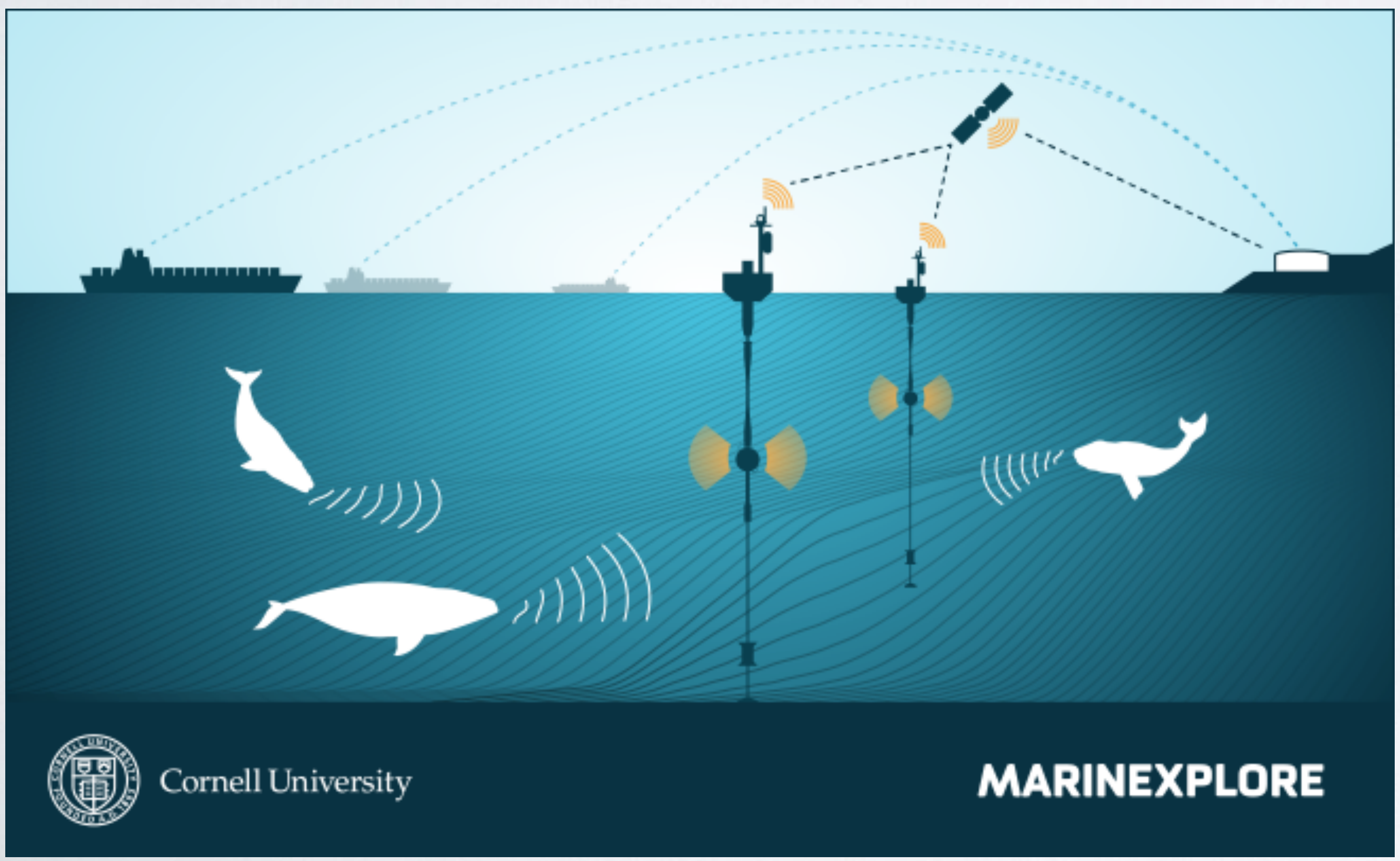
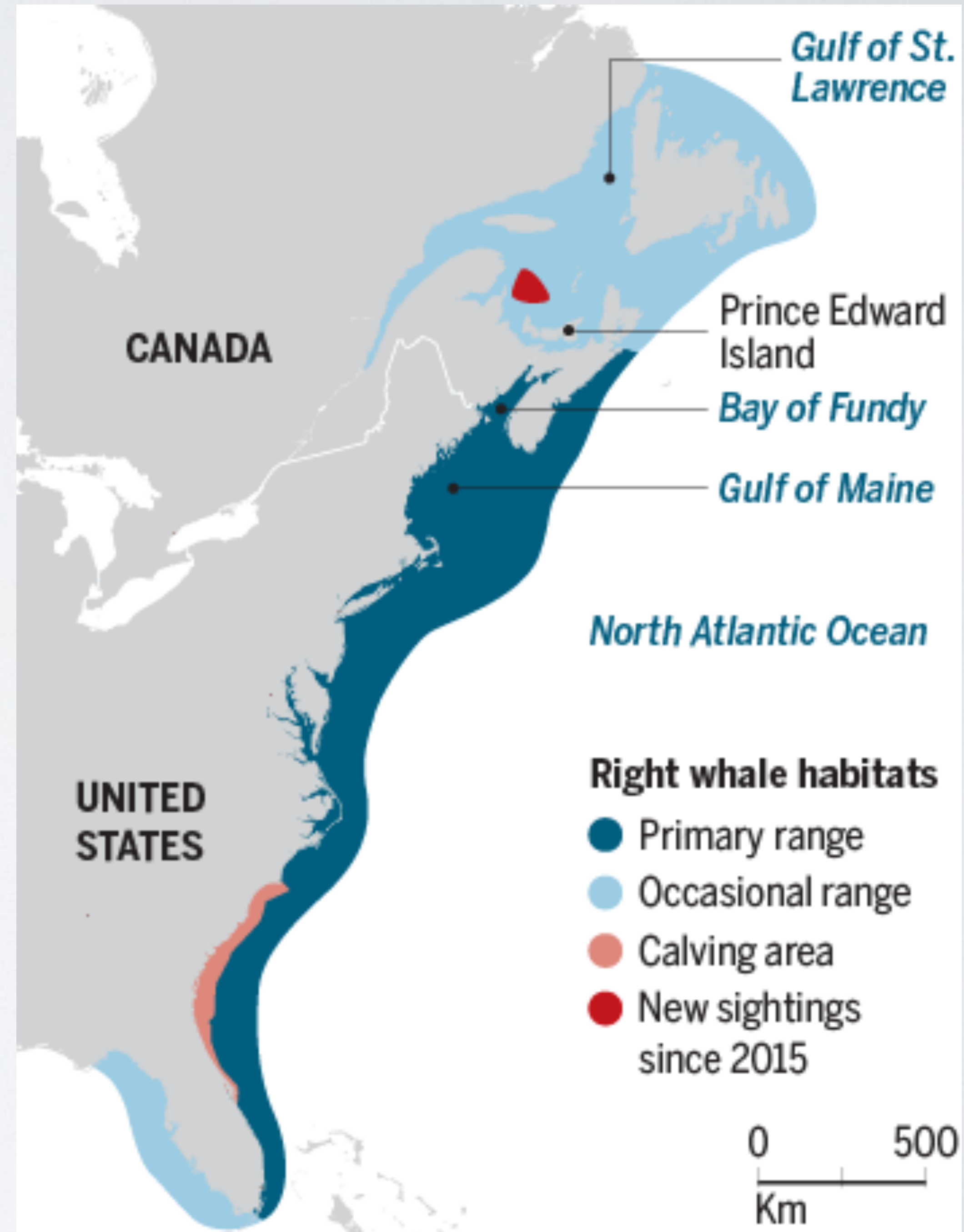
MOBY DICK:
WHALE DETECTION

Alberto Mur
Javier Antorán

THE RIGHT WHALE: (NARW)



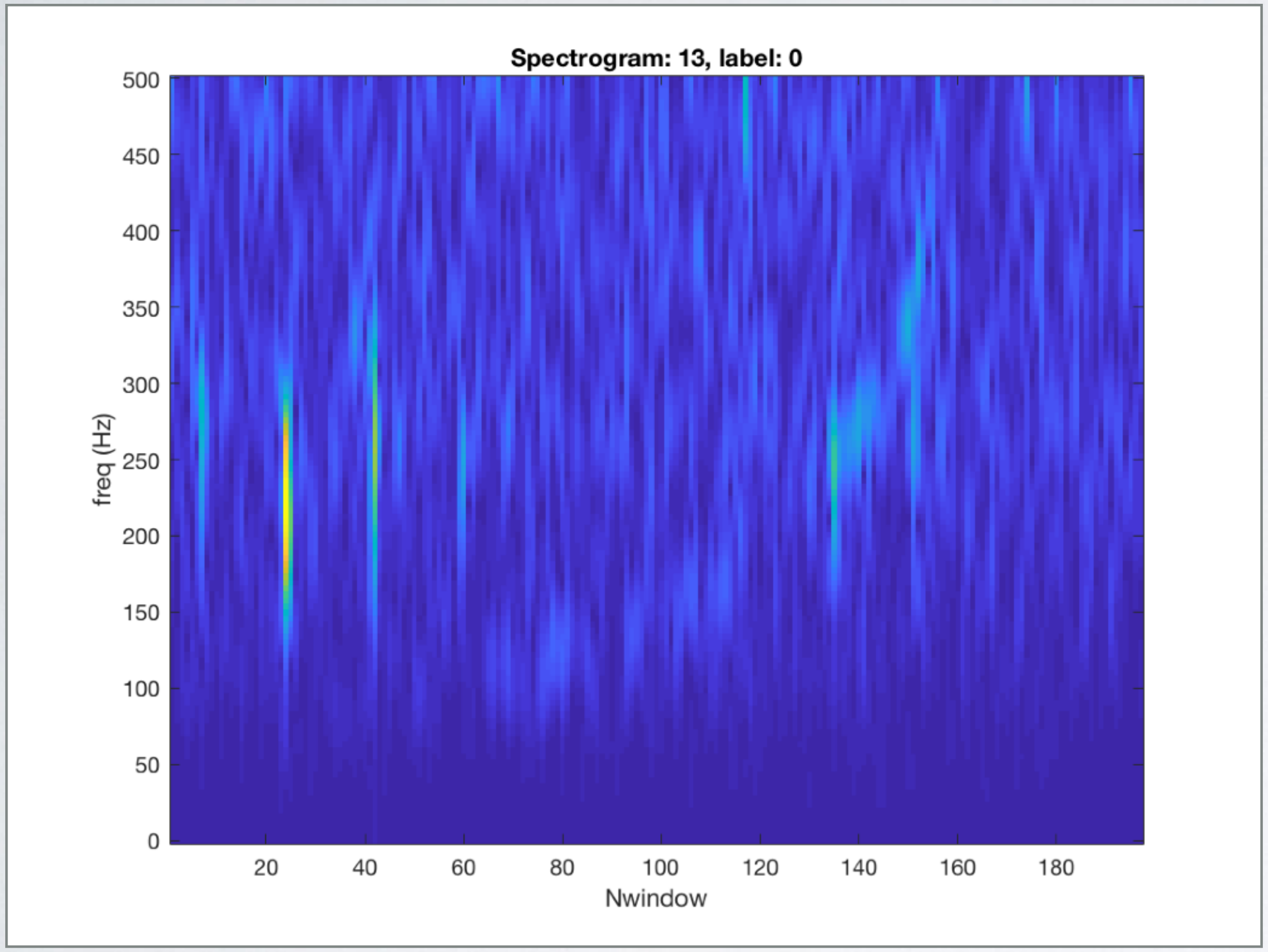


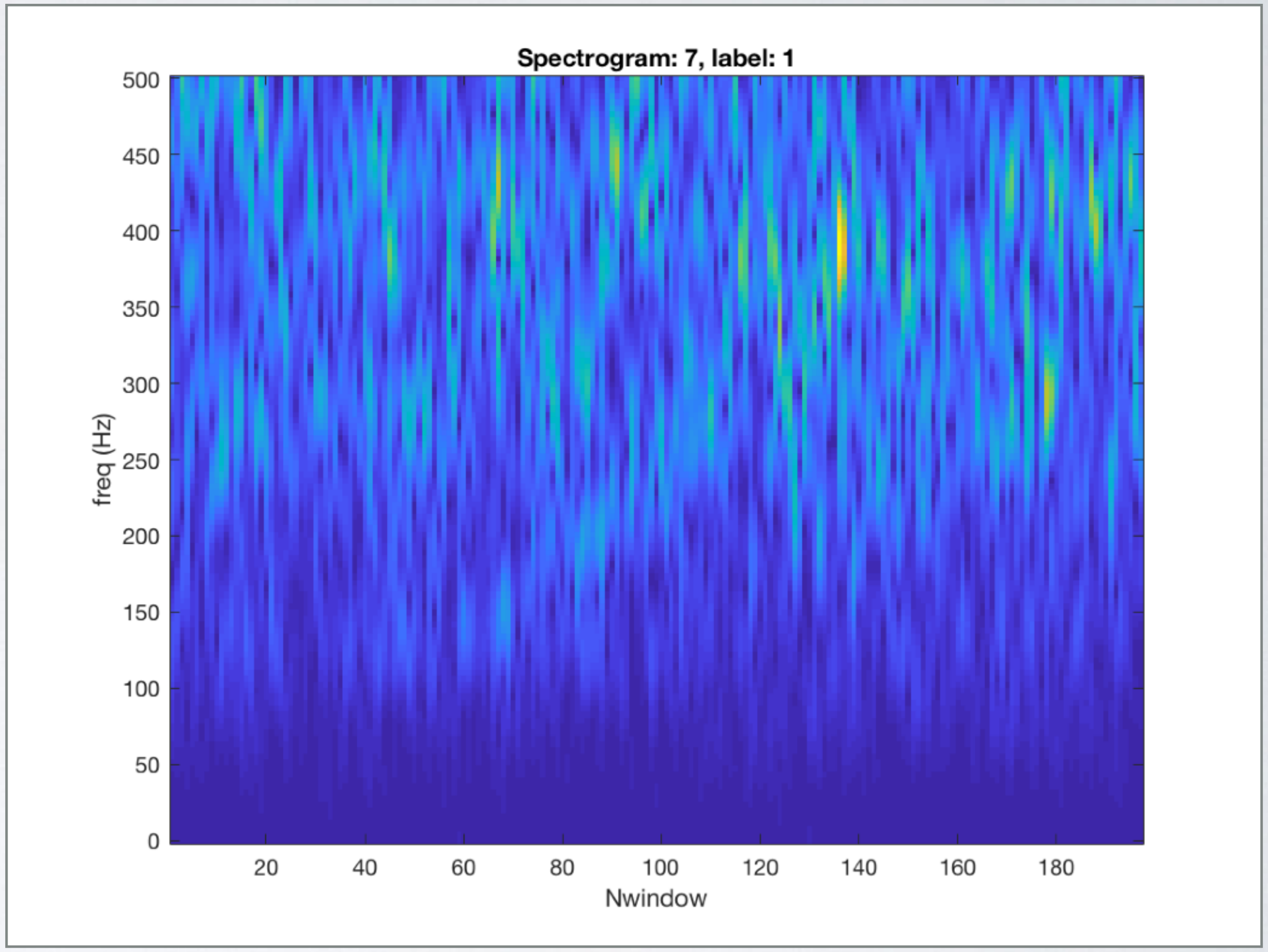
CORNELL CHALLENGE

- 2s audio clips x Fs: 2k = 4000 samples/audio
- 30000 train audio clips (7027 positives)
- 70000 test audio clips (no labels) [Unused]
- Winning result: 0.9834 ROC-AUC (template method)
- https://vimeo.com/227009627

kaggle™

UPCALL SPECTROGRAM





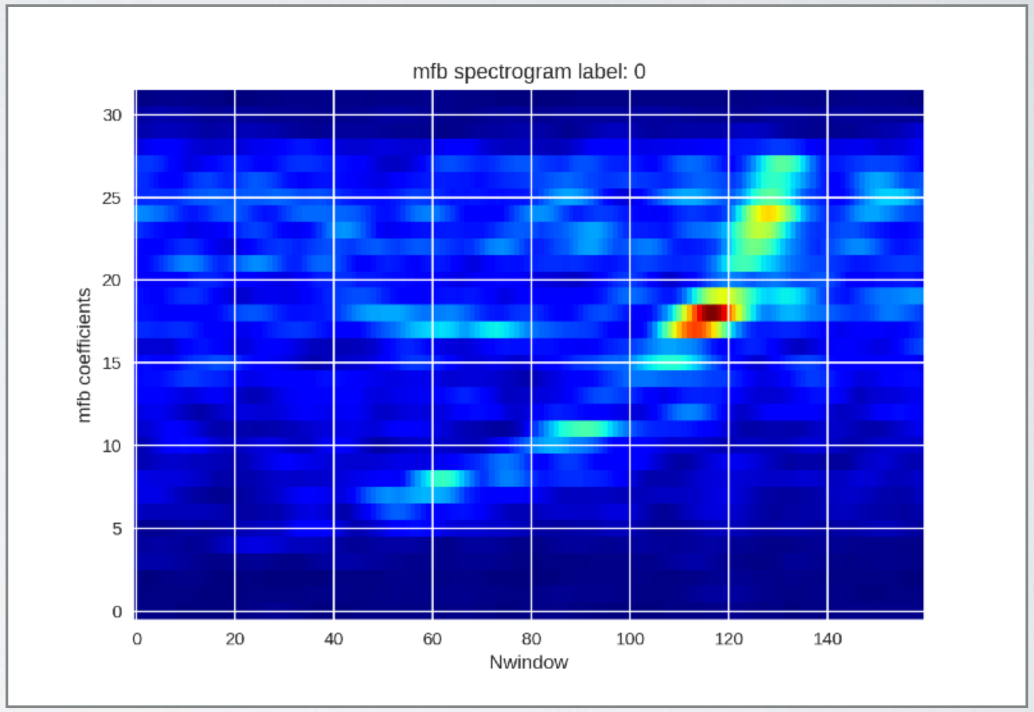
STRATEGIES & TECHNOLOGIES

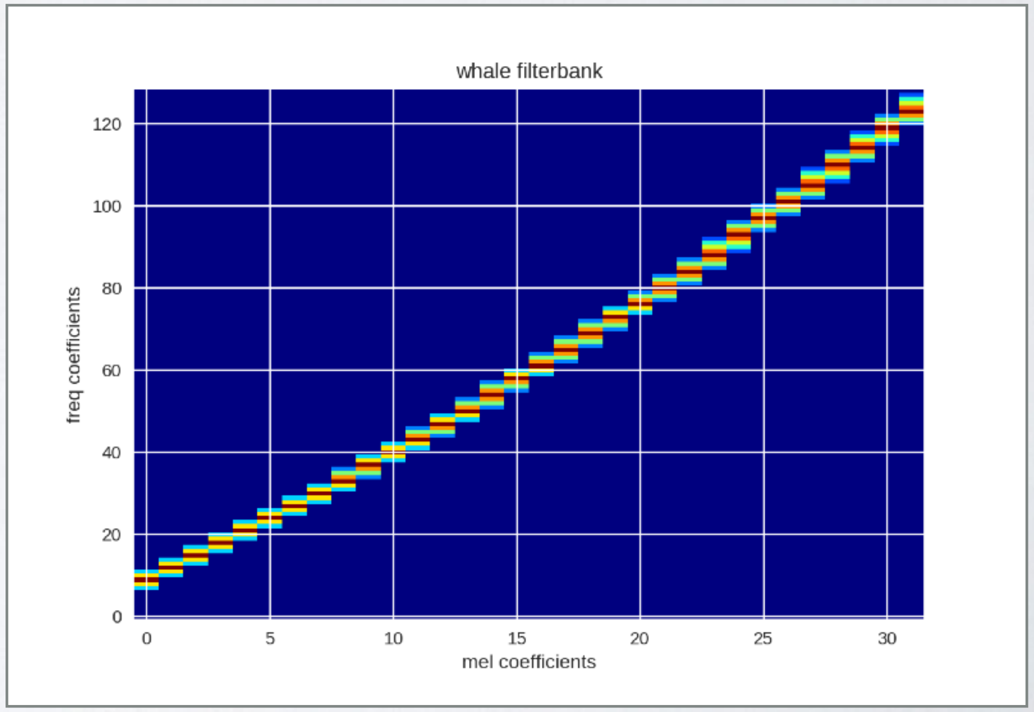
- Manual Feature Engineering
- CNN
- HMM + GMM / Normalizing Flows
- Gradient Boosting

Computing Cluster

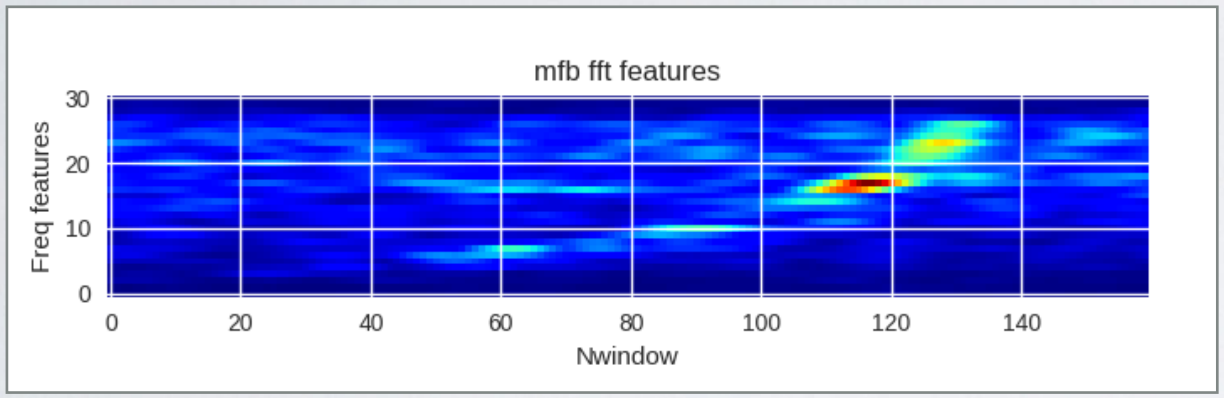
CNN FEATURE ENGINEERING: SPECTROGRAM

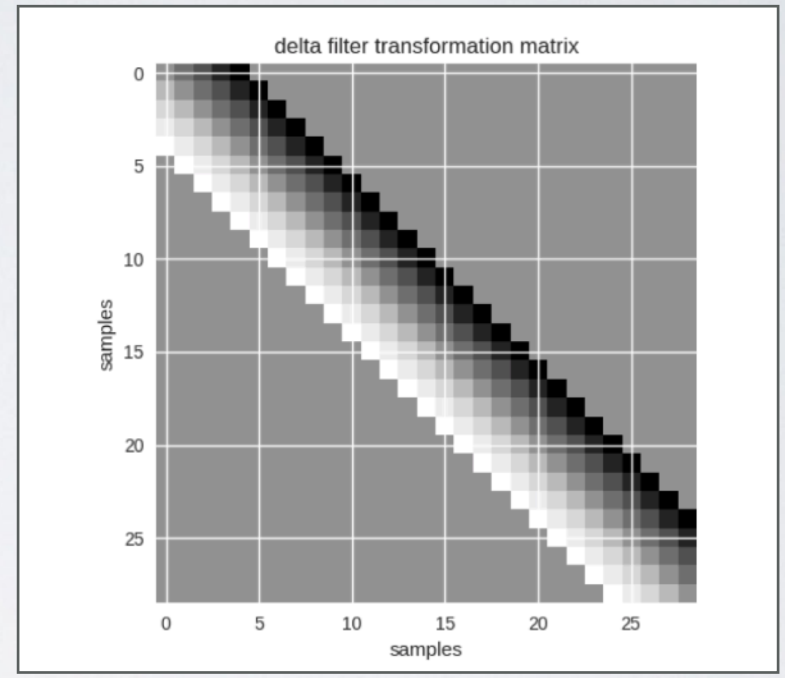
- Downsampling from 2k to 1k.
  Call range: 50-450 Hz
- Hamming window
- Whale-filterbank ~ mfb
  (coefficient reduction)
- Multiple time scale analysis.
- Win duration: 250ms, 11ms advance

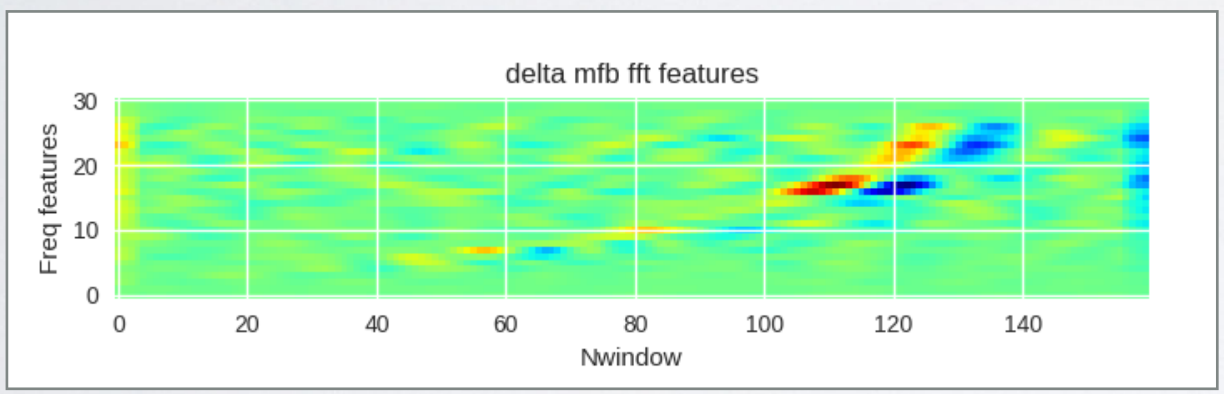


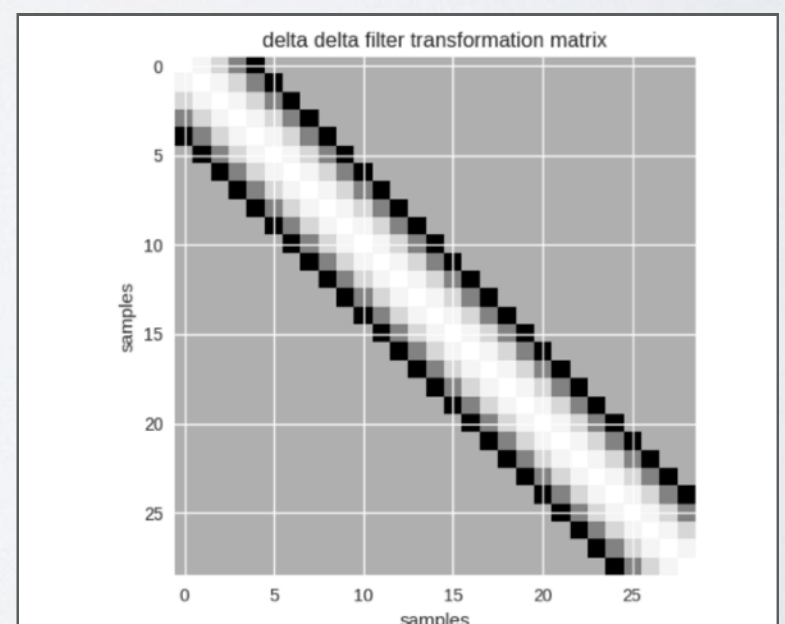


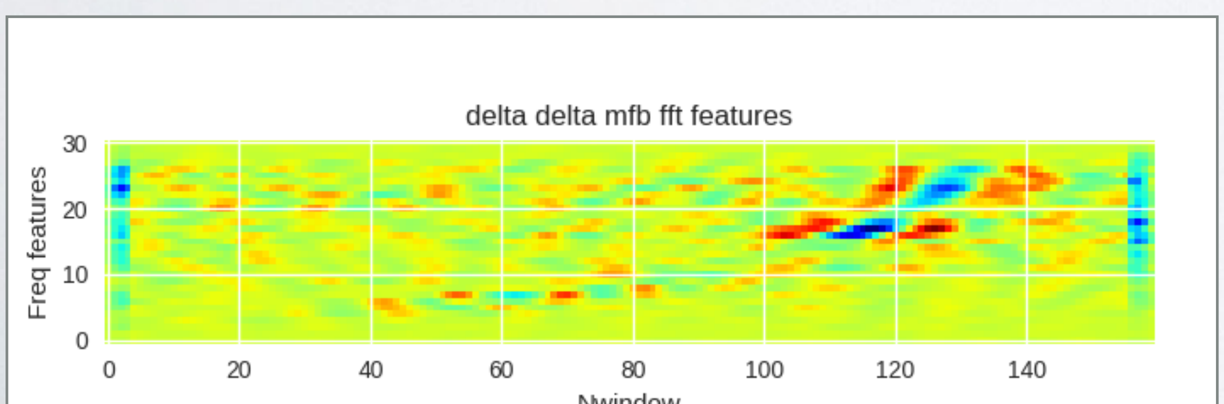
CNN FEATURE ENGINEERING: DELTA FEATURES





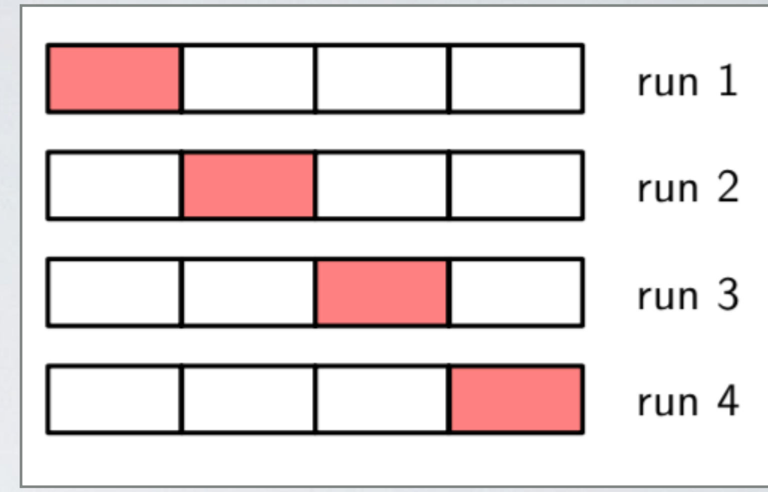


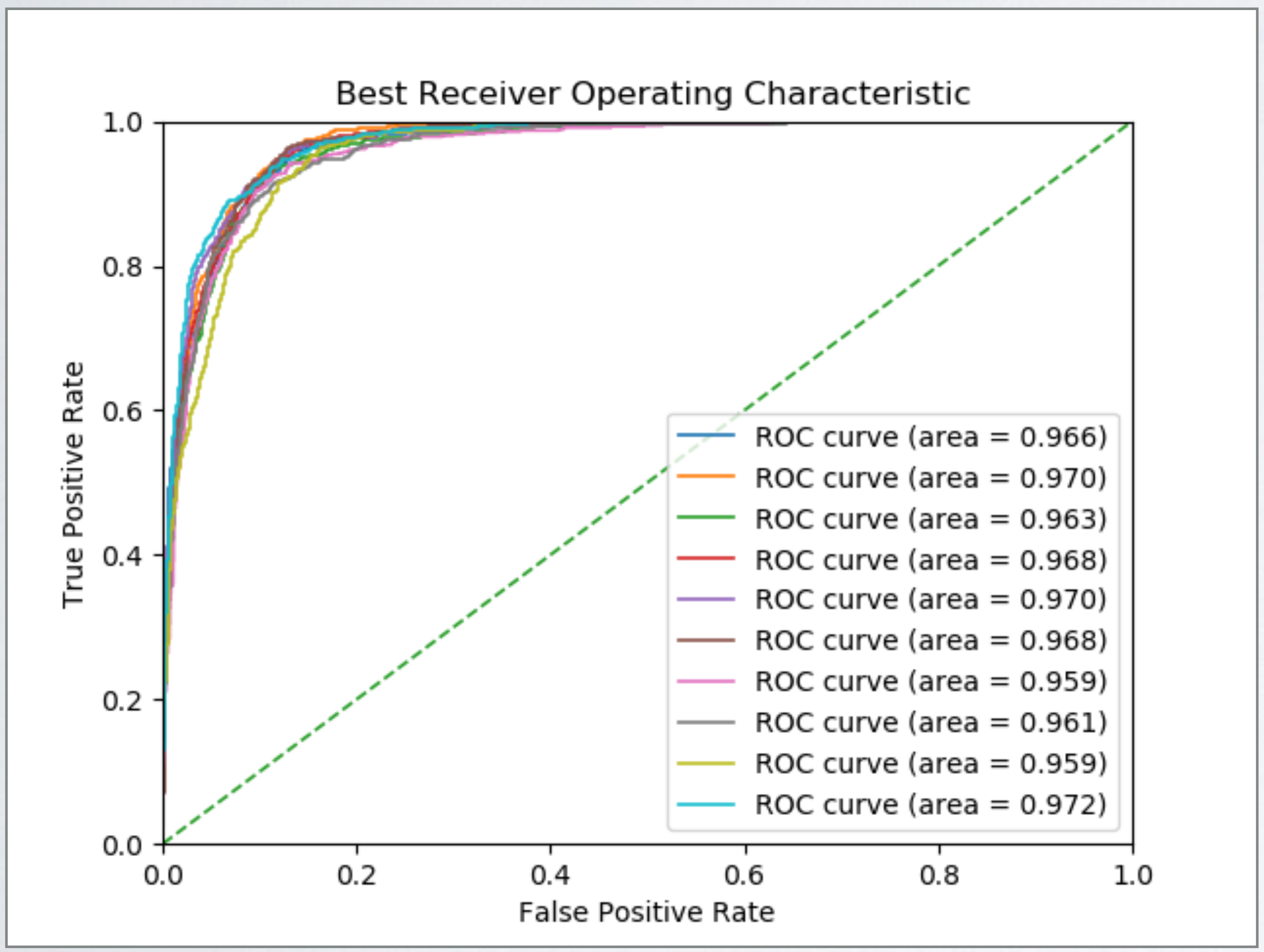




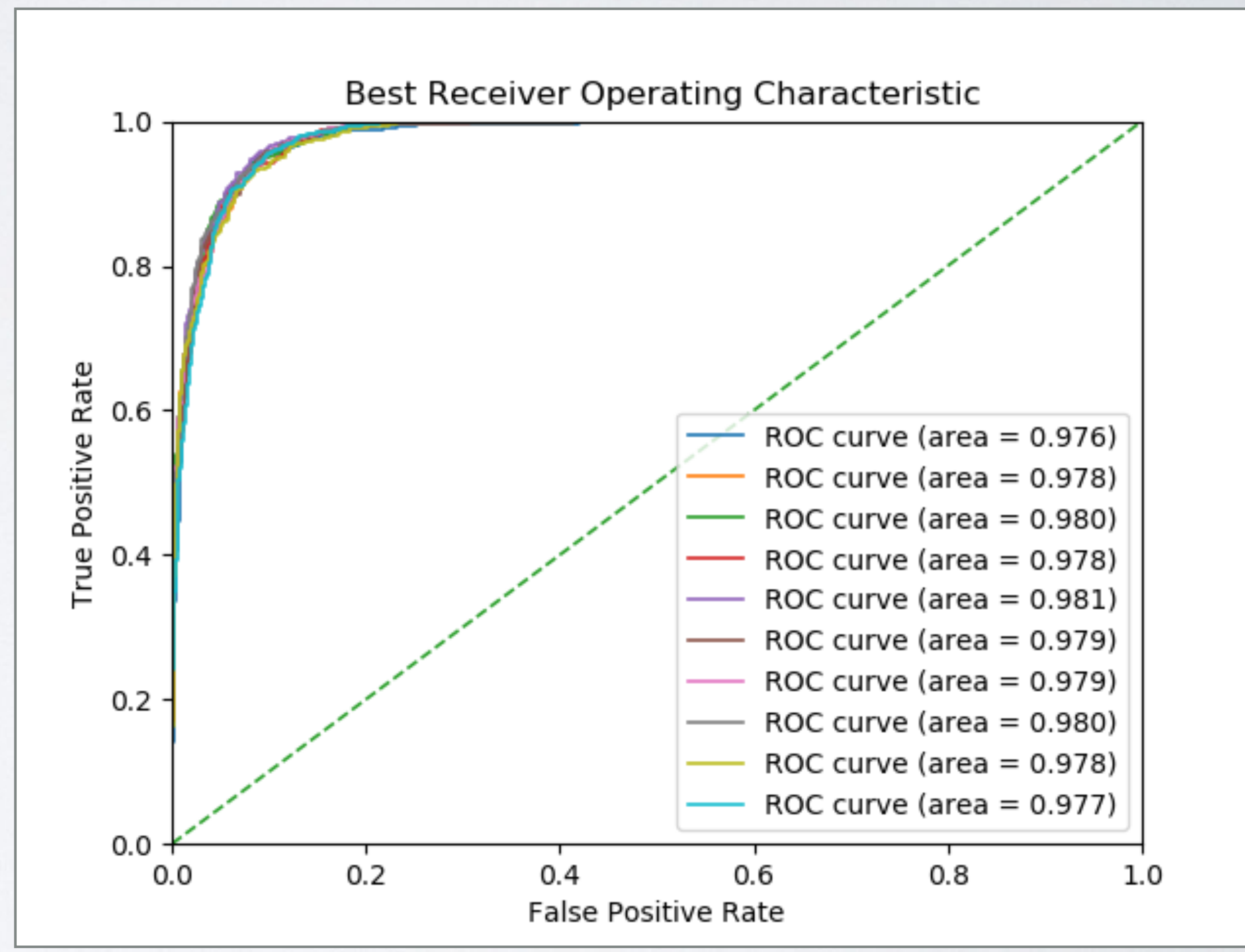
CNN RESULTS

- 10-fold cross validation





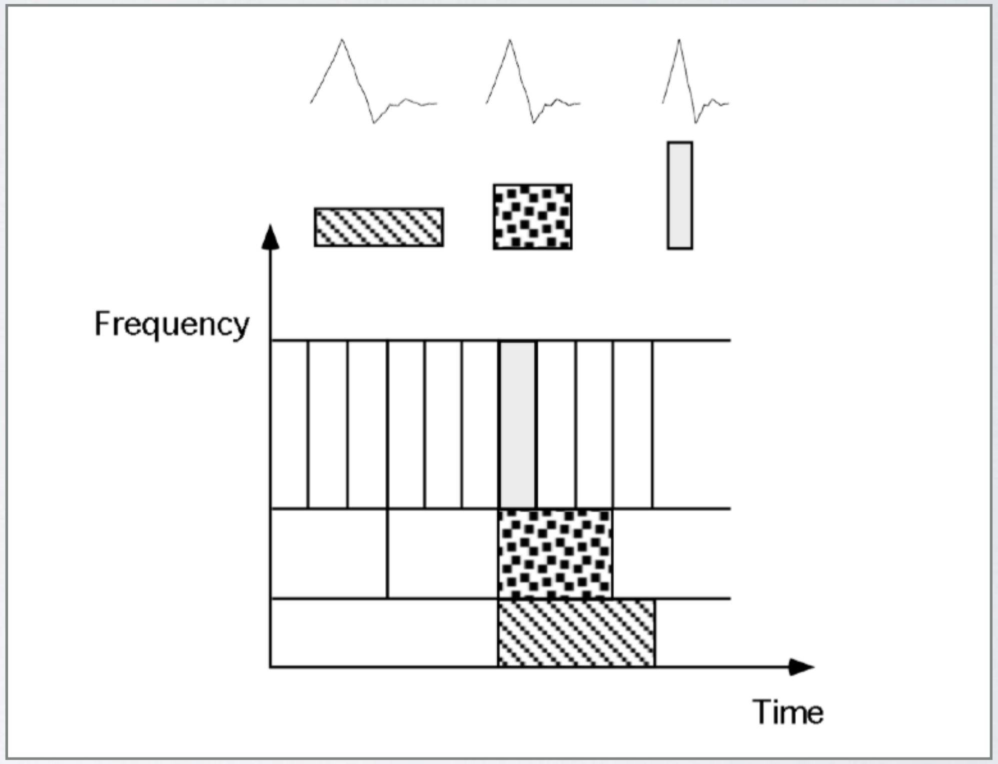
25ms window



250ms window

HMM FEATURE ENGINEERING: WAVELET TRANSFORM (DWT)

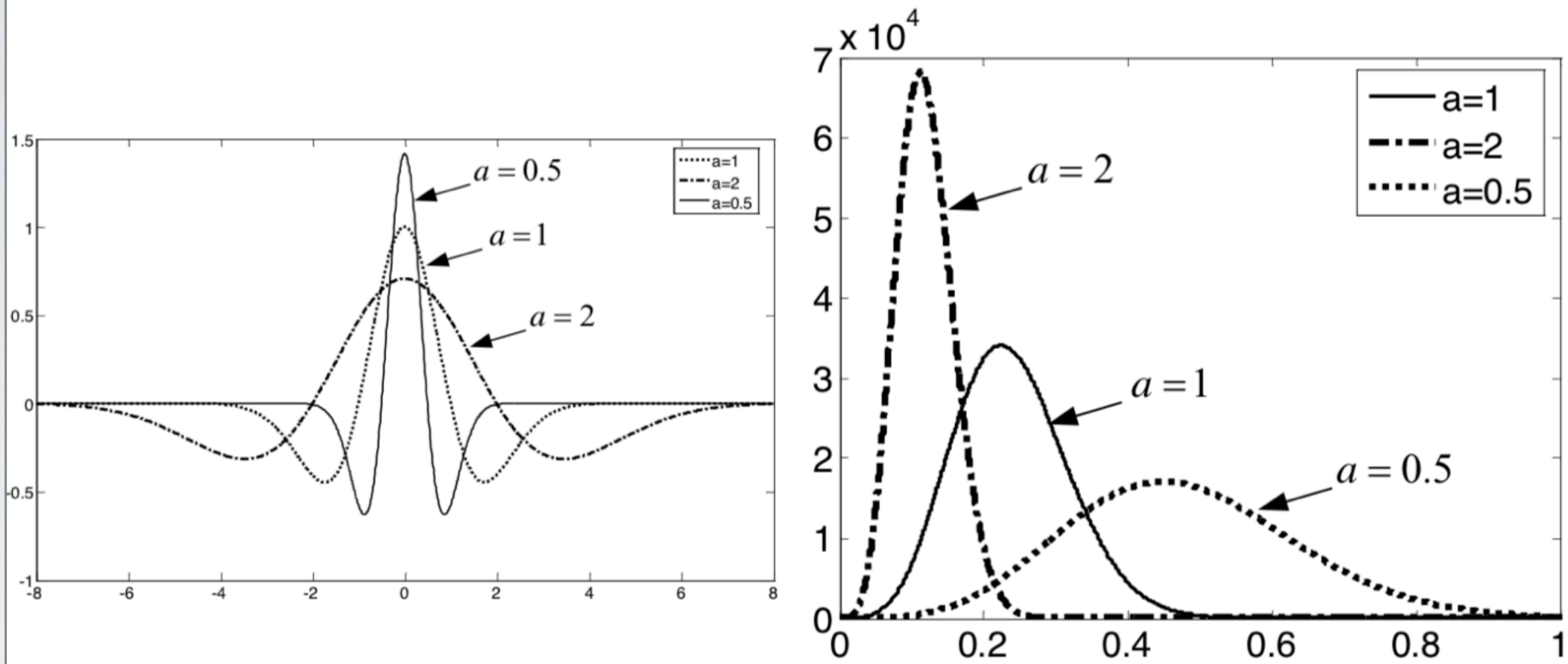
- Spectrogram's resolution is frequency dependent
- Heisenberg's uncertainty theorem
- Frequency dependent window size
- Use different wavelet functions



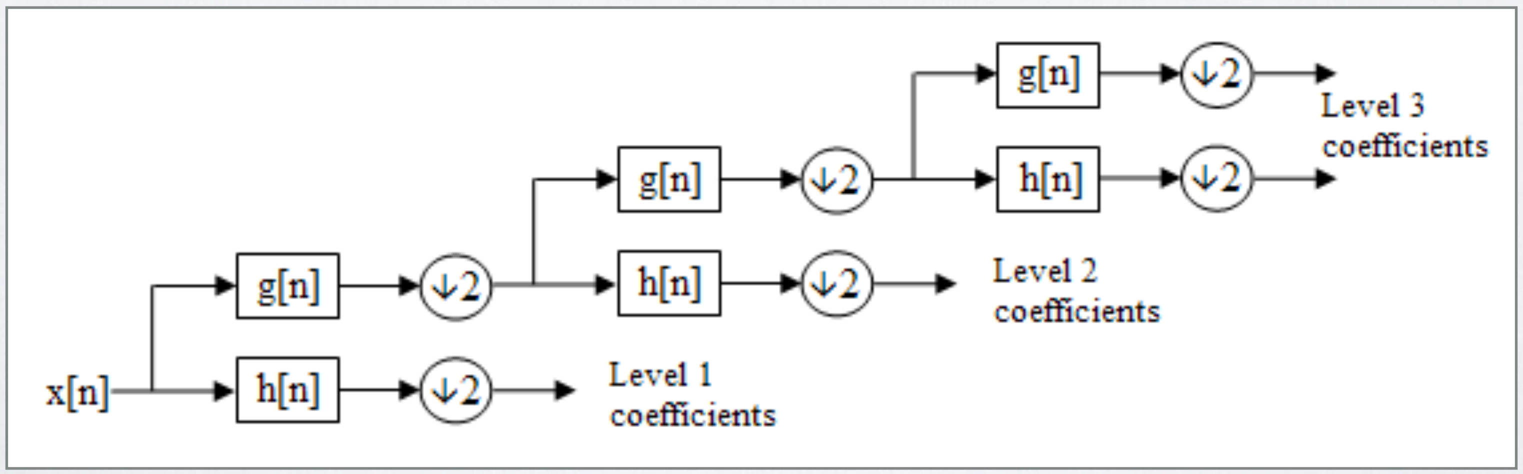
$$\Delta t \Delta f \geq \frac{1}{4\pi}$$

$$CWT(\tau, \alpha) = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{\infty} f(t) g^*(\frac{t-\tau}{\alpha}) e^{-2\pi i k 0 \frac{t-\tau}{\alpha}} dt$$

DWT: MULTI SCALE FILTERING

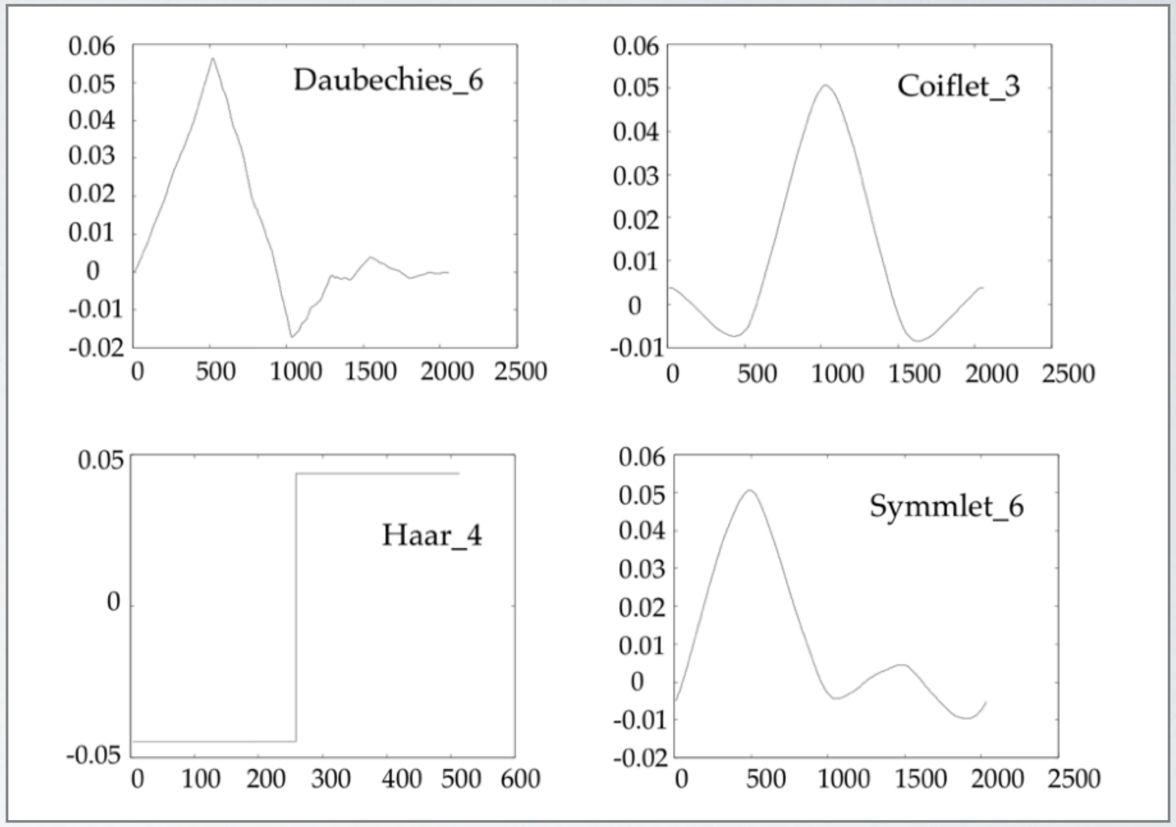


Low frequency components => long duration
High frequency components => short duration



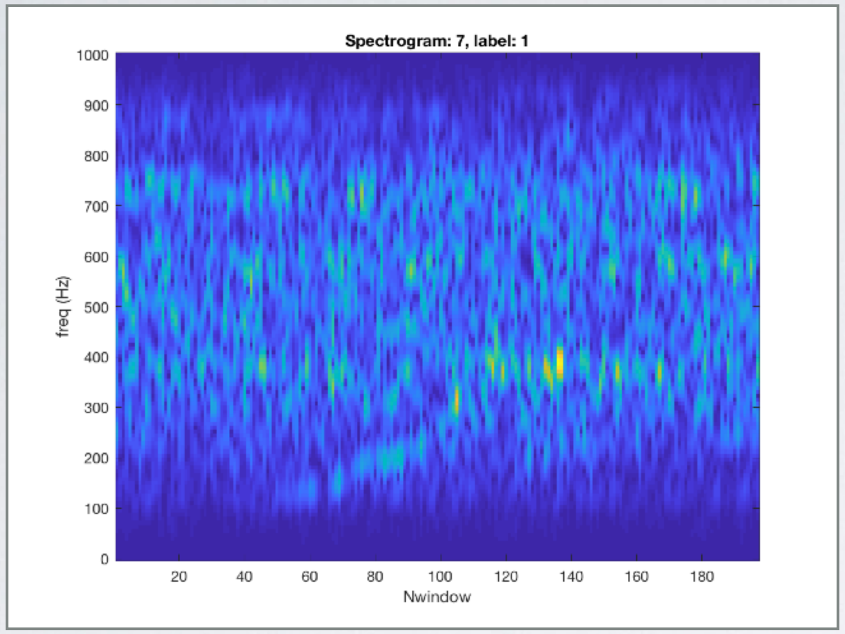
DWT WAVELETS

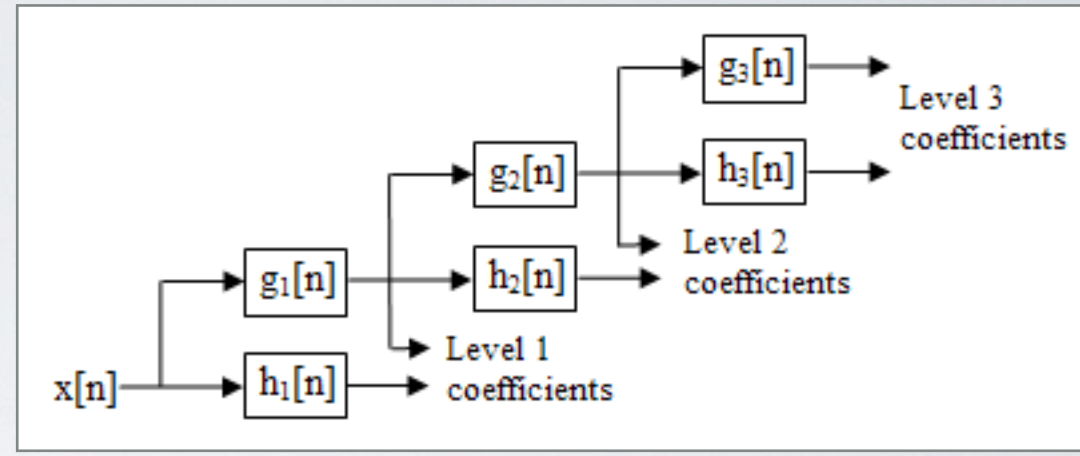
- BPF-LPF wavelet filters
- We use: sym8, db4
- Low-pass averaging filters
- Used for Denoising & Compression



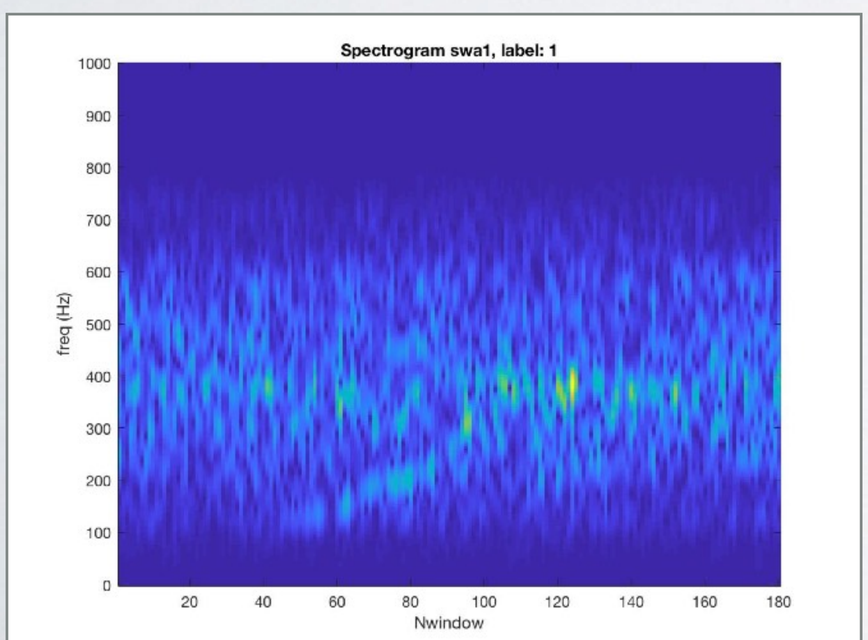
$$CWT_h(\tau, \alpha) = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{\infty} f(t) h^*\left(\frac{t-\tau}{\alpha}\right) dt$$

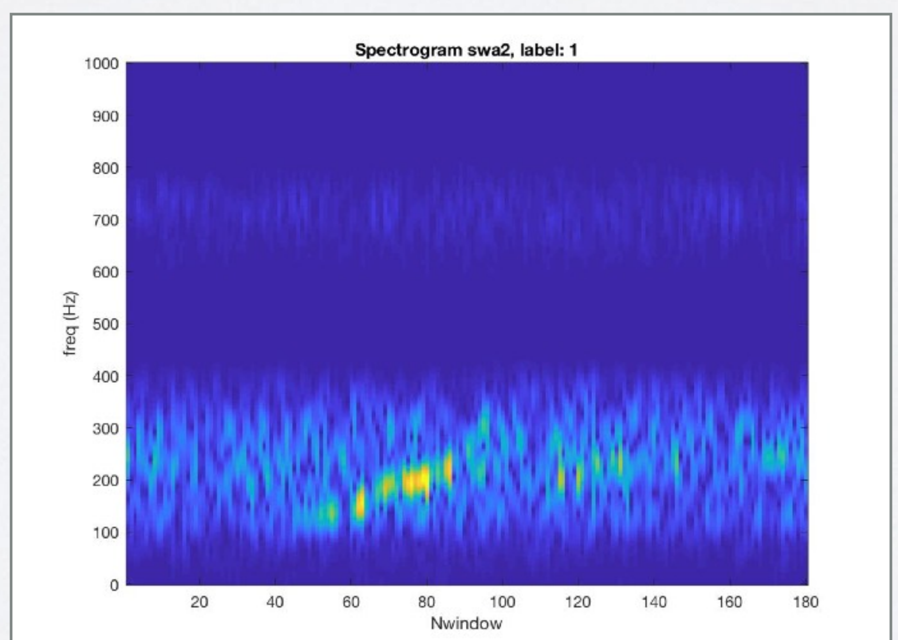
SWT: MULTIRESOLUTION

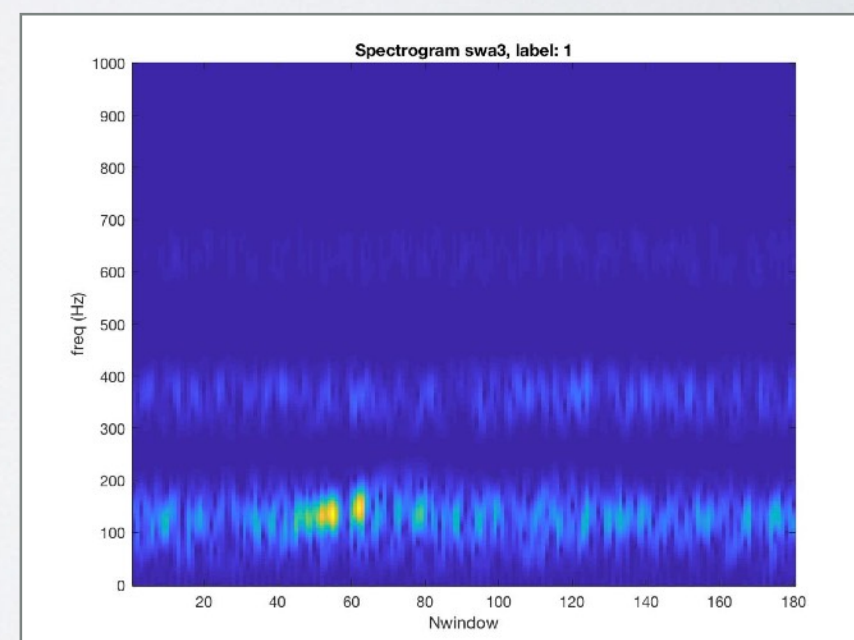




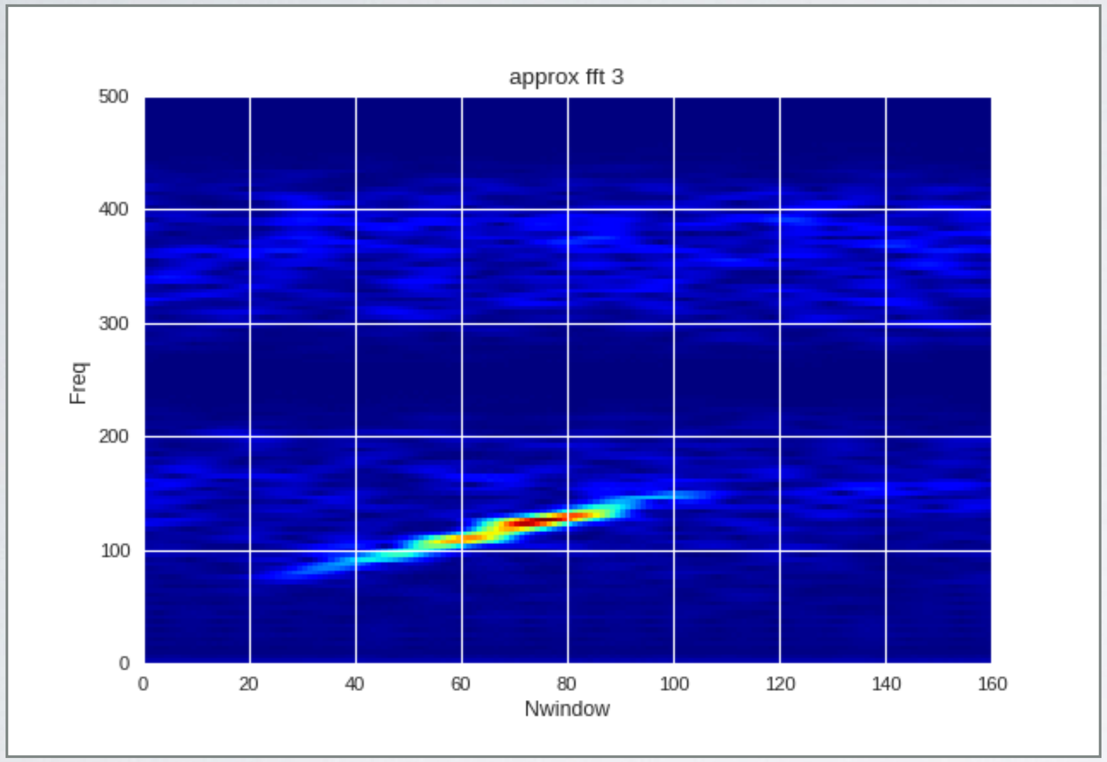
- Approximation coefficients 1,2,3

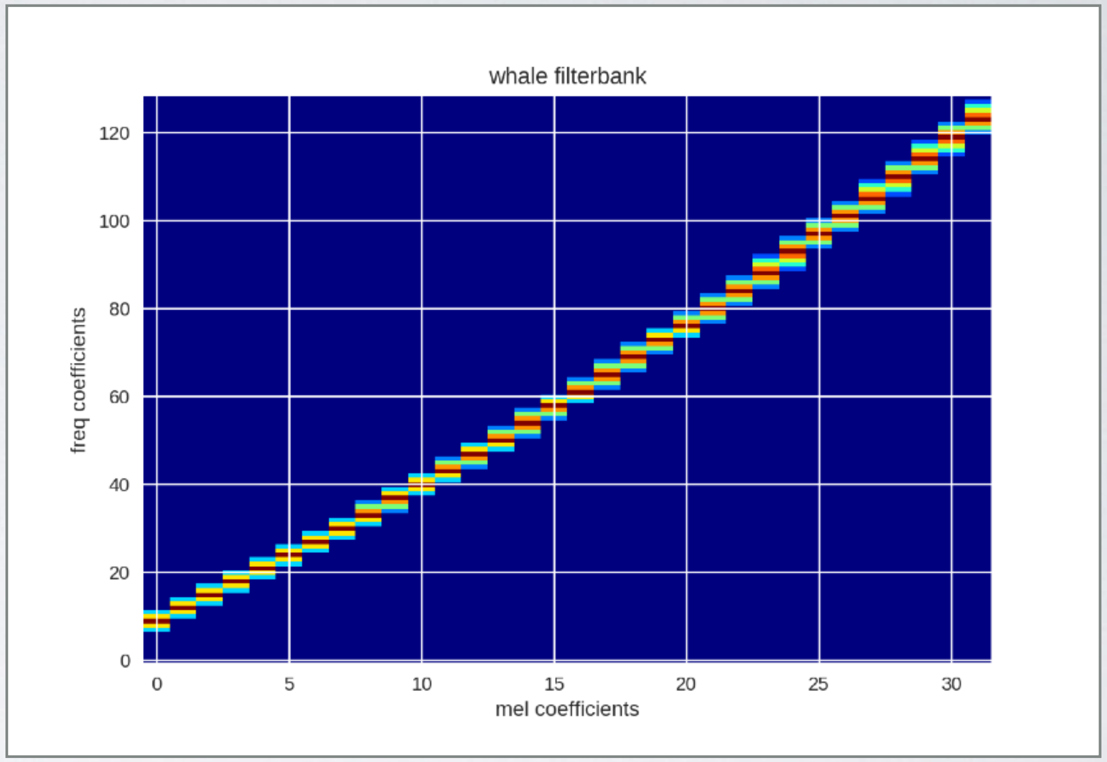


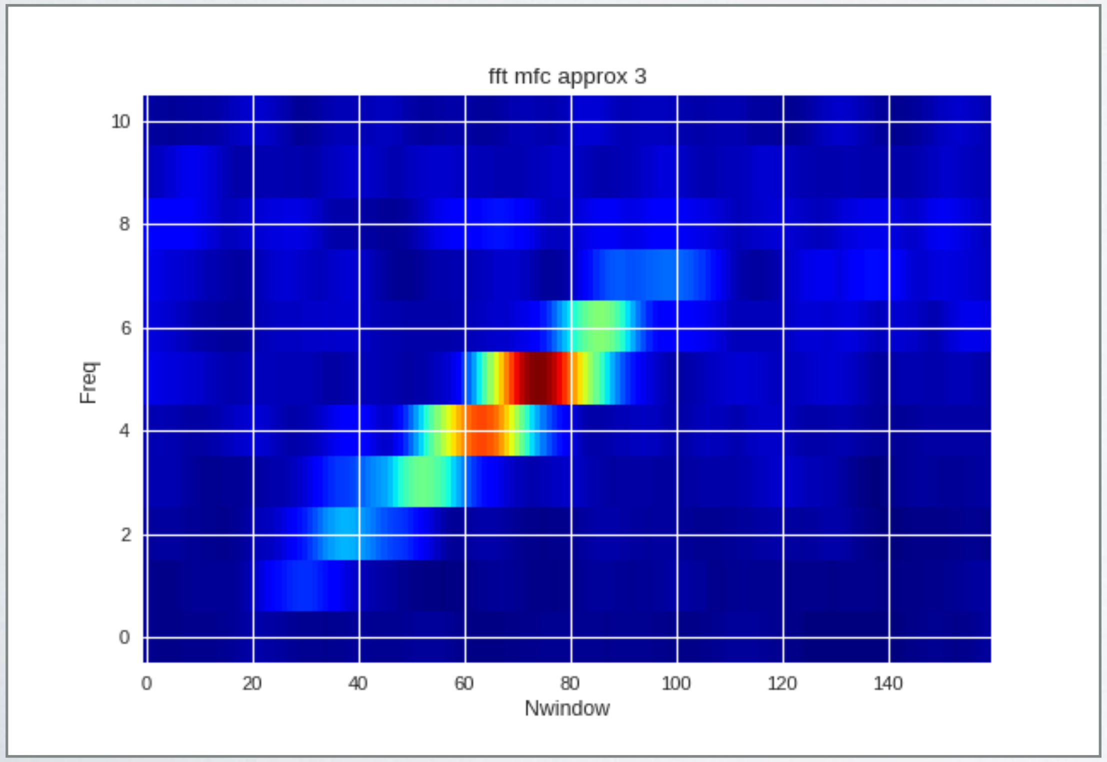


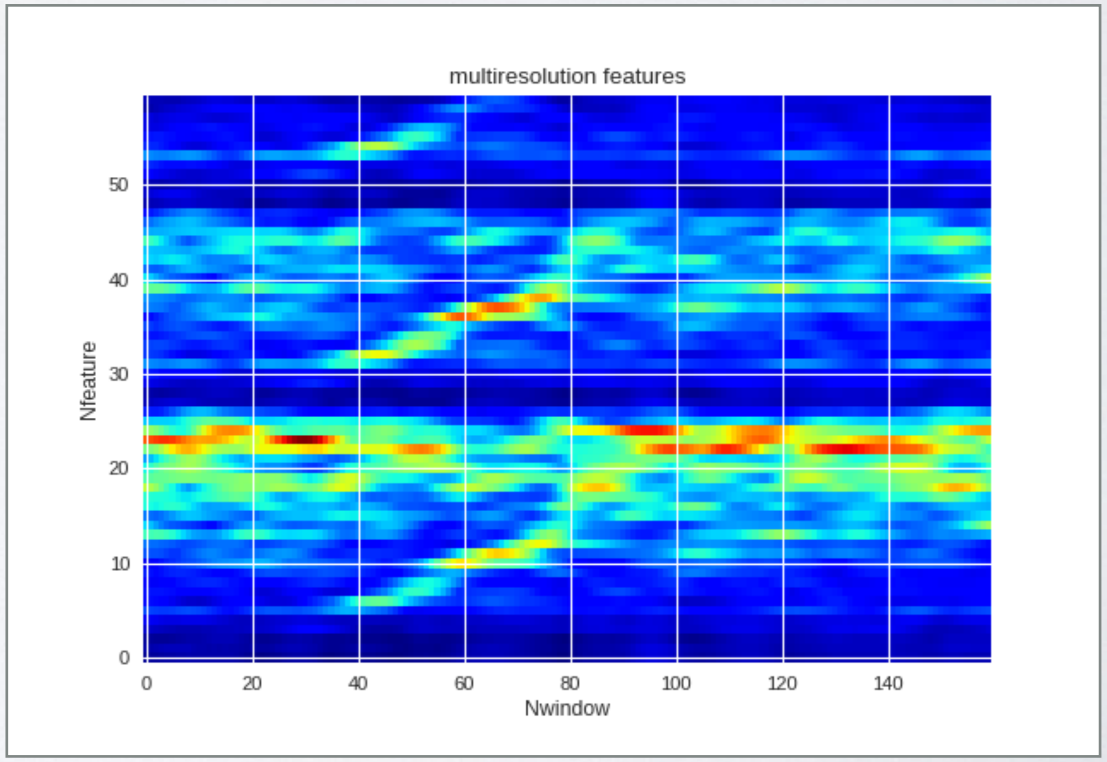


SWT: FEATURE EXTRACTION



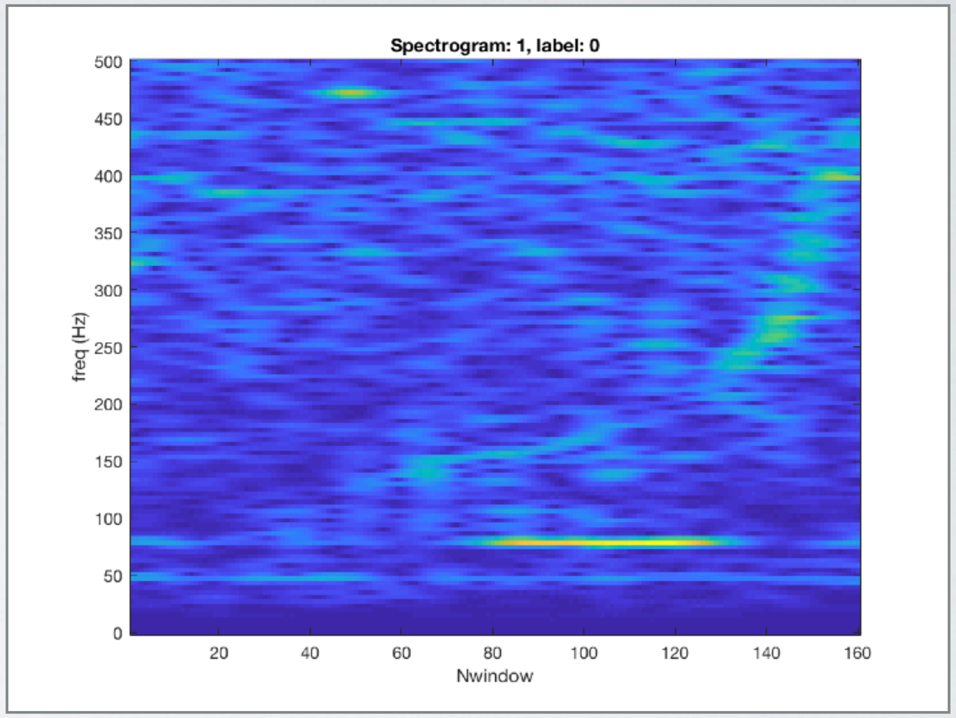


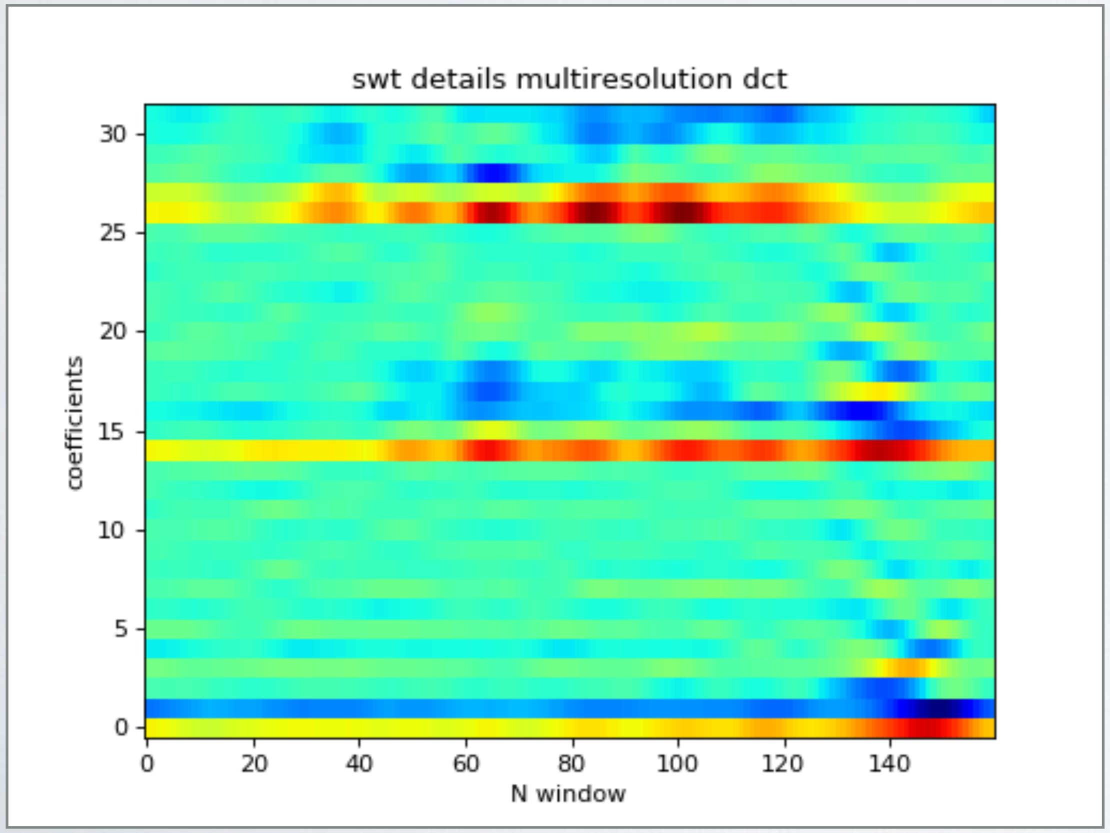


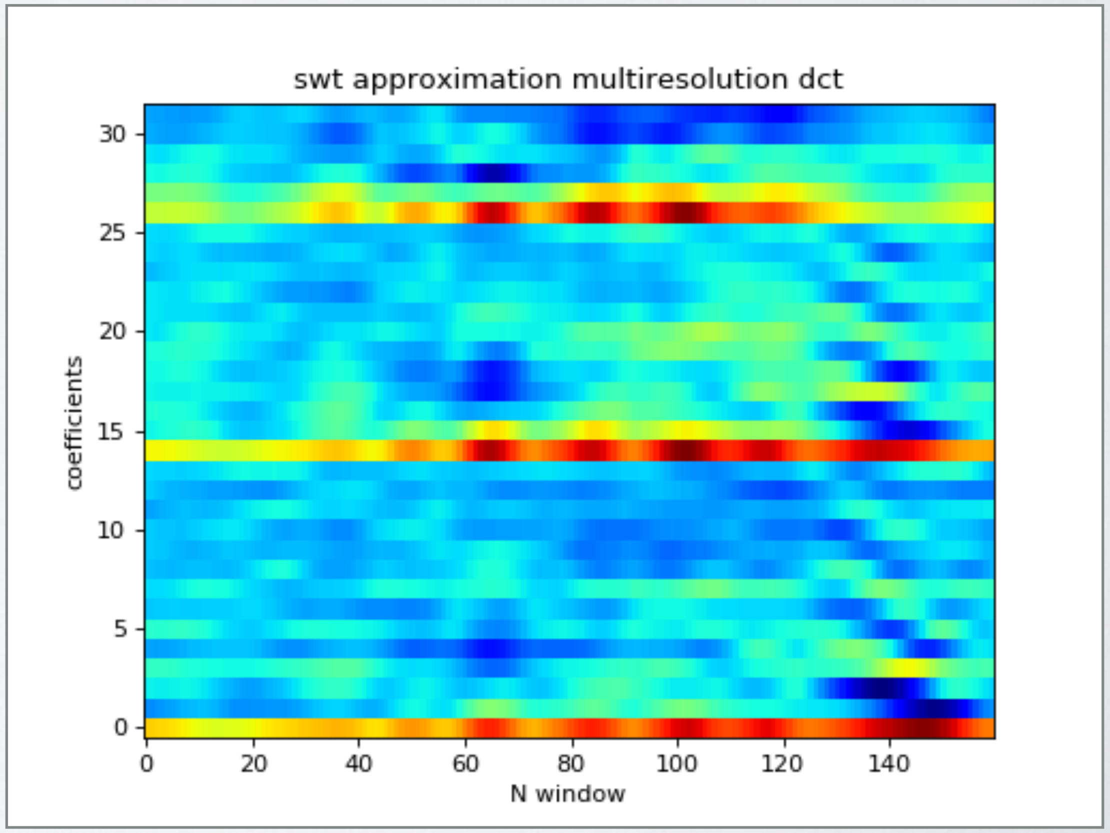


RESULTING FEATURES

- DCT decorrelation
- Parallel Generation
- Mean, Std normalized

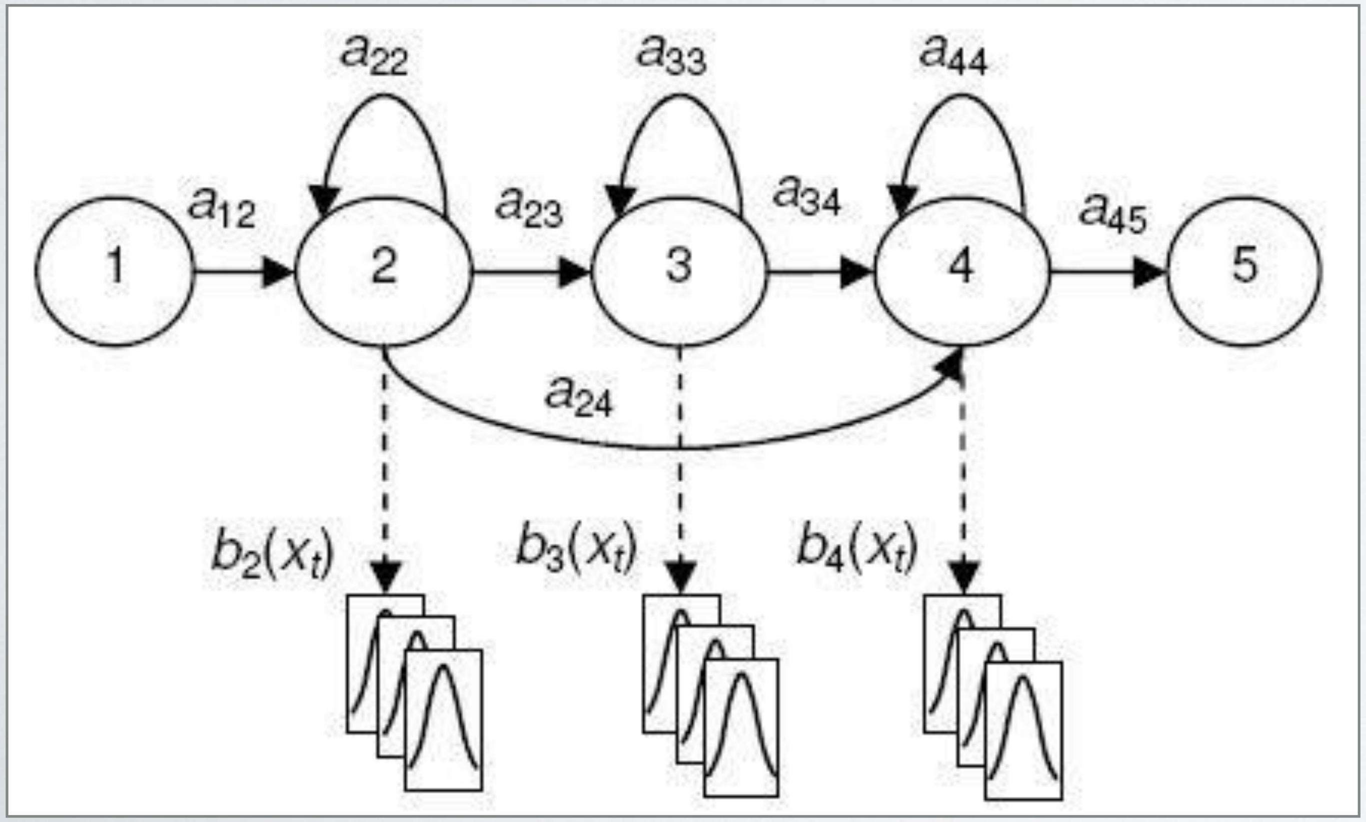


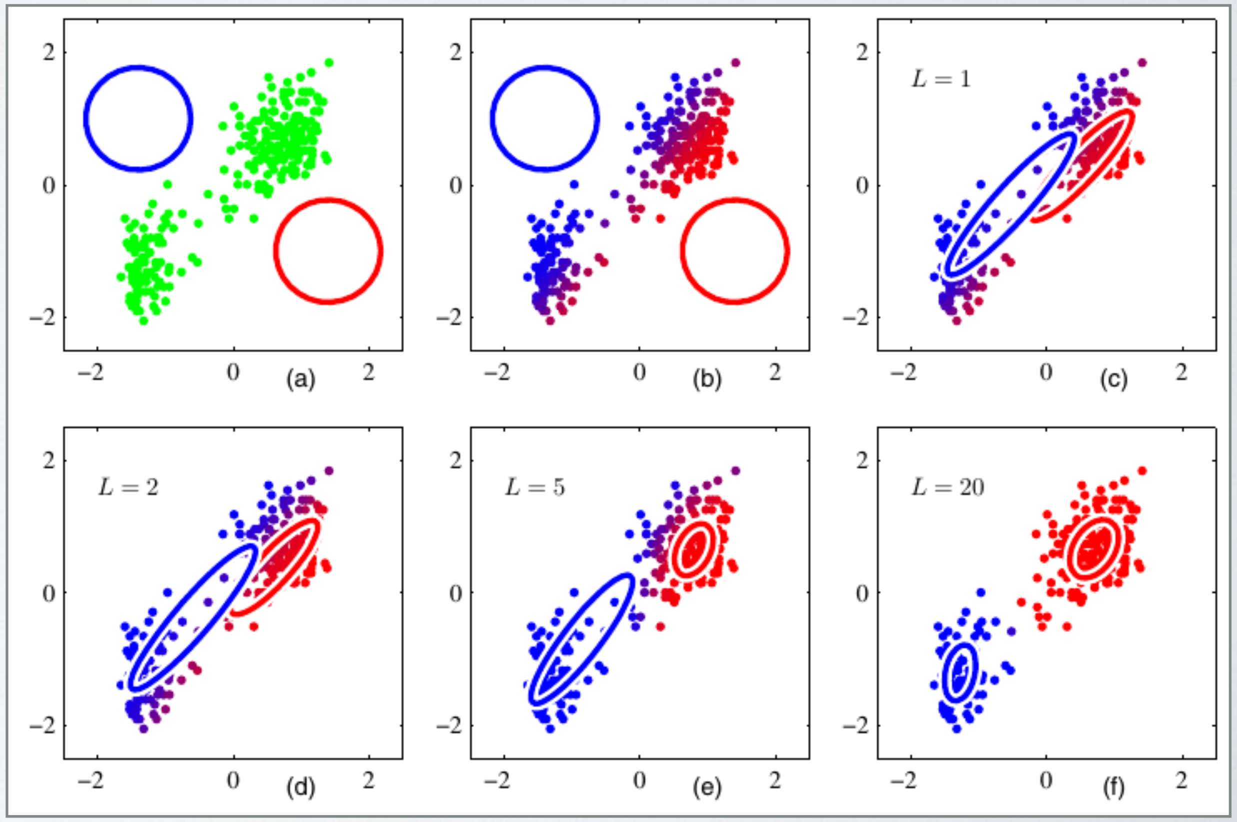




HIDDEN MARKOV MODEL

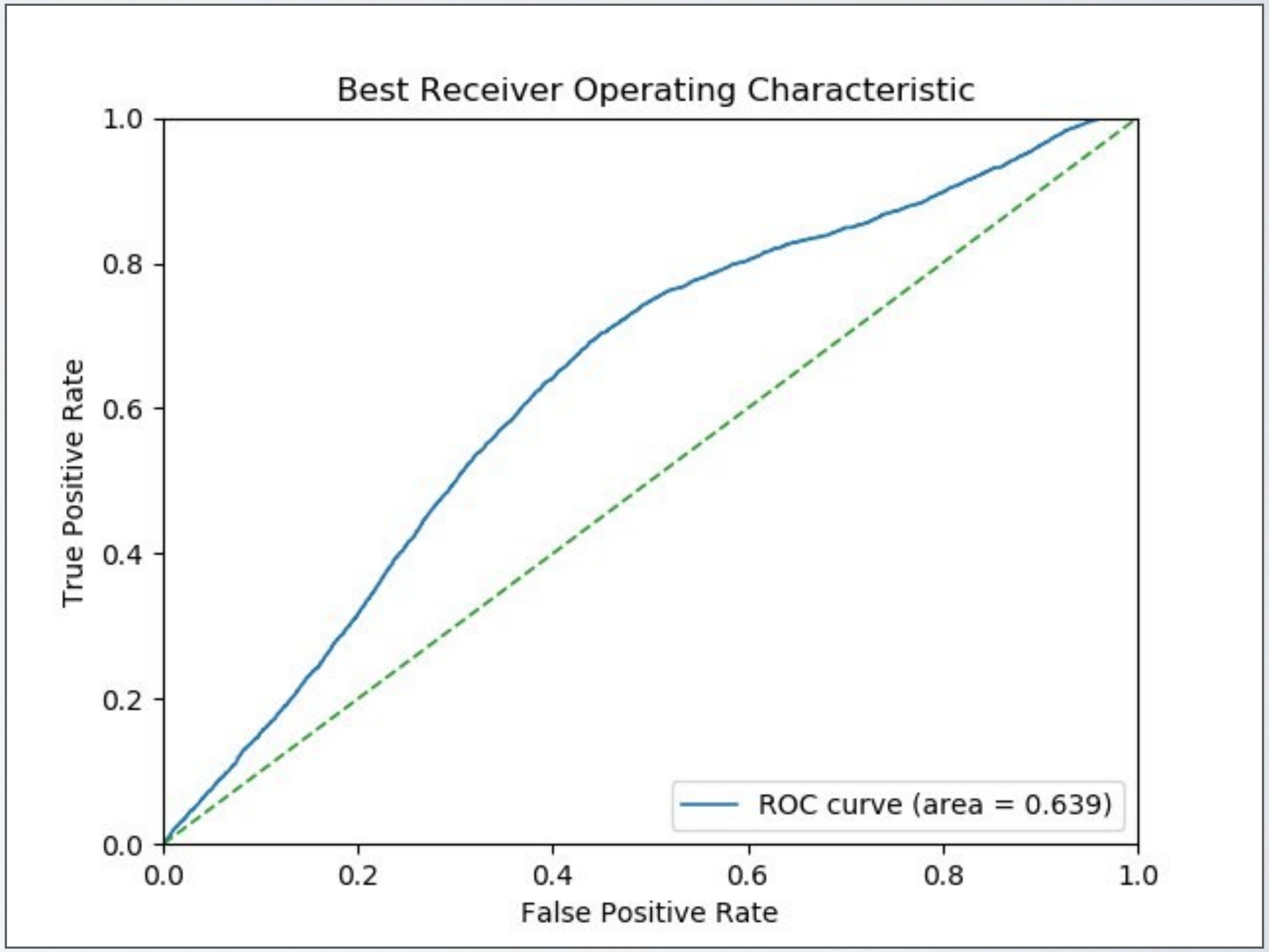
- Model to learn time sequences
- HMM + Density at each node
- GMM density or Normalizing Flows
- Trained with EM algorithm
- Dirichlet prior + GMM regularization + GMM pre-training





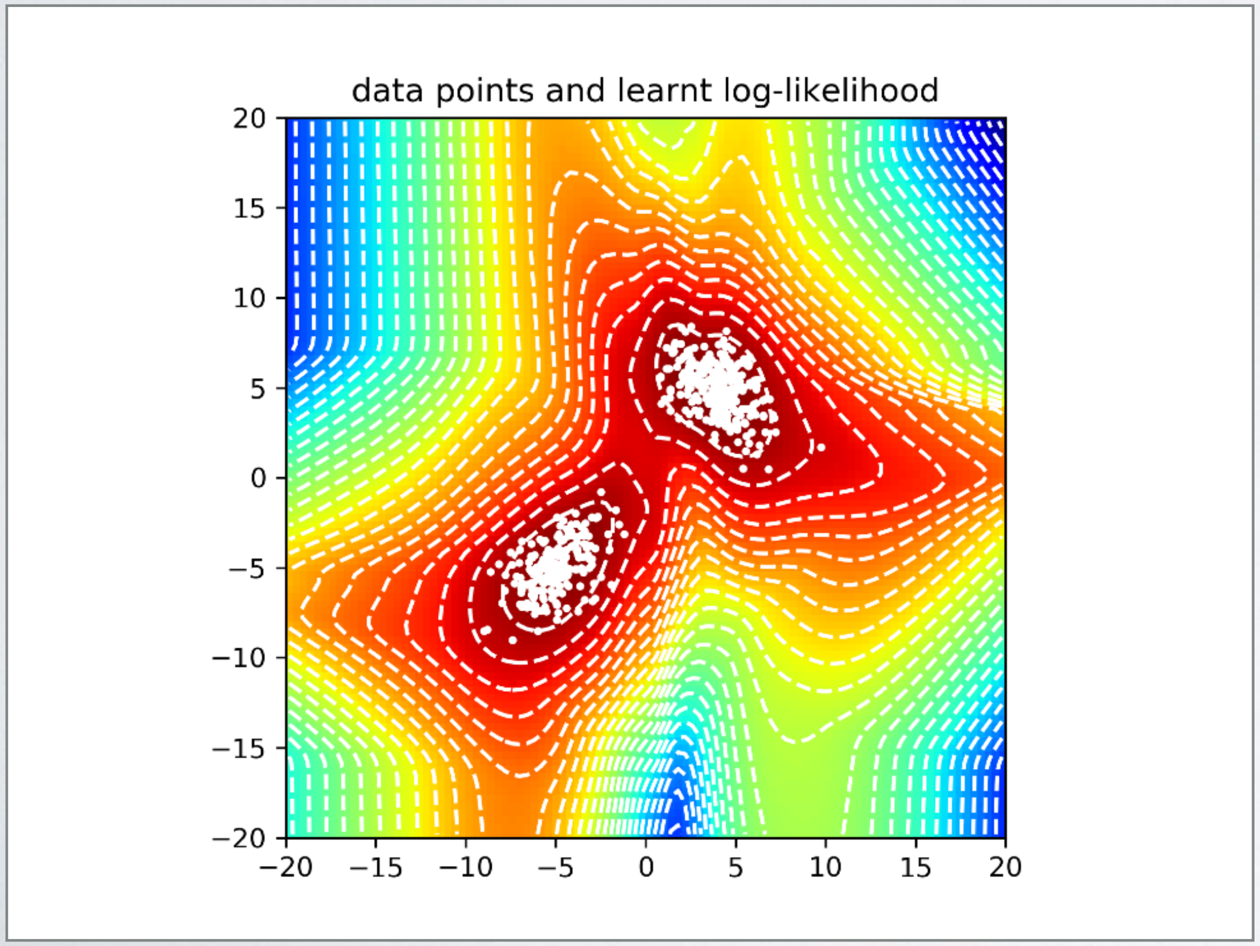
HMM RESULTS

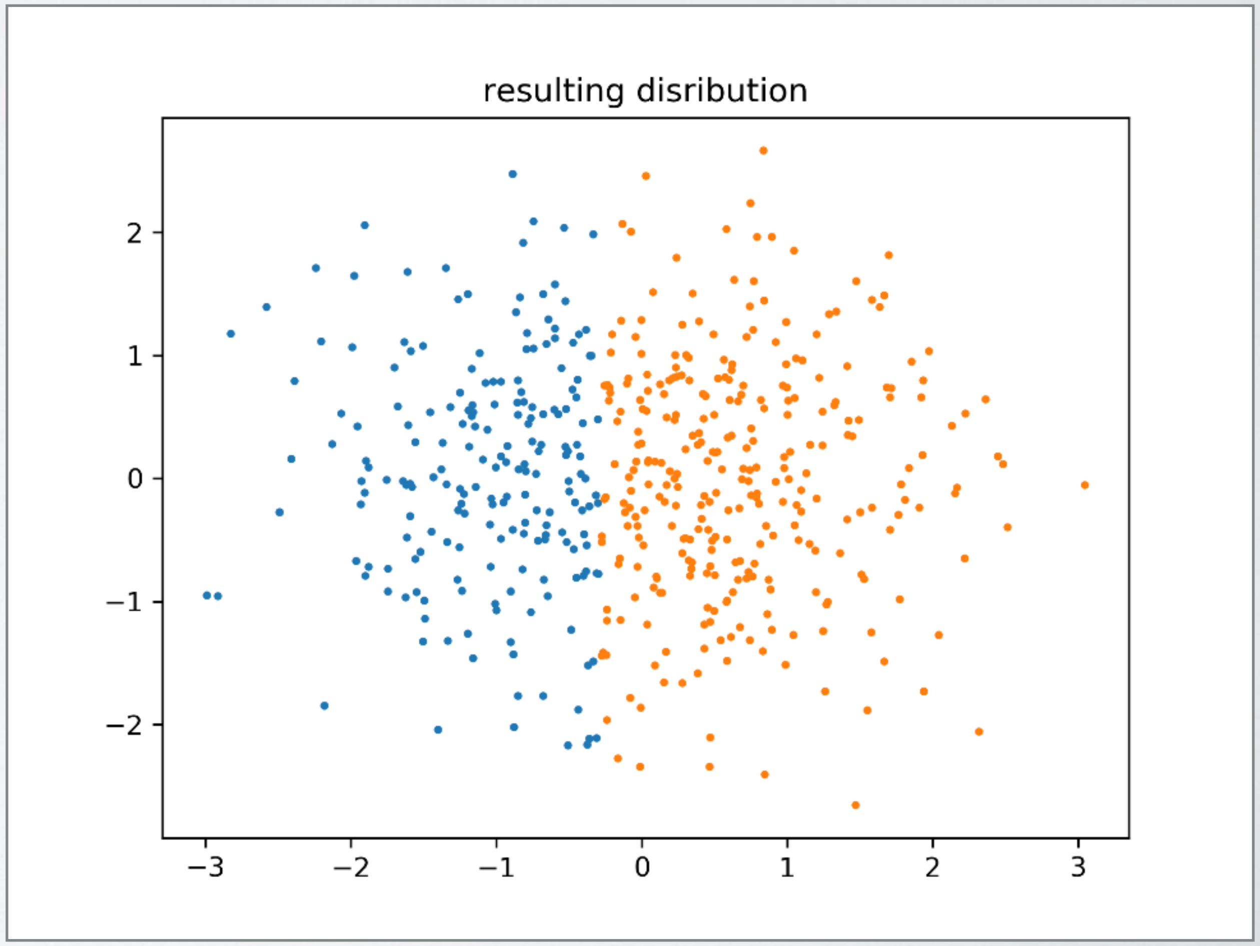
- Trained 2 HMMs: one for each label
- p(whale|data) = softmax of normalized HMM log_like
- Using HMMs is <u>Hard</u>



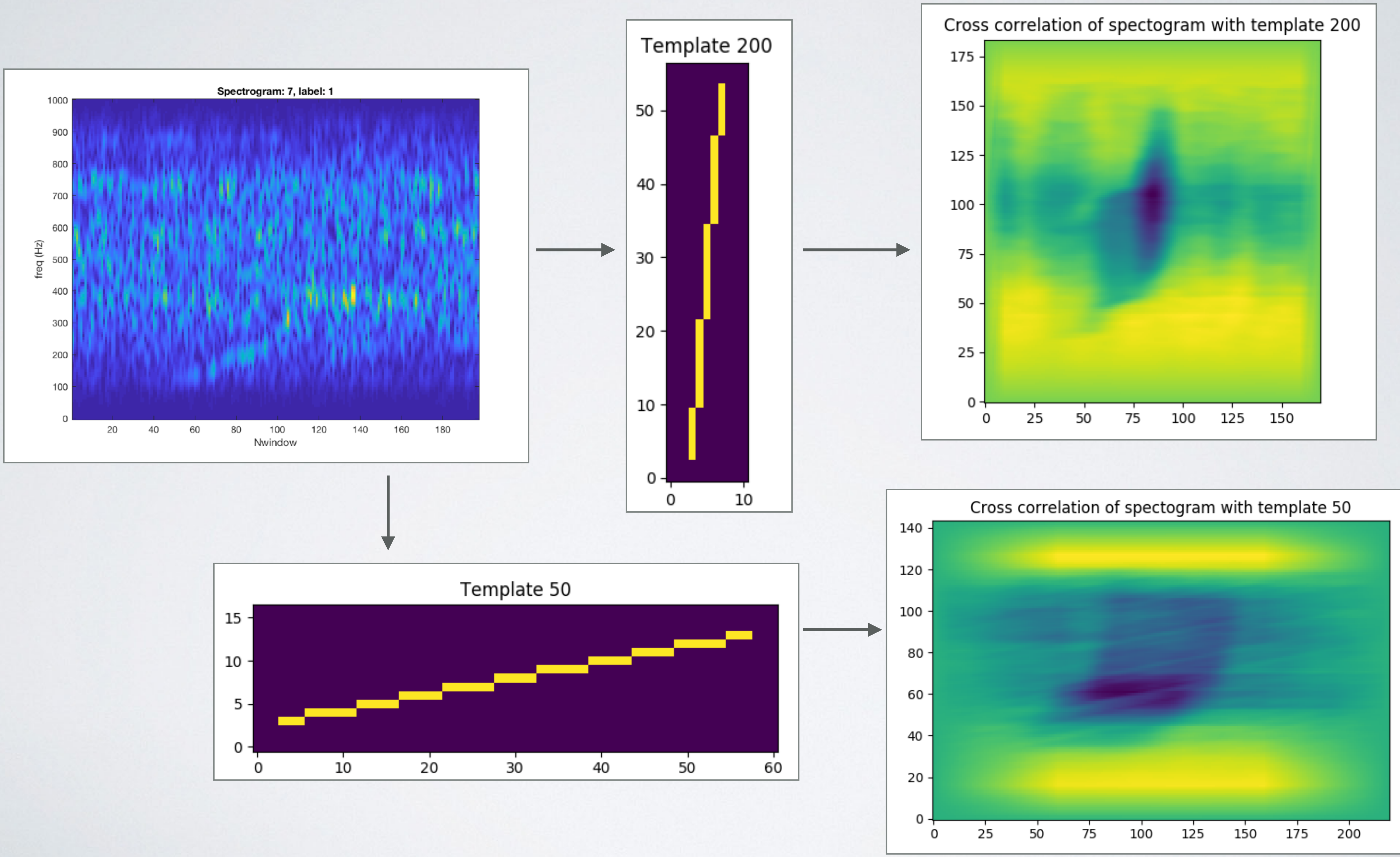
FUTURE WORK: NORMALIZING FLOWS

- Neural autoregressive flow
- Transform complex spaces to known densities
- Drop-in replacement for GMM in HMM





300 DIAGONAL LINE TEMPLATES

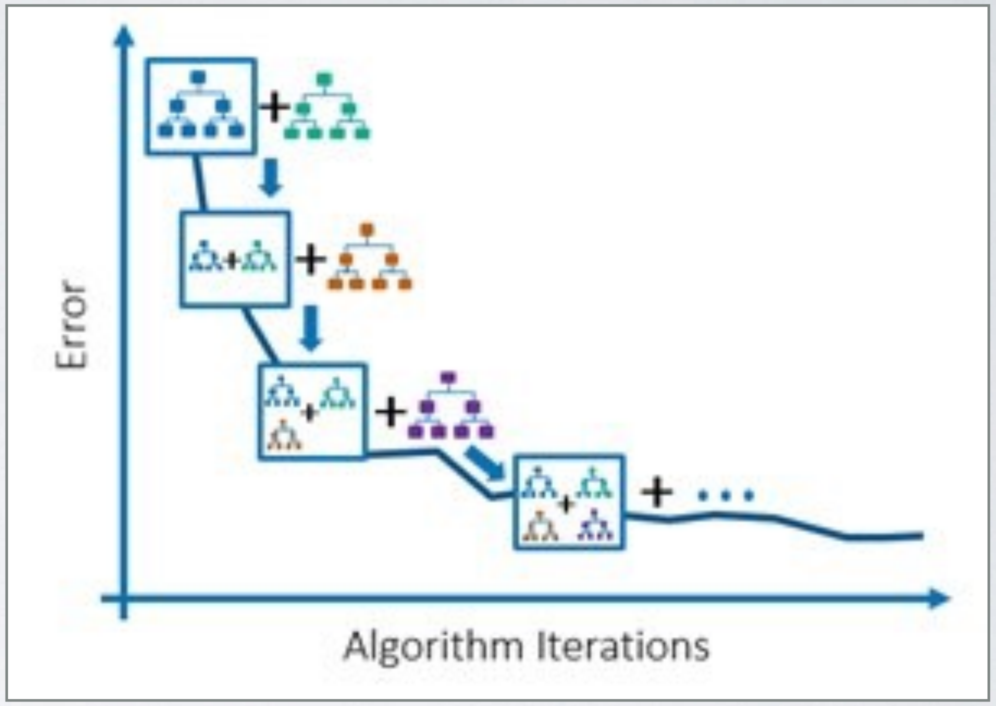


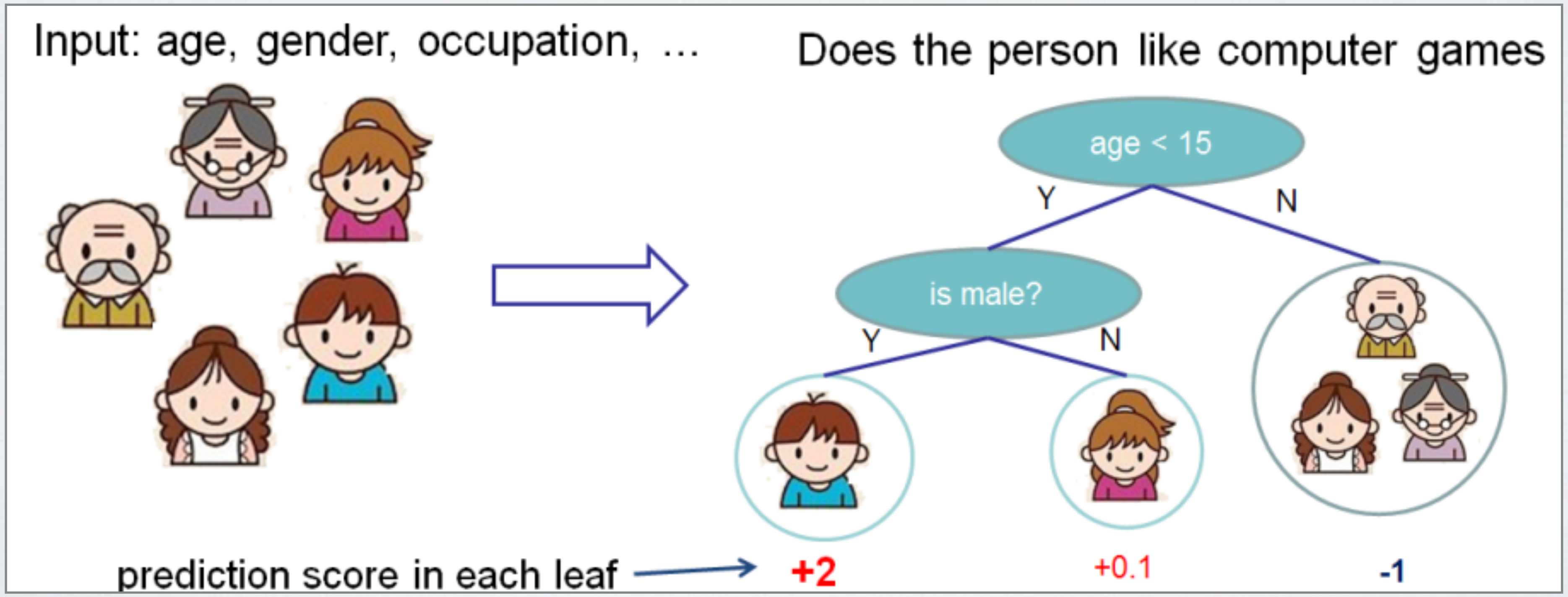
PARALLEL FEATURE EXTRACTION

- 30000 spectrograms x 300 templates = 9000000 2d correlations: Intractable on a single cpu
- Divided into 100 tasks and ran on ViVoLab cluster
- Features: max, std, mean + axis-wise: centroids, std, skewness and kurtosis on x, y axes: <u>3600 feats/spec</u>
- Future work: Input these features to CNN

GRADIENT BOOSTING TREES

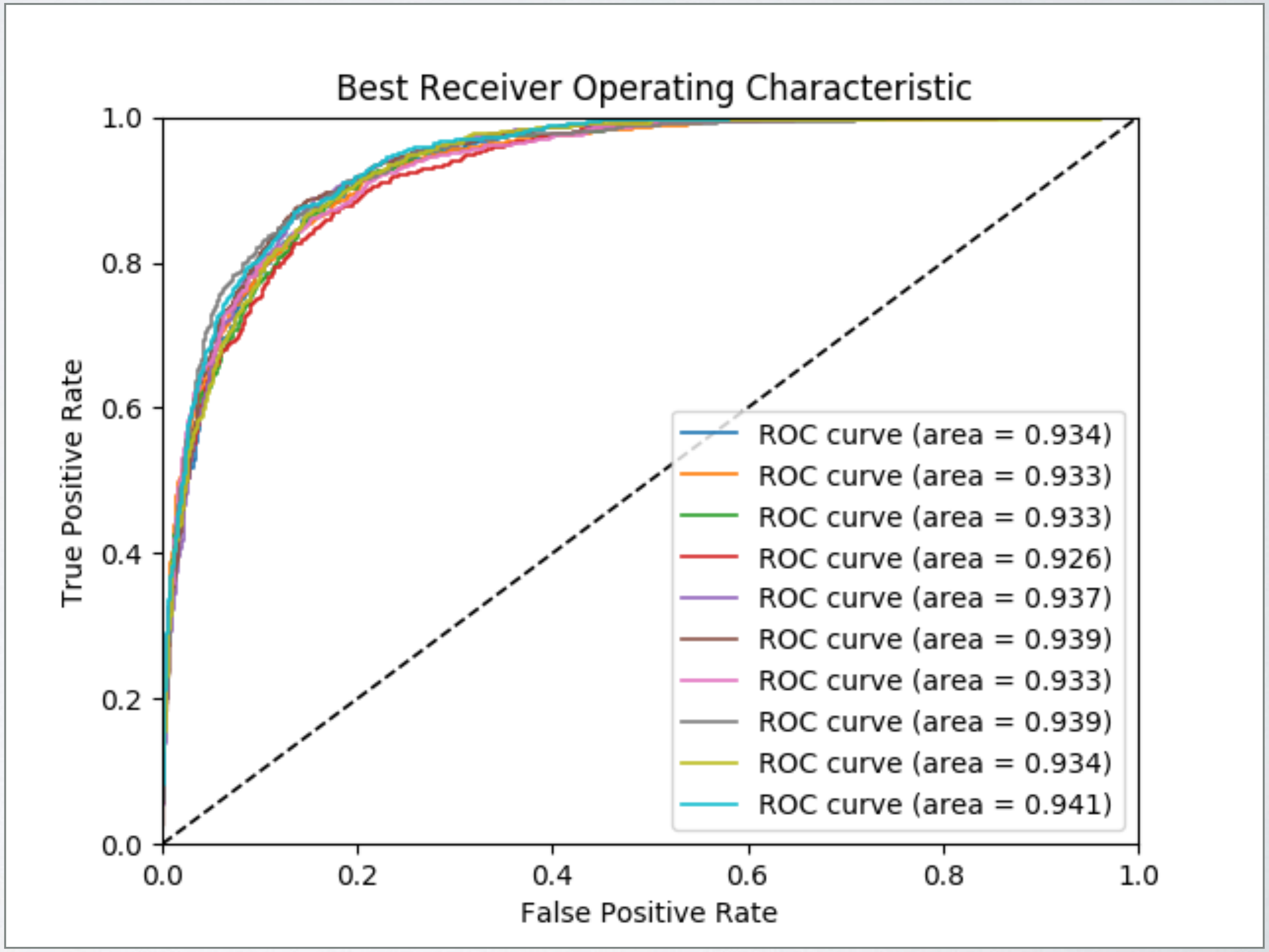
- New trees predict the residuals of other trees
- XGBoost: regularized boosting
- Only add tree if improvement larger than complexity cost





TEMPLATES + XGBOOST RESULTS

- Max tree depth: 3
- 60% data subsampling
- 90% feature subsampling
- Ratio adjusting
- 100 iterations



10-FOLD CROSS VALIDATED ROC-AUC MEAN+STD

ROC-AUC	CNN 25ms window deltas	CNN 250ms window deltas	HMM swt mfb multiresolution	Templates + XGboost
mean	0.9656	0.9786	0.6101	0.9347
std	0.0045	0.0014	0.0605	0.0040

Sources

+ Papers: → https://github.com/JavierAntoran/moby_dick