Underwater Object Classification Using Scattering Transform of Sonar Signals

David Weber and Naoki Saito

Department of Mathematics
University of California, Davis

August 6, 2017

Signal Invariants

How to construct a feature extractor that is invariant to task-irrelevant deformations in the data? Some examples:

Translation	$T_c[f] = f(x-c)$
Modulation	$M_\omega[f] = \mathrm{e}^{\mathrm{i}\omega t} f(x)$
Scaling	$\mathscr{S}_a[f] = f\left(x/a\right)$
Amplitude	$A_a[f] = a f(x)$

Signal Invariants

How to construct a feature extractor that is invariant to task-irrelevant deformations in the data? Some examples:

Translation	$T_c[f] = f(x-c)$
Modulation	$M_\omega[f] = e^{i\omega t} f(x)$
Scaling	$\mathscr{S}_a[f] = f\left(x/a\right)$
Amplitude	$A_a[f] = af(x)$

[Amari, 1968] and [Otsu, 1973] demonstrated that only trivial linear features are absolutely invariant to even just translation, so they use *relative invariance* of feature extractor ρ:

$$\rho[T_c f] = \eta(c)\rho[f]$$

Signal Invariants

How to construct a feature extractor that is invariant to task-irrelevant deformations in the data? Some examples:

Translation	$T_c[f] = f(x-c)$
Modulation	$M_\omega[f] = \mathrm{e}^{\mathrm{i}\omega t} f(x)$
Scaling	$\mathscr{S}_a[f] = f\left(x/a\right)$
Amplitude	$A_a[f] = a f(x)$

[Amari, 1968] and [Otsu, 1973] demonstrated that only trivial linear features are absolutely invariant to even just translation, so they use *relative invariance* of feature extractor ρ:

$$\rho[T_c f] = \eta(c)\rho[f]$$

They establish that the only linear feature extractors $\rho[f] = \langle f, \rho \rangle$ that are relatively invariant w.r.t. both amplitude and translation deformations are Fourier-Laplace type, i.e. for some $z \in \mathbb{C}^d$

$$\int_{\mathbb{R}^d} f(x) c_1 \mathrm{e}^{z \cdot x}\, \mathrm{d}x$$

Generalized Scattering Transform

A single propagating layer $u[q_i]f$ of the generalized scattering transform is a vector consisting of semi-discrete shift invariant frame transforms $\psi_{\lambda_i^{(m)}} \star f$ indexed by $\lambda_i^{(m)} \in \Lambda_m$, a pointwise nonlinearity σ_m with Lipschitz constant γ_m, and a subsampling factor $r_m \geq 1$.

Generalized Scattering Transform

A single propagating layer $u[q_i]f$ of the generalized scattering transform is a vector consisting of semi-discrete shift invariant frame transforms $\psi_{\lambda_i^{(m)}} \star f$ indexed by $\lambda_i^{(m)} \in \Lambda_m$, a pointwise nonlinearity σ_m with Lipschitz constant γ_m, and a subsampling factor $r_m \geq 1$.Putting together k indices in a path $q = (\lambda_{i_m}^{(m)}, \ldots, \lambda_{i_1}^{(1)})$ gives the propagating value at layer m:

$$u[\lambda_i]f := \frac{1}{r_1^{d/2}} \sigma\big(\psi_{\lambda_i} \star f\big)(r_1 \cdot) \qquad m = 1$$

$$u[\lambda_i^{(2)}, \lambda_j^{(1)}]f := \frac{1}{r_1^{d/2}} \sigma\Big(\psi_{\lambda_i^{(2)}} \star \frac{1}{r_1^{d/2}} \sigma\big(\psi_{\lambda_j^{(1)}} \star f\big)(r_1 \cdot)\Big)(r_2 \cdot) \quad m = 2$$

And so on, until the desired depth.

Generalized Scattering Transform

A single propagating layer $u[q_i]f$ of the generalized scattering transform is a vector consisting of semi-discrete shift invariant frame transforms $\psi_{\lambda_i^{(m)}} \star f$ indexed by $\lambda_i^{(m)} \in \Lambda_m$, a pointwise nonlinearity σ_m with Lipschitz constant γ_m, and a subsampling factor $r_m \geq 1$.Putting together k indices in a path $q = (\lambda_{i_m}^{(m)}, \ldots, \lambda_{i_1}^{(1)})$ gives the propagating value at layer m:

$$u[\lambda_i]f := \frac{1}{r_1^{d/2}} \sigma\big(\psi_{\lambda_i} \star f\big)(r_1 \cdot) \qquad m = 1$$

$$u[\lambda_i^{(2)}, \lambda_j^{(1)}]f := \frac{1}{r_1^{d/2}} \sigma\Big(\psi_{\lambda_i^{(2)}} \star \frac{1}{r_1^{d/2}} \sigma\big(\psi_{\lambda_j^{(1)}} \star f\big)(r_1 \cdot)\Big)(r_2 \cdot) \quad m = 2$$

And so on, until the desired depth. The output $s^m[f]$ is taken by averaging $u[f]$ for every path q of depth m with one atom ϕ_m, and then subsampling:

$$s_m[\lambda_{i_m}^{(m)}, \ldots, \lambda_{i_1}^{(1)}]f := \phi_m \star u[\lambda_{i_m}^{(m)}, \ldots \lambda_{i_1}^{(1)}]f$$

Generalized Scattering Transform

A single propagating layer $u[q_i]f$ of the generalized scattering transform is a vector consisting of semi-discrete shift invariant frame transforms $\psi_{\lambda_i^{(m)}} \star f$ indexed by $\lambda_i^{(m)} \in \Lambda_m$, a pointwise nonlinearity σ_m with Lipschitz constant γ_m, and a subsampling factor $r_m \geq 1$.Putting together k indices in a path $q = (\lambda_{i_m}^{(m)}, \ldots, \lambda_{i_1}^{(1)})$ gives the propagating value at layer m:

$$u[\lambda_i]f := \frac{1}{r_1^{d/2}} \sigma\big(\psi_{\lambda_i} \star f\big)(r_1 \cdot) \qquad m = 1$$

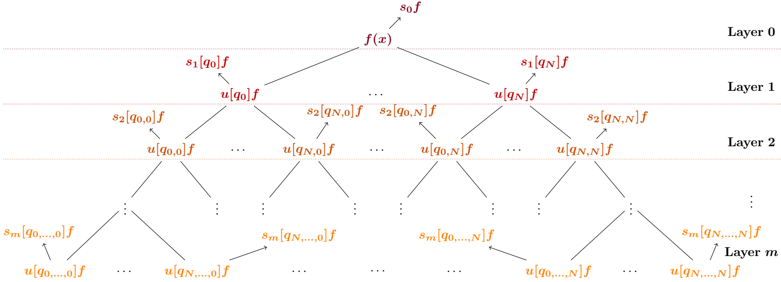
$$u[\lambda_i^{(2)}, \lambda_j^{(1)}]f := \frac{1}{r_1^{d/2}} \sigma\Big(\psi_{\lambda_i^{(2)}} \star \frac{1}{r_1^{d/2}} \sigma\big(\psi_{\lambda_j^{(1)}} \star f\big)(r_1 \cdot)\Big)(r_2 \cdot) \quad m = 2$$

And so on, until the desired depth. The output $s^m[f]$ is taken by averaging $u[f]$ for every path q of depth m with one atom ϕ_m, and then subsampling:

$$s_m[\lambda_{i_m}^{(m)}, \ldots, \lambda_{i_1}^{(1)}]f := \phi_m \star u[\lambda_{i_m}^{(m)}, \ldots \lambda_{i_1}^{(1)}]f$$

If we write s without an index, this is the collection of outputs at all layers up to some desired M: $0, 1, \ldots, M$.

Generalized Scattering Transform



The original scattering transform specifies that $\sigma_m = |\cdot|$, indexes by $\Lambda_m = \{a^{j/Q_m} h\}_{j>-J_m, h\in H_m}$ for some rotation h in the discrete rotation group H_m, subsamples only the output, and has strong conditions on the parent wavelets ψ and ϕ. Q_m is the quality factor, which can vary by layer.

The first results on Scattering transforms were from [Mallat, 2012] and his group. A more recent generalization for "weakly admissible" frames, and not just wavelets, that increasing the depth m increases translation invariance:

Theorem (Depth translation invariance, [Wiatowski and Bölcskei, 2015])

As long as the frames have upper frame bounds b_m satisfying $\max\{b_m, \gamma_m b_m / r_m^d\} \leq 1$, the features at depth m satisfy:

$$S^m[T_c f] = T_{\frac{c}{r_1 \cdots r_{m-1}}} S^m[f]$$

Further if the output atoms satisfy $\widehat{\phi_m}|\omega| \leq K$, this implies a bound on the difference in norm:

$$\|S^m[f] - S^m[T_c f]\| \leq \frac{2\pi|c|K}{r_1 \cdot r_m} \|f\|_2$$

In addition to this quasi-translation invariance, this generalized scattering transform is stable under space and frequency modulations:

$$F_{\tau,\omega}[f](x) = \mathrm{e}^{\mathrm{i}\omega(x)} f\big(x - \tau(x)\big)$$

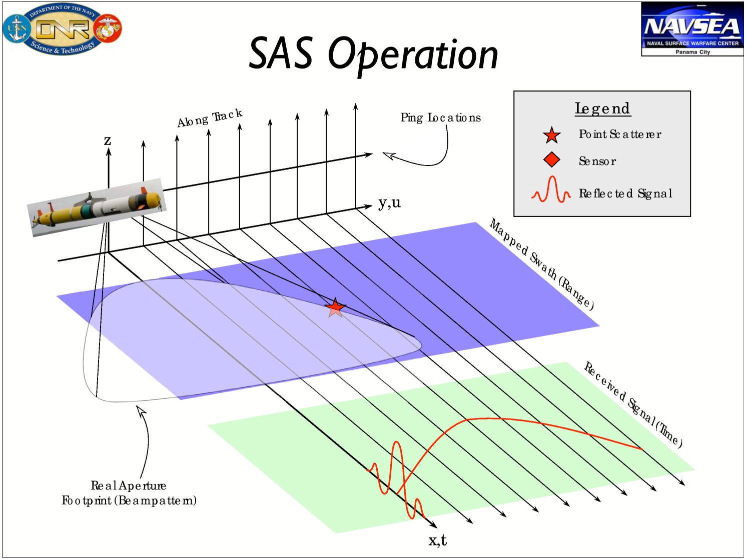
Theorem (Stability, [Wiatowski and Bölcskei, 2015])

If f is a band limited function, ω and τ are continuous, τ is once differentiable and $\|\nabla\tau\|_\infty \le \frac{1}{2d}$, there is a C independent of S so that

$$\Big\| S[f] - S\big[F_{\tau,\omega}[f]\big] \Big\|_2 \le C\|f\|_2\big(R\|\tau\|_\infty + \|\omega\|_\infty\big)$$

where the norm on S is just $\|\cdot\|_2$ on each output element

Sonar Scattering



Sonar Scattering

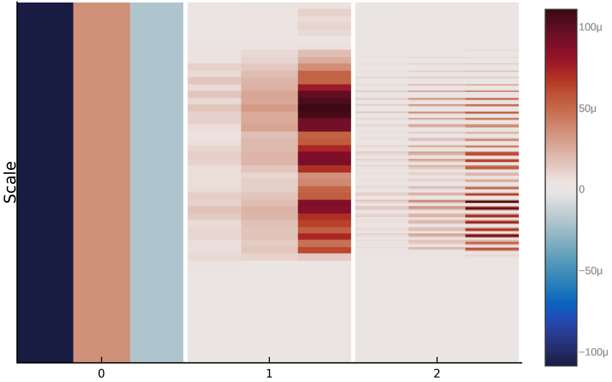


Figure: 1D scattering transform of the shell with $m \leq 2$, $Q_1 = 8, Q_2 = 1$, and the averaging function ϕ having width 80% of the total width. Q_m here is the number of wavelets per octave in layer m, i.e. the scale factor is $a = 2^{j/Q}$

1D Classifier

Consider each 2D wavefield as a set of 1D signals, perform a scattering transform using Morlet Wavelets, and use the resulting vectors as the input to a linear classifier, in our case Sparse logistic regression.

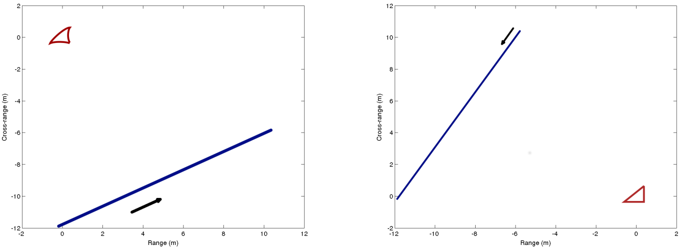


Figure: The triangle, the sharkfin and observation paths

Implementation Details

- As m increases, the remaining energy is concentrated at coarser scales, so only those S_J^m with increasing scales at deeper layers are kept for computational reasons (e.g. $|\phi \star |\psi_4 \star |\psi_1 \star f|||$ is kept, while $|\phi \star |\psi_1 \star |\psi_4 \star f|||$ is not).
- Discrimination is also concentrated towards coarser scales empirically

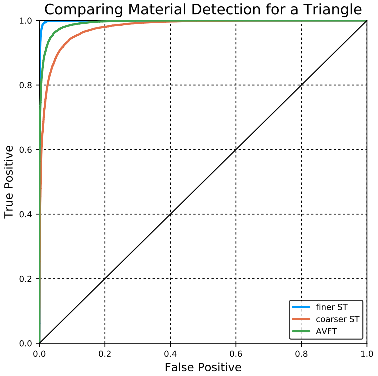


Figure: The ROC curve for detecting the material difference in a triangle, for speeds of sound $c_1 = 2000\text{m/s}$ and $c_1 = 2500\text{m/s}$.

Synthetic Experiments
Shape Discrimination

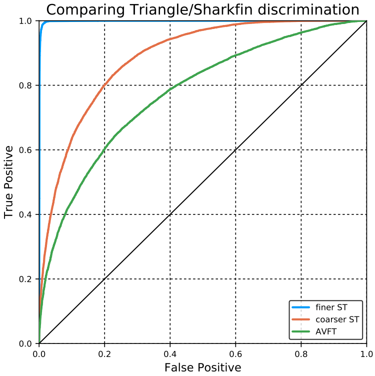


Figure: The ROC curve for discriminating a shark-fin from a triangle where both have a speed of sound fixed at $2000 \mathrm{m/s}$.

Real Experiments
UXO Detection

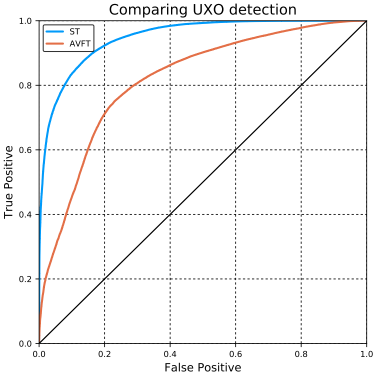
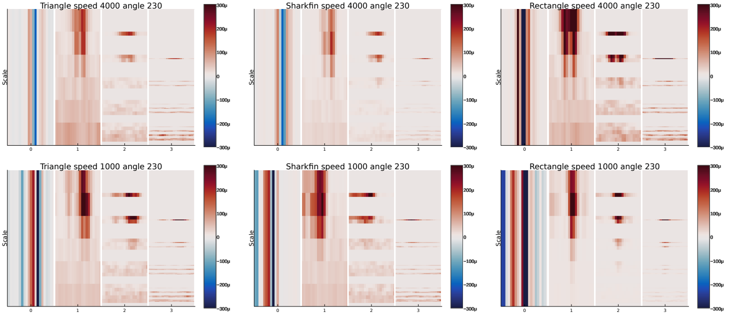


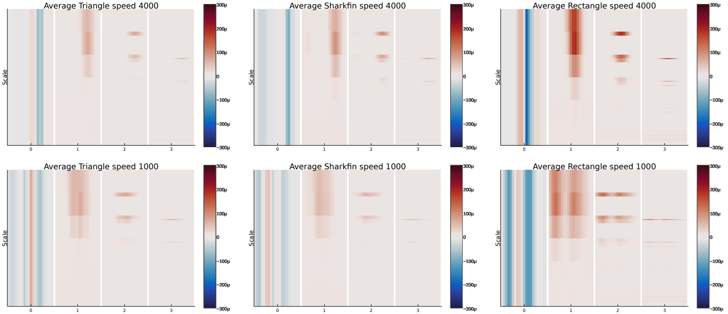
Figure: The ROC curve for detecting UXOs. The scattering transform has two layers, with quality factors $Q_1 = Q_2 = 8$

Synthetic Experiments
Specific examples



The results from a depth 3 scattering transform with ($Q_1 \equiv 8$, $Q_2 = 8$, $Q_3 = 1$, $0 \leq m \leq 3$) on various materials and shapes, for a fixed angle ($230°$) and track position ($n = 200$ out of 481)

Synthetic Experiments
Averages



The same scattering transform, but each plot is now averaged over rotation and translation

Shape Discrimination Coefficients

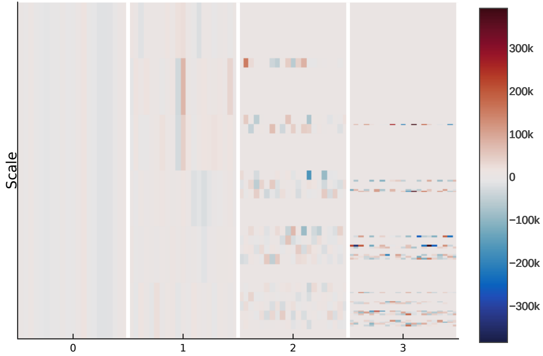
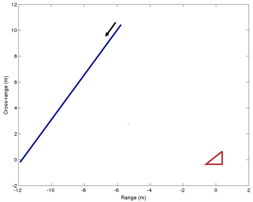
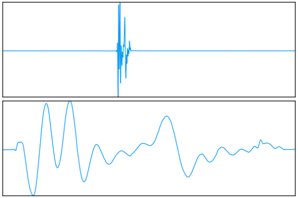


Figure: The GLMNET coefficients selected in one run. Red coefficients correspond to the triangle class, while blue to the sharkfin

Object Domain vs Signal Domain



(a) object domain



(b) signal domain

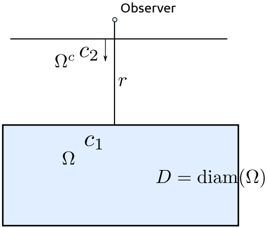
The invariants discussed in the first part of the talk are in the signal domain $f(t-c)$. What happens when we move or deform the triangle?

- Translation perpendicular to the rail
- Translation along rail
- Rotation
- Shape deformation
- Material deformation

Snell's law: Given a wave perpendicular to a boundary going from Ω^c to Ω, in the case of a rectangular region of width D, we have

Reflected magnitude	$V_{2,1} = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1}$
Transmitted magnitude	$W_{2,1} = \frac{2\rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1}$
peak after n internal reflections	$A_n = W_{2,1} V_{1,2}^n W_{1,2} \frac{A_0}{r^2 D^n}$

As important, the distance between peaks is $c_1 D$, giving signals that are pseudo-periodic with decreasing magnitude. The frequency of the end result corresponds directly to the speed of sound in the material



The point: speed shows up clearly in the Fourier domain, so even the AVFT will do quite well.

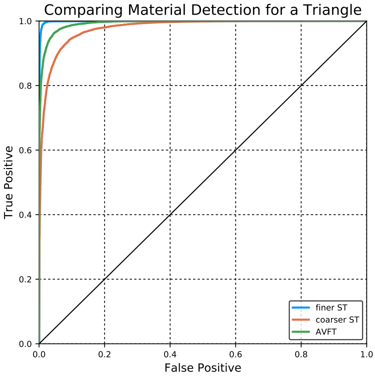
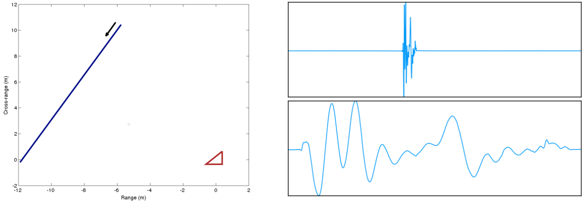


Figure: The ROC curve for detecting the material difference in a triangle, for speeds of sound $c_1 = 2000\mathrm{m/s}$ and $c_1 = 2500\mathrm{m/s}$.

Object Domain vs Signal Domain



(a) object domain

(b) signal domain

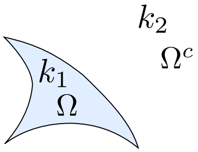
The invariants discussed in the first part of the talk are in the signal domain $f(t-c)$. What happens when we move or deform the triangle?

- Translation perpendicular to the rail
- Translation along rail
- Rotation
- Shape deformation
- Material deformation

Helmholtz equation

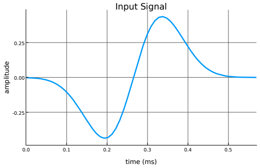
For a pure sinusoid of frequency ω, the amplitude of the returned signal is given by the Helmholtz equation:

$$\begin{aligned} \Delta u_\omega + k_1^2 u_\omega &= 0 \quad \text{in } \Omega \\ \Delta v_\omega + k_2^2 v_\omega &= 0 \quad \text{in } \Omega^c \\ u_\omega - v_\omega &= g \quad \text{on } \partial\Omega \\ \partial_\nu u_\omega - \partial_\nu v_\omega &= \partial_\nu g \quad \text{on } \partial\Omega \\ \sqrt{|x|}\left(\partial_{|x|} - \mathrm{i}k_2\right) v_\omega(x) \to 0 & \text{ as } |x| \to \infty \end{aligned}$$



Object Domain
Helmholtz equation solution

Real signals aren't pure sinusoids, but instead a compactly supported "chirp" $s(t)$



The time varying solution at a point $\vec{x}$ is given as an integral across frequencies by

$$f(t,\vec{x}) = \int_{-\infty}^{\infty} \widehat{s}(\omega)\, u_\omega(\vec{x}) e^{-i\omega t}\, d\omega$$

Far field gives an approximate form in terms of a radial solution $\Theta(\omega,\theta)$, that depends on the geometry, and a radial function $R(\omega,r) = J_0(k_2 r)$, which is always a Bessel function of the first kind (saw yesterday)!

Object Domain
Rotation

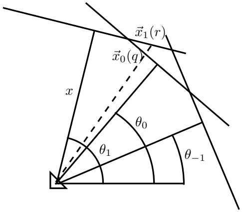


Figure: Synthesizing the rail x_0 from its neighbors x_1 and x_{-1}

Use the far field approximation to obtain a first order estimate of the dependence on rotation

$$f(r,\vec{x}_0(q)) \approx \int_{-\infty}^{\infty} \widehat{s}(\omega) H(q,\omega) K\big(\omega, \vec{x}_1\big(T(q)\big)\big) \mathrm{e}^{-\mathrm{i}\omega(t)} \, \mathrm{d}\omega$$

$H(q,\omega)$ is a ratio of Bessel functions, while $T(q)$ gives the mapping from q to r. We give conditions for when H can be written as a translation diffeomorphism.

Acknowledgements

- ONR Grants: N00014-12-1-0177; N00014-16-1-2255
- Vincent Bodin (former intern with Saito, responsible for initial database code)
- Jim Bremer and Ian Sammis (UC Davis, wrote the fast Helmholtz solver)
- Bradley Marchand
- Frank Crosby, Julia Gazagnaire (NSWC-PCD, FL, real dataset)
- GLMNET (T. Hastie and his group)
  - Simon Kornblith for his Julia wrapper for the fortran GLMNET
- SCATNET (S. Mallat and his group)

References I

Amari, S. (1968).
Invariant structures of signal and feature space in pattern recognition problems.
RAAG Memoirs, 4(1-2):553–566.

Mallat, S. (2012).
Group invariant scattering.
Comm. Pure Appl. Math., 65(10):1331–1398.

Naoki Saito, D. W. (2017).
Underwater object classification using scattering transform of sonar signals.

Otsu, O. (1973).
An invariant theory of linear functionals as linear feature extractors.
Bulletin of the Electrotechnical Laboratory, 37(10):893–913.

References II

Wiatowski, T. and Bölcskei, H. (2015).
Deep convolutional neural networks based on semi-discrete frames.
IEEE Int. Symp. on Info. Theory, pages 1212–1216.

Synthetic Experiments
Setup Description

We use Mallat's framework with Morlet Wavelets, and compare with the absolute value of the Fourier transform (AVFT). We use two scattering transforms:

Type	Q_1	Q_2	Q_3
Finer	8	8	1
Coarser	8	4	4

- Each signal is normalized so the maximum amplitude is 1
- White Gaussian noise is added to get average SNR is about 5dB.
- Multiclass logistic regression with Lasso (via GLMNET) is used as a feature extractor and a classifier.
- Perform 10-fold cross validation, i.e., repeat the classification 10 times by randomly splitting the whole dataset into training and test sets with 50/50.

[Naoki Saito, 2017]