**Music Informatics**

1. Music Representations
   1. Sheet music representations
   2. Symbolic representations
   3. Audio representations
2. Signal
   1. A signal is a function that conveys information about the state or behaviour of a physical system.
   2. Example: time-varying sound pressure, motion of a particle, distribution of light, sequence of images...
   3. Here, we consider audio signals, which depict the amplitude of air pressure over time.
   4. Two different types of signals: analog and digital
3. Analog Signals
   1. Analog Signal: An analog signal is a function f : R → R which assigns an amplitude value f(t) contained within R to each time point t contained within R.
   2. Periodic Signal: A signal f is called periodic with period lambda contained within R > 0 if f(t) = f(t + lambda) holds for all t contained within R.
   3. Sinusoid: A periodic function defined by f(t) := A sin(2 pi (omega \* t − phi)). A: amplitude; omega: frequency; phi: phase
4. Digital Signals
   1. Computers can only store and process a finite number of values, therefore we need to convert waveforms into a discrete representation (“digitisation”).
   2. Digitisation typically involves two steps: sampling and quantisation.
   3. Sampling:
      1. The process of reducing a continuous-time signal to a discrete-time signal: x(n) := f(n · T) , where T is the sampling period and n contained within Z.
   4. Sampling is a lossy operation, in that the original analog signal cannot be recovered from its sampled version.
   5. Sampling might cause an effect known as aliasing, where certain frequency components of the signal become indistinguishable.
5. Quantisation:
   1. Replacing the continuous range of possible amplitudes by a discrete range of possible values (typically rounding off to some unit of precision).
   2. The difference between the actual analog value and the quantised value is called the quantisation error.
6. Intermission:
   1. Complex numbers:
      1. Complex numbers extend real numbers by introducing the imaginary number i := sqrt(-1) with the property i^2 = −1.
      2. Each complex number can be written as c = a + ib, where a is the real part and b the imaginary part of c.
      3. The set of all complex numbers is written as C, and can be represented using polar coordinates: |c| = sqrt(a^2 + b^2);
7. Fourier Transform Idea:
   1. Decompose a given signal into a superposition of sinusoids (elementary signals).
   2. Interpretation: The magnitude A reflects the intensity at which the sinusoid of frequency omega appears in the signal.
   3. The phase phi reflects how the sinusoid has to be shifted to best correlate with the signal.
8. The Role of the Phase
   1. The degree of similarity between the signal and a sinusoid of fixed frequency crucially depends on the phase.
9. Computing similarity with integrals
   1. Assume two functions f and g - what does it mean for f and g to be similar?
   2. The joint behaviour of these functions can be captured by forming the integral of the product of the two functions:  
        
      integral f(t) \* g(t) dt
10. Fourier transform:
    1. c\_w = hat(f)(w) = integral f(t) exp(-2 pi \* i \* omega \* t): The values c\_w are called the Fourier coefficients.
    2. It can be seen that:  
         
       hat(f)(w) = integral f(t) \* cos(-2 pi \* omega \* t)dt + I \* integral f(t) \* cos(-2 pi \* omega \* t)  
         
       The absolute value |hat(f)(omega)| is also called the magnitude of the Fourier coefficient.
11. Fourier representation
    1. The original signal can be reconstructed from its Fourier transform:  
         
       f(t) = integral c\_w \* exp(2 pi \* i \* omega \* t) d omega
12. Fourier transform examples
    1. The Fourier transform tells which frequencies occur, but does not tell when the frequencies occur.
13. Discrete Fourier Transform
    1. Making the Fourier transform work with discrete signals x(n) = f(n \* T), where x(n): sample; Fs = 1/T: sampling rate
    2. Sampling theorem:
       1. The original analog signal f can be reconstructed perfectly from its sampled version x if f does not contain any frequencies higher than Omega = F\_s/2 Hz.
    3. (if it does, sampling would cause aliasing) In the following, we assume that f is bandlimited and that f has a finite duration.
    4. Assume a discrete signal x(n) with relevant samples x(0), x(1), . . . , x(N − 1).
    5. Discrete Fourier Transform (DFT):
       1. X(k) = sum of (x(n) \* exp((-2 \* pi \* I \* k \* n) / N))
    6. Linking indices k of X(k) with physical frequencies:  
         
       F\_coef(k) = k \* F\_s / N
    7. Linking the DFT with the Fourier transform:  
         
       X(k) = 1/T \* hat(f) \* ((k/N) / (1/T))  
         
       Given that X(k) = X(N − k), coefficients X(k) are redundant for k = ⌊ N/2 ⌋ +1, . . . , N − 1
14. Fast Fourier Transform
    1. The DFT requires O(N^2) multiplications and additions. Solution: Fast Fourier Transform (FFT) - a recursive algorithm with O(N log N) multiplications and additions.
15. Short-Time Fourier Transform (STFT)
    1. Idea: Consider only a small section of the signal for the spectral analysis → recovery of time information
    2. Formally, we consider a window function; and multiply the original signal with the window function to yield a windowed signal.
    3. However: the STFT reflects not only the properties of the original signal but also those of the window function (length and shape of the window).
    4. As with the Fourier transform, there is an analog and a discrete version of the STFT.
    5. Discrete STFT:  
         
       X(m,k) = sum of (x(n + mH) \* w(n) \* exp((-2\*pi\*i\*k\*n) / N))  
         
       where w is a window function of length N, m denotes the time frame, and H is the hop size.
    6. So, X (m, k) denotes the kth Fourier coefficient for the mth time frame.
    7. For each time frame m, one obtains a spectral vector which can be computed with the FFT.
    8. Linking STFT coefficients with time positions:  
         
       T\_coef(m) = (m\*H) / F\_s [sec]
    9. Linking STFT coefficients with physical frequencies:  
         
       F\_coef(m) = (k\*F\_s) / N [Hz]
    10. A common choice for hop size is H = N/2, as a trade-off between temporal resolution and data volume.
16. Spectrogram Spectrogram
    1. A two-dimensional representation of the squared magnitude of the STFT:
       1. Y(m, k) := |X (m, k)|^2
    2. Time-Frequency Localisation
       1. Size of window constitutes a trade-off between time localisation and frequency localisation:
          1. Large window: poor time localisation, good frequency localisation
          2. Small window: good time localisation, poor frequency localisation
    3. A variant of the Heisenberg Uncertainty Principle states that there is no window function that simultaneously localizes in time and frequency with arbitrary precision.
17. Log-frequency Spectrogram
    1. Notes of the equal-tempered scale depend on their center frequencies in a logarithmic fashion:   
         
       F\_pitch(p) = 2^{(p−69)/12} · 440
    2. This logarithmic perception of frequency motivates the use of a representation with a logarithmic frequency axis:  
         
       y\_LF(n,p) = sum of (|X(n,k)|^2)
18. Constant-Q Transform
    1. STFT drawback: linearly spaced time/frequency bins → constant time-frequency resolution
    2. Constant-Q Transform (CQT): time-frequency representation where:
       1. Frequencies are logarithmically spaced
       2. Bins have a constant ratio of center frequencies over bandwidth (Q-factor)
    3. In effect, that means that the frequency resolution is better for low frequencies and the time resolution is better for high frequencies Also: musically and perceptually motivated representation
    4. CQT drawbacks:
       1. CQT is computationally more intensive than the STFT/FFT CQT produces a data structure that is more difficult to handle than a linear frequency spectrogram
19. Mel scale
    1. Perceptual scale of pitches judged by listeners to be equal in distance from one another:   
         
       m = 1000 log\_2 ( 1 + f/1000) or m = 2595 log\_10( 1 + f/700)
    2. as proposed by Fant (1973) and O’Shaughnessy (1987), respectively.
20. Mel spectrogram:
    1. mapping an STFT spectrogram onto the mel scale. → commonly used for deep learning-based MIR systems Other commonly used psychoacoustic pitch scales: Bark scale, Equivalent Rectangular Bandwidth (ERB) scale.