Auditory Filter Bank Lab

RealSimPLE Project*
Ryan J. Cassidy and Julius O. Smith III
Center for Computer Research in Music and Acoustics (CCRMA), and the
Department of Electrical Engineering
Stanford University
Stanford, CA

Abstract

This laboratory activity guides the student through an explanation of and experiments relating to a provided auditory filter bank implementation.

Contents

*Work supported by the Wallenberg Global Learning Network

1 Summary of Objectives

- Understand the general function of a filter bank.
- Understand the background and details of the filter bank model of the human auditory periphery.
- Differentiate between the various auditory filter bank models, along with their applications in contemporary scientific inquiry.
- Generate and analyze sound spectrograms corresponding to the output of a provided auditory filter bank implementation. These spectrograms will allow the student to explore the auditory filter bank representations of various vowel sounds.

2 Background and Theory

2.1 What is a Filter Bank?

Consider an input audio signal $x(n)$. This signal might have come from a microphone, or the pickup on an electric guitar. To learn more about the sampling process used to obtain $x(n)$, consult the digital waveguide model laboratory assignment[1].

As discussed briefly in the monochord laboratory assignment[2], the spectrum of a signal gives the distribution of signal energy as a function of frequency. One commonplace situation where the concept of a spectrum arises involves the tunable equalizers found on many home and car stereo systems. In their simplest form, these equalizers may consist of two controls to adjust the level of bass and treble in the audio signal played through the system speakers. The bass control allows the user to adjust the level of the lower-frequency energy in the signal spectrum, whereas the treble control allows for the adjustment of higher frequency energy in the spectrum. Other equalizers are more advanced; many often have several controls to adjust the strength of various separate regions in the signal spectrum. In all cases, however, it is necessary to think of signal energy as a function of frequency, as provided by the spectrum concept.

In order to separate energy from a frequency region of a signal's spectrum, a bandpass filter may be used. An ideal bandpass filter rejects all input signal energy outside of a desired frequency range, while giving as output all input signal energy within that range. The range of accepted frequencies is often referred to as the band, or passband. The frequency boundaries defining the band, f_{cl} and f_{ch}, are known as the lower and upper cutoff frequencies (respectively). These are also referred to as the band edges. The difference between the upper and lower cutoff frequencies is known as the bandwidth:

$$BW = f_{ch} - f_{cl}. \tag{1}$$

The midpoint of the band edges is known as the center frequency f_c of the bandpass filter. Finally, the ratio of the center frequency f_c to the bandwidth BW of the filter is called the quality factor:

$$Q = \frac{f_c}{BW} \tag{2}$$

[1]http://ccrma.stanford.edu/realsimple/waveguideintro/
[2]http://ccrma.stanford.edu/realsimple/lab_inst/

A sketch showing the frequency response of an ideal bandpass filter with key features labeled is shown in Figure 1.

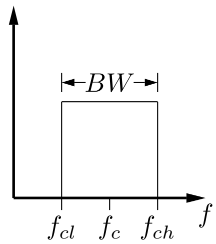

Figure 1: Sketch of the frequency response of an ideal bandpass filter, with key features labeled.

A filter bank is a system that divides the input signal $x(n)$ into a set of analysis signals $x_1(n), x_2(n), \ldots$, each of which corresponds to a different region in the spectrum of $x(n)$. Typically, the regions in the spectrum given by the analysis signals collectively span the entire audible range of human hearing, from approximately 20 Hz to 20 kHz. Also, the regions usually do not overlap, but are lined up one after the other, with edges, touching, as shown in Figure 2. The analysis signals $x_1(n), x_2(n), \ldots$ may be obtained using a collection of bandpass filters with bandwidths $BW_1, BW_2, \ldots$ and center frequencies $f_{c1}, f_{c2}, \ldots$ (respectively).

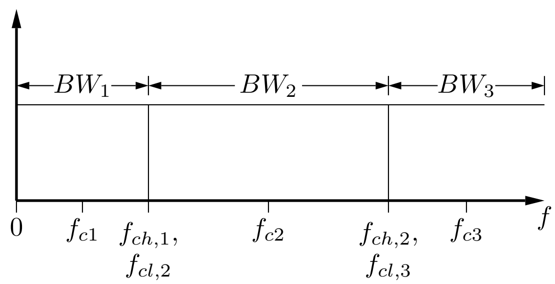

Figure 2: Sketch showing the bands of a three-band filter bank, with adjacent band edges touching but not overlapping. Together, the 3 bands span the frequency range from $f_{cl,1} = 0$ Hz to $f_{ch,3} = f_{\max}$, where $f_{\max}$ is the maximum frequency of interest (not shown).

3 The Auditory Filter Bank

It has been proposed that the peripheral auditory system effectively applies a filter bank to the acoustic signal reaching the ear drum. The output of this filter bank, consisting of the analysis signals $x_1(n), x_2(n), \ldots$, then affects the information transmitted to the auditory organs of the brain. In simulating the human auditory periphery, much scientific work to date thus employs some form of filter bank to emulate the characteristics of this proposed auditory bank. Some natural questions arise regarding this proposal:

- How many bands should an auditory filter bank have?

- What should the bandwidth of each band be?
- Should there be any overlap between adjacent bands?
- What should the bandwidths of each band be?

In the following sub-sections, we present the characteristics of two popular types of auditory filter banks: a more traditional, third-octave filter bank, and a more recently proposed Equivalent Rectangular Bandwidth (ERB) filter bank.

3.1 Third-Octave Filter Banks

Third-octave filter banks have historically been popular in audio analysis, as the bandwidths of these types of banks have been shown to loosely approximate the measured bandwidths of the auditory filters. Third-octave banks have also been internationally standardized for use in audio analysis [1]. In a third-octave filter bank, the center frequencies of the various bands $f_c[k]$ are defined relative to a bandpass filter centered at $f_c[0] = 1000$ Hz, by the following formula:

$$f_c[k] = 2^{k/3} 1000\text{Hz}. \tag{3}$$

The upper and lower band edges in the kth band are further given by the geometric means

$$f_{ch}[k] = \sqrt{f_c[k] f_c[k+1]}, \tag{4}$$

and

$$f_{cl}[k] = \sqrt{f_c[k-1] f_c[k]}, \tag{5}$$

respectively. From the above equations, it may be shown that the bandwidth of the kth band is given by

$$BW[k] = f_c[k] \frac{2^{1/3} - 1}{2^{1/6}}. \tag{6}$$

It may be shown that as the bandwidth in Equation (6) above is proportional to center frequency, the quality factor of each third-octave band filter is independent of k. As a result, filter banks such as the third-octave bank are referred to as *constant-Q* filter banks.

3.2 Equivalent Rectangular Bandwidth (ERB) Filter Bank

After the development of third-octave filter banks, psychoacousticians performed further studies to obtain more accurate estimates of the auditory filter bandwidths. Most recently, they arrived at a formula they use to refer to *Equivalent Rectangular Bandwidth (ERB)*. While a formula to convert frequency values into ERB-based frequencies is provided in psychoacoustics laboratory assignment[3], the bandwidth of an ERB filter centered at a given frequency f_c is

$$BW_{ERB} = 24.7\left(0.00437 f_c + 1\right). \tag{7}$$

It is important to note that the formula above converts a frequency (in Hz) to a bandwidth (also in Hz). To convert a frequency in Hz to a frequency in units of ERB-bands, the formula from psychoacoustics laboratory assignment[4] should be used, namely

$$ERB_{rate} = 21.4 \log\left(0.00437 f_c + 1\right) \tag{8}$$

[3] http://ccrma.stanford.edu/realsimple/psychoacoustics/

[4] http://ccrma.stanford.edu/realsimple/psychoacoustics/

3.3 Questions

- What is the quality factor of the third-octave filters whose bandwidth is given in Equation (6)?
- What is the bandwidth of an ERB filter centered at 1 kHz? What is the quality factor of this filter?
- Centered at 1 kHz, what is the difference between the bandwidth of an ERB filter and a third-octave filter, expressed as a percentage of center frequency?
- Repeat the previous question for a filter centered at 250 Hz.

4 Procedure

In this laboratory, you will analyze a variety of sound recordings. Each recording contains a single spoken word, featuring a single vowel sound. For each sound, you will use provided third-octave filter bank software to visualize the different sounds.

- If you have not already done so, install the program Octave on your computer. Start the program by typing `octave` on the command line. Also, download the archive of source code required for this lab[5], and uncompress the archive into the directory in which you will be running Octave.
- First, we need to load the first sound into Octave so that we can analyze it. To do this, enter the following command:

```
> x1 = wavread('vowel1.wav');
```

- To listen to the recording you have just loaded, enter the following command:

```
> play_sound(x1);
```

  Can you identify the vowel sound? To learn about the technical terminology used to describe various vowel sounds, visit the following online article: http://en.wikipedia.org/wiki/Vowel. What is the technical term and symbol for this vowel sound?

- Next you can plot a third-octave analysis of the vowel sound you just played by entering the following command:

```
> third_octave_analysis(x1);
```

  On the plot, the x-axis shows the time in seconds, and the y-axis shows a third-octave band index ℓ, where $\ell = k - 11$, where k is defined as in Equation (3). In other words, $\ell = 11$ corresponds to a frequency of 1 kHz. The orange-red colors on the plot indicate areas of high energy, whereas the blue-dark colors on the plot indicate areas of low energy. Which frequency index ℓ shows approximately the darkest red patch in the middle of the recording?

[5] http://ccrma.stanford.edu/realsimple/aud_fb/tova.zip

- Now load a second recorded vowel sound (this is a different vowel than the first):

  ```
  > x2 = wavread('vowel2.wav');
  ```

  Again, you can play this second recording using the `play_sound()` command:

  ```
  > play_sound(x2);
  ```

  What is the technical term and symbol for this vowel?

- Finally, you can perform a third-octave analysis on this recording by issuing a familiar command:

  ```
  > third_octave_analysis(x2);
  ```

  In this plot, you should notice a different pattern of orange-red patches. How does this pattern differ from the previous pattern? Do you think you could tell the vowels apart by looking only at the pictures?[6]

References

[1] *IEC 61260: Electroacoustics – Octave-band and Fractional-Octave-Band Filters*, Geneva, Switzerland: International Electrotechnical Commission, 1995.

[6] In fact, engineers and scientists often use an analysis roughly similar to third-octave analysis as part of their strategy for automatic speech recognition by computer.