4 MODULATION FILTERBANK MODEL

A computational model of temporal-envelope processing was used. This model mimics the different stages of the auditory processing in the temporal domain by human listeners. This is the MFB model and it is inspired by the early work of Dau et al. (1997a,b). It was first developed by Wallaert et al. (2017) for analysis of AM detection thresholds and then extended by Wallaert et al. (2018) for analysis of FM detection thresholds. The version of the MFB model used for this study is taked from the code of Wallaert et al. (2018) with some small modifications in order to address the research questions. This model was coded in Matlab.

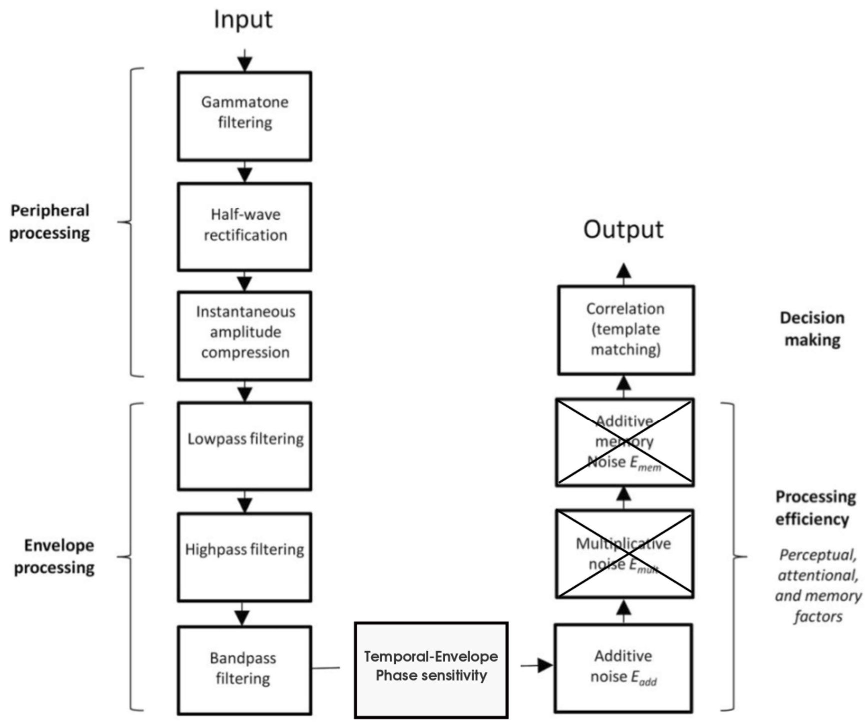

Figure 4.1: Schema of the MFB model. *Credits: Modified from fig.3 Wallaert et al., (2017)*

The MFB model is composed of four main consecutive stages (a schema is presented in figure 4.1):

1. Peripheral processing: It consist on three stages (cochlear filtering, amplitude compression using a power law and half-wave rectification) in which the input signal is modified in such a way that resembles the process that occurs in the peripheral auditory system in humans.

a) To simulate the bandpass filtering of the basilar membrane in the cochlea, five linear gammatone filters were used following Patterson et al. (1995). This means that for one input signal there is 5 different outputs that will be referred to as channels. One filter (on-frequency filter) was centered on the carrier frequency (500 Hz) while the other four filters (off-frequency filters) were centered one and two Cams[1] above and below the carrier frequency.

Cams are units of the ERB number scale (Glasberg and Moore, 1990)

b) The output of the on-frequency filter was processed by a broken-stick function that applied compression only for levels above a given threshold (40 dB SPL). This compression was done using a power law with exponent 0.3. This was the only channel compressed. This crudely simulates the physiological finding that compression in the cochlea occurs mainly for input frequencies close to the characteristic frequency of the place whose response is being measured (Robles and Ruggero, 2001).

c) After compression, all five channels were half-wave rectified by just keeping the positive amplitudes and making the negatives 0.

2. Temporal-envelope processing: It consist on three filtering stages (lowpass, highpass and bandpass filtering) and envelope extraction. All these are applied to the signal to mimic some effects that occur during the processing of the temporal envelope cues evoked by AM and FM signals for humans (e.g; hair cell transduction, adaptation, etc...).

   a) The rectified signals were first lowpass filtered to simulate hair cell transduction. This was done with a 1st order lowpass Butterworth filter with a cutoff 1000 Hz.

   b) Then they were highpass filtered to simulate adaptation (Tchorz and Kollmeier, 1999). This was done with a 1st order highpass Butterworth filter with a cutoff 2.5 Hz.

   c) The resulting signals of each channel were passed through a set of 10 bandpass filters (1st order Butterworth bandpass filters) to simulate frequency-selective processing in the AM domain (Dau et al., 1997a,b). These filters were spaced logarithmically with center frequencies between 2 and 120 Hz (Moore et al., 2009), each with a Q value of 1 and a rolloff of 620 dB/decade (Ewert and Dau, 2000; Ewert et al., 2002; Lorenzi et al., 2001; Sek and Moore, 2002). This produced 50 channels in total (5 gammatone channels x 10 modulation channels) which means that the internal representation of the signal is a matrix of 5x10xK_t dimensions.

This are called the modulation filterbank.

K_t is equal to the number of bins used to sample the time interval

1 Each 1-Cam step represents a distance of about 0.89 mm along the basilar membrane (Moore, 1986).

d) For each modulation filter centered below 10 Hz, the waveform at the output of the filter was passed on for further processing, while for each filter centered at and above 10 Hz only the Hilbert envelope of the output was passed on. This was done to simulate the loss of sensitivity to envelope phase for rates above 10 Hz (Dau et al., 1997a,b). Each envelope of the envelope (the so-called "venelope" Ewert et al., 2002) was scaled so that the root-mean-square value at the output of modulation filters was the same before and after the Hilbert transformation. Before passing them on, both the envelope and the venelope were down-sampled by a factor 10 to improve computational speed.

3. Processing efficiency: It consists on a stage in which the temporal-envelope information extracted at the preceding stage is degraded using Gaussian white noise. This degradation is required to decrease the performance of the model and it represents the loss of modulation information caused by internal noise.

   In this study only one noise was used: additive noise. It was modeled as a Gaussian noise with zero mean and different standard deviations σ_{int} depending on the condition (more details in section 5.4). Using this noise distribution, 5x10xK_t different and independent random values were chosen, one for each component of the internal (matrix) representation of the signal. These values were added to the signal using only one noise value per signal component.

   K_t is equal to the number of bins used to sample the time interval

   The signal at the end of this stage is the final output of all the processing done to a given stimulus. This was the signal used in the final stage corresponding to the detection of the modulation.

4. Decision making (2): This last stage consists in decision through a template-matching strategy. This stage implements the hypothesis that listeners weight more or less appropriately the available modulation cues extracted and represented at the preceding stages to reach a decision in the detection task.

   This stage was realized as a simplified version of the optimal detector described by Dau et al. (1997a,b). First, a template was derived at the beginning of each adaptive staircase as the difference between the model output of the Target and Comparison stimuli (described in chapter 3), both considering $\sigma_{\text{int}} = 0$. Then, in each trial, both complete stimuli were processed by the preceding stages of the model. The template was then correlated (using the Pearson correlation coefficient) with the model output for each stimuli. The decision made by the model (concerning which interval corresponds to the modulated stimulus) is defined as the interval which stimulus lead to the highest correlation.

   The specific details of this stage would depend on the task that is used to detect the modulations. The details provided here are for a 2I, 2AFC adaptive task (all the information about the task used in this study is given in section 5.3).

In order to characterize the effects of variability (through different internal and external noise values) and hearing loss on MDT, some parameters of the model were manipulated. The parameters modified were those controlling for: the variance of the noise added to the output of modulation filters in stage 3 to degrade the internal representation of modulation information (internal noise: σ_{int}), width of cochlear filters (the width of the 5 Gammatone filters used in the stage 1), and amplitude compression (also applied to the signal in stage 1). The last two parameters were associated to hearing loss and they were modified accordingly to what has been reported in the literature and to what is expected to happen with ageing and hearing loss. More details about the manipulation of these parameters is provided in section 5.4.