**The Psychoacoustic Phenomenon of Auditory Masking**

Adar Guy

*University of Victoria*

**Definitions SLIDE 1**

* Masking is the process by which the detection threshold of a signal is increased by the presence of a stronger signal (masker).
* A ‘not-masked’ threshold is the quietest level of the signal which can be perceived in quiet
* A masked threshold is the quietest level of a signal that can be perceived when accompanied by noise
* The ‘amount of masking’ is the difference between the two, and is defined as the increase (in dB) in the detection threshold of a signal due to the presence of a masker
  + a non-masked threshold of 20dB and masked threshold of 15 dB means an amount of masking of 5 dB

**Types of Masking SLIDE 2**

* Light blue is the masking signal, and the dark blue is the masked signal over time

**Simultaneous Masking SLIDE 3**

* Simultaneous masking occurs in the frequency domain and is a swamping or suppression of neural responses
* This is the more intuitive masking phenomenon where a low level signal like small-band noise can be made inaudible by a simultaneously occurring stronger signal
* Masking thresholds depend on sound pressure level and the characteristics of both the masked and masking signals
  + Without a masker a signal is inaudible if its SPL is below the threshold of ‘quiet’
  + The slope of the masking threshold is steeper towards lower frequencies meaning that higher frequencies are more easily masked

**Simultaneous Masking SLIDE 4**

* Fletcher believed that the auditory system behaved like a series of overlapping band pass filters called auditory filters. He stated that masking thresholds occur when the acoustic power of the signal at the filter output is proportional to the acoustic power of the masker at the filter output. The shape of the filter can be determined by looking at the shapes of psychoacoustic tuning curves
* The amount of masking increases with increasing masking energy that gets through the filter up to the bandwidth of the filter and then any increase in noise energy (bandwidth) does not effect the masking energy
* The most efficient mask-filter ratio is the second one called an equivalent rectangular bandwidth (ERB) which has an equal height and total area to the filter (same energy) but different shape

**Forward Masking SLIDE 5**

* It is a time domain phenomenon that is not as easily explained as simultaneous masking
* It is associated with 4 neurophysiological properties that exist in the auditory system
  + Central Inhibition is the result of the interaction of processes at the level of the neuron, and nerve center. In this case it is a reorganization of the neural structure that exists in a state of excitation or delayed inhibition.
  + The response of the Basilar Membrane does not end immediately after the offset of the stimulus, instead persisting over a period of time (fig on slide)
  + ­The neural response is similar to the basilar membrane as it also continues for a short period (fig on slide)
  + Nerve Fiber Adaptation causes a signal to invoke a higher nerve response if the signal is more isolated in time. This means that a signal that is close to another (masking) signal will cause a lesser nervous response than that of a more separated (isolated) signal to the masker

**Backward Masking SLIDE 6**

* It is the other time domain phenomenon that has very little information about it
* But it has some very important roles in production for hiding unwanted artifacts like pre-echoes
* It along with post masking are used to develop ISO/MPEG audio coding algorithms

**MPEG Audio Compression SLIDE 7**

* Lossy audio data compression like MPEG removes sounds that are already masked anyway.
* Some general rules for masking are established:
  + A lower tone can effectively mask a higher tone (but the opposite is not true)
  + The greater energy that the masking signal has, the wider its influence is, which means that it has a broader range of frequencies that it can mask (but as a consequence, if two tones are widely separated in frequency then little masking will occur)
  + The greater the difference between masking signal energy and masked signal energy, the larger stimulation will occur and it will take longer for the energy to dissipate (persistence) – and the opposite is true and important as well
* MPEG applies a filter bank to the input to break it up into its frequency components as a psychoacoustic model is implemented to the data for bit allocation. The number of bits allocated are used to quantize info from the filter bank which provides the compression
* MPEG defines 3 layers for audio where the basic model is the same of the complexity of the codec increases with every layer
  + In the first layer there is only 1 frame in a DCT (discrete cosine transform) filter. Equal frequency is spread per band and the psychoacoustic model only uses simultaneous frequency masking
  + The second layer consists of 3 frames in the filter (before, current, next) developed by the psychoacoustic model which uses both pre and post masking
  + The third layer is MP3 and it uses a good critical band filter which accounts non-equal frequencies and takes into account stereo redundancy. It uses a Huffman coder (prefix frequency based mapping algorithm) and a psychoacoustic model that brings together both temporal and frequency masking

Some things I didn’t get a chance to speak about are frequency selectivity and auditory scene analysis (heuristic processes of the brain to determine what separate sounds are being heard and their properties (space, band), as well as the anatomy of the ear where masking takes place