Big Idea (Week 6): Pychoacoustic Compression

MP3 encoding at a low bitrate demands carefully engineering to allocate the limited bits available to describe the audio stream. A good encoder will use a psychoacoustical model to drive an intelligent search to determine where to best allocate bits to minimize the subjective noise introduced.

MP3 encoding starts by breaking the original sequence of amplitude samples up into frames representing a window of time (similar to a frame in a movie), such as 0.026 seconds. It then transforms the amplitude samples in the frame into a collection of frequency coefficients. Since we cannot typically code the frequency components at full precision and meet a target bitrate, we must then determine how to encode them.

To save space we can consider encoding options such as:

- adjusting the quantization level for a frequency
- omitting a frequency
- applying lossless compression to compactly encode frequent values

We know the human ear has different sensitivity to different frequencies. MP3 encoding controls the quantization level for frequency coefficients at the level of the critical band—*i.e.*, a single quantization level is used for the set of frequencies in each critical band. Furthermore, we know that masking phenomena occur within critical bands, suggesting we may only need to keep the must prominent frequencies within each band. Our psychoacoustical model can be used to estimate the subjective noise introduced by a particular choice (*e.g.* omission or quantization of frequency), and a trial encoding will tell us the number of encoded bits. Together, this gives us:

1. a palette of adjustments we can make in representing each frame
2. a model of their impact on perceived sound quality
3. a way to compute the bits required for a particular sound frame and set of adjustments

With this formulation, encoding reduces to the challenge of selecting adjustments to maximize perceived sound quality within the allocated number of bits per frame. This demands an effective and efficient algorithm for navigating the space of adjustments. At one extreme we might search all possible adjustments to definitively select the optimal adjustment set. However, the search space is too large for this to be practical. At the other extreme we might try to create an algorithm that computes the adjustments directly. However, the space of choices does not appear well structured enough for this to consistently produce good results. A typical compromise is to explore the adjustments and iteratively refine them until the bitrate is met (*e.g.*, if the bitrate is too high, try to identify the next adjustment that reduces the bitrate while introducing the least subjective noise). In any case, the best set of adjustments will vary from frame-to-frame, adapting to changes in the frequency content of the sound.