Discrete Auditory Transforms

Jack Xin and Yingyong Qi

Outline

- **Invertible but Redundent Auditory Transforms (Frames).**
- **Orthogonal Discrete Auditory Transform (DAT).**
- **Comparison with FFT.**
- **Conclusions and Future Work.**

Supported in part by NSF-ITR and the Guggenheim Foundation.

From Waves to Transform

Represent ear response to single input frequency f by:

$$U(x, f)\, e^{i2\pi f t} + c.c.$$

Here t is time, x is spatial correspondance of frequency, e.g., distance along Basilar Membrane (BM); c.c denotes complex conjugate of preceding term. Let:

$$m^2 = m^2(f) = \int_0^L |U(x, f)|^2\, dx. \tag{1}$$

Continuous Transform

$$S(t,x) = \int_{R^1} d\tau\, s(\tau) \int_{R^1} df\, \frac{U^*(x,f)}{m(f)}\, e^{2\pi f i(t-\tau)}.$$

$*$ **complex conjugate operation.**

Transform is a temporal convolution of s **with kernel:**

$$K(t,x) \equiv \int_{R^1} df\, \frac{U^*(x,f)}{m(f)}\, e^{2\pi f i t}.$$

Continuous Transform

$$S(t,x) = \int_{R^1} df\, e^{2\pi f i t} \frac{U^*(x,f)}{m(f)} \int_{R^1} d\tau\, s(\tau)\, e^{-2\pi f i \tau} \tag{2}$$

$$= \int_{R^1} df\ \hat{s}(f)\, \frac{U^*(x,f)}{m(f)}\, e^{2\pi f i t}, \tag{3}$$

hat denotes Fourier transform.

Continuous Transform

• **Transform is a superposition of time harmonics with proper normalization.**

• **Transform formally reduces to wavelet transform, if** $\frac{U(x,f)}{m(f)}$ **is approximated as** $\Psi\left(x - \log f\right)$**, for some nonlinear function** Ψ**, as observed by (I. Daubechies, 92).**

• **Transformed variable** S **contains information on both frequency (in** x**) and time** t**.**

Inversion

Inversion formula:

$$s(t) = \int_0^L dx \int_{R^1} d\tau\, S(\tau, x) \int_{R^1} df\, \frac{U(x,f)}{m(f)}\, e^{2\pi i f(t-\tau)}.$$

The integral over τ is Fourier transform of $S(\tau, x)$ in τ. In view of (3), right hand side equals:

$$\begin{aligned} &\int_0^L dx \int_{R^1} \frac{df}{m(f)} U(x,f)\, e^{2\pi i f t}\, \frac{\hat{s}(f)}{m(f)} U^*(x,f), \\ &= \int_{R^1} df\, e^{2\pi i f t}\, \hat{s}(f) = s(t). \end{aligned} \tag{4}$$

thanks to normalization (1).

Energy Conservation

$$\int_0^L dx \int_{R^1} dt\, |S(t,x)|^2 = \int_0^L dx \int_{R^1} df\, \frac{|\hat{s}(f)|^2}{m^2(f)} |U(x,f)|^2$$
$$= \int_{R^1} df\, |\hat{s}(f)|^2 = \int_{R^1} s^2(t)\, dt.$$

Transform conserves L^2 norm.

RDAT

$$\mathbf{S_{j,m}} \equiv \sum_{\mathbf{l=0}}^{\mathbf{N-1}} \mathbf{s_l} \sum_{\mathbf{n=0}}^{\mathbf{N-1}} \mathbf{X_{m,n}\, e^{i(2\pi(j-l)n/N)}},$$

$X_{m,n}$ **square sum equal to one in** m **(**$m = 1 : M$**,** $M = N/2$**):**

$$\sum_{m=0}^{M-1} |X_{m,n}|^2 = 1, \quad \forall n.$$

Let (Discrete Fourier Transform –DFT):

$$\hat{s}_k = \sum_{n=0}^{N-1} s_n\, e^{-i(2\pi nk/N)},$$

$$S_{j,m} = \sum_{n=0}^{N-1} \hat{s}_n\, X_{m,n}\, e^{i(2\pi nj/N)}. \tag{5}$$

Inverse RDAT

$$\mathbf{s_j = \frac{1}{N^2} \sum_{m=0}^{M-1} \sum_{l=0}^{N-1} S_{l,m} \sum_{n=0}^{N-1} X^*_{m,n} \, e^{i(2\pi(j-l)n/N)}}.$$

Proof: Consider the sum in l. In view of (5),

$$\frac{1}{N} \sum_{l=0}^{N-1} S_{l,m} \, e^{i(2\pi(j-l)n/N)} = e^{2\pi ijn/N} \, \hat{s}_n \, X_{m,n}.$$

So the right hand side equals:

$$\frac{1}{N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |X_{m,n}|^2 \, e^{2\pi ijn/N} \, \hat{s}_n,$$

summing over m and using normalization property of $X_{m,n}$:

$$\frac{1}{N} \sum_{n=0}^{N-1} e^{2\pi ijn/N} \, \hat{s}_n = s_j.$$

Model and Transform Kernel

- **DFT of a real vector s satisfies the symmetry property $\hat{s}_k = \hat{s}^*_{N-k}$, $k = 1, 2 \cdots, N-1$.**

- **Transform kernel $X_{m,n}$ respects the symmetry while spectrally spreading.**

- **Discrete signal s has sampling frequency F_s (Hz). DFT component $\hat{s}_n$ $(0 \leq n \leq N/2)$ corresponds to frequency:**

$$f_n = F_s \cdot n/N, \quad n \leq N/2.$$

x_m's are sampled frequencies or corresponding place locations.

Model and Transform Kernel

- **For x_m, $0 \leq m \leq N/2-1$, define $X_{m,n}$:**

$$X_{m,0} = \frac{U^*(x_m, f_1)}{m_f(f_1)},$$

$$X_{m,n} = \frac{U^*(x_m, f_n)}{m_f(f_n)}, \quad 1 \leq n \leq N/2-1,$$

$$X_{m,n} = \frac{U^*(x_m, f_{N-n})}{m_f(f_{N-n})}, \quad N/2 \leq n \leq N-1,$$

m_f **function is:**

$$m_f(f) = \left(\sum_{m=0}^{N/2-1} |U(x_m, f)|^2 \right)^{1/2}. \tag{6}$$

- $X_{m,n}$ **is (1) symmetric in n with respect to $N/2$, (2) periodic in n, (3) square sum to one along m.**

Other Choices of Kernel

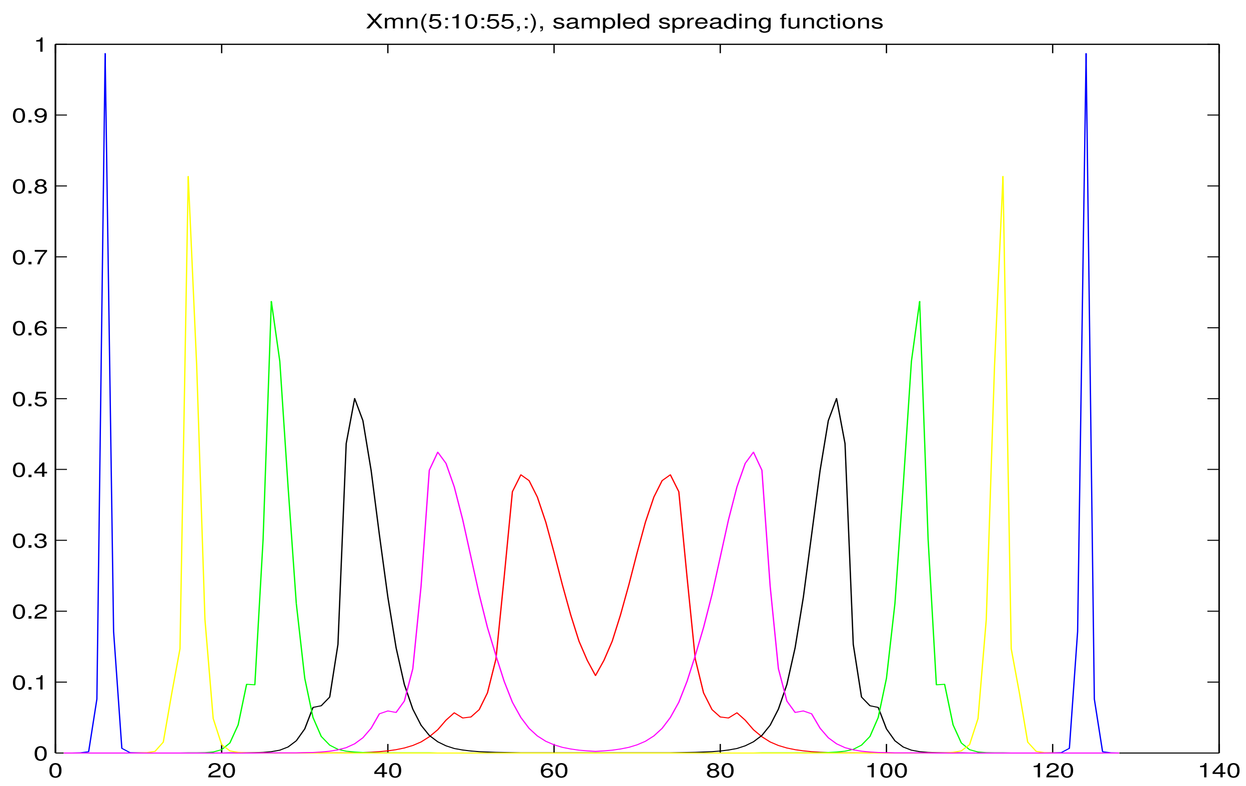
- **Use Schroeder-Atal-Hall's psycho-acoustic spreading function, a function of f - f', with f and f' on the Bark frequency scale.**

Bark scale has a one-to-one mapping to linear frequency scale.

- **Sample this function in $f = (f_i)$, $f' = (f_j)$, f_j linearly distributed, to form the matrix X_{mn}.**

- **Use hearing data (tuning curves) to infer X_{mn}.**

Schroeder Kernel (N=128)



Orthogonality

- **Matrices X_{mn} constructed from physiological models contain filtering charateristics of ear; however, are far from being orthogonal, a desired mathematical property for signal processing.**

- **If $X_{m,n}$ were orthogonal, the transfrom simplifies to:**

$$T_m \equiv S_{0,m} = \sum_{n=0}^{N} \hat{s}_n X_{m,n}.$$

Transform is simply DFT times auditory matrix $(X_{m,n})$. Redundancy is removed, $(T_m) \in C^N$.

- **Question: how does one construct an Orthogonal Auditory Matrix ?**

Schrödinger Equation

For complex scalar function $u = u(x,t)$, consider evolution equation ($*$ convolution):

$$i\,u_t = H\,u \equiv u_{xx} + V(x) * u, \qquad u(x,0) = u_0(x), \;\; x \in R^1, \; t \geq 0.$$

- **Schrödinger map SM is: $u_0(x) \longrightarrow u(x,1)$.**
- **If $V(x) = V(-x)$ (even) and real, then:**

$$\frac{d}{dt}\int_{R^1} |u|^2(x,t)\,dx = 0.$$

SM is orthogonal in L^2, and time reversible.

Convolution with V captures the long range auditory spreading (across critical bands).

Schrödinger Equation

• **Schrödinger evolution has spreading (dispersive smoothing) property. In the absence of V (Jensen, 86):**

$$\|e^{itH}\|_{SM(H^{0,k};H^{k,-k})} \leq C\,(|t|^{-k} + |t|^{k}),$$

$t \neq 0$, k **a positive interger,**

$$H^{m,s} = \{\psi \in L^2(R^n) : \ \|(1+|x|^2)^{s/2}(1-\Delta)^{m/2}\psi\|_2 < \infty.\}$$

Δ**:** n **dimensional Laplacian.**

Similar estimates for Schrödinger with nonlocal potential ($\in H^{k,0}_{\infty}$) and cubic nonlinearity, Hayashi and Ozawa (89).

Self-adjoint with variable second order coefficient, Kapitanski and Safarov (96).

Discrete Schrödinger

For (lattice) constants $\sigma_1 > 0$, $\sigma_2 > 0$, coupled (ODE) system ($n = 1 : N$):

$$i\, u_{n,t} = \sigma_1 \left(u_{n+1} - 2\, u_n + u_{n-1}\right) + \sigma_2 \sum_{m=1}^{N} V_{m,n}\, u_m,$$

or in matrix form:

$$i\, \mathbf{u_t} = \left(\sigma_\mathbf{1}\, \mathbf{A} + \sigma_\mathbf{1}\, \mathbf{B}\right) \mathbf{u},$$

A**,** B**, are real and symmetric** $N \times N$ **matrices.**

Discrete Schrödinger

- **Schrödinger Map is:**

$$SM = \exp\{i(\sigma_1\, A + \sigma_1\, B)\}.$$

$$SM * SM^{'} = \exp\{i(\sigma_1\, A + \sigma_1\, B)\}\ \exp\{-i(\sigma_1\, A + \sigma_1\, B)\} = Id.$$

Prime $'$ is complex conjugate transpose.

- **Orthogonal matrices resembling auditory responses are also constructed based on compactly supported wavelets (I. Daubechies, 88). However, they are not as flexible for injecting knowledge as Schrödinger based constructions.**

Orthogonal DAT

- **Let Schrödinger map** SH **act on** $\hat{s}_k$**,** $k = 2, \cdots, N/2$**, and the reversely rearranged Schrödinger map** $\widehat{SH}$ **(unitary too) acts on** $\hat{s}_k$**,** $k = N/2 + 2, \cdots, N$**. Leave DC and Nyquist modes invariant.**

- **Orthogonal auditory matrix in block diagonal form:**

$$OAM = \mathrm{diag}\{1, SH, 1, \widehat{SH}^*\}.$$

- **Orthogonal DAT is the product OAM · DFT.**

- **Use Schroeder's spreading function to construct** $V_{m,n}$**.**