Time Frequency Representation

Juan Pablo Bello
MPATE-GE 2623 Music Information Retrieval
New York University
Discrete Signal and Sampling

\[ \hat{x}(t_n) = Q(x(t) \times p(t)) \]
\[ x(t) \]  
Amplitude  
Time

\[ X(\omega) \]  
Magnitude  
Frequency

\[ \hat{x}(t_n) \]  
Amplitude  
Time

\[ \hat{X}(\omega_k) \]  
Magnitude  
Frequency
Aliasing

\[ X(\omega) \]

\[ P(\omega) \]

\[ \hat{X}(\omega_k) \]

\[ \hat{X}(\omega_k) \]

Aliasing
Aliasing

\[ \omega = 2\pi f \]

\[ \omega_k = 2\pi (f_s - f) \]
Discrete Fourier Transform (DFT)

\[ X(\omega_k) \equiv \sum_{n=0}^{N-1} x(t_n) e^{-j\omega_k t_n} \]

\( x = \) input signal
\( t_n = \frac{n}{f_s} = \) discrete time (s), \( n \geq 0 \) is an integer
\( f_s = \) sampling rate (Hz)
\( X = \) spectrum of \( x \)

\( \omega_k = k\Omega = \) discrete frequency (rad/s), \( k \geq 0 \) is an integer
\( \Omega = 2\pi \left( \frac{f_s}{N} \right) = \) frequency sampling interval (rad/s)
\( N = \) number of time/frequency samples

Simple form:
\[ X(k) \equiv \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \]
Discrete Fourier Transform (DFT)

This can also be written as:

$$X(k) = \langle x(n), s_k(n) \rangle$$

Which can be formulated as a matrix multiplication:

$$\begin{bmatrix}
X(0) \\
X(1) \\
\vdots \\
X(N-1)
\end{bmatrix} = \begin{bmatrix}
s_0^*(0) & s_0^*(1) & \cdots & s_0^*(N-1) \\
s_1^*(0) & s_1^*(1) & \cdots & s_1^*(N-1) \\
\vdots & \vdots & \ddots & \vdots \\
s_{N-1}^*(0) & s_{N-1}^*(1) & \cdots & s_{N-1}^*(N-1)
\end{bmatrix} \begin{bmatrix}
x(0) \\
x(1) \\
\vdots \\
x(N-1)
\end{bmatrix}$$
Discrete Fourier Transform (DFT)

where,

\[ s_k(n) = e^{j2\pi nk/N} = \cos(2\pi nk/N) + jsin(2\pi nk/N) \]

is the set of the sampled complex sinusoids with a whole number of periods in \( N \) samples (Smith, 2007).
Discrete Fourier Transform (DFT)

The $N$ resulting $X(k)$ are complex-valued vectors $X_R(k) + jX_I(k)$ such that, $\forall k = 0, 1, \cdots, N - 1$:

$$|X(k)| = \sqrt{X_R^2(k) + X_I^2(k)}$$

$$\angle X = \phi(k) = \tan^{-1} \frac{X_I(k)}{X_R(k)}$$

Furthermore, if $x(n)$ is real-valued, then:

$$X(k) = X^*(N - k)$$

The DFT of an audio signal is half-redundant!
Discrete Fourier Transform (DFT)

Spectrum

Conjugate Pairs
The IDFT and FFT

• The Inverse DFT (IDFT) is defined as:

\[ x(n) \equiv \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi nk/N}, \quad n = 0, 1, \ldots, N - 1 \]

• The DFT needs on the order of \( N^2 \) operations for its computation.

• The Fast Fourier Transform (FFT) is an efficient implementation of the DFT, that only requires on the order of \( N \log_2 N \) operations when \( N \) is a power of 2.

• The FFT is so fast that it can be used to efficiently perform time-domain operations such as convolution.
Spectral Leakage

Whole number of periods

Fractional number of periods

Amplitude

Time

Magnitude

Frequency

N

N

glitch

Leakage
Zero padding

N-long signal

Zero padded to 8xN length

Spectrum

Interpolated Spectrum

Amplitude

Time

Magnitude

Frequency
Windows

- We are effectively using a rectangular window: \( w(n) \)

- Spectrum = convolution of \( X(k) \) and \( W(k) \)

- Ideal window: narrow central lobe; strong attenuation in sidebands

- Figures show Magnitude in dB:

\[
\text{dB}(X) = 20 \times \log_{10}(X)
\]
Windowing

Rectangular

Blackman

Magnitude (dB)

Frequency

Magnitude (dB)

Frequency
Short-Time Fourier Transform (STFT)

$X_n(k) = |X_n(k)|$
Spectrogram
Time vs Frequency Resolution

- Frequency increases upwards.
- Time decreases downwards.

The diagram shows the trade-off between time and frequency resolution, where achieving high frequency resolution requires sacrificing time resolution, and vice versa. The color gradient represents the intensity of the frequency content over time.
The instantaneous frequency for frequency bin $k$ at time instant $mh$ can be defined as (Arfib et al., 2003):

$$ f_{i,k}(m) = \frac{1}{2\pi} \frac{d\phi_k(m)}{dt} = \frac{1}{2\pi} \frac{\Delta \phi_k(m)}{h/f_s} $$

where,

$$ \Delta \phi_k(m) = \Omega_k h + \text{princarg}\left[\phi_k(m) - \phi_k(m - 1) - \Omega_k h\right] $$

and

$$ \text{princarg}(x) = \pi + [(x + \pi) \mod (-2\pi)] $$

wraps the phase to the $(-\pi, \pi]$ range.
Instantaneous Frequency
Instantaneous Frequency
Sinusoidal Modeling

- The signal is approximated as a sum of time-varying sinusoidal components plus a residual:

\[ x(n) \approx \sum_{k=0}^{K} a_k(n)\cos(\phi_k(n)) + e(n) \]

where,

\[ a_k = \text{instantaneous amplitude} \]
\[ \phi_k = \text{instantaneous phase} \]
\[ e(n) = \text{residual (noise)} \]
Peak picking + interpolation

- Sinusoidal components are peak-picked.
- Instantaneous magnitude and phase values are obtained by interpolation.
- Components are tracked over time
Sinusoidal tracking
References


