UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

MPEG Digital Audio Coding

Alberto Carini

Bit-rates

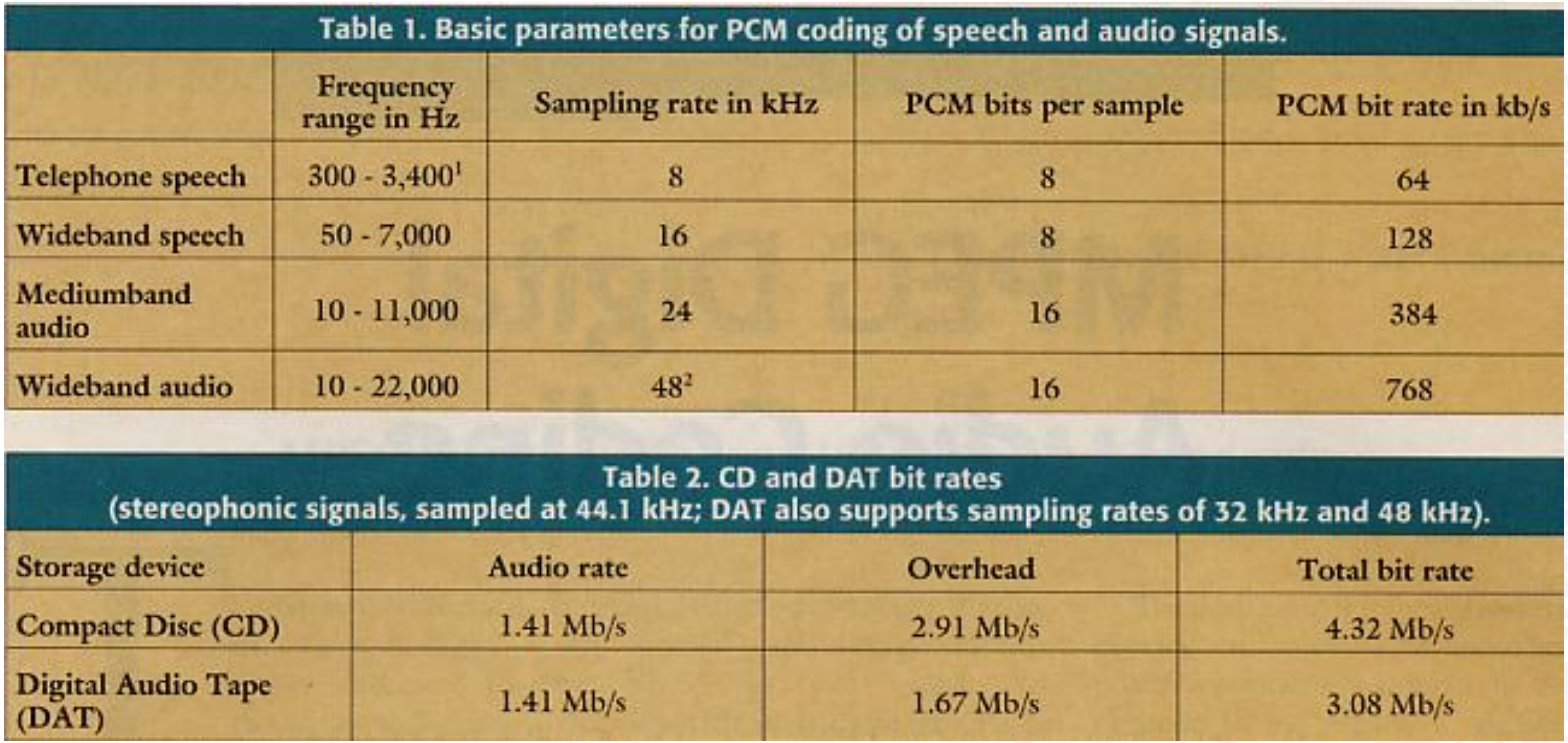


Table 1. Basic parameters for PCM coding of speech and audio signals.

	Frequency range in Hz	Sampling rate in kHz	PCM bits per sample	PCM bit rate in kb/s
Telephone speech	300 - 3,400[1]	8	8	64
Wideband speech	50 - 7,000	16	8	128
Mediumband audio	10 - 11,000	24	16	384
Wideband audio	10 - 22,000	48[2]	16	768

Table 2. CD and DAT bit rates (stereophonic signals, sampled at 44.1 kHz; DAT also supports sampling rates of 32 kHz and 48 kHz).

Storage device	Audio rate	Overhead	Total bit rate
Compact Disc (CD)	1.41 Mb/s	2.91 Mb/s	4.32 Mb/s
Digital Audio Tape (DAT)	1.41 Mb/s	1.67 Mb/s	3.08 Mb/s

Bit-rates

Differences between audio speech signals are manifold. However, audio coding implies higher sampling rates, better amplitude resolution, higher dynamic range, larger variations in power density spectra, stereophonic and multichannel audio signal representations, and, finally, higher quality expectations.

Bit-rates

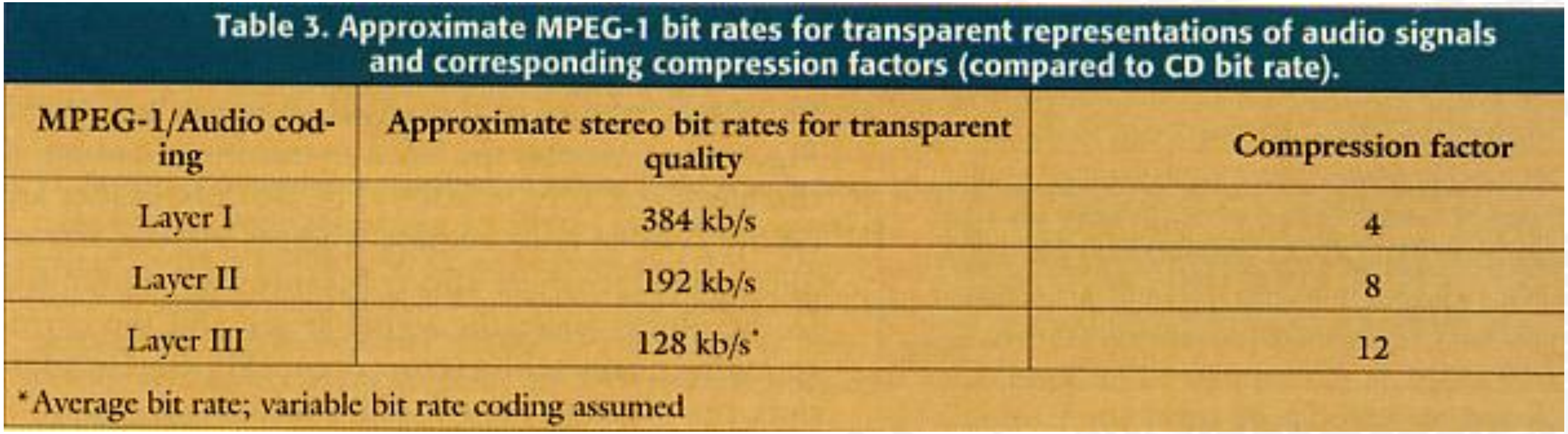
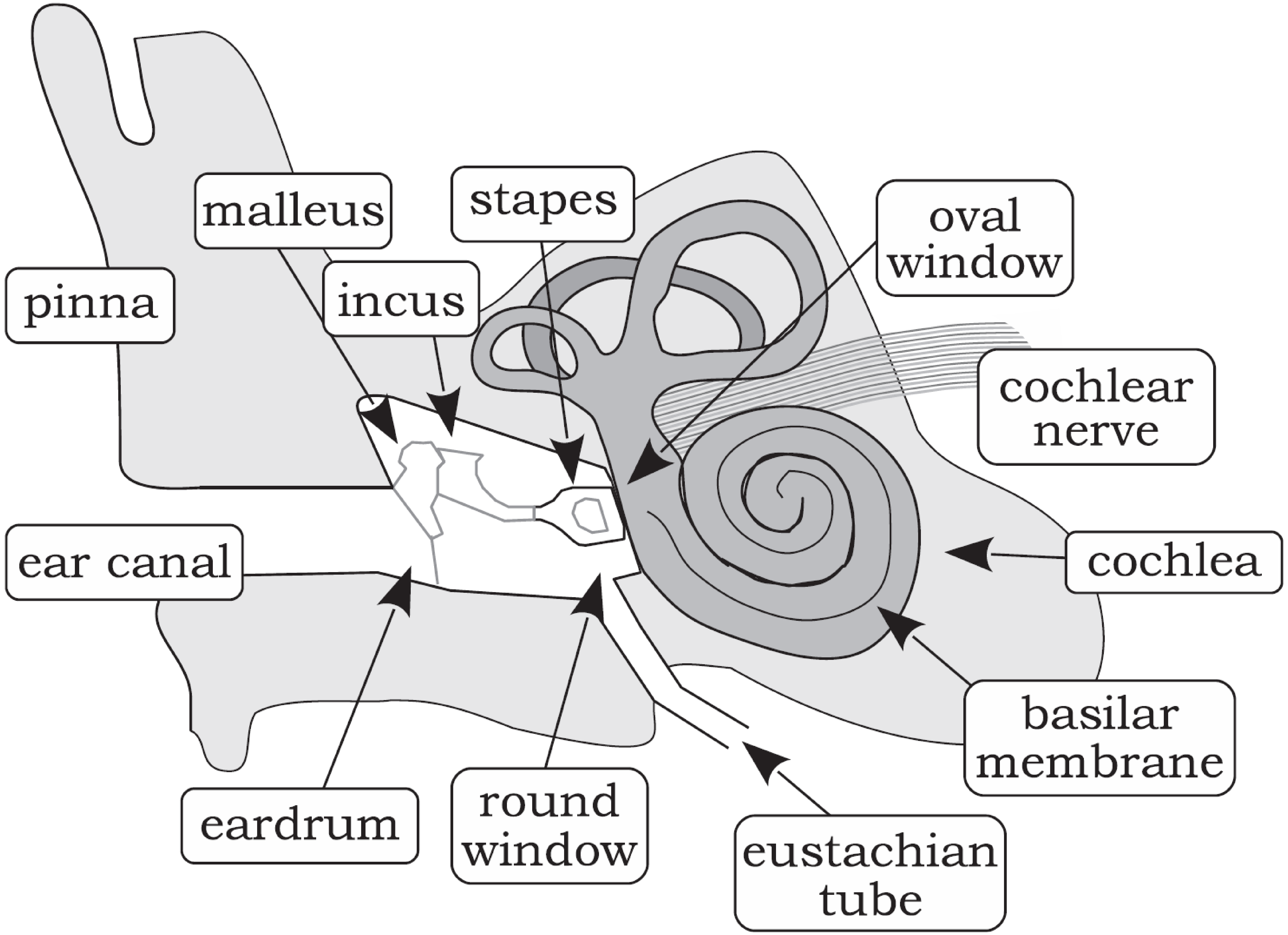


Table 3. Approximate MPEG-1 bit rates for transparent representations of audio signals and corresponding compression factors (compared to CD bit rate).

MPEG-1/Audio coding	Approximate stereo bit rates for transparent quality	Compression factor
Layer I	384 kb/s	4
Layer II	192 kb/s	8
Layer III	128 kb/s*	12

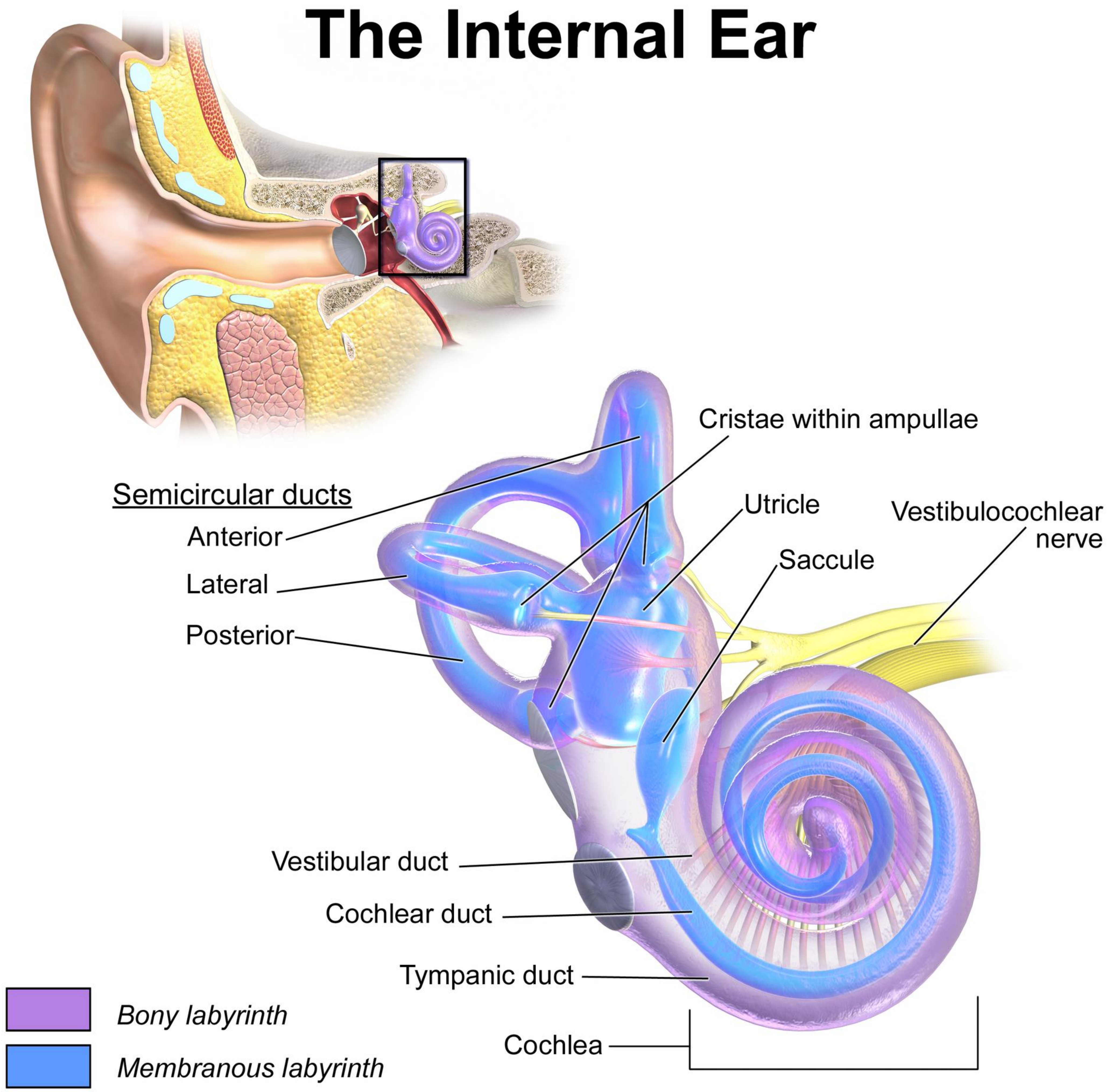
*Average bit rate; variable bit rate coding assumed

The ear

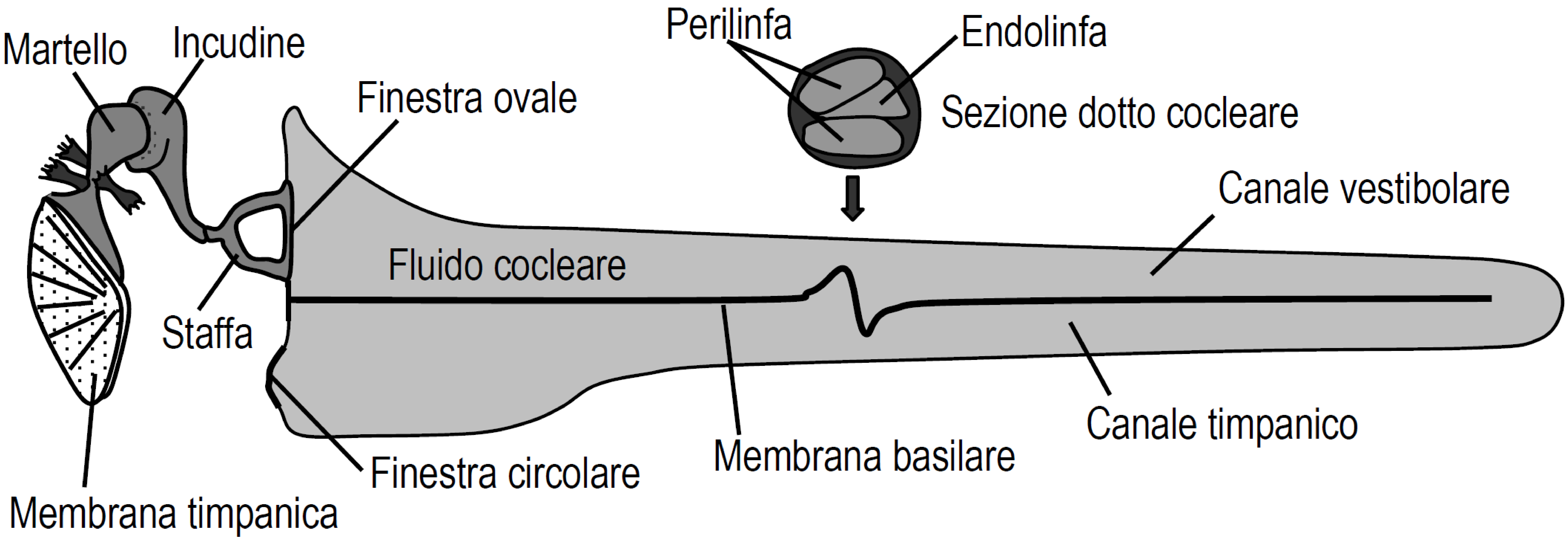


malleus
stapes
oval window
pinna
incus
cochlear nerve
ear canal
cochlea
basilar membrane
eardrum
round window
eustachian tube

The ear

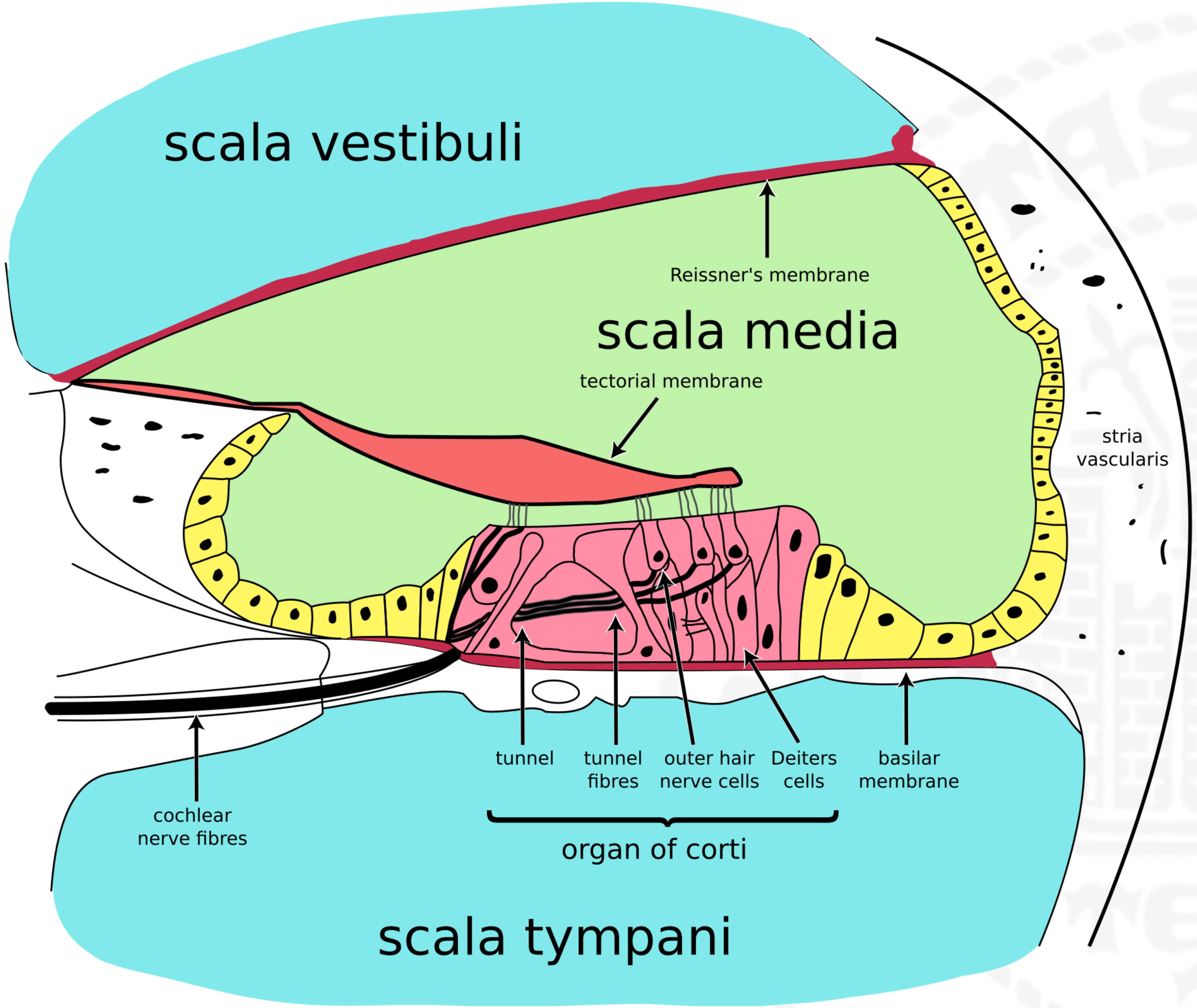

The Internal Ear
Cristae within ampullae
Semicircular ducts
Anterior
Lateral
Posterior
Utricle
Vestibulocochlear nerve
Saccule
Vestibular duct
Cochlear duct
Tympanic duct
Cochlea
Bony labyrinth
Membranous labyrinth

The ear



Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

The ear


scala vestibuli
Reissner's membrane
scala media
tectorial membrane
stria vascularis
tunnel
tunnel fibres
outer hair nerve cells
Deiters cells
basilar membrane
cochlear nerve fibres
organ of corti
scala tympani

Audible region

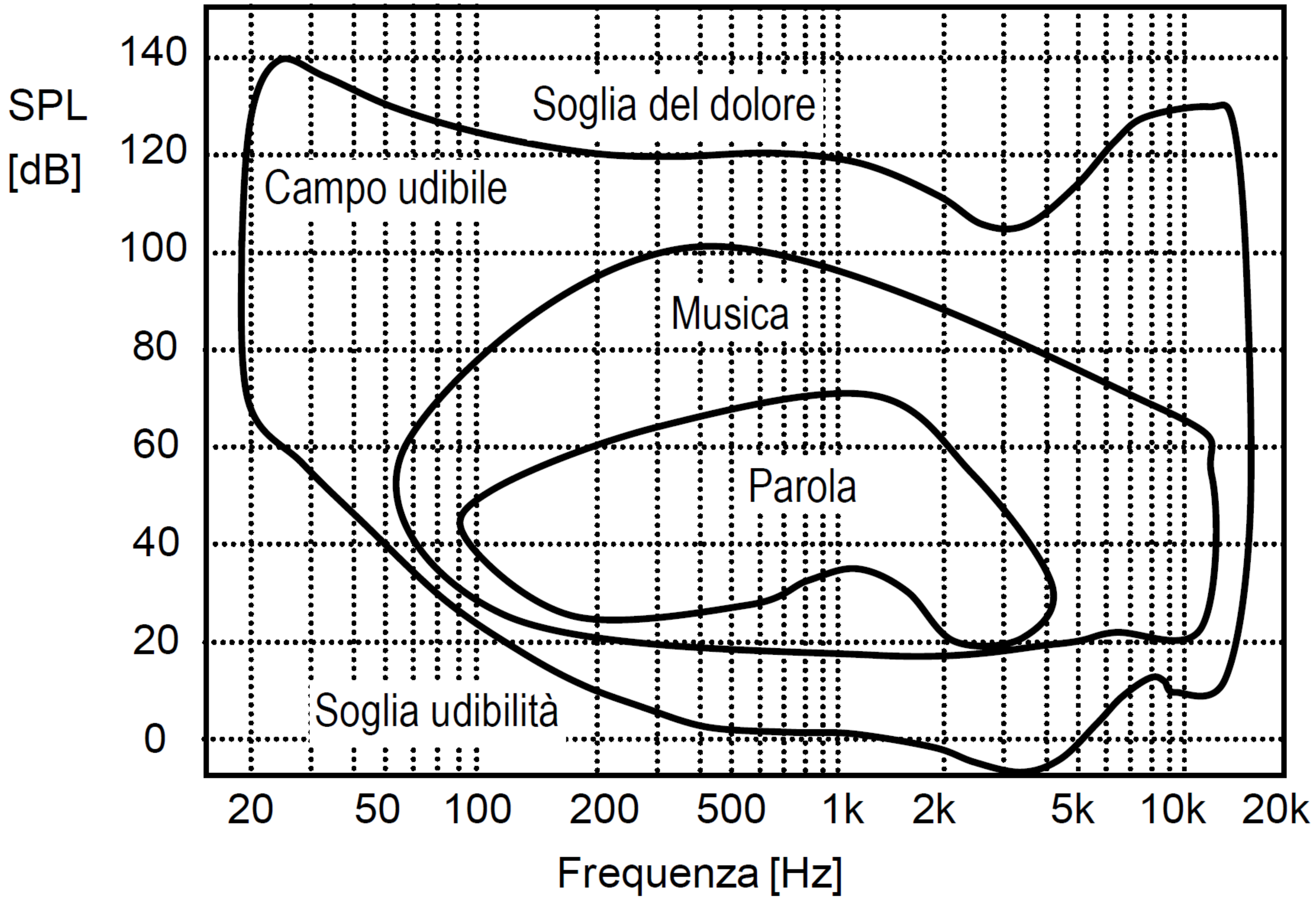


Figura 3.3 Gamma di frequenza e intensità acustica dei più comuni messaggi sonori. Limite superiore: soglia del dolore; limite inferiore: soglia di audibilità.

Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Equal loudness curves

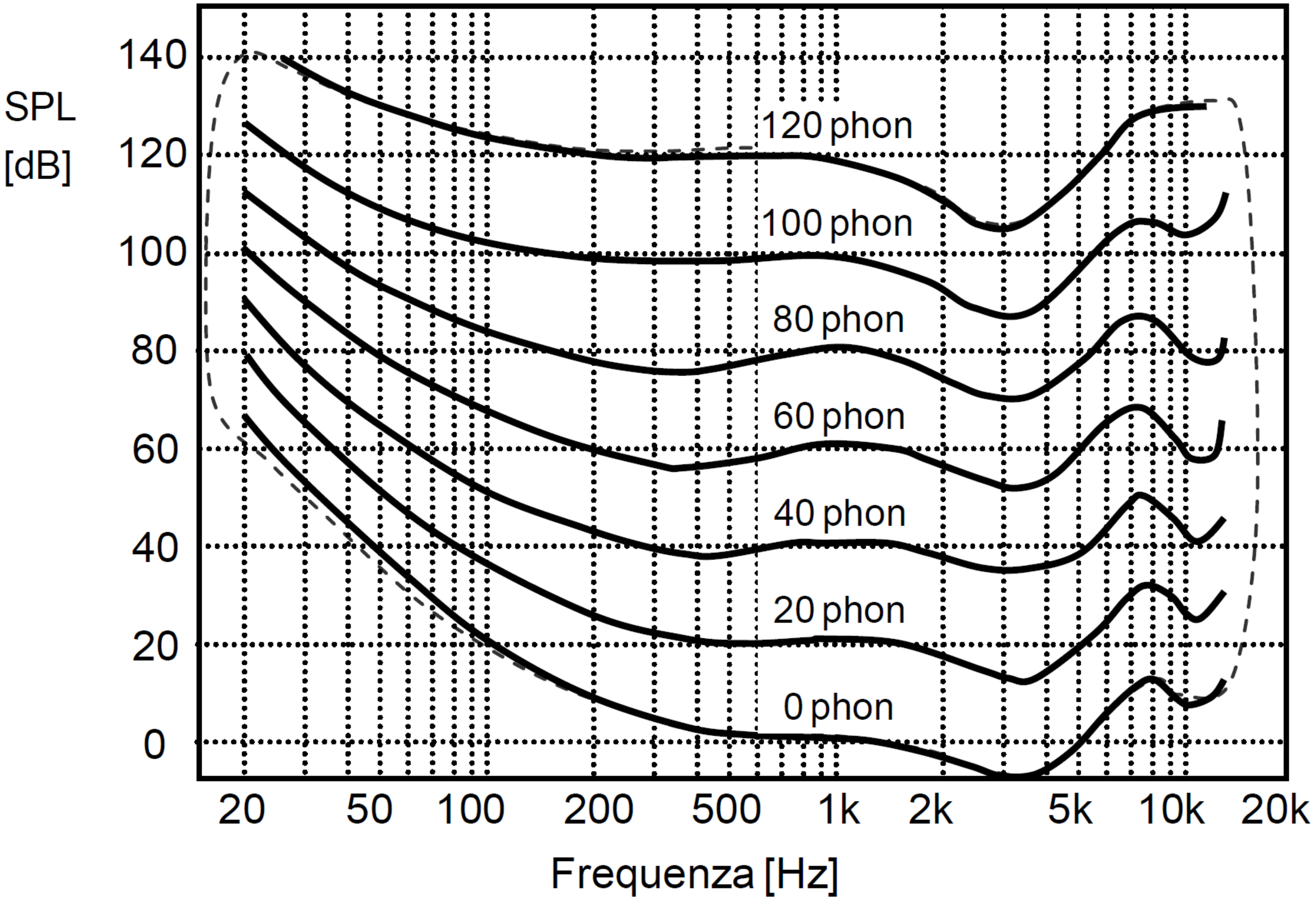


Figura 3.4 Curve isofoniche o di *loudness* di Fletcher-Munson. La famiglia di curve isofoniche prende anche il nome di *audiogramma normale* [1].

Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Sound pitch in mel

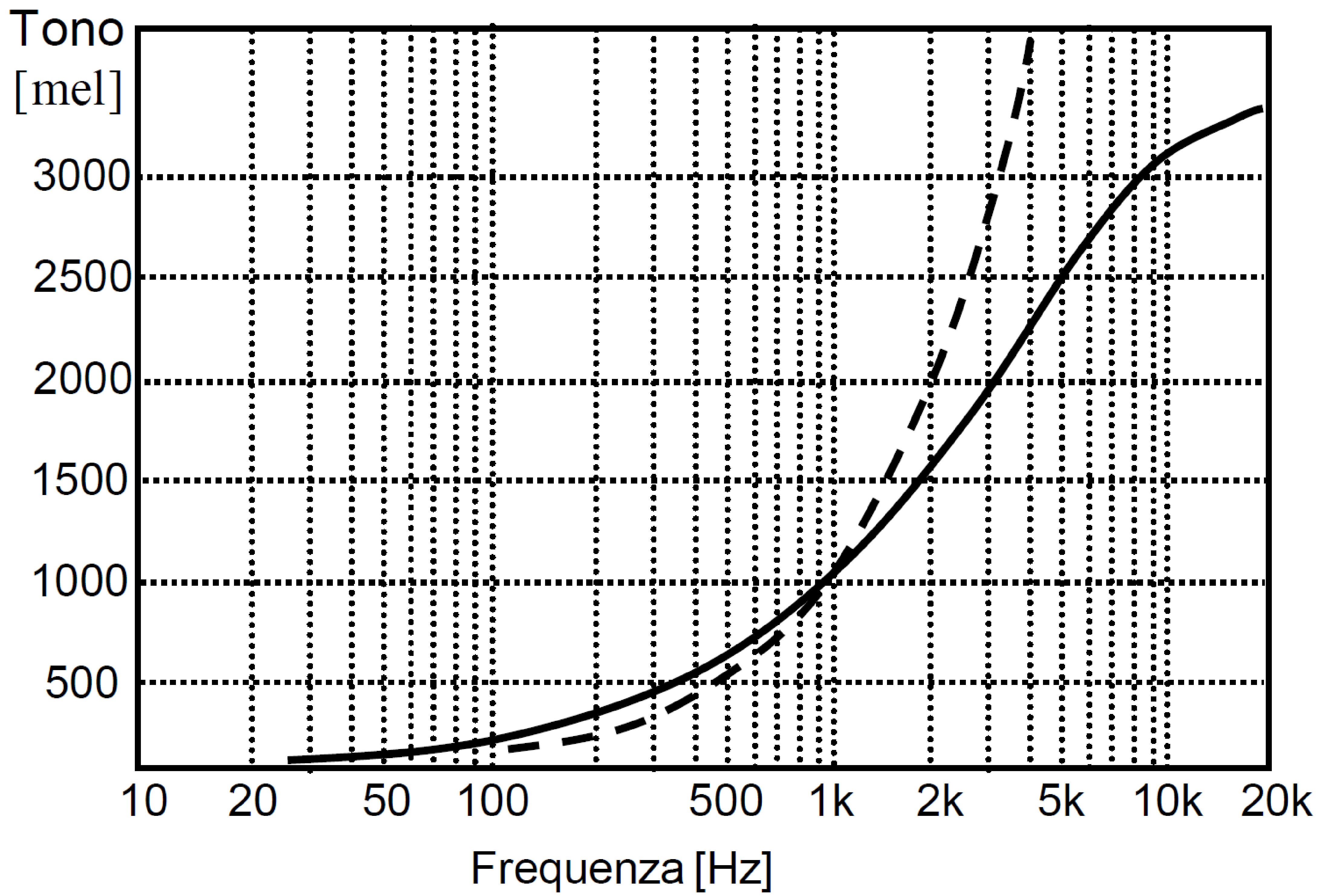


Figura 3.6 Giudizio medio di tono in [mel] vs. frequenza effettiva dell'oscillazione [Hz]. Linea continua *I* = 60 dB (curva mel). Linea tratteggiata *I* > 60dB.

Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Critical bands

- The cells of the organ of Corti work in groups of ~1300, each of which physically occupies about ~1.3mm of the basilar membrane and covers a frequency of ~1/3 of an octave.
- Each of these groups constitutes a critical band.
- When two frequencies are close enough to stimulate the same group of cells, and therefore fall into the same critical band, distinguishing them becomes difficult.
- Consider a simple experiment: we add two sinusoidal signals with close frequencies:
- As the difference between the frequencies increases, we first observe beating, then harsh dissonant sound, and finally two distinct, non-unpleasant, or consonant sounds.

Critical bands

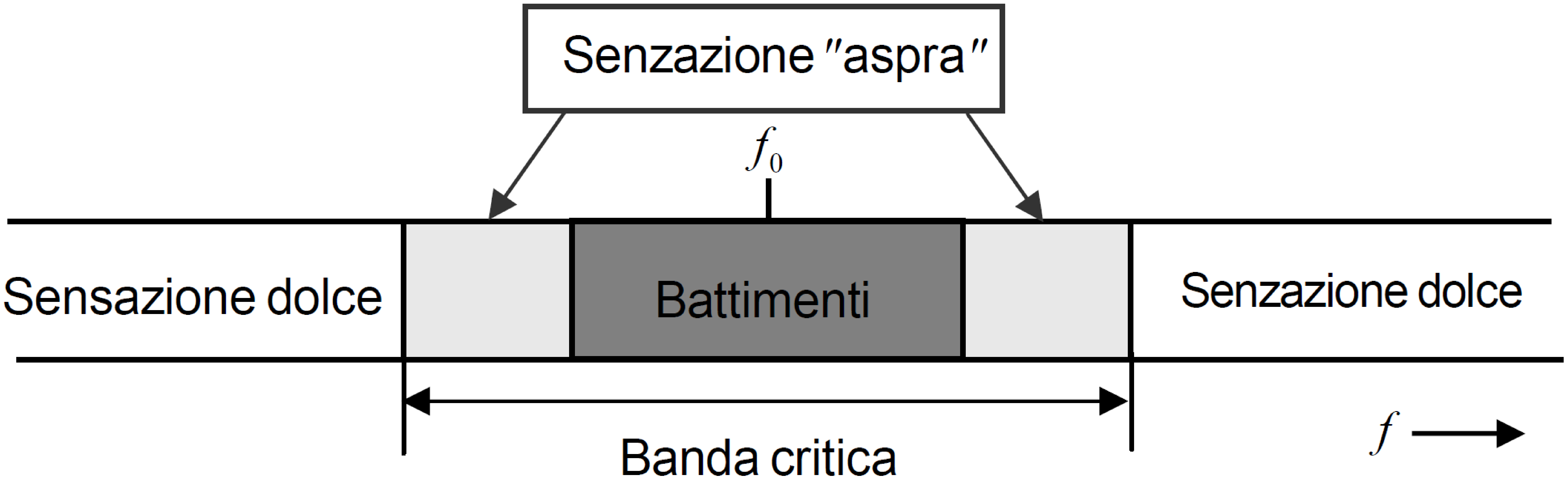


Figura 3.7 Fenomeno dei battimenti e banda critica.

Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Critical bands

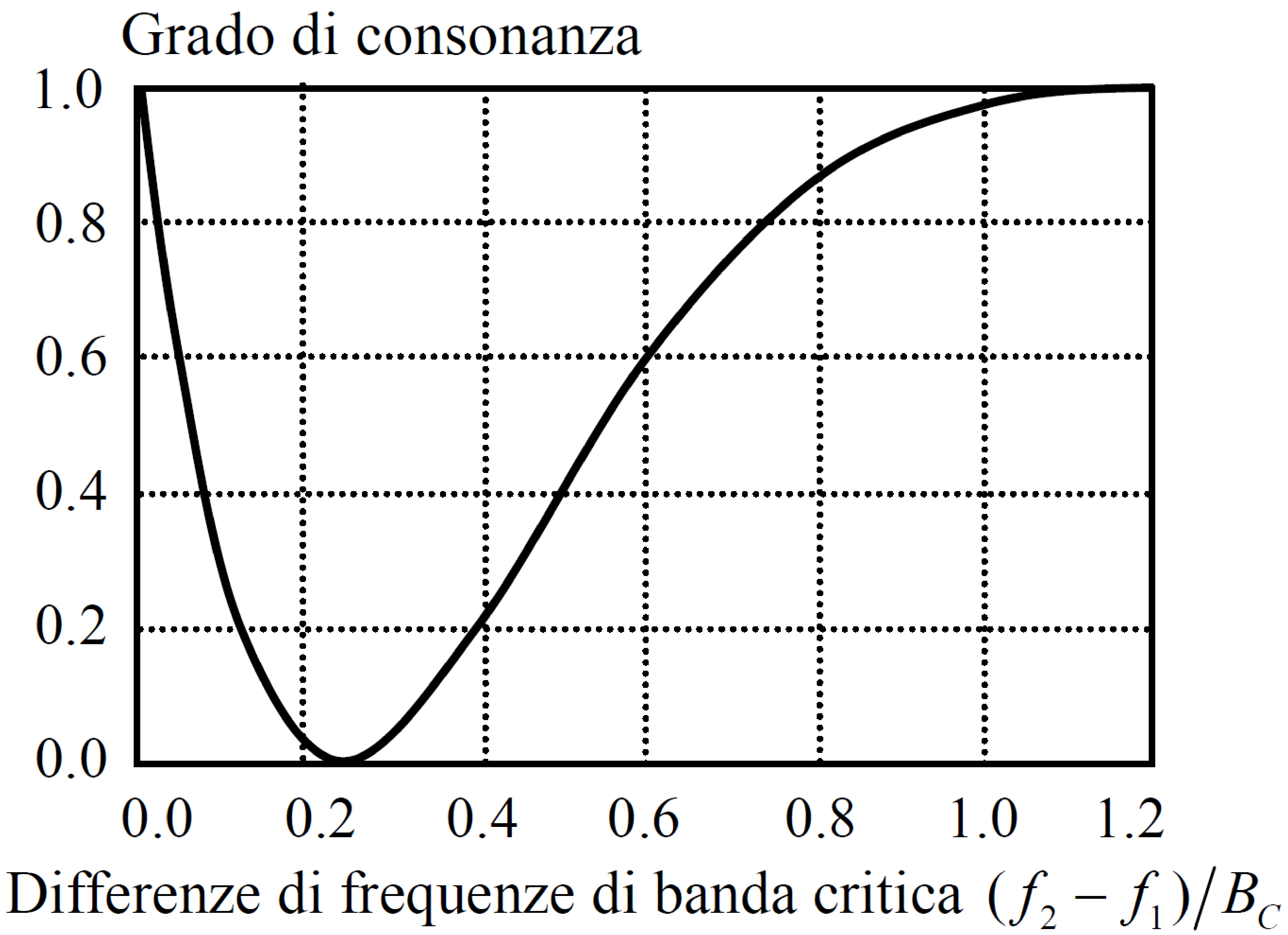
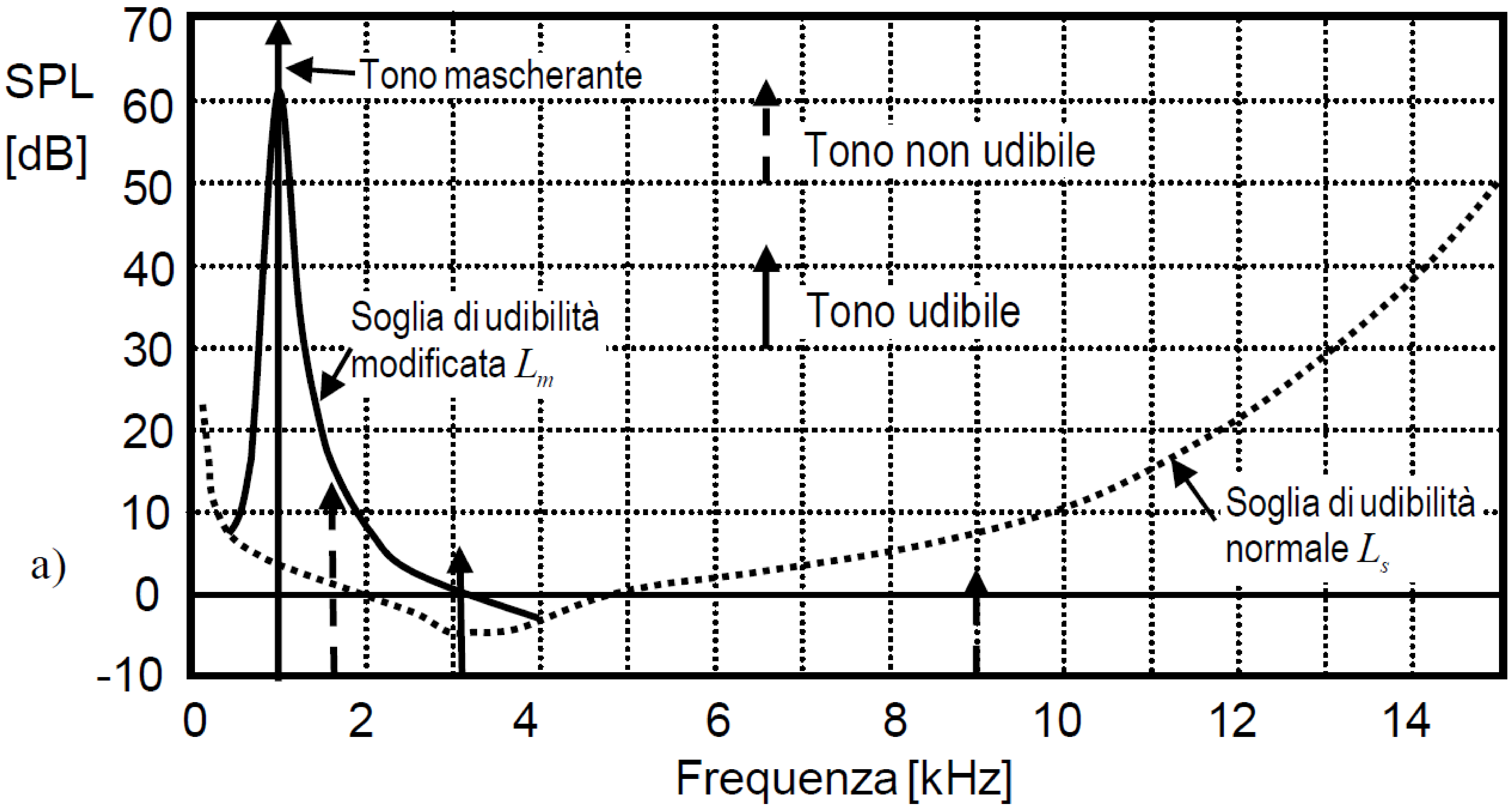


Figura 3.8 Curva di consonanza di due suoni puri (Misure di Plomp and Levelt (1965)).

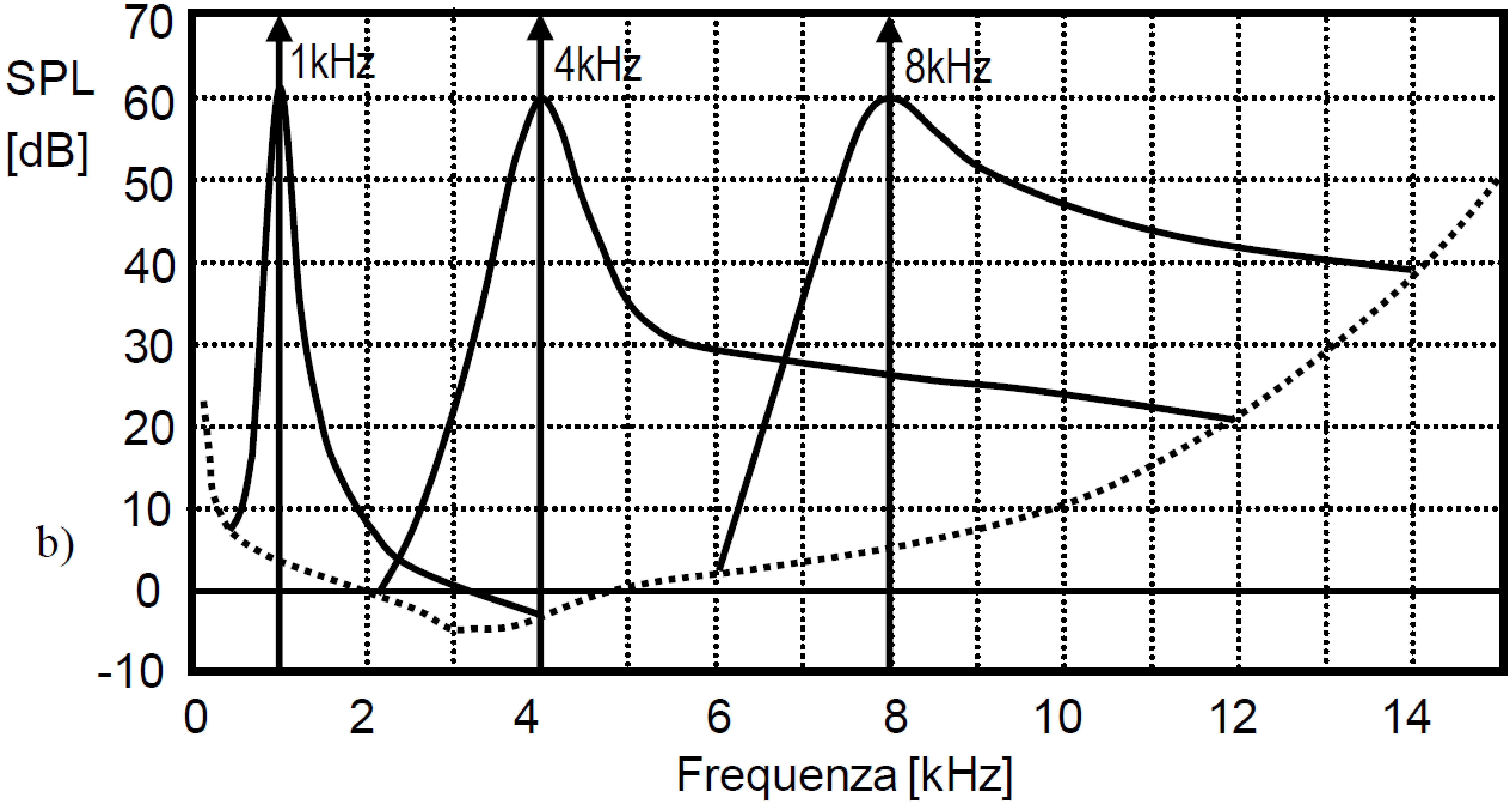
Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Simultaneous Masking



Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Simultaneous Masking



Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Non Simultaneous Masking

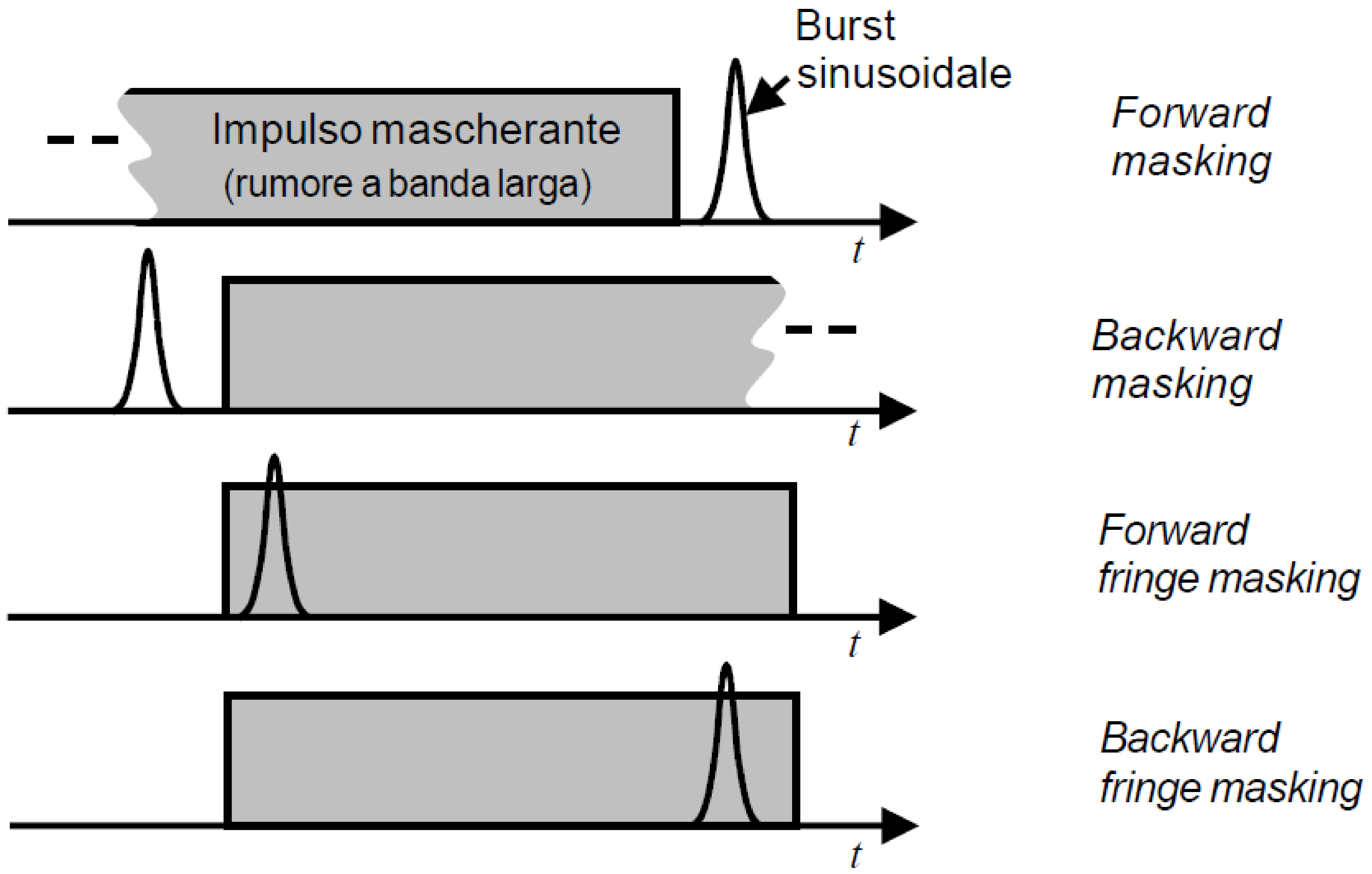


Figura 3.17 Successione temporale dei segnali da somministrare ai soggetti per lo studio dinamico del fenomeno di mascheramento.

Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Positional 2D model – duplex theory

- It is based on two quantities:
- Interaural Time Difference (ITD): the time difference with which the sound waveform reaches the two ears.
- Interaural Level Difference (ILD): the difference in intensity, in dB, perceived by the two ears.
- Both quantities are important for discriminating the origin of a sound in the azimuthal plane.

Positional 2D model – duplex theory

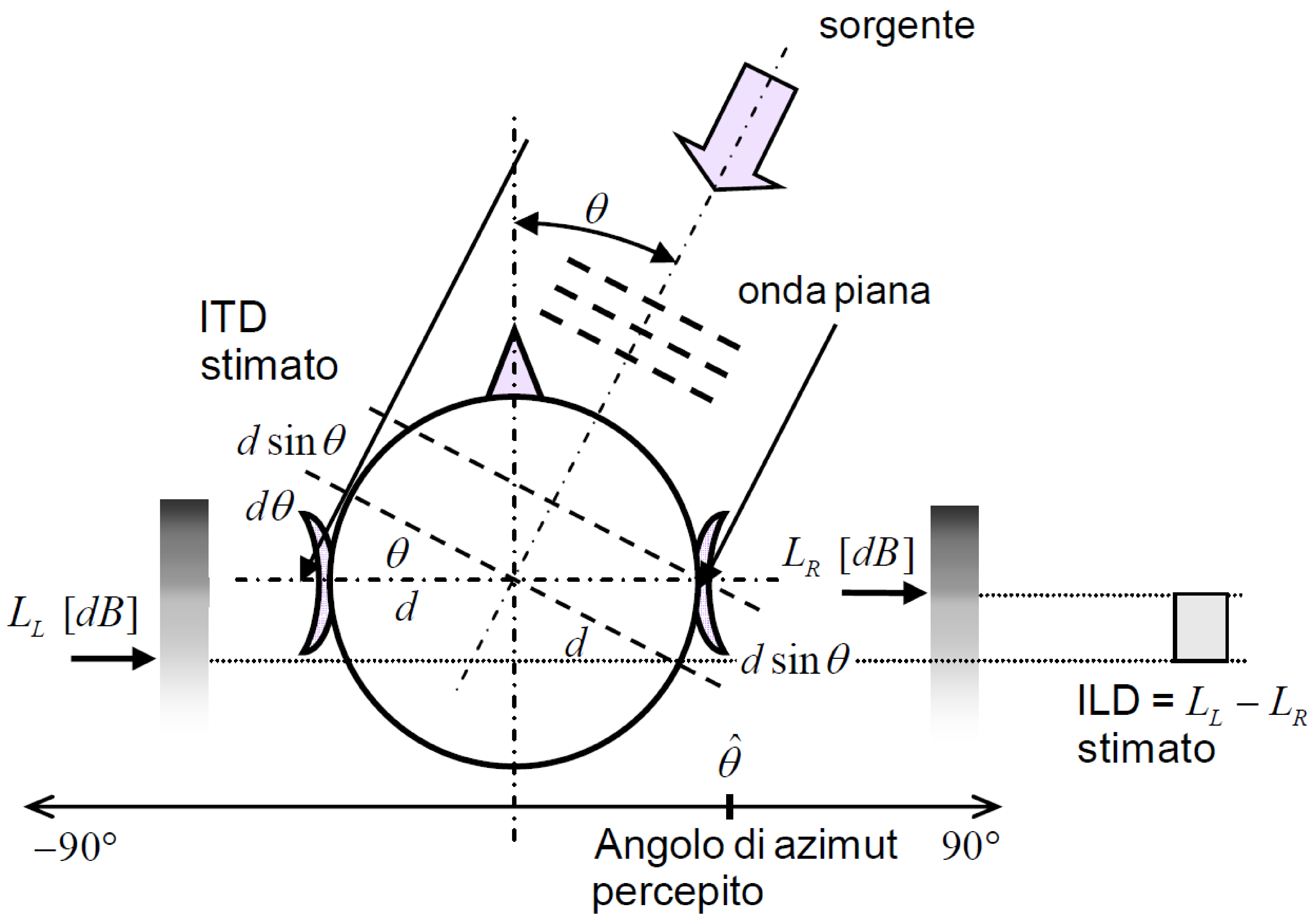


Figura 3.20 Geometria del modello Duplex. Differenza di intensità interaurale (ILD) e differenza di tempo di arrivo (ITD). Nel modello Duplex la direzione dall'angolo azimutale è valutata per mezzo di tali quantità.

Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Positional 2D model – duplex theory

- For frequencies < 1.5 kHz and a plane incident wave, the Woodworth formula applies:

$$\mathrm{ITD} = \frac{d}{c}\left(\sin\theta + \theta\right)\cos\phi; \qquad -90° \le \theta \le 90°\,;$$

- where c is the speed of sound, d is the radius of the sphere, θ is the azimuthal angle, and ϕ is the elevation angle.

Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Positional 2D model – duplex theory

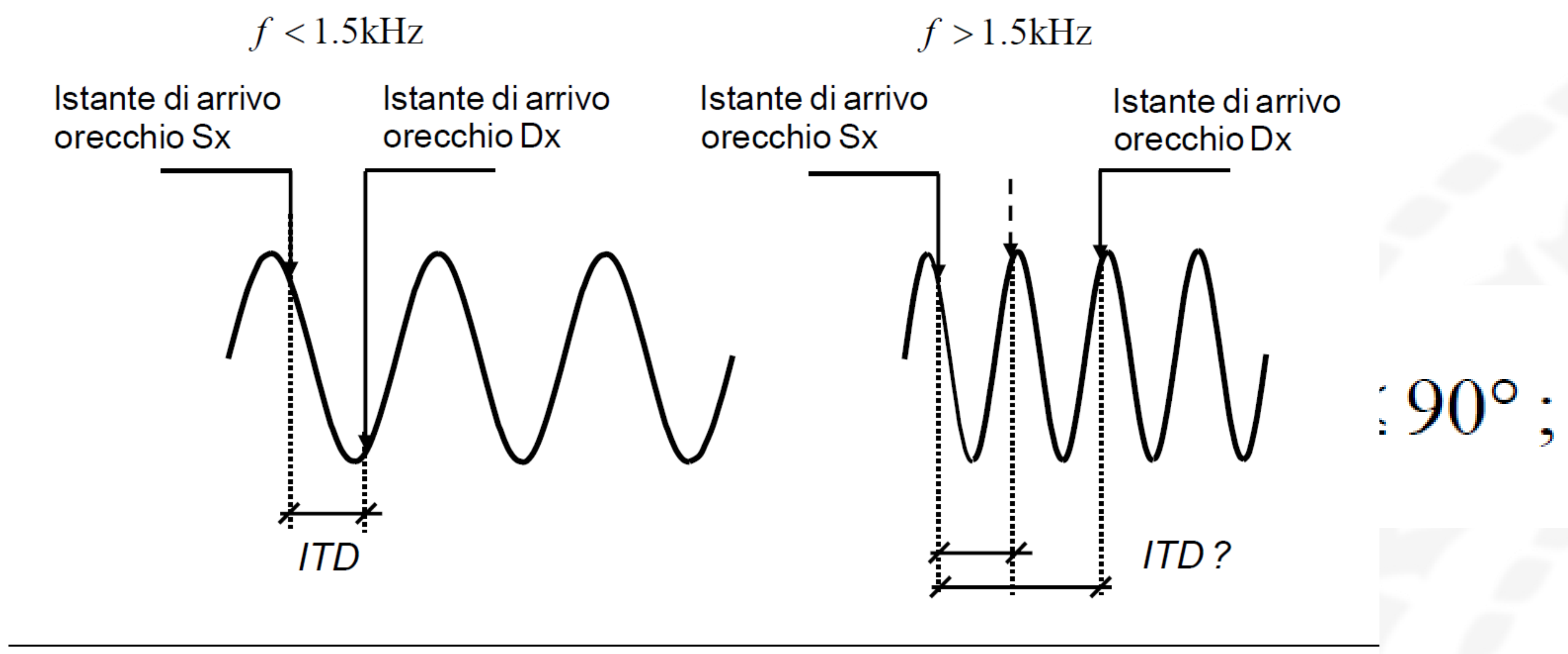

$f < 1.5\text{kHz}$
Istante di arrivo orecchio Sx
Istante di arrivo orecchio Dx
ITD
$f > 1.5\text{kHz}$
Istante di arrivo orecchio Sx
Istante di arrivo orecchio Dx
ITD ?
90° ;

Figura 3.22 Ambiguità della ITD per frequenze > 1500 Hz (*aliasing*).

- For frequencies > 1.5 kHz the ILD becomes fundamental.
- The relationship between the perception of direction and ILD is nonlinear and frequency-dependent. Approximately, the azimuth varies linearly with the logarithm of the ILD.

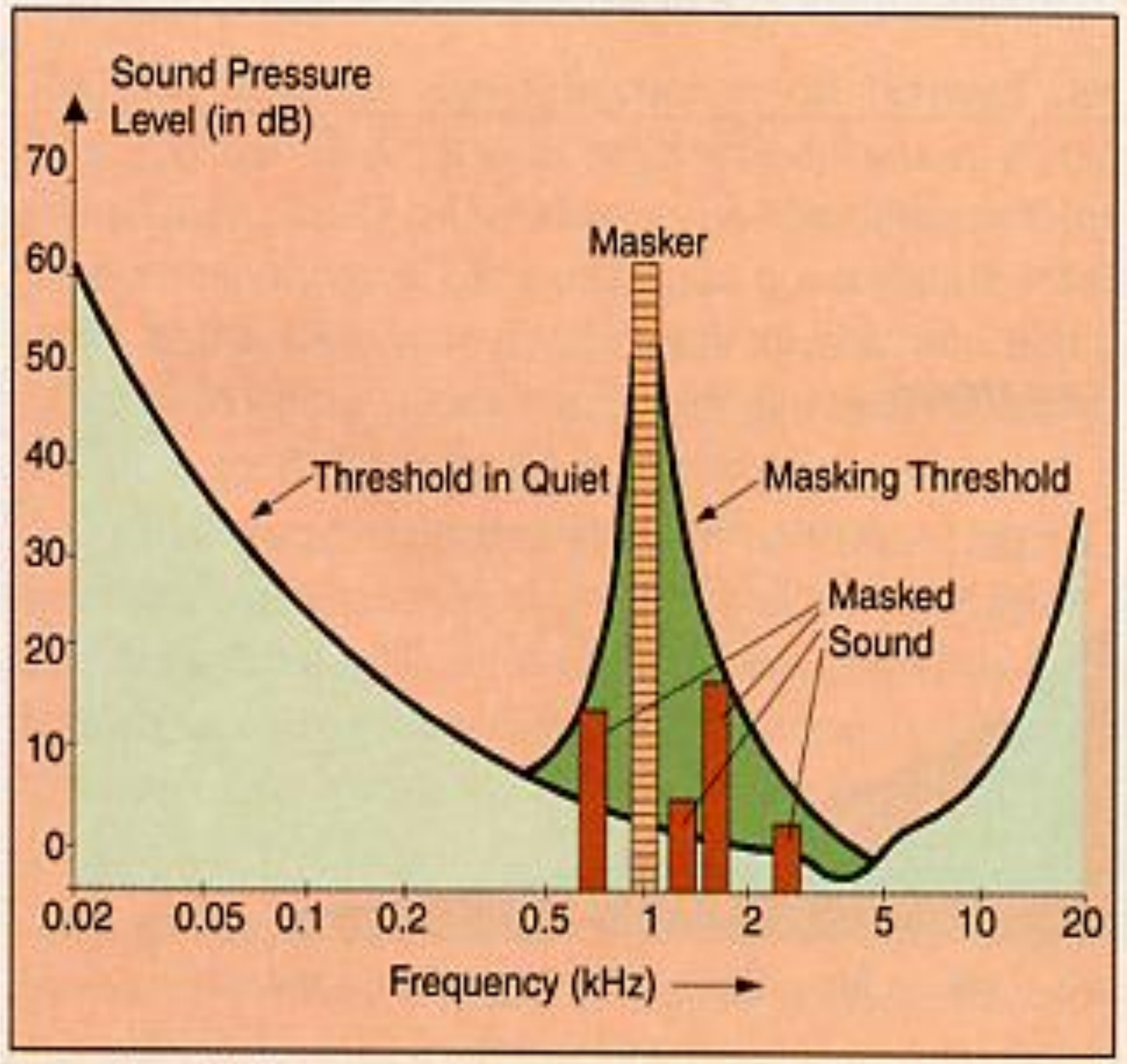
Da: A. Uncini «Audio digitale»
McGraw-Hill, 2006

Key tecnologies in Audio Coding

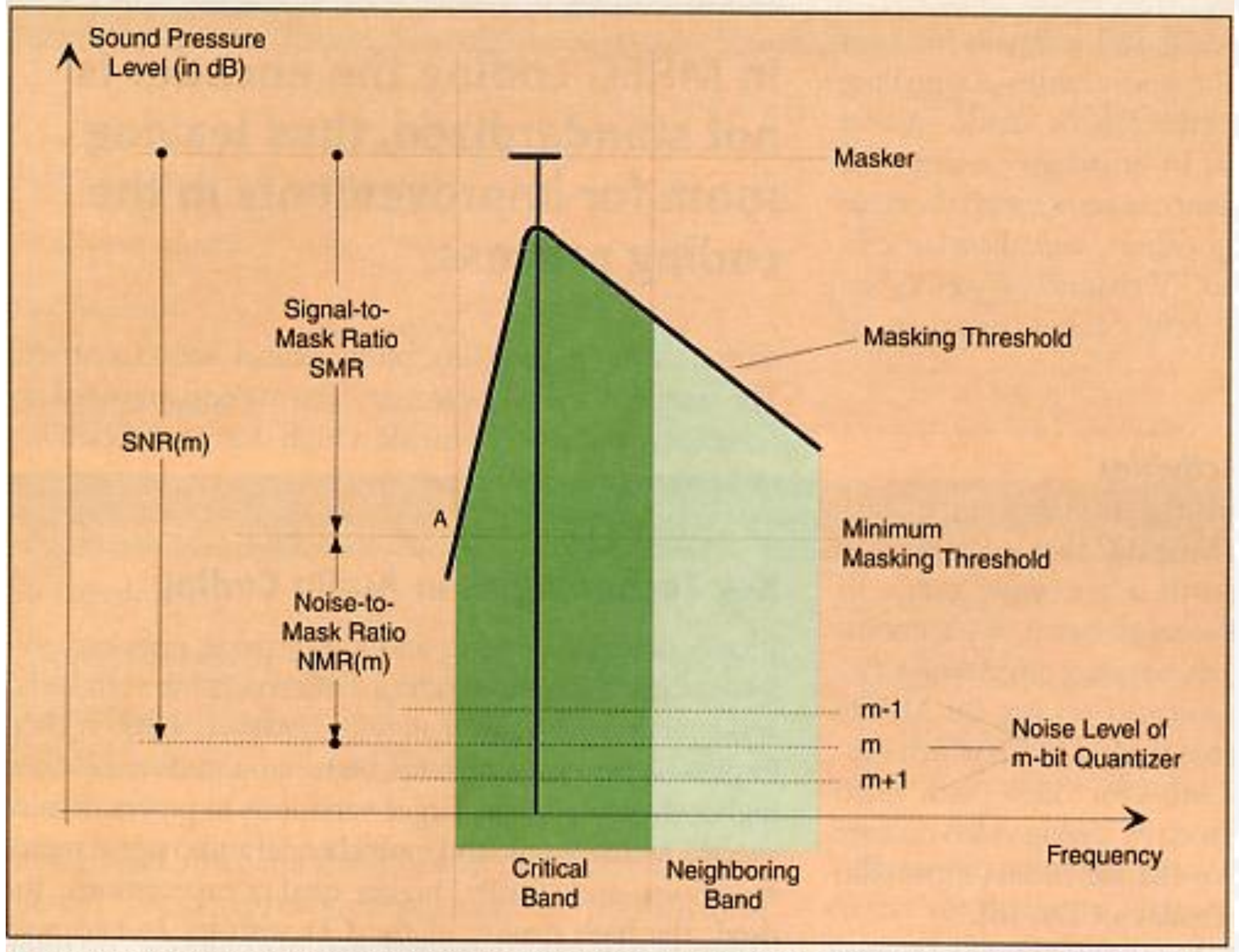
Four key technologies play an important role:

perceptual coding,
frequency-domain coding,
window switching, and
dynamic bit allocation.

Perceptual Coding

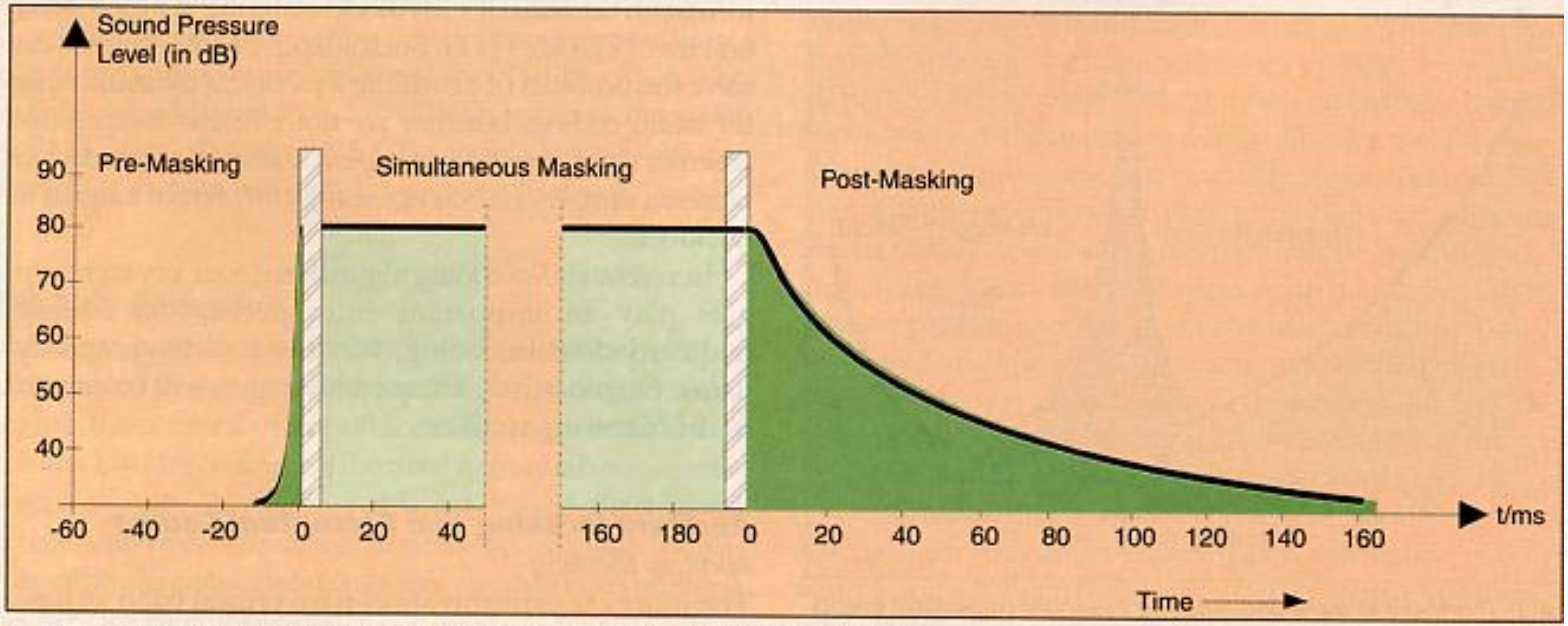


▲ *1. Threshold in quiet and masking threshold (acoustical events in the gray areas will not be audible).*



▲ *2. Masking threshold and signal-to-mask ratio (SMR) (acoustical events in the gray areas will not be audible).*

Perceptual Coding



▲ *3. Temporal masking (acoustic events in the gray areas will not be audible).*

Perceptual Coding

Assuming an m-bit quantization of an audio signal, within a critical band the quantization noise will not be audible as long as its signal-to-noise ratio (SNR) is higher than its signal-to-mask ratio (SMR).

Noise and signal contributions outside the particular critical band will also be masked, although to a lesser degree, if their sound pressure level (SPL) is below the masking threshold.

Defining SNR(m) as the SNR resulting from an m-bit quantization, the perceivable distortion in a given subband is measured by the noise-to-mask ratio (NMR):

$$\mathrm{NMR}(m)=\mathrm{SMR}-\mathrm{SNR}(m) \text{ (in dB)}$$

NMR(m) describes the difference in dB between the SMR and the SNR ratio to be expected from an m-bit quantization. The NMR value is also the difference (in dB) between the level of quantization noise and the level where distortion may just become audible in a given subband.

Within a critical band, coding noise will not be audible as long as NMR(m) is negative.

Perceptual Coding

We have just described masking by only one masker. If the source signal consists of many simultaneous maskers, each has its own masking threshold, and a global masking threshold can be computed that describes the threshold of just-noticeable distortions as a function of frequency.

In addition to simultaneous masking, the time-domain phenomenon of temporal masking plays an important role in human auditory perception. It may occur when two sounds appear within a small interval of time. Depending on the individual SPLs, the stronger sound may mask the weaker one, even if the maskee precedes the masker.

Temporal masking can help to mask pre-echoes caused by the spreading of a sudden large quantization error over the actual coding block.

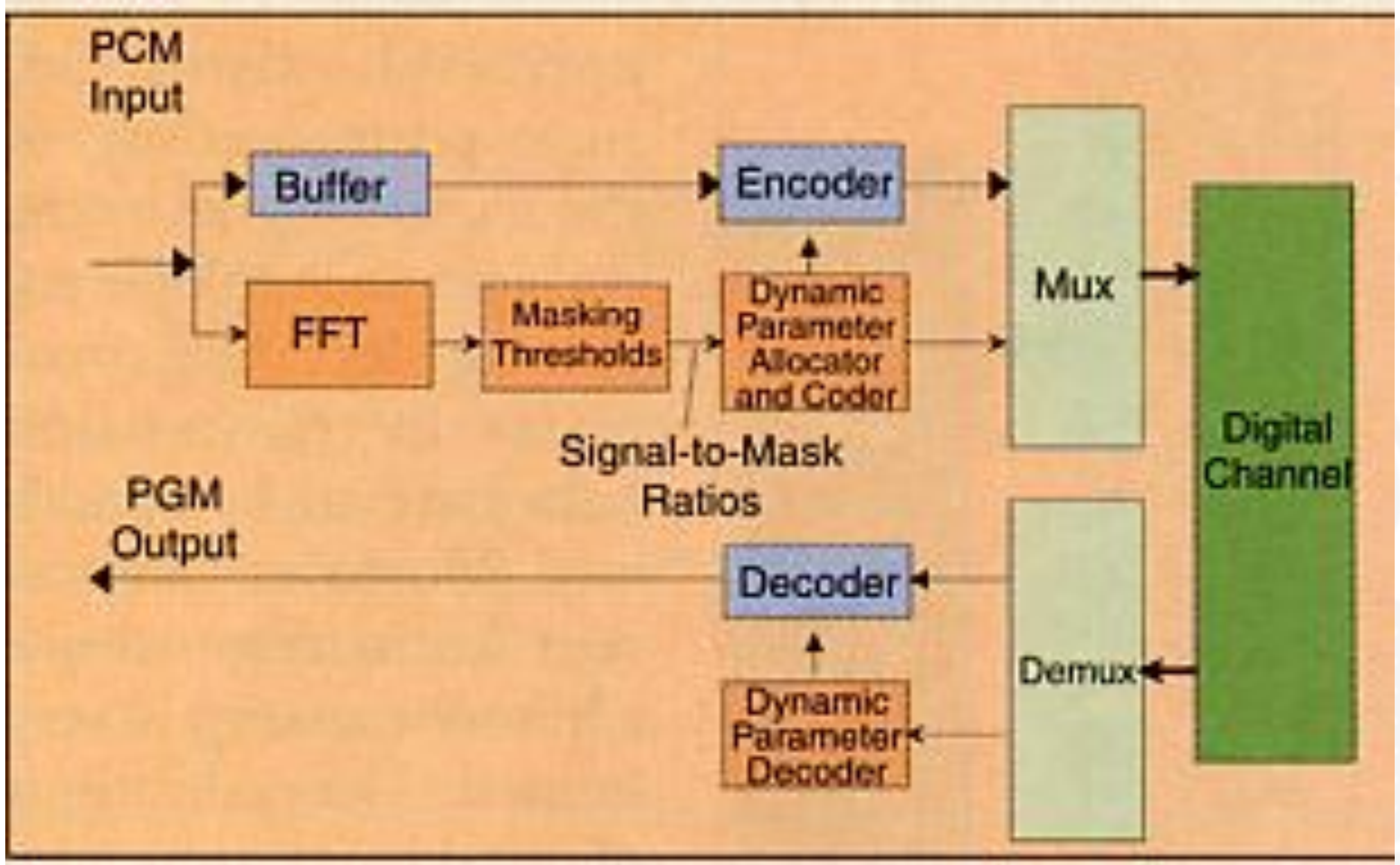
Perceptual Coding

An efficient source coding algorithm will

(i) remove redundant components of the source signal by exploiting correlations between its samples and

(ii) remove components that are perceptually irrelevant to the ear.

Perceptual Coding



▲ *4. Block diagram of perception-based coders (acoustical events in the gray areas will not be audible).*

Frequency Domain Coding

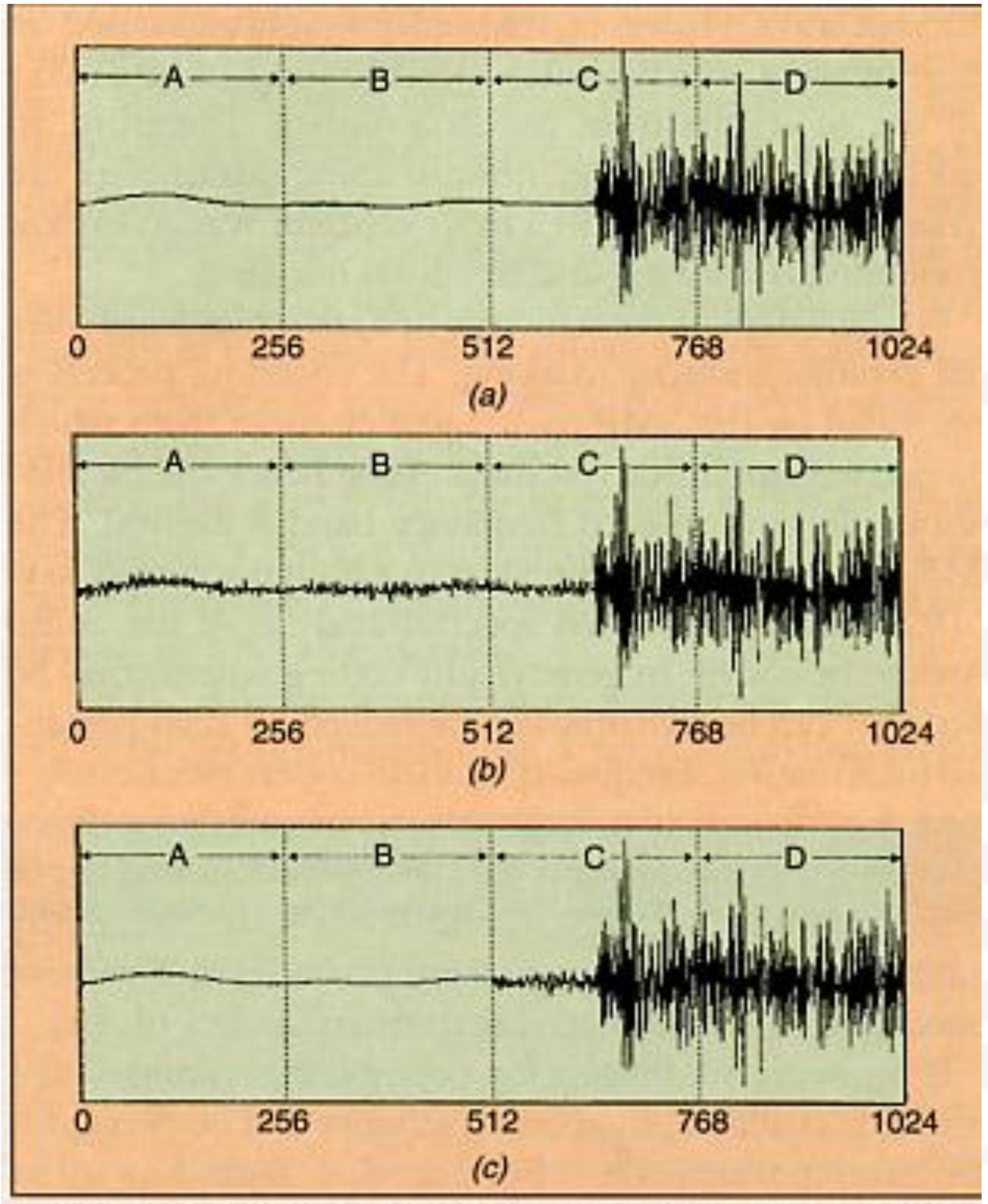
In all frequency-domain coders, redundancy (the non flat short-term spectral characteristics of the source signal) and irrelevancy (signals below the psycho acoustical thresholds) are exploited to reduce the transmitted data rate with respect to PCM.

This is achieved by splitting the source spectrum into frequency bands to generate nearly uncorrelated spectral components and by quantizing these components separately.

Two coding categories exist, **transform coding** (TC) and **subband coding** (SBC).

Hybrid coding provides a combination of the two.

Window switching

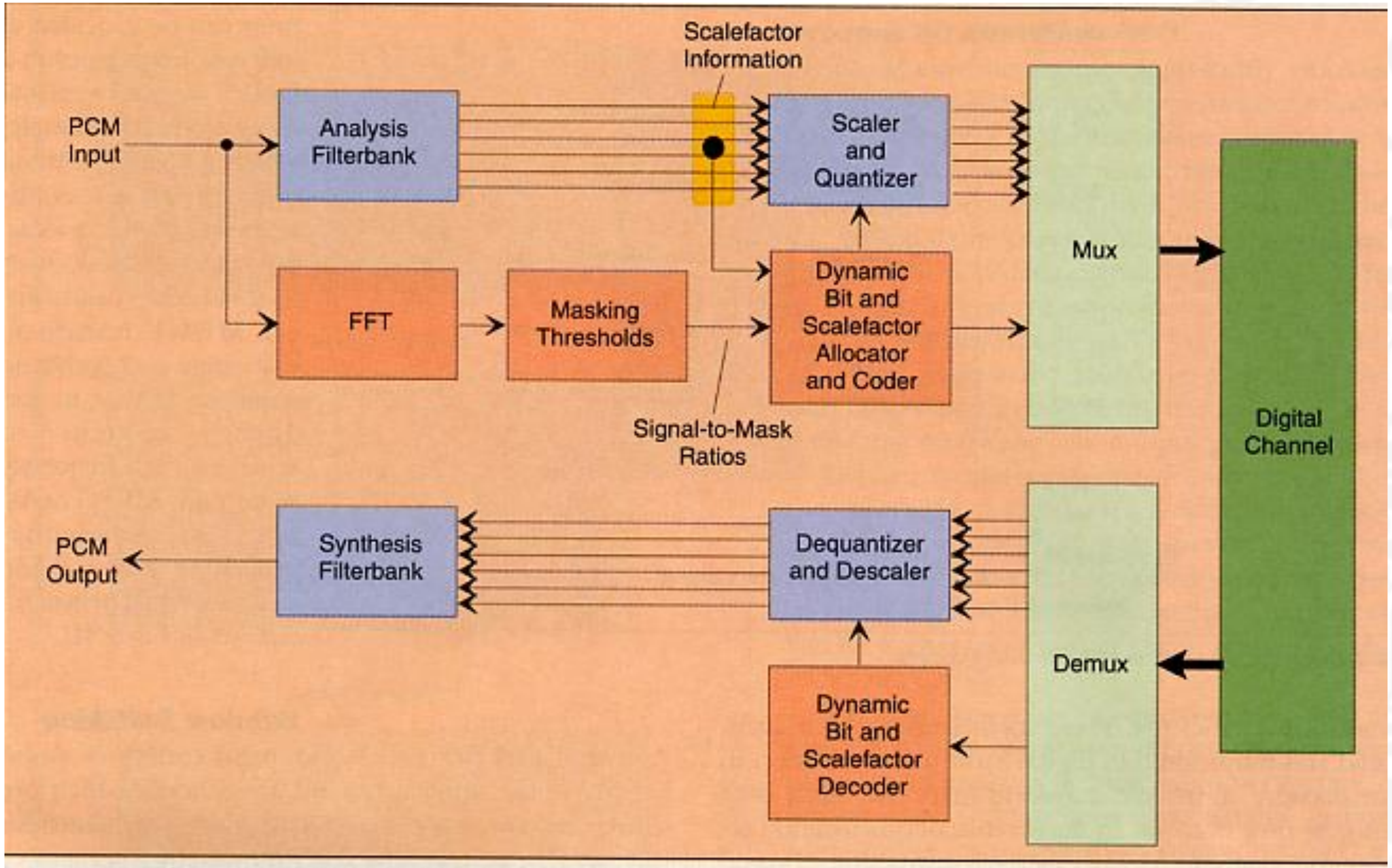


▲ *5. Window switching: (a) source signal, (b) reconstructed signal with block size N = 1024, (c) reconstructed signal with block size N = 256. (Source: Iwadare et al. [25].)*

Dynamic Bit Allocation

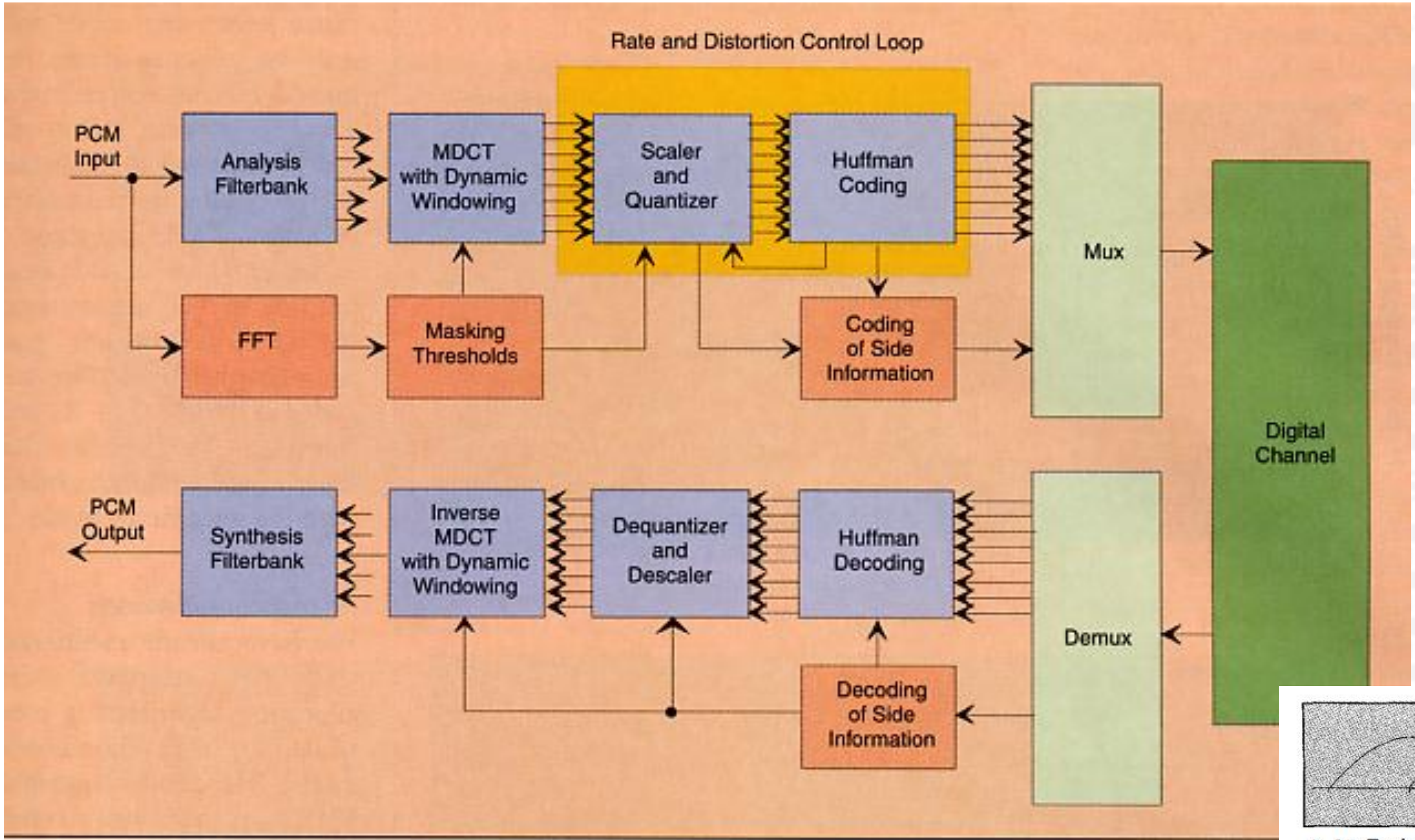
Frequency-domain coding significantly gains in performance if the number of bits assigned to each of the quantizers of the transform coefficients is adapted to the short-term spectrum of the audio coding block on a block-by-block basis.

MPEG-1 Layer I and II

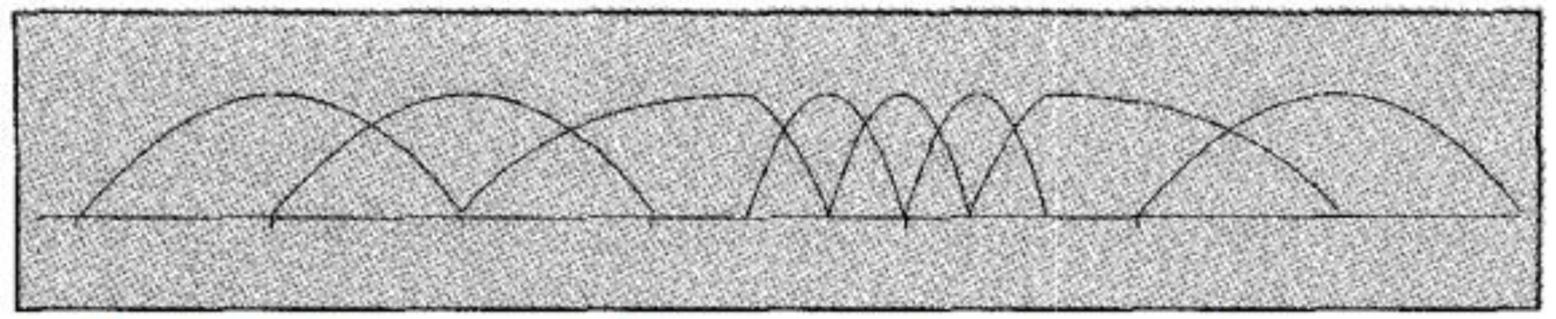


▲ *8. Structure of MPEG-1 audio encoder and decoder (Layers I and II).*

MPEG-1 Layer I and II



▲ *14. Structure of MPEG-1 audio encoder and decoder (Layer III).*



▲ *15. Typical sequence of windows in adaptive window switching.*

Study:

- Noll, P. (1997). MPEG digital audio coding. IEEE signal processing magazine, 14(5), 59-81.