Psychoacoustic Consequences of Compression in the Peripheral Auditory System

Brian C. J. Moore
University of Cambridge

Andrew J. Oxenham
Institute for Perception Research

Input-output functions on the basilar membrane of the cochlea show a strong compressive nonlinearity at midrange levels for frequencies close to the characteristic frequency of a given place. This article shows how many different phenomena can be explained as consequences of this nonlinearity, including the "excess" masking produced when 2 nonsimultaneous maskers are combined, the nonlinear growth of forward masking with masker level, the influence of component phase on the effectiveness of complex forward maskers, changes in the ability to detect increments and decrements with level, temporal integration, and the influence of component phase and level on the perception of vowel-like sounds. Cochlear hearing loss causes basilar-membrane responses to become more linear. This can account for loudness recruitment, linear additivity of nonsimultaneous masking, linear growth of forward masking, reduced temporal resolution for sounds with fluctuating envelopes, and reduced temporal integration.

Schematic illustrations of input-output functions are given in Figure 1. The solid line shows the type of function that would be observed for a tone at the characteristic frequency of the place being studied. Note that for midrange levels the output grows by less than 10 dB for each 10-dB increase in input level, indicating compression. This function can be approximated for midrange levels (30-70 dB) by a power function; the short-dashed line shows a power function with an exponent, $p$, of 0.25, meaning that a 10-dB range of input levels is compressed into a 2.5-dB range of response levels. The long-dashed line shows the type of function that would be observed for a tone with frequency well below the characteristic frequency. In this case, the response grows in a linear manner ($p = 1$); each 10-dB increase in input level gives rise to a 10-dB increase in response.

It is widely believed that the compression on the basilar membrane depends on an "active" physiological mechanism within the cochlea; it does not depend simply on the passive mechanical properties of the basilar membrane and surrounding structures (Yates, 1995). Running along the length of the basilar membrane are two types of receptor cells, which are called hair cells because of the hairlike stereocilia at their tips. The inner hair cells lie closest to the inside of the spiral-shaped cochlea and they form a single row. They are thought to be responsible for the detection of mechanical vibrations on the basilar membrane and for transforming these vibrations into action potentials in the auditory nerve. In other words, the inner hair cells act as transducers. Most, if not all, of the information carried by the auditory nerve is conveyed via the inner hair cells.

The outer hair cells lie closest to the outside of the spiral-shaped cochlea and they form three to five rows. The outer hair cells appear to play a crucial role in the active mechanism. They are able to change their length in response to electrical or chemical stimulation, and they appear to apply forces to the basilar membrane in such a way as to enhance the sharpness of tuning and to amplify the response to weak sounds. In other
Figure 1. The solid line shows a schematic illustration of an input-output function on the basilar membrane for a tone with frequency close to the characteristic frequency. The short-dashed line shows that a power function with exponent, \( p \), equal to 0.25 provides a reasonable approximation to the input-output function for midrange sound levels. The dashed line shows a linear input-output function (\( p = 1 \)). This is typical of what might be observed for a tone with frequency well below the characteristic frequency.

words, the outer hair cells play a motor role. Exactly how they do this is not fully understood, but many experiments have shown that damage to the outer hair cells causes dramatic changes in the mechanical responses of the basilar membrane to low-level sounds (Ruggiero, 1992; Ruggiero & Rich, 1991; Ruggiero, Rich, Robles, & Recio, 1996); the tuning is broadened, and the sensitivity is decreased.

Of particular interest for us in this article is that the compression on the basilar membrane depends on the active mechanism, mediated by the outer hair cells. The active mechanism amplifies the response to weak sounds but applies progressively less amplification as the sound level increases (Yates, 1995). The ordinate in Figure 1 has been chosen so that, for the tone at the characteristic frequency, an input level of 100 dB gives an output of 100 dB, consistent with the idea that the active mechanism plays little role at high levels (Yates, 1995). The "gain" for input levels around absolute threshold is about 55 dB, consistent with physiological estimates of the gain provided by the active mechanism at medium frequencies (Putzu, Yates, & Johnston, 1989; Yates, 1990, 1995).

The nonlinear processing probably does not have clear effects in simultaneous masking of a tone by broadband noise because the components of the noise that dominate the masking fall in the same frequency region as the signal and are subject to the same compression as the signal. In this case the compression does not change the effective signal:masker ratio, and this is the quantity that is usually assumed to determine the threshold in simultaneous masking (Fletcher, 1940; Patterson & Moore, 1986). However, the compression may influence the ability to detect changes in level (Moore, Peters, & Glasberg, 1996a), and it may affect simultaneous masking when the masker and signal frequencies are very different (Nelson & Schroder, 1996; Oxenham & Moore, 1995). Compression may have large effects when the signal and masker are presented nonsimultaneously, as in forward masking or backward masking. This is based on the assumption that the compression is fast acting, so that for nonsimultaneous stimuli each stimulus is independently compressed, even when they are close together in time; this is consistent with the physiological data (Yates, 1995). The compression may also have substantial effects on the internal representation of sounds, thus influencing timbre perception.

Cochlear hearing loss, also sometimes called sensory hearing loss or sensorineural hearing loss, is usually associated with damage to the hair cells in the cochlea. It appears that the outer hair cells are generally more susceptible to damage than the inner hair cells. Hence, it is relatively common to find cases in which the damage is mainly confined to the outer hair cells, but it is rare to find cases in which the damage is confined to the inner hair cells; often damage to both inner and outer hair cells occurs, with the latter predominating (Borg, Canlon, & Engström, 1995; Liberman, Dodds, & Learson, 1986; Moore, 1995b). Damage to the outer hair cells results in a loss of sensitivity and in a linearization of the basilar-membrane transfer function; essentially, the active mechanism appears to be rendered inoperative (Ruggiero, 1992; Ruggiero & Rich, 1991; Ruggiero et al., 1996). Hence, psychoacoustic phenomena observed for normal-hearing people that depend on the compression on the basilar membrane should be different or absent for people with cochlear hearing loss. Indeed, the differences between normal-hearing people and people with cochlear hearing loss can give insight into the importance of the compressive nonlinearity.

In this article we consider some of the psychoacoustic consequences of compression on the basilar membrane for normal-hearing people and the consequences of loss of the compression in people with cochlear hearing loss. We show that many apparently disparate phenomena can be explained in terms of this single underlying mechanism. In the work described below on people with cochlear hearing loss, it is assumed that a substantial part of the hearing loss is due to outer hair cell damage and the consequent loss of basilar-membrane compression. The people tested always demonstrated loudness recruitment, and this almost certainly is associated with outer hair cell damage (Moore & Glasberg, 1997).

The Additivity of Nonsimultaneous Masking

According to an energy detection model (Green & Swets, 1974), if two different maskers are equally effective (i.e., each produces the same amount of masking), then the combination of the two maskers should result in a doubling of the signal energy required for threshold; this corresponds to an increase in the signal level at threshold of 3 dB. In simultaneous masking, this prediction sometimes holds, but often the combined masker produces more masking than predicted (Lutfi, 1983; Moore, 1985). The amount by which signal thresholds exceed the energy summation prediction is referred to as excess masking. Various mechanisms have been put forward to account for the excess masking; for a review see Moore (1995a). However, the fact that excess masking does not always occur suggests that it cannot be explained in terms of a peripheral physiological process like compression.

A different argument can be made for nonsimultaneous masking. Two types of nonsimultaneous masking can be distinguished; in forward masking the signal is presented just after the masker, whereas in backward masking it is presented just before the masker. Combining two equally effective nonsimulta-
neous maskers (one forward and one backward) consistently results in excess masking, usually of 7–12 dB at moderate sound levels (Cokely & Humes, 1993; Wilson & Carhart, 1971). This excess masking can be explained if it is assumed that each stimulus (the forward masker, signal, and backward masker) is subjected to a compressive nonlinearity before the effects of the stimuli are combined in a linear temporal integrator (Penner, 1980; Penner & Shiffrin, 1980), as illustrated in the model of temporal resolution shown in Figure 2.

In the model, there is an initial stage of bandpass filtering, reflecting the action of the basilar membrane; each point on the basilar membrane behaves like a bandpass filter. For simplicity, only one filter is shown; in reality there is an array of parallel channels, corresponding to different places along the basilar membrane. Each filter is followed by a compressive nonlinearity that mimics the compression occurring on the basilar membrane. The output of the compressive nonlinearity is fed through a "smoothing" device often referred to as a temporal integrator. This sums the output of the compressive nonlinearity over a certain time interval or "window." Usually, the window is characterized as a weighting function such as an exponential or a pair of exponentials (Moore, Glasberg, Plack, & Biswas, 1988; Oxenham & Moore, 1994; Plack & Moore, 1990). However, the duration of the window can conveniently be described by its equivalent rectangular duration (ERD; Moore et al., 1988), which is typically about 7–11 ms. The window is assumed to slide in time so that the output of the temporal integrator is like a running average of the input. This has the effect of smoothing rapid fluctuations while preserving slower ones. Usually, the smoothing device is thought of as occurring after the auditory nerve; it is assumed to reflect a relatively central process. The output of the smoothing device is fed to a decision device. The decision device may use different "rules" depending on the task required. For example, if the task is to detect a brief temporal gap in a signal, the decision device might look for a "dip" in the output of the temporal integrator (Moore, 1997). For high-frequency signals, the response of the initial bandpass filter is very fast, and so the characteristics of the temporal integrator play a major role in limiting temporal resolution. For example, a gap in a signal may be smoothed by the temporal integrator leading to only a very small dip in the output of the temporal integrator.

Consider now the application of this model to excess masking. If two equally effective nonsimultaneous maskers (one forward and one backward) are presented together, then they will be compressed independently (as they do not overlap in time), and then their effects will be summed in the temporal integrator. The integrator itself is a linear device, and so the internal effect evoked by the two maskers is simply double the effect evoked by either alone. Thus, to reach threshold, the level of the signal has to be increased relative to the level required for a single masker. In fact, to reach the signal threshold, the internal effect of the signal must also be doubled. This requires more than a 3-dB increase in signal threshold because the signal itself is independently compressed.

This theoretical approach was taken by Oxenham and Moore (1994). In their work, the compression was modeled by a simple power function (as illustrated by the short-dashed line in Figure 1). A good fit to their forward and backward masking data was obtained when the rectified stimulus amplitude was raised to a power between 0.5 and 0.7 (depending on the participant). This corresponds to raising intensity to a power between 0.25 and 0.35. If, for example, the stimulus intensity is raised to the power 0.3, then a tenfold increase in power (corresponding to 10 dB) would be needed to double the internal effect of the signal. Thus, for two equally effective maskers, one forward and one backward, excess masking of 7 dB is predicted.

If excess masking is due to compression on the basilar membrane, and if cochlear hearing loss is usually accompanied by a loss of compression, then hearing-impaired participants should show less excess masking than normal-hearing participants. This hypothesis was evaluated in an experiment of Oxenham and Moore (1995, 1997). They tested 3 normal-hearing and 3 hearing-impaired participants. The latter were 73–77 years old and were diagnosed as having a moderate cochlear hearing loss with loudness recruitment; there was no evidence for any conductive component to their hearing loss. The hearing losses were about 30–40 dB at low frequencies, increasing to 55–65 dB at the 4-kHz test frequency.

Participants were tested in forward and backward masking conditions. The signal was a 4-kHz sinusoid with a 2-ms steady-state portion, gated with 2-ms raised-cosine ramps. The maskers (both forward and backward) consisted of low-pass filtered white noise with a cutoff frequency of 9 kHz. The maskers had a steady-state duration of 200 ms and were gated with 1-ms raised-cosine ramps.

Initially, the threshold of the signal in quiet was measured for each participant. Then the levels of the forward and backward maskers were varied to individually produce 5, 10, 15, 20, and 25 dB of masking. Forward masker levels were determined for masker–signal intervals of 5, 10, and 25 ms (defined as the interval between the 0-voltage points of the envelope of the electrical signal), and backward masker levels were determined for signal–masker intervals of 1 and 5 ms. Once the masker levels for each participant had been established, participants were presented with pairs of equally effective maskers (one forward and one backward), and signal thresholds in these combined masking conditions were measured.

The mean data, pooled across subject and condition, are shown in Figure 3. Mean thresholds in the presence of two

![Figure 2. A schematic model of temporal resolution in the auditory system.](image-url)
PSYCHOACOUSTIC CONSEQUENCES OF COMPRESSION

Equally effective maskers are plotted against thresholds in the presence of a single masker (in decibels sensation level [SL]). Thresholds for the hearing-impaired group are denoted by the open circles, and thresholds for the normal-hearing group are denoted by the filled circles. The error bars represent plus-and-minus one standard deviation. The solid line indicates no change in threshold in the presence of two maskers, and the short-dashed line indicates a 3-dB increase, as predicted by an energy detection model. Thus, the amount of excess masking is given as the amount by which the data points lie above this short-dashed line.

For the normal-hearing participants, the amount of excess masking increases with level, reaching 9 dB at levels of 20 and 25 dB SL. The long-dashed curve, which fits the data well, was calculated using a model of the type proposed by Humes and Jesteadt (1989). They modified a simple power function by including a constant to account for the absolute threshold. When applied to the case of two equally effective maskers, the equation becomes

\[ X_2 = 10 \log \left[ 2 \left(10^{X_1/p} - 1\right)^{1/p} \right], \]

where \( X_2 \) is the signal threshold (in decibels SL; that is, decibels relative to the absolute threshold) in the presence of two maskers: \( X \) is the signal threshold in the presence of a single masker, and \( p \) is the compressive exponent. A value of \( p \approx 1 \) corresponds to linear additivity (energy summation), and excess masking is predicted if \( p < 1 \). The best fitting value of \( p \) for the normal-hearing participants was 0.2, corresponding to a compression ratio of 5:1.

The data from the hearing-impaired participants showed little or no excess masking at any level. These results are consistent with the idea that excess masking results from compression in the cochlea and that loss of compression reduces or abolishes the excess masking. Note that the results for the hearing-impaired participants show roughly linear additivity of masking when intensity-like quantities are used. In other words, when compression on the basilar membrane is small or absent, the central temporal integrator appears to integrate an intensity-like quantity.

The Growth of Masking in Forward Masking

In forward masking, increments in masker level often do not produce equal increments in amount of forward masking. For example, if the masker level is increased by 10 dB, the masked threshold may only increase by 3 dB (Jesteadt, Bacon, & Lehman, 1982; Moore & Glasberg, 1983). This contrasts with simultaneous masking, for which, at least for wideband maskers, thresholds usually correspond to a constant signal:masker ratio. This effect can be quantified by plotting the signal threshold (in decibels) as a function of the masker level (in decibels). The resulting function is called a growth-of-masking function. In simultaneous masking such functions would have slopes close to one. In forward masking the slopes are less than one, and the slopes decrease as the time delay between the masker and signal increases. This is illustrated in Figure 4 (data are from Moore & Glasberg, 1983). The masker was a broadband noise, and the signal was a brief 4-kHz sinusoid presented at various times following the end of the masker.

Oxenham and Moore (1995) have suggested that the shallow slopes of the growth-of-masking functions can be explained, at least qualitatively, in terms of the compressive input–output function of the basilar membrane. Such an input–output func-
tion is shown schematically in Figure 5. It has a shallow slope for medium input levels, but a steeper slope at very low input levels. Assume that, for a given time delay of the signal relative to the masker, the response evoked by the signal at threshold is directly proportional to the response evoked by the masker. Assume, as an example, that a masker with a level of 50 dB produces a signal threshold of 12 dB. Consider now what happens when the masker level is increased by 20 dB. The increase in masker level, denoted by $\Delta M$ in Figure 5, produces a relatively small increase in response, $\Delta O$. To restore the signal to threshold, the signal has to be increased in level so that the response to it increases by $\Delta O$. However, this requires a relatively small increase in signal level, $\Delta S$, as the signal level falls in the range in which the input–output function is relatively steep. Thus, the growth-of-masking function has a shallow slope.

According to this explanation, the shallow slope arises from the fact that the signal level is lower than the masker level, so the masker is subject to more compression than the signal. The difference in compression applied to the masker and signal increases with increasing difference in level between the masker and signal. Hence the slope of the growth-of-masking function should also decrease with increasing difference in level between the masker and signal. This can account for the progressive decrease in the slopes of the growth-of-masking functions with increasing time delay between the signal and masker (see Figure 4); longer time delays are associated with greater differences in level between the signal and masker.

A prediction of this explanation is that the growth-of-masking function for a given signal time delay should increase in slope if the signal level is high enough to fall in the highly compressive region of the input–output function. Such an effect can be seen in the growth-of-masking function for the shortest delay time in Figure 4; the function steepens for the highest signal level. The effect is also apparent, but less clearly so, for the longer time delays of the signal.

Oxenham and Plack (1997) have investigated forward masking for a 6-kHz sinusoidal masker and a signal of the same frequency. They showed that under certain conditions, if the signal is made very brief and the time delay between the masker and signal is very short, the level of the signal at threshold is approximately equal to the masker level. Under these conditions, the growth-of-masking function has a slope of one; each 10-dB increase in masker level is accompanied by a 10-dB increase in signal level. This is illustrated in Figure 6 (filled symbols), and it is consistent with the explanation offered earlier. When a masker frequency well below the signal frequency was used (3 kHz instead of 6 kHz), the growth-of-masking function had a slope much greater than one; a 10-dB increase in masker level was accompanied by a 40-dB increase in signal level, as shown by the open symbols in Figure 6. This can be explained in the following way: The signal threshold depends on the response evoked by the masker at the signal characteristic frequency. The growth of response on the basilar membrane for tones well below the characteristic frequency is linear (see the long-dashed line in Figure 1). Thus, the signal is subject to compression whereas the masker is not (essentially the opposite of the situation illustrated in Figure 5). This gives rise to the steep growth-of-masking function. The slope value of around 4 corresponds to a compressive power exponent of about 0.25, which is in good agreement with the estimates of Oxenham and Moore (1994, 1995).
have slopes close to unity, except when the signal is very close to its absolute threshold. Furthermore, the slope should remain close to unity, regardless of the relative frequencies of the masker and signal, as all frequencies should be processed linearly. Empirical data have confirmed these predictions (Oxenham & Moore, 1995, 1997; Oxenham & Plack, 1997). This is illustrated in Figure 7, which shows individual data from 3 participants with moderate cochlear hearing loss in the same conditions as those used for the normal-hearing participants in Figure 6. In contrast to Figure 6, all 3 hearing-impaired participants in Figure 7 show linear growth-of-masking functions for both the 6-kHz masker and the 3-kHz masker. This is consistent with the view that cochlear damage results in a loss of basilar-membrane compression.

The Influence of Component Phase on the Effectiveness of Complex Forward Maskers

The envelope of a sound resembles a kind of "outline" of the sound. It may be determined graphically by drawing a smooth curve through the peaks in the fine structure of the waveform of the sound; for a more rigorous way of defining the envelope see Glasberg and Moore (1992). By manipulating the phases of the components in a complex sound, it is possible to create waveforms that differ markedly in their envelopes, even though the sounds have identical power spectra. Sounds differing in their envelopes may be processed differently by the auditory system, especially if the waveforms are subjected to fast-acting compression on the basilar membrane. However, in this case, the important factor is not the envelopes of the stimuli themselves but the envelopes after filtering on the basilar membrane.

Kohlrausch and Sander (1995) used complex tones containing many equal-amplitude harmonics of a 100-Hz fundamental. The phase, $\theta_n$, of the $n$th harmonic was determined according to equations described by Schroeder (1970):

$$\theta_n = +\pi(n + 1)/N \quad \text{or} \quad \theta_n = -\pi(n + 1)/N,$$

where $N$ is the number of harmonics in the complex sound. In this equation, if a plus sign is used (Schroeder-positive phase), the phase increases progressively with increasing harmonic number; whereas if a minus sign is used (Schroeder-negative phase), the phase decreases progressively with increasing harmonic number. Both of these phases give waveforms with almost flat envelopes; one waveform is a time-reversed version of the other; see Figure 8. The Schroeder-negative wave is like a repeated upward frequency sweep (up chirp), whereas the
Schroeder-positive wave is like a repeated downward frequency sweep (down chirp).

Kohlrausch and Sander (1995) presented evidence that the two waveforms give very different responses on the basilar membrane at a place tuned to 1100 Hz; the Schroeder-positive phase leads to a very "peaky" waveform, whereas the Schroeder-negative phase leads to a waveform on the basilar membrane for which the envelope remains nearly flat. They showed this by using the complex tones as maskers and by measuring the threshold for a brief 1100-Hz sinusoidal signal presented at various times during the 10-ms period of the masker. For the Schroeder-negative phase, the threshold of the signal varied little with the time of presentation. For the Schroeder-positive phase, the threshold varied over a considerable range, indicating that the internal representation of the masker was fluctuating within one period.

Carlyon and Datta (1997) took this experiment a stage further. They reasoned that if the waveforms were subjected to fast-acting compression, then the more peaky waveform (Schroeder-positive phase) should lead to lower average excitation in the auditory system than the less peaky waveform (Schroeder-negative phase). To test this idea, they used the two complex sounds as forward maskers. They reasoned that the amount of forward masking should provide an index of the average excitation evoked by each masker. For medium-to-high masker levels, the masker with Schroeder-positive phase gave more masking than the masker with Schroeder-negative phase, consistent with their prediction. However, at low levels, for which the input-output function on the basilar membrane is less compressive, the difference was absent.

These results confirm that the compression on the basilar membrane is fast acting, being able to influence the effective waveform within each 10-ms period of the stimulus. This is consistent with other psychoacoustic data (Moore, Wojtczak, & Vickers, 1996b); see also the experiments on vowel perception described later. Indeed, physiological data (Ruggiero, 1992; Ruggiero et al., 1996) suggest that the compression can act over time scales much shorter than 10 ms, operating on individual cycles of tones close to the upper frequency limit of hearing.

Detection of Increments and Decrements in Sinusoids

The compression on the basilar membrane can also influence the way that the ability to detect small changes in level is affected by overall level and frequency. Moore et al. (1996a) measured thresholds for the detection of brief decrements in level of sinusoidal signals as a function of decrement duration, level (70, 80, and 90 dB SPL), and frequency (250, 500, 1000, 2000, and 4000 Hz). Thresholds for detecting a 10-ms increment produced a change (at the output of the temporal integrator) equal to a criterion value, AO. The decrement detection thresholds plotted as a function of decrement duration and increment (Inc) detection thresholds for the single level, and frequencies away from the nominal frequency, an effect called spectral splatter. To prevent detection of the splatter, these authors presented the sinusoids in a background noise that would mask components away from the signal frequency. The noise also served a second purpose. When a sinusoid is changed in level, the response of the basilar membrane changes more at places tuned well above the signal frequency than at places tuned close to the signal frequency. This is another manifestation of cochlear nonlinearity; see Figure 1 and Oxenham and Moore (1995) and Nelson and Schroder (1996). To detect the change in level, participants monitor places on the basilar membrane tuned well above the signal frequency, a process known as off-frequency listening. The background noise in the experiment of Moore et al. would have prevented off-frequency listening, forcing the participants to monitor a place on the basilar membrane tuned close to the signal frequency.

Moore et al. (1996a) found that performance improved with increasing frequency for all decrement and increment durations, and it also tended to improve with increasing level—especially at 2000 and 4000 Hz. The results obtained for 4 participants at 4000 Hz are shown in Figure 9. The results were analyzed using a four-stage model similar to that shown in Figure 2. It consisted of a bandpass filter centered on the signal frequency, a compressive nonlinearity, a sliding temporal integrator, and a decision mechanism. It was assumed that threshold was reached when the increment or decrement produced a change (at the output of the temporal integrator) that equaled a criterion value, AO.

Initially, the results were analyzed by assuming a simple power-law compression; in such a model, the amount of compression does not decrease at high levels (see the short-dashed line in Figure 1). The improvement in performance with increasing level could, in principle, be accounted for by assuming that the equivalent rectangular duration (ERD) of the temporal integrator decreases with level, the value of AO decreases with level, or both (Moore et al., 1996a; Peters, Moore, & Glasberg, 1995). However, Moore et al. (1996a) found that the effects of level could be accounted for even better by assuming that the compressive nonlinearity was similar in form to the basilar membrane input-output function (as illustrated by the solid line in Figure 1). In this case, a very good fit to the data could be obtained when both the ERD of the temporal integrator and the value of AO were assumed to be invariant with level (although they could vary across frequency).

The goodness of fit of the model obtained in this way was
generally markedly better than obtained when a simple power-law nonlinearity was assumed and the ERD and ΔO were allowed to be different for each level. This happened despite the fact that the power-law analysis used five free parameters (one value of the ERD for each level and two parameters defining the variation of ΔO with level), whereas the analysis that was based on the basilar-membrane input–output function used only two (one value of the ERD and one value of ΔO). To fit the data at 250 Hz, Moore et al. (1996a) had to assume that the nonlinearity was somewhat less compressive than at higher center frequencies. This is consistent with physiological data (Neely, 1993; Patuzzi & Robertson, 1988; Patuzzi et al., 1989; Yates, 1990, 1995). However, the ERD was found to be about 7 ms, regardless of center frequency.

In summary, changes in increment and decrement detection with level and frequency can be accounted for using a relatively simple model in which the characteristics of the temporal integration process are invariant with level and frequency. The model uses a form for the compressive nonlinearity that is based on the basilar-membrane input–output function.

**Temporal Integration**

The absolute threshold for detecting a sound depends on the duration of the sound. For durations up to a few hundred milliseconds, the intensity required for threshold decreases as the duration increases. For durations exceeding about 500 ms, the sound intensity at threshold is roughly independent of duration (Florentine, Fasti, & Buus, 1988). Many researchers have investigated the relation between threshold and duration for tone pulses, over a wide range of frequencies and durations. The early work of Hughes (1946) and Garner and Miller (1947) indicated that over a reasonable range of durations, the ear appears to integrate the intensity of the stimulus over time in the detection of short-duration tone bursts. This is often called **temporal integration**. In other words, the threshold corresponds to a constant energy rather than to a constant intensity.

It seems very unlikely that the auditory system would actually integrate stimulus energy; it is almost certainly neural activity that is integrated (Penner, 1972; Zwislocki, 1960). It may also be the case that the auditory system does not actually perform an operation analogous to integration. Rather, it may be that the threshold intensity decreases with increasing duration partly because a longer stimulus provides more detection opportunities (more chances to detect the stimulus through repeated sampling). This idea is sometimes called **multiple looks** (Vieregger, 1991).

For people with cochlear damage, the change in threshold intensity with signal duration is often, but not always, smaller than it is for normal-hearing people. If the thresholds are plotted on decibel versus log-duration coordinates, the slopes are usually much less in absolute value than the typical value of −3 dB per doubling of signal duration found for normal-hearing people. This is often described as **reduced temporal integration** (Carlyon, Buus, & Florentine, 1990; Chung, 1981; Elliott, 1975; Gengel & Watson, 1971; Hall & Fernandes, 1983; Pedersen & Elberling, 1973). There is a trend for higher absolute thresholds to be associated with flatter slopes. In other words, the greater the hearing loss, the more reduced is the temporal integration.

A number of explanations have been advanced to account for reduced temporal integration in people with cochlear damage (Florentine et al., 1988). One of the more plausible of these is that it results from a reduction or complete loss of the compressive nonlinearity on the basilar membrane (Moore, 1995b). This leads to steeper rate-versus-level functions in the auditory nerve. According to the models of temporal integration proposed by Zwislocki (1960) and by Penner (1972), this will automatically lead to reduced temporal integration. Figure 10 illustrates schematically two rate-versus-level functions: The left-hand curve shows a typical function for a low-threshold neuron in a normal auditory system; the right-hand curve shows a typical function for a neuron in an auditory system with cochlear damage. The curve is shifted to the right, reflecting a loss of sensitivity, and is steeper, reflecting loss of the compressive nonlinearity on the basilar membrane. Note that it is assumed here that there is some residual compression on the normal basilar membrane at levels close to threshold; although the input–output function steepens at low levels (as illustrated in Figure 1), it does not become completely linear.

Assume that, to a first approximation, the threshold for detecting a sound requires a fixed number of neural spikes to be evoked by that sound. Assume also that the neurons involved in detection at absolute threshold are relatively homogeneous in terms of their input–output functions. The lower dashed horizontal line in Figure 10 indicates the number of neural spikes per second, N1, needed to achieve absolute threshold for a long-duration sound; in practice of course, the absolute threshold depends on the activity of many neurons, but if they are all similar, then the argument can be illustrated by considering just one. If the duration of the sound is decreased by a factor, R, then the level has to be increased to restore the total number of spikes evoked by the sound. Assume that the higher spike rate needed for the shorter duration sound is N2, where N2 = R × N1 — R

![Figure 10](image_url)  
**Figure 10.** Schematic illustration of rate-versus-level functions in single neurons of the auditory nerve for a normal ear (left curve) and an impaired ear (right curve). The horizontal dashed lines indicate the mean firing rate needed for threshold for a long-duration sound (lower line, rate N1) and a short-duration sound (upper line, rate N2) on the assumption that threshold corresponds to a fixed total number of spikes. See text for explanation of ΔL.
For example, if the duration is halved, the spike rate has to be increased by a factor of two to achieve the same total spike count. The increase in level, $\Delta L$, needed to achieve this increased rate is greater for the normal than for the impaired ear because the rate-versus-level function is steeper for the impaired ear, and this explains the reduced temporal integration.

If changes in basilar-membrane compression can affect temporal integration, then temporal integration should be dependent on level in normal-hearing participants, as basilar-membrane compression is thought to be maximal at medium sound levels. This idea was tested by Oxenham, Moore, and Vickers (1997) by measuring thresholds for detecting a 6.5-kHz sinusoid in the presence of a simultaneous noise masker. Consistent with predictions, maximum integration (i.e., the greatest rate of change of threshold with duration) was found at a medium masker level. However, the effect of level was only observed for signal durations less than 20 ms. This may mean that true integration occurs only over fairly short durations, up to 10 or 20 ms. By using the model described in the previous section, these authors found that the data could best be fitted by assuming that the basilar membrane was about four times as compressive at medium levels as at very high or low levels. Assuming that the basilar-membrane response is almost linear at very high and low levels, they derived a power exponent of 0.25 at medium levels, which is in good agreement with estimates of the maximum basilar-membrane compression obtained using other techniques, as described earlier (Oxenham & Moore, 1994, 1995; Oxenham & Plack, 1997).

Oxenham et al. (1997) used the model described in the previous section to fit their data for signal durations up to 10 ms. The mean data (symbols), together with the model predictions (solid lines), are shown in Figure 11. To achieve this fit, stimuli at the lowest and highest masker levels were assumed to be processed linearly while the stimuli at the medium masker level were assumed to be compressed using a power-law exponent of 0.25 before being integrated within the sliding temporal integrator. The model fits the data rather well; all points but one lie within one standard error of the mean data.

This approach is successful in accounting for changes in temporal integration at short durations. However, differences in temporal integration between normal-hearing and hearing-impaired participants over longer signal durations could not be described with the same model parameters. It is quite likely that there is some residual compression at very low and high levels in normal-hearing participants, which produces greater integration than for hearing-impaired participants. However, this aspect has not yet successfully been accounted for in quantitative terms.

Perception of Timbre in Vowellike Sounds

Under some circumstances it appears that the compression on the basilar membrane can have a marked effect on the perception of the timbre of complex sounds. This is especially the case when the waveforms of the sounds have high "peak factors" (the peak factor is the ratio of the peak value of the waveform to its root-mean-square value). If the waveform remains "peaky" after filtering on the basilar membrane, then the compression on the basilar membrane has the effect of increasing the salience of low-level portions of the waveform.

![Figure 11](image_url)

**Figure 11.** Data from Oxenham, Moore, and Vickers (1997) showing thresholds for detecting a 6.5-kHz sinusoid in a broadband noise as a function of signal duration. The noise spectrum level is given in each panel. Mean data from 4 normal-hearing participants (symbols) and model predictions (solid lines) are compared. Error bars represent plus-and-minus one standard error of the mean. SPL = sound pressure level.

Alcántara, Holube, and Moore (1996) examined the role of component phase and level on the identification of vowellike sounds. Normal-hearing participants were asked to identify which of six possible vowellike harmonic complexes was presented on each trial. The stimuli were complex tones containing the first 35 harmonics of a 100-Hz fundamental. All of the harmonics below 3000 Hz were equal in amplitude except for three pairs of successive harmonics, at frequencies corresponding to the first three formants of six vowels, which were incremented in level relative to the background harmonics by 1, 2, 4, 8, and 16 dB; the amount of increment was referred to as spectral contrast. The spectra of two of the stimuli are illustrated schematically in Figure 12.

The components in the harmonic complexes were added in
four different starting phase relationships: cosine, random, Schroeder-positive, and Schroeder-negative phases. For cosine phase, all of the components have a starting phase of 90° or $\pi/2$ radians. Schroeder phase (Schroeder, 1970) was described earlier; see Equation 2. These phases were chosen to give several distinct temporal patterns at the outputs of the basilar-membrane filters (i.e., the temporal pattern at a given place on the basilar membrane would be different for each of the four phases). The stimuli were presented at three overall levels: 85, 65, and 45 dB SPL.

The mean results (averaged across subjects and vowels) are shown in the upper panels of Figure 13. Performance was similar for the random and Schroeder-negative phases and did not vary as a function of level. Performance for the cosine and Schroeder-positive phase conditions was better than for the other two phase conditions but decreased as the level was reduced. Performance for all four phase conditions was equivalent for the lowest level.

Alcantara et al. (1996) argued that the good performance observed for the cosine-phase stimuli and Schroeder-positive-phase stimuli at high levels occurred because these stimuli produce waveforms on the basilar membrane with distinct low-amplitude portions. This is illustrated in Figure 14, which shows outputs from a basilar-membrane model (see below for details). The left column shows responses for “reference” cosine-phase stimuli, which have zero spectral contrast. When formant harmonics are incremented in level, this produces changes in the low-amplitude portions of the waveforms at points on the basilar membrane tuned close to the formant frequencies, as illustrated in the middle column of Figure 14.

The results were analyzed using a model with the following stages: (a) a bank of nonlinear basilar-membrane filters with level-dependent amplitude characteristics, a compressive input-output function, and a realistic phase response (Strube, 1985, 1986); (b) a threshold device (having an effect similar to rectification); (c) a square law device (as Oxenham and Moore, 1995, had found that the temporal integrator operates linearly on an intensity-like quantity); (d) a sliding temporal integrator; (e) an across-channel comparator; and (f) a decision device. It was assumed that performance depended on the difference in response between channels (corresponding to different places on the basilar membrane) responding mainly to the incremented harmonics and channels responding mainly to the background harmonics. This difference is shown in the right column of Figure 14. The maximum value of the difference was called the maximum internal contrast, and it corresponds to the peak values in the right column of Figure 14. If performance depends on the maximum internal contrast, then the results for all of the different conditions should collapse onto a single function if they are expressed as percentage correct versus maximum internal contrast. Figure 15 shows that this was indeed the case. The lower panels of Figure 13 show the predictions of this model for mean performance as a function of spectral contrast, overall level, and phase. The predictions correspond very well with the data in the upper panels.

Alcantara et al. (1996) found that the effects of level could not be accounted for using a linear basilar-membrane model as the initial stage; it was important to use a model with a realistic nonlinear input-output function. This suggests that the compressive nonlinearity on the basilar membrane plays an important role in producing the observed phase and level effects.

**Loudness Recruitment**

Most, if not all, people with cochlear hearing loss show a phenomenon referred to as loudness recruitment (Fowler, 1936; Steinberg & Gardner, 1937). The phenomenon may be described as follows. The absolute threshold is higher than normal. However, when a sound is increased in level above the absolute threshold, the rate of growth of loudness level with increasing sound level is greater than normal. When the level is sufficiently

![Figure 12. Schematic spectra of two voweilike sounds, showing the background and formant harmonics. The formant harmonics are indicated by dashed lines.](image)
high, usually around 90–100 dB SPL, the loudness reaches its "normal" value; the sound appears as loud to the person with impaired hearing as it would to a normal-hearing person. With further increases in sound level above 90–100 dB SPL, the loudness grows in an almost normal manner.

Two factors have been suggested to contribute to loudness recruitment. The first is a reduction in, or loss of, the compressive nonlinearity in the input–output function of the basilar membrane; this is usually associated with damage to the outer hair cells. In an ear where the damage is confined largely to the outer hair cells, with inner hair cells intact, the transformation from basilar-membrane velocity or amplitude to neural activity (spike rate) probably remains largely normal. However, in an ear with damage to the inner hair cells as well, the transduction process may also be affected.

If the input–output function on the basilar membrane is steeper (less compressive) than normal in an ear with cochlear damage, this would be expected to lead to an increased rate of growth of loudness with increasing sound level. However, at high sound levels, around 90–100 dB SPL, the input–output function becomes almost linear in both normal and impaired ears. The magnitude of the basilar-membrane response at high sound levels is roughly the same in a normal and an impaired ear (Ruggero & Rich, 1991). This could explain why the loudness in an impaired ear usually "catches up" with that in a normal ear at sound levels around 90–100 dB SPL.

Another factor that may contribute to loudness recruitment is reduced frequency selectivity; the filtering on the basilar membrane is less sharp in an impaired than in a normal cochlea (Sellick, Patuzzi, & Johnstone, 1982), and this can be measured behaviorally in humans as a broadening of the auditory filters (Glasberg & Moore, 1986; Pick, Evans, & Wilson, 1977). For a sinusoidal stimulus, reduced frequency selectivity leads to an excitation pattern that is broader (spreads over a greater range of characteristic frequencies) in an impaired than in a normal ear. Kiang, Moxon, and Levine (1970) and Evans (1975) suggested that this might be the main factor contributing to loudness recruitment. They suggested that, once the level of a sound exceeds threshold, the excitation in an ear with cochlear damage spreads more rapidly than normal across the array of neurons,

**Figure 13.** The upper panels show results of Alcántara, Holube, and Moore (1996). Percentage correct scores are plotted as a function of spectral contrast level (1, 2, 4, 8, and 16 dB) for the four phase conditions (cosine, random, Schroeder positive, and Schroeder negative). Results are shown for three different overall levels. Error bars indicate plus-and-minus one standard error across subjects. Error bars are omitted when they would be smaller than the symbol used. The lower panels show predictions of a model that are described in the text. SPL = sound pressure level; +ve = positive; —ve = negative.
and this leads to the abnormally rapid growth of loudness with increasing level. However, experiments using filtered noise to mask signal-evoked excitation at characteristic frequencies remote from the signal frequency indicate that reduced frequency selectivity is not a major contributor to loudness recruitment (Hellman, 1978; Hellman & Meiselman, 1986; Moore, Glasberg, Hess, & Birchall, 1985; Zeng & Turner, 1991); such noise has only a small effect on the perceived loudness of the signal. Although spread of excitation may play a small role, the loss of compressive nonlinearity on the basilar membrane seems to be a more dominant factor.

This conclusion is reinforced by a model of loudness perception applied to cochlear hearing loss, as proposed by Moore and Glasberg (1997). The model attempts to partition the overall hearing loss between the loss due to outer hair cell damage and the loss due to inner hair cell damage. The former is associated with loss of compression and reduced frequency selectivity, whereas the latter is associated simply with reduced sensitivity. The model accounts well for measures of loudness recruitment in people with unilateral cochlear hearing loss. It also accounts for the fact that changes in loudness with stimulus bandwidth (Zwicker, Flottorp, & Stevens, 1957) are smaller in hearing-impaired than in normal-hearing people.

Temporal Resolution for Sounds With Fluctuating Envelopes in People With Cochlear Hearing Loss

For certain types of sounds, the temporal resolution of people with cochlear hearing loss is worse than normal, even when the stimuli are well above threshold and when all of the components of the stimuli fall within the audible range. This happens mainly for stimuli that contain slow, random fluctuations in amplitude, such as narrow bands of noise. For such stimuli, people with cochlear damage often perform more poorly than normal in tasks.

Figure 14. Temporal waveforms at the output of a nonlinear basilar-membrane filter with a characteristic frequency of around 2150 Hz for a cosine-phase reference stimulus (left column) and a test stimulus with 4-dB spectral contrast (middle column). The right-hand column shows the variation in internal contrast as a function of time for two periods of the stimulating waveforms. The time axes are aligned with respect to one another. Each row shows results for one level. Note that the amplitude scale in each of the rows is linear and varies across rows. SPL = sound pressure level.
Figure 15. Results of Alcantara, Holube, and Moore (1996) showing the percentage correct for each condition (a specific combination of spectral contrast level, phase, and overall level) plotted as a function of the maximum internal contrast derived from the model. The phases are indicated as follows: c = cosine; r = random; s+ = Schroeder positive; s− = Schroeder negative.

such as gap detection (Buus & Florentine, 1985; Fitzgibbons & Wightman, 1982; Florentine & Buus, 1984; Glasberg, Moore, & Bacon, 1987). However, gap detection is not usually worse than normal when the stimuli are sinusoids, which do not have inherent amplitude fluctuations (Moore, 1997; Moore & Glasberg, 1988). It seems likely that the poor temporal resolution for sounds with fluctuating envelopes is a consequence of a loss of, or reduction in, the compressive nonlinearity on the basilar membrane (Glasberg et al., 1987; Moore & Glasberg, 1988). When compression is reduced, the inherent amplitude fluctuations in a narrowband noise would result in larger-than-normal loudness fluctuations from moment to moment (Moore et al., 1996b) so that inherent dips in the noise might be more confusable with the gap to be detected. It should be noted that most everyday sounds, including speech, fluctuate in level from moment to moment, so the poor temporal resolution found for such sounds probably leads to difficulties in everyday life.

To assess the idea that steeper input–output functions lead to impaired gap detection for stimuli with fluctuating envelopes, Glasberg and Moore (1992) processed the envelopes of narrow bands of noise so as to modify the envelope fluctuations. The envelope was processed by raising it to a power, N. For stimuli of constant amplitude (e.g., sinusoidal tones), the level (in decibels) of stimuli processed in this way is a linear function of the level in the unprocessed stimuli, with slope N. This reproduces one of the features of loudness recruitment seen in people with unilateral hearing loss; for levels below about 90–100 dB SPL, the level of sound needed in the normal ear to match the loudness of a sound in the impaired ear is roughly a linear function of the level in the impaired ear.

The effect of the processing on the envelope of a narrowband noise is shown in Figure 16 for values of N of 2 (squared envelope), 0.5 (square root of envelope), and 1 (unprocessed, normal envelope). If N is greater than unity, this has the effect of magnifying fluctuations in the envelope, thus simulating the effects of loss of compression or, equivalently, loudness recruitment; higher powers correspond to greater losses of compression. If N is less than unity, fluctuations in the envelope are reduced. This represents a type of signal processing that might be used to compensate for recruitment; it resembles the operation of a fast-acting compressor or automatic gain control (AGC) system.

Values of N used in the experiment were 0.5, 0.66, 1.0, 1.5, and 2. A value of N of 2 simulates the type of recruitment typically found in cases of moderate-to-severe cochlear damage, for which, for example, a 50-dB range of stimulus levels gives the same range of loudness as a 100-dB range of stimulus levels in a normal ear. To prevent the detection of spectral splatter associated with the gap or with the envelope processing, the stimuli were in a continuous background noise. The spectrum of the noise was chosen so that it would be as effective as possible in masking the splatter while minimizing its overall loudness.

Figure 17 shows an example of results obtained using a participant with unilateral hearing loss of cochlear origin, aged 74. The hearing loss in his impaired ear was reasonably uniform across frequency and was 55 dB at the 1-kHz test frequency. The stimuli were presented at 85 dB SPL, a level that was well above the absolute threshold for both the normal and impaired ear.
ears (although the sensation level was lower in the impaired ear). The results for the normal ear were very similar to those of 3 normal-hearing participants who were also tested. Gap thresholds increased significantly with decreasing noise bandwidth. This is as expected because the inherent fluctuations in the noise are slower, and more confusable with the gap to be detected, when the bandwidth is narrow (Moore, 1997; Shailer & Moore, 1983).

For all noise bandwidths, gap thresholds increased as $N$ increased. This effect was particularly marked for the smaller noise bandwidths. There was a significant interaction between bandwidth and $N$, reflecting the fact that changes in gap threshold with $N$ were greater for small bandwidths. This supports the idea that fluctuations in the noise adversely affect gap detection; greater fluctuations lead to worse performance, especially when the fluctuations are slow.

Gap thresholds were larger for the impaired than for the normal ear. The overall geometric mean gap threshold was 12.8 ms for the normal ear and 27.2 ms for the impaired ear. Performance for the normal ear with $N = 2$ was roughly similar to performance for the impaired ear with unprocessed noise bands ($N = 1$); geometric mean gap thresholds were 26.9 ms for the former and 26.5 ms for the latter. Thus, the simulation of loss of compression in the normal ear was sufficient to produce impaired gap detection, comparable to that actually found in the impaired ear.

Reduction of fluctuations, by raising the envelope to a power less than 1, produced a small improvement in performance for the normal ear. Values of $N$ less than 1 gave a marked improvement for the impaired ear. Performance for the impaired ear with the envelope raised to the power 0.5 was comparable to, or even slightly better than, that for the normal ear with unprocessed stimuli; geometric mean thresholds were 11.6 ms for the former and 12.5 ms for the latter. Thus, the impaired detection of gaps in noise bands occurring for an impaired ear can be restored to normal by appropriate compression of fluctuations in the envelopes of the stimuli.

For both normal-hearing and hearing-impaired participants, the effects of changing $N$ decreased with increasing noise bandwidth. One reason for this is that slow fluctuations can be followed by the auditory system, whereas rapid fluctuations are smoothed to some extent by the central temporal integration process described earlier. The rapid fluctuations of the wider noise bands are smoothed in this way, thus reducing their influence on gap detection.

The results suggest that for most people with cochlear damage, a reduction in the peripheral compressive nonlinearity may provide a sufficient explanation for increased gap thresholds. Thus, it is not usually necessary to assume any abnormality in temporal processing occurring after the cochlea.

Conclusions
The basilar membrane in the normal cochlea shows a strong compressive nonlinearity for tones close to the characteristic...
frequency at midrange levels. In this article we have reviewed some of the psychoacoustic consequences of this nonlinearity, and of the loss or reduction of the nonlinearity in people with cochlear hearing loss.

Many of the properties of nonsimultaneous masking may be accounted for by the compressive nonlinearity. When two equally effective nonsimultaneous maskers are combined, the resultant masking is usually markedly greater than would be predicted from a simple intensity summation of the effects of the individual maskers, an effect referred to as excess masking. This can be accounted for by assuming that stimuli are compressed before being combined in a linear temporal integrator. In hearing-impaired people, excess masking is reduced or absent, probably because the compression on the basilar membrane is absent.

The nonlinear growth of forward masking can partially be accounted for by assuming that the masker usually falls in a range of levels where the basilar-membrane response is strongly compressive, whereas the signal usually falls in a lower range where the basilar-membrane response is less compressive. The growth of masking becomes more linear when the masker and signal levels are nearly equal, as can be achieved by using very brief signals, close in time to the masker. The amount of forward masking produced by a harmonic complex masker can be altered by manipulating the phases of the components, holding the power spectrum constant. This can be explained by assuming that the waveform on the basilar membrane is subjected to fast-acting compression. Phases leading to peaky waveforms lead to less average excitation and less forward masking than phases leading to waveforms with relatively flat envelopes.

The results of tasks measuring temporal resolution, such as the detection of brief decrements and increments, can be accounted for using a nonlinearity resembling the form of the input–output function on the basilar membrane. The variation in performance with level and frequency can be explained in this way without assuming any variation with level of the central processes involved.

The influence of component phase and level on the perception of vowel-like sounds can be accounted for using a basilar membrane model incorporating a compressive nonlinearity. More generally, the timbre of sounds with strongly fluctuating envelopes (on the basilar membrane) is influenced by the compression, as the low-amplitude portions are made more salient.

The loss of compressive nonlinearity in people with cochlear hearing loss may be responsible for many of the perceptual difficulties that they encounter in everyday environments. It appears to be the main cause of loudness recruitment. It also contributes to reduced changes in loudness with stimulus bandwidth and to reduced temporal integration. Finally, it can also result in impaired temporal resolution for sounds with fluctuating envelopes.

References


Schroeder, M. R. (1970). Synthesis of low peak-factor signals and bi-


Received December 11, 1996
Revision received March 5, 1997
Accepted March 18, 1997