Introduction to Time-Frequency Distributions

Selin Aviyente
Department of Electrical and Computer Engineering
Michigan State University

January 19, 2010
Motivation for time-frequency analysis

- When you listen to music, you hear the time variation of the sound frequencies. These localized frequency events are not pure sinusoids but packets of close frequencies.
- We need to decompose signals over elementary functions that are well concentrated in time and frequency.
- There is no unique choice for the ’elementary functions’. Two important classes of local, linear time-frequency decomposition methods are short-time Fourier transform and wavelet transform.
- Desired properties for the elementary functions are localization in time and frequency, orthogonal/orthonormal basis for $L_2$, and should quantify the time variations of instantaneous frequency.
The short-time Fourier transform is the most widely used method for studying nonstationary signals. The basic idea of STFT is to break up the signal into small time segments and Fourier analyze each time segment to determine the frequencies that existed in that segment. The totality of such spectra indicates how the spectrum is varying with time. Applications: Music analysis (to localize the different instruments), speech analysis (to localize the different vowels or different speakers)
Definition

- STFT is a linear time-frequency transform that correlates the signal with a family of waveforms that well concentrated in time and frequency. (Compare to Fourier Transform which correlates the signal with eternal sinusoids, good frequency localization, poor time localization)

- Consider a family of time-frequency atoms \( \{ \phi_{\gamma} \}_{\gamma \in \Gamma} \) where \( \gamma \) might be a multi-index parameter. Suppose that \( \phi_{\gamma} \in L_2(\mathbb{R}) \) and \( \| \phi_{\gamma} \| = 1 \).

- The corresponding linear time-frequency transform is defined as \( < f, \phi_{\gamma} > = \int f(t) \phi_{\gamma}^*(t) dt \).

- Gabor introduced STFT to measure the frequency variations of sounds [1946].
In STFT, the time-frequency atoms are chosen such that they are localized around a time center and frequency center, \( \phi_\gamma(\tau) = g_{t,\omega}(\tau) = e^{j\omega\tau} g(\tau - t) \).

The window \( g(\tau) \) is real, symmetric and normalized \( \| g_{t,\omega} \| = 1 \). It is translated by \( t \) and modulated by the frequency \( \omega \).

Therefore, Short-Time Fourier transform of signal \( f(t) \) is given as:

\[
S(t, \omega) = \langle f, g_{t,\omega} \rangle = \int f(\tau) g(\tau - t) e^{-j\omega\tau} d\tau \tag{1}
\]

By Parseval’s theorem, \( \int f(t) \phi_\gamma^*(t) dt = \frac{1}{2\pi} \int F(\omega) \Phi_\gamma^*(\omega) d\omega \).

Therefore, STFT can also be written as:

\[
S(t, \omega) = \frac{1}{2\pi} \int F(\lambda) G(\lambda - \omega) e^{i(\lambda - \omega)t} d\lambda \tag{2}
\]
One can define an energy density, called the spectrogram, by:

\[ P_S(t, \omega) = |S(t, \omega)|^2 = |\int f(\tau)g(\tau - t)e^{-j\omega \tau} d\tau|^2 \]  

The spectrogram measures the energy of the signal in the time-frequency neighborhood of \((t, \omega)\).
The location of information provided by $\langle f, g_{t,\omega} \rangle$ depends on the time-frequency spread of the window $g_{t,\omega}$.

Since $\|g_{t,\omega}\|^2 = \int |g(\tau - t)|^2 d\tau = 1$, we can interpret $|g_{t,\omega}|^2$ as probability distribution.

The center of this distribution is at $t$ since $g$ is even symmetric.

The time spread around $t$ is:

$$\sigma_{\tau}^2 = \int (\tau - t)^2 |g_{t,\omega}(\tau)|^2 d\tau = \int \tau^2 |g(\tau)|^2 d\tau \quad (4)$$
Similarly, we can determine the localization in frequency.

The Fourier transform of $g$ is real and symmetric since $g$ is real and even. The Fourier transform of $g_{t,\omega}$ is

$$G_{t,\omega}(\lambda) = G(\lambda - \omega) \exp(-jt(\lambda - \omega))$$  \hspace{1cm} (5)

Its frequency center is $\omega$.

The frequency spread around $\omega$ is

$$\sigma^2_\lambda = \frac{1}{2\pi} \int (\lambda - \omega)^2 |G(\lambda)|^2 d\lambda = \frac{1}{2\pi} \int \lambda^2 |G(\lambda)|^2 d\lambda$$  \hspace{1cm} (6)

Therefore, $g_{t,\omega}$ corresponds to a Heisenberg box of area $\sigma_T \sigma_\lambda$ centered at $(t, \omega)$. The size of this box is independent of $(t, \omega)$ which means that STFT has the same resolution across the time-frequency plane.
Examples

Example 1: Find the STFT of $f(t) = \exp(j\omega_0 t)$:

$$S(t, \omega) = \int e^{j\omega_0 \tau} g(\tau - t) e^{-j\omega \tau} d\tau$$

$$= \int g(\tau - t) e^{-j(\omega - \omega_0)\tau} d\tau$$

$$= G(\omega - \omega_0) \exp(-jt(\omega - \omega_0)) \quad (7)$$

Example 2: Find the STFT of $f(t) = \delta(t - t_0)$:

$$S(t, \omega) = g(t_0 - t) \exp(-j\omega t_0) \quad (8)$$

Example 3: Find the STFT of a linear chirp $f(t) = \exp(jat^2)$ for a Gaussian window $g(t) = (\pi\sigma^2)^{-1/4} \exp(-t^2/(2\sigma^2))$:

$$P_S(t, \omega) = \left(\frac{4\pi\sigma^2}{1 + 4a^2\sigma^4}\right)^{1/2} \exp\left(-\frac{\sigma^2(\omega - 2at)^2}{1 + 4a^2\sigma^4}\right) \quad (9)$$
You can recover back the signal from its STFT as:

\[ f(t) = \frac{1}{2\pi} \int \int S(\tau, \omega) g(t - \tau) e^{j\omega t} d\omega d\tau \]  \hspace{1cm} (10)

Parseval’s Theorem: The energy is preserved by STFT:

\[ \int |f(t)|^2 dt = \frac{1}{2\pi} \int \int |S(t, \omega)|^2 d\omega dt \]  \hspace{1cm} (11)

Energy preservation implies that \( S(t, \omega) \in L_2(\mathbb{R}^2) \).
Properties of STFT

- STFT represents a one-dimensional signal by a two-dimensional function, \( S(t, \omega) \in L_2(\mathbb{R}^2) \). Because STFT is redundant, it is not true that any \( \Phi \in L_2(\mathbb{R}^2) \) is the STFT of some function, \( f \in L_2(\mathbb{R}) \). STFT is usually complex valued.

- Time marginal is not satisfied:

\[
P(t) = \int |S(t, \omega)|^2 d\omega = \frac{1}{2\pi} \int f(\tau)g(\tau - t)f^*(\tau')g^*(\tau' - t)e^{-j\omega(\tau - \tau')}d\tau d\tau' d\omega \\
= \int f(\tau)g(\tau - t)f^*(\tau')g^*(\tau' - t)\delta(\tau - \tau')d\tau d\tau' \\
= \int |f(\tau)|^2|g(\tau - t)|^2 d\tau
\]

(12)
Similarly, frequency marginal is not satisfied, i.e.
\[ \int \int |S(t, \omega)|^2 dt \neq |S(\omega)|^2. \]

Spectrogram scrambles the energy distributions of the window with those of the signal. This introduces effects unrelated to the properties of the original signal.
The resolution in time and frequency of STFT depends on the spread of the window in time and frequency. This can be measured from the area $\sigma_t\sigma_\omega$.

In numerical applications, $g$ must have compact support.

For good time resolution, we have to pick a narrow window in the time domain.

For good time frequency resolution, the bandwidth of the window should be small, the ratio of the first sidelobes to the peak of $G(\omega)$ should be large, the rate of decay of $G(\omega)$ should be fast.

There is time-frequency resolution tradeoff. The degree of trade-off depends on the window, signal, time and frequency.

The minimum time-frequency bandwidth product is for Gaussian window.
Consider discrete signals of period $N$. The window is a symmetric discrete window, $g[n]$, with unit norm $\|g\| = 1$.

The time-frequency atoms are $g_{m,l}[n] = g[n - m] \exp\left(\frac{i2\pi ln}{N}\right)$.

The discrete-time STFT is given as:

$$S[m, l] = \langle f, g_{m,l} \rangle = \sum_{n=0}^{N-1} f[n]g[n - m] \exp\left(-\frac{j2\pi ln}{N}\right)$$  \hspace{1cm} (13)

for $0 \leq m < N$ and $0 \leq l < N$.

The computational complexity is $N$ FFTs of length $N$ requiring a total of $O(N^2 \log N)$ operations.
The discrete time reconstruction is:

\[
f[n] = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} S[m, l] g[n - m] \exp\left(\frac{j2\pi ln}{N}\right)
\]  \hspace{1cm} (14)

Unique reconstruction is only possible if appropriate sampling in time and frequency are chosen.

The sampling period in time, \(T\), and sampling rate in frequency \(\Omega\) should satisfy the relationship \(\Omega T \geq 2\pi\).
Summary of STFT and Spectrogram

- Linear
- STFT is complex, spectrogram is always positive.
- Preserves the energy and can be inverted.
- It is time and frequency shift invariant.
- Suffers from time-frequency resolution tradeoff.
- Does not satisfy the marginals, not a proper distribution.
The goal of time-frequency analysis is to track the time variation of frequency. Therefore, we need to define a concept of instantaneous frequency.

For a signal \( f(t) = A \cos(\omega_0 t + \phi_0) = A \cos(\phi(t)) \), \( \phi(t) = \omega_0 t + \phi_0 \) is called the instantaneous phase.

The instantaneous frequency is then defined as the rate of change of phase: \( \frac{d\phi(t)}{dt} = \omega_0 \).

Any signal can be written in the form \( A(t) \cos(\phi(t)) \) where \( A(t) \geq 0 \) (AM-FM decomposition)

However, \( A(t) \) and \( \phi(t) \) are not unique for a given \( f(t) \).
Analytic Signal

- A particular decomposition is obtained from the analytic part of a signal.
- Analytic signal, \( f_a(t) \) is defined as:
  \[
  F_a(\omega) = \begin{cases} 
    2F(\omega) & \omega \geq 0 \\
    0 & \omega < 0 
  \end{cases} 
  \]  
  (15)
- This complex signal is represented as 
  \( f_a(t) = A(t) \exp(j\theta(t)) \) and \( A(t) \) is called the analytic amplitude, \( \theta'(t) \) is called the instantaneous frequency.
- Example: If \( f(t) = acos(\omega_1 t) + acos(\omega_2 t) \), then
  \[
  f_a(t) = a \exp(j\omega_1 t) + a \exp(j\omega_2 t) \\
  = 2acos(\frac{1}{2}(\omega_1 - \omega_2)t) \exp(\frac{j}{2}(\omega_1 + \omega_2)t) 
  \]  
  (16)
- Therefore, the instantaneous frequency is \( (\omega_1 + \omega_2)/2 \) and the amplitude is \( 2a|cos(\frac{1}{2}(\omega_1 - \omega_2)t)|. \)
A lot of real life signals can be modeled as sums of sinusoidal partials

\[ f(t) = \sum_{k=1}^{K} a_k(t) \cos(\theta_k(t)) \]

such as musical sounds and voiced speech segments.

Therefore, it is important to be able to identify the multiple frequencies instead of the average frequency.

One approach is to determine the local maxima of \( P_S(t, \omega) \) to identify the instantaneous frequencies.

\( P_S(t, \omega) \) measures energy of the signal in a neighborhood of \((t, \omega)\).

For a general symmetric window \( g_s(t) = \frac{1}{\sqrt{s}} g(t/s) \) (scaled versions of a window function), we can define STFT as

\[ < f, g_s, \omega, t > = \int f(\tau) g_s(\tau - t) e^{-j\omega \tau} d\tau. \]
For a signal \( f(t) = a(t) \cos(\theta(t)) \),
\[
S(t, \omega) = \frac{s}{2} a(t) \exp(j[\theta(t) - wt]) (G(s(\omega - \theta'(t))) + \epsilon(t, \omega)).
\]
The maximum amplitude will be achieved around \( G(0) \) since for a window function \( |G(0)| \geq |G(\omega)| \).

\( G(0) \) corresponds to the frequency \( \omega = \theta'(t) \), i.e. the instantaneous frequency, and the amplitude of the ridge is \( a(t)\sqrt{s}/2 \).

If we have multiple frequencies:
\[
S(t, \omega) = \frac{\sqrt{s}}{2} a_1(t) \exp(j(\theta_1(t) - \omega t)) G(s(\epsilon - \theta'_1(t)))
+ \frac{\sqrt{s}}{2} a_2(t) \exp(j(\theta_2(t) - \omega t)) G(s(\epsilon - \theta'_2(t))),
\]

The two ridges will be discriminated if
\[ G(s|\theta'_1(t) - \theta'_2(t)|) \ll 1 \text{, i.e., } |\theta'_1(t) - \theta'_2(t)| \geq \Delta \omega / s. \]