# Sparse Representation Classification of Dolphin Whistles Using Gabor Wavelets


Mahdi Esfahanian, Hanqi Zhuang, and Nurgun Erdol

Department of Computer and Electrical Engineering and Computer Science

Southeast National Marine Renewable Energy Center

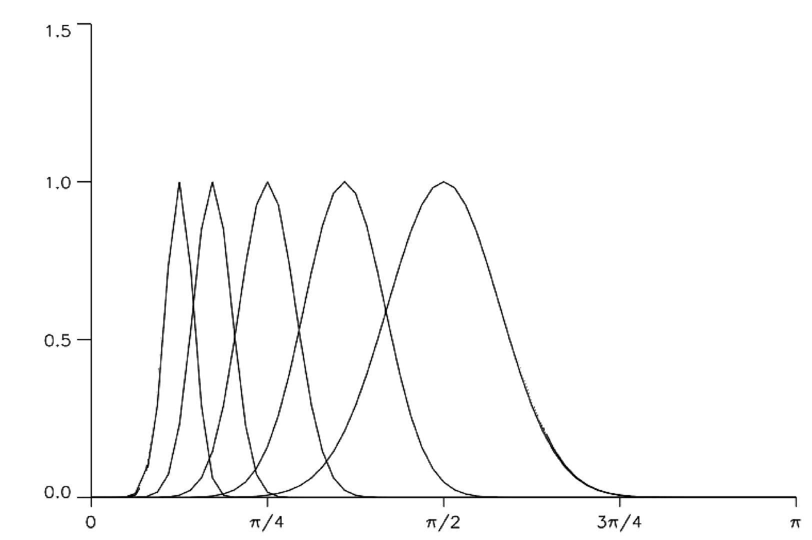

## MOTIVATION

- A major task in bioacoustic research is analyzing recorded repertoires of various species to either be associated with a behavior or determine animal identity.
- Bottlenose dolphins can produce whistles that are narrow-band long duration sounds from 1 kHz up to 24 kHz.
- Marine mammal vocalizations are usually corrupted by different acoustic sources making the pre-processing task necessary and costly.
- Most feature extraction techniques require tracing the contour of whistles in order to obtain distinguishing information for recognition purposes.

## METHODOLOGY

Pre-processing Contour tracing → Feature extraction → Classification

- Pre-processing stage may implement band-pass filtering with cut-off frequencies at 4 kHz and 16 kHz as well as spectral denoising.
- If needed to contour tracing, the highest frequency peak of isolated fundamental whistle is picked at each time instance.
- Feature extraction methods can be relied on the whistle contour or not.
- **Contour-based:**
  - Time-Frequency Parameters (TFPs)
  - Fourier Descriptors (FDs)



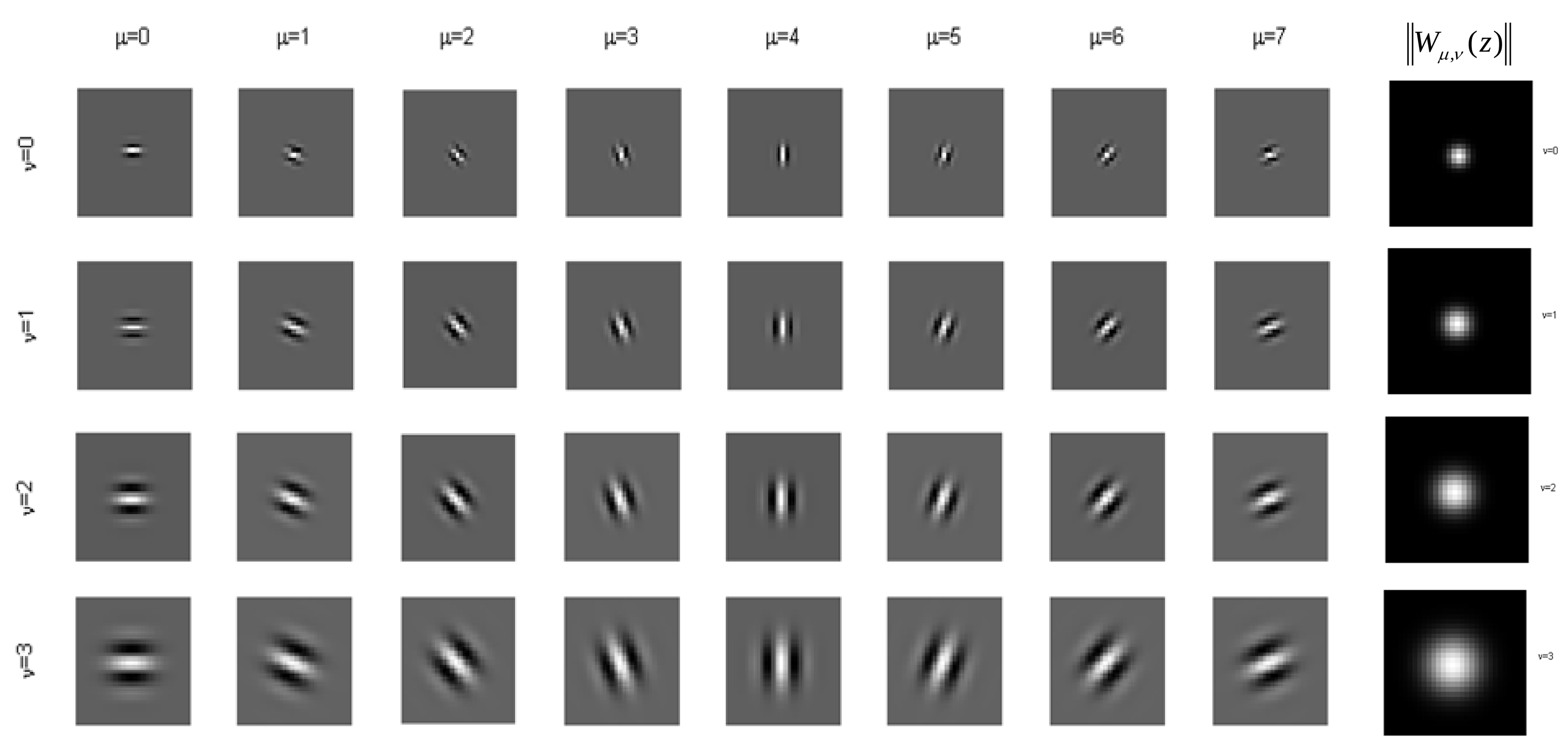

$$k_\nu = \frac{\pi/2}{\sqrt{2}^{\nu}} \quad for\ \nu = 0,1,2,3$$

✓ Fourier transform of all kernels where variations inside the critical frequency range between $\pi/16$ and $\pi/2$ are small.

The Gabor wavelet representation is the convolution of the spectrogram with a family of Gabor kernels:

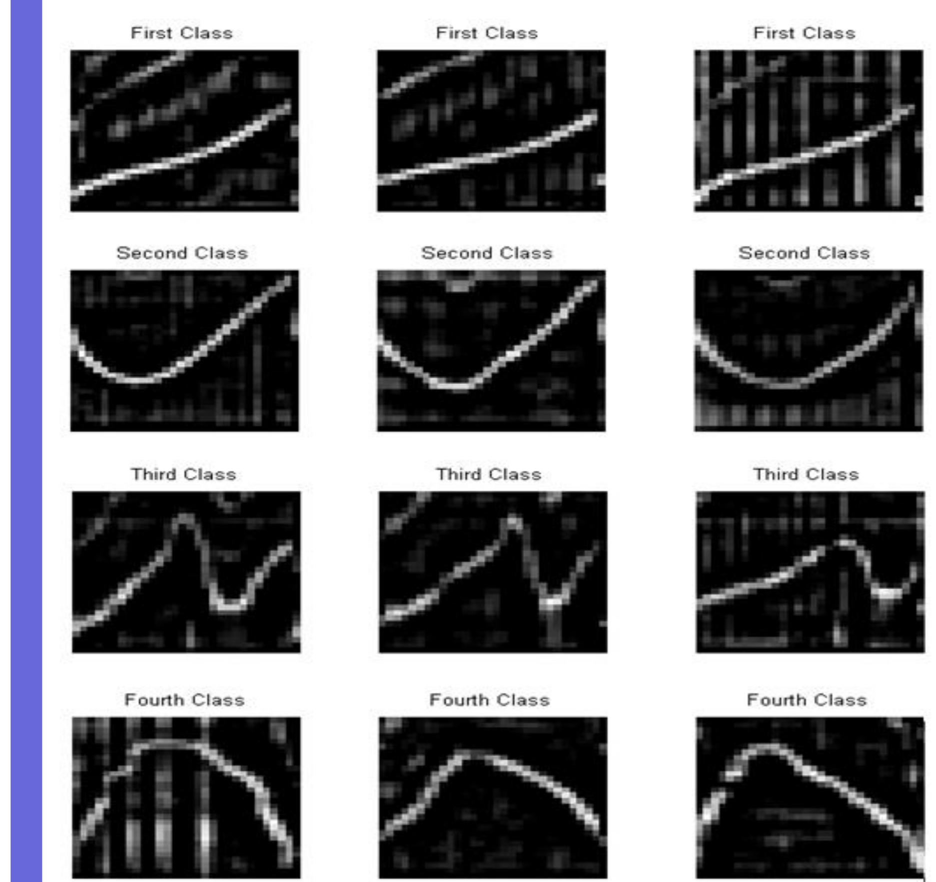
$$W_{\mu,\nu}(z) = I(z) * \psi_{\mu,\nu}(z)$$

$I(z)$ denotes the spectrogram image.



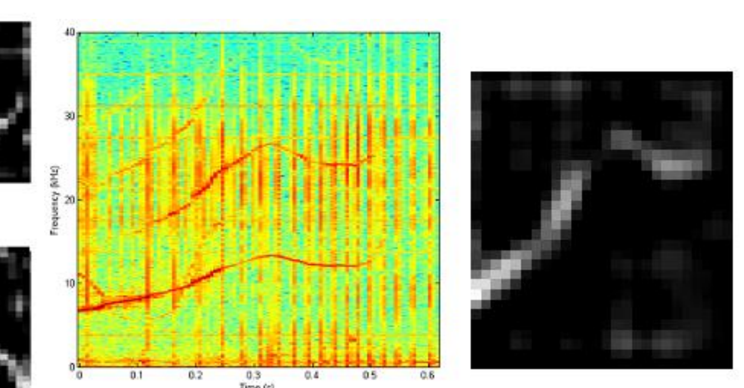

## RESULTS

**Examples of Gabor wavelet outputs:**




**Confusion matrix of SRC+Gabor (98%)**

| | First class (%) | Second class (%) | Third class (%) | Fourth class (%) |
|---|---|---|---|---|
| 1st class | 100 | 0 | 0 | 0 |
| 2nd class | 0 | 100 | 0 | 0 |
| 3rd class | 0 | 0 | 92 | 8 |
| 4th class | 0 | 0 | 0 | 100 |


(a) (b)

**Confusion matrix of SRC+TFPs (92%)**

| | First class (%) | Second class (%) | Third class (%) | Fourth class (%) |
|---|---|---|---|---|
| 1st class | 85 | 15 | 0 | 0 |
| 2nd class | 0 | 100 | 0 | 0 |
| 3rd class | 0 | 0 | 83 | 17 |
| 4th class | 0 | 0 | 0 | 100 |

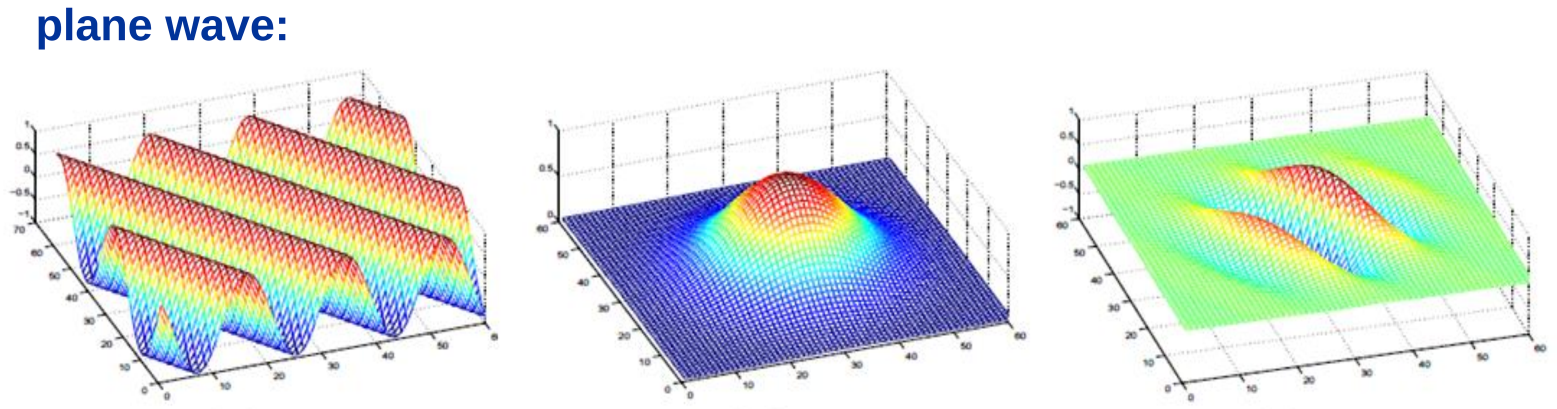
**Confusion matrix of SRC+FDs (88%)**

| | First class (%) | Second class (%) | Third class (%) | Fourth class (%) |
|---|---|---|---|---|
| 1st class | 100 | 0 | 0 | 0 |
| 2nd class | 0 | 94 | 0 | 6 |
| 3rd class | 33 | 0 | 58 | 9 |
| 4th class | 0 | 0 | 0 | 100 |

## FEATURE EXTRACTION

- **Non-contour-based:**
  - Gabor wavelets:

2-D Gabor filter is a Gaussian kernel function modulated by a sinusoid plane wave:


(a) (b) (c)

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2\|z\|^2}{2\sigma^2}} [e^{ik_{\mu,\nu}z} - e^{-\sigma^2/2}]$$

$$\phi_\mu = \pi\mu/N \qquad k_\nu = k_{\max}/f^{\nu} \qquad k_{\mu,\nu} = k_\nu e^{i\phi_\mu}$$

$\mu, \nu$ : the orientation and scale of Gabor filer, respectively.
$Z=(x,y)$ : the Cartesian coordinate.
$k_{\mu,\nu}$ : the wave vector.
$k_{\max}$ : the maximum frequency and $N$ is the number of orientations.
$f$ : spacing factor in between kernels in the frequency domain.
$\sigma$ : standard deviation of the Gaussian window .

**Parameter selection:**

$\mu \in \{0,1,2,3,4,5,6,7\}$ ; $\nu \in \{0,1,2,3\}$ ; $\sigma = \pi$ ; $f = 2$^0.5 ; $k_{max} = \pi/2$

## CLASSIFICATION METHOD

**Sparse Representation Classifier (SRC)**

The application of compressive sensing to classification has been recently emerged where sparse representation of the test sample is obtained using a training data set.

➢Consider a dictionary $U = [U_1, U_2, \ldots, U_d]$ where each class consists of $n_i$ training vectors $U_i = [u^1, u^2, \ldots, u^{ni}]$ for $i = 1,2,\ldots,d$. A test sample $y$ from $i$-th class can be represented as a linear combination of all training samples from $d$ distinct classes as:

$$y = \sum_{i=1}^{n_1}\beta_1^i u_1^i + \sum_{i=1}^{n_2}\beta_2^i u_2^i + \ldots + \sum_{i=1}^{n_d}\beta_d^i u_d^i \quad \text{or} \quad y = U\alpha_0$$

In this case, all the elements of weighting vector $\alpha_0$ should be zero but those associated with the same class:

$$[0,\ldots,0,\beta_i^1,\ldots,\beta_i^{n_i},0,\ldots,0]$$

➢Now the signal $y$ is deduced to be sparse in domain $U$ and the convex optimization problem can be solved:

$$\bar{\alpha} = \arg\min \|\alpha_0\|_1 \quad subject\ to \quad \|y - U\alpha_0\|_2 \le \varepsilon$$

➢Once the sparse solution is obtained, the classification procedure of SRC is summarized as follows:

1) For each class, find the residual between the reconstructed sample and the test sample:

$$r_i = \|y - U_i\bar{\alpha}_i\|_2 \quad for \quad i = 1,2,\ldots,d$$

2)The class of test sample is determined by:

$$class(y) = \arg\min_i (r_i)$$

## CONCLUSIONS

- A new algorithm that combines a compressive-sensing based technique called Sparse Representation Classifier and a feature vector set derived from Gabor wavelets was developed and applied to the classification of dolphin whistle types.
- The proposed approach avoids the need for tedious preprocessing and contour tracing steps while achieving a superior classification performance in comparison with two other contour-based feature extraction techniques.
- The accuracy obtained by combination of Gabor wavelet and SRC classifier is comparable with the accuracy achieved with LBP and SVM classifier reported in an earlier paper by authors.

## REFERENCES


[1] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, 2009.
[2] D. L. Donoho, "Compressed Sensing," IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289-1306, 2006.
[3] R. G. Baraniuk, "Compresseive sensing," IEEE Signal Processing Magazine, vol. 24, no. 4, pp. 118-121,2007.
[4] E. J. Candès, and M. B. Wakin, "An introduction to compressive sensing," IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 21-30, 2008.
[5] M. Esfahanian, H. Zhuang, and N. Erdol, "Using local binary patterns as features for classification of dolphin calls," Journal of Acoustical Society of America, vol. 134, no. 1, 2013.
[6] M. Esfahanian, H. Zhuang and N. Erdol, "On contour-based classification of dolphin whistles by type," Journal of Applied Acoustics, vol. 67, pp. 276-279, 2014.
[7] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," IEEE Transactions on Acoustics, Speech, and Signal Processing (ICASSP), vol. 36, no. 7, pp. 1169-1179, 1988.


## ACKNOWLEDGMENT


•This project was partially supported by SNMREC.
• The data used in this research was provided by Woods Hole Oceanographic Institution (WHOI).