Appendix: Notes on signal processing

Capturing the Spectrum: Transform analysis: The discrete Fourier transform

A digital speech signal such as the one shown in Fig. 1 is a sequence of numbers.

Fig. 1:

Transform analysis decomposes this sequence of numbers into a weighted sum of other (component) time series. The component time series must be precisely defined. Different transform analyses are based on different definitions of component time series. The most popular transform used is the Fourier Transform. In a Fourier transform, the component time series are complex exponentials. The transform analysis determines the weights of the component time series that comprise the given signal being analyzed.

The complex exponential

The complex exponential is a complex sum of two sinusoids.
\[ e^{in} = \cos \alpha + j \sin \alpha \]
The real part is a cosine function. The imaginary part is a sine function. A complex exponential time series is a complex sum of two time series
\[ e^{iwt} = \cos(wt) + j \sin(wt) \]
Two complex exponentials of different frequencies are “orthogonal” to each other. i.e.
\[ \int_{-\infty}^{\infty} e^{i\omega t} e^{i\beta t} dt = 0 \quad \text{if} \ \alpha \neq \beta \]

Figure 2 shows a set of orthogonal complex exponential time series of the same frequency.

Fig. 2:
A signal such as the one in Fig. 1 is expressed as a sum of several such complex exponential time series, of different frequencies. The number of such time series (and therefore the number of frequencies into which the signal is analyzed) is decided by the algorithm used to obtain the transform. Fig. 3 shows three sets of complex exponential time series.

Fig. 3:

The discrete Fourier transform

Fourier transform of a discrete signal is often called Discrete Fourier Transform, or DFT. In Fig. 3, the coefficients (or weights) A, B, and C, for example, would be obtained by a DFT. The discrete Fourier transform decomposes the signal into the sum of a finite number of complex exponentials. In fact, it decomposes a signal into exactly as many exponentials as there are samples in the signal being analyzed. An aperiodic signal cannot be decomposed into a sum of a finite number of complex exponentials. Or into a sum of any countable set of periodic signals. The discrete Fourier transform actually assumes that the signal being analyzed is exactly one period of an infinitely long signal. In reality, it computes the Fourier spectrum of the infinitely long periodic signal, of which the analyzed data are one period.

Consider the signal in Fig. 4.

Fig. 4:
The discrete Fourier transform of the above signal actually computes the Fourier spectrum of the periodic signal shown in Fig. 5. Note that the spectrum extends from −infinity to +infinity. The period of this signal is 31 samples in this example.

Fig. 5:

The kth point of a Fourier transform is computed as:

\[
X[k] = \sum_{n=0}^{M-1} x[n] e^{-j2\pi kn/M}
\]

\[x[n] \text{ is the n}^{\text{th}} \text{ point in the analyzed data sequence. } X[k] \text{ is the value of the k}^{\text{th}} \text{ point in its Fourier spectrum. } M \text{ is the total number of points in the sequence. Note that the } (M+k)^{\text{th}} \text{ Fourier coefficient is identical to the k}^{\text{th}} \text{ Fourier coefficient}
\]

\[
X[M+k] = \sum_{n=0}^{M-1} x[n] e^{-j2\pi (M+k)n/M} = \sum_{n=0}^{M-1} x[n] e^{j2\pi kn/M} e^{-j2\pi nk/M}
\]

\[= \sum_{n=0}^{M-1} x[n] e^{-j2\pi nk/M} = \sum_{n=0}^{M-1} x[n] e^{j2\pi kn/M} = X[k]
\]

Discrete Fourier transform coefficients are generally complex. \(e^{jn}\) has a real part \(\cos q\) and an imaginary part \(\sin q\)

\[e^{jn} = \cos q + j \sin q\]

As a result, every \(X[k]\) has the form

\[X[k] = X_{\text{real}}[k] + jX_{\text{imaginary}}[k]\]

A magnitude spectrum represents only the magnitude of the Fourier coefficients

\[X_{\text{magnitude}}[k] = \sqrt{X_{\text{real}}[k]^2 + X_{\text{imag}}[k]^2}\]

A power spectrum is the square of the magnitude spectrum

\[X_{\text{power}}[k] = X_{\text{real}}[k]^2 + X_{\text{imag}}[k]^2\]

For speech recognition, we usually use the magnitude or power spectra

A discrete Fourier transform of an M-point sequence will only compute M unique frequency components, i.e. the DFT of an M point sequence will have M points. The M-point DFT represents frequencies in the continuous-time signal that was digitized to obtain the digital signal. The 0th point in the DFT represents 0Hz, or the DC component of the signal. The (M-1)th point in the DFT represents (M-
1)/M times the sampling frequency. All DFT points are uniformly spaced on the frequency axis between 0 and the sampling frequency.

Fig. 6 (a) shows a 50 point segment of a decaying sine wave sampled at 8000 Hz. The corresponding 50 point magnitude DFT is shown in Fig. 6(b). The 51st point (shown in red) is identical to the 1st point.

Fig. 6: (a)

(b)

The Fast Fourier Transform (FFT) is simply a fast algorithm to compute the DFT. It utilizes symmetry in the DFT computation to reduce the total number of arithmetic operations greatly. The time domain signal can be recovered from its DFT as:

\[
x[n] = \frac{1}{M} \sum_{k=0}^{M-1} X[k] e^{j2\pi kn/M}
\]

**Windowing**

The DFT of one period of the sinusoid shown in the Fig. 6 computes the Fourier series of the entire sinusoid from \(-\infty\) to \(+\infty\).

Fig. 6: (a) a sinusoid; (b) one period of the sinusoid; (c) DFT of (b)

(a)

(b)
The DFT of any sequence computes the Fourier series for an infinite repetition of that sequence.

The DFT of a partial segment of a sinusoid (Fig. 7) computes the Fourier series of an infinite repetition of that segment, and not of the entire sinusoid. This will not give us the DFT of the sinusoid itself!

Fig. 7: (a) Partial segment of a sinusoid; (b) corresponding infinite periodic signal; (c) DFT of (b); (d) DFT of the “correct” sinusoid
The difference between Fig. 7 (c) and Fig. 7 (d) occurs due to two reasons: The transform cannot know what the signal actually looks like outside the observed window. Rather, it infers what happens outside the observed window from what happens inside. As a result, a signal such as Fig. 8 cannot be inferred.

Fig. 8: The transform cannot infer the signal outside the seen window as such. It infers the signal shown in 7(b) instead.

The implicit repetition of the observed signal introduces large discontinuities at the points of repetition. These are shown encircled in green in Fig. 8. This distorts even our measurement of what happens at the boundaries of what has been reliably observed. The actual signal (whatever it is) is unlikely to have such discontinuities.

Fig. 8: discontinuities at the points of replication in the signal inferred by the transform

While we can never know what the signal looks like outside the window, we can try to minimize the discontinuities at the boundaries. We do this by multiplying the signal with a window function, as in Fig. 9(a). We call this procedure windowing. We refer to the resulting signal as a “windowed” signal.

Fig. 9: (a) windowing; (b) change in the central regions of the selected segment due to windowing; (c) inferred windowed signal

(a)
Windowing attempts to keep the windowed signal similar to the original in the central regions, as shown in Fig. 9(b), and reduce or eliminate the discontinuities in the implicit periodic signal, as in Fig. 9(c).

The DFT of the windowed signal shown in Fig. 10(a) is shown in Fig. 10(b). It does not have any artefacts introduced by discontinuities in the signal. Often it is also a more faithful reproduction of the DFT of the complete signal whose segment we have analyzed.

Fig. 10: (a) a windowed signal; (b) magnitude spectrum of the windowed signal in (a)
Fig. 11 summarizes the advantages of windowing in terms of the changes achieved in the signal spectrum:

Fig. 11: (a) Magnitude spectrum of original segment; (b) Magnitude spectrum of windowed signal; (c) Magnitude spectrum of complete sine wave

As we see in Fig. 9, Windowing is not a perfect solution. The original (unwindowed) segment is identical to the original (complete) signal within the segment. The windowed segment is often not identical to the complete signal anywhere. Several windowing functions have been proposed that strike different tradeoffs between the fidelity in the central regions and the smoothing at the boundaries. Fig. 9(a) uses a Hamming window. This is one of a class of windows called cosine windows. Some cosine windows are:

(In the following, window length is M, Index begins at 0)

Hamming: \( w[n] = 0.54 - 0.46 \cos(2\pi n / M) \)

Hanning: \( w[n] = 0.5 - 0.5 \cos(2\pi n / M) \)
Blackman: $0.42 - 0.5 \cos(2\pi n/M) + 0.08 \cos(4\pi n/M)$

Geometric windows are another category of common windows. Some of these are shown in Fig. 12.

Fig. 12: Geometric windows: (a) Rectangular (boxcar); (b) Triangular (Bartlett); (c) Trapezoid

---

**Zero Padding**

We can pad zeros to the end of a signal to make it a desired length. This is useful if the FFT (or any other algorithm we use) requires signals of a specified length. (one example is a radix-2 FFT computation algorithm : it requires signals of length $2^n$, where $n$ is a natural number). The consequence of zero padding is to change the periodic signal whose Fourier spectrum is being computed by the DFT. Zero padding is shown in Fig. 13, which shows a zero-padded signal and its DFT. The DFT of the zero padded signal in Fig. 13(b) is essentially the same as the DFT of the unpadded signal, with additional spectral samples inserted in between. It does not contain any additional information over the original DFT. It also does not contain less information.

Fig. 13: an example of a zero-padded signal; (a) the signal; (b) its DFT

(a) [Image]

(b) [Image]
Fig. 14 further illustrates the consequences of zero padding.

Fig. 14: The left panels show the signals, and the right panels show the magnitude spectra. The effects of windowing are not the same as the effects of zero padding. The DFT of the zero padded signal is essentially the same as the DFT of the unpadded signal, with additional spectral samples inserted in between. It does not contain any additional information over the original DFT. It also does not contain less information.

Fig. 15 illustrates the special case of zero padding windowed signals. While windowing results in signals that appear to be less discontinuous at the edges, the “regularization” of the signal is only illusory. We also do not introduce any new information into the signal by merely padding it with zeros.
Fig. 15: (a) zero-padded signal (b) signal as perceived by the transform (c) magnitude spectrum of the signal

Other examples of magnitude spectra are shown in Fig. 16.

Fig. 16: Left panels show the signals and the right panels show the corresponding magnitude spectra.
Number of points in a DFT

Fig. 17(a) shows 128 samples from a speech signal sampled at 16000 Hz. Fig. 17(b) and 17(c) show the first 65 points of a 128 point DFT, and the first 513 points of a 1024 point DFT respectively. The magnitude spectrum is are more detailed in 17(c).

Fig. 17:
(a)
(b)