Notes on Auditory Sensitivity.

Pulstation Thresholds

Louis D. Braida
Sensory Communication Group
Research Laboratory of Electronics
and
Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

July 2, 2013
1 Introduction

In this section we discuss some terminology.

1.1 Sound Level

A sinewave having a pressure waveform

\[ p(t) = \sqrt{2} P \cos (2\pi ft + \theta) \]

has an rms pressure \( P \) (Newtons per square meter, or Pascals).

The rms sound pressure is conveniently described in terms of the Sound Pressure Level (SPL) as the ratio of \( P \) to a reference pressure \( P_0 \) of \( 2 \times 10^{-5} \) newtons per square meter, 20\( \mu \) Pascals, or 0.0002 dynes per square centimeter. This is conveniently measured on a logarithmic scale (hence the term Level) in decibels (dB):

\[ \text{dB SPL} = 20 \log_{10} \frac{P}{P_0} \]

The intensity, \( I \), of a sound is the rate of energy flow across a unit area and is measured in watts per square meter. The Sound Intensity Level (SIL) of a sound is the logarithm (again the term Level is used to emphasize that this is a logarithmic measure) of the ratio of \( I \) to \( I_0 \), where \( I_0 = 10^{-12} \) watts per square meter. This is conveniently measured in decibels:

\[ \text{dB SIL} = 10 \log_{10} \frac{I}{I_0} \]

The reference for sound pressure, \( P_0 \) is compatible with the reference for sound intensity \( I_0 \) in air, thus

\[ 20 \log_{10} \frac{P}{P_0} = 10 \log_{10} \frac{I}{I_0} \]

1.2 Sensation Level

The detection threshold of a sound is often a convenient reference for measuring sound pressures. Not only are many phenomena simpler to describe when expressed relative to detection thresholds, but is is fairly straightforward to measure, unlike absolute pressures or intensities.

If the detection threshold of a sound has an rms pressure \( P^* \) the Sensation Level (SL) of the sound is defined as the logarithm (note again the term Level) of the ratio of the pressure of the sound, \( P \) to \( P^* \). This is conveniently measured on in decibels:

\[ \text{dB SL} = 20 \log_{10} \frac{P}{P^*}. \]
It is easy to show that
\[ \text{dB SL} = 20 \log_{10} \frac{P}{P_0} - 20 \log_{10} \frac{P^*}{P_0}. \]

In other words, the Sensation Level of a sound (in dB) is the difference between the Sound Pressure Level (in dB) of the sound and the Sound Pressure Level (in dB) of the sound at threshold.

### 1.3 Spectra

#### Periodic Signals

A periodic pressure waveform
\[ p(t) = \sum_{k=1}^{\infty} \sqrt{2} P_k \cos(2\pi k f_0 t + \theta_k) \]

with fundamental frequency \( f_0 \) and period
\[ T = \frac{1}{f_0} \]

is said to have a “line power spectrum” with components \( P_k^2 \) at frequencies \( k f_0 \). The power in the signal between frequencies \( f_A \) and \( f_B \) is
\[ P(f_A, f_B) = \sum_{k=k_A}^{k_B} P_k^2 \]

where \( k_A \) is the largest \( k \) for which \( k f_0 \leq f_A \) and \( k_B \) is the smallest \( k \) for which \( f_B \leq k f_0 \).

#### Noise

Noise is often modelled as a random process with a well defined, continuous, non-negative power spectrum, \( N(f) \geq 0 \) and a random phase spectrum. The power spectrum of the noise process \( n(t) \) is defined so that if \( n(t) \) is filtered by an ideal bandpass filter that passes frequencies \( f_A \leq f \leq f_B \), the power \( Q_{\text{out}} \) in the process that emerges from the output of the filter is
\[ Q_{\text{out}} = \int_{f_A}^{f_B} N(f) df \quad (1) \]

Noise that has the property that \( N(f) = N_0 \) is said to be white noise with spectrum level \( N_0 \). This terminology is used even when the noise bandwidth is finite, provided \( N(f) = N_0 \) over the frequency range of interest.

A constant- or flat-spectrum noise of spectral density \( N_0 \) over the frequency range \( f_A \) to \( f_B \) thus has total power
\[ Q = N_0 (f_B - f_A) \]
The units of $Q$ are watts. $Q$ is more often expressed in terms of dB SIL

$$\text{dB SIL} = 10 \log_{10} \frac{Q}{I_0} = 10 \log_{10} \frac{N_0 (f_B - f_A)}{I_0}$$

If we filter noise with an ideal bandpass filter of width $\Delta = f_B - f_A$ and measure the power that comes out of the filter, $Q_{\text{out}} (f_A, f_B)$, we can estimate

$$N \left( \frac{f_A + f_B}{2} \right) \approx \frac{Q_{\text{out}} (f_A, f_B)}{\Delta}$$

with the approximation getting better as $\Delta \to 0$. The Spectrum Level (again note the use of the term Level) of the noise is defined to be

$$S_X (f) = 10 \log_{10} \left( \lim_{\Delta \to 0} \frac{Q_{\text{out}} (f, f + \Delta) / \Delta}{I_0 / \Delta} \right).$$

Since

$$Q_{\text{out}} (f, f + \Delta) \approx N (f) \Delta$$

$$S_X (f) = 10 \log_{10} \left( \frac{N (f)}{I_0 \Delta_0} \right)$$

Thus for example, when Hawkins and Stevens (1950) cite the noise level as -20 dB (per cycle or in modern terminology per Hz) they mean

$$-20 = 10 \log_{10} \left( \frac{N (f)}{I_0 \Delta_0} \right)$$

or

$$N (f) \Delta_0 = 10^{-2} I_0$$

where $\Delta_0 = 1$ Hz. Because the noise spectrum is flat, extending to 10,000 Hz, the total power in the noise is

$$10^4 N_0 = 10^2 I_0$$

which corresponds to an rms pressure of 20 dB SPL.
2 Absolute Detection of Sound

The minimum pressure (or equivalently, minimum intensity) of a sound stimulus that can be heard by a given individual under specified listening conditions is said to specify the detection threshold for those conditions. When listening in perfectly quiet conditions, the sound level is said to be at absolute threshold or the threshold of audibility or the absolute detection threshold for the sound under those conditions.

2.1 Methods

The minimum detectable pressure can be measured using a variety of methods. Some of these methods are objective, with a correct response known by the experimenter. Some are subjective, without a correct response.

Method of Adjustment

In the method of adjustment the listener controls the pressure of a sound and adjusts it to determine the minimum audible pressure. Typically the listener brackets the threshold of the sound, i.e., determining a pressure for which the sound is clearly audible and a pressure for which the sound is clearly inaudible. Then the listener reduces the range of bracketing until a single pressure is determined that is “just audible”.

Method of Limits

In the method of limits, commonly used in clinical audiometry, the minimum audible pressure is determined by presenting a sound of varying pressure and the listener is asked to indicate whether he/she “heard” the sound or not. The pressure of the sound is either systematically decreased (descending limits) or systematically increased (ascending limits) from presentation to presentation. In descending limits the sound pressure is sufficiently high that the listener can hear it, and the pressure is decreased at fixed intervals until the listener responds “No” to the question, “Do you hear the sound?” The “descending threshold” for a given test is taken to be the pressure at which the responses change from “Yes” to “No”. In ascending limits the pressure of the sound is sufficiently low that the listener cannot hear it and then increased at fixed intervals until the subject responds “Yes,” determining the ascending threshold.

The ascending threshold is usually higher than the descending threshold. Ascending and descending thresholds are usually averaged to determine a single estimate of the detection threshold. The rationale for this is that when pressures are presented in descending order, there is a bias for the response “Yes”. When the tone pressures are presented in ascending order, there is a bias for the response “No”. Therefore, averaging the ascending and descending thresholds tends to cancel the biases.
Bekesy Audiometry

In a variant that combines the above two methods, *Bekesy audiometry*, is used to measure the threshold of hearing for tones. The tone frequency is swept continuously and the listener controls the pressure of tone with a button. The pressure is reduced when the button is depressed and increased when the button is released. Usually there is a difference of 5–10 dB between the points at which the button is pressed and released, and an average of these two points is taken as threshold.

Method Of Constant Stimuli

In the method of constant stimuli, the minimum pressure is measured by presenting the sound or no sound with equal probabilities on successive trials of an experiment and determining the pressure required to achieve a specified proportion of correct responses (e.g., 75%).

Objective Adaptive Methods

In objective adaptive methods, the minimum pressure is typically measured by presenting the sound or no sound with equal probabilities on successive trials of an experiment. The pressure required to achieve a specified proportion of correct responses is determined by varying the stimulus level. The stimulus presented on a given trial is determined by the listener’s responses on previous trials. The rules for adjusting the stimulus are intended to achieve a specified level of performance on the part of the listener.

2.2 Study of Watson, Franks and Hood (1972)

In one study [33], tone pulses were presented under headphones to listeners with “clinically normal” hearing. A two-alternative temporal forced-choice experiment was performed. On each trial of the experiment, two 150 ms light flashes were presented, separated by 500 ms. During one of the light-flash intervals, selected at random, a fixed-frequency tone was presented, during the other there was no tone. Listeners were asked to specify the interval in which a tone was presented. They were trained for 10 hours before data were collected. Feedback was not provided on a trial-by-trial basis, but listeners were told their scores at the end of each 100 trial block.

The data consist of percentage correct scores. Psychometric functions were fitted to the scores in the range 60–99% correct (Fig. 1) and averaged (parametrically) to establish each psychometric function. The psychometric functions which relate proportion of correct responses to sound pressure were similar in shape for all frequencies (Fig. 2). The absolute threshold is typically taken as the value of sound pressure leading to a specified percentage of correct responses (e.g., 75%). The positions of the psychometric functions on the sound pressure axis (corresponding to absolute threshold values) were found to be strongly dependent on frequency, but the relative positions were not very dependent on the proportion of correct responses selected.
2.3 The Physical Audiogram

Absolute thresholds for tones are typically plotted as a function of frequency to form a physical audiogram. Fig. 3 presents averages of measured thresholds measured under carefully calibrated earphones for young adult listeners with “normal” hearing. Note that, while strictly speaking there are no limits on the ability to hear low or high frequency sounds, the thresholds of hearing typically exceed 100 dB for tones below 20 Hz and tones above 20 kHz. Also note that there is considerable variability in the threshold of hearing across listeners. At 8 kHz, for example, the standard deviation in the threshold is roughly 9 dB. Since the distribution of thresholds is roughly normal, this implies that 95% of these “normal hearing” listeners have hearing thresholds at 8 kHz that lie in a 36 dB band about 19 dB SPL, while 5% lie outside that band.

2.4 MAF and MAP

Measurements of the minimum detectable sound pressure at the listener’s eardrum are generally called Minimum Audible Pressures. Measurements of the minimum detectable sound pressure made in sound fields, called Minimum Audible Fields, refer to the sound pressures at the location of the center of the head when the head is absent.

Sivian and White (1933) measured MAFs using the following procedure. The tone was on for 2 s and off for 2 s. The listener depressed a key whenever (and as long as) the tone was
audible. The experimenter reduced sound in 5 and 1 dB steps until threshold was determined. The experimenter judged from the ability of the listener to follow the interruptions with the key whether the tone was heard. Pure tones were used to measure 100-800 Hz thresholds, warble (±50 to ±146 Hz at a 10 Hz rate) tones were used for 1100-15000 Hz thresholds. The sound was presented from a loudspeaker located 1 m in front of the listener in highly sound absorbing structure. Sivian and White reviewed a wide variety of measurements of MAPs, rather than measured them themselves.

Sivian and White (1933) found that MAFs are typically lower than MAPs by roughly 6-10 dB, e.g., Fig. 4.\(^1\) The difference between MAP and MAF thresholds is thought to reflect acoustic diffraction effects (in the field), the masking effects of physiological noise (under earphones), and differences related to monaural vs. binaural listening.

### 2.5 Audiogram Microstructure

Audiograms are typically measured at fairly widely separated frequencies, e.g., at octave or half-octave intervals. Thresholds measured with Bekesy Audiometry on a more fine grained frequency scale, called the micro-audiogram, (Fig. 5) can show systematic variations of 10–20 dB over a frequency range of 100-200 Hz. This variation is not seen at all frequencies, nor in all listeners with normal hearing. Audiograms with such microstructure are generally not seen in listeners with sensorineural hearing impairment.

### 2.6 Variability

According to Fig. 3 at a given frequency 95% of the hearing thresholds for young adult males lie within roughly \(\pm 15 – 20\) dB of the mean value for the group. This variability reflects

\[^1\]Note that 1 bar corresponds to \(10^5\) Pa, or \(5 \times 10^9\) times the SPL reference (0.0002 dynes/sq-cm) so that 1 bar corresponds to roughly 194 dB SPL.
Figure 3: Absolute hearing thresholds for tones of young adult listeners (solid lines). Low-frequency thresholds [36] are average values for 10 listeners. Mid-frequency thresholds [2] are median values for 99 men (198 otologically normal ears) in the age range 18-25 y. High-frequency thresholds [8] are average values for 37 listeners aged 18-26 y who were required to have thresholds of 15 dB HL or less at audiometric frequencies of 8 kHz and less. Only 15 listeners contributed data to the 20 kHz threshold. The dotted line is the standard deviation of the thresholds across listeners.

- **Within listener factors**
  - The intrinsic variability of measurements based on psychometric functions.
  - Temporal variation in hearing sensitivity.

- **Across listener factors**
  - Differences in audiogram fine structure
  - Differences in listening strategies
  - Differences in listener basic acuity
Figure 4: Minimum Audible Pressure and Minimum Audible Field thresholds (dB SPL) reported by Sivian and White (1933).

Fig. 92.—Hearing Acuity for Typical Young Observers.
Figure 5: Measurements of the microstructure of the audiogram reported by Long et al. (1984). Open circles represent the results of a ultra fine grained threshold measurements made using Bekesy audiometry. Filled circles represent coarser grained measurements made with a 2I-2AFC technique. Note that while the measurements made via Bekesy Audiometry agree closely with the 2I-2AFC estimates of threshold for listener SN, who does not show audiogram microstructure over the frequency range 1000-1300 Hz, they are up to 10 dB higher in the case of the other listeners, who show the microstructure over the frequency ranges indicated.
3 Discrimination of Sound Differences

Two sounds that can be distinguished with a specified degree of reliability are said to be discriminable. If the sounds differ only in the value of a single physical parameter (e.g., intensity, frequency, duration) the difference in the values of the parameter across the two sounds is often called the Just Noticeable Difference or JND (sometimes the Difference Limen or DL) in the parameter. In current parlance, listener’s are said to have a finite resolution for the parameter.

3.1 Intensity

The ability of listeners to discriminate differences in sound pressure, or equivalently the intensity of sound, has been studied systematically for at least 80 years. Studies of intensity discrimination have utilized different psychoacoustic techniques. The “best method” is still somewhat elusive.

Consider two stimuli that differ only in rms pressure:

\[ p_1(t) = Pu(t) \]
\[ p_2(t) = (P + \Delta P)u(t) \]

where \( u(t) \) has unit rms pressure and \( \Delta P > 0 \). The equivalent intensities are proportional to the square of pressures:

\[ I_1 \propto P^2 \]
\[ I_2 = I_1 + \Delta I \propto (P + \Delta P)^2. \]

In the case where \( \Delta P \) is small relative to \( P \)

\[ \Delta I \propto 2P \Delta P. \]

Measurements of intensity discrimination are typically reported as

\[ \frac{\Delta I}{I} = \frac{(P + \Delta P)^2 - P^2}{P^2} \approx \frac{2 \Delta P}{P}, \]

or as

\[ \frac{I + \Delta I}{I} = 1 + \frac{\Delta I}{I}. \]

3.2 Riesz and the Method of Just-Noticeable Beats

The first systematic attempts to measure discrimination for sound pressure were made by R.R. Riesz (1928). Possibly for reasons of simplicity of generating the pressures, Riesz produced pressure variations by adding two sinusoidal voltages and then applying the sum to an earphone. Riesz’s stimulus can thus be described as

\[ p(t) = P_1 \cos(2\pi f_1 t) + P_2 \cos(2\pi f_2 t) \]
which Riesz expressed (in the case $P_1 \gg P_2$) as

$$p(t) = m(t) \cos(2\pi f_1 t + \Phi),$$

where

$$m^2(t) = P_1^2 + P_2^2 + 2P_1P_2 \cos(2\pi (f_2 - f_1) t)$$

In Riesz’s experiments $\Delta f = f_2 - f_1$ was small compared to $f_1$ or $f_2$ so that the listener heard a tone-like stimulus having a frequency of very nearly $f_1$ whose amplitude fluctuates periodically between a maximum of $P_1 + P_2$ and a minimum of $P_1 - P_2$ at a rate of $\Delta f = f_2 - f_1$ beats per second.

Riesz had his subjects determine the minimum value of $P_2$ that caused the beats to be audible and reported his results in terms of the difference between the tone intensities correspondingly to the maximum and minimum tone pressure amplitudes:

$$\Delta I = \frac{(P_1 + P_2)^2 - (P_1 - P_2)^2}{(P_1 - P_2)^2} = 4 \frac{P_1P_2}{(P_1 - P_2)^2}$$

which, in the case $P_2 \ll P_1$ is approximately

$$\frac{\Delta I}{I} \approx 4 \frac{P_2}{P_1}$$

Riesz reported his results in terms of

$$\Delta \alpha = 10 \log_{10} \left(1 + \frac{\Delta I}{I}\right).$$

As shown in Fig 6, Riesz found that for 1000 Hz tones, the minimum value of $\Delta \alpha$ corresponding to audible beats depended on $\Delta f$ as well as on $P_1$ expressed in sensation level (i.e. in dB relative to the threshold SPL for 1000 Hz tones). Clearly, while $\Delta \alpha$ depends on $P_1$, there is a minimum value of $\Delta \alpha \approx 1$ dB that occurs at $\Delta f \approx 3$ Hz, roughly independent of the value of $P_1$.

While holding the rate of beating constant at 3 beats/sec, Riesz determined the differential sensitivity, the value of $\Delta \alpha$, systematically as a function of the intensity (expressed in terms of sensation level) and the frequency of the more intense tone (Fig. 7). Generally, $\Delta \alpha$ decreases as the value of intensity increases, with most of the decrease occurring in the first 20 dB above threshold. The function relating the value of $\Delta \alpha$ to intensity is not a strong function of frequency $f$ when intensity is expressed in sensation level.

---

2Intensity is proportional to the square of pressure.

3Riesz used $E$ for intensity $I$. 
Figure 6: Dependence of the minimum perceptible change in the intensity of 1000 Hz tones, \( \Delta \alpha = 10 \log_{10} (1 + \Delta I/I) \), as measured by beat detection, on beat rate at \( P_1 \) levels of 25 and 50 dB SL. From Riesz (1928).

**Controversy**

Riesz’s results are not without controversy, however. He determined the conditions under which a listener was able to distinguish between a pure tone

\[
p_1(t) = A \cos \omega t
\]

where \( \omega = 2\pi f \) and a tone to which a second tone of different frequency was added:

\[
p_2(t) = A \cos \omega t + B \cos (\omega + \delta) t
\]

where \( \delta = 2\pi \Delta f \). While \( p_2(t) \) can exhibit beats for small \( \delta \) it is not a purely amplitude modulated tone, such as \( p_3(t) \):

\[
p_3(t) = A \cos \omega t + \frac{B}{2} \cos (\omega - \delta) t + \frac{B}{2} \cos (\omega + \delta) t
\]

In particular, \( p_3(t) \) has zero crossings that occur with the same period as \( p_1(t) \),

\[
p_3(t) = A \left( 1 + \frac{B}{A} \cos \delta t \right) \cos \omega t
\]

whereas \( p_4(t) \) does not (e.g., Fig. 8).

In the case \( B \ll A \) it can be shown that

\[
p_2(t) \approx A [1 + \frac{B}{A} \cos \delta t] \cos \left( \omega t + \frac{B}{A} \sin \delta t \right)
\]
so that \( p_2(t) \) has a time varying amplitude

\[
A[1 + \frac{B}{A} \cos \delta t]
\]

and a time varying frequency

\[
\omega + \frac{B}{A} \delta \cos \delta t.
\]

The ratio of the maximum to minimum amplitudes is

\[
\frac{A + B}{A - B} = \frac{1 + B/A}{1 - B/A} \approx 1 + 2 \frac{B}{A}
\]

and the ratio of the maximum to minimum instantaneous frequencies is

\[
\frac{\omega + B\delta/A}{\omega - B\delta/A} \approx 1 + 2 \frac{B \delta}{A \omega}.
\]

Thus in discriminating \( p_2(t) \) from \( p_1(t) \), the listener can base judgments on both amplitude and frequency changes.
Weber’s Law and Deviations

Weber, a 19th century German physicist, measured the ability to distinguish between lifted weights. He concluded that there was a range of weights $W_{\text{min}} \leq W \leq W_{\text{max}}$ for which the “just noticeable difference” (JND, $\Delta W$) between two weights $W$ and $W + \Delta W$ was proportional to the base weight, $W$, i.e.

$$\frac{\Delta W}{W} = kW$$

A similar relationship has been found for many “intensive” physical properties (vibration amplitude, light intensity, ...) provided $W > W_{\text{min}}$, where $W_{\text{min}}$ is typically well (e.g., 20 dB) above the absolute detection threshold for the stimulus.

The JND in intensity for wideband stimuli such as broadband noise has been found to be reasonably well described by Weber’s Law at high intensities.

Miller (1947) presented two listeners with a continuous flat spectrum noise to which 25 increments (1.5 s duration) were added at intervals of 4.5 s. The subjects were asked whether they could hear each increment and the percentage “heard” was tabulated. Four such series were performed for each intensity increment and 5–8 increments were measured for each intensity. The DL was obtained by linear interpolation as that increment that listener could hear 50% of time.

Miller (1947) determined that his measurements of the DL for the intensity of white noise (Fig. 9) could be well described by

$$\frac{I + \Delta I}{I} = 1 + C + \frac{I_0}{I}$$

where $I$ is the intensity of the noise, $I_0$ is the detection threshold of the noise, $C$ is a constant independent of $I$ and $\Delta I$ is the DL for intensity. Clearly, at sufficiently high intensities,

$$\frac{I + \Delta I}{I} \to 1 + C$$

or

$$\frac{\Delta I}{I} \to C.$$

This is in agreement with Weber’s Law.

Modern methods of measuring the JND for sound pressure or intensity typically require the listener to compare two comparatively (e.g. ≤ 1 sec) short pulses of sound presented with a brief inter-pulse interval. Furthermore, the methods used are objective (in the sense that the experimenter can determine whether the observer’s response is correct or not) and (often) adaptive.

For narrow-band stimuli, such as tones, the range over which Weber’s law applies to intensity discrimination is fairly restricted (Fig. 10).
Dependence on Frequency and Level

Jesteadt, Wier, and Green (1977) determined the dependence of tone intensity discrimination on the intensity and frequency of the tone. They measured intensity discrimination for 500 ms bursts of tone using an adaptive two-interval forced-choice procedure that converged on the 70.7% correct point. Each of three listeners were tested in four or five consecutive 100 trial adaptive blocks. Stimuli were presented in a low level (0 dB Spectrum Level) noise that was low-pass-filtered to 10 kHz. The noise raised thresholds by 5–15 dB relative to the ISO standard. Estimates of the intensity DL were obtained at distinct frequencies between 200 and 8000 Hz and over the range 5–80 dB SL.

When measured by the pulse comparison method, JND for intensity exhibits essentially the same dependence on intensity (specified in sensation level) for a wide range of frequencies (Fig. 11).
Figure 8: Comparison of a true Amplitude Modulated waveform (upper plot in both panels) with the sum of two sinusoids (lower plot in both panels). The carrier of the AM waveform is a 30 Hz sinusoid, the modulating waveform is a 10 Hz sinusoid, and the depth of modulation is 0.3. The dotted envelopes indicate the modulating waveform. The time scale in the lower panel is expanded by a factor of 10. The dashed lines in the lower panel show that whereas the zero crossings of the AM waveform occur with the same periodicity as the carrier, those of the sum of two sine-waves do not.
Figure 9: JNDs for intensity discrimination of white noise (flat spectrum within ±5 dB between 150 and 7000 Hz), from Miller (1947).

Figure 10: Measurements of the Weber fraction for intensity \( W(I) = \frac{\Delta I}{I} \) plotted on a logarithmic scale. From Viemeister (1988).
Figure 11: Dependence of $\Delta I/I$ on $I$ for 200–8000 Hz tones as a function of sensation Level. From Jesteadt et al., (1977).
3.3 Frequency

Like intensity discrimination, frequency discrimination has been studied using a variety of methods. In the AX method (or the classical method of constant stimulus differences) two tone bursts are presented in temporal sequence. The first tone (A) is fixed in frequency. The second (X) tone differs from the first only in frequency: its frequency is greater or less than that of A. The task of the listener is to determine whether the frequency of the second tone is higher or lower than that of the first. In the ABX method (Munson and Gardner, 1950), three tones differing only in frequency are presented in temporal sequence. The task of the observer is to determine whether the third tone (X) is more similar to the first tone (A) or the second (B). Surprisingly, most listeners require a greater frequency difference (between A and X, or between A and B) for the ABX than the AX method to achieve a given level of performance.

Rosenblith and Stevens, 1953

![Figure 12: The JND in frequency for pure tones as a function of frequency. Data from the AX method are shown by filled circles, from the ABX method by open circles, from a frequency modulation procedure (Shower and Biddulph, 1931) by crosses, and by the quantal method by squares. From Rosenblith and Stevens (1953).](image)

The results of several sets of measurements of the JND or Difference Limen for frequency are shown in (e.g., Fig. 12). In this Figure, the closed circles represent data taken by the AX method, the open circles the ABX method, data taken by a frequency modulation (Shower...
and Biddulph (1931) procedure by crosses, and data taken by a quantal method by squares. Almost all the data lie within the dotted lines, which represent a spread of one decade.

Perhaps more informative are plots of the relative Difference Limen $\Delta F/F$ for both the AX Method and for a frequency modulation technique (Fig. 13). These plots suggest that whereas the relative DL decreases by roughly an order of magnitude between 0.125 and 2.0 kHz when measured by the frequency modulation technique, it varies by only a factor of two when measured by the AX Method.

![Figure 13: Measurements of the JND in frequency carried out using a the AX Method (circles, Rosenblith and Stevens, 1953) and by a frequency modulation procedure (squares - Shower and Bidulph, 1931).](image)

Moore, 1973

The width of the spectrum of a tone pulse decreases monotonically with the duration of the pulse. Thus one might expect that frequency discrimination would be a function of duration. Measurements of the effect of duration on frequency discrimination were made by Moore (1973).

Moore measured frequency discrimination tone bursts that were switched on and off at random phase with 2 ms rise/fall times. Data were taken using a two-interval forced-choice procedure, without feedback. Two listeners were tested using a fixed frequency difference for blocks of 100 trials. Three blocks were measured (on different days) defining one point on the empirical psychometric function. One listener was tested using an adaptive procedure (PEST; Taylor and Creelman, 1967) set to converge on the 75% correct point. The listeners exhibited performance improvements (learning or practice effects) over a period of two weeks of 2 hours per day testing. After no further improvement could be detected, the two listeners who used the fixed procedure were tested for roughly 100 hours and the listener who used the PEST procedure was tested for roughly 200 hours.
Stimuli were presented at a constant (across frequency) loudness level of roughly 50 phons, but the presentation levels of the two tones being compared were not randomized. The measurements of $\Delta F/F$ for the subject who was tested using the PEST method are shown in Fig. 14. Results for the other listeners are essentially similar. Measurements indicate that $\Delta F/F$ decreases as a function of duration for all frequencies, although the changes as duration is increased from 100 to 200 ms are small. For each duration $\Delta F/F$ decreases as $F$ increases from 250 Hz to 2000 Hz, and then increases. For all listeners there is a sharp increase in $\Delta F$ as $F$ increases from 4 to 6 kHz. Moore attributes these results to the use of place cues only for frequency discrimination at frequencies above 5 kHz, while both timing and place cues are used below 5 kHz.

**Wier, Jesteadt, and Green, 1977**

Data on the effect of intensity and base frequency on frequency discrimination obtained by Wier et al. (1977) are shown in Figure 15.

They measured frequency discrimination for 500 ms bursts of tone using an adaptive two-interval forced-choice procedure that converged on the 70.7% correct point. The presentation level of the tones were not randomized. Four listeners received four or five consecutive 100 trial adaptive blocks to measure the frequency DL. Stimuli were presented in a low level (0 dB Spectrum Level) noise that was low-pass-filtered to 10 kHz. The noise raised thresholds by 5–15 dB relative to the ISO standard. Estimates of the frequency DL were obtained at distinct frequencies between 200 and 8000 Hz and over the range 5–80 dB SL.

The data show a large effect of frequency and a smaller effect of intensity on frequency discrimination.
discrimination. The effect of frequency is approximated by the dotted lines in Fig. 15.

\[
\log \Delta F = S(L) + T(L) \sqrt{F}
\]

or

\[
\Delta F = C(L) e^{T(L) \sqrt{F}}
\]

where \( L \) is the Sensation Level (in dB). This dependence of \( \Delta F \) on frequency is roughly the same as that found by others who used pulsed tones (Fig. 16), in particular it is fairly consistent with the measurements of Rosenblith and Stevens (1953) and Moore (1973). The dependence of frequency discrimination on sensation level is frequency dependent. While the \( \Delta F \) decreases with sensation level for \( L \leq 40 \text{ dB} \), it increases or remains constant at 80 dB SL for \( F = 400, 600, \text{ and } 8000 \text{ Hz} \), but continues to decrease for the other frequencies tested.

Figure 15: The JND in frequency for pure tones as a function of frequency and sensation level as obtained using an adaptive 2AFC procedure. The abscissa is proportional to \( \sqrt{F} \). The dotted line is a fit to the critical band data from Zwicker, Flottorp, and Stevens (1957). From Wier, Jesteadt and Green (1977).
Figure 16: The JND in frequency for pure tones as a function of frequency. From Wier, Jesteadt and Green (1977).
4 Masking

Masking is the obscuring or interference produced by one sound (the masker) in the perception of a second sound (the maskee or target). The effects of a masker on a maskee can be very complex. For example, in addition to altering the threshold of detectability, the masker can alter the loudness, pitch, and apparent duration of the maskee.

If the intensity required to make the maskee audible in quiet is $I_Q$, and the intensity required to achieve the same degree of audibility in the presence of a masker is $I_M$, then the amount of masking produced by the masker is

$$M = \frac{I_M}{I_Q}.$$ 

If $I_0$ is any reference intensity then

$$M = \frac{I_M/I_0}{I_Q/I_0},$$

so that the amount of masking (in dB) is the difference between the masked and quiet thresholds when expressed in dB SPL.

The study of masking is one of the major endeavors in the study of hearing. Nearly every sound can interfere to some extent with the perception of another sound, and the two sounds need not always overlap in frequency or time for this interference to occur.

The study of masking has proven to be both of substantial practical and great theoretical importance. We rarely are fortunate enough to communicate in a totally quiet environment, so background interference can be a major limitation on communication. The study of masking has also provided insight into the functional organization of the auditory system, particularly with respect to frequency selectivity.

Much of the available data on masking has been obtained using procedures that would today be considered “informal” e.g., the method of limits and the method of adjustment. Since the 1960’s, most studies have tended to use objective methods, and most recently, adaptive objective methods. In the presence of a masker, a maskee may be detected by a wide variety of perceptual cues. The percept of the masker may be altered, new spectral components may be created by nonlinear interactions, etc. In the older studies, experimenters often used introspection to identify the cue(s) used in detection. It is much more difficult to obtain such introspection in modern studies that use objective methods, since any cue that helps identify the presence of the maskee can assist the listener in making a correct response. In addition, objective methods often require far more time to estimate the masked threshold value, so that fewer data points can be obtained. Several recent studies that focussed on fine-grain measurements have reverted to the use of subjective methods to overcome this problem.
5 Simultaneous Masking

In our daily life, the most deleterious effects of masking occur when the masker is present simultaneously with the maskee.

5.1 Tones Masked by Tones

In terms of the physical components of the stimulus, the simplest instance of masking to consider is the case of two tones. In the study of Wegel and Lane (1924), the masker is referred to as the primary tone and the maskee as the secondary tone. In this study the primary tone was presented at a constant frequency ($f_P = 1200$ Hz) and level ($L_P = 80$ dB SL) while the secondary tone was varied in level $L_S$ and frequency $f_S$.

For each value of $f_S$ there is a level of the secondary tone (reported in dB SL in Fig. 17) below which the listener reported hearing only the primary tone. Plotted as a function of frequency, this level defines the masked threshold of the secondary tone. Above threshold, the listener reported hearing

1. Beats.
2. Only the primary and secondary tones.
3. The primary and secondary tones plus a difference tone.
4. The primary tone plus a difference tone of frequency $|f_P - f_S|$.
5. A mixture of tones.

Beats can generally be heard when the difference $|f_P - f_S|$ is less than 40 Hz, although as shown by Riesz (1928, Fig. 6) the detection threshold increases as the frequency difference deviates from 3–4 Hz. What is surprising is the audibility of beats at $f_S \approx 2f_P$ and $f_S \approx 3f_P$.

To the extent that these beats do not represent distortions introduced by the stimulus generation apparatus, they suggest that the function of the ear is inherently nonlinear.

The nonlinearity of the ear is also required to account for the listener’s observations that the presence of the secondary tone was cued by the audibility of a difference tone (frequency $|f_P - f_S|$). Indeed, when $f_P = 1200$ Hz with $L_P = 80$ dB SL, and $f_S = 700$ Hz with $L_S = 70$ dB SL, tones of frequency 200 ($2f_S - f_P$), 500 ($f_P - f_S$), 700 ($f_S$), 900 ($3f_S - f_P$), 1000 ($2f_P - 2f_S$), 1200 ($f_P$), 1400 ($2f_S$), 1700 ($2f_P - f_S$), 1900 ($f_P + f_S$), 2100 ($3f_S$), 2400 ($2f_P$), 2600 ($2f_S + f_P$), 2800 ($4f_S$), 2900 ($3f_P - f_S$), 3100 ($2f_P + f_S$), 3300 ($3f_S + f_P$), 3600 ($3f_P$), 3800 ($2f_P + 2f_S$), and 4300 ($3f_P + f_S$) Hz were found to be perceived by the ear. Tones of frequency $mf_P \pm nf_S$ are said to be combination (or difference) tones and were first discovered by the violinist Tartini around 1750. They are caused by distortion either in the acoustic system that produced the primary and secondary tones or in the ear itself.

Additional evidence for auditory nonlinearity is provided by measurements of the rate of growth of masked threshold values of $L_S$ as $L_P$ increases. As can be seen in Fig. 17, when $f_S < f_P$, $L_S$ grows more slowly than $L_P$, when $f_S \approx f_P$, $L_S$ grows at the same rate as
Figure 17: Top panel: percepts elicited by the simultaneous presentation of a 1200 Hz primary tone at 80 dB SL and a secondary tone of frequency $f_S$ and sensation level $L_S$. The solid curve is the masked threshold of the secondary tone. Lower left panel: as in top panel, but for primary tones of different levels. Lower right panel: dependence of the masked threshold of the secondary tone on the level of the primary tone for various secondary tone frequencies. Adapted by Fletcher (Fletcher, 1953) from the data of Wegel and Lane (1924).
$L_P$, when $f_S > f_P$, $L_S$ initially grows more slowly than $L_P$, but then grows more rapidly. These nonlinear patterns of growth may be associated with “two-tone inhibition” seen in the response of cochlear nerve fibers.

### 5.2 Tones Masked by Narrow-band Noise

Figure 18: Left panels: amount of masking of secondary tones of various frequencies in the presence of a 400 Hz pure tone (top panel) or 90 Hz wide narrow band noise centered at 410 Hz (bottom panel). The number under each curve is the sound pressure level of the masking stimulus. The absolute threshold at 400 Hz was measured to be 15 dB SPL. Right panels: dependence of amount of masking on the level of the masker for various secondary tone frequencies. Data from one listener. From Egan and Hake (1950).

The pattern of masking produced by a narrow band of noise should differ from that produced by a pure tone. In particular the deep notches in the masked threshold curve for
\( f_S = k f_P \) seen for a tonal masker should be absent. This is because the time waveform for a narrow-band noise can be expressed as

\[ n(t) = a(t) \cos(\omega_0 t + \phi(t)), \]

where \( a(t) \) is a low-pass signal having a bandwidth half that of the narrow-band noise. Thus the envelope of \( n(t) \) is not constant, as in the case of a pure tone, but fluctuates randomly, making the detection of the secondary tone by the creation of beats difficult.

Figure 19: Comparison of masking for secondary tones of various frequencies in the presence of an 80 dB SPL 400 Hz pure tone with an 80 dB SPL 90 Hz wide narrow band noise centered at 410 Hz. Data are the average of five listeners with normal hearing. From Egan and Hake (1950).

Egan and Hake (1950) measured detection thresholds for tones masked by 400 Hz tones and by 90 Hz wide\(^4\) noise centered at 410 Hz. Tones were 700 ms in duration and alternated with 700 ms periods of silence. Five listeners used the method of adjustment with a criterion of “detection of anything.”

\(^4\)3 dB down from the peak.
At least two measurements were made at each frequency by each listener, except that only one measurement was made per listener for frequencies near 400 Hz with the tone masker.

The results of one listener are shown in Fig. 18; the other four listeners showed roughly the same results. Note that at low masker levels the masking function is roughly symmetrical when plotted on a logarithmic frequency axis for both types of masker. At higher intensities the masking pattern is asymmetric, particularly in the case of tonal maskers.

The dependence of masked threshold on masker level is different for tone and narrow-band noise maskers. For noises whose spectra include the frequency of the tone, the amount of masking increases 1 dB for each 1 dB increase in masker level. The equation of the straight line is

\[ M = \frac{I_M \Delta F_K}{I_0 \Delta F_M} \]

where \( M \) is the masking, \( I_M \) is the overall intensity of the masking noise, \( \Delta F_M \) is the bandwidth of the masking noise, \( I_0 \) is the quiet threshold of the masked tone, and \( \Delta F_K \) is the critical ratio (Sec. 5.4) at the noise center frequency. The same slope is seen only at intermediate intensities of the tone masker. For tones whose frequency are above the spectra the noise, masking increases more than 1 dB for each 1 dB increase in noise level at sufficiently high masker levels. For the tone masker, masking increases even more rapidly.

Averaged masking patterns produced by 80 dB SPL maskers for the five listeners are compared in Fig. 19. The masking pattern for the tonal masker exhibits the effect of “beats” for frequencies near 400, 800, and 1200 Hz. The masking function passes through a minimum between 400–800 Hz because the difference tone, of frequency \( f_S - f_P \) is detectable before the masked stimulus \( f_S \) is heard. The masking pattern produced by narrow band noise is essentially free of the effects of beats, harmonics, and difference tones.

**Confirmation**

Recently Alcántara et al. (2000), have used low frequency noise to mask combination tones and a pair of high- or low-frequency tones to interfere with the detection of beats (modulation detection interference, Yost et al., 1989) to understand the shapes of tone-on-tone masking functions similar to those shown in Fig. 17. They concluded that the shapes of the masking patterns for a \( f_P = 2000 \) Hz masking tone are influenced by the detection of beats for \( |f_S - f_P| < 300 \) Hz and by the detection of combination tones for \( 300 < f_S - f_P < 1000 \) Hz.
5.3 Masking of Tones by Wideband Noise

Noise is often modelled as a random process with a well defined power spectrum, $N(f)$ and a random phase spectrum. The power spectrum of the noise process $n(t)$ is defined so that if $n(t)$ is filtered by an ideal bandpass filter that passes frequencies $f_A \leq f \leq f_B$, the power in the process that emerges from the output of the filter is

$$P_{\text{out}} = \int_{f_A}^{f_B} N(f) \, df$$

where $N(f) \geq 0$.

Noise that has the property that $N(f) = N_0$ is said to be white noise with spectrum level $N_0$. This terminology is used even when the noise bandwidth is finite, provided $N(f) = N_0$ over the frequency range of interest.

If white noise with spectrum level $N_0$ is filtered by an ideal bandpass filter with center frequency $f_0$ and bandwidth $\Delta f$, the resulting noise process will have a flat power spectrum over the range $f_0 - \Delta f/2 \leq f \leq f_0 + \Delta f/2$. If the resulting signal is filtered by an ideal bandpass filter with center frequency $f_0$ and bandwidth $\delta f$, the power in resulting process is

$$P = \begin{cases} 
N_0 \delta f & \text{if } \delta f \leq \Delta f \\
N_0 \Delta f & \text{if } \Delta f \leq \delta f 
\end{cases}$$

so that $P = N_0 \min(\delta f, \Delta f)$.

5.4 Critical Ratio

Hawkins and Stevens (1950) measured tone detection threshold as a function of $f_T$ and the level of broadband noise. Thresholds in the absence of noise were the levels at which the tones had definite pitch. The noise spectrum was measured to be constant ±2 dB over the range 0.1–9 kHz (Fig. 20) and so may be approximated to be a constant $N_0$. At low noise levels, the tone detection threshold $I_M (f_T)$ approaches the detection threshold in quiet $I_Q (f_t)$. At high noise levels the ratio of the tone detection threshold $I_M (f_T)$ to the noise power spectral density $N_0$ approaches a frequency dependent constant, independent of $N_0$ that is now called the Critical Ratio:

$$\frac{I_M (f_T)}{N_0} = \text{CR} (f_T) .$$

Hawkins and Stevens also report (Fig. 21) how the masking $M$ produced by the flat spectrum noise on a tone of frequency $f_T$ depends on the effective level $Z$ of the masking noise

$$Z (f_T) = \frac{N_0 \text{CR} (f_T)}{I_Q (f_T)} .$$

Hawkins and Stevens interpreted this ratio as the critical bandwidth and published estimates of the ratio by averaging results at the four highest noise levels shown in Fig. 20.
Figure 20: Masked thresholds of tones in broadband flat-spectrum noises differing in power spectral density, as determined by Hawkins and Stevens (1950). Thresholds are averages of measurements made using the method of adjustment by four listeners whose average thresholds for tones in quiet were similar to those reported by Sivian and White (1933). Five-six determinations of thresholds in quiet and two-five determinations of the threshold contour for each of 8 noise levels were made by each listener.
Figure 21: Dependence of the masking $M$ for tones in broadband flat-spectrum noise on the effective level ($Z$, the ratio of the noise power in a bandwidth equal to the Critical Ratio to the quiet threshold for a tone in the center of the band), as determined by Hawkins and Stevens (1950).
Since $I_M(f_T) \approx N_0 CR(f_T)$ at high noise levels, $M = Z$ at high noise levels. At low noise level levels, the ratio $I_M/I_Q$ approaches unity as $Z$ is decreased. This effect is often modelled by assuming that there is an “internal noise” with a non-uniform spectrum whose power adds to the external masking noise.

5.5 The “Critical Band”

Figure 22: Dependence of thresholds for detection of tones of various frequencies on noise bandwidth in simultaneous masking. In each case the noise band is centered on the frequency of the tone. $I_m$ is the tone level at masked threshold; $I_f$ is the noise spectral level (power in a 1 Hz band). From Fletcher (1940).

Fletcher (1940) measured the masked thresholds of tones of various frequencies in noise that was either wideband or filtered to various bandwidths ($W_N$) that were, in each case, centered on the tone frequency $f_T$. His results (Fig. 22) indicate that as the noise bandwidth is decreased from a large value, the masked threshold $P_T$ of the tone is essentially unchanged, suggesting that noise components spectrally remote from the tone frequency do not affect the detection of the tone. If the noise bandwidth is reduced sufficiently, however, the detection
threshold is reduced, eventually becoming roughly proportional to the noise bandwidth.

Fletcher interpreted his results as reflecting the joint operation of his external filter and an internal auditory filter of bandwidth $W_C(f_T)$ that he called the critical band. Fletcher did not estimate the dependence of $P_T$ on $W_N$ extensively. Rather he assumed that when $W_N < W_C(f_T)$, the tone threshold equals

$$P_T = N_0 W_N$$

for all values of $f_T$. He then determined value of $W_C(f_T)$ from the intersection of the two curves:

$$P_T(W_N) = W_N N_0 \text{ for } W_N \leq W_C(f_T)$$

$$P_T(W_N) = W_C(f_T) N_0 \text{ for } W_N \geq W_C(f_T)$$

At the apparent breakpoint in Fletcher’s curves $W_C(f_t) = W_N$.

It should be noted that Fletcher’s analysis assumed that the process of detecting a tone depended on the linear filtering operation of a rectangular filter centered on the tone frequency, with tone power equal to noise power at the output of the filter at the masked threshold. Although subsequent research has called into question many details of Fletcher’s analysis, his fundamental ideas have stood the test of time and have proven to be among the most influential in auditory research.

### 5.6 Width of the Critical Band

In the early studies, the critical band was treated as an ideal bandpass filter centered on the frequency of the tone $f_T$, i.e., the magnitude of the attenuation characteristic $|H(f)|$ was assumed to be

$$|H(f)| = \begin{cases} 
1 & \text{if } |f - f_T| \leq W_{CB}/2 \\
0 & \text{otherwise} 
\end{cases}$$

This filter is “ideal” in two senses:

1) Input components that fall in the pass band of the filter (Eq. 5.6) are not attenuated.

2) Input components that fall in the stop band of the filter (Eq. 5.6) are attenuated completely.

If the input to the filter $x(t)$ consists of the sum of a tone and a noise

$$x(t) = \sqrt{2} A \cos 2\pi f_T t + n(t),$$

where $n(t)$ is a sample of a noise process with power spectrum $N(f)$

$$N(f) = \begin{cases} 
N_0 & \text{if } |f - f_T| \leq W_N/2 \\
0 & \text{otherwise} 
\end{cases}$$
then the tone power at the output of the filter is \( A^2 \) and the noise power is

\[
N = \int_0^\infty |H(f)|^2 N(f) df
\]  

\[
= \begin{cases} 
N_0 W_N & \text{if } W_N \leq W_{CB} \\
N_0 W_{CB} & \text{otherwise}
\end{cases}
\]

The ratio of tone to noise power (\( S/N \)) at the output of the filter is

\[
\frac{S}{N} = \begin{cases} 
\frac{A^2}{N_0 W_N} & \text{if } W_N \leq W_{CB} \\
\frac{A^2}{N_0 W_{CB}} & \text{otherwise}
\end{cases}
\]

**Fletcher’s Theory**

Fletcher (1940) assumed that the tone would be audible in the noise when \( S/N = 1 \) for all values of \( N_0 \) and \( W_N \), and that \( W_{CB} \) could depend on \( f_T \). The masking data that Fletcher used to estimate \( W_{CB} \) are roughly consistent with the (non-)dependence of detection thresholds on noise bandwidth (Eq. 5) for large \( W_N \), but are inadequate to define the breakpoint implied by the above equations. Instead Fletcher relied on Eq. 5 to estimate \( W_{CB} \). For a sufficiently small value of \( W_N (30 \text{ Hz}) \) and fixed \( N_0 \), the data indicated that the tone threshold was independent of \( f_T \). Fletcher used Eq. 5 and the tone threshold at \( W_N = 30 \text{ Hz} \) to estimate the tone threshold at other values of \( W_N \leq W_{CB} \) (the sloping line in Fig. 22). He estimated the width of the critical band \( W_{CB} \) as the bandwidth that satisfies Eq. 5.

**Moore’s Correction**

The assumption stated in Eq. 2 has not been accepted by all researchers. Moore (2003) used Fletcher’s band-narrowing technique, but with a larger range of bandwidths. As in Fletcher’s data, he found that when a 2000 Hz tone is masked by noises of various bandwidths, thresholds (Fig. 23) are roughly independent of bandwidths for widths of 400 Hz and above, but decline when \( W_N < 400 \text{ Hz} \). These results suggest that \( W_C = 400 \text{ Hz} \) at 2000 Hz, a value substantially larger than that reported by Fletcher (roughly 100 Hz). Note, however, that thresholds only decreases by 4 dB as the noise bandwidth decreases from 400 Hz to 50 Hz, whereas Eq. 2 would have predicted a decrease of roughly 9 dB.

The data reported by Moore (2003) may represent a fortuitous arrangement of conditions, listeners, and responses. Other data (Fig. 24) published by Schooneveldt and Moore (1989), are only roughly consistent with Moore’s (2003) findings.

### 5.7 Relation of the Critical Band to the Critical Ratio

According to Fletcher’s thinking, the Critical Ratio \( CR(f_T) \) should equal \( W_C(f_T) \). Indeed, Fletcher included wide band noises in his experiments. A comparison of the dependence of widths of critical bands and of the size of the critical ratio on frequency is shown in Fig. 25.
Figure 23: Masked thresholds for 2000 Hz tones in noises of varying bandwidth and constant spectrum level. Data replotted from Moore, (2003).

Figure 24: Masked thresholds for 2000 Hz tones in noises of varying bandwidth and constant spectrum level. Data from Schooneveldt and Moore, (1989).
Figure 25: Estimates of the width of critical bands obtained by Fletcher (1940) and of the critical ratio (expressed as a bandwidth) by Hawkins and Stevens (1950).
Figure 26: The internal auditory filter attenuation characteristic is represented by the unfilled curve, which has a width of $W_C$ and is centered on the tone frequency $f_T$. The noise power spectrum, which has a width of $W_N$ and is centered on $f_N$, is represented by the filled rectangle. In the upper panel represents the case $W_N < W_C$, in the lower $W_N > W_C$.

5.8 Masking by Noise as a Function of Bandwidth

An extensive set of measurements of masked thresholds for tones in bandlimited flat-spectrum noise has been made by Greenwood (1961) using a modified form of Bekesy audiometry. For a given noise base frequency $f_B$, bandwidth $W_N$, center frequency $f_C = f_B + W_N/2$ and spectral level $N_0$, measurements were made for tone frequencies $f_T$ both within and outside of the noise band (Fig. 27).

These measurements can be understood by assuming that

1) The tone and filtered noise are further filtered by an internal filter of width $W_C (f_T)$ that is centered on and symmetrical about $f_T$.

2) Detection is determined by the ratio of the tone power to noise power at the output of the filter.

Two cases (Fig. 26) yield different predictions:

1. $W_C \geq W_N$. In this case the (doubly) filtered noise power is maximal when $f_T = f_C$ and increases as $W_N$ increases.

2. $W_C \leq W_N$. In this case there is a range of $f_T$ for which the (doubly) filtered noise power is the same. The power is greater than for $f_T$ outside of this range.

Greenwood identified the critical bandwidth with the highest value of $W_N$ for which the tone threshold curve resembles an inverted V, i.e., with no flat top. For this and smaller
values of $W_N$ the peak tone threshold occurs when $f_T$ is roughly in the middle of the noise band, corresponding to maximal of noise power passed through the auditory filter. Because the low-frequency cutoff of the noise band was fixed as the noise bandwidth increased, the value of $f_T$ with the highest threshold increased as $W_N$ increased.

Greenwood’s measurements have been used to estimate $W_C$ and produce values that are roughly twice those reported by Fletcher.
Figure 27: Masked thresholds for tones in bandlimited flat-spectrum noise of various spectrum levels, from Greenwood (1962). The different panels correspond to different listeners and/or different spectrum levels of the masking noise. The different curves within each panel correspond to different noise bandwidths (and hence different total noise powers). The noise bands all share the same lower cutoff frequencies (roughly 2992 Hz) and extend to different higher cutoff frequencies as the noise bandwidth increases. The data points for each curve are the thresholds for tones at the frequency specified by the abscissa.
5.9 Complex Tones

Other attempts to characterize the properties of the critical band filter have used stimuli $s(t)$ consisting of tone complexes rather than single tones:

$$s(t) = \sum_{k=k_M}^{k_N} A \sin (2\pi (f_c + k\Delta f) t).$$

The total average power in such a tone complex is

$$P = \frac{k_N + 1 - k_M A^2}{2}$$

In a typical experiment, $k_M$, $k_N$, and $f_c$ are chosen so that a set of tone complexes maintain a common mean frequency as the number of components varies.

Thresholds of Complex Tones

![Figure 28: Schematic illustration of the interaction of G"assler's tone complexes (circle-topped vertical lines) with internal auditory critical-band filter of assumed rectangular shape. The power of the tone complex is maintained constant by adjusting the amplitudes of the components as the number of components is increased from 1 to 4, as shown in panels a–d. In panels a–c, the total bandwidth of the complex is less than the width of the filter passband, so that the power at the filter output is the same. In panel d, the total bandwidth of the complex exceeds the width of the filter passband, causing the power at the filter bandwidth to be less than in panels a–c.](image)

G"assler (1954) varied the bandwidth $B = (k_N + 1 - k_M) \Delta f$ of the complex and determined the minimum value of the power in the complex $P(B)$ required for the complex of bandwidth $B$ to be audible. Consider the case in which the tone complex is presented in a noise background. For a fixed internal filter bandwidth, $W$, the noise power at the filter output is $N_0 W$. As can be seen in Fig. 28, the tone power at the output of the filter is equal.

\[^6\text{In the quiet case, the absolute detection threshold can be thought of as produced by an internal noise.}\]
to the total power of the complex if $B \leq W$, or a decreasing fraction of that power if $B > W$ in which case one or more components of the complex are not passed by the filter.

Results obtained both for tone complexes presented in quiet and also for complexes presented in flat spectrum noises (Fig. 29) indicate that as the bandwidth $B$ of the complex increases, detection thresholds (reported as total power of the complex) are the same as for a single tone for small values of $B$. This is presumably because all of the power of the complex is passed by the internal filter. As the bandwidth continues to increase, however, the power in the complex must be increased for it to be detectable. Presumably some components are outside the passband of the filter and the components passed by the filter are detectable only if they are amplified.

Thus when $B \leq B_C$, the detection threshold of the complex is roughly the same, in terms of power, as for a single tone at the center of the complex. When $B > B_C$, detection requires that the power of the complex exceed the power of a single tone by an amount that increases as $B$ increases.

The abscissa of the breakpoints in Fig. 29 defines the width of the internal critical band filter: it is the largest bandwidth of the tone-complex for which the threshold power is the same as that for a single tone. These breakpoints imply that the width of filter at 1000 Hz is $B_C \approx 150$ Hz both in quiet and for noise spectral levels up to at least 50 dB SPL/Hz. Note that this width is roughly twice that derived by Fletcher. Although it might be argued that the structure of Gässlér’s experiment favors the use of a filter with a greater width (to sum the power of the components of the tone complex, as opposed to rejecting noise) other types of experiments yield estimates of the filter width that are compatible with Gässlér’s.

---

7The powers of these components are integrated less efficiently than the powers of the components within the passband.
Figure 29: Thresholds for tone complexes consisting of equal amplitude tones spaced symmetrically by 20 Hz steps around 1000 Hz, from Gässler (1954).
Loudness of Complex Tones

A second set of experiments using tone complexes of differing bandwidths is concerned not with detection, but rather with the loudness of the complex or the noise band.

For pure tones, loudness is often assumed to be a power function of intensity

$$L(I) = kI^\gamma,$$

where the exponent $\gamma$ decreases from $\gamma \approx 1$ at intensities near threshold to $\gamma \approx 0.3$ at high intensities. The growth of loudness for sounds whose bandwidth is less than the width of a critical band is roughly the same as for tones.

According to loudness summation theory (Zwicker, Flottorp, and Stevens, 1957), the loudness of sounds is equal to the sum of the loudnesses (in isolation) of the sounds that fall within each critical band. Thus as the bandwidth of a sound increases while the power is kept constant, loudness is expected to remain constant as long as the bandwidth of the sound is less than the width of the critical band. Evidence for this constancy was obtained by Zwicker, Flottorp, and Stevens (1957) for both tone complexes (Fig. 30) and bandlimited noise. Again, note that the bandwidth at which constant power does not ensure constant loudness is roughly twice as large as that derived by Fletcher.
Figure 30: Intensity of a 1000 Hz tone matched in loudness to a four-tone complex of various widths centered on 1000 Hz.
A. M. vs Q. F. M

A final measure of the critical bandwidth is provided by experiments that measure the ability to discriminate phase differences (Zwicker, 1952). Sounds of the form

\[ s_{AM}(t) = A(\alpha \cos 2\pi (f_C - \Delta f) t + \cos 2\pi f_C t + \alpha \cos 2\pi (f_C + \Delta f) t) \]

are amplitude-modulated constant frequency tones, whereas

\[ s_{FM}(t) = A(\beta \sin 2\pi (f_C - \Delta f) t + \cos 2\pi f_C t - \beta \sin 2\pi (f_C + \Delta f) t) \]

are quasi-frequency modulated tones with nearly constant amplitude (provided \( \beta \ll 0.5 \)). Note that when \( \alpha = \beta \), \( s_{AM}(t) \) and \( s_{FM}(t) \) have the same power spectrum, but different phase spectra.

When \( \Delta f \) is small, the smallest \( \beta \) that allows \( s_{FM}(t) \) to be distinguished from a pure tone is larger than the smallest \( \alpha \) that allows \( s_{AM}(t) \) to be distinguished from a pure tone. However as \( \Delta f \) increases, the the smallest values of \( \beta \) and \( \alpha \) that allow these distinctions to be made converge. For this critical value of \( \Delta f \), and larger values, it is not possible to distinguish between \( s_{FM}(t) \) and \( s_{AM}(t) \).

Twice the critical frequency difference is taken as a measure of the width of the critical band centered on \( f_C \). The fact that the same values of \( \alpha \) and \( \beta \) are found for \( \Delta f \) greater than a critical bandwidth is thought to be evidence for the monaural phase insensitivity of the ear.

The dependence of this estimate of the critical bandwidth on frequency is similar to that obtained from measurements of masking by Greenwood, for the detection of tone complexes by Gässler, and for loudness summation by Zwicker, Flottorp, and Stevens (Fig. 31).

Summary

The experimental results described above indicate that the notion of internal auditory filters considerable explanatory power that extends far beyond Fletcher’s masking experiments. More recent research has attempted to determine the shape (attenuation characteristic) of these internal filter as well as their width. Shape parameters are typically derived from measurements of the masked thresholds of tones presented in noise with deep spectral notches. At the present time, the phase characteristics of the filters are also being characterized using maskers consisting of tone complexes rather than noise.
Figure 31: Comparison of the frequency dependence of the critical bandwidth, the critical ratio, and the frequency difference limen, from Zwicker, Flottorp, and Stevens (1957).
The Auditory Filter

According to the Power Spectrum Model of Masking, a tone masked by a noise of spectrum \( N(f) \) is at its detection threshold if the tone power \( P_T \) satisfies

\[
P_T = K \int_{0}^{\infty} N(f) H(f) \, df
\]

where \( K \) is a constant and \( H(f) \) is the power attenuation characteristic of the auditory filter. This is a generalization of Fletcher’s (1940) picture of the masking of tones. Fletcher treated the auditory filter as an ideal bandpass filter centered on the frequency of the tone \( f_T \). The magnitude of the attenuation characteristic \( |H(f)| \) was assumed to be

\[
|H(f)| = \begin{cases} 
1 & \text{if } |f - f_T| \leq W_{CB}/2 \\
0 & \text{otherwise}
\end{cases}
\]

Moreover, Fletcher assumed that \( K = 1 \). More recent direct measurements (e.g. Greenwood, 1962) suggest that \( K \approx 0.2 \).

Although early studies of the masking of tones by noise focussed on the width of the internal (critical band) auditory filter, \( W_{CB} \), more recent research has focussed on estimating the shape (i.e., attenuation characteristic) of the internal “auditory filter”.

Early attempts to determine the detailed shape of the auditory filter used noise that was processed by analog bandpass filters. Predicted values of masked tone thresholds were derived for various assumed filter shapes, with filter parameters (e.g., bandwidth, slopes of filter skirts, etc.) fit to measurements tone detection thresholds.

This approach was limited by the difficulty of filtering noise to have a precise bandpass characteristic (e.g., Eqs. 6) using analog techniques. In particular, if the stop band of the filter provides only finite (frequency dependent) attenuation, estimates of the shape of the “tails” of the attenuation characteristic of the auditory filter will be in error. Moreover, since the noise power in the stop-bands is typically only a small fraction of the total noise power at the output of the filter, this error will have only a small effect on predicted thresholds.

6.1 “Notched Spectrum” Noise

More recent attempts to determine the shape of the auditory filter have used a different approach based on measuring the amount of masking produced by notched noise whose power spectrum is given by

\[
N(f) = \begin{cases} 
N_0 & \text{if } |f - f_T| \geq \Delta f \\
0 & \text{otherwise}
\end{cases}
\]

Use of digitally-generated notched noise permits the shapes of the tails of the attenuation characteristic of the auditory filter to be determined more precisely. When the bandwidth of the notch \( W_{NN} \) (2\( \Delta f \) in Fig. 32) is large, the power passed by the internal auditory filter is determined by the noise power in the stop band of the auditory filter since there is no noise power in the pass band, e.g., Fig. 32.
6.2 Rounded Exponential Filters

Patterson (1976) argued that in detecting a tone of frequency \( f_T \), the listener employs an internal auditory filter with the attenuation characteristic:

\[
W(g) = (1 + pg) e^{-pg}
\]

where

\[
g = \frac{|f - f_T|}{f_T}.
\]

This filter passes the tone with a gain of unity, but progressively attenuates spectral components with frequencies different from \( f_T \). The parameter \( p \) determines both the width and the steepness of the filter. The shape of the attenuation characteristic of the filter is symmetric about the tone frequency \( f_T \) on a linear frequency scale. Since \( e^{-x} \approx 1 - x \) for small values of \( x \), the filter characteristic is parabolic for frequencies close to the tone, i.e.,

\[
W(g) \approx 1 - (pg)^2.
\]

The equivalent rectangular bandwidth (ERB)\(^8\) of the filter is \( 4f_T/p \).

Several modifications of this shape have also been considered. To account for possible asymmetry in the shape of the filter, different values of \( p \) are allowed for frequencies below \( (p_l) \) and above \( (p_u) f_T \). In general \( p_u \) and particularly \( p_l \) are dependent on the level of the signal applied to the filter (Fig. 33).

To account for the effect of absolute thresholds, the dynamic range of the filter is limited to a maximum attenuation factor of \( r \).

\[
W(g) = (1 - r)(1 + pg)e^{-pg} + r
\]

Whereas different values of \( p \) may be required for frequencies above and below \( f_T \), the same value of \( r \) is typically used in both cases.

\(^8\)The equivalent rectangular bandwidth (ERB) of a filter is the bandwidth of an ideal filter whose output, when driven by flat spectrum noise, is the same as that of the filter.
6.3 Parameter Estimation

The parameters of the auditory filter are generally determined by measuring tone detection thresholds in the presence of a notched noise masker. Detection is assumed to require a threshold ratio of tone power $P_S$ to noise power $P_N$ at the output of the filter,

$$P_N = N_0 \int_0^{g_L} (1 + p_L g) e^{-p_L g} dg + N_0 \int_{g_H}^\infty (1 + p_u g) e^{-p_u g} dg$$

where $g_L = 1 - f_L/f_T$ and $g_H = f_H/f_T - 1$.

Data from several experiments, e.g., Fig. 35 indicate that under conditions where $p_L \approx p_u$, the equivalent rectangular bandwidth $4 f_T/p$ is related to the tone frequency $f_T$ by

$$\text{ERB}(f_T) = \frac{4 f_T}{p} = 0.1079 f_T + 24.7$$ \hspace{1cm} (8)

or

$${\frac{1}{p}} = 0.026975 + \frac{6.175}{f_T}.$$  

Thus the ERB is roughly 25 Hz at low frequencies, 125 Hz at 1 kHz, and is proportional to frequency at frequencies above 1 kHz. This dependence was determined at fairly high noise levels.

Figure 33: Dependence of the shape of the auditory filter at 1 kHz on input sound level.
Figure 34: Derivation of the shape of the rounded exponential filter from thresholds measured using notched-noise maskers.

spectrum levels, e.g., 58, 55, 52, 49, 30, 35, and 50 dB/Hz at 100, 200, 400, 800, 1000, 8000, and 10000 Hz, respectively. This relationship leads naturally to the \textit{ERB Scale}. The number $N_E$ of ERBs below frequency $f$ being given by

$$N_E(f) = 21.4 \log_{10} (0.00437 f + 1.0)$$

where $f$ is in kilohertz.

6.4 Level Dependence

The level dependence of the sharpness of tuning on the low-frequency side of the filter is accounted for by assuming that $p_l$ changes linearly as the level of the stimulus (in dB SPL) changes:

$$p_l(X) = p_l(51) - \frac{K}{q_l(51)} (X - 51)$$

where $X$ is the effective level (in dB/ERB), $p_l(X)$ is the value of $p_l$ at level $X$, $q_l(51)$ is the value of $p_l$ at 1 kHz, and $K$ is a constant in the range $0.35 \leq K \leq 0.38$. For example, at
Figure 35: Equivalent rectangular bandwidths of rounded exponential filters as a function of center frequency.

1 kHz, where $p_l(51) \approx 30$, $p_l(X)$ decreases by half when $X \approx 94$ dB (the filters become broader, corresponding to the “upward spread of masking”) and increases by half when $X \approx 8$ dB.
7 Non-simultaneous masking

Figure 36: The envelopes of the sound pressure $p(t)$ waveforms are shown in panels A (forward masking) and B (backward masking), with the vertically hatched envelope representing the masker and the horizontally hatched waveform representing the maskee. Panels C (forward masking) and D (backward masking) represent the “effective” stimuli, which are assumed to be prolonged relative to the pressure waveforms.

Although the study of masking in conditions when the masker and maskee occur simultaneously has proven of great practical significance, the study of non-simultaneous masking has proven of great theoretical significance. In non-simultaneous masking, for example, the possibility that detection is based on distortion products formed between the masker and the maskee (e.g., Fig 17) is eliminated.

Studies of non-simultaneous masking have demonstrated the existence of a suppressive phenomena (corresponding somewhat to physiological suppression of neural firing rates). Under appropriate conditions, the masking produced by a primary tone $f_P$ on a secondary tone $f_S$ can be reduced (suppressed) by the introduction of a third tone $f_X$. Unlike masking, suppression is thought to occur instantaneously when the suppressor is present. Thus in an experiment on simultaneous masking the maskee may suppress the masker, or some spectral components of the masker may suppress others. In non-simultaneous masking, the occurrence of suppression is excluded.

A suggestion of how forward and backward masking might arise is illustrated in Fig. 36. The prolongation of the effective stimuli relative to the physical stimuli allows the tail of the masker to reduce the effectiveness of the head of the maskee, giving rise to forward masking. This prolongation allows the head of the masker to reduce the effectiveness of the tail of the maskee, giving rise to backward masking.

7.1 Time Course

Forward masking occurs when the masker is presented before the maskee, backward masking when the masker follows the maskee. For both types of masking the amount of masking
decreases as the time interval between the masker and the maskee increases. As seen in Fig. 37, the time constant for forward masking (roughly 30 ms) is substantially longer than for backward masking (roughly 10 ms). One analysis of the difference between these two types of non-simultaneous masking postulates that backward masking reflects the time decay of excitation produced by the maskee while forward masking reflects the time decay of excitation produced by the masker plus physiological adaptation.

7.2 Munson and Gardner, 1950

The first attempt to measure the frequency dependence of the amount of forward masking produced by a tonal masker (Munson and Gardner, 1950) was made with a relatively long duration maskee (70 ms) that began its rise roughly 20 ms after the cessation of the 1000 Hz. 400 ms duration masker. They described their results (Fig. 38) in terms of the amount of (residual\textsuperscript{9}) masking caused by the 1000 Hz masker. When the masker was presented at 70 dB SL, they found that the amount of masking reached a peak at the frequency of the masker and was highly asymmetric: maskee tones with frequencies below the masker frequency were hardly masked at all, while those with frequencies above 1000 Hz were masked substantially. When the masker level was increased to 100 dB SL, the asymmetry persists, but maximum masking occurs at roughly 1500 Hz rather than at the frequency of the masker. In addition, the width of the masking pattern increases as the intensity of the masker increases, similar

\textsuperscript{9}They used this term to indicate the relationship of the masking to “short duration auditory fatigue.”
Figure 38: Amount of forward masking produced by 400 ms bursts of 1000 Hz tones on 70 ms tones of various frequencies that occur roughly 20 ms after the cessation of the masker. Average results for a 70 dB SL masker are shown on the left, and for a 100 dB SL masker are shown on the right. From Munson and Gardner, (1950).

to simultaneous masking.

Note that there are clear differences between the amount of forward as opposed to simultaneous masking produced by a given stimulus. A 70 dB SL tone produces a maximum of 14 dB of forward masking and a 100 dB SL tone produces a maximum of roughly 29 dB of forward masking. Increasing the level of the tone by 30 dB, thus increases the amount of forward masking only by roughly 15 dB, half as much. By contrast a 70 dB SL tone produces about 40 dB of on-frequency simultaneous masking and a 100 dB SL tone produces about 70 dB of masking. Thus, not only is less forward masking produced by tones, but it grows less rapidly with masker intensity.
7.3 Psychophysical Tuning Curves

Figure 39: Psychophysical tuning curves measured in simultaneous- and forward-masking at 1 and 6 kHz. The duration of the probe was 34 ms, consisting of 17 ms rise and fall segments. The masker had a total duration of 334 ms with 17 ms rise and fall segments. In the simultaneous case the cessation of the probe coincided with the cessation of the masker. In the forward-masking case, the probe began to rise immediately after the cessation of the masker completed. Data from Moore (1978). Points near 1000 Hz in the left hand panel for simultaneous masking represent the variable effects of phase interactions between the masker and the probe.

Non-simultaneous masking has been used to derive the shapes of “psychophysical tuning curves” that bear a striking resemblance to curves measured in physiological experiments. In the physiological case, the frequency \( f_T \) of a tonal stimulus is varied while the level \( L_T(f_T) \) of the stimulus is adjusted to produce a constant response by a neural element. The dependence of \( L_T \) on \( f_T \) defines the shape of the physiological tuning curve. In the perceptual case, a brief tonal probe tone of frequency \( f_P \) and (typically low) level \( L_P \) is masked by a second tone of frequency \( f_M \) and level \( L_M(f_M) \). The dependence of \( L_M \) on \( f_M \) defines the psychophysical tuning curve. The relationship between these two types of tuning curves hinges on the notion that a fixed set of neural elements are stimulated to a constant degree as \( f_T \) and \( f_M \) are varied. Note, however, that in the physiological case the stimulus consists of a single tone, whereas in the perceptual case it consists of a pair of tones.

Psychophysical tuning curves are in principle related to tone-on-tone masking functions, such as those studied by Wegel and Lane (Fig. 17, 1924) and by Egan and Hake (Fig. 18, 1950). In the masking studies, the value of \( L_P \) is determined as a function of \( f_P \) while \( L_M \) and \( f_M \) are kept constant. In the psychophysical tuning curve studies, the value of \( L_M \) is determined as a function of \( f_M \) while \( L_P \) and \( f_P \) are kept constant.
Psychophysical tuning curves have been measured under conditions of simultaneous masking and non-simultaneous (usually forward) masking. Psychophysical tuning curves measured with simultaneous masking are prone to many of the same problems seen in measurements of tone-on-tone masked thresholds: beats when \( f_p \approx n f_M \) and nonlinear effects due to the creation of combination tones. In addition, psychophysical tuning curves measured with simultaneous masking are generally broader than those measured with non-simultaneous masking (Fig. 39). The latter curves are generally more similar to physiological tuning curves (Fig. 40).

**FIG. 10.** Forward-masking curves for probes at 10 dB SL. Subject SB. Frequencies of the probes, with peak SPL’s at absolute threshold in brackets, were 0.5 (31 dB), 1.0 (21 dB), 2.0 (19 dB), 4.0 (15 dB), and 8.0 (20 dB) kHz.

Figure 40: Psychophysical tuning curves measured in forward-masking at a variety of probe frequencies. Data from Moore (1978).
7.4 Level Dependence of Forward Masking

Figure 41: Level and time dependence of forward masking (from Moore, 1978).

Fig. 41 shows the level and time dependence of the forward masking of a 2 kHz probe produced by a noise masker. For the range of time intervals considered, forward masking decreases roughly in proportion to the logarithm of the interval between the masker and the probe. Masking grows as a power function with an exponent of roughly 1/2, although smaller exponents are seen when the interval between the masker and the probe increases.
7.5 Oxenham and Shera, 2003

Recently Oxenham and Shera (2003) derived the shapes of human auditory filters using methods that suggest that the filters may be substantially narrower than is conventionally thought. Their methods differed from those used conventionally to measure auditory filter shapes in the following ways:

a) They varied the spectral level of noise to just mask (70.7% correct responses in three-interval three-alternative forced choice experiments) a fixed level signal. Many previous measurements varied the signal level in a fixed level of noise.

b) They held the tone level constant at 10 dB SL and varied the masker intensity as a function of the spectral composition of the masker. Many previous measurements varied the signal level as the spectral composition of the noise varied.

c) They used forward masking techniques to measure filter shapes and consequently they measured the noise level that just masked brief duration (10 ms) tonal stimuli. Many previous measurements determined filter shapes using data from simultaneous masking.

d) They used notched noises of different notch widths and symmetries to mask the tones.

For comparison, they measured filter shapes using more conventional techniques: simultaneous masking for the same short tone bursts levels of 10 and 35 dB SL.

![Figure 42: Mean masker spectrum levels plotted as a function of notch width for simultaneous masking of 10 dB SL (inverted triangles) and 35 dB SL tone bursts (up-pointing triangles), and for forward masking (squares). Oxenham and Shera, 2003.](image-url)
They obtained measurements with forward masking on 8 listeners, and with simultaneous masking at 10 dB SL on 4 of those listeners and at 35 dB SL on the other 4. The results they obtained for 8 kHz tone bursts and symmetrical notches are shown in Fig. 42 (the results measured for frequencies of 1, 2, 4, and 6 kHz do not differ appreciably from those discussed below). All data points show that with increasing masker notch width, the masker level necessary to mask the signal increases. The measurements made for simultaneous masking with no notch in the noise are roughly consistent with the results of Hawkins and Stevens (1950). However the slopes of the simultaneous and forward masking curves differ at this point: roughly 200 dB/unit for the forward masking and 150 dB/unit for simultaneous masking. These differences in slope suggest a difference in frequency selectivity, depending on whether the signal was presented simultaneously or following the masker. In particular the increase in masker level with increasing notch width appears shallower in the simultaneous masking cases, suggesting poorer effective frequency selectivity.

Oxenham and Shera fit a variety of filter functions (e.g., Eq. 6,7 to their masking data. The principal findings did not depend on which filter shape was fit. They report results in terms of

\[ Q_{\text{ERB}}(f_T) = \frac{f_T}{\text{ERB}(f_T)} \]

where \( f_T \) is the frequency of the tone, \( \text{ERB}(f_T) \) is the estimated width of the filter (in ERBs, and \( Q_{\text{ERB}}(f_T) \) is the quality factor of the filter. Estimated \( Q_{\text{ERB}} \)s for both simultaneous and forward masking are shown in Fig. 43. The three estimates of \( Q_{\text{ERB}} \) obtained in simultaneous masking are in rough agreement. The results also show that \( Q_{\text{ERB}} \)s are substantially larger in forward masking than has been found in studies using simultaneous masking. Furthermore, in contrast to earlier studies, the sharpness of tuning doubles over the range of frequencies tested, giving \( Q_{\text{ERB}} \) values of about 10 and 20 at signal frequencies of 1 and 8 kHz, respectively. The estimates of auditory filter bandwidth made by Oxenham and Shera (2003) may provide a more accurate estimate of human cochlear tuning at low levels than earlier estimates using simultaneous masking at higher levels.

Also shown in Fig. 43 Qs estimated by more conventional forward-masking techniques by Glasberg and Moore (1990)

\[ Q_{\text{GM}} = \frac{9.27f_T}{f_T + 228.9} \]

Note that \( Q_{\text{GM}} \) grows from roughly 7.5 to a maximum of 9.27 and are remarkably similar to the findings of Oxenham and Shera (2003).
Figure 43: Mean $Q_{\text{ERB}}$ values in simultaneous masking with signal levels of 10 (inverted triangles) and 35 dB SL (up-pointing triangles) and in forward masking (squares) of 10 DB SL tonal probes, Oxenham and Shera, 2003. The solid curve without symbols plots results from Glasberg and Moore (1990). The filled circles (and vertical ranges) are estimates of $Q_{10\text{dB}}$ from Moore (1978) based on the widths of forward-masked psychophysical tuning curves for 10 dB SL tonal probes.
8 Pulsation Thresholds

Houtgast (1974) developed a technique that permits rapid measurement of non-simultaneous masking effects based on the perception of pulsations. In this method, masker and maskee are presented in relatively brief pulses (e.g., 125 ms with 20 ms rise and fall times) that alternate in time (e.g., Fig. 44). Each fourth burst of the maskee is deleted. For most maskee levels, the listener hears pulsations, three brief maskee bursts in a 1 s cycle. However it is possible to adjust the level of the maskee to minimize the pulsations, achieving “continuity” corresponding to the perception of one long maskee burst in a 1 sec cycle. Houtgast’s measurements suggest that the pulsation thresholds can be used to investigate the existence of a variety of phenomena observed in experiments on non-simultaneous masking.

The measurement of pulsation thresholds reflects the ability of the listener to focus on one component of a complex, time varying percept and to determine when there is minimal temporal fluctuation in the strength of this component. At threshold, the neural excitation produced by the two temporally alternating components of the stimulus is presumably equal.

8.1 Tone on Tone Masking

In one of the simplest demonstrations of the use of the pulsation threshold technique, the “signal” is a tone of frequency $f_S$ and the “masker” is also tone of the same frequency. In the case of simultaneous masking, the experiment is formally equivalent to an intensity discrimination experiment for tonal signals. The addition of the signal increases the intensity of the masker from $I$ to $I + \Delta I$. Since tone intensity discrimination is roughly described by Weber’s Law,

$$\frac{\Delta I}{I} \approx k_W \approx 0.1$$

so that the power increase associated with the addition of the signal is roughly $0.1I$. If the masker and signal are in phase, then

$$p_M(t) = \sqrt{2}P_M \cos 2\pi f_st$$
$$p_S(t) = \sqrt{2}P_S \cos 2\pi f_st,$$
Figure 45: Demonstration of the effect of suppression by the measurement of tone on tone masking. The panels in the left column illustrate the stimulus configurations; the panels in the middle column present the masked threshold levels ($L_T$) of the 1000 Hz maskee in the presence of a 1000 Hz masker of level $L_1$; the panels in the right column present the masked threshold levels ($L_T$) of the 1000 Hz maskee in the presence of a 1000 Hz masker of indicated level plus a 60 dB masker of frequency $f_2$. Data points are the average of four measurements from each of two listeners. Figure 5.1 from Houtgast (1974).
and $I$ is proportional to $P_M^2$ and $I + \Delta I$ is proportional to $(P_M + P_S)^2$. Thus

$$\frac{(P_M + P_S)^2}{P_M^2} \approx 1.1 P_M^2$$

and

$$20 \log P_S \approx 20 \log P_M - 26 \text{dB}$$

If on the other hand,

$$\frac{\Delta I}{I} \approx k_W \approx 0.05,$$

(reflecting the improvement of intensity discrimination at high tone intensities, then

$$20 \log P_S \approx 20 \log P_M - 32 \text{dB}$$

In cases of non-simultaneous masking (forward masking and gap masking) the dependence of the masked threshold on masker level is weaker than that seen in simultaneous masking: thresholds increase by roughly 0.5 dB per 1 dB increase in masker level. This “compressive” growth of masking is similar to that reported by Moore and discussed above.

The addition of a second masker tone simultaneous with the 1000 Hz masker has distinctly different effects on masked thresholds in simultaneous and non-simultaneous masking. In simultaneous masking, the addition is only found to produce increases in masked thresholds. When $f_2 \approx f_1$, the powers of the two maskers appear to add, with the sum playing the same role as the single-tone power in the case of a single masker. In non-simultaneous masking, the addition of a second masker reduces masked thresholds for some range of $f_2$ relative to $f_1$. This reduction in masked thresholds has been interpreted as a suppression of the effective strength of the masker tone of frequency $f_1$ by the masker tone at frequency $f_2$.

According to Houtgast’s results, the suppression effect is largest, both in terms of the magnitude of the reduction in masked threshold (up to roughly 20 dB) and frequency range ($1100 < f_2 < 2000 \text{ Hz}$) in the case of pulsation thresholds.

### 8.2 Auditory Mach Bands

In vision, lateral suppression is thought to be responsible for phenomena such as Mach Bands, the apparent increase in contrast at the edges of a region of uniform luminance gradient. Such phenomena had never been found in the auditory sense before Houtgast measured tone thresholds for frequencies close to the edge of a low-pass noise that had been filtered with a steep attenuation characteristic (Fig. 46).

In simultaneous masking, masked thresholds in the unfiltered noise exhibit roughly the same dependence on frequency as reported by Hawkins and Stevens (1950). Masked thresholds in the filtered noise are roughly the same as in unfiltered noise for tone frequencies below the low-pass cutoff frequency, but are lower than those thresholds for tone frequencies above the cutoff frequency.
Figure 46: Demonstration of auditory effects similar to the Mach band in vision by the measurement of the masking of tones by broadband noise and sharply low-pass-filtered noise. The top left panel illustrates the phenomenon of Mach Bands, the top right panel shows the configuration of the acoustic stimuli. The remaining panels in the left column illustrate the threshold measurement conditions, with thresholds shown in the right panels. Data points are the average of two measurements from one listener. Fig. 3.1 from Houtgast (1974).
In non-simultaneous masking, a different relationship is seen between thresholds in the presence of filtered and unfiltered noise: for a range of frequencies near the low-pass cutoff frequency, thresholds in the low-pass noise are higher than in the unfiltered noise. The removal of noise components above the cutoff frequency increases the masking produced by noise components below the cutoff frequency.

Figure 47: Tone thresholds in noises of varying bandwidths as determined with direct (simultaneous) masking and pulsation threshold techniques. Data points are the average of two (direct masking) or three (pulsation thresholds) measurements from each of two listeners. Fig. 8.3 from Houtgast (1974).

8.3 Dependence of Pulsation Thresholds on Noise Bandwidth

The increase in thresholds that accompanies attenuation of noise components above the low-pass cutoff frequency seen in Fig. 47 for the non-simultaneous masking conditions is in sharp contrast to the findings of Fletcher (1940) in the course of his attempts to measure the width
of critical bands. Fletcher’s estimates were based on the results of simultaneous masking experiments in which tones were detected in the presence of noises of varying bandwidths. Fletcher found that attenuating noise components had no effect on thresholds when the masker components were spectrally remote from the masker, or reduced thresholds when they were spectrally close to the masker.

However different masking patterns are obtained when Fletcher’s stimuli are used in an experiment in which the noise masker and tone signal presented non-simultaneously (Fig. 47). In particular, whereas widening the noise band beyond the width of a critical band has little effect on the threshold in simultaneous masking, it lowers the threshold in non-simultaneous masking.

8.4 Tuning Curves

The assumption of equal stimulation during the two phases of the pulsating stimulus is similar to the assumption that the neural stimulation produced by the probe is constant along the psychophysical tuning curve. Indeed, similar tuning curve shapes are obtained forward masking and pulsation threshold experiments (Fig. 48).

8.5 Two-tone Suppression

In physiological experiments, the neural firing rate produced by a tone whose frequency is that to which an auditory neuron is most sensitive is reduced if a second tone of different frequency is added at an appropriate level. The phenomenon, two-tone rate suppression, was discovered by Sachs and Kiang (1968). Houtgast (1974) has demonstrated that similar regions can be determined using measurements of Pulsation Thresholds (Fig. 49).
Figure 48: Psychophysical tuning curves measured using simultaneous (direct) and non-simultaneous masking and by the pulsation threshold technique. Three measurements from one listener for each condition. Fig 4.1 from Houtgast (1974).
Figure 49: Determination of regions of two-tone suppression using the pulsation threshold technique. Fig 5.3 from Houtgast (1974).
References


