CHAPTER 7

QUADRILATERAL FILTER BANK STRUCTURE

7.1 INTRODUCTION

This chapter introduces a new quadrilateral shaped Mel scale filter bank structure to better cover the Mel frequency scale, so that speaker dependent features can be faithfully extracted from this filter bank structure. Some of the speaker dependent information which are scattered around the lower frequency range of each filter bin is not given any priority during design of Triangular and Gaussian filter bank structures. Hence, a new filter bank which covers more low frequency range for every filter bin is designed and the resultant filter bank structure is named as Quadrilateral Filter bank.

For designing the filter bank, its center frequency and Equivalent Rectangular Bandwidth (ERB) are essential to determine the lower, upper and intermediate frequencies of every filter bin. Since speech signal contains more information in the low frequency range, there is a need for more number of filter bins to be provided in the low frequency range. This is not possible with uniform filter bank. Hence a non uniform filter bank with quadrilateral shaped filter bins is constructed for speaker identification analysis.

7.2 HUMAN HEARING SYSTEM

The human ear is a complex organ. To make things even more difficult, the information from two ears is combined in a difficult neural network. The hearing system performs a spectrographic analysis of any
auditory stimulus. The cochlea is a liquid filled tube roughly 2 mm in diameter and 3 cm in length. Contained within the cochlea is the basilar membrane, the supporting structure for about 12,000 sensory cells forming the cochlear nerve. The basilar membrane acts as a frequency spectrum analyzer. The cochlea can be regarded as a bank of filters whose outputs are ordered tonotopically, so that a frequency-to-place transformation is effectuated. The filters closest to the cochlear base respond maximally to the highest frequencies and those closest to its apex respond maximally to the lowest frequencies. Tuning is the characteristic of the cochlea that allows the brain to distinguish a low- from a high-pitched sound. A small region of the cochlea will vibrate in response to a high-frequency sound. Tuning is important for the perception of speech since many of the cues humans use to distinguish one word from the other are based on the ability to discriminate different frequencies. The hearing system can also be said to perform a temporal oscillographic analysis of the set of neural signals that originate in the cochlea in response to an auditory stimulus. This process is important for frequencies below 500 Hz and it contributes to frequency resolution up to about 1.5 kHz.

Pitch is our perceptual interpretation of frequency. Ideal human hearing ranges from 20 to 20,000 Hz, yet we have our greatest sensitivity to frequencies which lie within 200 to 2000 Hz, which takes up two-thirds of the distance on the basilar membrane. In order to understand the perception of pitch, it is also necessary to consider the facts that are given by the harmonic structure of many sounds, such as the voiced sounds of speech. In the range of 20 to 2000 Hz, the ear has the ability to fuse harmonically-related frequencies into a single entity with a fundamental frequency, even if the fundamental is missing. The perceived pitch of a sinusoidal tone is not strictly given by its frequency, but it is also marginally influenced by its intensity level: An increase in level produces a slightly more extreme sensation of pitch. Thus,
the pitch of a sound is specified by the frequency of a pure tone whose pitch is judged to be the same as the pitch of that sound.

7.3 AUDITORY PROPERTIES

Auditory scales of frequency representation are required in models of auditory perception. Some of the representations of auditory scales of frequency are given below,

Period is the duration of one cycle in a repeating event. Period is represented with \( p \). It is more directly relevant than frequency in describing the temporal analysis in hearing. Period is measured in second or millisecond. Frequency is the number of occurrences of a repeating event per unit time. Frequency is represented with \( f \). Frequency is preferred over period as the default choice in describing acoustic phenomena. For counts per unit of time, the unit for frequency is hertz (Hz).

\[
\log (p) \text{ and } \log (f):
\]

These logarithmic measures are fully equivalent for most applications, since they have the same absolute value, \( \log (f) = -\log (p) \), as time and frequency have inverse relationship between them. A musical octave scale can be obtained by choosing 2 as the base of log.

7.3.1 Ratio pitch

Humans perceive pitch logarithmically in relations to frequency. Every doubling in Hz is perceived as an equivalent octave. It is thought that because a doublings of frequency causes a response at equal distance on the basilar membrane, humans hear octaves as related. In fact, because of the logarithmic spacing of pitch placement on the membrane it can be extrapolated that we perceive differences in pitches not as differences in
frequency, but as the ratio of pitches separating them. Ratio pitch is measured in Mel. One Mel was defined as one thousandth of the pitch of a 1 kHz tone. It is closely related with the critical-band rate scale.

### 7.3.2 Critical Band rate $z$

Critical band represents the frequency resolution of the auditory system. Psycho-acoustically the critical bandwidth can be measured by masking and loudness comparison methods. Measurement of the classical "Critical Bandwidth" (Zwicker et al 1957 and Traunmüller 1990) typically involves loudness summation experiments. The critical band rate scale differs from Stevens' Mel-scale mainly in that it uses the CB as a natural scale unit. The relation between frequency $f$ and CB-rate $z$ is given by the equations below.

$$\text{Critical band rate } z \text{ (in bark): } z = \left[ \frac{26.81}{1+1960} / f \right] - 0.53 \quad (7.1)$$

$$\text{Inverse: } f = 1960 / \left[ 26.81 / (z + 0.53) - 1 \right] \quad (7.2)$$

Critical bandwidth (in Hz): $B_z = 52548 / (z^2 + 52.56z + 690.39) \quad (7.3)$

where $f$ is in Hz and $z$ is in bark.

### 7.3.3 Equivalent Rectangular Bandwidth (ERB) rate

Measuring the critical bands below 500 Hz seems to be quite difficult, because, at low frequencies, the sensitivity and the efficiency of the auditory system rapidly diminish. More accurate measurements of the auditory-filter bandwidth have lead to the ERB-rate scale. The auditory frequency selectivity can be described in terms of ERB as a function of center frequency (Moore & Glasberg 1983). In general, on this ERB-rate scale the auditory-filter bandwidth, expressed in equivalent rectangular bandwidth (ERB), is smaller than on the Bark scale. The CB and the ERB have been
found to be proportional for center-frequencies above 500 Hz. For lower frequencies, the ERB decreases along with decreasing center-frequency, while the CB remains close to constant.

Equivalent rectangular bandwidth (in Hz):

\[ B_e = 6.23 \times 10^{-6} f^2 + 9.339 \times 10^{-2} f + 28.52 \]  \hspace{1cm} (7.4)

ERB-rate \( E \) (in ERB units) = \[ 11.17 \ln[(f + 312)/(f + 14675)] + 43.0 \]  \hspace{1cm} (7.5)

These equations are valid within the frequency range from 0.1 to 6.5 kHz (Moore & Glasberg 1983). It should be noted that the physiological frequency analysis is different from the standard Fourier decomposition of a signal into its frequency components. A key difference is that the auditory system’s frequency response is not linear. The relationship between the centre frequency of the analysis filters and their location along the BM is approximately logarithmic in nature (Johnson 2003).

In the speech and speaker recognition based researches, most of the works rely on extracting cepstral coefficient features from triangular shaped Mel scale filter bank. But this bank of triangular shaped Mel scale filters may lose some useful information at the sharp boundaries between the triangular filter bins. To avoid this, a filter structure having no sharp boundaries – Gaussian shaped was introduced by the research community. The quadrilateral shaped filters introduced in this chapter cover most of the low frequency than the high frequency components. A single bin of the quadrilateral filter is shown in Figure 7.1.
Figure 7.1 Single Quadrilateral Filter bin

The first and last filter bin’s center frequency is determined from Moore & Grasberg's ERB expression as given in Equation (7.6) and (7.7).

\[
af_{c_i}^2 + bf_{c_i} + c = \frac{1}{2}(f_{\text{high}_i} - f_{\text{low}_i}) \tag{7.6}
\]

\[
ERB_i = 24.7(0.00437f_{c_i} + 1) \tag{7.7}
\]

For the remaining filter bins except the first and last quadrilateral filter bins, the center frequency is found by using Equation (7.8),

\[
f_{c_i} = f_{\text{low}_i} + (i - 1)\left[\frac{f_{\text{high}_i} - f_{\text{low}_i}}{N + 2}\right] \tag{7.8}
\]

The lower and upper frequency of each filter bin is given as

\[
(700 + f_{c_i})^2 = (700 + f_{\text{low}_i} + 2ERB_i)(700 + f_{\text{low}_i}) \tag{7.9}
\]

\[
f_{\text{high}_i} = f_{\text{low}_i} + 2ERB_i \tag{7.10}
\]
The two intermediate frequencies \( f_{int1} \) and \( f_{int2} \) which forms the middle two vertices of the quadrilateral are found using the equations (7.11) and (7.12),

\[
\begin{align*}
    f_{int1} &= 0.25 \left( f_{high_i} + f_{low_i} \right) \\
    f_{int2} &= 0.75 \left( f_{high_i} + f_{low_i} \right)
\end{align*}
\] (7.11) (7.12)

where,

\[ a = 6.23 \times 10^6; \quad b = 93.39 \times 10^{-3}; \quad c = 28.52 \]

\( f_c \) is the \( i^{th} \) center frequency of the filter bin,

\( f_{high} \) is the upper frequency range of the \( i^{th} \) filter bin,

\( f_{low} \) is the lower frequency range of the \( i^{th} \) filter bin,

\( f_{int1} \) is the first intermediate frequency of the \( i^{th} \) filter bin,

\( f_{int2} \) is the second intermediate frequency of the \( i^{th} \) filter bin,

\( i = 1, 2, \ldots \) \( N \), \( N \) is the total number of filter bins in the filter bank.

After the construction of a first single bin, the second bin is constructed such that its lower frequency is the first intermediate frequency of the previous filter bin. The amplitude of the four vertices in each quadrilateral bins are \([0, 0.7, 1, \text{ and } 0]\). The value 0.7 is found to be optimum height of the second vertices after a series of tests for the values in between the range \([0.5, 1.0]\). The resultant designed filter bank will be placed in linear frequency scale as shown in Figure 7.2. But, the response human auditory system is non linear. Hence, the linearly spaced filters should be converted into non linear scaling before further processing. The designed Quadrilateral filter bank structure is placed in Mel frequency scaling in order to closely approximate
the human cochlear membrane and the resultant filter bank structure is shown in Figure 7.3.

![Linear Quadrilateral filter bank](image)

**Figure 7.2 Linear Quadrilateral Filter bank Structure**

In the linear filter bank all filter bins of the quadrilateral filter bank is arranged at uniform intervals. In Mel quadrilateral filter bank, filter bins are arranged non-uniformly Mel frequency mapping. Mel mapping places more filter bins in low frequency and few filter bins in high frequency. Quadrilateral shape is provided with wide width at low frequency and narrow width at high frequency for every filter bin in the filter bank. Altogether, Mel quadrilateral filter bank filter bins help in acquiring most of the speech energies by encompassing low frequency region with more filter bins.
7.4 SUMMARY

A new Quadrilateral structure Mel scale filter bank design is discussed in this chapter. Mel frequency scale is linear below 1 KHz and logarithmic above 1 KHz, covering low frequency region with more filter bins whereas, quadrilateral shape of every filter bin covers more low frequency within its first two vertices. Compared to the conventional triangular and Gaussian shaped Mel filter bank, newly introduced quadrilateral shaped Mel filter bank, because of its shape, encompasses more low frequencies than high frequencies in adjacent filters in the filter bank. Since speech signal has more energy in the low frequency region, the newly introduced quadrilateral shaped filter is capable of acquiring more energy, in turn more useful speaker specific information, to increase the identity level in speaker identification task.