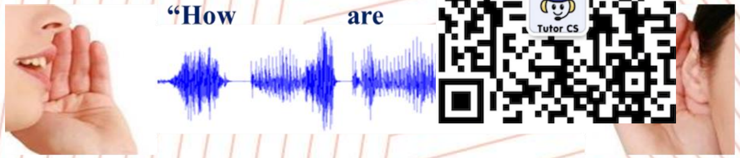
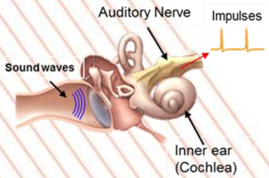
ELEC3104: Mini-Project – Cochlear Signal Processing





TLT – Level 4 (Distinction Level): Incorporate temporal masking properties of the cochlea in the **FIR** filter bank model developed in Level 2 and synthesise the signal.

Complete TLT-Level 3 first and ensure that you are on the right track before proceeding to TLT – Level 4

Prof. Eliathamby Ambikairajah, School of EE&T

Term 3, 2024

Analysis and Synthesis Filter Bank

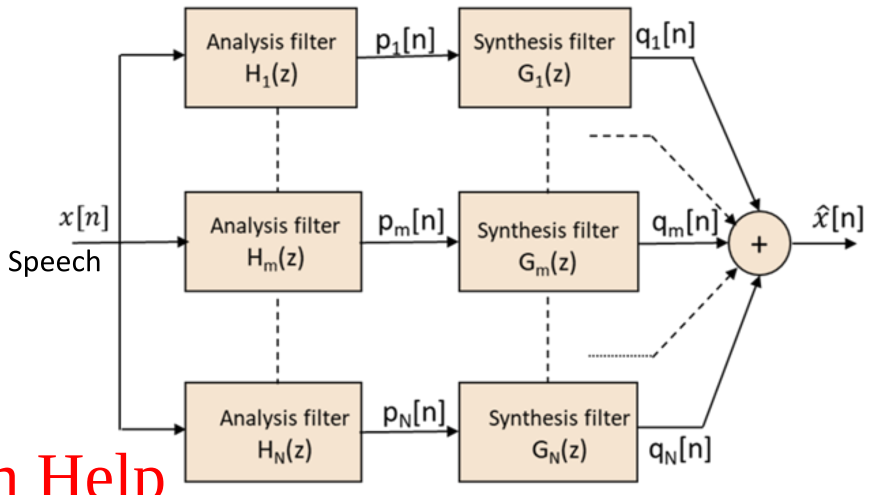
FIR Analysis and Synthesis Filters (Level 2)

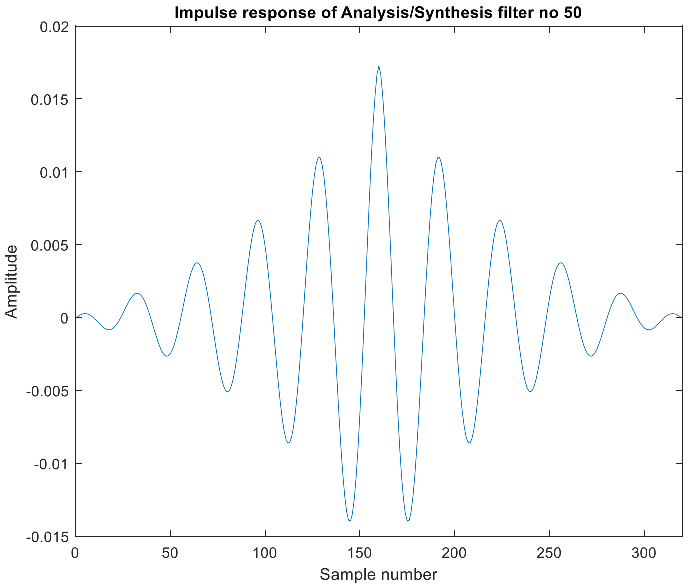
- ✓ In TLT Level 2, we learnt that by using analysis filters and synthesis filters as shown on the right, we can reconstruct the the input signal $x[n]$ by simply summing the Synthesis filter outputs to ob[illegible] $_{=1}\, q_m[n]$. The impulse response of the analysis filter $H_{m(Z)}$, is $h_m[$[illegible]pulse response of the synthesis filter, $G_{m(Z)}$, is $g_m[n] \;=\; h_m[-n]$ [illegible]e-reversed impulse response of the analysis filter.
- ✓ The impulse responses of the synthesis filters are a time-reversed version of the corresponding analysis filters and thus the filters are non-causal. To make the synthesis filters causal, a delay is introduced. That is,

  $g_m[n] \;=\; h_m[L-n]$ for m = 1, 2….N , where L is the delay
- ✓ In our example we use 160 coefficients, so L = 160.
- ✓ See the diagram on the right, here we will use $H_m(z)$ and $G_m(z)$; but note their impulse resoponses are identical . They are obtained by convolving $h_m[n] * gm[n]$which provides a linear phase characteristics and are symmetrical around n = 160 . (Note: $H_1[z] = G_1\,[z]$, … $H_m[z] = G_m\,[z]$, …. $H_N[z] \;=\; G_N\,[z]$)
- ✓ The FIR filter coefficients of all the filters $H_1[z] = G_1\,[z]$, … $H_m[z] = G_m\,[z]$, …. $H_N[z] \;=\; G_N\,[z]$) are symmetrical around n=160. An example is shown on the right.

Note: Reconstructed speech $\hat{x}[n]$should be scaled so that it lies between -1 and +1

Analysis/Synthesis Filter Bank





Auditory masking

Auditory masking

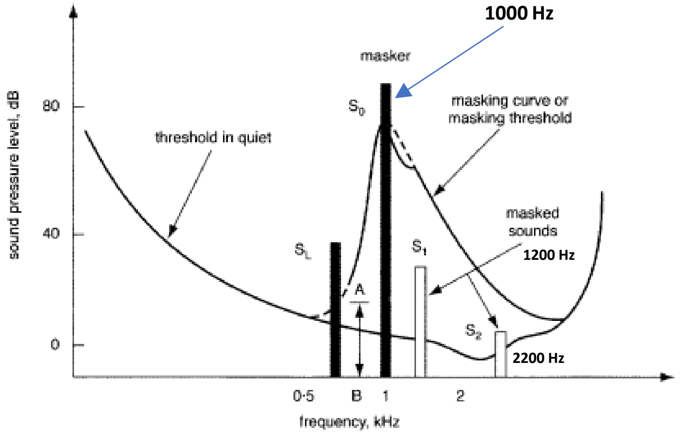
Auditory masking is a phenomenon in which the perception of one sound is affected by the presence of another sound. Essentially a louder sound can "mask" a quieter or softer sound making it diffi[illegible]ible to hear. Auditory masking occurs primarily in two ways:

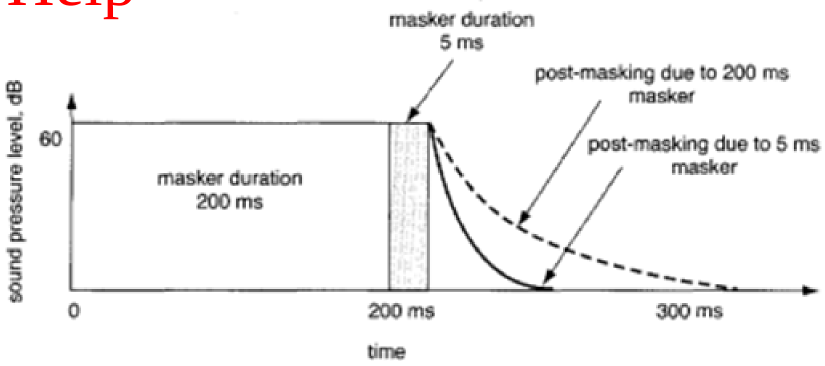
- **Simultaneous Masking (Frequency Ma[illegible]**[illegible]ccurs within a frequency band when two sounds are [illegible]aneously and the presence of a louder sound at a certain frequency makes it harder to hear a weaker sound at a nearby frequency. For example, if a loud sound at 1000 Hz is being played one might not be able to hear a quieter sound at 1200 Hz.

- **Temporal Masking (Time Masking)**: This occurs when a loud sound makes it difficult to hear a softer sound immediately before or after it even if they are at different frequencies. For instance, if a loud sound occurs it can mask a low amplitude sound that appears shortly after it in time, even if that low amplitude sound occurs after the loud sound has stopped.

Note: Auditory masking plays a significant role in audio compression algorithms (such as MP3), which use auditory masking to reduce file size by removing sounds that are likely to be masked and therefore inaudible to the listener.

Simultaneous Masking





Temporal Masking

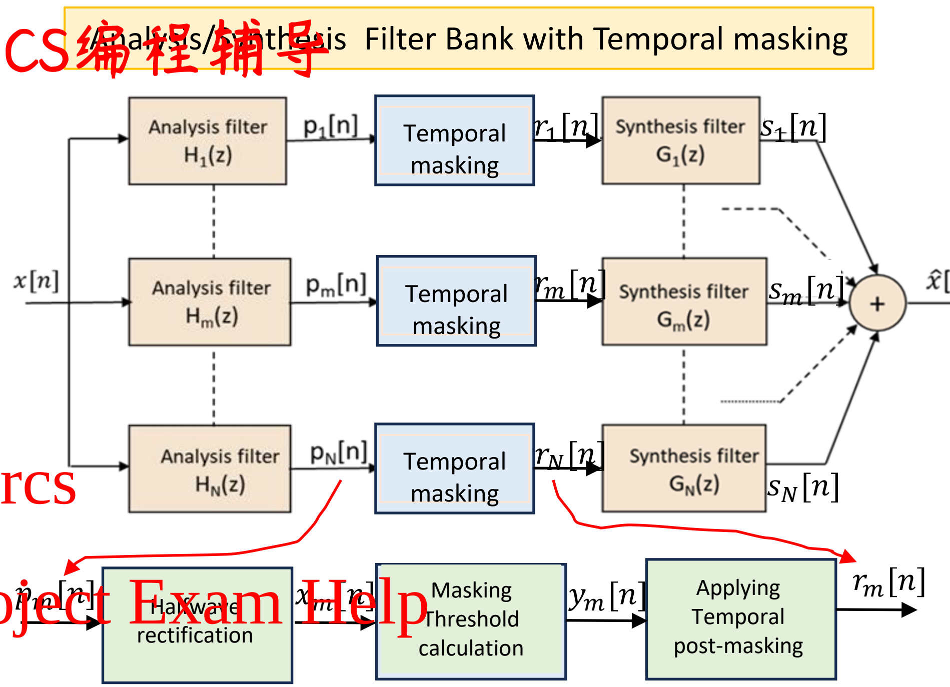
Temporal post-masking

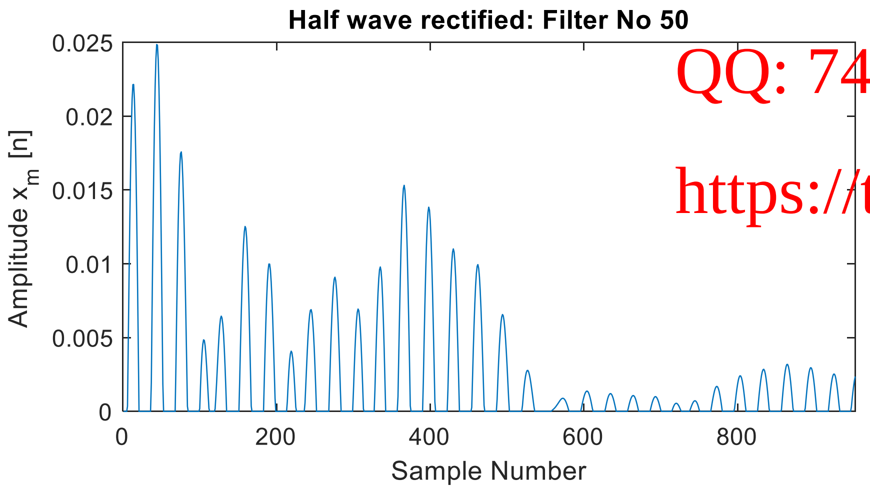
Step1: Temporal post-masking

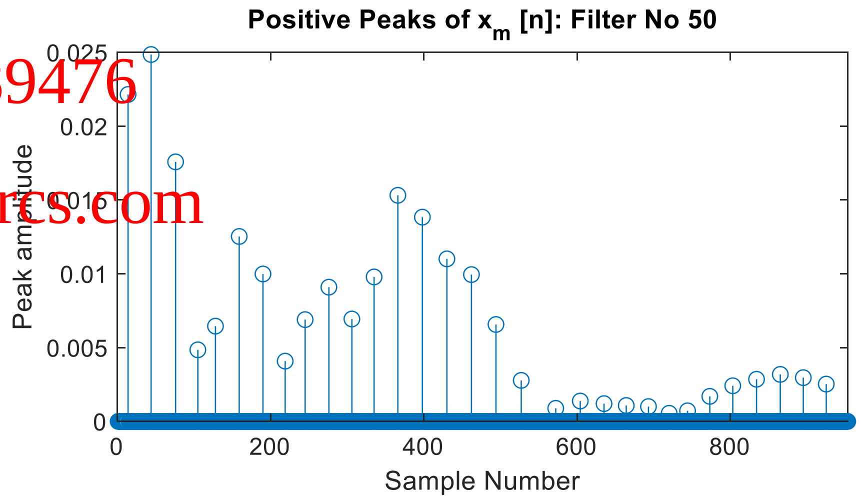
✓ In this project, we will focus solely on temporal post-masking incl. Simultaneous masking will not be considered.

The first step in implementing **temporal post-m** … follows:

- The output of each analysis filter is half-wa … nd **only** the positive peaks of the band signals … d retained. Physically, the half-wave rectification process simulates the action of the inner hair cells, which respond only to the movement of the basilar membrane in one direction. This process results in a series of pulse trains, $x_m[n]$, for each analysis band (see below for one particular band) , where the pulses retain the amplitudes of the corresponding band signals from which they were derived.







Temporal post-masking

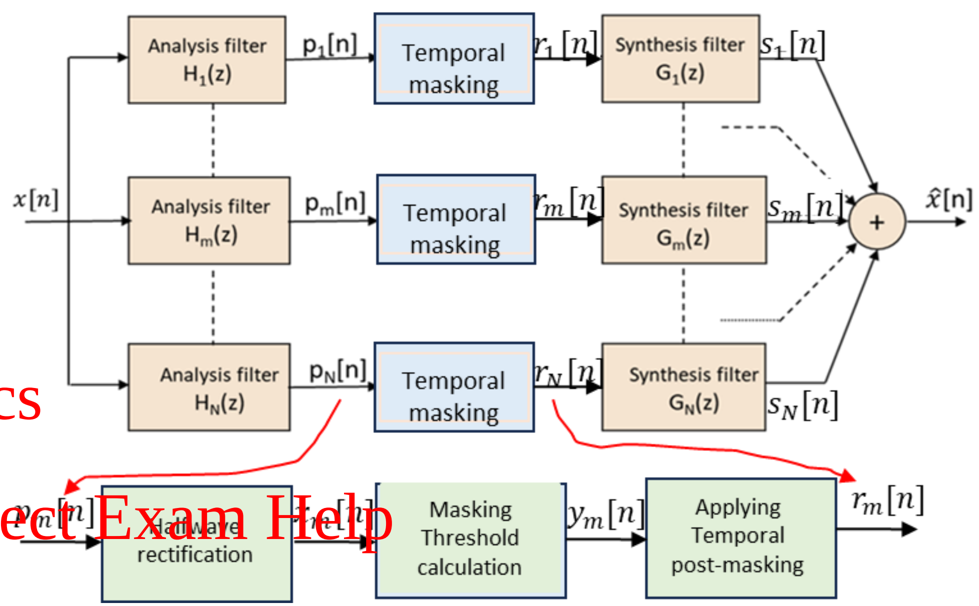
Step2: Temporal post-masking

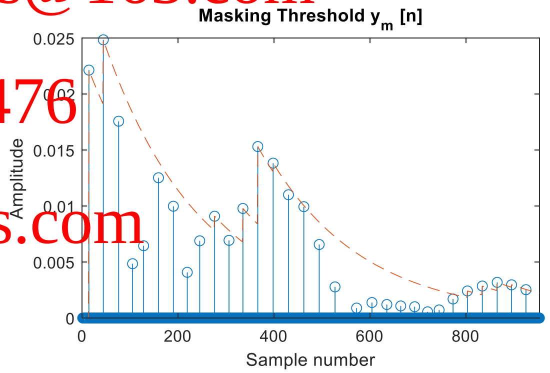
- ✓ A temporal post-masking effect occurs as a result of weaker signal components being rendered inaudible by stronger signal components in the same band that precede the[illegible] second step in implementing **temporal post-mas[illegible]**[illegible]ws:
- The temporal post-masking threshold $y_m[n]$ dec[illegible]ately exponentially following each positive pulse. A cr[illegible]ation to this masking threshold $y_m[n]$ is given by,

$$y_m[n] = \begin{cases} x_m[n] & x_m[n] > c_o\, y_m[n-1] \\ c_o\, y_m[n-1] & otherwise \end{cases}$$

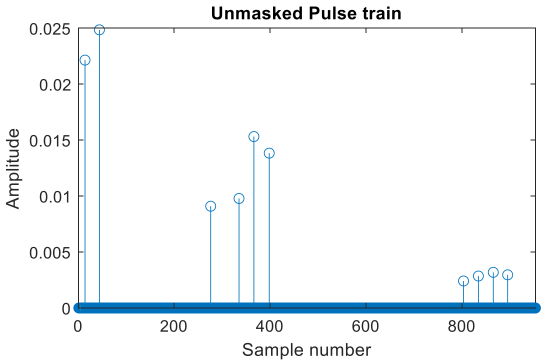
where $x_m[n]$ is the m[th] band pulse train, c_o = exp(−τ) and n is the discrete time sample index. The time constant (τ) which varies as a function of amplitude could be determined empirically by listening the quality of the reconstructed speech $\hat{x}[n]$. For example, τ could be in the region 0.01 (Filter 1) to 0.008 (Filter N).

- Before applying the temporal post-masking, the peaks of the sub-band signals $p_m[n]$ were extracted to form a sub band pulse train $x_m[n]$. All pulses with amplitude less than the masking threshold $y_m[n]$ were discarded, as seen in Figure on the right.
- The resulting sub band masked pulse trains $r_m[n]$ characterise both the spectral envelope and the temporal information in the speech signal. From these pulse trains, speech signal could be resynthesised using the synthesis filter bank $G_1[z]$ *to* $G_m[z]$.





A sub band pulses (solid), and corresponding temporal masking threshold (dashed).

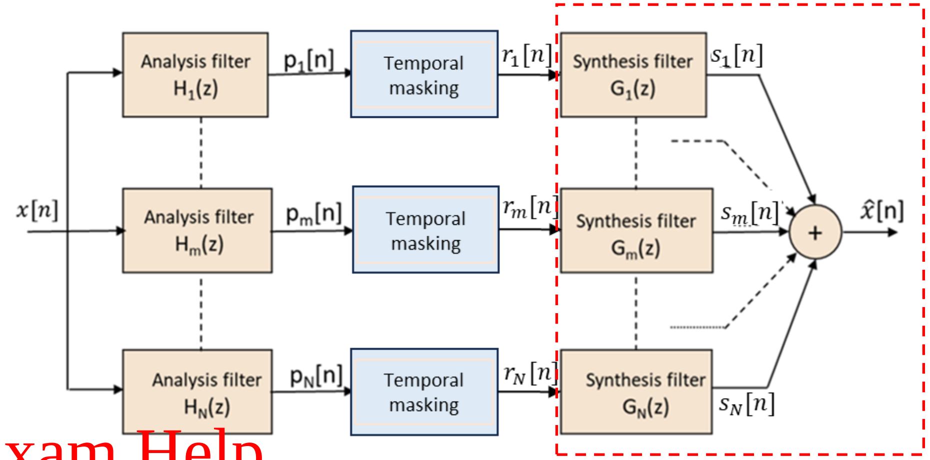


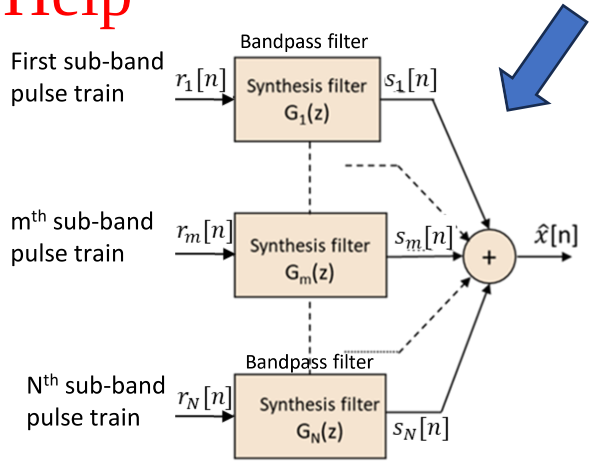
The resulting sub band pulse train after temporal post-masking.

Speech and Audio Synthesis: Final implementation

- ✓ Using MATLAB, filter the speech signal (saved in speech.wav and music.wav from 'Wav files for DSP Labs' on Moodle) sampled at 16 kHz using the analysis filter bank $H_1[z]$ to $H_m[z]$. The outputs $p_m[n]$ of these analysis filters can then be fed as an input to the synthesis filters G [illegible] (**without using Temporal masking module**) which have the same [illegible] impulse responses as the analysis filters. The synthesis fil[illegible] $[n]$ are added ($\sum_{m=1}^{N} s_m[n]$) to produce synthesised speech/mu[illegible]
- ✓ Using the *soundsc* command in MATLAB, **compa**[illegible] of the reconstructed speech/music with that of the original $x[n]$. **Discuss** any noticeable differences. Explain what you notice?
- ✓ Now repeat the above step, but with Temporal masking module introduced. As described earlier the sub band signals can be reconstructed from the pulse trains $\{r_1[n]\ to\ r_N[n]\}$ by means of bandpass filtering $\{\ G_1[z]\ to\ G_m[z]\ \}$
- ✓ Using the *soundsc* command in MATLAB, **compare** the quality of the reconstructed speech/music with that of the original $x[n]$. **Discuss** any noticeable differences and explain your observation? Vary the values of τ and determine which values of τ the reconstructed quality is good in your opinion. Discuss your finding with the lab demonstrator.
- ✓ Finally, calculate the average percentage of pulses removed by the application of temporal masking ensuring the reconstruction quality remains high. First count the number of pulses $x_m[n]$ for m[th] band and then count $r_m[n]$, then compute the reduction. Repeat this process for all bands and then average the results. What percentage of the pulses have been removed by temporal post-masking effect.

Analysis/Synthesis Filter Bank with Temporal masking





Reconstruction of the speech or audio signal from the sub band pulse trains is achieved simply and at low complexity using the filter bank scheme shown above