Whale Song Unit Classification

Preliminary Exploration
Using Linear Prediction Vector Quantization
and Hidden Markov Modeling

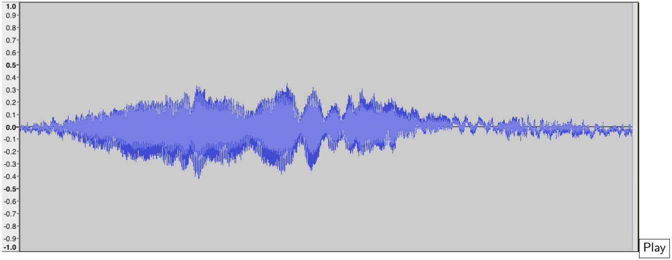
Carlos Rueda - MBARI

The Problem

Following Mitchell (1997), we state the problem as follows:

- Task:
  Classify whale song unit instances
  according to a given vocabulary of whale song unit types
- Performance measure:
  Percent of instances correctly classified
- Training experience:
  A database of labelled whale song unit instances

Whale Song Units



A "modulated cry" instance, from `HBSe_20151207T070326.wav, 124.5sec-126.5sec`

- Acoustic signal:

$$\mathbf{x} = \langle x_1, x_2, \ldots, x_N \rangle \qquad (N = 66,283)$$

Linear Predictive Coding and Vector Quantization

- Acoustic signal:

$$\mathbf{x} = \langle x_1, x_2, \ldots, x_N \rangle$$

- Transformed into a sequence of predictor vectors:

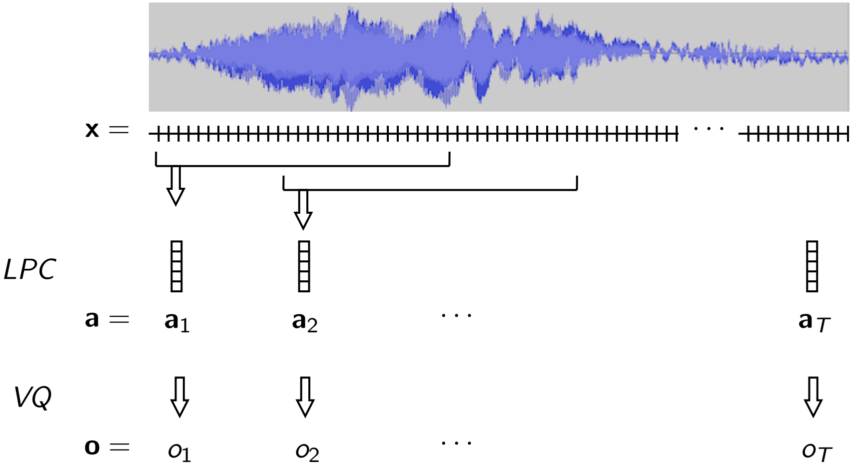
$$\mathbf{a} = \langle \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_T \rangle$$

- Transformed into a sequence of symbols:

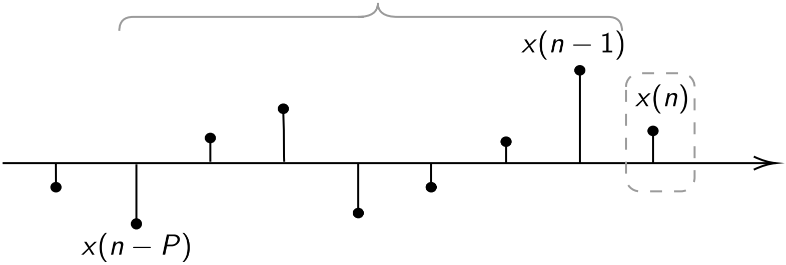
$$\mathbf{o} = \langle o_1, o_2, \ldots, o_T \rangle$$

where $o_t \in \{1, 2, \ldots, M\}$

Linear Predictive Coding and Vector Quantization



Linear Prediction



- Estimate $x(n)$ as a linear combination of P previous samples:

$$\hat{x}(n) = -\sum_{i=1}^{P} a_i x(n-i)$$

Linear Prediction

- Error:

$$
\begin{aligned}
e(n) &= x(n) - \hat{x}(n) \\
&= \sum_{i=0}^{P} a_i x(n-i) \quad (a_0 \equiv 1)
\end{aligned}
$$

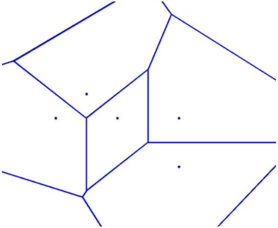
- Find prediction coefficients by minimizing the sum of the squared error over the signal interval:

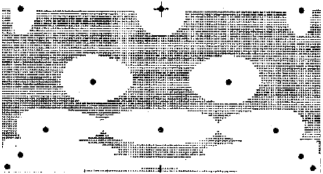
$$
\begin{aligned}
E &= \sum_n e^2(n) \\
&= \sum_n \left(\sum_{i=0}^{P} a_i x(n-i) \right)^2
\end{aligned}
$$

- Resulting coefficients form prediction vector over the interval

$$
\mathbf{a} = \langle a_1, a_2, \ldots, a_P \rangle
$$

Vector Quantization





- Partition the P-dimensional space into M regions that "best" represent the predictor vectors arising from song units instances
- "Best" in terms of minimizing some overall distortion measure

$$d(\mathbf{a}, \mathbf{b}) \equiv \text{Distance or dissimilarity between } \mathbf{a} \text{ and } \mathbf{b}$$

Vector Quantization

- Let $\mathbf{C} \equiv \langle C_1, C_2, \ldots, C_M \rangle$ be the set of centroids that best represent the space
- Quantization of a given $\mathbf{a}$:

$$o_t \leftarrow \operatorname*{argmin}_{k=1}^{M} d(\mathbf{a}, C_k)$$

Hidden Markov Model

- π, initial state probability distribution:

$$\pi = (\pi_1, \pi_2, \ldots, \pi_N)$$

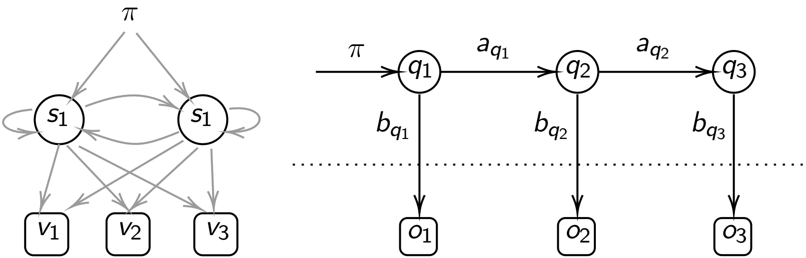
- A, state transition distributions:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1N} \\ a_{21} & a_{22} & \ldots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \ldots & a_{NN} \end{bmatrix}$$

- B, observation symbol distributions:

$$B = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1M} \\ b_{21} & b_{22} & \ldots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \ldots & b_{NM} \end{bmatrix}$$

Hidden Markov Modeling



- $\mathbf{q} = \langle q_1, q_2, \ldots, q_T \rangle$: hidden state sequence
- $\mathbf{o} = \langle o_1, o_2, \ldots, o_T \rangle$: our acoustic signal

HMM Operations

With $\lambda = (\pi, A, B)$ denoting an HMM

- For given $\mathbf{o}$ and λ, compute $P[\mathbf{o}|\lambda]$
- For a given $\mathbf{o}$, learn a model $\lambda^* \equiv (\pi^*, A^*, B^*)$ such that $P[\mathbf{o}|\lambda^*]$ is maximized
- For given $\mathbf{o}$ and λ, estimate a most likely state sequence $\mathbf{q}^*$

Implementation

ECOZ Software[1]

- `lpc` - Performs LPC on wav files
- `vq.learn` - Trains codebooks for vector quantization
- `vq.quantize` - Generates observation sequences
- `hmm.learn` - Trains HMM model
- `hmm.classify` - HMM based classification of observation sequences
- `vq.classify` - VQ based classification of predictor files

[1] https://github.com/ecoz2/ecoz2

Preliminary Exercises

- LPC order: 36
- Analysis window size: 45ms ($1,440$ samples)
- Window offset: 15ms (480 samples)
- Pre-emphasis
- Hamming weighting

"whale1": On a Selection of Units from a Song File

Song file: `HBSe_20151207T070326.wav`

- Classes with at least 6 unit instances
- Approximately 80% for training, 20% for a testing
- 69 training song unit instances
- $N = 16$ states, $M = 512$ symbols
- 8 classes:

`descending_moan`	`groan`	`gurgle`	`modulated_cry`
`descending_shriek`	`groan_+_purr`	`gurgle?`	`purr`

"whale1": Classification on 69 Training Sequences

Confusion matrix:

		0	1	2	3	4	5	6	7	tests	errors
descending_moan	0	8	0	0	0	0	0	0	0	8	0
descending_shriek	1	0	4	0	0	1	0	0	0	5	1
groan	2	0	0	8	0	1	0	0	0	9	1
groan_+_purr	3	0	0	1	5	0	0	0	0	6	1
gurgle	4	0	0	0	0	20	1	0	0	21	1
gurgle?	5	0	0	0	0	0	5	0	0	5	0
modulated_cry	6	0	0	0	0	0	1	9	0	10	1
purr	7	0	0	0	0	0	0	0	5	5	0

	class	accuracy	tests	candidate order							
descending_moan	0	100.00%	8	8	0	0	0	0	0	0	0
descending_shriek	1	80.00%	5	4	1	0	0	0	0	0	0
groan	2	88.89%	9	8	1	0	0	0	0	0	0
groan_+_purr	3	83.33%	6	5	1	0	0	0	0	0	0
gurgle	4	95.24%	21	20	1	0	0	0	0	0	0
gurgle?	5	100.00%	5	5	0	0	0	0	0	0	0
modulated_cry	6	90.00%	10	9	1	0	0	0	0	0	0
purr	7	100.00%	5	5	0	0	0	0	0	0	0
	TOTAL	92.75%	69	64	5	0	0	0	0	0	0

"whale1": Classification on 12 Test Sequences

Confusion matrix:

		0	1	2	3	4	5	6	7	tests	errors
descending_moan	0	0	0	1	0	0	0	0	0	1	1
descending_shriek	1	0	1	0	0	0	0	0	0	1	0
groan	2	0	0	1	0	0	0	0	0	1	0
groan_+_purr	3	0	0	1	0	0	0	0	0	1	1
gurgle	4	0	0	0	0	4	0	0	0	4	0
gurgle?	5	0	0	0	0	0	1	0	0	1	0
modulated_cry	6	0	0	0	0	0	0	2	0	2	0
purr	7	0	0	0	0	0	0	0	1	1	0

	class	accuracy	tests	candidate order							
descending_moan	0	0.00%	1	0	0	1	0	0	0	0	0
descending_shriek	1	100.00%	1	1	0	0	0	0	0	0	0
groan	2	100.00%	1	1	0	0	0	0	0	0	0
groan_+_purr	3	0.00%	1	0	1	0	0	0	0	0	0
gurgle	4	100.00%	4	4	0	0	0	0	0	0	0
gurgle?	5	100.00%	1	1	0	0	0	0	0	0	0
modulated_cry	6	100.00%	2	2	0	0	0	0	0	0	0
purr	7	100.00%	1	1	0	0	0	0	0	0	0
	TOTAL	83.33%	12	10	1	1	0	0	0	0	0

"whale10": On a Selection of Units from 10 Song Files

- Classes with at least 20 unit instances
- Approximately 80% for training, 20% for a testing
- 752 training song unit instances
- $N = 64$ states, $M = 2048$ symbols
- 13 classes:

`ascending_moan`	`descending_shriek`	`gurgle`	`trill`
`ascending_shriek`	`groan`	`modulated_cry`	
`cry`	`grunt`	`modulated_moan`	
`descending_moan`	`grunts`	`purr`	

"whale10": Classification on 752 Training Sequences

Confusion matrix:		0	1	2	3	4	5	6	7	8	9	10	11	12	tests	errors
ascending_moan	0	32	0	1	0	0	0	0	0	0	0	0	0	0	33	1
ascending_shriek	1	0	74	0	0	0	0	0	0	1	0	0	0	0	75	1
cry	2	0	0	42	0	0	0	0	0	0	0	0	0	0	42	0
descending_moan	3	0	0	5	116	0	0	0	0	0	0	2	0	0	123	7
descending_shriek	4	0	0	0	0	21	0	0	0	0	0	0	0	0	21	0
groan	5	0	0	0	0	0	32	0	0	0	0	0	0	0	32	0
grunt	6	0	0	0	0	0	0	80	0	1	0	0	0	0	81	1
grunts	7	0	0	0	0	0	0	0	19	0	0	0	0	0	19	0
gurgle	8	0	0	0	0	0	1	0	0	172	0	0	0	0	173	1
modulated_cry	9	0	0	0	0	0	0	0	0	0	23	0	0	0	23	0
modulated_moan	10	0	0	0	0	0	1	0	0	0	0	66	0	0	67	1
purr	11	0	0	1	0	0	0	0	0	0	0	0	28	0	29	1
trill	12	0	0	0	0	0	0	0	0	0	0	1	0	33	34	1

	class	accuracy	tests	candidate order												
ascending_moan	0	96.97%	33	32	0	0	0	0	0	1	0	0	0	0	0	0
ascending_shriek	1	98.67%	75	74	0	0	0	0	0	0	0	0	0	0	1	0
cry	2	100.00%	42	42	0	0	0	0	0	0	0	0	0	0	0	0
descending_moan	3	94.31%	123	116	4	2	0	0	1	0	0	0	0	0	0	0
descending_shriek	4	100.00%	21	21	0	0	0	0	0	0	0	0	0	0	0	0
groan	5	100.00%	32	32	0	0	0	0	0	0	0	0	0	0	0	0
grunt	6	98.77%	81	80	0	1	0	0	0	0	0	0	0	0	0	0
grunts	7	100.00%	19	19	0	0	0	0	0	0	0	0	0	0	0	0
gurgle	8	99.42%	173	172	1	0	0	0	0	0	0	0	0	0	0	0
modulated_cry	9	100.00%	23	23	0	0	0	0	0	0	0	0	0	0	0	0
modulated_moan	10	98.51%	67	66	0	1	0	0	0	0	0	0	0	0	0	0
purr	11	96.55%	29	28	0	0	0	1	0	0	0	0	0	0	0	0
trill	12	97.06%	34	33	0	0	1	0	0	0	0	0	0	0	0	0
	TOTAL	98.14%	752	738	5	4	1	1	1	1	0	0	0	0	1	0

Classification on 178 Test Sequences

Confusion matrix:		0	1	2	3	4	5	6	7	8	9	10	11	12	tests	errors
ascending_moan	0	1	0	4	1	0	0	0	1	0	0	0	0	0	7	6
ascending_shriek	1	1	11	0	1	0	0	0	0	5	0	0	0	0	18	7
cry	2	0	0	4	2	0	0	0	0	1	0	1	2	0	10	6
descending_moan	3	1	2	2	17	0	1	0	2	0	0	1	1	3	30	13
descending_shriek	4	0	1	0	0	4	0	0	0	0	0	0	0	0	5	1
groan	5	0	0	1	0	0	3	0	0	3	0	0	0	0	7	4
grunt	6	0	0	0	1	0	2	13	0	3	0	1	0	0	20	7
grunts	7	0	0	0	0	0	0	0	4	0	0	0	0	0	4	0
gurgle	8	0	1	3	3	0	2	2	0	27	0	2	1	1	42	15
modulated_cry	9	0	0	0	0	1	0	0	0	0	4	0	0	0	5	1
modulated_moan	10	0	0	2	3	0	1	1	0	0	0	9	0	0	16	7
purr	11	0	0	0	1	0	1	0	0	2	0	0	2	0	6	4
trill	12	0	0	0	1	0	1	1	0	1	1	0	0	3	8	5

	class	accuracy	tests	candidate order												
ascending_moan	0	14.29%	7	1	2	0	0	2	1	0	0	0	1	0	0	0
ascending_shriek	1	61.11%	18	11	0	2	2	0	0	0	2	0	1	0	0	0
cry	2	40.00%	10	4	2	1	1	1	1	0	0	0	0	0	0	0
descending_moan	3	56.67%	30	17	6	1	3	3	0	0	0	0	0	0	0	0
descending_shriek	4	80.00%	5	4	1	0	0	0	0	0	0	0	0	0	0	0
groan	5	42.86%	7	3	0	2	2	0	0	0	0	0	0	0	0	0
grunt	6	65.00%	20	13	4	0	2	1	0	0	0	0	0	0	0	0
grunts	7	100.00%	4	4	0	0	0	0	0	0	0	0	0	0	0	0
gurgle	8	64.29%	42	27	9	1	0	1	1	1	1	1	0	0	0	0
modulated_cry	9	80.00%	5	4	1	0	0	0	0	0	0	0	0	0	0	0
modulated_moan	10	56.25%	16	9	2	2	1	0	1	1	0	0	0	0	0	0
purr	11	33.33%	6	2	1	1	1	0	1	0	0	0	0	0	0	0
trill	12	37.50%	8	3	0	1	0	1	0	1	1	0	0	0	1	0
	TOTAL	57.30%	178	102	28	11	12	9	5	3	4	1	2	0	1	0

Some Remarks

- Exercises so far mainly intended to validate the software revision
- Labelled data used as given
- SNR varies significantly
- Large scale model training and tuning not considered at all
- Lots of interesting approaches out there!