# 3 Application of Psychoacoustic Principles: ISO 11172-3 (MPEG-1) PSYCHOACOUSTIC MODEL 1

- It is useful to consider an example of how the psychoacoustic principles described thus far are applied in actual coding algorithms. The ISO/IEC 11172-3 (MPEG-1, layer 1) psychoacoustic model 1 determines the maximum allowable quantization noise energy in each critical band such that quantization noise remains inaudible.
- In one of its modes, the model uses a 512-point DFT for high resolution spectral analysis (86.13 Hz), then estimates for each input frame individual simultaneous masking thresholds due to the presence of tone-like and noise-like maskers in the signal spectrum. A global masking threshold is then estimated for a



subset of the original 256 frequency bins by (power) additive combination of the tonal and non-tonal individual masking thresholds.

- This section describes the step-by-step model operations. The five steps leading to computation of global masking thresholds are as follows:

1. Spectral Analysis and SPL (Sound Pressure Level) Normalization
2. Identification of Tonal and Noise Maskers
3. Decimation and Reorganization of Maskers
4. Calculation of Individual Masking Thresholds
5. Calculation of Global Masking Thresholds



## 3.1 Spectral Analysis and SPL Normalization

First, incoming audio samples of $b$ bit integer, $s(n)$, are normalized according to the FFT length, $N$ , and the number of bits per sample (signed integer), $b$, using the relation

$$x(n) = \frac{s(n)}{N\,(2^{b-1})}$$

Normalization references the power spectrum to a 0-dB maximum.

The normalized input, $x(n)$, is then segmented into 12 ms frames (512 samples) using a 1/16th overlapped Hann window such that each frame contains 10.9 ms of new data. A power spectral density (PSD) estimate, $P(k)$, is then obtained using a 512-point FFT.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi nk}{N}}$$

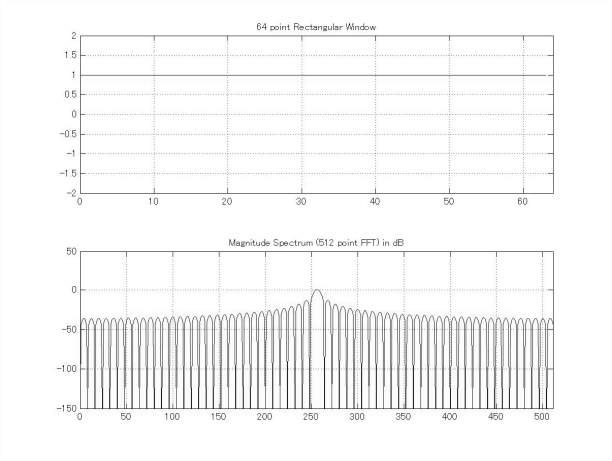


$$X(k) = \sum_{n=0}^{N-1} x(n) w(n) e^{-j\frac{2\pi nk}{N}}.$$

The ~~Hanning window~~ (Hann window) defined by

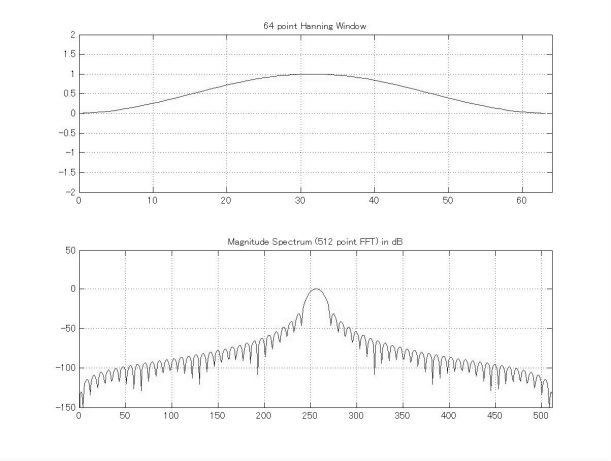
$$w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N}\right)\right]$$

is used to reduce the spectrum leakage from other frequencies to the analysing frequency.




Spectrum of Rectangular (time) Window






Spectrum of the Hanning Window



A power spectral density (PSD) estimate, $P(k)$, is then obtained from $X(k)$ computed by a 512-point FFT (Fast Fourier Transform), a fast algorithm to compute DFT (Discrete Fourier Transform). PSD resulting from 512 FFT has 256 spectral components (harmonics).

$$P(k) = PN + 10\log_{10}|X(k)|^2 \quad \text{for} \quad 0 \le k \le \frac{N}{2}$$

where the power normalization term, $PN$, is the reference sound pressure level of 90.3 dB

> Because playback levels are unknown during psychoacoustic signal analysis, the normalization procedure and the parameter PN are used to estimate SPL conservatively from the input signal [Spanias 2000]



## 3.2 Identification of Tonal and Noise Maskers

After PSD estimation and SPL normalization, tonal and non-tonal masking components are identified.

Tonal maskers

Local maxima in the sample PSD which exceed neighboring components within a certain bark distance by at least 7 dB are classified as tonal. Specifically, the tonal set, $S_T$, is defined as

$$S_T = \left\{ P(k) \text{ such that } \begin{array}{l} P(k) > P(k \pm 1) \\ P(k) > P(k \pm \Delta_k) + 7\text{dB} \end{array} \right\}$$

where,

$$\Delta_k \in \begin{cases} 2 & 2 < k < 63 & \text{0.17-5.5 KHz} \\ (2,3) & 63 \leq k < 127 & \text{5.5-11 KHz} \\ (2,\cdots,6) & 127 \leq k < 256 & \text{11-20 KHz} \end{cases}$$

Tonal maskers, $P_{TM}(k)$, are computed from the spectral peaks listed in $S_T$ as follows

$$P_{TM}(k) = 10\log_{10}\sum_{j=-1}^{+1} 10^{0.1P(k+j)} \quad \text{dB}$$

for each neighborhood maximum, energy from three adjacent spectral components centered at the peak are combined to form a single tonal masker

Noise maskers

A single noise masker for each critical band, $P_{NM}(\bar{k})$, is then computed from (remaining) spectral lines not within the $\pm\Delta_k$



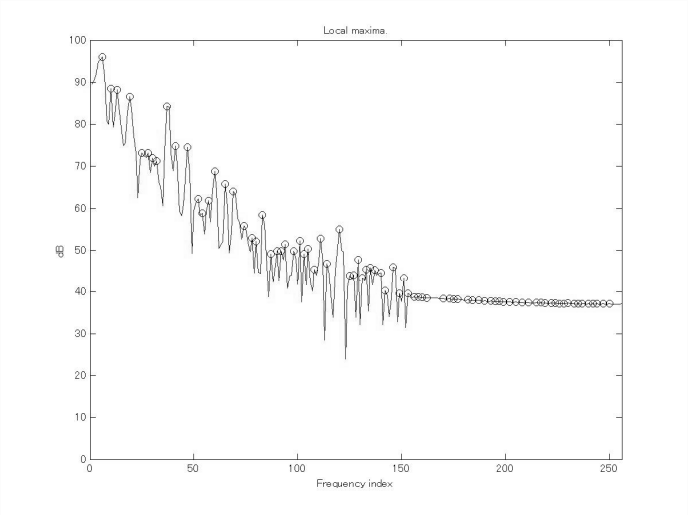
neighborhood of a tonal masker using the sum,

$$P_{NM}(\bar{k}) = 10\log_{10}\sum_{j} 10^{0.1P(j)} \quad \text{dB}$$

for all $P(j)$ not the member of $P_{TM}(k, k\pm1, k\pm\Delta_k)$

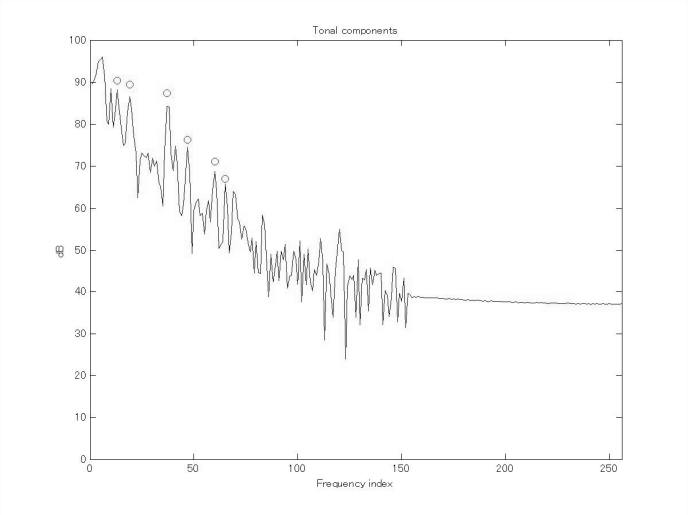
where, $\bar{k} = \left(\prod_{j=l}^{u} j\right)^{\frac{1}{u-l+1}}$ and $l$ and $u$ are the lower and upper spectral line boundaries of the critical band, respectively.



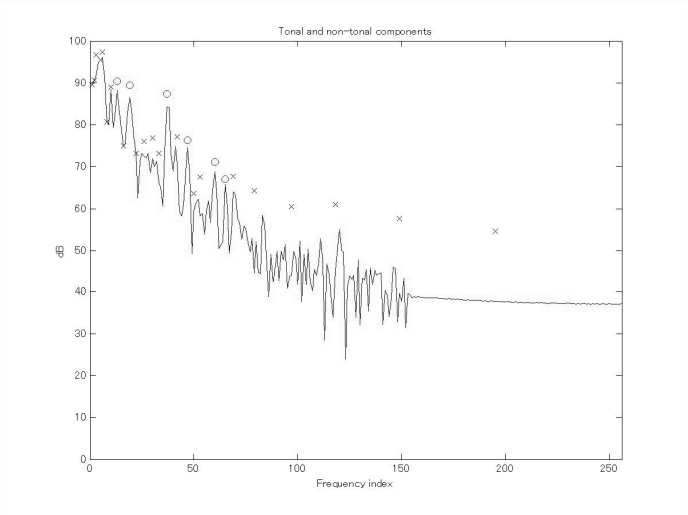



(1) local maxima






(2) tonal components




(3) tonal and non-tonal components of Eine Kleine Nachtmusik



reorganized according to the subsampling scheme,

$$P_{TM,NM}(i) = \begin{cases} P_{TM,NM}(k) & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}$$

The net effect is 2:1 decimation of masker bins in critical bands 18-22 and 4:1 decimation of masker bins in critical bands 22-25 , with no loss of masking components. This procedure reduces the total number of tone and noise masker frequency bins under consideration from 256 to 106. An example of decimation for the equal SPL is shown in the table below.



## 3.3 Decimation and Reorganization of Maskers

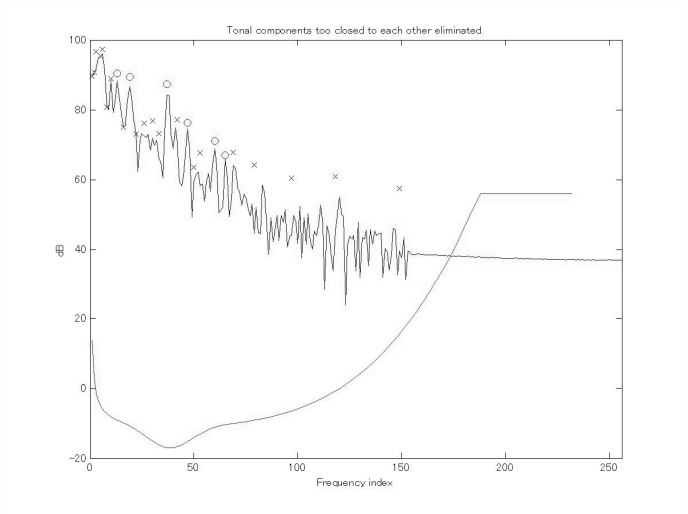
In this step, the number of maskers is reduced using two criteria. First, any tonal or noise maskers below the absolute threshold are discarded, i.e., only maskers which satisfy

$$P_{TM,NM}(k) \geq T_q(k)$$

are retained, where $T_q(k)$ is the SPL of the threshold in quiet at spectral line $k$. Next, a sliding 0.5 Bark-wide window is used to replace any pair of maskers occurring within a distance of 0.5 Bark by the stronger of the two.

After the sliding window procedure, masker frequency bins are






Tonal and non-tonal maskers after decimation. Only one non-tonal masker SPL under the absolute threshold was eliminated.

## 3.4 Calculation of Individual Masking Thresholds

Having obtained a decimated set of tonal and noise maskers, individual tone and noise masking thresholds are computed next. Each individual threshold represents a masking contribution at frequency bin i due to the tone or noise masker located at bin j (reorganized during step 3). Tonal masker thresholds, $T_{TM}(i,j)$, are given by

$$T_{TM}(i,j) = P_{TM}(j) - 0.275z(j) + SF(i,j) - 6.025 \quad \text{dB}$$

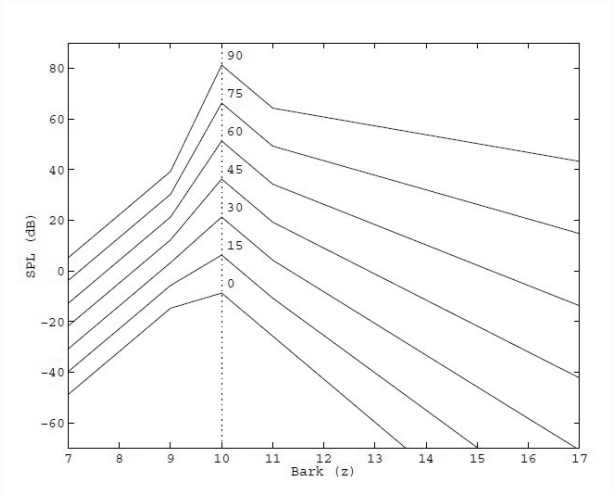
where $P_{TM}(j)$ denotes the SPL of the tonal masker in frequency bin $j$, $z(j)$ denotes the Bark frequency of bin $j$,



and the spread of masking from masker bin $j$ to maskee bin $i$, $SF(i,j)$, is modeled by the expression,

$$SF(i,j) = \begin{cases} 17\Delta_z - 0.4P_{TM}(j) + 11 & -3 \le \Delta_z < -1 \\ (0.4P_{TM}(j) + 6)\Delta_z & -1 \le \Delta_z < 0 \\ -17\Delta_z & 0 \le \Delta_z < 1 \\ (0.15P_{TM}(j) - 17)\Delta_z - 0.15P_{TM}(j) & 1 \le \Delta_z < 8 \end{cases} \quad \text{dB}$$






Prototype spreading functions at $z$=10 as a function of masker level



$SF(i,j)$ is a piecewise linear function of masker level, $P_{TM}(j)$, and Bark maskee-masker separation, $\Delta_z = z(i) - z(j)$. $SF(i,j)$ approximates the basilar spreading (excitation pattern) given. As shown in the figure, the slope of $T_{TM}(i,j)$, decreases with increasing masker level. This is a reflection of psychophysical test results, which have demonstrated that the ear's frequency selectivity decreases as stimulus levels increase. It is also noted here that the spread of masking in this particular model is constrained to a 10-Bark neighborhood for computational efficiency. This simplifying assumption is reasonable given the very low masking levels which occur in the tails of the basilar excitation patterns modeled by $SF(i,j)$.

Individual noise masker thresholds, $T_{NM}(i,j)$, are given by

$$T_{NM}(i,j) = P_{NM}(j) - 0.175z(j) + SF(i,j) - 2.025 \quad \text{dB}$$

where $T_{NM}(i,j)$ denotes the SPL of the noise masker in frequency bin $j$, $z(j)$ denotes the Bark frequency of bin $j$, and $SF(i,j)$ is obtained by replacing $P_{TM}(j)$ with $P_{NM}(j)$.

**Problem**

A subroutine Individual_masking_thresholds.m contained in the MP3 spychoacoustic masking simulation program Matlab_MPEG_1_2_4.zip calculates individaul masking thresholds of tonal maskers $T_{TM}(i,j)$, and non-tonal maskers $T_{NM}(i,j)$ using the spreading function $SF(i,j)$. Apply this program to a music piece in *.wav chosen in the previous **Problem** to plot the individaul masking thresholds of a frame.



## 3.5 Calculation of Global Masking Thresholds

In this step, individual masking thresholds are combined to estimate a global masking threshold for each frequency bin in the subset given by Eq. 3.4. The model assumes that masking effects are additive. The global masking threshold, $T_g(i)$, is therefore obtained by computing the sum,

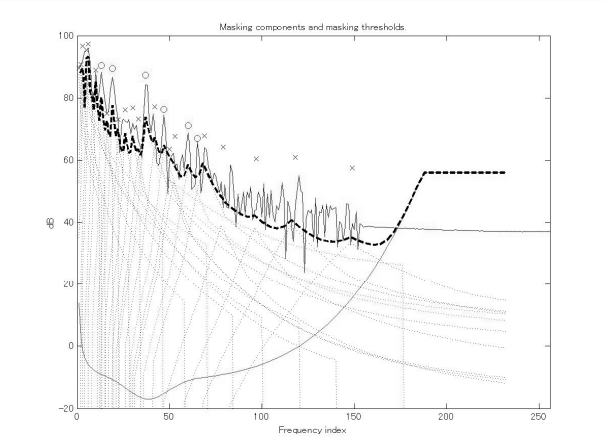
$$T_g(i) = 10\log_{10}\left(10^{0.1T_q(i)} + \sum_{l=1}^{L} 10^{0.1T_{TM}(i,l)} + \sum_{m=1}^{M} 10^{0.1T_{NM}(i,m)}\right) \quad \text{dB}$$

where $T_q(i)$ is the absolute hearing threshold for frequency bin $i$, $T_{TM}(i,l)$ and $T_{NM}(i,m)$ are the individual masking thresholds, and $L$ and $M$ are the number of tonal and noise maskers, respectively, identified previously.



In other words, the global threshold for each frequency bin represents a signal dependent, power additive modification of the absolute threshold due to the basilar spread of all tonal and noise maskers in the signal power spectrum. The next Fig. shows global masking threshold obtained by adding the power of the individual tonal and noise maskers to the absolute threshold in quiet.






Individaul masking thresholds for both tonal and non-tonal maskers. The global masking threshold is the sum of all individual masking thresholds.