

PSYCHOACOUSTICS AND AUDIBILITY - FUNDAMENTAL ASPECTS OF THE HUMAN HEARING

Lecture note for the course TI-EAKU

*Lars G. Johansen
University College of Aarhus, 2006*

-
- 1 Understanding the human hearing by modelling
 - 2 Loudness, pitch, and timbre perception
 - 3 Audibility of effects present in listening rooms
 - 4 Audibility of effects in generalised transfer functions
 - 5 Subjective perception of physical measures
 - 6 Parameters of subjective evaluation
-

1 Understanding the human hearing by modelling

It is safe to say that the human hearing does not act as an ordinary and simple measurement device. Fortunately, from loads of experiments in the last century we know quite a lot the behaviour of the human hearing. If no particular references in the text below are given, the facts referred are mainly extracted (and condensed heavily) from the textbooks [1], [2], and [3].

Auditory filters

It is a well known fact that the human perception of sound does not have a linear dependency on the commonly used Hertz frequency scale. From analyses of the physical parts of the human ear and hearing system, the predominant understanding of the peripheral auditory system is that of a bank of filters with overlapping pass bands. These are commonly referred to as auditory filters, see e.g. the descriptions in [3].

The masking effects

Masking is a psychoacoustic effect that has been used to determine the critical bandwidth (CB) of the auditory filters. Masking is the effect by which one sound, the masker, raises the threshold of audibility of another sound. It can be used in a quantitative way to find the amount by which the threshold of audibility of a sound is raised by the presence of another masking sound. Masking occurs in both time and frequency domains.

Temporal masking basically appears in two forms; pre-masking and post-masking. Pre-masking occurs when the audibility of sounds, which has already appeared, are masked by a sound appearing now. This effect can be observed up to 20 ms before the louder sound appears. Normally, pre-masking only occurs within a region of 5 ms, and pre-masking only occurs when the difference in level is substantial. It has been suggested that trained subjects will not report pre-masking at all, see [1], the explanation being that relatively untrained listeners simply confuse the two signals. Post-masking (also known as forward masking) occurs when the masker precedes the signal. In studies of this phenomenon the duration of the masker is altered and it is observed which levels and time gaps will result in masking of the signal. Post-masking effects are observed for time gaps up to about 200 ms depending on masker and masked signal levels.

In frequency masking (also known as simultaneous masking) two sounds appear at the same time. There are several forms of simultaneous masking. One is known as broadband masking in which the presence of a pure tone is effectively oppressed by a broadband noise. Signal-to-masker levels are determined by presenting a tone in a noisy background and adjusting the tone to be just detectable. Another form of simultaneous masking is narrow band masking in which the presence of one pure tone masks signals of lower intensity in the same frequency range. This masking mechanism is not symmetrical around the masking tone, a phenomenon known as the ‘upward spread of masking’ is observed. It means that the high frequency side of the masker is more effectively masked than the lower side.

Critical bands (CB)

The concept of a critical bandwidth is used to describe the bandwidth at which a masker ceases to increase the threshold of detection of the masked sound. For frequencies from 0.5-1 kHz and above the human hearing is primarily sensitive to signal energy, [1], which, for most ordinary listening rooms, complies nicely with the Schroeder frequency beyond which we can consider the sound field in a room diffuse and hence treat it statistically, e.g. with energy measures. An analytic expression for the critical bandwidth as a function of frequency is given by eq. 1.

$$\Delta f_{CB} = 25 + 75 \left[1 + 1.4 \left(f \left[\text{kHz} \right] \right)^2 \right]^{0.69} \quad (1)$$

The CB Bark scale

A scale based on the critical bands (CB) was defined by Zwicker in 1961 known as the *Bark* scale. Through eq. 2, the ordinary frequency scale is mapped into the *Bark* scale defining 24 equal relative bandwidth bandpass filters of approximately 1/3 octave above 500 Hz - modelling the energy sensitivity. Below 500 Hz the critical bandwidth is almost constant - about 100 Hz. Instead of the Bark critical bandwidth, the Equivalent Rectangular Bandwidth (ERB) has been proposed by Moore and Glasberg. Below 500 Hz the bandwidth of ERB is narrower compared to CB. Above 500 Hz they appear to be fairly equal.

$$v = 13 \arctan(0.76 f [kHz]) + 3.5 \arctan\left(\frac{f [kHz]}{7.5}\right)^2 \quad (2)$$

Thus, the ear performs a spectral analysis using a filter bank, critical bands described by the Bark scale. The critical band model is primarily based on loudness of stationary broad band signals but is valid for many other psycho-acoustic experiments, for example it is believed that tonal balance (timbre) is related to a 1/3 octave analysis. For different signals within one critical band the auditory system works as an energy integrator expressed by a summation of the individual sound pressures. The effective value of the perceived sound pressure within one critical band is thus given in eq. 3.

$$p_{tot} = \sqrt{p_1^2 + p_2^2 + \dots + p_n^2} \quad (3)$$

Frequency energy integration

The human auditory system has been modelled in various ways with respect to frequency resolution and frequency dependent sound pressure sensitivity. There is some agreement that for many auditory phenomena and at least for frequencies above 500 Hz a filter bank model seems appropriate. This modelling is primarily based on perceived energy considerations. For low frequencies the human auditory system is considerably discriminative and sensitive to narrow band phenomena such as the modal resonance frequencies in closed rooms which cause more or less narrow peaks and dips in the magnitude spectrum. The human hearing seems to be more sensitive to peaks than to dips, which is fortunate in the sound field correction quest since damping is generally more desirable (and easier to accomplish) than amplification.

Frequency resolution

At 100 Hz the frequency selectivity is about 1 Hz, i.e. two tones at 100 Hz and 101 Hz can be distinguished from one another by the auditory system. This fine frequency resolution is however not only due to frequency analysis in the auditory system. A kind of temporal analysis is also applied. The time pattern of neural activity also carries information but probably only below 4-5 kHz.

Temporal energy integration

It is assumed that the auditory system contains a temporal energy integrator, i.e. it performs a summation of the input signal. A simple way of estimating the relationship between thresholds and durations is to plot threshold against duration on a dB vs. logarithmic-time scale. Data will fall roughly on a straight line with a slope of -3dB per doubling of log duration. Letting J represent the integration time of the auditory system, several scientists have estimated its magnitude, - the estimated values lying in the region of 50-200 ms. Some researchers report that J is greater at low frequencies while others have found no frequency dependency.

Another way of determining the integration ability of the auditory system is to present equal energy tone bursts of different durations. An ideal energy integration would imply that the detectability of these tone bursts would be independent of duration. According to Green, this is only the case in the region 15-150 ms outside of which the detectability will fall off. The fall off at long durations indicates that the integration operation is time delimited, while the fall off at very short durations might be a result of the spread of energy over the frequency range that occurs for short duration pulses. Other scientists have found similar results, but again there is some variation in the results, - some scientists reporting frequency dependency (low frequency, long duration and vice versa) and others do not. According to [4], and essentially also [1], the integration time is frequency dependent, about 60 ms up to 1000 Hz decreasing linearly to around 10 ms at 5 kHz. This is fairly consistent with an effective time constant (or integration time) for speech around 20 ms. According to Niese, speech intelligibility can be predicted using an integration of so-called useful energy in the range up to 17 ms (full weight) and a linearly decreasing weight factor in the range 17-30 ms.

Temporal resolution

In order to estimate the temporal resolution of the auditory system a number of experiments have been carried out, in which the subjects are asked to detect a temporal gap in a sound signal. If the gap is introduced in a broadband noise, e.g. white noise, there will be no effect on the magnitude spectrum. In an investigation performed by Plomp the subjects were presented with two successive white noise signals in random order, one of which contained the gap, and were asked to judge which one contained the gap. This typically showed gap thresholds of 2-3 ms. Investigations based on broadband signals inherently lack information on the frequency dependency of the temporal resolution. To account for this Shailer and Moore in 1983 devised an experiment in which the signal was a noise signal with a bandwidth of half the investigated centre frequency. The results showed decreasing gap thresholds with increasing centre frequencies, the threshold being 22 ms at 200Hz decreasing monotonically to about 3 ms at 8 kHz. This finding suggests that the ringing of the auditory filters is important in temporal resolution as the bandwidth and ringing are inversely proportionally related, see [5].

2 Loudness, pitch, and timbre perception

Loudness perception

Loudness is the perceived level of sound, i.e. a subjective measure. The loudness perception is determined by both level and bandwidth of the sound, but these are not the only factors. If a sound exists within one critical band, it only excites a single auditory filter. If however it has a bandwidth exceeding a critical band, the loudness perception is increased since more than one auditory filter will be activated. For loudness the ear employs an integration time in the order of 100 ms (the IEC “fast” integration time being standardised to 125 ms). An estimate of loudness of a sound can be calculated, see [3].

Pitch perception

Pitch is essentially that auditory sensation which enables ordering of sounds on a musical scale. It is a subjective measure and will normally be found by specifying that pure tone to which a signal corresponds. The American Standards Association, ASA, has defined pitch as:

"that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale."

When a sound consists of clicks at a given frequency interval, the spectral pattern will in theory include an infinite number of harmonic components at integer multiples of the fundamental frequency. When a tone complex of this sort is presented, the pitch of the signal will be judged to be that of the fundamental even when the fundamental is missing.

Timbre perception

Timbre is perhaps the single most important subjective acoustical measure of a sound. The ASA has defined timbre as:

"that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar."

In [3], timbre is described as a residual basket of sensations containing anything but loudness and pitch, i.e. a vague definition of a very complex and extensive auditory sensation. The perception of timbre is not yet fully understood, it is however clear that one has to be familiar with a sound in order to judge the timbre properly. It is the harmonic patterns of most natural sounds we can hear out, thus the perception of timbre is closely related to spectral distribution. Timbre can be split into at least two important and inversely related sensations, *sharpness* and *pleasantness*, the latter is not an independent measure but is again related to *roughness*, *loudness*, and *tonalness*.

Other aspects of the hearing

Other aspects of the hearing are the dynamic range described by Fletcher and Munson, the linearity (or non-linearity), possible interaction between the two slightly different perceived sounds by the two ears (binaural perception), and the issues of “combined time-frequency” hearing which may lead us to understand why some phenomena are difficult to examine in one single domain.

3 Audibility of effects present in listening rooms

Audibility of a single reflection

The audibility of a single reflection depends on the level, delay-time, direction, and signal type. For speech in a living room (reverberation time 0.8 s), echo disturbance is observed for a single reflection if the delay time exceed approximately 30 ms, see [3], (reflection 0 dB re. to the direct sound and both arriving from the frontal direction). The threshold is about 10 dB lower if the reflection arrives from about 45 degrees. For short pulses, echo disturbance can be observed for delay times of about 10 ms. The threshold, $L = -0.6 t_0 - 8$ dB, can be found in [6] where t_0 is time delay in ms (both signals arriving from the front). For classical music, echo disturbance is observed at about 80 ms delays.

Audibility of multiple reflections

Some investigations indicate that individual reflections may not entirely determine the perceived quality of reproduced sound. Instead the whole pattern of early reflections should be considered as a key to perceived quality, - at least when dealing with average size listening rooms. A study is reported by Bech in [7]. See also the results from the Athena project as reported in [8].

Colouration by comb filtering

A closely related issue is colouration caused by a single reflection, or multiple reflections from two parallel walls (flutter). The impulse response for a direct sound combined with a reflection is given by $h(t) = \delta(t) + a \delta(t-t_0)$. This filter is known as a comb-filter, the frequency response is a curve with equidistant dips, the distance between the dips, f_{comb} , is given by $1/t_0$. The audibility depends on the delay time and the level of the delayed component (reflection). Human beings are particularly sensitive to colouration caused by delays of about 5 ms. Colouration can be regarded as a frequency domain effect (a change in timbre) for delay times up to approximately 25 ms, see [9] and [10]. If t_0 exceeds 25 ms the perception changes from colouration to a rough character effect where the regular repetition is detected as a time domain effect. This is a very important observation, since the ear probably performs some kind of joint time/frequency analysis based on short-time spectral analysis.

The image shift effect

If a direct sound and a lateral reflection with a delay in the range from 0-2 ms (maybe up to 5 ms) is combined an image shift can be observed depending on the level of the lateral reflection (summing localisation).

The precedence effect

Localisation of a sound source in reflecting surroundings is an important issue. Localisation errors are suppressed substantially according to the "law of the first wave front" (or precedence effect). The auditory "processor" is able to extract the direct sound and use this information for sound localisation. Reflections delayed in the interval 5-50 ms are suppressed in the sound localisation process. The exact boundaries depend on the type of signal and cognitive factors, and the precedence processor requires a "transient signal", e.g. onset/offset of a signal, to be effective and it works best for short broadband excitation signals.

4 Audibility of effects in generalised transfer functions

Frequency magnitude deviations

Taking a starting point in a transfer function, e.g. a loudspeaker/room transfer function (or a purely synthetic one), it is of significant interest to what extent the magnitude deviations from an ideal transfer function (flat magnitude spectrum and linear phase) are audible. The transfer functions dealt with in this work are always characterised by low frequency peaks and notches simply because of the room acoustics.

Audibility of peaks and notches

The audibility of peaks and notches in a magnitude spectrum is not equal. Bücklein [11] made some experiments revealing this fact, and later on these early studies have been examined more thoroughly by Toole and Olive in [12] and by Olive et al. in [13] (see also [5]). The two main conclusion are:

- notches are far less audible than peaks,
- peaks and notches become increasingly audible as bandwidth is increased.

Peaks usually occur as a result of resonances, and notches (or dips or troughs) usually result from anti-resonances. The audibility of resonances and the way they modify timbre is thoroughly investigated by Toole and Olive in [12]. They describe resonances as a function of centre frequency, Q-value, time delay, programme material, loudspeaker directivity, and reverberation added to recording or reproduction. The main conclusions of their work are summarised below.

One primary factor controlling the audibility of resonances is the Q-value. It is found that resonances of low Q (broad and soft peaks) are more audible than high-Q ones. It is also found that the duration (the ringing) of the resonance is not a good measure of the audibility. This is for example the case when some other part of the spectrum partly or entirely masks the resonance. The audible effects of electronic manipulations (e.g. equalising) are found to depend on the amount of reverberation in the programme material and in the surroundings. Further investigations by Olive et al. [13] show that the centre frequency of a resonance also has significant effect on the audibility. For Q-values around 1, the threshold of audibility for resonances is constant as a function of centre frequency, but for Q-values above approximately 10 the threshold of detectability is reduced by 0.5-2.0 dB for each decrease of an octave below 500 Hz. In [13] it is also stated that resonances and anti-resonances of similar amplitudes are equally detectable for Q-values around 1. For Q-values above 10 however, anti-resonances become significantly less detectable than resonances - at least when using pink noise as excitation material. Below approximately 100 Hz high-Q notches are found to be essentially inaudible. The programme material is of great importance when experimenting with the detectability of resonances. Pink noise and white noise are the most sensitive excitation signals while speech and musical signals are progressively less usable for evoking audible resonances. Based on this programme material dependency, [13] concludes that the thresholds found in the experiments, which for the

major part are carried out using pink/white noise presented through headphones, will probably be reduced greatly when performing the same experiments using loudspeakers in listening rooms and/or when using natural signals.

Phase and group delay

In [14] Karjalainen *et al.* among other issues discuss the audibility of phase distortion. It seems like that the just noticeable difference for group delay is about 2 ms, mostly excited by transient and impulse-like signals. The human hearing seems to be much more tolerant to group delay differences when using wide-band steady-state signals. For further discussions on the audibility of phase distortion see [15], [16], and [17].

Excess phase audibility

Until recently no experiments have been reported on the audibility of the excess phase in loudspeaker/room transfer functions. The theory of minimum and excess phase splitting of such transfer functions will not be discussed in detail in this section. Instead the reader should pay attention to the paper [18] of the author published as an AES preprint in 1996. In [18] is shown that, under certain conditions, the isolated excess phase is audible. Two listening tests with modification filters were performed in this piece of work. The first one showed that when extracting the excess phase part of a recorded loudspeaker/room impulse response and imposing this on an anechoically recorded signal, the difference between the compound and the raw unmodified signal is indeed audible. The second listening test showed that the minimum phase part produces some masking of the excess phase part. This result emerges from imposing the same minimum phase part of the impulse response on both signals in experiment one - thus yielding a signal with the entire room impulse response and one without the excess phase part.

Two different room impulse responses were used, one from a room having reverberation time $T_{60} = 450$ ms and one with $T_{60} = 600$ ms. Excess phase was in both experiments more audible using the 600 ms room. Impulse response length was also varied, 150 ms and 300 ms. Again the longer the response, the more audible the excess phase part. In fact, the longer the impulse response, the more excess phase will be present. Lastly, the experiments showed that when using music as source material, the excess phase became considerably less audible than when using male or female voices.

5 Subjective perception of physical measures

Room size and volume

One of the most important parameters is still the reverberation time T_{60} , perception of room size/volume is well correlated with T_{60} . It also accounts for the subjective impression of reverberation as a diffuse decaying process. Especially the early decay time, EDT, correlates generally well with the subjective impression of room acoustic quality. The masking properties of the reverberation process are likewise described by T_{60} to a first approximation.

Early/late energy relations

It has also been shown that the classic room acoustic parameters Clarity, Definition, and DR to some extent correlate with perceived quality. A more general approach though is based on joint time/frequency analysis. The Short Time Fourier Transform is a simple way to form a time frequency distribution, while the Wigner-Ville distribution might enable important resolution enhancement.

Distance perception

The parameter DR is important in relation to the distance perception. In an investigation by Michelsen, see [19], it is found that Clarity (C80) also has a substantial influence on the distance perception. In this investigation it is also suggested that distance perception is highly dependent on early reflections.

6 Parameters of subjective evaluation

Colouration

Colouration can be described as a characteristic change of the signal spectrum. Whether an ideal comb filter (or room with similar impulse response behaviour) will produce audible colouration or not depends on the delay time T_L and the Q-factor. If T_L exceeds a certain value, about 25 ms, the effect does not appear subjectively as colouration (change of the timbre of sound). The sound will tend to have a rough character, we simply “become aware” of the regular repetition of the signal in the time domain.

Roughness

The psychoacoustic parameter *roughness* is associated with modulation frequencies in the frequency range 30-300 Hz and maximum at about 70 Hz. The observed change in the sound character for delay times around 25 ms corresponding to 40 Hz between succeeding peaks could be described as roughness.

Sharpness

Sharpness is a relatively independent sensation, affecting timbre though. It is measurable and hence the sharpness of two sounds can be compared. The most important variables that influence sharpness are:

- spectral envelope - the spectral content and the centre frequency of narrow band sounds. Sharpness increases when the centre frequency increases,
- bandwidth - sharpness is increased when bandwidth is increased.

Randomness

The human ear is not only a frequency analyser but is also sensitive to the temporal structure, e.g. periodic events. The most adequate technique for investigation of the lack of such (i.e. the randomness or pseudo-randomness) of an impulse response is based on autocorrelation analysis. A quantitative measure for *randomness* can be obtained using the autocorrelation function of an impulse response. This measure is the temporal diffusion index TD. The perception threshold for comb filtering of white noise is known to be a function of t_r (the repetition period) and q (relative level of repeated event), see [6]. The maximum sensitivity for repetition lies in the range 1-20 ms. Based on

this a criterion for audibility of periodic components can be developed, see eq. 4. Essentially, it says that repeated events must be attenuated more than 12 dB in order not to be audible.

$$\Phi_{xxw}(t_r) \geq 0.06 \Phi_{xx}(0) \quad (4)$$

Spatial impression and more ...

Lateral reflections, delayed in the range 5-80 ms, are very important for the perception of space (spatial impression), see [20]. Barron and Marshall here propose a physical measure of Spatial Impression (SI), the level of lateral sound energy integrated up to 80 ms being the important factor. Spatial impression can be decomposed in more subjective attributes, e.g. apparent source width (ASW). ASW is mainly related to the interaural crosscorrelation (IACC) and early lateral energy. Listener envelopment (LEV) is related to late-lateral-energy-level and to some extent also the reverberation time, see [21].

REFERENCES AND OTHER LITERATURE

- [1] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press 1991
- [2] J. Blauert, *Spatial Hearing*, S. Hirzel Verlag 1996
- [3] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*, Springer-Verlag 1990
- [4] E. Corliss, "The Ear as a Mechanism of Communication," *J. Audio Eng. Soc.*, **38**(9), 1990 September
- [5] M. J. Shailer and B. C. J. Moore, "Gap Detection as a Function of Frequency, Bandwidth, and Level," *J. Acoust. Soc. Am.*, **74**(2), 1983 August
- [6] H. Kuttruff, *Room Acoustics*, Applied Science Publishers Ltd. 1991
- [7] S. Bech, "Perception of Reproduced Sound: Audibility of Individual Reflections in a Complete Sound Field, III," *a preprint of the Audio Eng. Soc. 100th Conv.*, Copenhagen 1996
- [8] P. L. Schuck *et al.*, "Perception of Perceived Sound in Rooms: Some results of the Athena Project," *proc. of the Audio Eng. Soc. 12th Int. Conf.*, Copenhagen 1993
- [9] P. Rubak and L. G. Johansen, "Colouration in Natural and Artificial Room Impulse Responses," to appear in *proc. of the Audio Eng. Soc. 23rd Int. Conf.*, Copenhagen 2003
- [10] M. Brüggén, "Coloration and Binaural Decoloration in Natural Environments," *Acta Acustica/Acustica*, vol. 87, pp. 400-406, 2001
- [11] R. Bücklein, "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.*, **29**(3), 1981 March
- [12] F. E. Toole and S. E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement," *J. Audio Eng. Soc.*, **36**(3), 1988 March
- [13] S. E. Olive *et al.*, "The Detection Thresholds of Resonances at Low Frequencies," *a preprint of the Audio Eng. Soc. 93rd Conv.*, San Francisco 1992
- [14] M. Karjalainen *et al.*, "Comparison of Loudspeaker Equalization Methods Based on DSP Techniques," *J. Audio Eng. Soc.*, **47**(1/2), 1999 January/February
- [15] H. Kuttruff, "On the Audibility of Phase Distortions in Rooms and its Significance for Sound Reproduction and Digital Simulation in Room Acoustics," *Acustica*, vol. 74, 1991
- [16] R. Greenfield and M. Hawksford, "The Audibility of Loudspeaker Phase Distortion," *a preprint of the Audio Eng. Soc. 88th Conv.*, Montreux 1990
- [17] D. Preis, "Phase Distortion and Phase Equalization in Audio Signal Processing - A Tutorial Review," *J. Audio Eng. Soc.*, **30**(11), 1982 November

BIBLIOGRAPHY

- [18] L. G. Johansen and P. Rubak, "The Excess Phase in Loudspeaker/Room Transfer Functions: Can it be Ignored in Equalization Tasks?," *a preprint of the Audio Eng. Soc. 100th Conv.*, Copenhagen 1996
- [19] J. Michelsen and P. Rubak, "Parameters of Distance Perception in Stereo Loudspeaker Scenario," *a preprint of the Audio Eng. Soc. 102nd Conv.*, Munich 1997
- [20] M. Barron and A. H. Marshall, "Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure," *J. Sound and Vibration*, **77**(2), pp. 211-232, 1981
- [21] D. Griesinger, "General Overview of Spatial Impression, Envelopment, Localization and Externalization," *proc. of the Audio Eng. Soc. 15th Int. Conf.*, Copenhagen 1998