Define a representation which explicitly reveal information for classification. How?

A possible answer: create sparsity with a dictionary.
- Minimise the representation size with a small error

Sparse decomposition of $f$ in a dictionary $\mathcal{D} = \{\phi_{\lambda}\}_{\lambda \in \Lambda}$:

$$f = \sum_{m=1}^{M} a_{\lambda_n} \phi_{\lambda_n} + \epsilon.$$  

Good for compression, denoising, source separation.

What about recognition and classification?
- Problem: needs to reduce variability.
• Images and sounds within a class have a considerable variability.
• A Euclidean norm does measure signal «similarities».
• Need to find a representation $\Phi$ (kernel) which:
  - Reduces intra-class variability (**invariants**)
  - Maintains discriminability (**informative**)

• Role of invariants in physics.

Linear classifiers: SVM, PCA ...
Audio Psychophysics

Reduction of processing rate

Cochlea: dilated wavelet filters
Filtering Cascade Models

Torsten Dau & Shihab Shamma

The following stage in the model, as shown in Fig. 1, is the modulation filterbank. The modulation filterbank was presented in Fassel and Puel. This stage will be called the modulation filterbank and finally added to internal noise; this processing transforms the signals into modulation-detection data with an optimal detector as decision device. The decision device is realized as an optimal detector in the same way as described in Dau et al.

Within the present model, it is postulated that the modulation-detection process is a sequential decision. The output of the 'preprocessing' stages can now be described for masking conditions using sinusoidal test signals. Limitations of resolution are again simulated by adding internal noise with a constant variance to each modulation channel. Limitations of resolution are again simulated by adding internal noise with a constant variance to each modulation channel. The 'key decision' stage will be realized as an optimal detector: a comparison of the current representation of the internal representation with the stored internal representation of the deterministic hypothesis is performed. The difference between the current representation and the stored representation is used as the basis for determining whether the hypothesis is accepted or rejected.

The output of the modulation filterbank will be treated as a sequence of events with a high temporal resolution. Each event is treated as a separate decision, and the optimal detector is used to determine whether the hypothesis is accepted or rejected.

The following stage in the model, as shown in Fig. 1, is the modulation filterbank. The modulation filterbank was presented in Fassel and Puel. This stage will be called the modulation filterbank and finally added to internal noise; this processing transforms the signals into modulation-detection data with an optimal detector as decision device. The decision device is realized as an optimal detector in the same way as described in Dau et al. Within the present model, it is postulated that the modulation-detection process is a sequential decision. The output of the 'preprocessing' stages can now be described for masking conditions using sinusoidal test signals. Limitations of resolution are again simulated by adding internal noise with a constant variance to each modulation channel. Limitations of resolution are again simulated by adding internal noise with a constant variance to each modulation channel. The 'key decision' stage will be realized as an optimal detector: a comparison of the current representation of the internal representation with the stored internal representation of the deterministic hypothesis is performed. The difference between the current representation and the stored representation is used as the basis for determining whether the hypothesis is accepted or rejected.

The output of the modulation filterbank will be treated as a sequence of events with a high temporal resolution. Each event is treated as a separate decision, and the optimal detector is used to determine whether the hypothesis is accepted or rejected.
Hypercolumns in V1: directional wavelets

- Non-linear
- Large receptive fields
- Some forms of invariance

Complex Cells

Simple cells Gabor linear models

ψ(x) = θ(x)e^{iξx}

«What» Pathway towards V4:
- More specialized invariance
- «Grand mother cells»
Perceptual Distance

- Invariance to local translations: variability reduction.
- Sensitive to elastic deformations: natural metric.
- Similar for audio.
• Signal \( f \in L^2(\mathbb{R}^d) \) whose norm is \( \|f\|^2 = \int |f(x)|^2 \, dx \).

• Normalized representation \( \Phi(f) \in \mathcal{H} : \|\Phi(f)\|^2_{\mathcal{H}} = \|f\|^2 \).

• Rigid translation invariance \( D_\tau f(x) = f(x - \tau) : \)

\[ \Phi(D_\tau f) = \Phi(f). \]

**Example:** registration \( \Phi(f) = f(x - c) \)

with center of mass: \( c = \frac{\int x |f(x)|^2 \, dx}{\|f\|^2} \)
Stability to Deformations

- $C^1$ diffeomorphisms define deformations:

$$D_\tau f(x) = f(x - \tau(x))$$

with a distance:

$$d(D_{\tau_1}, D_{\tau_2}) = \sup_x |\tau_1(x) - \tau_2(x)| + \sup_x |\nabla \tau_1(x) - \nabla \tau_2(x)|$$

$$= \|\tau_1 - \tau_2\|_\infty + \|\nabla \tau_1 - \nabla \tau_2\|_\infty.$$

- Representation invariance to local translations:

$$\|\Phi(f) - \Phi(D_\tau f)\| \leq C \|\nabla \tau\|_\infty.$$
Groups Beyond Translations

- Invariant to progressively more complex groups
  - Roto-translation (non commutative): removes local rotations
    \[ \| \Phi(f) - \Phi(D_\tau f) \| \leq C \| \nabla \tau + \nabla \tau^t \|_\infty. \]
  - Roto-translation-scaling (for images)
  - More complex specialized groups that are learned from data («group of bicycles»)
  - Group invariants define a «Gestalt»
Part I: Represent without Learning

- Local translation invariance and stability to deformations.

- Averaging and the Fourier failures

- A good representation: Mel Frequency Spectrum

- From Mel Spectrum to Wavelets

- Lost co-occurrence and interference information.

- Scattering, modulus filter banks and Convolution Networks
Part II: Applications and Learning

• Audio reconstruction and phase retrieval from modulus.

• Representation of stationary processes

• PCA, DCT and Mel Frequency Cepstrum Coefficients (MFCC)

• Classification of Sounds and Images

• Learning with sparsity
• Averaging kernel: \( \phi_J(x) = 2^{-J} \phi(2^{-J} x) \).

\[
f(x) \quad \quad \quad \Phi_J(f) = f \ast \phi_J \longrightarrow \int f(u) \, du \quad J \to \infty
\]

\[
\| \Phi_J(D_\tau f) - \Phi_J(f) \| \leq C \left( 2^{-J} \| \tau \|_\infty + \| \nabla \tau \|_\infty \right).
\]

• Invariance when \( J \) increases but \( \Phi_J(f) = f \ast \phi_J \) looses too much information.
**Fourier modulus is invariant to translations**

If \( D_\tau f(x) = f(x - \tau) \) then \( \hat{D_\tau f}(\omega) = e^{-i\tau \omega} \hat{f}(\omega) \).

\[
|\hat{D_\tau f}(\omega)| = |\hat{f}(\omega)| : \Phi(f) = |\hat{f}|.
\]

- For deformations \( D_\tau f(x) = f(x - \tau(x)) \)
  
  \( |\hat{f}(\omega)| \) is unstable at high frequencies \( \xi \).

  Scaling example: \( \tau(x) = \epsilon x \):

\[
\| |\hat{D_\tau f}| - |\hat{f}| \| \sim \| f \| \| \nabla \tau \cdot \xi \|_\infty
\]

----

*Deformation Instability of Fourier*
The loss of the Fourier phase eliminates too much information. 

\( \delta(x) \) and \( e^{ix^2} \) have same Fourier modulus (constant).

\[
\hat{f}(\omega) = \hat{g}(\omega)
\]
• Spectrogram with $w_T(t) = T^{-1}w(T^{-1}t)$ of length $T$:

$$|\hat{f}_T(u, \xi)|^2 = \left| \int f(t) w_T(u - t) e^{-iu\xi} \, du \right|^2$$

Spectrogram time-frequency resolution
Spectrogram averaging over Mel frequency bands:

\[ M_T(t, j) = \int |\hat{f}_T(u, \xi)|^2 |\hat{\psi}_j(\xi)|^2 \, d\xi \]

with Q-constant frequency bandwidth filters \( \hat{\psi}_j \)
Example of Mel Spectrum

**Spectrogram**

\[ T = 50\text{ms} \]

**Mel Spectrum**

\[ T = 800\text{ms} \]
• Dilated wavelets: \( \psi_j(t) = a^{-j} \psi(a^{-j} t) \) with \( a > 1 \).

• Wavelet transform of \( f \) at a scale \( T = a^J \):

\[
W_J f(t) = \left( \begin{array}{c}
 f \star \phi_J(t) \\
 f \star \psi_j(t)
\end{array} \right)_{j < J}
\]
Energy Conservation

\[
W_J f(t) = \left( \begin{array}{c} f \star \phi_J(t) \\ f \star \psi_j(t) \end{array} \right)_{j < J}
\]

Proposition: If \( \forall \omega \geq 0 \) , 
\[|\hat{\phi}_J(\omega)|^2 + \sum_{j<J} |\hat{\psi}_j(\omega)|^2 = 1\]

then the wavelet transform is unitary

\[
\|W_J f\|^2 = \|f \star \phi_J\|^2 + \sum_{j<J, \gamma \in \Gamma} \|f \star \psi_{j,\gamma}\|^2 = \|f\|^2
\]

and hence invertible.
Wavelets for Images

- In 2D, dilated and rotated wavelets:

\[ \psi_j,\gamma(x) = 2^{-2j} \psi(2^{-j} R_\gamma x) \]

\[ \hat{\psi}_{j,\gamma}(\omega) = 2^{-j} \hat{\psi}(2^{-j} R_\gamma \omega) \]

- Wavelet transform of \( f \) for all \( \gamma \in \Gamma \) and \( 2^j < 2^J \)

\[ W_J f(x) = \left( \begin{array}{c} f * \phi_J(x) \\ f * \psi_{j,\gamma}(x) \end{array} \right)_{j<J, \gamma \in \Gamma} \]
• Average wavelet coefficient amplitudes over $T = a^J$:

$$|f * \psi_j| * \phi_J(t)$$

Wavelet time-frequency

Mel Spectrum time-frequency

• Mel-Spectrum are averaged wavelet coefficients locally translation invariant and stable to deformations.
Scalogram to Mel Spectrum

\[ |f \ast \psi_{j_1}|(t) \quad \text{T = 50ms} \quad |f \ast \psi_{j_1}| \ast \phi_{J}(t) \]

T = 800ms
Wavelet Module: Mel Freq. Spectrum

Joakim Anden

\[ f(t) \]

\[ |f \ast \psi_{j_1}|(t) \]

\[ |f \ast \psi_{j_1}| \ast \phi_J(t) \]

Tuesday, October 4, 2011
Wavelet Modulus: SIFT

\[ f(x) \]

\[ |f \star \psi_{j_1, \gamma_1}(x)|, \ \forall j_1, \gamma_1 \]

\[ |f \star \psi_{j_1, \gamma_1}| \star \phi_J(x), \ \forall j_1, \gamma_1 \]
Invariance by Time Resolution Loss

• Example on short transients:

  ![Graph of f(x)](image1)
  ![Graph of f * ψ_j(x)](image2)

  ![Graph of |f * ψ_j(x)|](image3)
  ![Graph of |f * ψ_j| * φ_J(x)](image4)

• Problem: Important loss of information by averaging.
• Mel Spectrum over short time windows of 23ms.
• Can we recover information that remains locally invariant?
Scattering Operators

\[ W_J(|f \ast \psi_{j_1}|) = \begin{pmatrix} |f \ast \psi_{j_1} \ast \phi_J| \\ |f \ast \psi_{j_1} \ast \psi_{j_2}| \end{pmatrix} \quad j_2 < J \]

Phase removal

\[ ||f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J| \]

Co-occurrence at scales \(2^{j_1}, 2^{j_2}\).

- Iterations on \(U_J f = \begin{pmatrix} f \ast \phi_J \\ |f \ast \psi_j| \end{pmatrix} \quad j < J\)
Scattering Modulus Filter Banks

Convolution Network: LeCun, Bengio, Hinton, Poggio...
Scattering Representation

\[ S_J f(x) = \begin{pmatrix} 
  f \ast \phi_J(x) \\
  |f \ast \psi_{j_1}| \ast \phi_J(x) \\
  |f \ast \psi_{j_1} \ast \psi_{j_2}| \ast \phi_J(x) \\
  \vdots \\
  |f \ast \psi_{j_1} \cdots \ast \psi_{j_m}| \ast \phi_J(x) 
\end{pmatrix} \quad \forall j_1 \cdots j_m \]
Scattering Co-occurrence

\[ |f \ast \psi_{j_1}|(t) \quad \forall j_1 \]

\[ |f \ast \psi_{j_1} \ast \phi_J(t) \]

\[ \|f \ast \psi_{j_1} \ast \psi_{j_2}(t) \quad \forall j_1, j_2 \]

\[ \|f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J(t) \]

Tuesday, October 4, 2011
Scattering Co-occurrence

\[ |f \ast \psi_{j_1, \gamma_1}(x)| , \ \forall j_1, \gamma_1 \]

\[ |f \ast \psi_{j_1, \gamma_1}| \ast \phi_J(x) , \ \forall j_1, \gamma_1 \]

\[ \|f \ast \psi_{j_1, \gamma_1} \ast \psi_{j_2, \gamma_2}(x)\|_{j_1, \gamma_1, j_2, \gamma_2} \]

\[ \|f \ast \psi_{j_1, \gamma_1} \ast \psi_{j_2, \gamma_2} \ast \phi_J(x)\| \]
Co-occurrence gives Interferences:

\[ f(t) = \sum_m a_m \cos(\omega_m t) \]

\[ a_{j,m} = a_m \hat{\psi}(2^j \omega_m) \]

**Energy:** \( e_j^2 \)

\[ |f \ast \psi_j(t)|^2 = \sum_m |a_{j,m}|^2 + 2 \sum_{m' \neq m} a_{j,m} a_{j,m'} \cos(\omega_m - \omega_m') t \]

**Interferences:** \( \epsilon_j(t) \)

\[ |f \ast \psi_j(t)| = e_j + \frac{\epsilon_j(t)}{2e_j} + O\left(\frac{\epsilon_j^2(t)}{e_j^3}\right) \]

Music chord:

- C Major
- Minor 3rd: \( \omega_3 - \omega_2 \)
- Major 3rd: \( \omega_2 - \omega_1 \)
- Perfect 5th: \( \omega_3 - \omega_1 \)
\( U_J \) is contracting: \( \| U_J f - U_J g \| \leq \| f - g \| \).

\( U_J \) preserves norms: \( \| U_J f \| = \| f \| \).
A scattering computes for all paths $p = (j_1, \ldots, j_m)$

$$S_J(p)f = |f \ast \psi_{j_1}| \ast \ldots \ast \psi_{j_m} \ast \phi_J$$
• Scattering norm:

\[ \| S_J f \|^2 = \sum_{\text{output } p} \| S_J(p) f \|^2. \]

**Contractive** because cascade of contractive operators \( U_J \):

\[ \| S_J f - S_J g \| \leq \| f - g \|. \]
If $\psi(x) = \theta(x)e^{i\xi x}$ then $\psi_j(x) = \theta_j(x)e^{i\xi_j x}$

with $\theta_j(x) = 2^{-dj} \theta(2^{-j}x)$ and $\xi_j = 2^{-j}\xi$

so $f \star \psi_j(x) = e^{i\xi_j x} f_j \star \theta_j(x)$ with $f_j(x) = e^{i\xi_j x} f(x)$

hence $|f \star \psi_j(x)| = |f_j \star \theta_j(x)|$. 

Modulus Demodulation

Low frequency mapping
Wavelet Transform

\[ \hat{f}(\omega) \]

Tuesday, October 4, 2011
Wavelet Modulus Propagator

\[ \hat{f}(\omega) \]

\[ \tilde{f} \star \phi_{J} \]

\[ |f \star \psi_{J-1}| \]

\[ |f \star \psi_{j+1}| \]

\[ |f \star \psi_{j}| \]

\[ |f \star \psi_{j-1}| \]

\[ 2^{-j+1}\pi \]

Tuesday, October 4, 2011
Wavelet Modulus Propagator

\[ \hat{f}(\omega) \]
\[ f \ast \phi_J \]
\[ f \ast \psi_{j-1} \]
\[ f \ast \psi_{j+1} \]
\[ f \ast \psi_j \]
\[ W_J(|f \ast \psi_j|) \]

\[ |f \ast \phi_J \ast \psi_{j+2}| \]
\[ |f \ast \phi_J \ast \psi_{j+1}| \]
\[ |f \ast \phi_J \ast \psi_j| \]
\[ |f \ast \phi_J \ast \psi_{j-1}| \]

Tuesday, October 4, 2011
Theorem: For appropriate complex wavelets

\[ \| S_J f \|^2 = \sum_{\text{output } p} \| S(p) \|^2 = \| f \|^2 . \]
Theorem: \( \lim_{J \to \infty} \| S_J f - S_J g \| \) converges and

\[
\lim_{J \to \infty} 2^{2J} |f \ast \psi_{j_1} \ldots \ast \psi_{j_m} \ast \phi_J(x) = \int |f \ast \psi_{j_1} \ldots \ast \psi_{j_m}(u)| \, du.
\]

if \( D_\tau f(x) = f(x - \tau) \) is a translation then

\[
\lim_{J \to \infty} \| S_J f - S_J(D_\tau f) \| = 0.
\]
Theorem \hspace{1em} \text{If } D_\tau f(x) = f(x - \tau(x)) \text{ with } \|\nabla \tau\|_\infty < 1 \\
then \text{ for } J > \log \frac{\|\tau\|_\infty}{\|\nabla \tau\|_\infty} \\
\|S_J f - S_J(D_\tau f)\| \leq C m \|f\| \log \left(\frac{\|\tau\|_\infty}{\|\nabla \tau\|_\infty}\right) \|\nabla \tau\|_\infty
Scattering coefficients $S_J f(x)$ are averaged by $\phi_J$.

If $f(n)$ is of size $N$

Compute only $S_J f(2^J n) : 2^{-2J} N$ scattering vectors.

$O(N)$ coefficients computed with $O(N \log N)$ operations.
• \( \lim_{J \to \infty} S_J f \) converges if it is renormalized.

• \( S_J(p)f(t) \) is represented by \( \frac{S_J(p)f(t)}{\mu_J(p)} \) over a rectangle of surface \( \mu(p)^2 \).

\[
f(t) = \cos(\omega_1 t)
\]

\[
|f \ast \psi_j \ast \phi_J| \quad ||f \ast \psi_j \ast \psi_{j_2} \ast \phi_J|
\]
\[ f(t) = \cos(\omega_1 t) \]

\[ |f \ast \psi_{j_1}| \ast \phi_J \quad ||f \ast \psi_{j_1} \ast \psi_{j_2}| \ast \phi_J \]
\[ f(t) = \cos(\omega_1 t) + \cos(\omega_2 t) \]
Scattering Zoom

\[ f(t) = \cos(\omega_1 t) + \cos(\omega_2 t) \]

\[ |f \ast \psi_{j_1} \ast \phi_J| \quad ||f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J|| \]
Scattering Multiple Interferences

\[ f(t) = \cos(\omega_1 t) + \cos(\omega_2 t) + \cos(\omega_3 t) \]

\[ |f \ast \psi_{j_1} \ast \phi_J| \quad |f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J| \]
Scattering Zoom

\[ f(t) = \cos(\omega_1 t) + \cos(\omega_2 t) + \cos(\omega_3 t) \]

\[ |f \ast \psi_{j_1} \ast \phi_J| \quad ||f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J|| \]
Scattering of an Attack

\[ f(t) = g(t) \cdot \cos(\omega_1 t) \]

\[ |f \star \psi_{j_1}| \star \phi_J \quad \| f \star \psi_{j_1} \star \psi_{j_2} \star \phi_J \]
Scattering Zoom

\[ f(t) = g(t) \cdot \cos(\omega_1 t) \]

\[ |f \star \psi_{j_1} \star \phi_J| \]

\[ ||f \star \psi_{j_1} \star \psi_{j_2} \star \phi_J|| \]
Audio Scattering

\[ f(x) \]

\[ |f \ast \psi_{j_1} \ast \phi_J| \]

\[ |f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J| \]
Scattering Zoom

\[ f(x) \]

\[ |f \ast \psi_{j_1} \ast \phi_J| \]

\[ |f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J| \]
Overview: Part II

- Audio reconstruction and phase retrieval from modulus.

- Representation of stationary processes

- PCA, DCT and Mel Frequency Cepstrum Coefficients (MFCC)

- Classification of Sounds and Images

- Learning with sparsity
\[ U_J \text{ is contracting: } \| U_J f - U_J g \| \leq \| f - g \|. \]

\[ U_J \text{ preserves norms: } \| U_J f \| = \| f \|. \]
Inverse Scattering

$U^{-1}_j$

$U^{-1}_j$

$U^{-1}_j$

$U^{-1}_j$

$|f \star \psi_{j_1}|$

$|f \star \psi_{j_2}|$

$|f \star \psi_{j_1}| \star \phi_J$

$|f \star \psi_{j_2}| \star \phi_J$

Input

Tuesday, October 4, 2011
Theorem  For appropriate wavelets

\[
U_J f = \left( f \ast \phi_J \bigg| f \ast \psi_j \right)_{j<J}
\]

is invertible and the inverse is continuous.
Fourier Phase Retrieval

• A signal $f(t)$ can not be recovered from $|\hat{f}(\omega)|$

• Only the minimum phase signal can be reconstructed

• No unique reconstruction from a spectrogram (needs complementary information).

• Alternate projection reconstruction (Griffin & Lim) algorithms with good perceptual results.
Reconstruct a wavelet transform \( \{g_j\}_{j \in \mathbb{Z}} \) such that
\[
\forall j \quad |g_j(x)| = |f \ast \psi_j(x)|. \quad C_1 : \text{non-convex.}
\]

\( \{g_j\}_{j \in \mathbb{Z}} \) is a wavelet transform a signal if and only if
\[
\forall j, x \quad g_j(x) = \sum_{j'} \int K(j, j', x - x') \, g_{j'}(x') \quad C_2 : \text{linear.}
\]

Alternate projection on constraints:
- Errors decreases through iterations
- Convergence depends upon initial point.
- Instabilities: similar to audio perception.
Original audio signal $f$

Reconstruction from $S_J f$ with $2^J = 3$ s

From coefficients of order 1:

From coefficients of order 1 and 2:
Texture Discrimination

- Textures are realizations of stationary processes $F$ but typically not Gaussian and not Markovian.
- An open problem for 30 years.
Scattering Stationary Processes

If $F(x)$ is a stationary process then

$$S_J(p)F = |\cdots|F \ast \psi_{j_1}| \ast \cdots | \ast \psi_{j_m}| \ast \phi_J$$

is stationary

$$E\{S_JF(x)\} = \begin{pmatrix}
E\{F\} \\
E\{|F \ast \psi_{j_1}|\} \\
\vdots \\
E\{|F \ast \psi_{j_1}| \ast \cdots \ast \psi_{j_m}|\} \\
\end{pmatrix}_{\forall j_1 \cdots j_m}$$

$$\forall \gamma_1 \cdots \gamma_m$$

does not depend upon $J$ and $x$.

**Theorem:** For appropriate complex wavelets

$$E\{|S_JF|^2\} = \sum_{\text{all } p} E\{|S_J(p)F|^2\} = E\{|F(x)|^2\}.$$
Conjecture: for a wide class of "ergodic" stationary processes

\[
\lim_{J \to \infty} \| S_J F - E\{ S_J F \} \| = 0 : \text{with probability 1.}
\]
• Usual approaches use high order moments: large variance estimators. Not enough training samples.

• Non-gaussian process models with first and second order moments of scattering coefficients: co-occurrence information (Bela Julesz conjecture for images).

• Non Gaussian property estimation with small variance.
Classification

- $K$ classes corresponding to $K$ (non stationary) processes $\{F_k\}_{k \leq K}$

- Scattering transformation.

- Two possible strategies: discriminant or generative classifiers.
  - Discriminant (e.g. SVM): asymptotically optimal.
  - Generative: can be better on small training sets or large number of classes.
Each class is represented by the centroid $E\{S_J F_k\}$ and a space $V_{d,k}$ of principal variance directions (PCA).

Affine space model $A_{d,k} = E\{S_J F_k\} + V_{d,k}$. 
Given \( f(t) = e \star h(t) \), with \( \hat{h}(\omega) \) smooth.

\[
|f \ast \psi_{j_1} \ast \phi_J(t)| = |\hat{h}(a^{-j_1} \pi)| \cdot |e \ast \psi_{j_1} \ast \phi_J(t)|
\]

\[
||f \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J(t)|| = |\hat{h}(a^{-j_1} \pi)| \cdot ||e \ast \psi_{j_1} \ast \psi_{j_2} \ast \phi_J(t)||
\]
Taking logarithm turns convolution into linear combination.

\[
\log | f \ast \psi_{j_1} | \ast \phi_J(t) = \log | \hat{h}(a^{-j_1} \pi) | + \log | e \ast \psi_{j_1} | \ast \phi_J(t)
\]

\[\begin{align*}
\text{smooth in } j_1 & \\
\text{not smooth in } j_1
\end{align*}\]

For fixed \( t \), DCT in \( j_1 \) separates \( h \) from \( e \).

\[
\text{DCT} \left\{ \log | f \ast \psi_{j_1} | \ast \phi_J(t) \right\} (k_1) = \text{DCT} \left\{ \log | \hat{h}(a^{-j_1} \pi) | \right\} (k_1)
\]
\[
+ \text{DCT} \left\{ \log | e \ast \psi_{j_1} | \ast \phi_J(t) \right\} (k_1)
\]

For \( q = 2 \), first DCT along \( j_2 \) (\( j_1 \) fixed), then \( j_1 \) (\( j_2 \) fixed).

1st order: keep 50%

2nd order: keep 20%
DCTs efficiently decorrelate log-scattering coefficients.

\[ N = 16384 \]
Classification

- Scale $J$:
  How does varying scale affect classification?
  What is the appropriate scale for a problem?

- Maximum order $m$:
  What is the impact of higher orders?
  Can we recover lost temporal structure?

- How does scattering compare with other methods?

**$\Delta$-MFCCs:** complement $M_k[n]$ with $M_k[n] - M_k[n-1]$.

\[
\begin{array}{c|c}
M_k[n-1] & M_k[n] \\
T & T \\
\end{array} \quad \rightarrow \quad \begin{array}{c|c|c}
M_k[n] & M_k[n] - M_k[n-1] \\
2T & 2T
\end{array}
\]
• Genre classification on GTZAN, 10 genres of 100 clips each.

• Bag-of-frames: independent modeling.

  Discriminative model: SVM
  Generative model: PCA, GMM

• Unknown clip classified by aggregating over frames.
• Larger scales, more global descriptors, better results.
• $\Delta$-MFCCs add temporal dynamics, but crudely.
• Second-order CLS provides much richer representation.
• State of the art at 7% (cortical representations) and 17% (AdaBoost on MFCCs and local spectral features)

<table>
<thead>
<tr>
<th>T (s)</th>
<th>0.02 PCA</th>
<th>0.2 PCA</th>
<th>1.5 PCA</th>
<th>1.5 SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFCC</td>
<td>46</td>
<td>36</td>
<td>37</td>
<td>28</td>
</tr>
<tr>
<td>$\Delta$-MFCC</td>
<td>37</td>
<td>33</td>
<td>30</td>
<td>26</td>
</tr>
<tr>
<td>CLS, m=1</td>
<td>46</td>
<td>36</td>
<td>36</td>
<td>28</td>
</tr>
<tr>
<td>CLS, m=2</td>
<td>34</td>
<td>23</td>
<td>23</td>
<td>18</td>
</tr>
</tbody>
</table>
Classification of Textures

CUREt database

61 classes

Rotations and illumination variations.

<table>
<thead>
<tr>
<th>Training per class</th>
<th>Scat PCA</th>
<th>Scat SVM</th>
<th>Mark. Rand.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 2$</td>
<td>$m = 2$</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.9%</td>
<td>3.3%</td>
<td>22.43%</td>
</tr>
<tr>
<td>46</td>
<td>0.09%</td>
<td>1.1%</td>
<td>2.46%</td>
</tr>
</tbody>
</table>

Cross-validation: $2^J = N$ and $d \approx 100$. 

Tuesday, October 4, 2011
### Digit Classification: MNIST

<table>
<thead>
<tr>
<th>Training Size</th>
<th>Conv. Net.</th>
<th>Scat PCA $m = 2$</th>
<th>Space dim. d</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>7.18</td>
<td>6.05</td>
<td>24</td>
</tr>
<tr>
<td>5000</td>
<td>1.52</td>
<td>1.22</td>
<td>40</td>
</tr>
<tr>
<td>400000</td>
<td>0.65</td>
<td>0.78</td>
<td>180</td>
</tr>
</tbody>
</table>

Cross-validation: $2^J = 8$. 

---

Tuesday, October 4, 2011
Combined Scattering

- Translation group scattering: not sufficient for complex classes

- Intra-class variability need to be further reduced:

\[ f \rightarrow S_J^{\text{Trans}} \rightarrow S_{J'}^G \rightarrow ... \]

- Scattering \( S_{J'}^G \) over a compact Lie group \( G \) with iterated wavelet transforms over \( G \) cascaded with modulus operators.

- How to learn appropriate "structured" Lie Groups \( G \)?
Sparse coding of natural images

- Natural images have a structure that is sparse

\[ I(x) = \sum_i u_i A_i(x) + \nu(x) \]

- Image patch
- Coefficients
- Features
- Noise

Sparse prior on coefficients
First Layer
Probabilistic Model

\[ P(I, a, \phi) \propto e^{-E_1} \]

\[
E_1 = \sum_t \sum_x \frac{1}{\sigma^2_N} \left[ I(x,t) - \sum_i \Re\{z_i^*(t) A_i(x)\} \right]^2 + \lambda_{Sp} \sum_{i,t} a_i(t) + \lambda_{Sl} \sum_{i,t} (a_i(t) - a_i(t-1))^2
\]

Reconstruction Error
Sparse
Slow
Adapt to Natural Movies
Learned Basis Functions

\[ a \quad \phi \quad u^R \quad u^I \quad A^R_{191} \quad A^I_{191} \quad I \]
Conclusion

• Representation for classification: variability reduction.

• Lie Group invariants that are stable and informative. A scattering provides such invariant with convolution networks.

• Complements Mel Spectrum Coefficients

• Impact on audio perception and processing ?

• Learning groups with sparsity: a major challenge

• Papers/softwares: www.cmap.polytechnique.fr/scattering