Big Idea (Week 5): Time-Frequency Representation and Transformation

There are many different ways to represent information. For example, we might represent time as (1) (year, date, month, and hours) in Universal Time, (2) elapsed seconds since midnight (UT), January 1, 1970, (3) years into the reign of the current monarch, days since the last equinox, and hours since sunrise or sunset, or (4) days, hours, and minutes to graduation. Which one is most convenient depends on what we want to do with the information. Some representations may be more natural for data collection, others may be more convenient for certain kinds of processing, and some representations may expose more structure in the data allowing greater compression or understanding.

In the case of sound, samples in time are a very direct way to capture and reproduce acoustic pressure waves. We can take these samples directly by converting analog voltages from a microphone to digital samples, and we can use them directly to drive a speaker by converting the digital samples back to voltages. However, time samples are not the most efficient way to represent the common structure that often occurs in the patterns of acoustic pressure waves that humans call "sound" because of the manner in which those patterns are perceived. For example, arrays of time samples do not offer the most directly compressible representation. For many operations it is more convenient to describe, work with, and store sounds using a representation based upon frequency content rather than time varying amplitudes. Musical notes, for example, can be represented directly by their frequencies, with one or a few frequencies often defining the sound wave exactly for a period of a very large number of time samples.

As an alternative to representing a sound sequence as a discrete set of amplitude samples in time, we can also represent sound as a weighted sum of "harmonics"— discrete sine and cosine waves whose frequencies are multiples of a base frequency, the "fundamental". That is, rather than representing a sound wave as a vector of time samples:

$$\{Sample[0], Sample[1], \ldots, Sample[N]\},$$

we can represent it as a vector of coefficients: $\{a_{c0}, a_{s1}, a_{c1}, a_{s2}, a_{c2}, ...a_{s_{N/2}}, a_{c_{N/2}}\}$ chosen such that:

$$\begin{aligned} Sample[k] &\approx a_{c0}\cos\left(0\cdot k\cdot\frac{2\pi}{N}\right) + a_{s1}\sin\left(k\cdot\frac{2\pi}{N}\right) + a_{c1}\cos\left(k\cdot\frac{2\pi}{N}\right) + a_{s2}\sin\left(2\cdot k\cdot\frac{2\pi}{N}\right) + ... \\ &\approx \sum_{n=0}^{n=N/2}\left(a_{cn}\cos\left(\frac{2\pi n\times k}{N}\right) + a_{sn}\sin\left(\frac{2\pi n\times k}{N}\right)\right) \end{aligned}$$

The relationship between time and frequency representations for sampled signals can be geometrically interpreted as a linear change of basis between different frames of reference in the (typically high dimensional) vector space comprised of all possible time samples. This brings the powerful and highly efficient computational machinery of linear algebra to bear on the problem of audio signal processing. This means we can convert between the time representation and the frequency representation by simply evaluating a dot product (shown above for frequency to time, below for time to frequency).

$$ac_n = \left(\frac{2}{N}\right)\sum_{k=0}^{k=N} Sample[k]\cdot\cos\left(\frac{2\pi n\times k}{N}\right) \qquad as_n = \left(\frac{2}{N}\right)\sum_{k=0}^{k=N} Sample[k]\cdot\sin\left(\frac{2\pi n\times k}{N}\right)$$