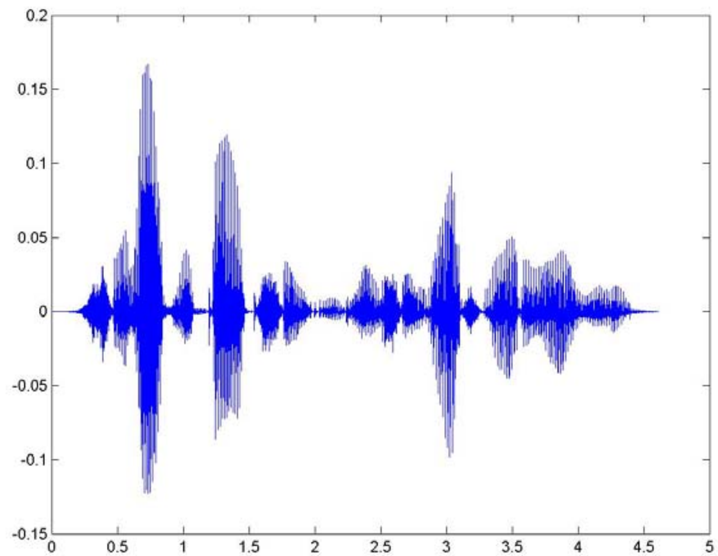
- Audio is sampled at $\Omega_T = 2\pi 44100$Hz so spectrum has up to 22050Hz

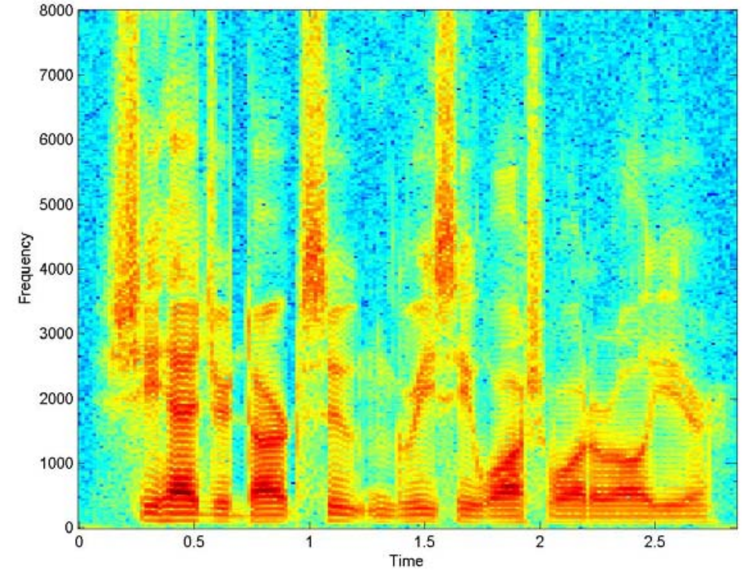
Looking at sounds-The Spectrogram

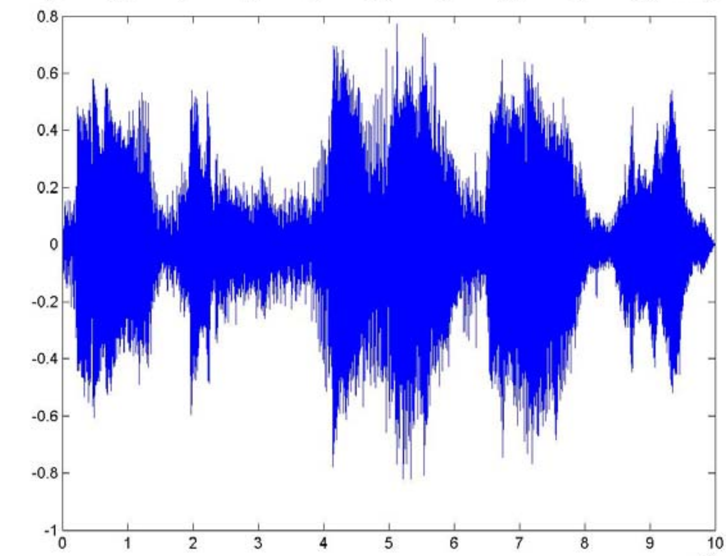
- Looking at energy in time and frequency

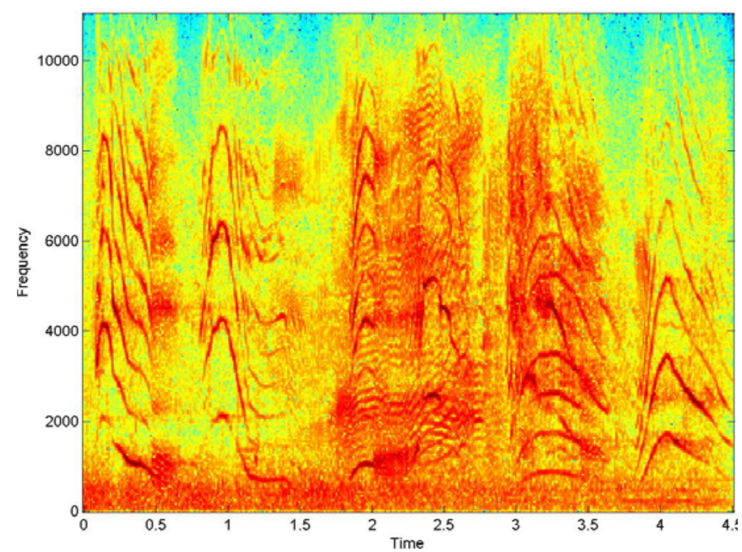


More on spectrograms

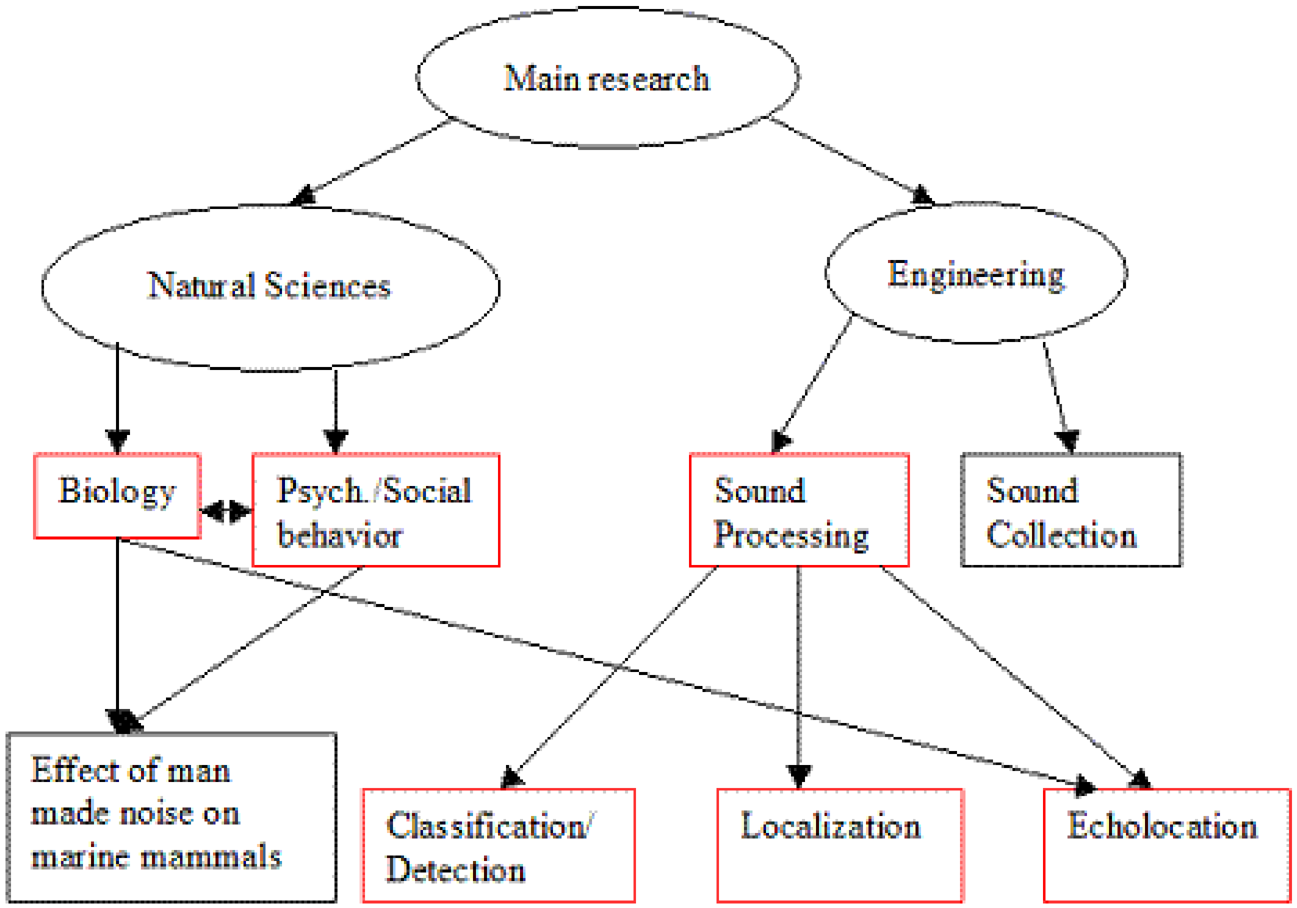








Overview of marine mammal research



Call detection

What is it good for…

- Detect different calls within the recording automatically
- Differentiate between species or identify the number of marine mammals in the region through overlapping of calls
- Tracking marine mammals through their calls
- Use calls to analyze and construct a possible language structure

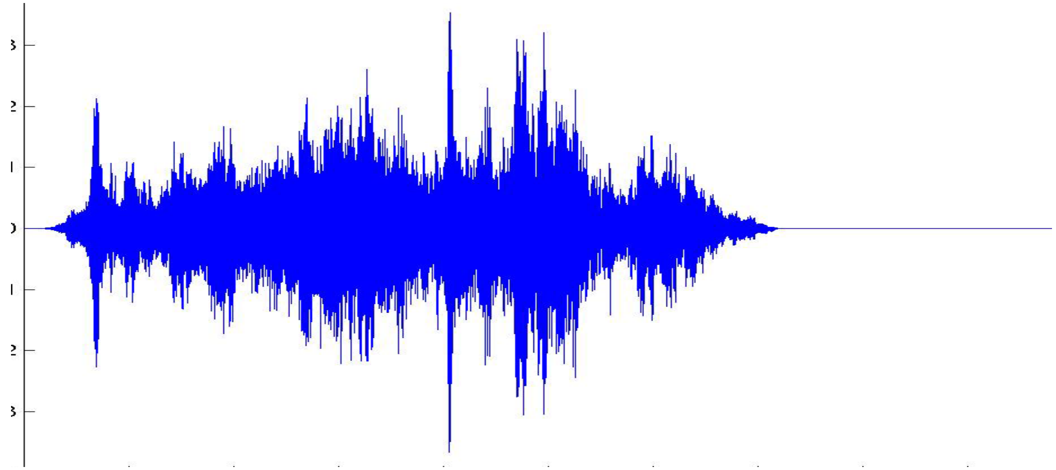
Problems

- Data, data, data…

Call detection approaches

- Noise is the biggest problem
- D. K. Mellinger et all use the cross-correlation approach

Cross-correlation is a way of measuring how similar two signals are



Call detection-kernel cross- correlation

- This method requires manual interference and is performed on the signal waveform

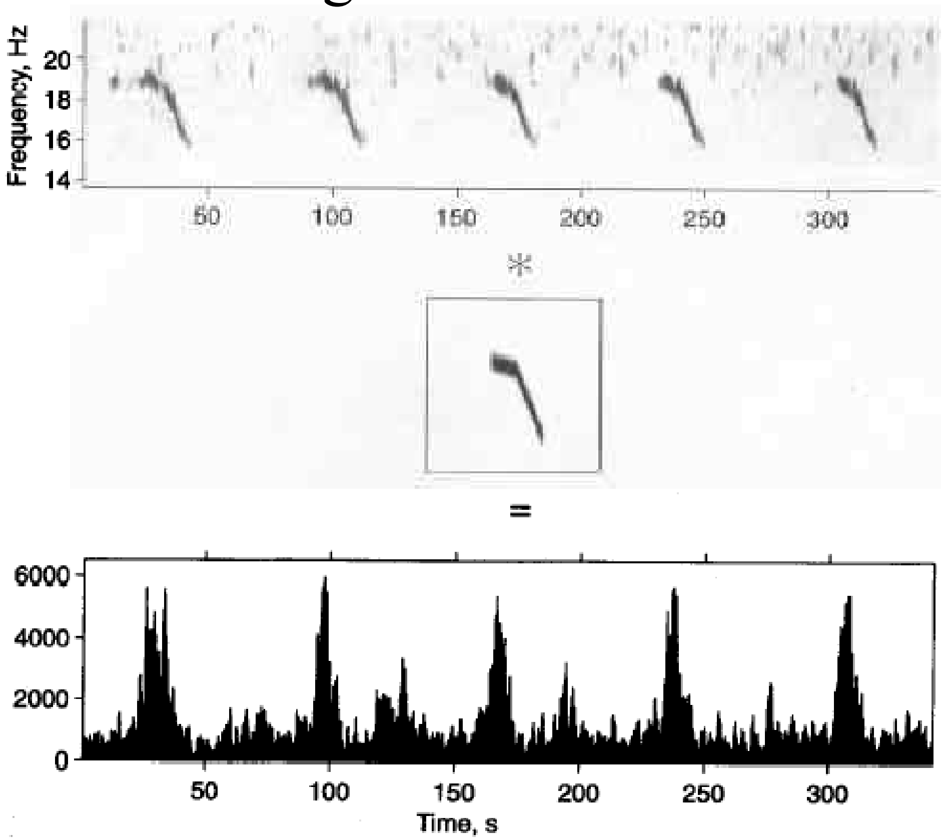


Image obtained by D. K. Mellinger and C. W. Clark. *"Methods for automatic detection of mysticete sounds"*, Mar. Fresh. Behav. Physiol. Vol. 29, pp. 163-181, 1997

Call detection-spectrogram correlation

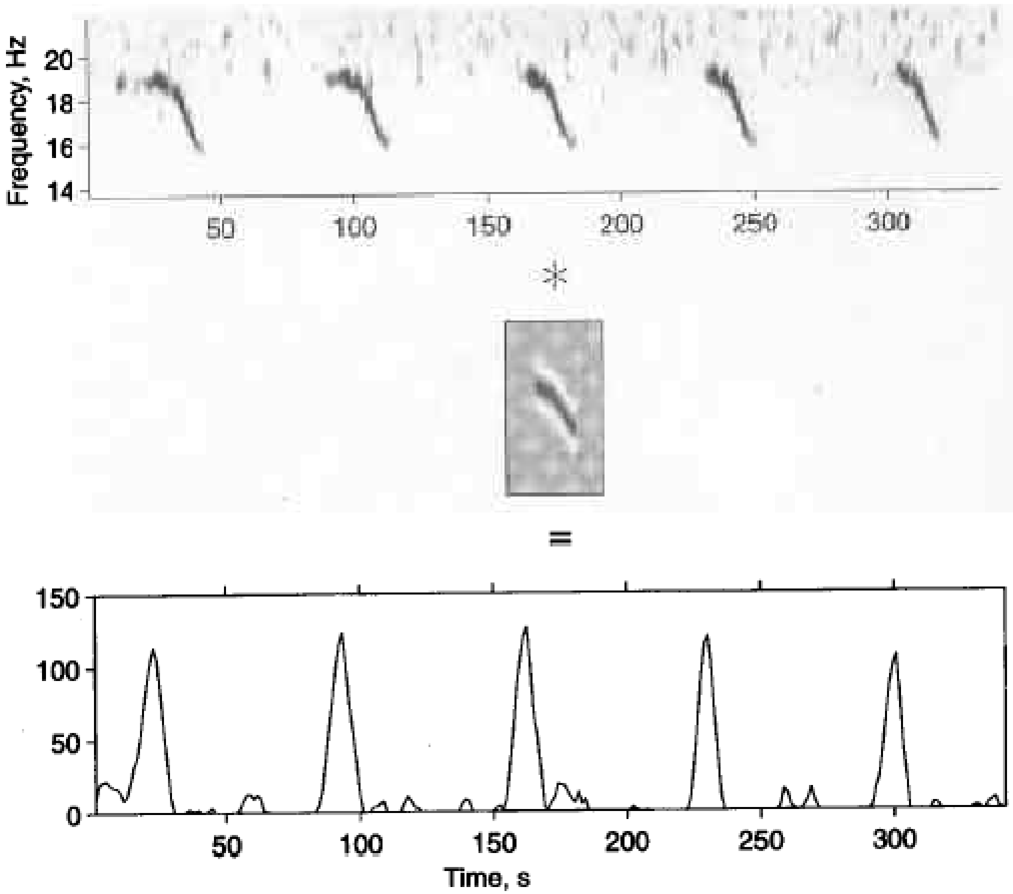
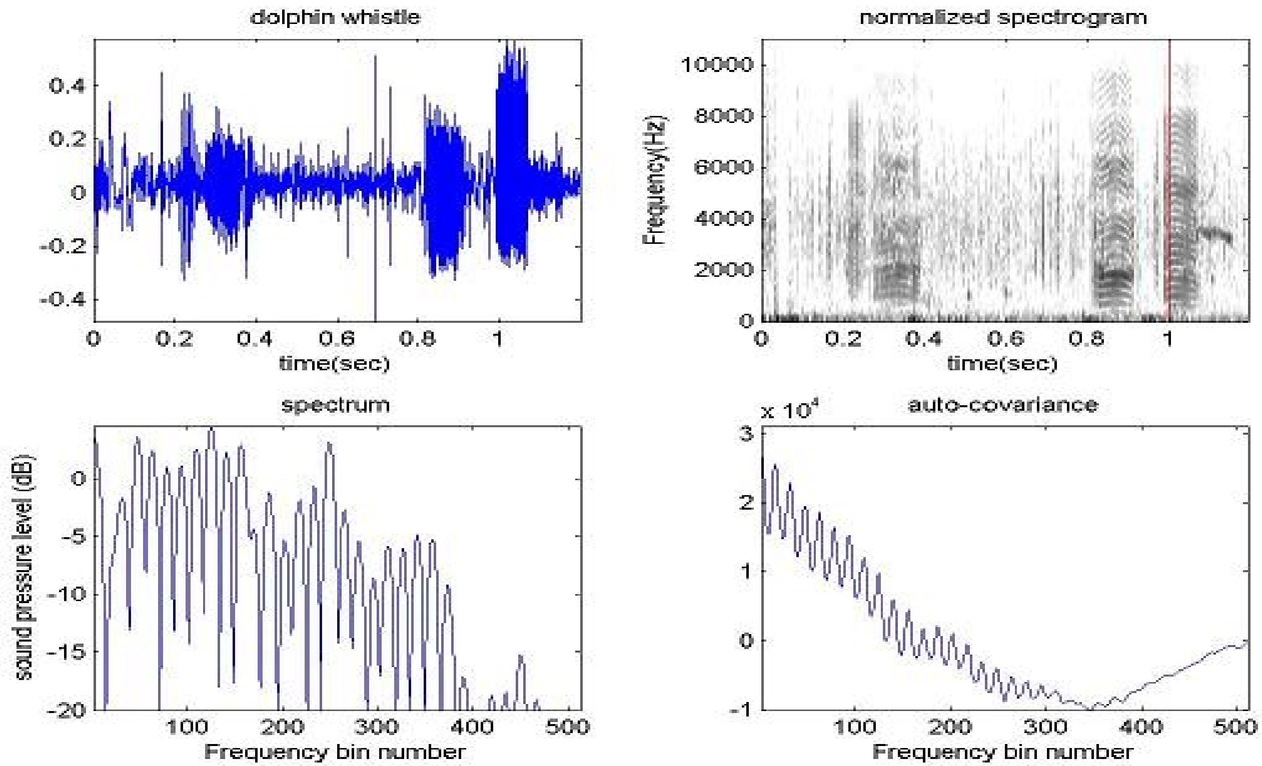


Image obtained by D. K. Mellinger and C. W. Clark. *"Methods for automatic detection of mysticete sounds"*, Mar. Fresh. Behav. Physiol. Vol. 29, pp. 163-181, 1997

"Voiced" calls

Energy appears in multiples of some frequency (=pitch)



Comments

•Both methods require manual measurements for the construction of the template

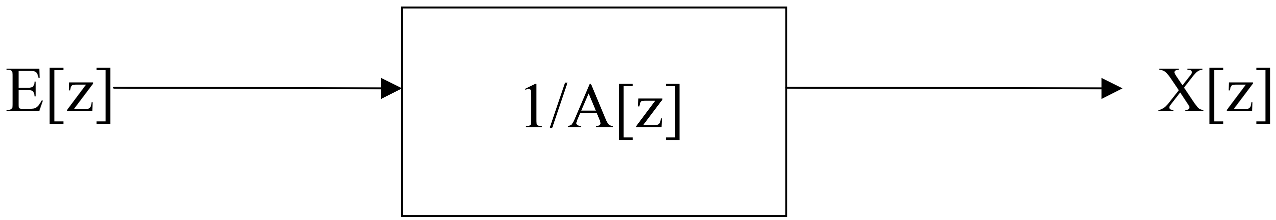
•The quality of the results depends highly on the noise present in the data

•Quality recordings at high sampling rates decide the course of action

•Correlation methods can’t capture all types of calls without constructing different kernels

Linear Predictive Coding

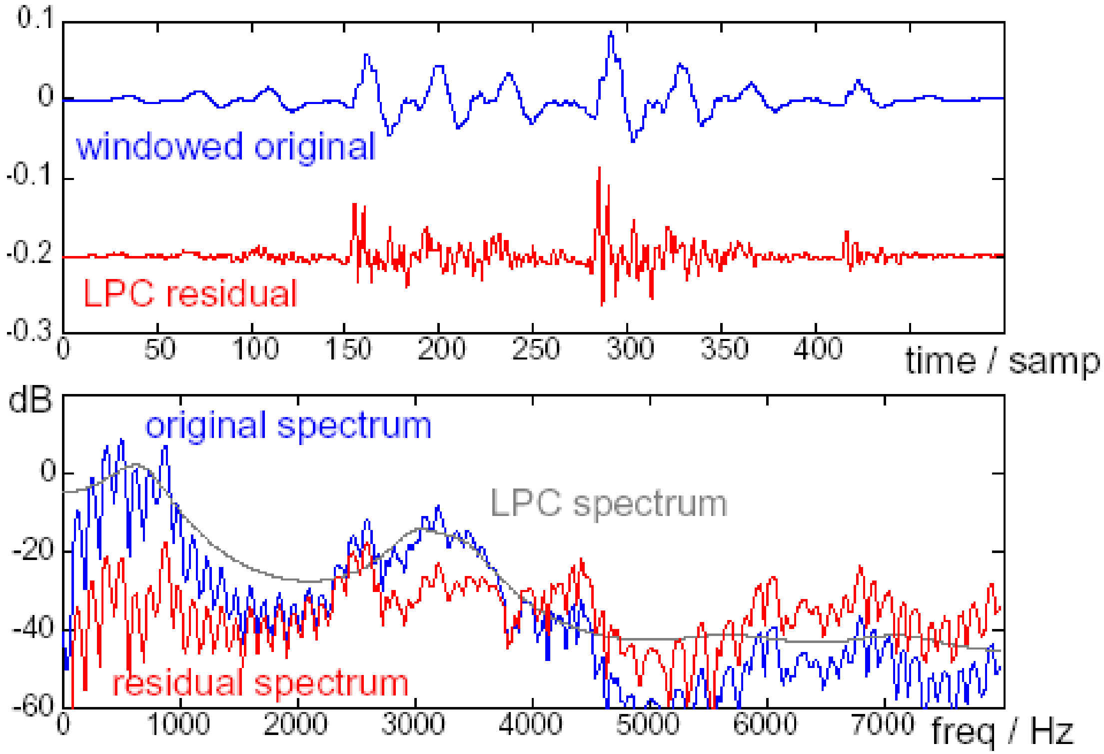
•Idea: the signal, x[n], is formed by adding white noise, e[n], to previous samples weighted by the linear predictive coefficients, *a*



•The number of coefficients defines the detail that we capture of the original signal

Linear Predictive Coding

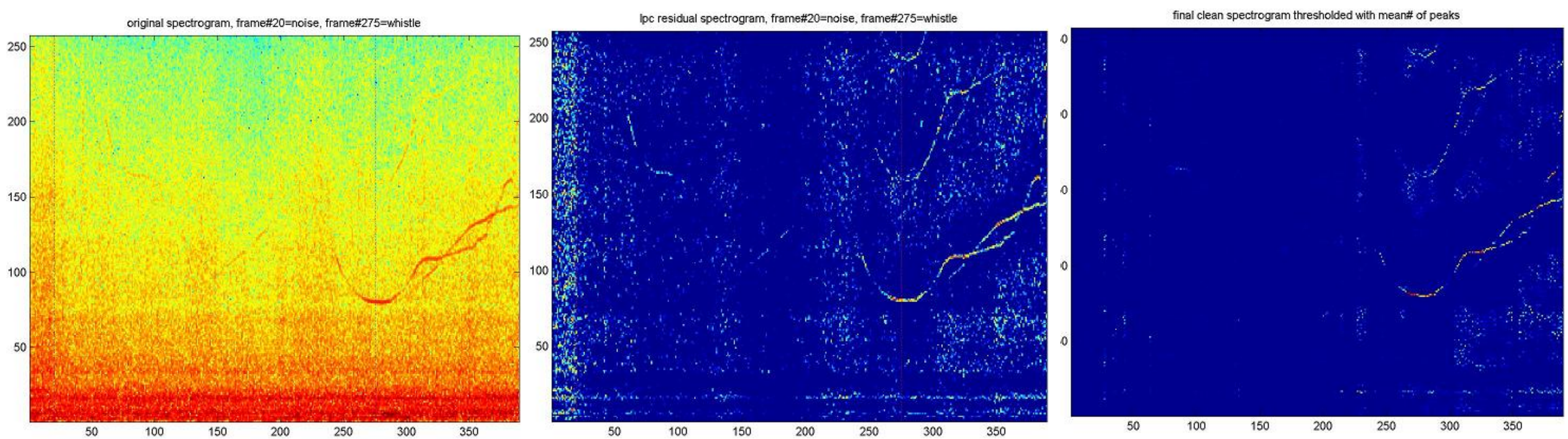
- Used in speech for transmission purposes
- Intuition: LPCs model the spectral peaks of your signal



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

LPCs in marine mammal recordings

- Model the peaks in the recordings that likely belong to calls that way we alleviate the problem of noise



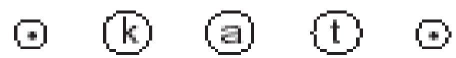
- Unveils harmonic structure not visible in original spectrogram

Hidden Markov Models

•Machine learning involves training a general model based on your data in order to extract and predict desired features

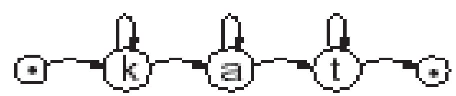
•HMMs, M_i are defined by:

- states q^i



- transition probabilities a_{ij}

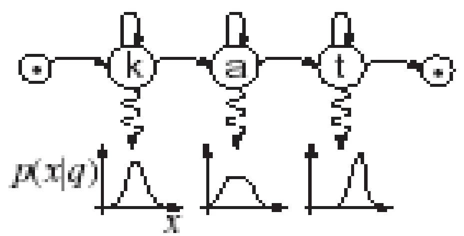
$$p(q_n^j | q_{n-1}^i) \equiv a_{ij}$$



	k	a	t	•
•	1.0	0.0	0.0	0.0
k	0.9	0.1	0.0	0.0
a	0.0	0.9	0.1	0.0
t	0.0	0.0	0.9	0.1

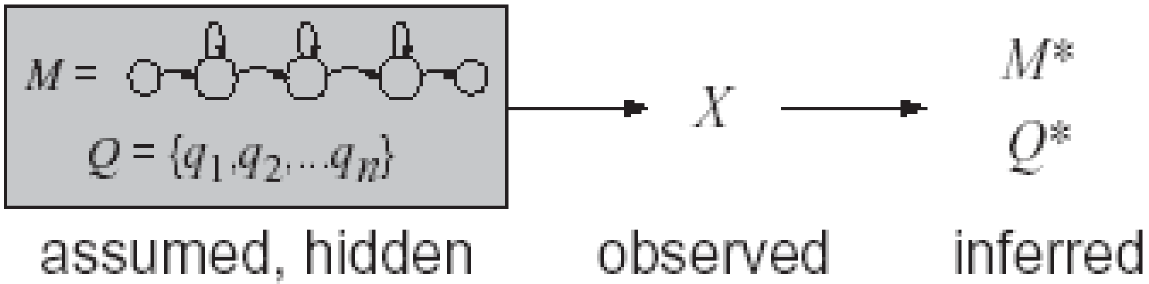
- emission distributions $b_i(x)$

$$p(x | q^i) \equiv b_i(x)$$



+ (initial state probabilities $p(q_1^i) \equiv \pi_i$)

HMMs some more…



•Training: getting the parameters of the model, *a, b, π*

•Evaluating: we are given a sequence of states we want to know if the model produced them

•Decoding: we have some observations and we want to find out the hidden states

HMMs in marine mammal vocalizations

- HMMs could provide a call detection tool
- The data has to be workable
- Use frequencies of the spectrogram as hidden states
- Observe the spectrogram and use it for learning
- Tracking the call in the spectrogram

References

- D. P. Ellis

www.ee.columbia.edu/~dpwe/e6820

www.ee.columbia.edu/~dpwe/e4810

- D. K. Mellinger and C. W. Clark. *"Methods for automatic detection of mysticete sounds"*, Mar. Fresh. Behav. Physiol. Vol. 29, pp. 163-181, 1997
- R. O. Duda, P. E. Hart, D. G. Stork. *Pattern Classification*, John Wiley & sons, inc. 2001