M.Tech. Credit Seminar, Electronic Systems Group, EE Department, IIT Bombay

Basics of Hearing

Rishabh Bhargava (Roll No: 03307425)

Supervisor: Prof. P. Rao

November 2003

Abstract

An understanding of human auditory system can lead to robust automatic speech recognition, efficient speech coders and high quality speech synthesis. The report presents the basics of hearing mechanism with reference to frequency-to-place transformation along basilar membrane (BM) in the inner ear. In this report, we follow the acoustic signal as it transcends the peripheral auditory system in the form of neural impulses for further processing by the brain. Further, various psychoacoustic observations such as superposition effects, frequency masking etc. which give insights into the perception of primary sound attributes are interpreted on the basis of neural response of the fibers along the BM.

1 Introduction

The human auditory system represents an amazing system which can detect, separate and recognize an astonishingly wide variety of complex sounds. Research in modeling and analysis of our hearing mechanism has been actively pursued over the last century. There has been tremendous progress in understanding of speech perception process, especially from the point of view of psychophysics, acoustics, linguistics, neural processing and auditory physiology. Despite such progress, speech processing community has not made adequate use of these results in speech recognition, synthesis and coding applications.

Nevertheless, over the last decade or so, it has been realized that taking cues from auditory models can lead to improved performance of the stated speech applications. The peripheral auditory system has been well understood and many computational models explaining the psychoacoustic observations have been proposed. However, our understanding of the neural processing of speech signal at higher centers of auditory processing in the brain is still rudimentary.

Section 2 gives an overview of the primary attributes of an aoustic signal viz. pitch, loudness and timbre, on the basis of which the auditory system classifies the signal. The functioning of the auditory periphery with emphasis on basilar membrane (BM) behavior is presented in Section 3. Section 4 presents an

Table 1: Dependence of subjective qualities of sound on physical parameters. Adapted from [1].

Subjective quality → **Physical parameter** ↓	Pitch	Loudness	Timbre
Frequency	***	*	*
Pressure	*	***	*
Spectrum	*	*	***
Envelope	*	*	**

* weakly dependent; ** moderately dependent; *** strongly dependent

interpretation of pitch perception based on excitation pattern evoked along the BM response. Section 5 explains the loudness perception mechanism in terms of neural firing rates. It also discusses various concepts related to loudness such as superposition and frequency masking effects.

2 Subjective Attributes of Sound

It is universally accepted that the human auditory system can unambiguously associate three primary attributes to a given sound:

1. Pitch
2. Loudness
3. Timbre

The assignment of pitch, loudness and timbre to a sound is a result of complex processing operations in the ear and in the brain. This perception is subjective and can't be physically measured.

These sensations can only be quantified by conducting psychophysical experiments. In such experiments, the *physical input stimuli* (acoustic signal) is applied to subjects' *sensory system* (auditory system) and the response is noted in terms of the *psychological sensations* expressed by the subject.

In spite of being subjective, each one of these sensations can be associated to some well-defined physical quantities of the original stimulus that can be measured and expressed numerically. Table 1 shows the interrelation of the above *subjective* sensations to the associated physically measurable quantities.

It can be concluded from Table 1 that the sensation of pitch is strongly related to the *frequency* (repetition rate of the vibration pattern), loudness to the *intensity* (pressure oscillation amplitude of the sound) and timbre to the *spectrum* (frequencies and amplitudes of all the components in the sound) and *temporal envelope* (the transient attack, the release and variations in amplitude).

2.1 Pitch

Pitch is that aspect of our sensation that allows us to follow the melody of a piece of music or the intonation of speech. The American National Standards Institute defines it as: "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low". This definition can be best interpreted in terms of a musical scale.

There are no physical limits to the repetition frequency of a sound. However there are limits to our sensation of pitch. At the lowest level, our sensation of pitch disappears when the sound frequency drops below about 20 Hz. Below this point the individual vibration patterns start being discerned as beats. At the high frequency end, the sensation of pitch gets gradually more confused above 4000 Hz. We can still differentiate between the pitch levels but it becomes it hard to recognize and associate the pitch of the sounds with frequencies higher than 4000 Hz. It is possible that the sense of tone quality (or timbre) merges with the sense of pitch above this limit [2].

The following points can help in qualitative appreciation of the pitch phenomenon:

- Pitch levels have a natural ordering from low to high. This implies that sounds can be ordered in terms of pitch, regardless of other properties. Sound waveforms that have a short repetition frequency tend to give the perception of a low pitch, while sound waveforms with a high repetition frequency tend to give the perception of a high pitch.

- Only signals that have some repetitive structure provide a clear sense of pitch. This leads to different strengths of pitch sensation. A waveform in which the repetitions are approximate, rather than exact, will have a weaker pitch than a perfectly regular waveform.

  For example, a *complex tone* (a periodic sound wave consisting of more than one frequency component occur at the harmonic frequencies of a fundamental frequency) has a clear pitch corresponding to the fundamental frequency present in the tone, whereas *whistle* (consisting of a random combination of sine waves of different frequencies) has a weak pitch. In the extreme case, *white noise* has virtually no associated pitch.

- The pitch of sounds is most noticeable when it is changing within time. Indeed it is the change in pitch that conveys information about the identity of a tune or provides cue about the information structuring of speech. [2]

2.2 Loudness

Loudness is an attribute related to the sensation of pressure variation of the sound. Unlike the physical scale of amplitude, loudness has two limits of sensitivity to a

tone of given frequency: a lower limit– *threshold of hearing*– representing the minimum just audible intensity; and an upper limit– *threshold of pain*– beyond which physiological pain is evoked, eventually leading to physical damage of the hearing mechanism. In general, for a tone of frequency of about 1000 Hz, the difference between the limits is largest. The range of intensities encompassed between the two limits of hearing is of the order of 10^{12}. Such wide range of pressure stimuli are more conveniently measured on a logarithmic scale. Figure 1 shows the limits of loudness sensitivity on an equal loudness curve with a 1000 Hz reference tone.

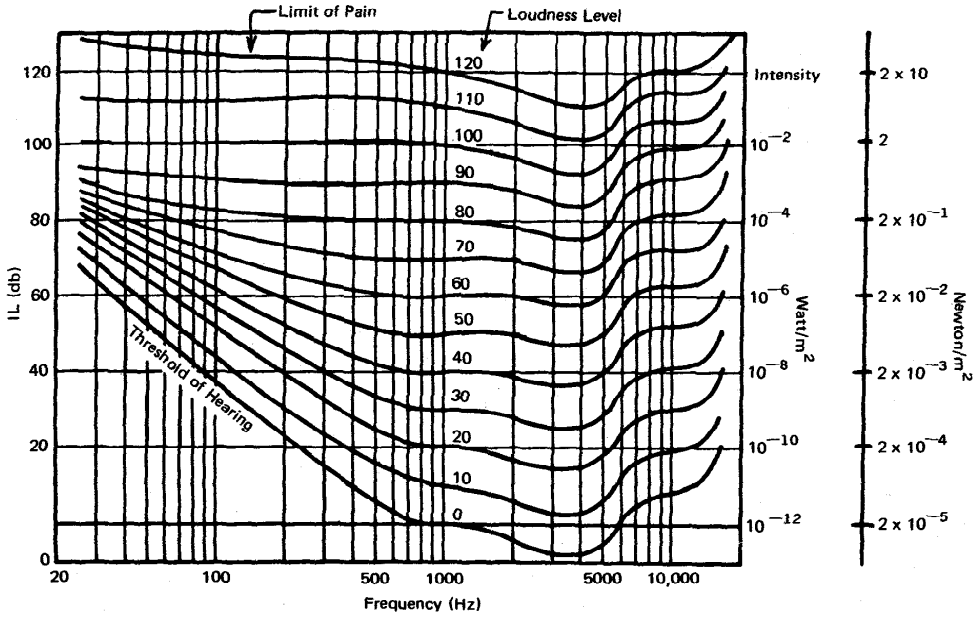

Figure 1: Curves of equal loudness in IL and associated frequency diagram [3].

2.2.1 Physical parameters of loudness

There are a number of physical quantities closely related to the *subjective* sensation of loudness:

- *Sound pressure level:* Sound pressure level (SPL) refers to the average pressure variation associated with a sound wave. SPL is defined with respect to an average pressure variation of $\Delta p_0 = 2 \times 10^{-5} Nm^{-2}$ corresponding to the minimum threshold of hearing:

$$SPL = 20\, log \frac{\Delta p}{\Delta p_0} \tag{1}$$

- *Sound intensity level:* Sound intensity level (IL) refers to the rate of energy flow across a unit area. The reference for measuring IL is $I_0 = 10^{-12}\, W m^{-2}$, and is defined as:

$$IL = 20\, log \frac{I}{I_0} \tag{2}$$

2.2.2 Subjective loudness

To represent the intensities or SPLs that are perceived equally loud by human auditory system, a subjective scale of loudness has been devised. The *subjective loudness* L is expressed in a unit called *sone*. The sone is defined as the loudness of a 100 Hz tone at a sound level of 40 db. It has been shown that the relation between L and the wave intensity I or the average pressure variation Δp, can be described approximately as [3]:

$$L = C_1 \sqrt[3]{I} = C_2 \sqrt[3]{\Delta p} \tag{3}$$

For a *traveling wave*, IL and SPL represent one and the same thing. However, for *standing waves*, there is no energy flow at all, and the intensity I as used in (2) cannot be defined; hence IL loses its meaning. Yet the concept of average pressure variation Δp, and hence SPL, at a given point in space is still meaningful.

It is notable that the ear does not respond to the total acoustical energy which reaches the eardrum. Rather, it is sensitive to the rate at which this energy arrives. This rate is what determines the sensation of loudness [3].

2.3 Timbre

Timbre is used to denote the *tone quality* of a sound. The American National Standards Institute defines it as: "that attribute of auditory sensation in terms of which a listener can judge two sounds, similarly presented and having the same loudness and pitch, as dissimilar".

Unlike pitch and loudness, timbre is dependent on a large number of parameters. Hence, it is more difficult to analyze in physical and mathematical terms. Timbre depends on the spectrum of the stimulus, the sound pressure, the frequency of the spectrum, and the temporal characteristics of the stimulus [1]. So, the judgment of timbre must take place under conditions of equal loudness and pitch.

3 Peripheral auditory system

Speech communication involves signal processing at two levels in the auditory system:

- *Peripheral auditory system*: It refers to the process of *hearing*, during which the pressure variations in the outer ear are represented by neural firing patterns on the auditory nerve.
- *Neural auditory system*: It refers to *perception*, which involves a higher level of processing in the auditory nervous system to translate the neural firings in the auditory nerve into perceptual information.

In this report, we limit our discussion to the *peripheral auditory system*.

3.1 Structure of the ear

The auditory periphery can be subdivided into three main sections, as depicted in Figure 2:

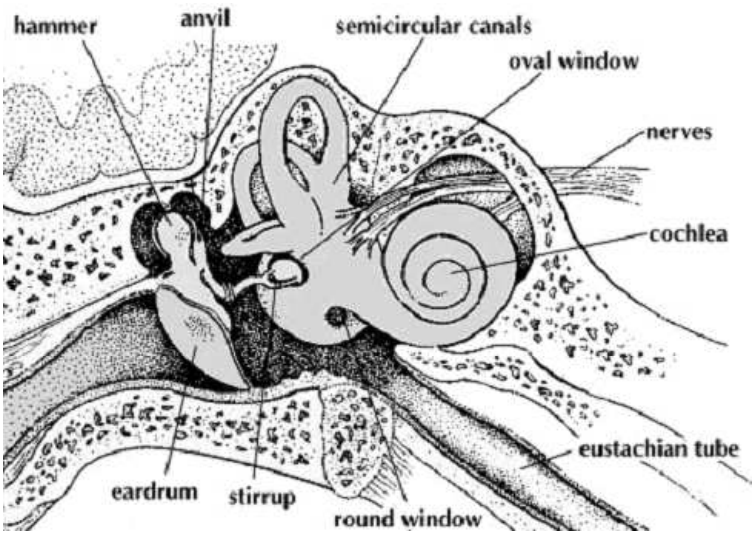

Figure 2: Structure of an ear [2].

1. Outer Ear– It consists of the external *pinna* and the *auditory canal*, which is terminated by the *eardrum*. The pinna helps in sound localization and helps in determining the direction of origin of sounds, especially at high frequencies. The auditory canal acts as a pipe resonator that aids perception of sounds having significant information at frequencies in the range of 3-5 kHz [4].
2. Middle ear– The eardrum marks the beginning of of the middle ear, an air filled cavity that contains three tiny bones (malleus, incus and stapes)

called *ossicles*. The ossicles transmit eardrum vibrations to the *oval window* membrane of the inner ear.

The eardrum changes the pressure variations of incoming sound waves into mechanical vibrations to be transmitted via ossicles to the inner ear. The ossicles acts as a mechanical transformer to change small pressure exerted by the sound wave on the eardrum into much greater pressure variations on the oval window. The middle ear also protects the delicate inner ear against very loud sounds and sudden pressure changes, by muscular contraction via an *acoustic reflex* mechanism.

3. Inner ear– The main hearing organ in the inner ear is a tapered tube, filled with fluids, called *cochlea*. It transforms mechanical vibrations at its oval window input into electrical excitation on its neural fiber outputs. The cross-section of the cochlea is divided into three distinct chambers by two membranes that run along the entire length: the *Reissner's membrane* and the *basilar membrane*. On the BM lies the organ of Corti, which contains about 30000 sensory *hair cells* arranged in several rows along the cochlea [4]. The endings of the auditory nerve terminate on these hair cells.

3.2 Basilar membrane behavior

The basic hearing mechanism can be explained by the response of the BM to vibrations in the fluids in the cochlear cavity. Due to gradual change in width and elasticity of the BM along its length, its frequency response varies accordingly. When the stapes vibrate against the oval window (at the base), hydraulic pressure waves are transmitted rapidly down the cavity, inducing ripples in the BM. As these waves travel down the apex, its amplitude increases to a maximum at a given place, which depend on the input frequency, and then dies out very quickly toward the apex, as shown in Figure 3. The hair cells pick up the motion of the BM and impart signals to the attached nerve cells, culminating in the generation of electrical signals in the acoustic nerve.

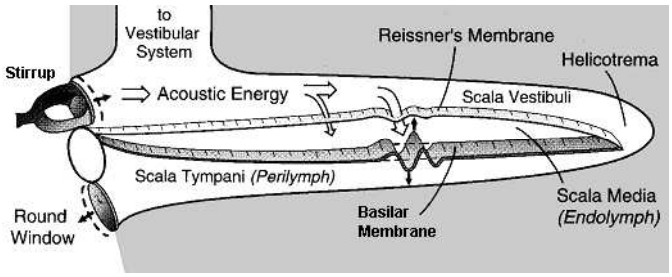

Figure 3: Vibration pattern along basilar membrane [2].

The interesting fact is that for a pure tone of given frequency, the maximum BM oscillations occur only in a limited region (called *resonance region*) of the membrane. The higher the frequency of the tone, the closer to the base (oval window) it is located (where the membrane is stiffest). Thus, the *spatial position* along the BM of the responding hair cells and associated neurons determine the primary sensation of pitch. Figure 4 shows this *frequency-to-place transformation* along the BM. Further, the displacement of the resonance region along the BM follows a *logarithmic* relationship with the frequency of the exciting tone.

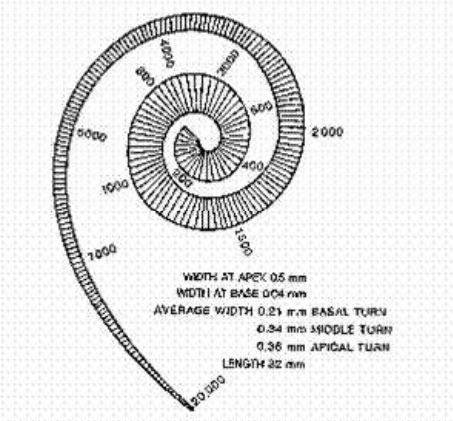

Figure 4: Frequency-to-place transformation along BM [5].

It has been found that a given nerve fiber, which innervates a particular region of the BM, has a lowest firing threshold for that acoustical frequency which evokes a maximum oscillation at that place of the BM. This frequency of maximum response is called the neuron's *characteristic frequency*. Thus, we can think of BM as a bank of filters, with response as given in Figure 5. This *filtering* allows the separation of various frequency components of the signal with a good signal-to-noise ratio. This interpretation is equivalent to that described in the previous paragraph, but has been reiterated for the sake of completeness.

4 Perception of Pitch

4.1 Pitch of pure tones

The human pitch detection mechanism involves the processing of the shape of excitation pattern of nerve firing along the BM, to infer the following pitch judgments of the pure tones [2]:

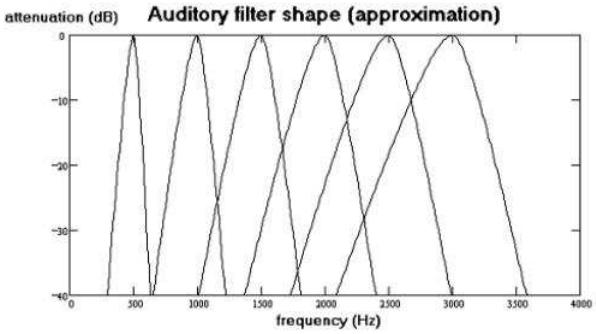

Figure 5: Auditory filter response of the BM [2].

1. *Identification of pitch level:* The position of the peak in the excitation pattern changes with the frequency of the stimulating tone as the resonant properties of the BM vary along its length. Tones that cause an excitation pattern that peaks toward the apex of the cochlea give a sensation of low pitch, while tones that cause a peak toward the base of the cochlea give a sensation of high pitch. Thus, excitation patterns that peak in similar places will have similar pitch, while patterns that peak at different places will have different pitch.

2. *Discrimination of pitch levels:* Tones of different frequency evoke excitation patterns on the BM which peak at different places. Due to background noise in the original signal and the uncertainty in the nerve firing mechanism, there is always some natural fluctuation in any excitation pattern. Therefore, two tones of different frequency can only be perceived as distinct if the patterns vary by more than, they would, if the two tones were the same frequency. It has been proposed that two excitation patterns can only be perceived as distinct if they vary somewhere along their length by at least 1dB.

3. *Assessment of the strength of pitch:* Pure tones evoke a distinct excitation pattern on the BM, with a single sharp peak, and sides of certain slope, steeper on the low-frequency side (see Figure 5). So, a signal composed of a combination of tones of different frequencies would excite a wider range of fibers resulting in a less sharp pattern with different slopes. The closer an excitation pattern is to this shape, the narrower the bandwidth of the signal and the more simply repetitive it is. This is used as the basis for determining whether the signal has a clear pitch.

4.2 Theories of pitch perception

Two disparate theories have been offered to explain the pitch perception mechanism in the auditory system:

1. *Place theory* maintains that the perception of pitch depends on the vibration of different portions of the the BM. That is, hair cells in each region of the membrane are specialized for the detection of specific sound frequency.

2. *Temporal theory* maintains that the pitch perception corresponds to the rate of vibration of hair cells all along the BM. This is based on the observation that hair cells tend to fire only at a particular phase of the BM vibration. This rate of vibration is then encoded in terms of neuronal firing rate for higher order frequency based processing in the brain.

There has been contention over the validity of each of the two theories. The actual findings suggest that both theories are in part valid in their explanation of the mechanism underlying pitch perception. Place theory is accurate, except that the hair cells along BM lack independence in response. They in fact vibrate together as suggested by the temporal theory. Likewise, the temporal theory was weakened when discovered that the hair cells are unable to fire at rates reflecting the higher frequency of hearing. Current thinking maintains that the sounds under 1 kHz are translated into pitch through temporal coding. Sounds between 1-5 kHz are coded via a combination of place and temporal coding. Finally, for sounds over 5 kHz, pitch is coded via place coding [2].

4.3 Superposition of pure tones

Sounds are seldom heard in isolation. The sensation of pitch of a pure tone very much depends on the frequency of the other interfering tone. The tone sensations evoked by the superposition of two pure tones of equal amplitude and of frequencies f_1 and $f_2 = f_1 + \Delta f$, respectively can be summarized as (Figure 6):

- At the unison, when $\Delta f = 0$, we hear one single tone of pitch corresponding to f_1 (or f_2) and loudness that will depend on the amplitudes of the superposed tones and their phase difference.

- When $\Delta f < 10Hz$, we continue hearing one single tone, but of slightly *higher pitch*, corresponding to the average frequency $f = f_1 + \Delta f/2$. The loudness of this tone will be beating with a frequency Δf.

- When $10Hz < \Delta f < 15Hz$, the beat sensation disappears, giving way to a quite characteristic *roughness* of the resulting tone sensation.

- When $\Delta f > \Delta f_D$, the *limit of frequency discrimination*, we begin to distinguish two separate tones of pitch corresponding to f_1 and f_2. At this moment, the two resonance regions on the BM have separated sufficiently to give two distinct pitch signals. However, the sensation of roughness still persists.

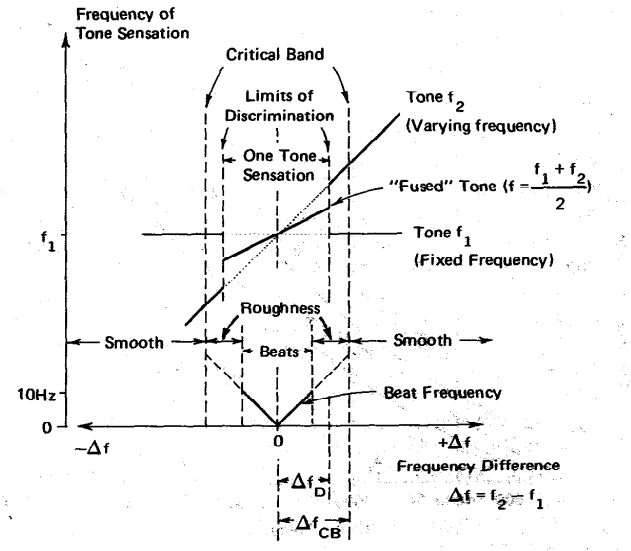

Figure 6: Tone sensations evoked by the superposition of two pure tones of nearby frequencies [3].

- Only when $\Delta f > \Delta f_{CB}$, the *critical band*, the roughness sensation disappears and both pure tones sound smooth and pleasing.

The existence of a finite Δf_D indicates that the resonance region on the BM, corresponding to a pure tone must have a finite extension. Otherwise, two superposed tones would always be heard as two separate tones, no matter how small their frequency separation be. Further, the persistence of roughness sensation, even beyond Δf_D is an indication that the two resonance regions still overlap and interfere to some extent, at least until Δf_{CB}.

The critical band represents a sort of "information collection and integration unit" on the BM. It has been found that it corresponds to an extension of BM innervated by a constant number of hair cells, all along the BM. The critical band is an important concept in pitch perception, which helps in understanding of many psychoacoustic phenomenon.

5 Perception of Loudness

The huge range of detectable sound intensities (of the order of 10^{12}) are mapped to the limited subjective scale of loudness via range compression by the loudness detection mechanism. The loudness perception in the auditory periphery is encoded in neural firings as follows [3]:

1. When a pure sound is present, the primary neuron fibers with the same characteristic frequency increase their firing rate above the spontaneous level. This increase is monotonic but a non-linear function of the stimulus amplitude.

2. At higher SPLs, a primary neuron's firing rate begins to saturate. The nerve fibers with similar characteristic frequencies have widely different firing thresholds. Indeed, it has been found that each inner hair cell receives fibers from three different kinds of auditory neurons. These neurons belong to one of the three groups: fibers with high spontaneous firing rates (up to 20 impulses per second) and low thresholds; fibers with very low spontaneous rates and high threshold values (up to 50-60 dB); and a group with intermediate spontaneous rates and thresholds. In this way, the ensemble of auditory neurons innervating from a given region on BM cover the wide dynamic range of perceivable sound levels.

3. As the intensity of a pure tone increases, the amplitude of the traveling wave increases all along the BM, not just in the peak resonance region. This gives a chance to neurons whose characteristic frequency is different from that of the incoming sound wave to increase their firing rate when their thresholds have been surpassed.

To summarize, the sound intensity information is encoded in: (1) *firing rate* of each neuron in the resonance region, (2) *type* of the acoustic nerve fiber carrying the information, and (3) *total number* of activated neurons [3].

5.1 Superposition effects on loudness

When two or more tones of the same frequency superpose, the perceived loudness depends on the intensities of the component tones [3]:

- If the frequencies of the component tones fall all within the critical band of the center frequency, the resulting loudness is given by the sum of the individual intensities (refer to (3))

$$L = C_1 \sqrt[3]{(I_1 + I_2 + I_3 + \ldots)} \tag{4}$$

  This result can be used for precise determination of the critical band.

- When the frequency spread of the multi-tone stimulus exceeds the critical band, the resulting subjective loudness is greater than that obtained by simple intensity summation, increasing with increasing frequency difference and tending toward a value given by the sum of the individual contributions from the adjacent critical bands

$$L = L_1 + L_2 + L_3 + \ldots \tag{5}$$

If the individual loudnesses differ considerably among each other, frequency masking effects must also be taken into account.

- When the frequency difference between individual tones is quite large, the situation becomes complicated. The mechanism tends to focus on only one of the component tones (such as the loudest or that of the highest pitch) and assigns the sensation of total loudness to just that single component

$$L = max\,(L_1, L_2, L_3 \ldots) \tag{6}$$

The relation between subjective loudness and total firing rate can very well explain the main characteristics of loudness summation. For simultaneous tones of frequencies differing more than a critical band, the grand total of transmitted neural impulses is roughly equal to the sum of the pulse rates evoked by each component separately; hence, the total loudness will tend to be the sum of the loudnesses of each tone. In contrast, for tones whose frequencies lie within a critical band, with resonance regions on the BM overlapping substantially, the total number of pulses will be controlled by the sum of the original stimulus intensities.

5.2 Frequency masking effects

The threshold of hearing of a single tone gets shifted upward when it is sounded in the presence of another. This observation can be explained by masking. It is defined as the minimum intensity level which the weaker *masked* tone must exceed in order to be heard individually in the presence of the louder *masking* tone. The most familiar experience of masking is that of not being able to follow a conversation in presence of a lot of background noise.

We restrict our discussion to *frequency masking*. The masking level of a pure tone of given frequency in presence of another pure tone of fixed characteristics is shown in Figure 7.

Many interesting conclusions can be inferred from typical frequency masking experiments [1]:

- Pure tones close together in frequency mask each other more than tones widely separated in frequency.
- A pure tone masks tones of higher frequency more effectively than tones of lower frequency.
- The greater the intensity of the masking tone, the broader the range of frequencies it can mask.
- If the two tones are widely separated in frequency, little or no masking occurs.

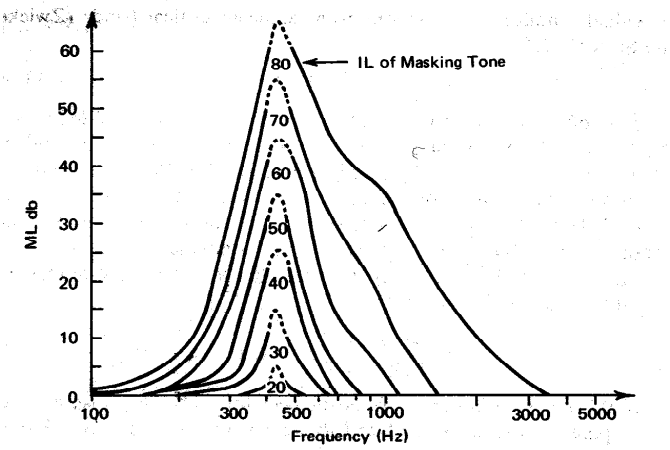

Figure 7: Masking level corresponding to a pure tone of 415 Hz for various sound level values of the masking tone [3].

The above conclusions can be well understood in the light of BM behavior to excitation by pure tones. We know from the discussion in Section 3.2 that high-frequency tones excite the BM near the oval window, whereas low-frequency tones create their greatest amplitude at the far end. Further, the excitation due to a pure tone is asymmetrical, having a tail that extends toward the high-frequency end. This may be accounted for by the fact that the frequency response of the BM is rather logarithmic, with higher frequencies being more closely spaced. Thus, it is easier to mask a tone of higher frequency than one of lower frequency. As the intensity of the masking tone increases, a greater part of its tail has amplitude sufficient to mask tones of higher frequency. Further, if the two tones are separated in frequency to such an extent that their excitation patterns do not interfere, no masking occurs.

6 Conclusions

Firstly, we identified the three subjective acoustic sensations perceivable by our auditory system, viz. pitch, loudness and timbre. A qualitative explanation of the above sensations in terms of physically measurable quantities was explored.

Further, a brief overview of auditory system with emphasis on signal processing in the auditory periphery was presented. The report discussed how the ear transforms acoustic pressure signal into a mechanical vibration pattern on the BM and then representing this pattern by a series of pulses for further processing in the brain. The key to this transformation lies in understanding the response of BM to the frequency components present in the incoming acoustic signal.

We then interpreted how the pitch information is encoded in the resonance

peaked excitation pattern evoked along the BM. Since all the observations related to perception of pitch cannot be fully explained by this peaked neural excitation pattern, an overview of the place coding and the frequency coding theories was given. A combination of these theories can satisfactorily explain the pitch perception process. To keep the discussion elementary, we have not incorporated modern pitch perception theories in the report. Some observations on the pitch sensations arising due to superposition of pure tones are also presented.

We also discussed the different ways in which loudness information is encoded in the excitation pattern. Finally, superposition and frequency masking effects on loudness perception were interpreted on the basis of neural firing patterns along the BM.

Acknowledgments

I will like to express my sincere thanks for Prof. P. Rao for her guidance all through the course of the seminar work. The lengthy, but useful, discussions with her helped me in comprehending the subtilities of the subject. I thank Vikash Sethia for his help in preparation of the report. I appreciate the support of Pradeep Kumar and Kotta Manohar of Digital Audio Processing lab during the seminar work.

References

[1] T. D. Rossing, *The Science of Music*. Reading: Addison-Wesley, 1990.

[2] M. Huckvale, "Web tutorial on pitch and loudness." http://www.phon.ucl.ac.uk/courses/spsci/b214/tutorial.htm, Nov. 2003.

[3] L. Roederer, *The Physics and Psychophysics of Music: An Introduction*. New York: Springer-Verlag, 1995.

[4] D. O'Shaughnessy, *Speech Communications: Human and Machine*, ch. 4, pp. 109–131. Hyderabad: University Press, 2001.

[5] P. Rao, "How do you hear it?." Presentation slides, 2003.