Chapter 6

Modeling loudness perception in the hearing impaired. I.

Abstract

The purpose of this chapter is two–fold. Firstly, different coding schemes which the auditory system could apply for coding sound intensity will be described. These different schemes are important for understanding the different "physiological" loudness models. Secondly, very different ways to quantitatively describe loudness perception and explain loudness recruitment will be described. Two types of very different models have been proposed to describe loudness perception in normal–hearing and hearing–impaired listeners. These models range from different simple power laws ("Stevens law") relating the physical magnitude of a stimulus directly to its subjective loudness impression without consideration of physiological findings ("psychological models"), to strongly physiologically oriented models which cannot yet be used to model loudness functions measured with magnitude estimation or categorical scaling ("physiological models"), since they do not relate nerve fiber output to a subjective loudness scale. The physiological models provide an excellent explanation for the occurence of loudness recruitment, while the psychological models provide the basis for calculating loudness on a subjective scale. Therefore, both approaches are important for the loudness model discussed in chapter 7.

6.1 Coding of sound level

One of the most remarkable characteristics of the auditory system is its ability to code sound pressure levels over the very large dynamic range of about 120 dB. Over that range, intensity discrimination performance, measured in terms of the Weber fraction, $\frac{\Delta I}{I}$ changes only by a small amount (Viemeister and Bacon, 1988). Several different theories have been proposed to explain this so called "dynamic range problem". These will be briefly described in this section. For extensive reviews on this topic see (Viemeister, 1988a; Viemeister, 1988b; Smith, 1988; Pickles, 1988; Moore, 1989; Plack and Carlyon, 1995).

There are at least three different means by which the auditory system might encode sound intensity. Firstly, the firing rate of auditory nerve fibers usually increases with increasing stimulus intensity. One problem with this coding mechanism is that the firing rates of individual neurons typically change over a dynamic range of only 20–40 dB, far less than the 120–dB dynamic range of auditory system. This problem was reduced by the discovery of several populations of neurons with different absolute thresholds[1] and shapes of rate/intensity (R/I) functions (Sachs and Abbas, 1974; Liberman, 1978; Liberman, 1982; Smith, 1988; Winter et al., 1990; Ruggero, 1992a; Yates et al., 1992). Although the dynamic ranges of these different populations cover the whole behavioral dynamic range of 120 dB, most neurons show saturation at relatively low sound levels, at least for tones with frequencies near the characteristic frequency (CF) of the neuron being considered. Several authors have proposed models for intensity discrimination based on firing rates (Winter and Palmer, 1991; Viemeister, 1988b). They showed that a very small number of neurons (about 20–30) within a small spectral range would suffice to provide intensity discrimination as good as that measured in psychoacoustical tasks. The auditory system appears to be able to code intensity over a very large dynamic range on the basis of information from a very small spectral range. Indeed, the information contained in the responses of all auditory nerve fibers (if used optimally) would provide a much better intensity discrimination than is actually seen.

A second way of encoding intensity information is by spread of excitation. For pure tones, the excitation on the basilar membrane (BM) spreads out with increasing stimulus level (Rose et al., 1971; Ruggero, 1992a). Thus information regarding the intensity of the stimulus, and regarding changes in intensity, is available in the firing rates of neurons with CFs above and below the stimulus frequency. The spread of excitation also gives rise to an increase in the number of activated neurons, which might also be used as a cue for intensity perception or discrimination (Florentine and Buus, 1981). Finally, intensity may also be coded using temporal cues (Rose et al., 1967; Javel et al., 1988). The precision of phase locking of neural spike trains to the stimulus fine structure increases with increasing level. However, the precision tends to saturate at relatively low sound levels, so this cue would not be very effective for coding absolute sound level. Temporal cues may be more important for coding the relative levels of components in complex sounds. The amount of phase locking to a given component depends on the level of that component relative to adjacent components. These temporal cues might play an important role for coding of loudness at very low frequencies.

Psychoacoustic experiments on intensity discrimination in a variety of conditions, like different frequency, masking / background noise, different signal durations, gating of stimuli, different levels, seem to indicate that a single code for loudness could be sufficient for a variety of experimental conditons, since the human performance is quite robust in these different experimental conditions (Carlyon and Moore, 1984; Viemeister, 1988b; Moore, 1989; Plack and Viemeister, 1993; Plack and Carlyon, 1995). These experiments further

[1]These neurons also differ in size, morphology and their projection to the cochlear nucleus.

show that neither spread of excitation nor phase locking are essential in order to maintain the large dynamic range over which loudness is coded. Spread of excitation, however, seems to be important to recover the slope of the loudness growth function. Hellman (1994) showed that, in subjects with a very steeply increasing, sloping hearing loss, loudness functions measured at the "cut–off" frequency (i.e., the frequency where the hearing loss starts to increase) grow at a lower than normal rate. Hellman explains these results as being caused by a restricted growth of spread of excitation induced by the steeply increasing hearing loss. A restricted spread of excitation can also be produced in normal–hearing subjects by masking the high frequency excitation by a broadband noise. Again a reduced growth of loudness function is also obtained (Hellman, 1978; Zwicker and Fastl, 1990; Schlauch, 1994).

In summary, several possibilities for the coding of loudness (sound intensity) exist. However, the most likely ones are the firing rate of auditory neurons and spread of excitation. In the next section these codes will play an important role for modeling loudness by physiological models and in explaining the occurrence of recruitment.

6.2 Physiological models

Several models, based on different intensity coding schemes, have been proposed to explain loudness perception and especially loudness recruitment on a physiological basis, although none has yet been extended to calculate loudness functions on a subjective scale. It is worth mentioning one assumption on which all physiological models are based without explicitly stating it. They all assume that recruitment can fully be accounted for by changes in cochlear mechanics. However, that is not necessarily the case (Boettcher and Salvi, 1993; Salvi et al., 1992). Changes in the input to higher stages of the auditory pathway due to damage of the cochleae could lead to alterations in the mode of operation of these higher stages. Thus, they might somehow contribute to loudness recruitment.

The first model was proposed by Evans (Evans, 1975); see also (Pickles, 1988; Phillips, 1987). It is based on the assumption that loudness is coded by the total amount of activity of auditory neurons. In impaired cochleae, neural tuning curves are broadened and do not show the sharp "tip" usually seen in normal cochleae. The idea is that, since the tips of tuning curves are missing, the neurons will not respond to a weak sound. However, once the intensity exceeds the threshold, the activity spreads rapidly across the array of neurons, because of the broad tuning (see Fig. 6.1). This yields a steep increase in percentage of activated fibers and thus of loudness function at threshold and above.

If spread of excitation were crucial for loudness coding, then the amount of recruitment should be reduced by presenting the signal in a notched noise, designed to mask activity in neurons with CFs far from the signal frequency. Moore et al. (1985) showed that this is not the case; the amount of recruitment was not reduced by such a background noise, indicating that abnormal growth of spread of excitation (i.e., the rapid increase of activity across different bands) is not the main factor producing recruitment.

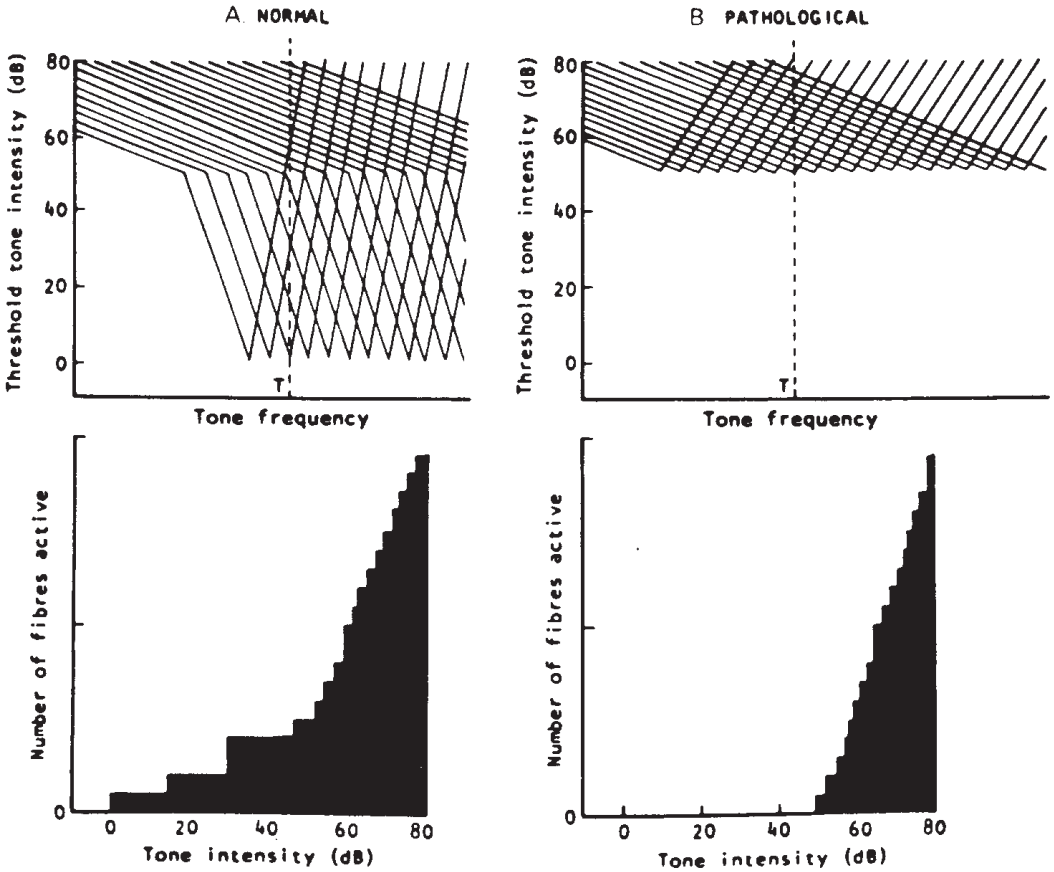

Fig. 6.1: *Neural model proposed by Evans (1975) to explain loudness recruitment. Since the narrow tuned (A) tips of the tuning curves are missing, loudness in the abnormal ear (B) grows abnormally quickly with intensity once threshold is exceeded.*

More recent models of loudness coding and recruitment are based on schemes other than spread of excitation coding. These models are all based on the assumption that recruitment is caused by the reduced nonlinearity on the BM. In fact, Evans' model is based on the same assumption, since the broadening of auditory filters is probably a result of less nonlinear BM vibration and thus less mechanical tuning.

Some models (Sachs et al., 1989; Yates, 1990; Zeng and Turner, 1991) use a firing–rate based code; thus, they assume that the loudness of a sinusoid is mainly determined by the firing rates of neurons with CFs close to the frequency of the sinusoid. Sachs and colleagues and Yates and colleagues developed very similar semi–empirical models to describe the different types of rate–intensity functions of auditory nerve fibers. Basically, their models consist of two stages (shown in the upper left and the large panel of Fig. 6.2). The large panel shows the BM input–output function at CF. This function consists of three parts, as measured physiologically (Sellick et al., 1982; Robles et al., 1986): a linear part at low levels (below about 40 dB SPL), a nonlinear (compressive) part at medium levels and a linear part at high

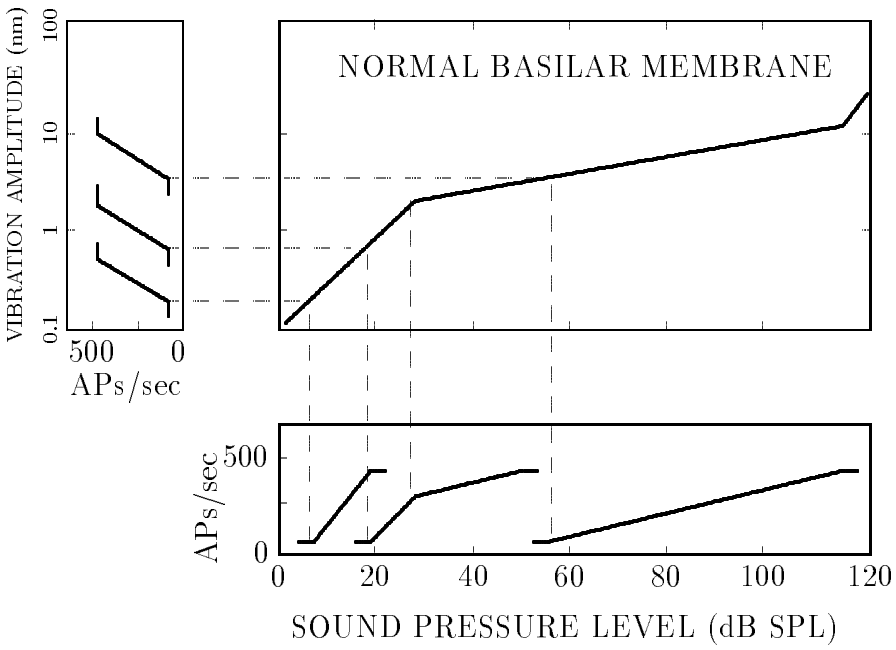

Fig. 6.2: *Schematic representation of the output of the model proposed by Sachs et al. and Yates et al.. Note how the nonlinear growth of basilar membrane input–output function (large panel) combined with sigmoidal shaped rate–intensity functions of auditory neurons (upper left panel) can account for different shapes of rate–intensity functions, indicating that an increase in firing rate (or action potential AP) exists over the whole dynamic range. Thus, an increase in firing rate could suffice for coding intensity over the large dynamic range of the auditory system. Adapted from Patuzzi (1992).*

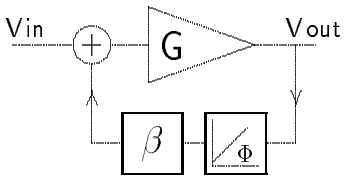

Fig. 6.3: *Electric analog of a feedback circuit describing basilar membrane vibration with mechanical support by OHC. For a detailed description see text. Adapted from Yates (1990).*

levels (above about 90 dB SPL). The transition from linear to nonlinear and back to linear behavior is believed to be correlated with the activity of OHCs (Patuzzi et al., 1989a; Patuzzi et al., 1989b; Patuzzi, 1992). In the model, the nonlinear behavior is introduced by an active feedback loop which contains a simple hyperbolic saturating nonlinearity Φ as is

shown in Fig. 6.3. This feedback path is assumed to reflect the mechanical support to BM vibration by OHCs. The shape of the saturating nonlinearity determines the shape of the BM transfer function: it grows linearly at low levels and is saturated at very high levels, together producing the linear increase of the BM transfer function with level at low and high levels. Between these two extremes, it grows nonlinearly with level and thus causes the nonlinear growth of BM transfer function with level. The insertion gain of the attenuator β is such that the output of the feedback path is slightly less then its input (without the saturating nonlinearity). The second stage of the model is shown in the upper left panel of Fig. 6.2. It describes the transformation from IHC activation (by BM vibration) to neural spike rate. It is assumed that the functions relating BM vibration to spike rate have a sigmoidal shape (deduced from measured R/I functions for frequencies in the tail of the tuning curve, where BM vibration is linear at all levels). The output of the first stage provides the input for the second; thus, BM vibration characteristics determine the shapes of R/I functions measured in the auditory neurons. The different types of R/I functions which originate from this interplay of BM mechanics and neural R/I functions, are shown in the lower panel of Fig. 6.2. Thus far, the model has only been used for modeling R/I functions and explaining their different shapes (Yates et al., 1990), and has not been extended to calculate loudness functions on a subjective scale like the sone or a categorical scale. Yates

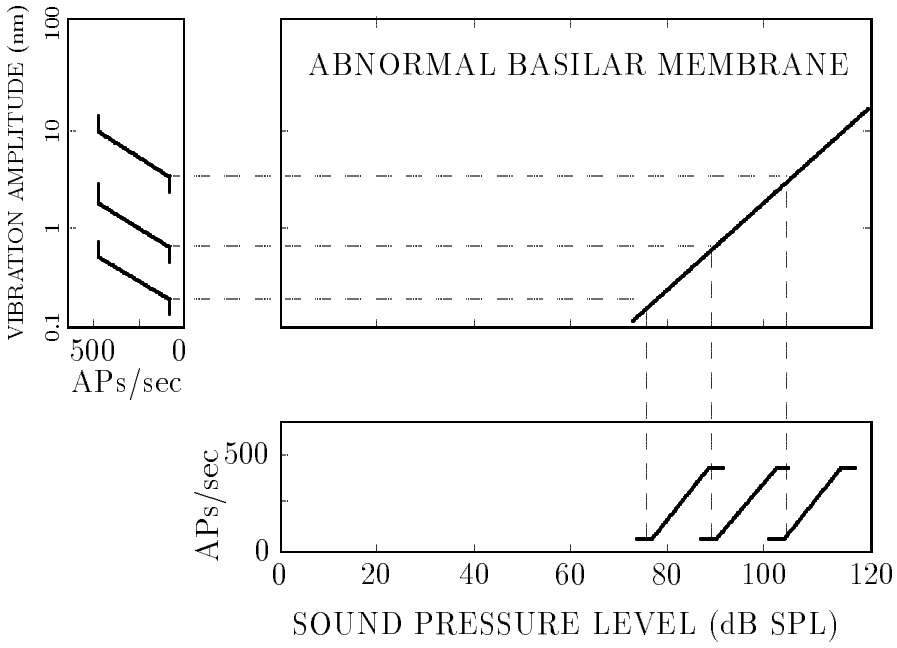

Fig. 6.4: *Same representation as in Fig. 6.2 but for a linearized BM transfer function, i.e., for an impaired cochlea. Assuming that loudness is determined by the increase in total firing rate (or action potential AP) of the auditory neurons, then this linearization should result in a steeper than normal loudness function. Adapted from Patuzzi (1992).*

(1990) and Patuzzi (1992) applied this model to explain the occurrence of recruitment and to model loudness functions of hearing impaired listeners measured with a loudness matching technique. Decreasing the amount of feedback, which corresponds to damage to the OHCs, provided to BM vibration via the feedback path shown in Fig. 6.3, yields a linear BM input–output function. This linearization yields a steeper BM input–output function and thus a steeper increase in firing rate, as illustrated in Fig. 6.2. If one assumes that loudness is related to the total firing rate of nerve fibers, then this steeper increase in firing rate should also yield a steeper loudness function, i.e., loudness recruitment.

The model proposed by Carney (1994) is based on a spatio–temporal coding scheme, which only works at low frequencies. The scheme combines spatial and temporal properties of nerve fiber responses, i.e., spatial and temporal codes for loudness. It is assumed that a kind of temporal activity pattern of nerve fiber responses at different center frequencies exists, which is crucially determined by nonlinearities in the peripheral processing. The model mainly consists of a time–varying peripheral filter, followed by a travelling wave delay, models for mechano–electrical transduction and a discharge generator simulating the transfer characteristics of auditory neurons. The bandwidth and the phase characteristics of the time–varying filters are determined by a nonlinear, level dependent feedback loop representing the nonlinear properties of peripheral processing. The temporal activity pattern of the model auditory nerve fibers provide the input for specialized cells working as coincidence detectors. There is evidence that such cells exist in the anteroventral cochlear nucleus (Carney, 1994). As level is increased the bandwidth and phase characteristics of the time-varying filters change causing a more similar timing of fibers across different frequencies. Therefore, fibers with different center frequencies discharge more synchronously. With increasing coincident input the output (i.e., the firing rate) of the coincidence detectors increases, which could be used as a loudness code. Again, recruitment could be explained by changing the amount of compression in the peripheral processing, i.e., by the amount of feedback provided to the time–varying filters. This model is heavily based on nonlinear processing in the cochlea even at low frequencies. Several authors reported that BM mechanics depends on frequency and is less compressive at low frequencies (Cooper and Yates, 1994; Patuzzi and Robertson, 1988).

The above described physiological models have not yet been extended to model psychoacoustic data, i.e., they cannot be used to describe loudness functions of hearing impaired listeners measured in different experimental conditions.

In summary, the physiological models all account for recruitment by linearizing the BM transfer function, i.e., by less compressive inner ear mechanics. Thus, according to these models, the main cause of recruitment is the loss of compressive nonlinearity. This has some implications for modeling loudness functions using Zwicker's model described in the chapter 7. In this model, the exponent of the power law is interpreted as reflecting nonlinear compressive processes in the cochlea. Thus, according to all physiological findings and models, this exponent should be modified when dealing with loudness functions for hearing impaired listeners.

6.3 Psychological models

The psychological models are based on Steven's power law relating subjective loudness to physical magnitude (Stevens, 1957; Luce and Krumhansl, 1988):

$$L = k \cdot I^{\alpha}, \tag{6.1}$$

where L is the loudness, I is the stimulus intensity and α is a constant equal to about 0.3. This power law relationship might be viewed as reflecting some of the compressive characteristics of the peripheral auditory system (Stevens, 1970; Hellman, 1991). For sounds close to absolute threshold, the change in loudness with level is more rapid than predicted by this simple power law. Different extensions to the power law have been proposed to correct for this (for a review see (Humes and Jesteadt, 1991)), which differ in the way they take the absolute threshold into account. Often, the absolute threshold is assumed to be produced by an internal (inaudible) noise which masks stimuli at very low levels. Hearing impairment might thus be modeled as a raised level of this internal noise.
The simplest way of taking absolute threshold into account is to apply a linear correction to the power law (Humes and Jesteadt, 1991):

$$L = k \cdot (I - I_{ThQ})^{\alpha}, \tag{6.2}$$

where I_{ThQ} is the linear intensity of sound at absolute threshold. Here, the intensity at absolute threshold is simply subtracted from the stimulus intensity before compression, i.e., the correction appears before the nonlinearity. This way of accounting for threshold yields only a weakly increased slope near threshold. It more or less only yields a parallel shift of the whole function.
A second way of taking absolute threshold into account is to subtract the two intensities after they have been separately compressed (Humes and Jesteadt, 1991):

$$L = k \cdot (I^{\alpha} - I_{ThQ}^{\alpha}), \tag{6.3}$$

This is called the compressed internal noise model. This second way of accounting for threshold differs from the first one in that it yields a more strongly steepened loudness function near threshold.
The third way, proposed by Zwislocki (1965), is to calculate "overall intensity" (stimulus + internal noise) first and then to calculate the loudness in the same way as for the compressed internal noise model, by subtracting the compressed internal noise from the compressed total intensity (Humes and Jesteadt, 1991). It is assumed that the level of the internal noise is equal to I_{ThQ}:

$$L = k \cdot ((I + C \cdot I_{ThQ})^{\alpha} - C \cdot I_{ThQ}^{\alpha}), \tag{6.4}$$

The multiplicative constant C depends on both the bandwidth of the stimulus and the center frequency of the stimulus tone. Thus it can be modified in order to account for loudness summation or broadened auditory filters in hearing impaired subjects.

These models (Eq. 6.2, 6.3, 6.4) mainly differ for sound levels near absolute threshold but yield the same results for sound levels that are well above threshold. Moreover, Eq. 6.3 and 6.4 mainly differ in the resulting slope of loudness functions near threshold.
When dealing with loudness functions of hearing–impaired subjects, the first model yields quite different results than the second or the third: it is the only one of the three where the exponent of the power law must be increased in order to get steeper loudness functions, since elevation of threshold only yields a parallel shift of the loudness function, resembling a conductive loss. In the latter two models, Eq. 6.3 and 6.4, the steepening of teh loudness function is solely achieved by the increased compressed internal noise, producing a steeper than normal slope at low and mid levels, but normal slopes at high levels, thus nicely describing the well known "catching up" of loudness in (complete) recruiting ears. Without changing the exponent, these models cannot describe various other effects seen in impaired listeners, such as over–recruitment, partial recruitment and individual differences in loudness functions. Especially in cases with a large hearing loss, the models with compressed internal noise can overestimate the amount of recruitment (Launer et al., 1994), which varies greatly even between subjects with a similar amount of hearing loss (as has been discussed in chapter 4).
Hellman and Meiselman (1990) used Zwislocki's model (i.e., Eq. 6.4) quite successfully to describe loudness functions averaged across subjects, leaving the exponent of the power law constant and fitting only the constant C to the mean data.

6.4 Conclusions

In this section two extreme types of models have been discussed. Firstly, physiologically oriented models have been described. These have mainly been applied to describe and model different forms of rate–intensity functions in the auditory nerve of the normal cochlea. Furthermore, it was shown how the occurrence of steeper loudness functions, i.e., recruitment is explained by these models. Assuming, that overall loudness is proportional to the total output of the auditory nerve, then recruitment is caused by a linearization of BM mechanics. In other words, the processing of the Organ of Corti becomes less commpressive causing a steeper increase in neural firing rate. All physiological models have this property in common, i.e., model recruitment by less compressive inner ear mechanics. In chapter 2 it was pointed out that this is a major alteration often seen in injured cochleae. However, these models cannot be applied to predict measured loudness functions in impaired listeners, since up to now no relation between the model code for intensity, e.g., neural firing rate, and subjective loudness has been proposed.
The other extreme of models, the psychological models, yield this relation between sound level and subjective loudness. However, these models do not incorporate any physiological findings. Thus, it is difficult to extend them for modeling hearing impairment.
In the next chapter a loudness model, originally proposed by Zwicker (1960), and two different modifications of Zwicker's model to account for hearing impairment, will be discussed.

One might think of Zwicker's model as a "link" between the two extreme forms of models presented in this chapter.