Psycho-Acoustics

Richard Heusdens

April 28, 2020

Signal and Information Processing Lab, Dept. of Mediamatics

Delft University of Technology

Outline

- Introduction
- Anatomy and physiology of the auditory system
- Physical versus subjective scales
- Spatial perception
- Masking (spectral and temporal)
- Perceptual audio coding
  - ISO MPEG perceptual model

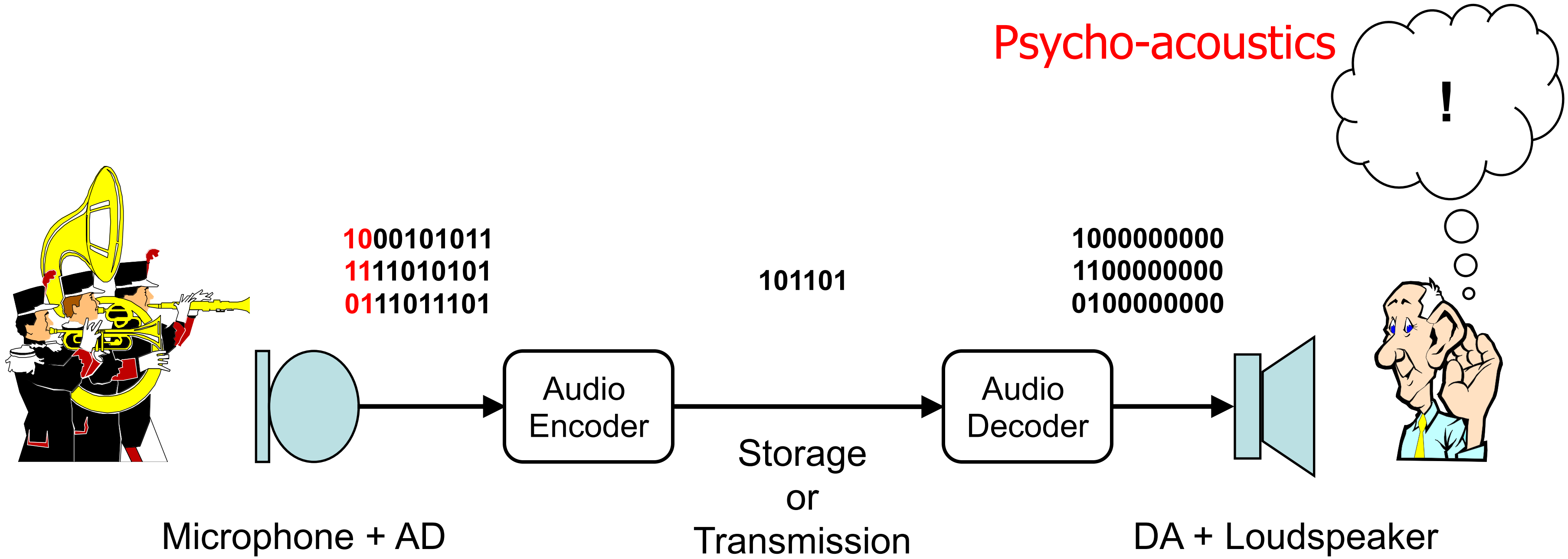
TU Delft

Psychoacoustics

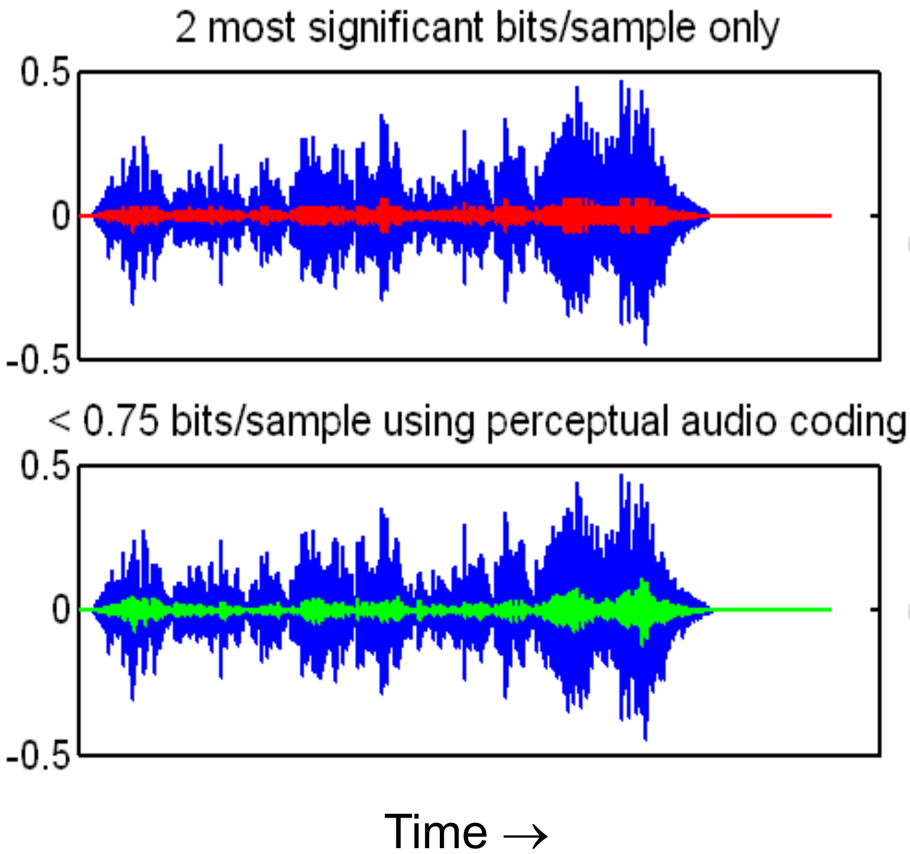
=

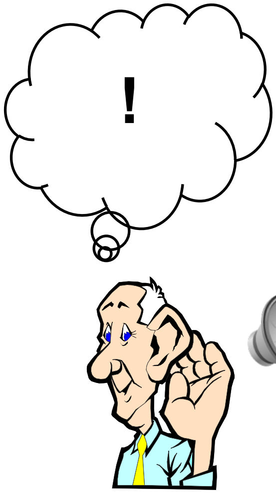
The scientific study of the perception of sound.

Digital Signal Processing



What you see is not what you hear





TUDelft

Anatomy and Physiology of the Human Auditory System

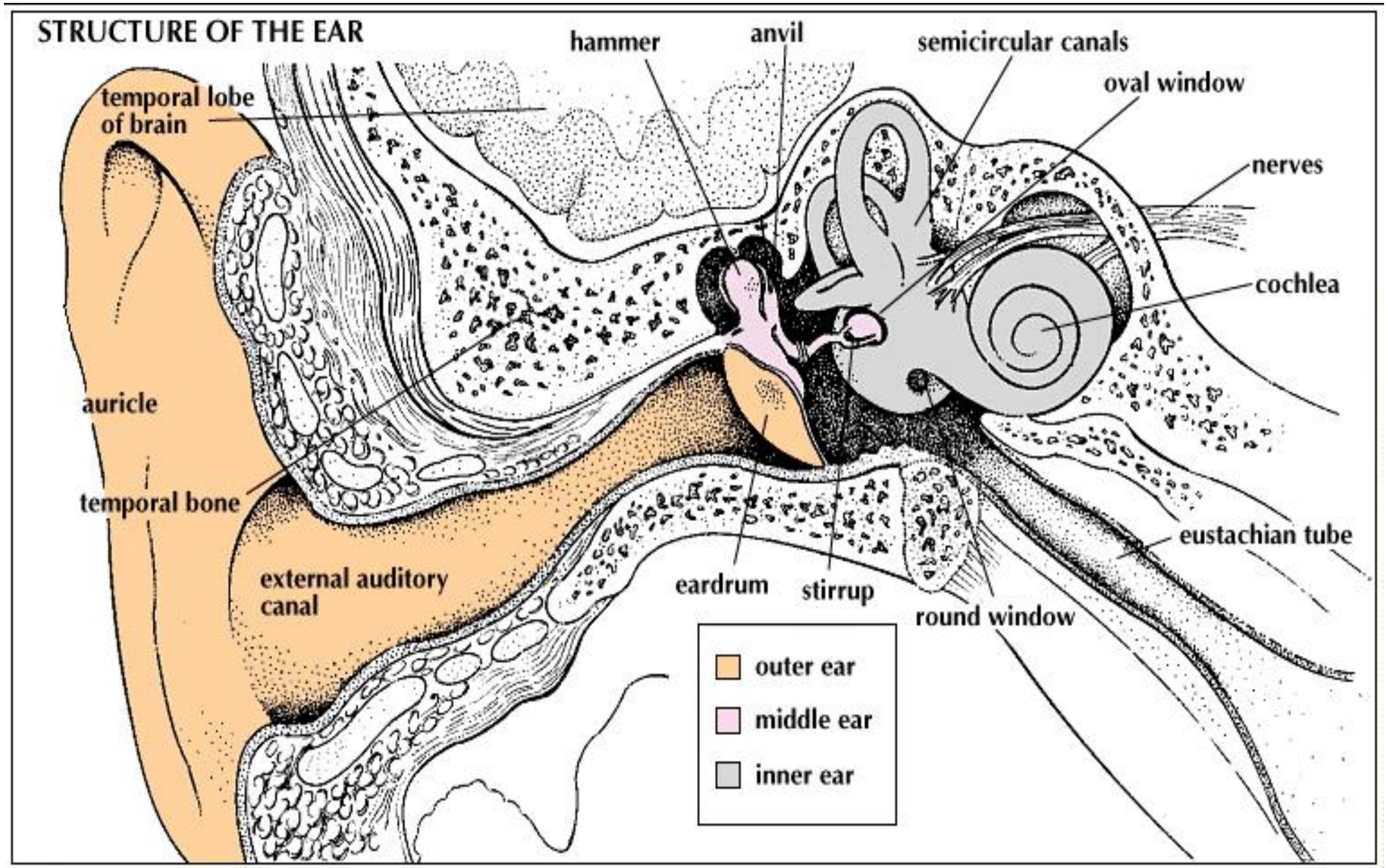
Literature:

James O. Pickles, “An Introduction to the Physiology of Hearing”, Academic Press, London, 1982

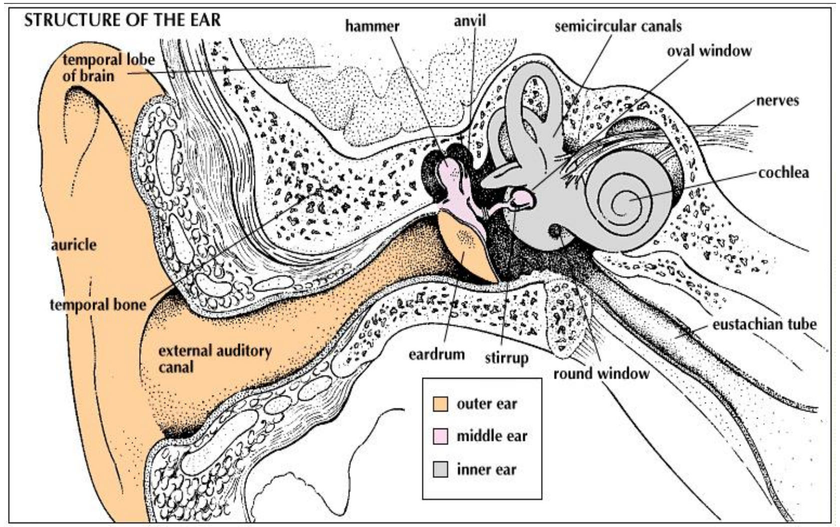
April 28, 2020

TU Delft
Delft University of Technology

The Peripheral Hearing System



Outer Ear

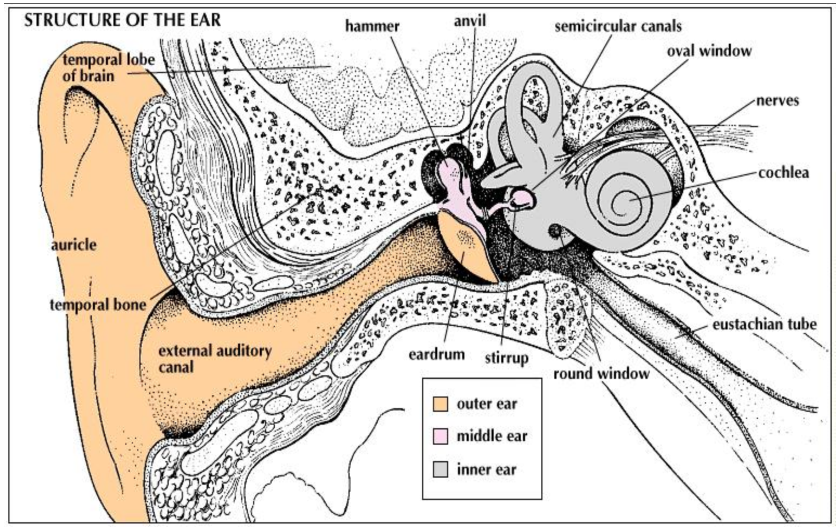


Acoustic energy is conducted to the eardrum

- Conversion of acoustic energy into mechanical energy
- Resonance in the range of 2 – 5 kHz (± 10 dB)
- Acoustic filtering is in part dependent on source direction

TU Delft

Middle Ear

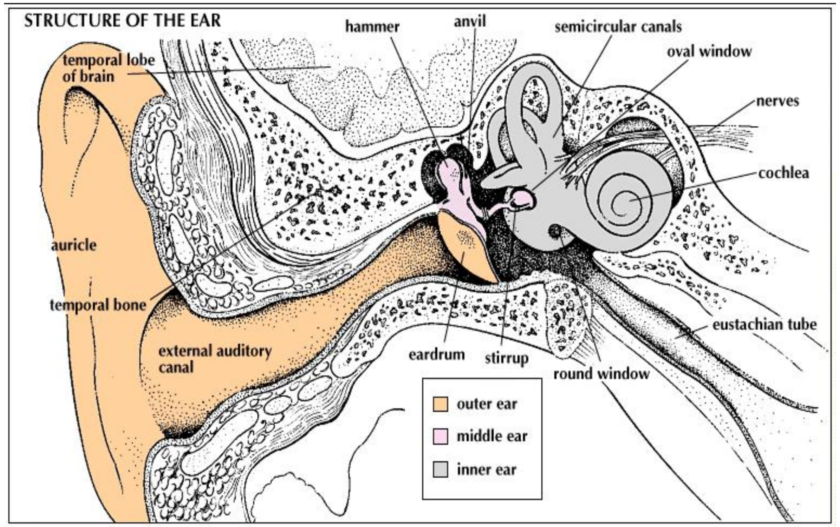


Impedance adjustment for effective energy transmission:

- Eardrum: Large volume displacement, little pressure
- Oval window: Little volume displacement, large pressure
- Band-pass filter (200 Hz - 8 kHz)

TU Delft

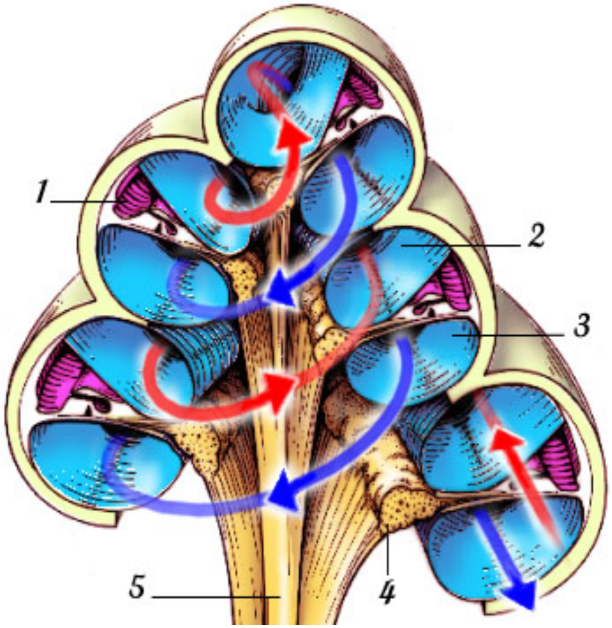
Inner Ear



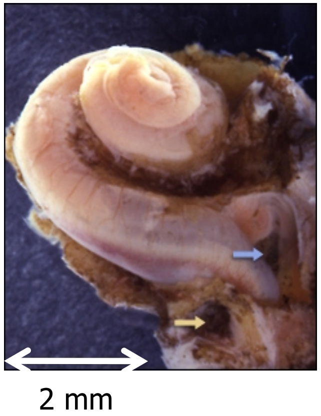
Cochlea:

- Mechanical energy (oval window) is converted into a neural signal (auditory nerve)
- Performs a time-frequency analysis

Cochlea



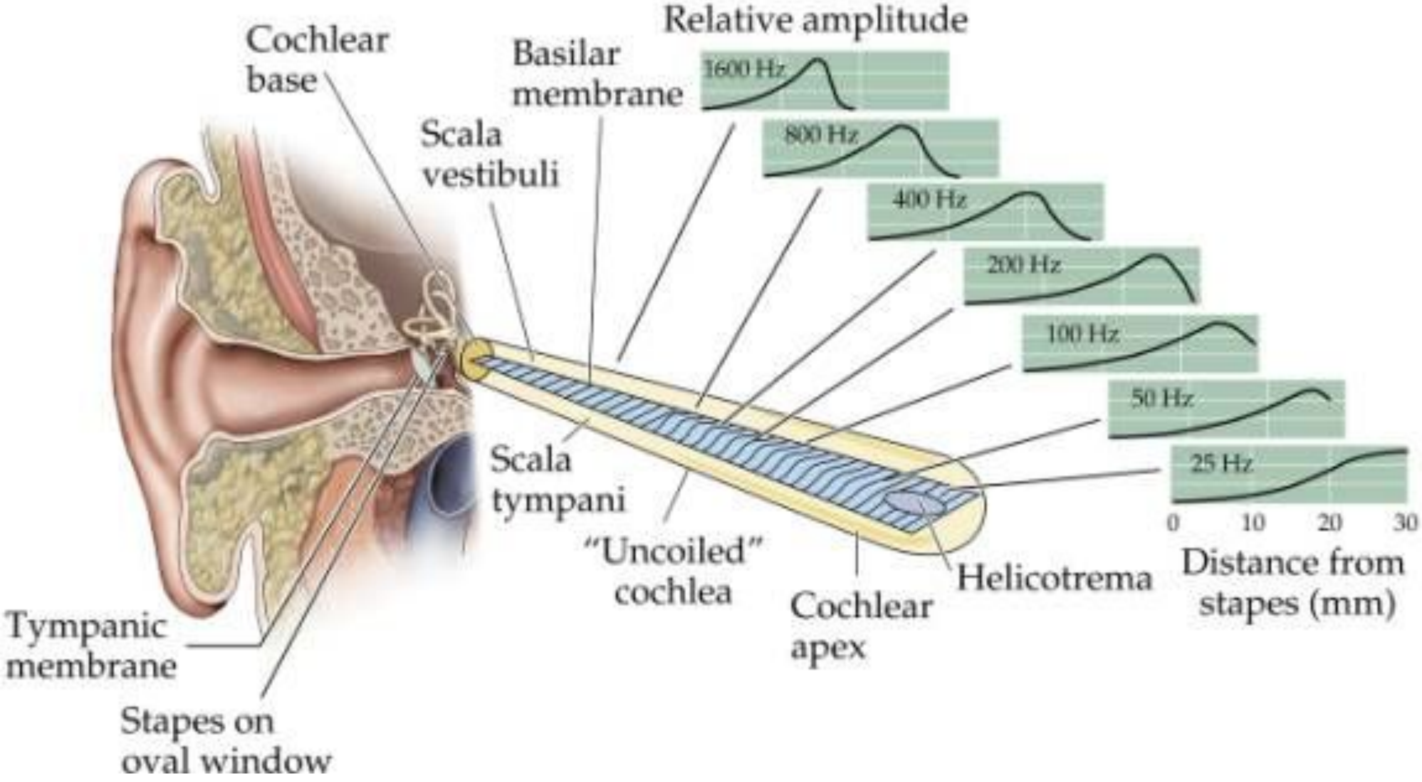
1. cochlear duct
2. scala vestibuli
3. scala tympani
4. spiral ganglion
5. auditory nerve fibres



- The red arrow is from the oval window
- The blue arrow points to the round window
- The cochlea is about 2 mm in diameter

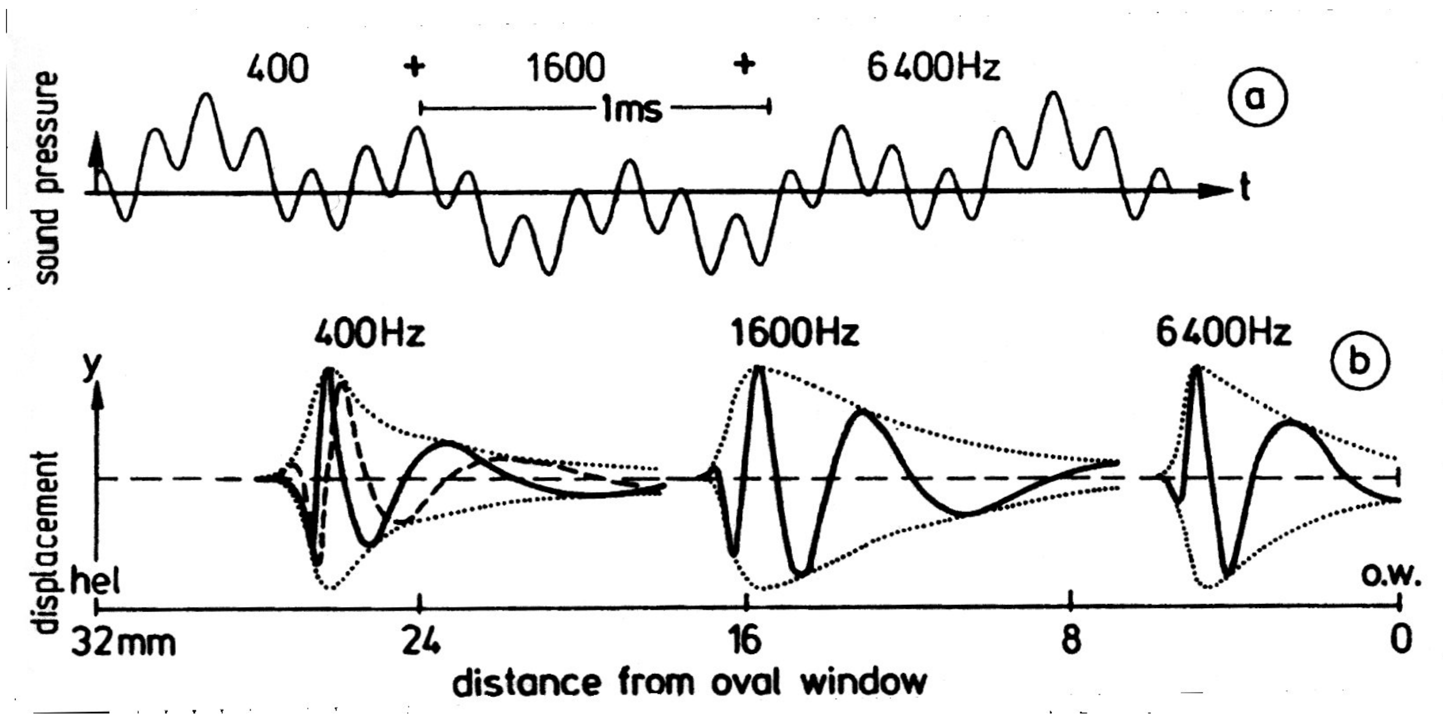
TU Delft

Basilar Membrane

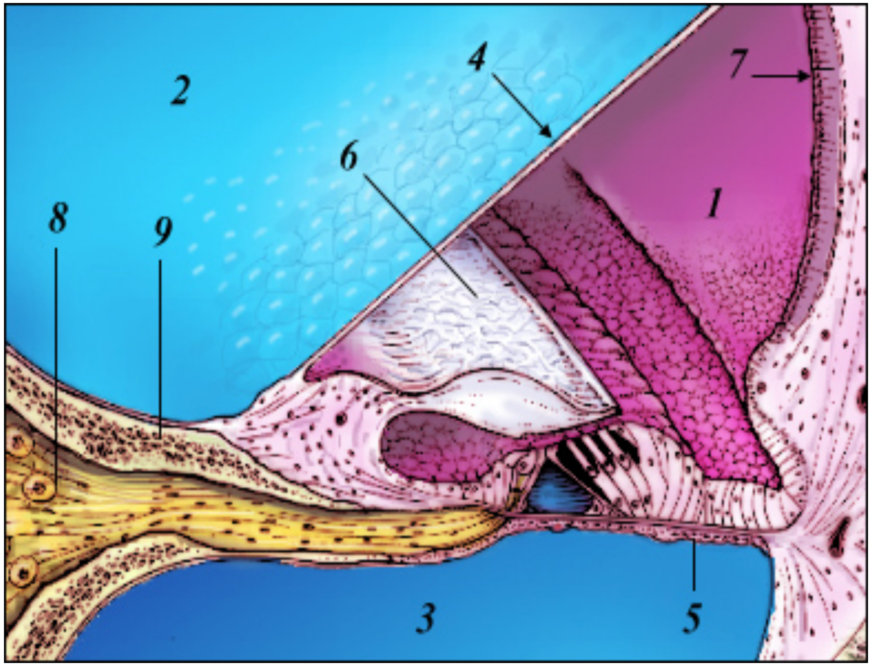


Basilar Membrane

Frequency-to-place transformation:



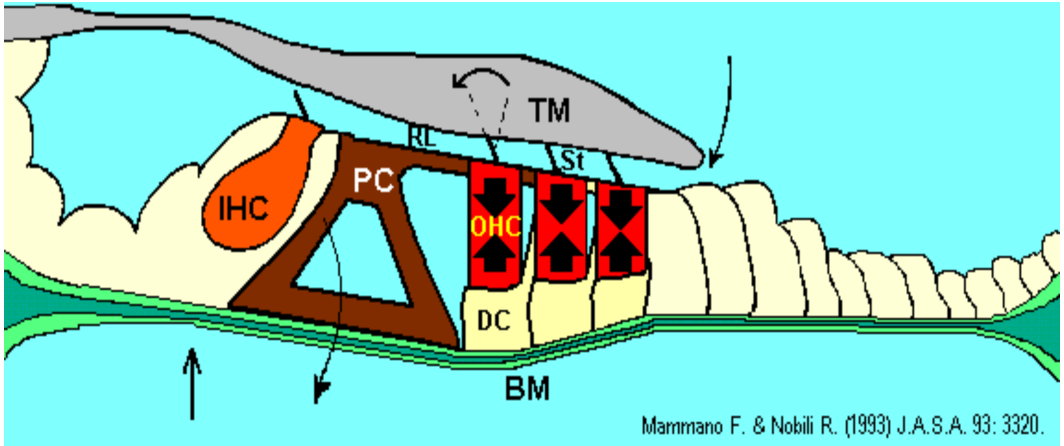
Cochlea



1. Cochlear duct
2. Scala vestibuli
3. Scala tympani
4. Reissner's membrane
5. Basilar membrane
6. Tectorial membrane
7. Stria vascularis
8. Nerve fibres
9. Bony spiral lamina

The organ of Corti is on top of the basilar membrane under the tectorial membrane

Organ of Corti

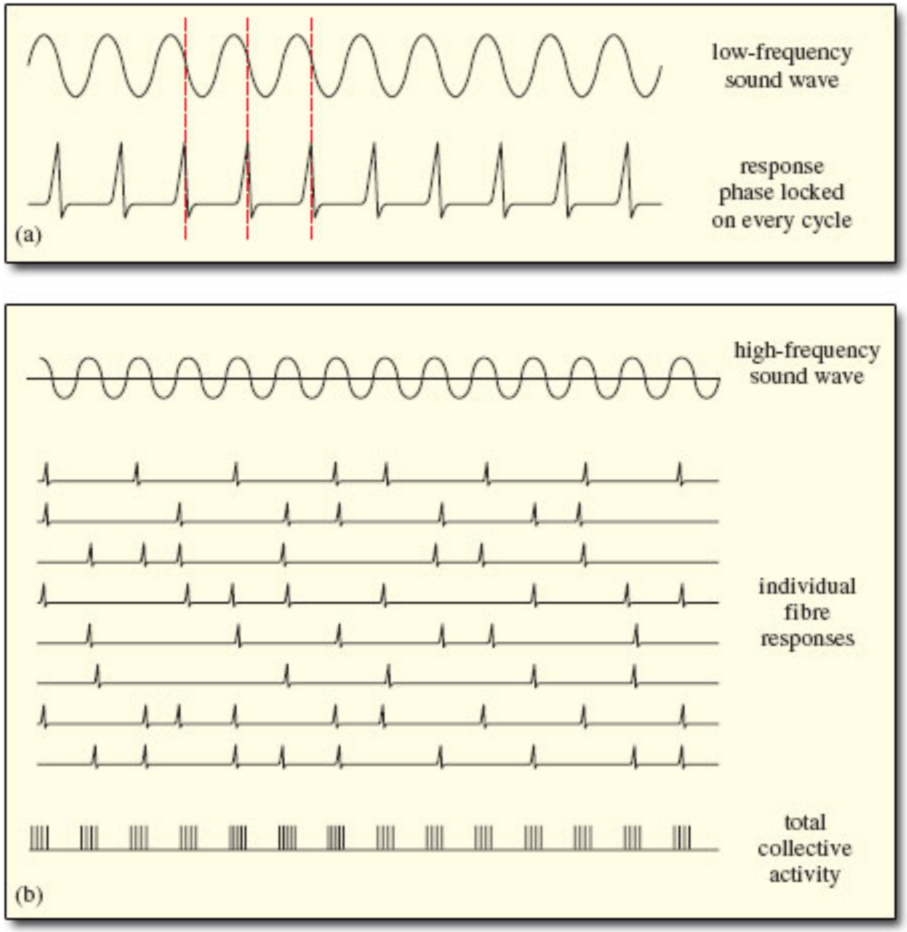


- Tectorial membrane (TM)
- Basilar membrane (BM)
- Inner haicells (IHC)
- Outer haircells (OHC)
- Deiters' cells (DC)
- Reticular lamina (RL)
- Pillar cells (PC)
- Stereocilia (St)

TU Delft

Auditory nerve

- Auditory nerve fiber responses are spiky
- Auditory nerve responses can follow the phase at low frequencies
- Auditory nerve responses need to recover at high frequencies
- Beyond about 2 kHz phase locking response is lost



TU Delft

Auditory Transduction

TUDelft

Dependence of Subjective Parameters on Physical Parameters

Literature:

Brian C.J. Moore, “An Introduction to the Psychology of Hearing”, Fourth edition, Academic Press, London, 1997

Physical vs. Subjective Scales

Physical		**Subjective**
Fundamental frequency (Hz)	$\leftrightarrow$	Pitch
Sound pressure (Pa, dB SPL)	$\leftrightarrow$	Loudness

Fundamental Frequency and Pitch

- Linear frequency increase over time
- Exponential frequency increase over time
- Mapping of frequency to pitch is approximately a logarithmic transformation
- Empirical finding from several experiments measuring frequency just noticable differences (JNDs):

$$\log \Delta f = 0.0264\sqrt{f} - 0.52$$

	f	f+1 JND	f+2 JND	f+3 JND
$f = 1$ kHz $\rightarrow \Delta f = 2$ Hz				
$f = 6$ kHz $\rightarrow \Delta f = 33$ Hz				

Sound Pressure and Loudness

- Linear sound pressure increase over time
- Expansive sound pressure increase over time

→ Mapping of sound pressure to loudness resembles a compressive transformation

Stevens Law (sones): $S = kI^{0.3}$

→ Thus each 10 dB increase in intensity leads to a doubling in loudness (= doubling in sones)

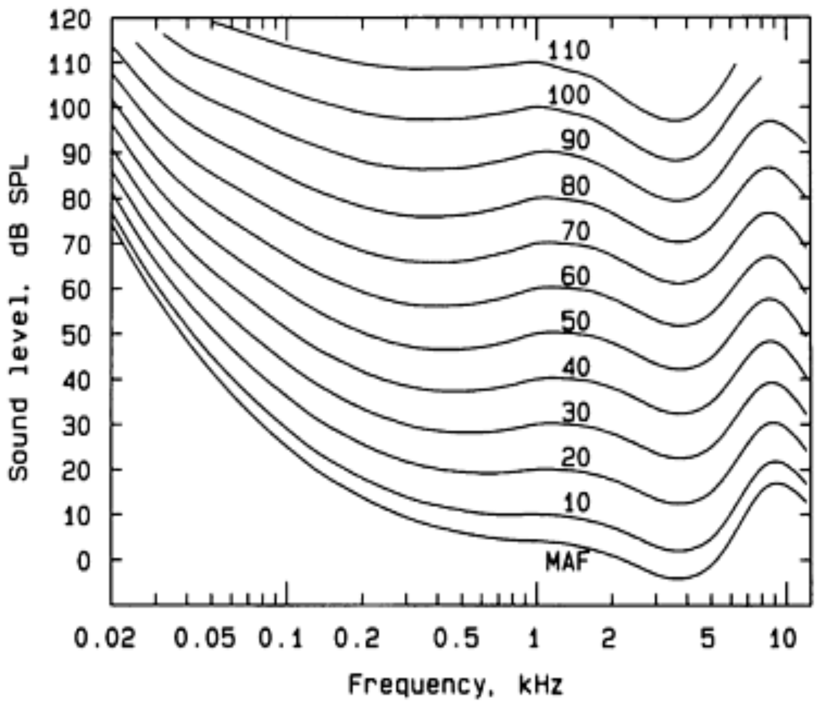
TU Delft

Source of sound	sound pressure	sound pressure level	loudness
	pascal	dB re 20 μPa	sone
threshold of pain	100	134	**~ 676**
hearing damage during short-term effect	20	approx. 120	**~ 256**
jet, 100 m away	6 ... 200	110 ... 140	**~ 128 ... 1024**
jack hammer, 1 m away / nightclub	2	approx. 100	**~ 64**
hearing damage during long-term effect	6×10^{-1}	approx. 90	**~ 32**
major road, 10 m away	2×10^{-1} ... 6×10^{-1}	80 ... 90	**~ 16 ... 32**
passenger car, 10 m away	2×10^{-2} ... 2×10^{-1}	60 ... 80	**~ 4 ... 16**
TV set at home level, 1 m away	2×10^{-2}	ca. 60	**~ 4**
normal talking, 1 m away	2×10^{-3} ... 2×10^{-2}	40 ... 60	**~ 1 ... 4**
very calm room	2×10^{-4} ... 6×10^{-4}	20 ... 30	**~ 0.15 ... 0.4**
leaves' noise, calm breathing	6×10^{-5}	10	**~ 0.02**
auditory threshold at 1 kHz	2×10^{-5}	0	**0**

sone	1	2	4	8	16	32	64	128	256	512	1024
phon	40	50	60	70	80	90	100	110	120	130	140

Equal Loudness Contours

Alternative scale for loudness is the phon:



- The loudness in phon of a specific tone is defined as the level in dB SPL of a 1 kHz reference tone that sounds equaly loud as the specific tone.
- MAF = minimum audible field (absolute threshold of hearing)

TU Delft

Sound Pressure and Loudness

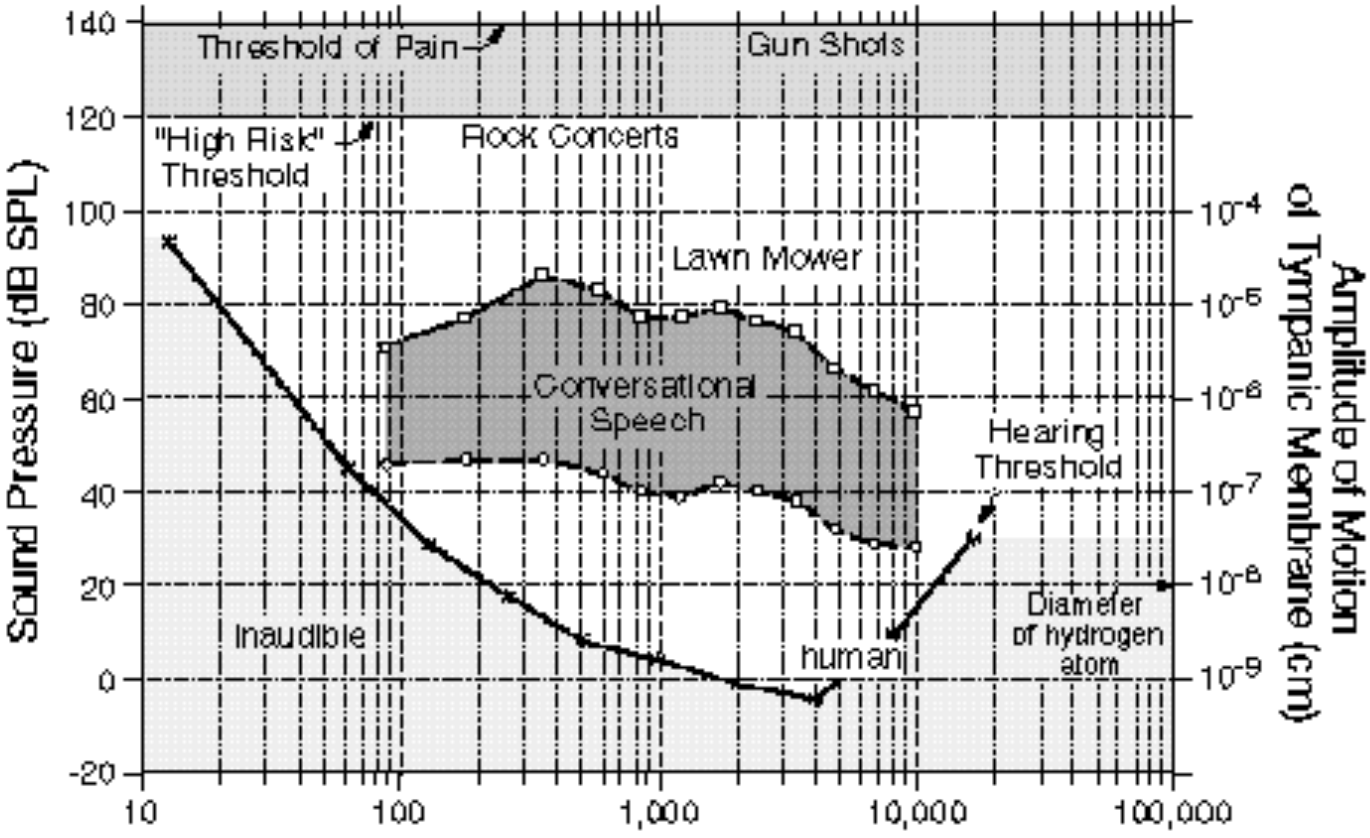


Fig. I-4

Weber's Law

- Weber's Law states that the just noticeable difference in level is a constant percetage of level:

$$\frac{\Delta I}{I} = \text{constant}$$

- For pure tones an increase of about 0.5 - 1 dB is just noticable

Spatial perception

April 28, 2020

Binaural Hearing

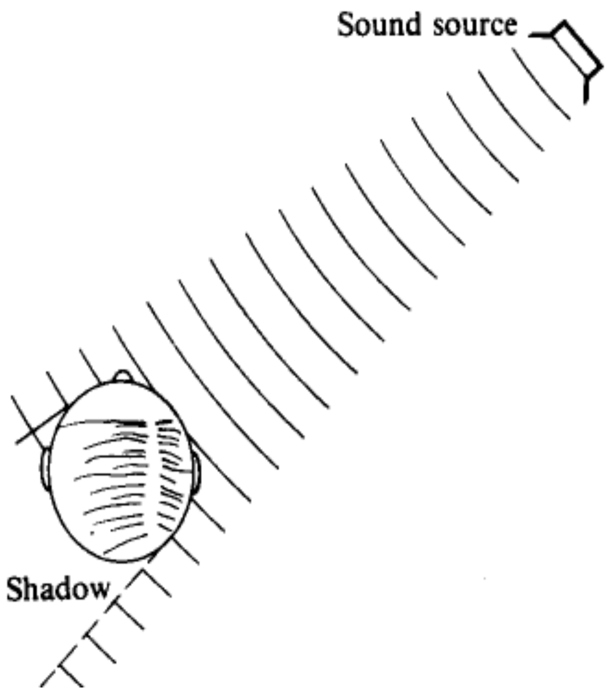
We have a remarkable ability to compare acoustic signals across the two ears. On headphones:

- Identical signals lead to a narrow sound image centered in our head (correlation is 1)
- An interaural level difference of 1 dB leads to a just noticeable shift in position towards the louder ear
- An interaural time difference of 15 μs leads to a just noticeable shift in position towards the leading ear
- When two signals are not identical (correlation <1), the sound image increases in width
- Changes in correlation are best audibile around values of 1.
- A reduction from 1 to 0.98 is audible, a reduction from 0.1 to 0 is not audible.

Correlation: $$C_{LR}(\tau) = \frac{\int L(t)R(t+\tau)dt}{\sqrt{\int L(t)^2 dt \int R(t+\tau)^2 dt}}$$

Spatial Hearing

Freefield listening: binaural cues help to localize sound source



- Low frequencies: Sound bends around the head (path length difference):
  Interaural *time* differences

- High frequencies: Sound is obstructed by the head:
  Interaural *level* differences

- Interaural time differences are not perceived above 2 kHz.
- Room reflections lead to reduction of coherence

TU Delft

Spectral and Temporal Masking

Literature:

Brian C.J. Moore, "An Introduction to the Psychology of Hearing", Fourth edition, Academic Press, London, 1997

T. Painter and A. Spanias, "Perceptual Coding of Digital Audio". *Proceedings of the IEEE,* Vol. 88, No. 4, pp. 451-513, April 2000

Masking

The phenomenon where a sound (maskee/test signal) that is perfectly audible in isolation is not audible due to the presence of a masking sound (masker)

Relevance for audio coding:

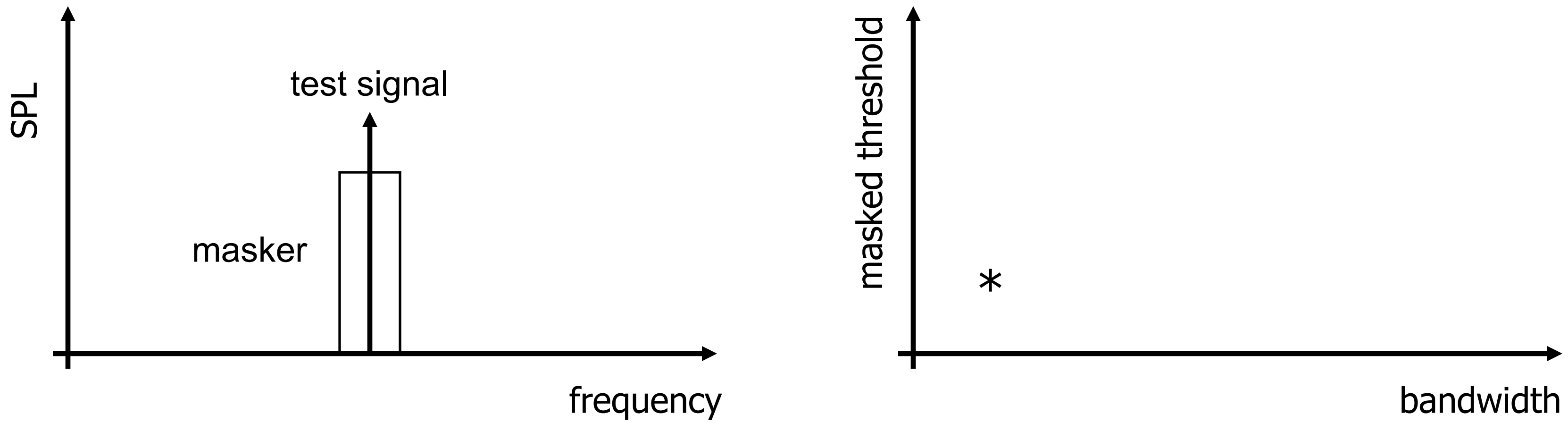
- Quantisation noise (*maskee*) that is introduced by the coding algorithm is masked by the signal which is coded (*masker*)
- By shaping the spectro-temporal shape of the distortion a very efficient coding of a signal is possible (1-2 bits/sample)

Encoded signal: (= Original + quantisation noise)

Quantisation noise:

Masking, Fletcher (1940)

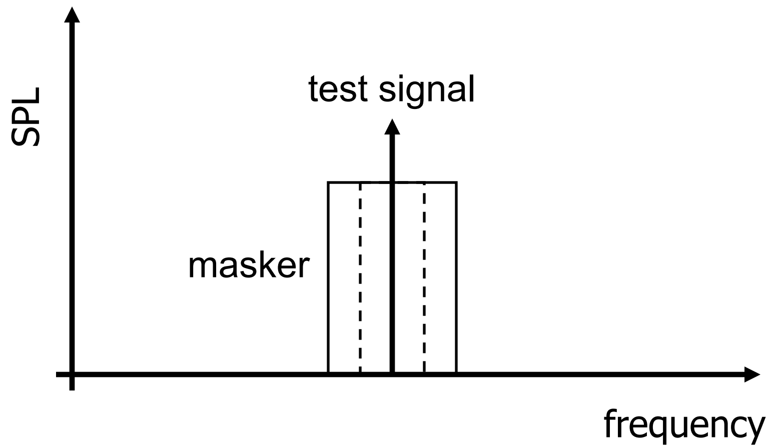
- Measure threshold of detectability of a tone masked by bandpass noise centred spectrally around the tone
- Measure thresholds as a function of masker bandwidth

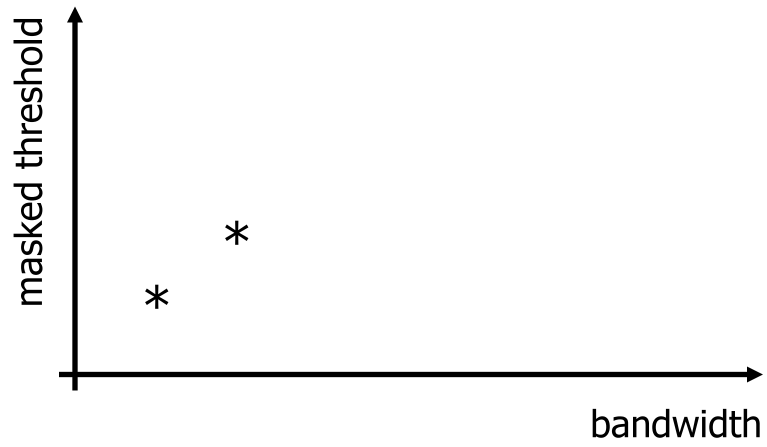


TUDelft

Masking, Fletcher (1940)

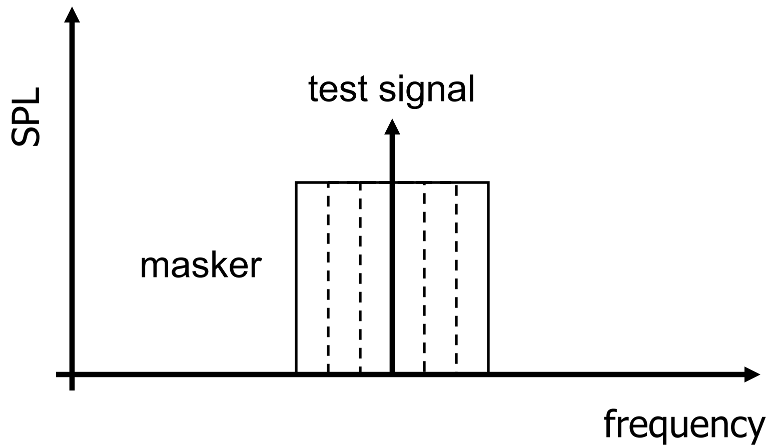
- Measure threshold of detectability of a tone masked by bandpass noise centred spectrally around the tone
- Measure thresholds as a function of masker bandwidth

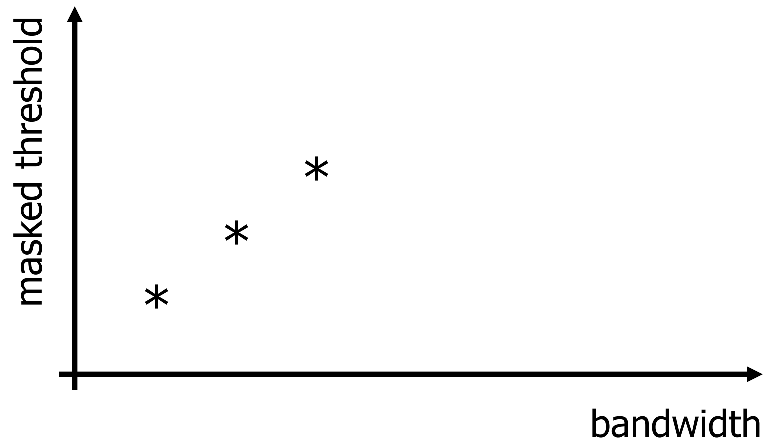




Masking, Fletcher (1940)

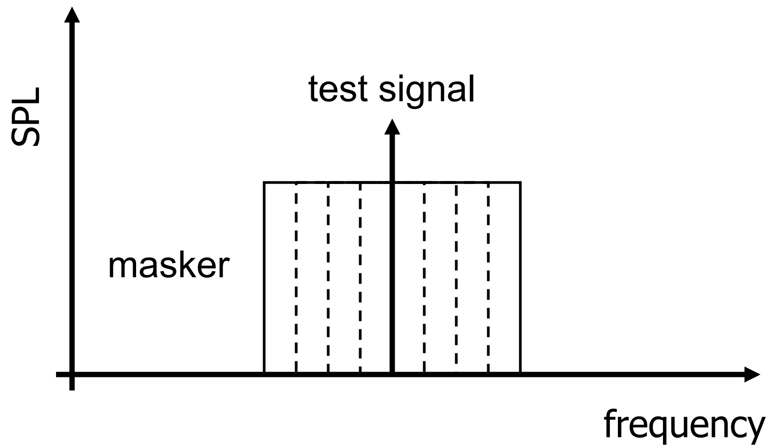
- Measure threshold of detectability of a tone masked by bandpass noise centred spectrally around the tone
- Measure thresholds as a function of masker bandwidth

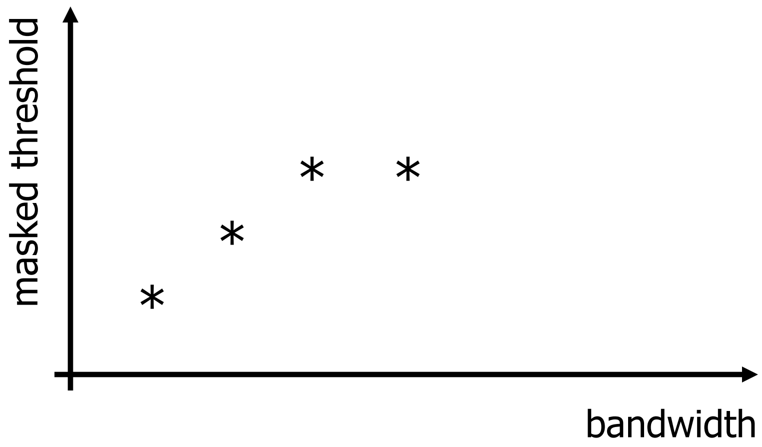




Masking, Fletcher (1940)

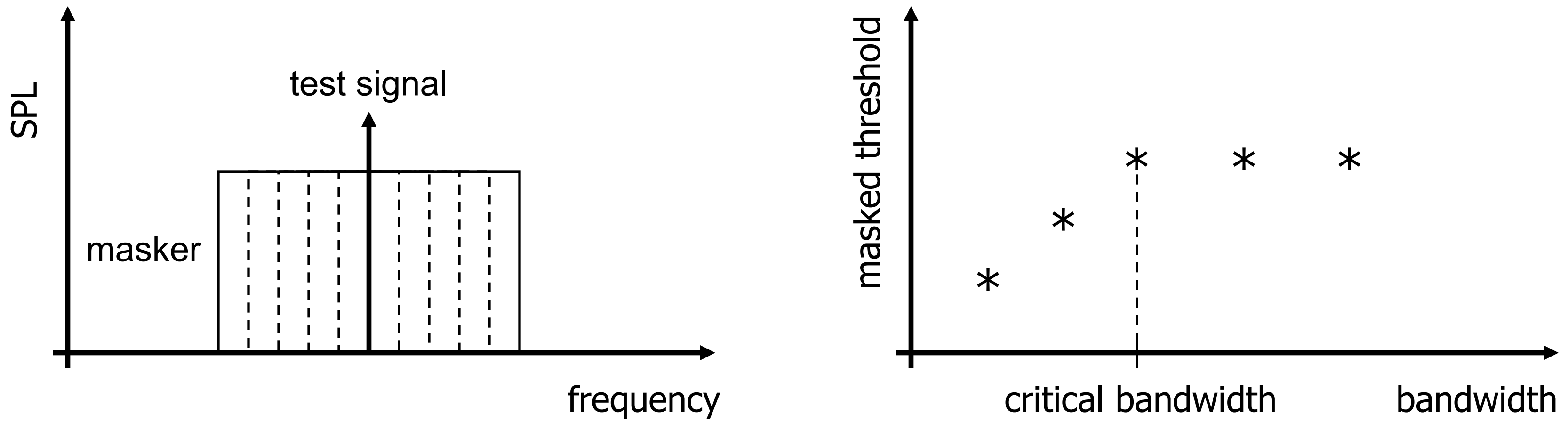
- Measure threshold of detectability of a tone masked by bandpass noise centred spectrally around the tone
- Measure thresholds as a function of masker bandwidth





Masking, Fletcher (1940)

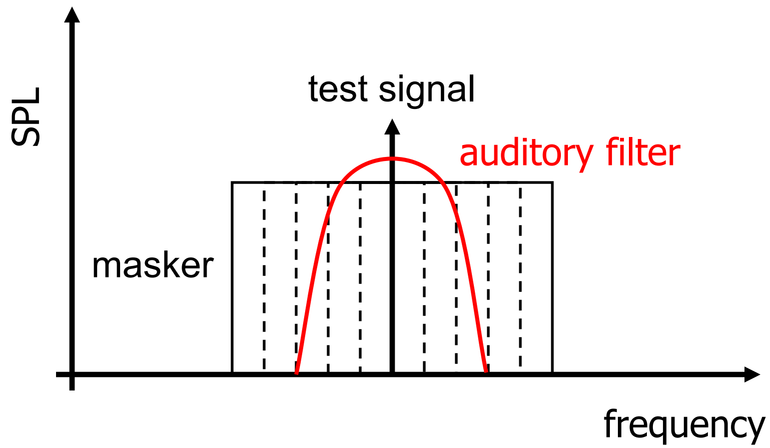
- Measure threshold of detectability of a tone masked by bandpass noise centred spectrally around the tone
- Measure thresholds as a function of masker bandwidth

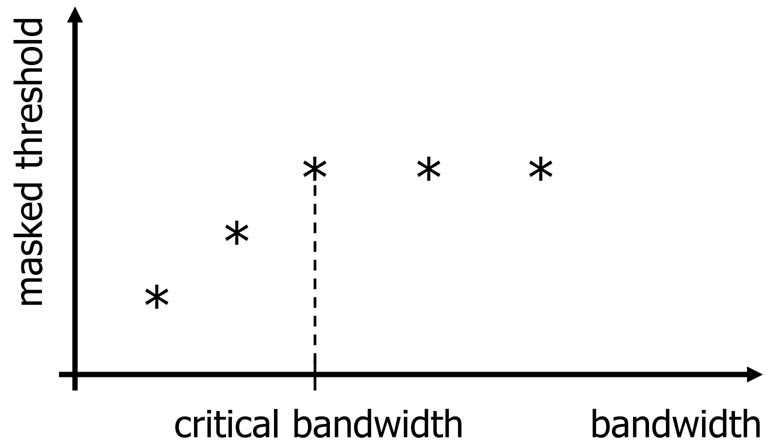


TU Delft

Masking, Fletcher (1940)

Assumption: signal is detected when the signal-to-noise ratio at the output of the auditory filter exceeds a certain criterion value





TU Delft

Demo: Critical Bands

Masking of a single 2000 Hz tone (decreasing in 10 steps of 5 dB) by spectrally flat (white) noise of different bandwidths:

- broadband noise
- bandwidth 1000 Hz
- bandwidth 250 Hz
- bandwidth 10 Hz

Bark Scale

- A scale that converts frequency (Hz) into units of critical bandwidth

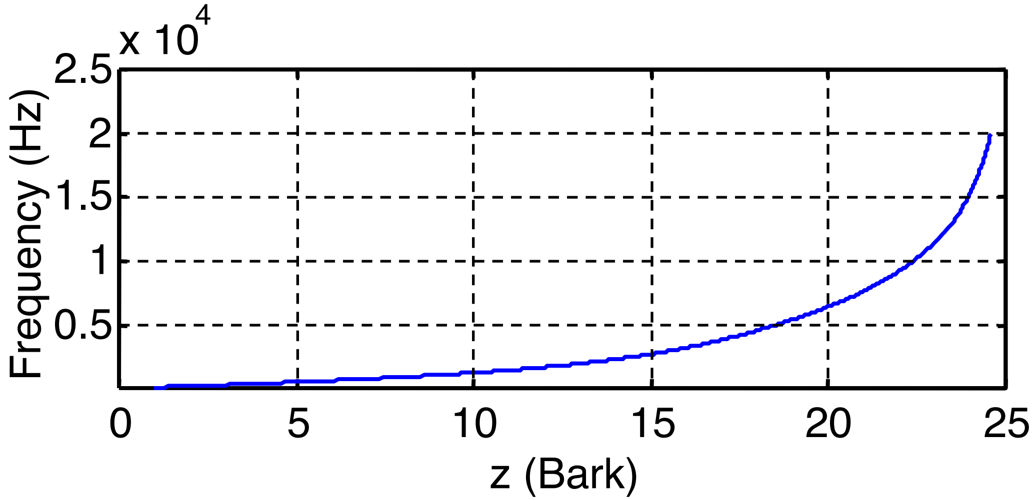
$$z(f) = 13\arctan(0.00076f) + 3.5\arctan\left[\left(\frac{f}{7500}\right)^2\right] \quad \text{(Bark)}$$

named after Heinrich Barkhausen (who proposed the first subjective measurements of loudness)

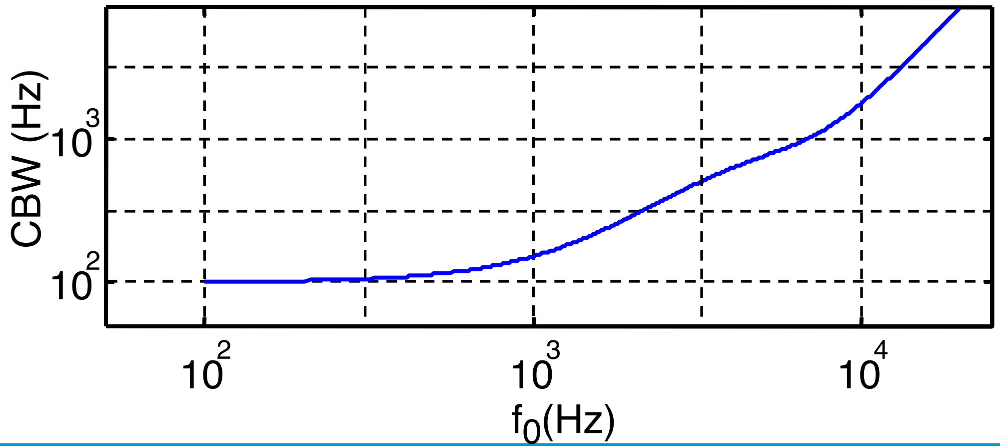
- Critical bandwidth:

$$\text{CBW} = \frac{\partial f}{\partial z}$$

Bark Scale and Critical Bandwidth



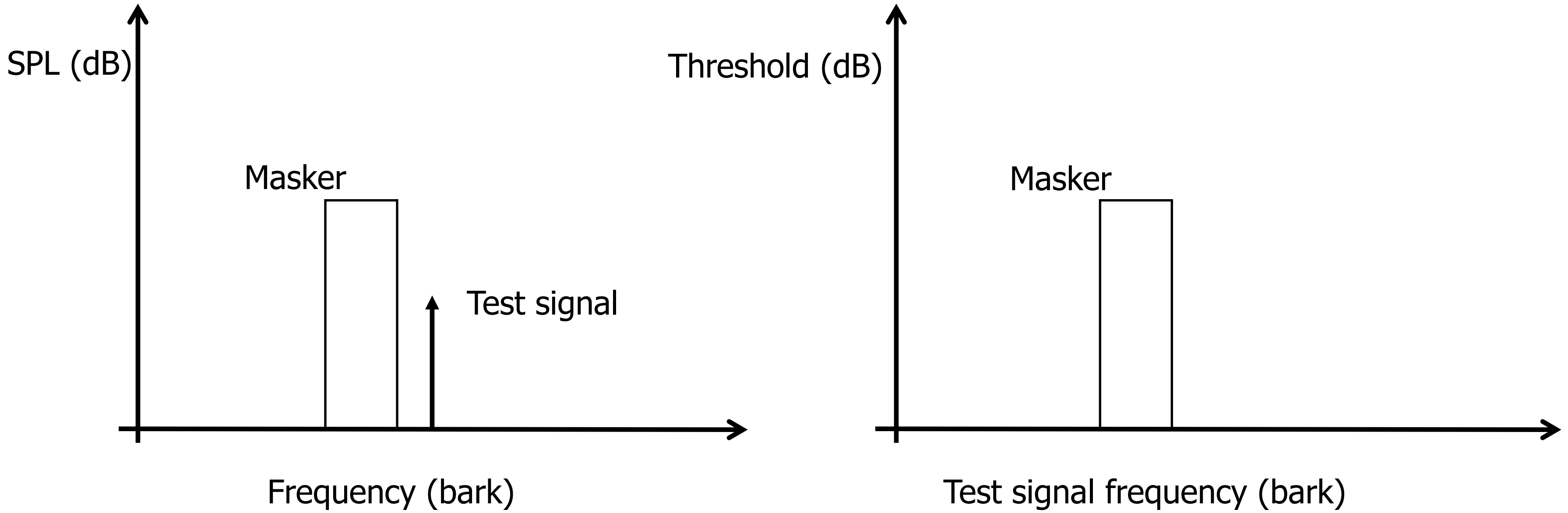
Bark scale



Critical bandwidth

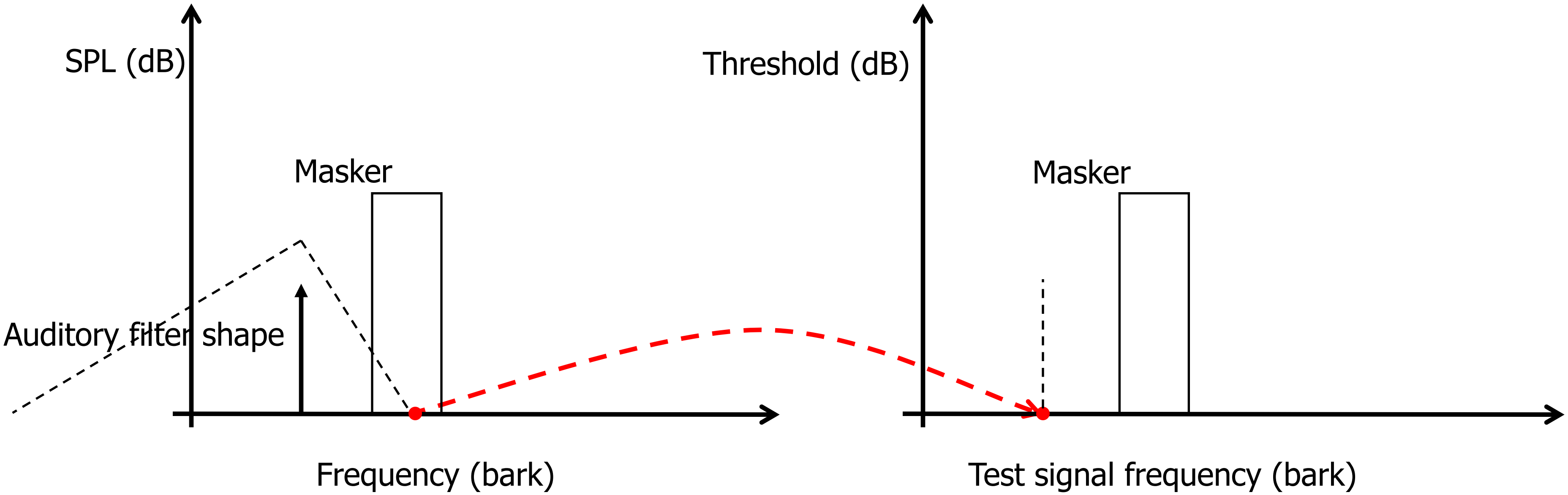
TU Delft

Spectral Spread of Masking



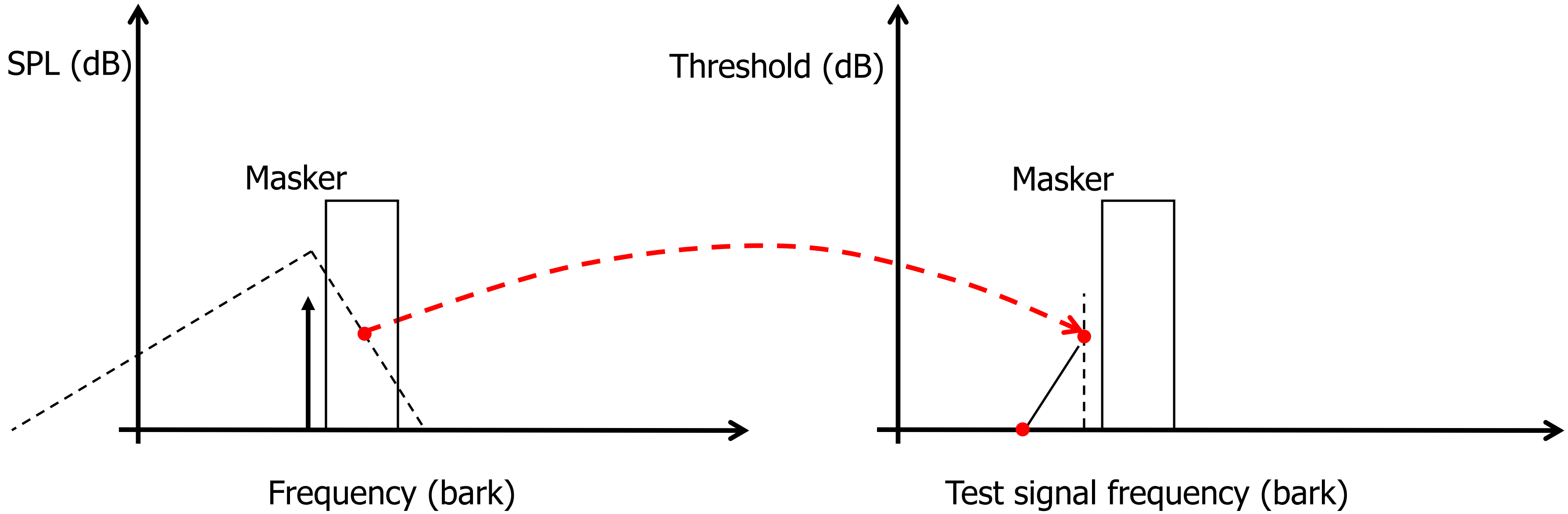
Test signal is detected when the test-signal-to-masker ratio at the output of the auditory filter exceeds a certain criterion value *k*

Spectral Spread of Masking



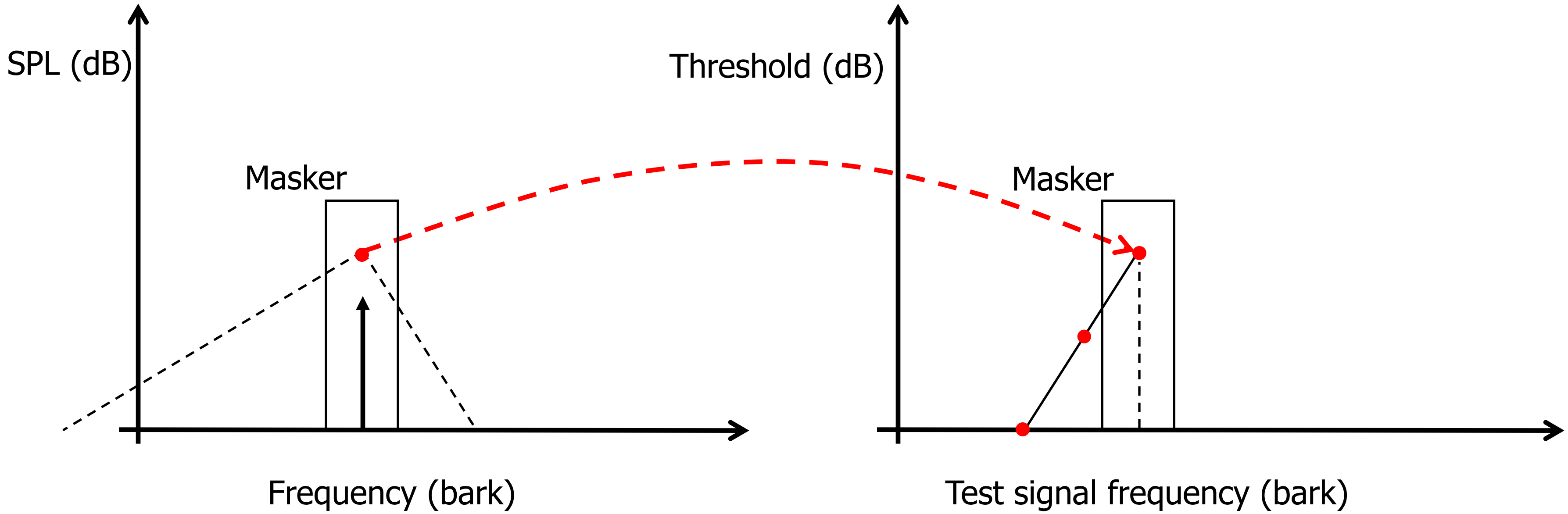
Test signal is detected when the test-signal-to-masker ratio at the output of the auditory filter exceeds a certain criterion value *k*

Spectral Spread of Masking



Test signal is detected when the test-signal-to-masker ratio at the output of the auditory filter exceeds a certain criterion value *k*

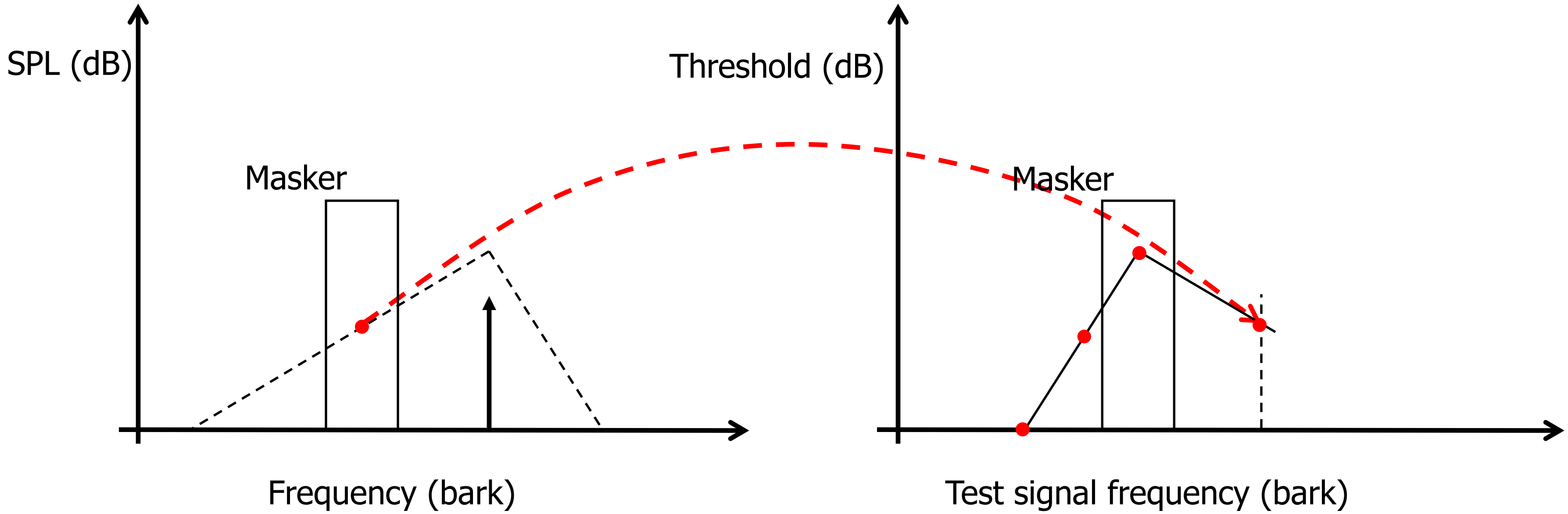
Spectral Spread of Masking



Test signal is detected when the test-signal-to-masker ratio at the output of the auditory filter exceeds a certain criterion value *k*

TU Delft

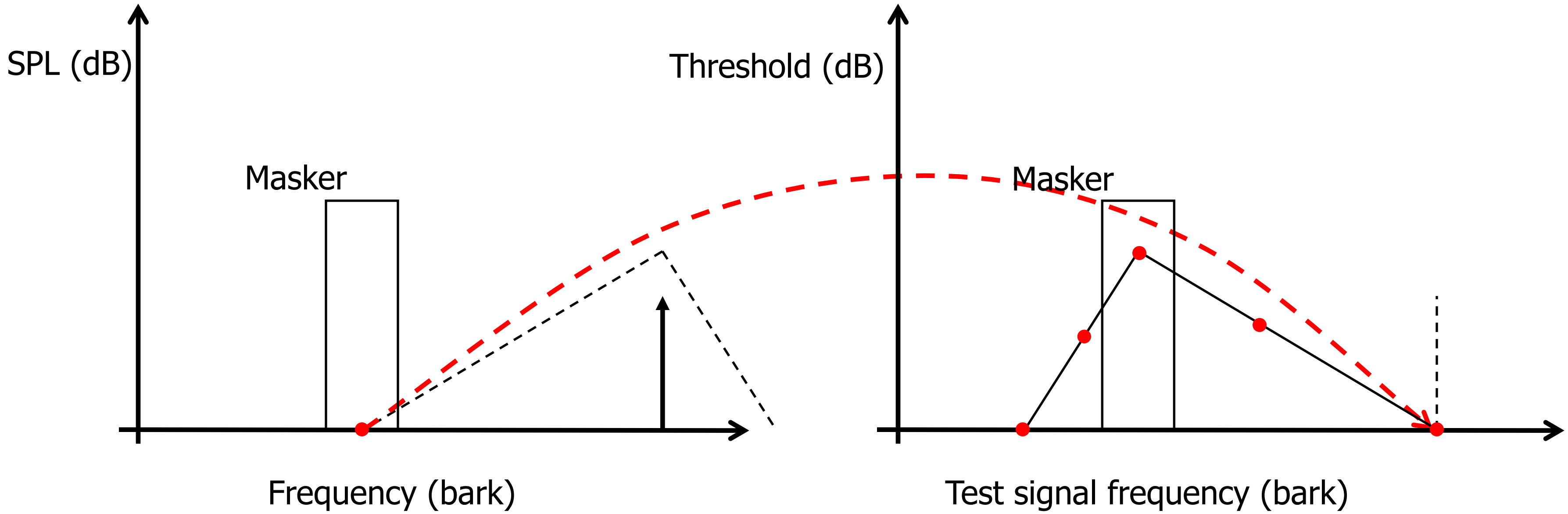
Spectral Spread of Masking



Test signal is detected when the test-signal-to-masker ratio at the output of the auditory filter exceeds a certain criterion value *k*

TU Delft

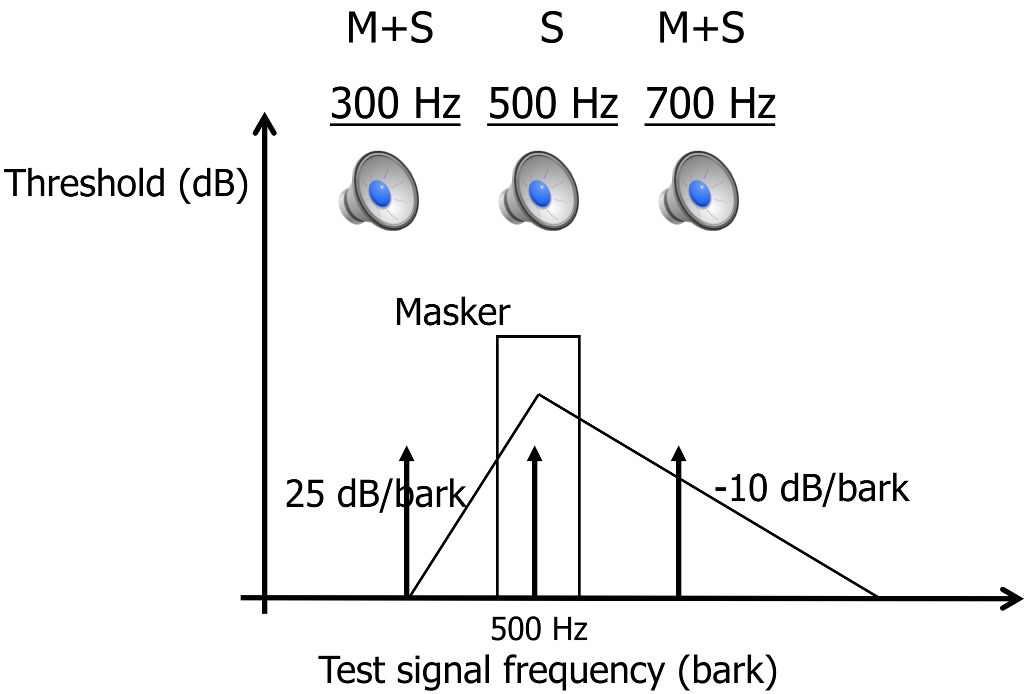
Spectral Spread of Masking



Test signal is detected when the test-signal-to-masker ratio at the output of the auditory filter exceeds a certain criterion value *k*

TUDelft

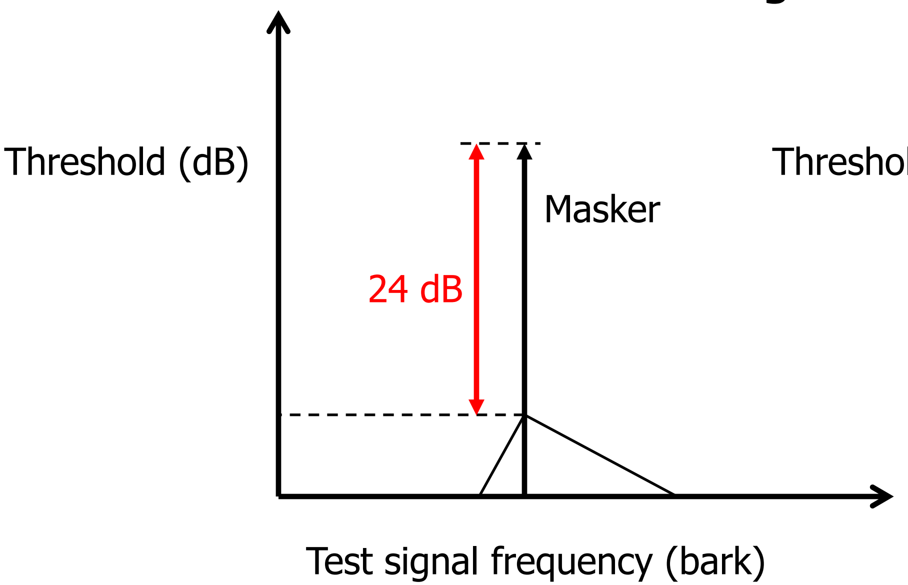
Spectral Spread of Masking



Test signal is detected when the test-signal-to-masker ratio at the output of the auditory filter exceeds a certain criterion value *k*

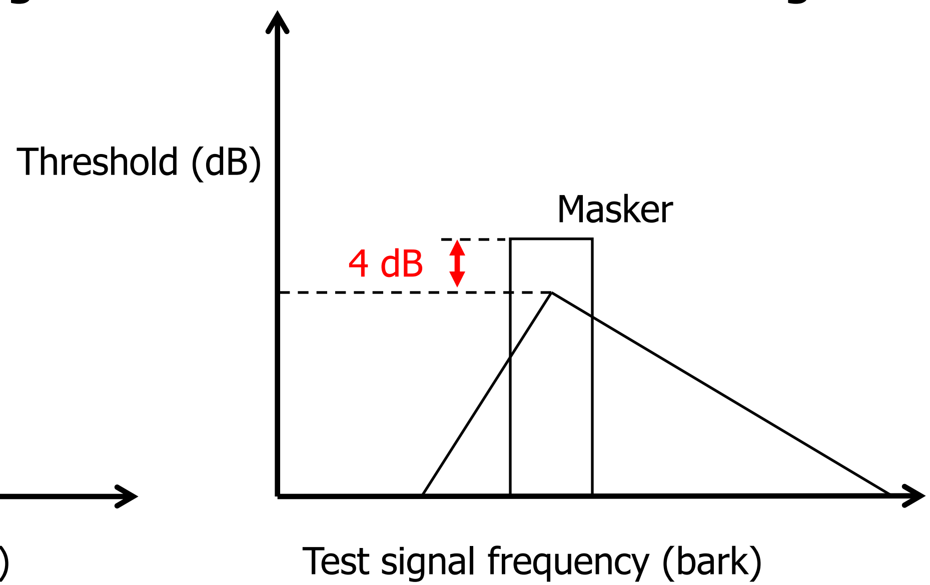
Tonal versus Noise Maskers

Tone-on-tone masking



Threshold is about 24 dB below masker (k = -24 dB)

Noise-on-tone masking



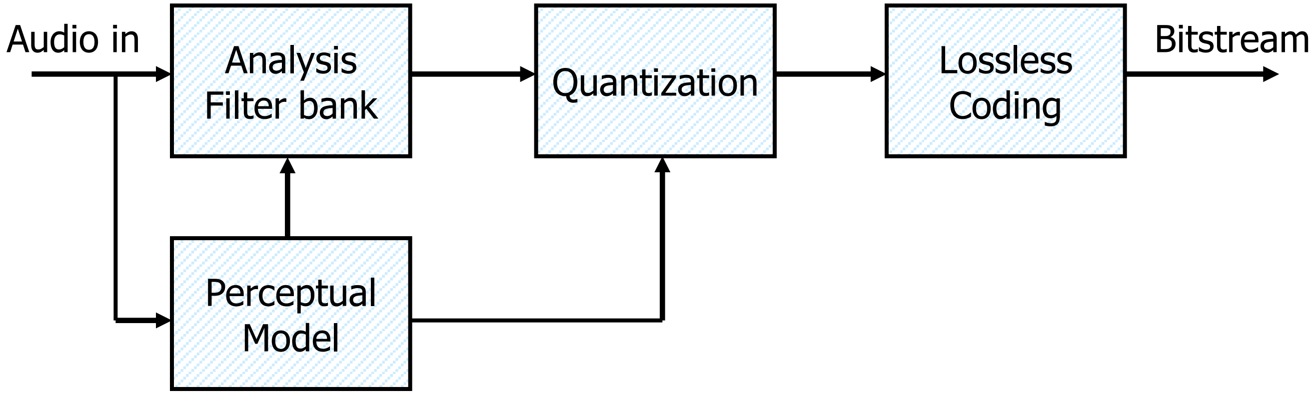
Threshold is about 4 dB below masker (k = -4 dB)

Perceptual Audio Coding

April 28, 2020

Perceptual Audio Coder

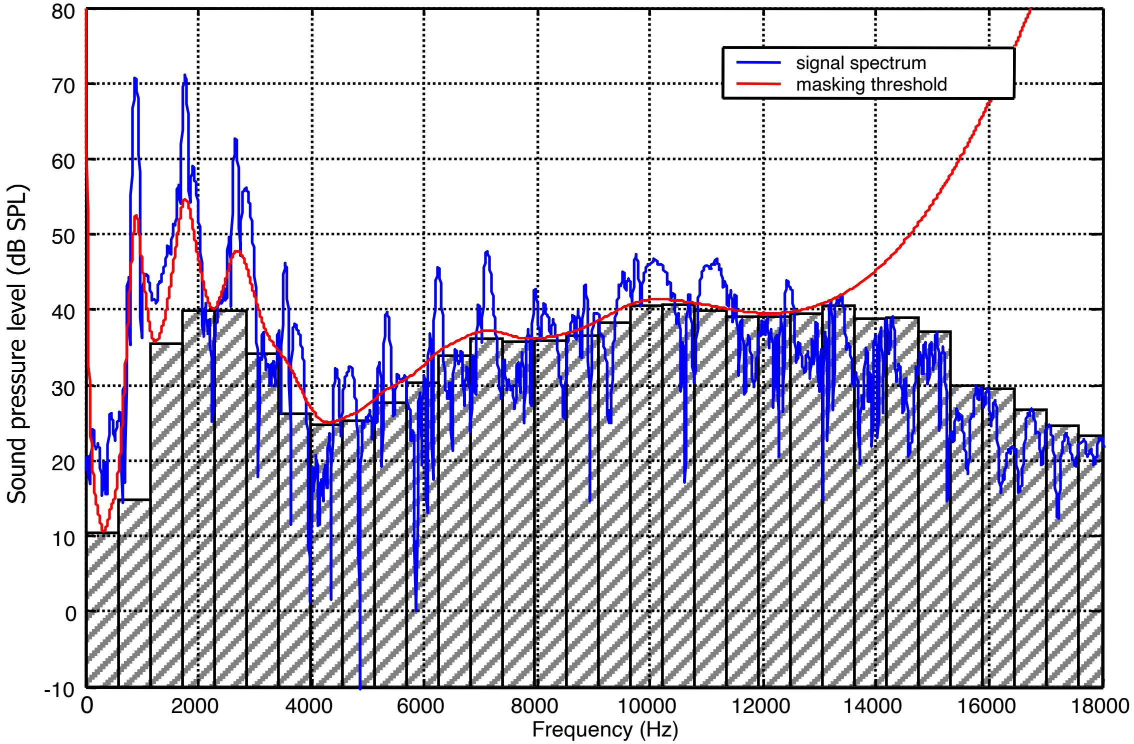
Encoder:



Decoder:

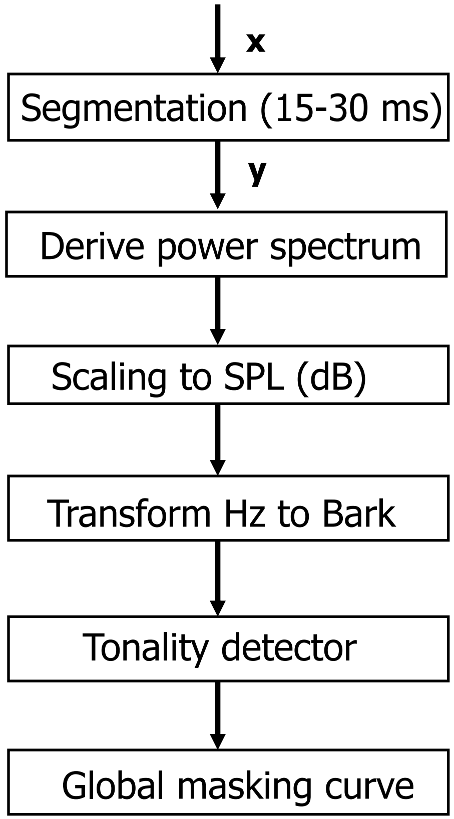
Irrelevancy Removal

Quantize every spectral component such that the quantization error stays below the masking threshold



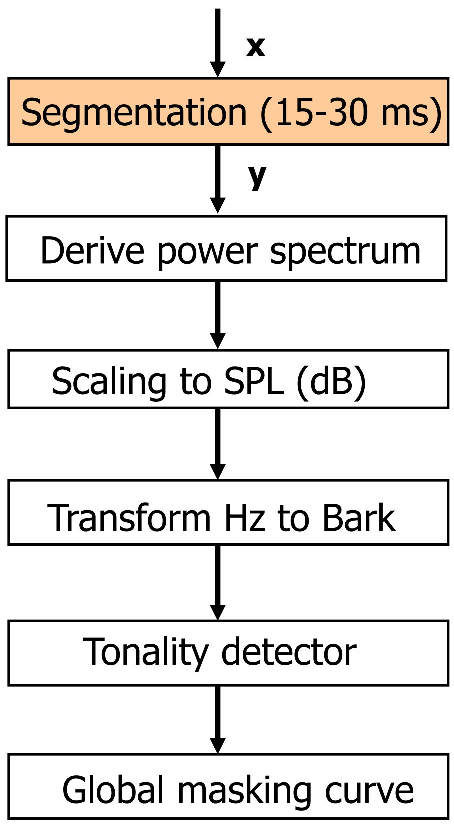
TU Delft

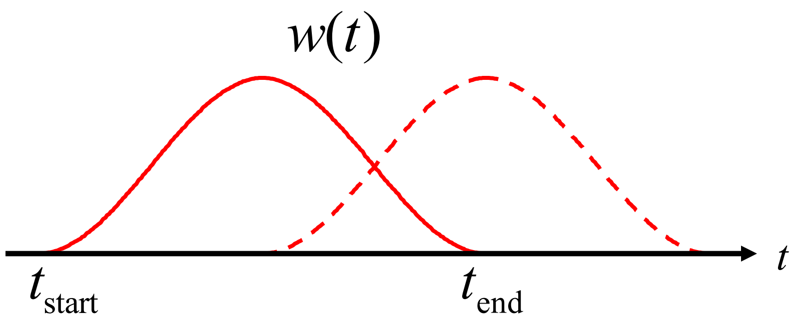
Spreading Function Based Masking Models (ISO MPEG Model)



TU Delft

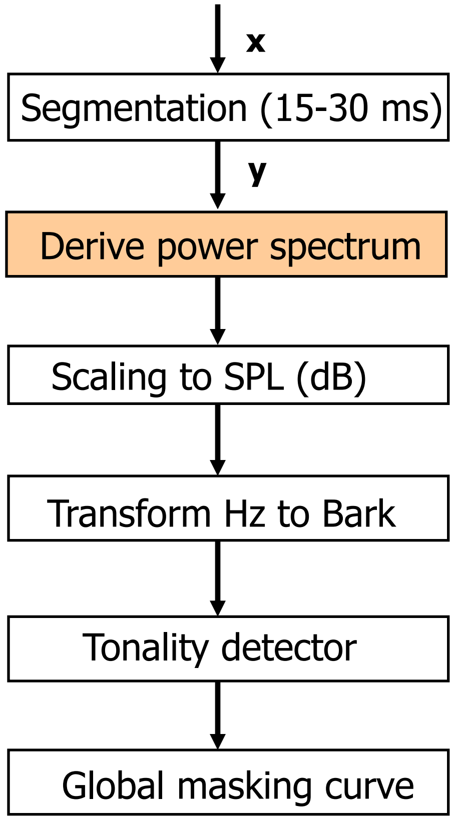
Spreading Function Based Masking Models (ISO MPEG Model)





$$y(t) = w(t)x(t), \quad t = t_{\text{start}} \ldots t_{\text{end}}$$

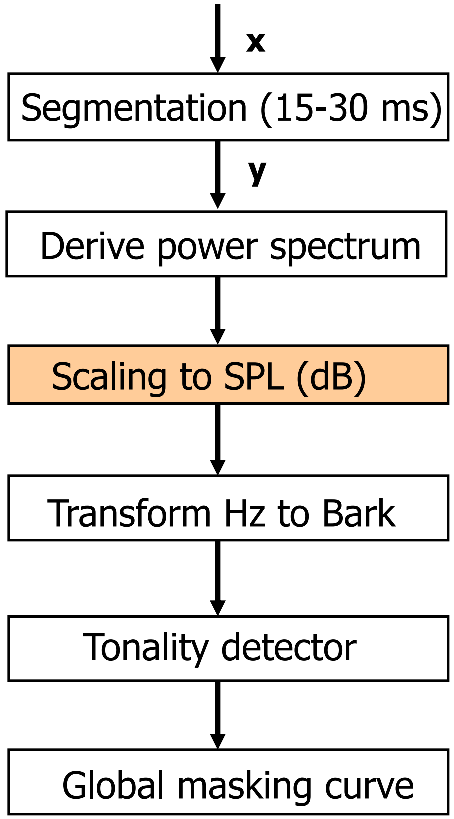
Spreading Function Based Masking Models (ISO MPEG Model)



$$\text{psd}(f_k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} y(n)e^{-j\frac{2\pi}{N}kn}\right|^2 \qquad f_k = \frac{kf_s}{N}$$

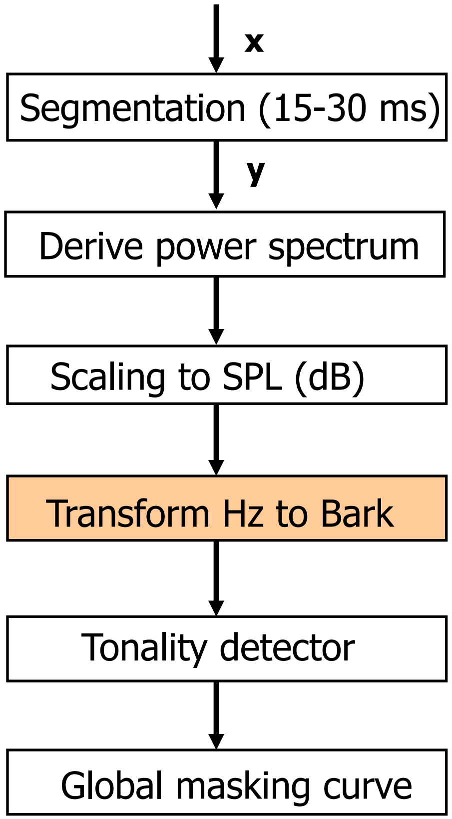
TU Delft

Spreading Function Based Masking Models (ISO MPEG Model)



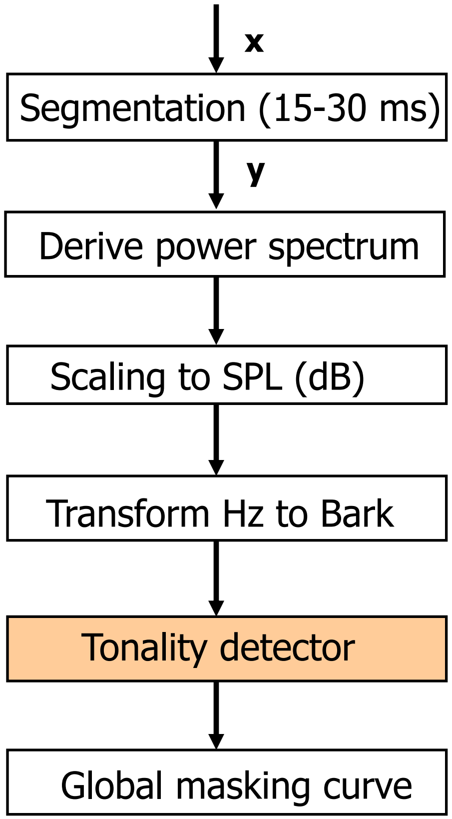
$$\mathrm{SPL}(f_k) = 10\log_{10}\left(\frac{\mathrm{psd}(f_k)}{\mathrm{psd}_{0\mathrm{dB}}}\right)$$

Spreading Function Based Masking Models (ISO MPEG Model)



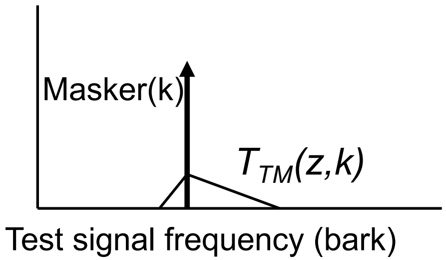
$$z(f) = 13\arctan(0.00076f) + 3.5\arctan\left[\left(\frac{f}{7500}\right)^2\right] \quad \text{(Bark)}$$

Spreading Function Based Masking Models (ISO MPEG Model)

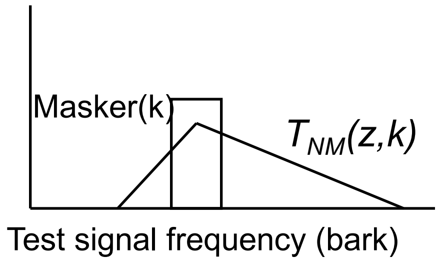


- Because of the difference in masking strength we need a classification of tonal and noisy components in a signal spectrum
- Spectral flatness measure: Tonal if the power spectral density of a component exceeds nearby components by more than e.g. 7 dB

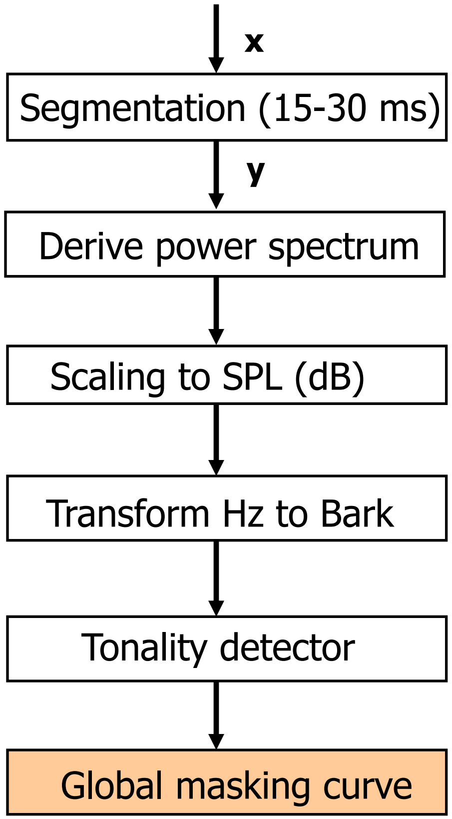
Tone on tone masking



Noise on tone masking



Spreading Function Based Masking Models (ISO MPEG Model)



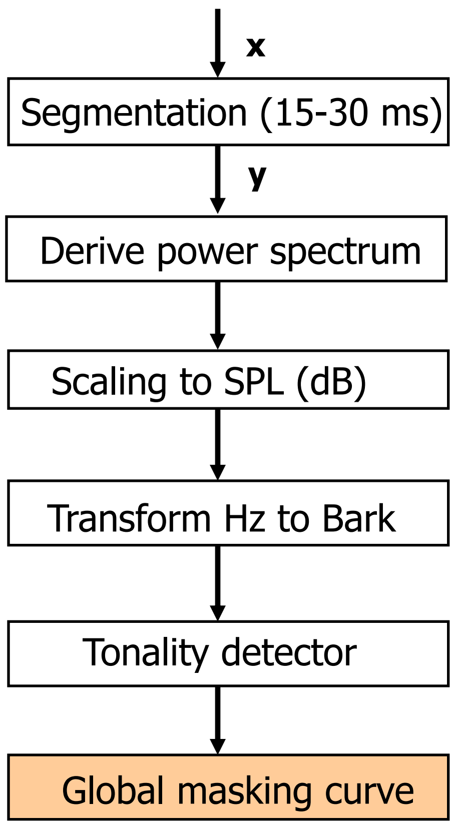
- Tonal maskers (k): Threshold for each *z*: $T_{TM}(z,k)$
- Noise maskers (k): Threshold for each *z*: $T_{NM}(z,k)$
- Threshold in quiet:

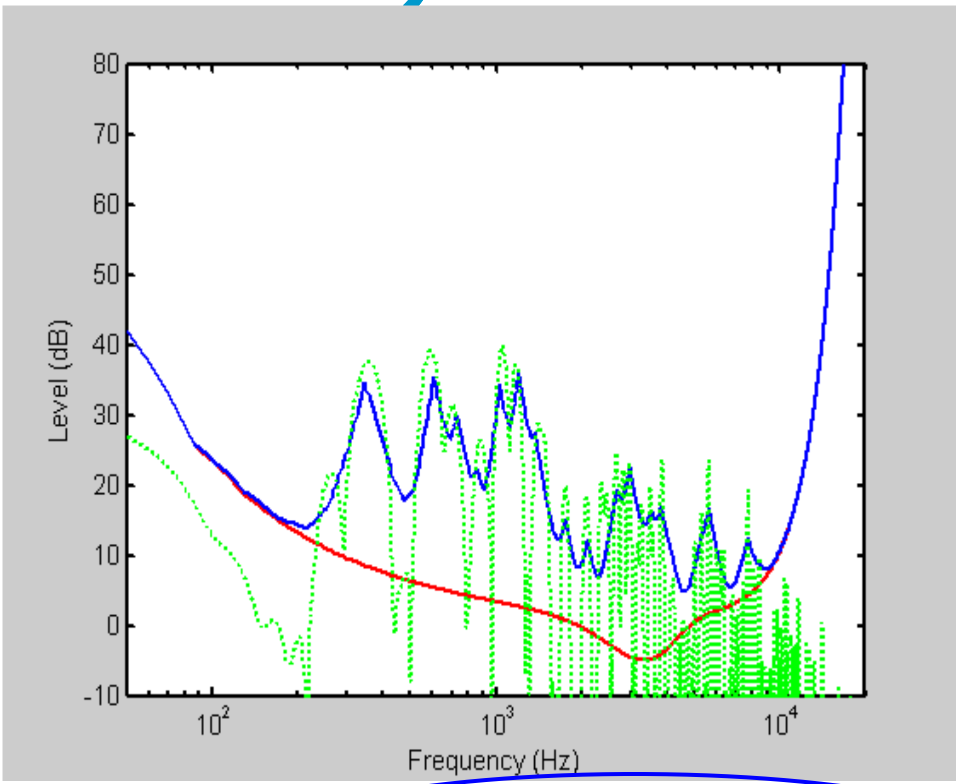
$$T_q(f) = 3.64 f^{-0.8} - 6.5 e^{-0.6*(f-3.3)^2} + 10^{-3} f^4$$

- Combining masking components by power addition:

$$T_g(z) = 10^{10}\log\left(10^{0.1T_q(z)} + \sum_k 10^{0.1T_{TM}(z,k)} + \sum_k 10^{0.1T_{NM}(z,k)}\right)$$

Spreading Function Based Masking Models (ISO MPEG Model)





$$T_g(z) = 10^{10}\log\left(10^{0.1T_q(z)} + \sum_k 10^{0.1T_{TM}(z,k)} + \sum_k 10^{0.1T_{NM}(z,k)}\right)$$

TU Delft

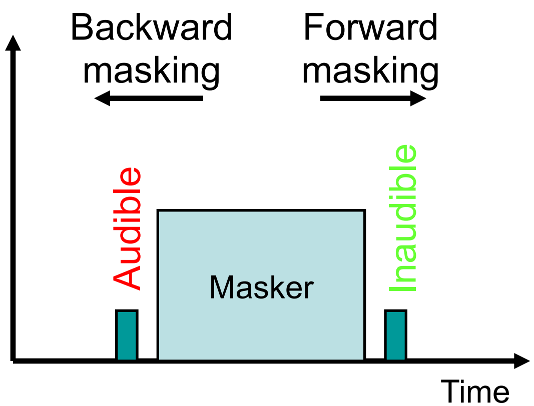
Meaning of Global Masking Curve

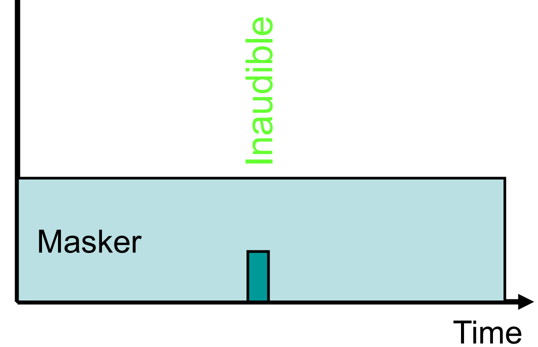
- The (quantisation) noise-to-masker ratio (NMR) specifies the audibility of the quantisation noise
- Quantisation noise inaudible when: NMR< 0 dB
- The NMR needs to be considered within each critical band

Forward and Backward Masking

Forward masking has a much stronger effect than backward masking

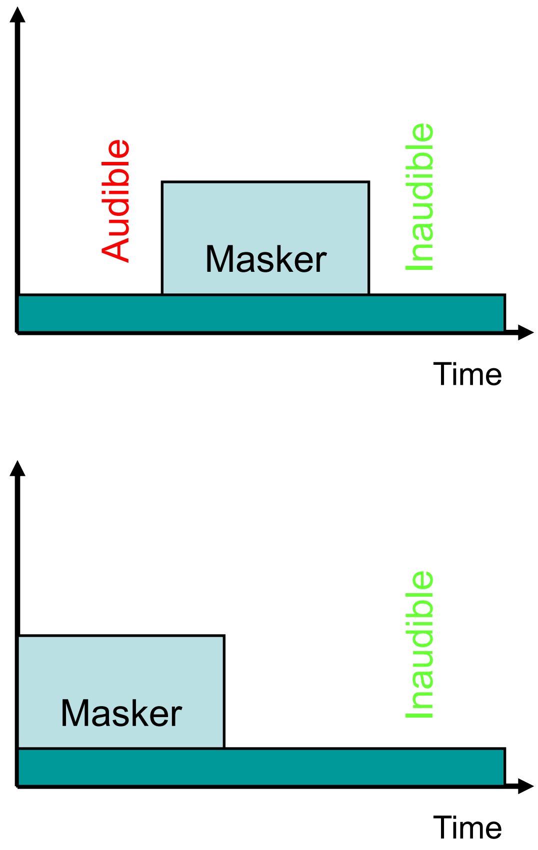
- Forward masking: Effect is observed until 100-200 ms after masker
- Backward masking: 10 ms, but sometimes not even present
- Simultaneous masking has the strongest effect





TUDelft

Forward and Backward Masking

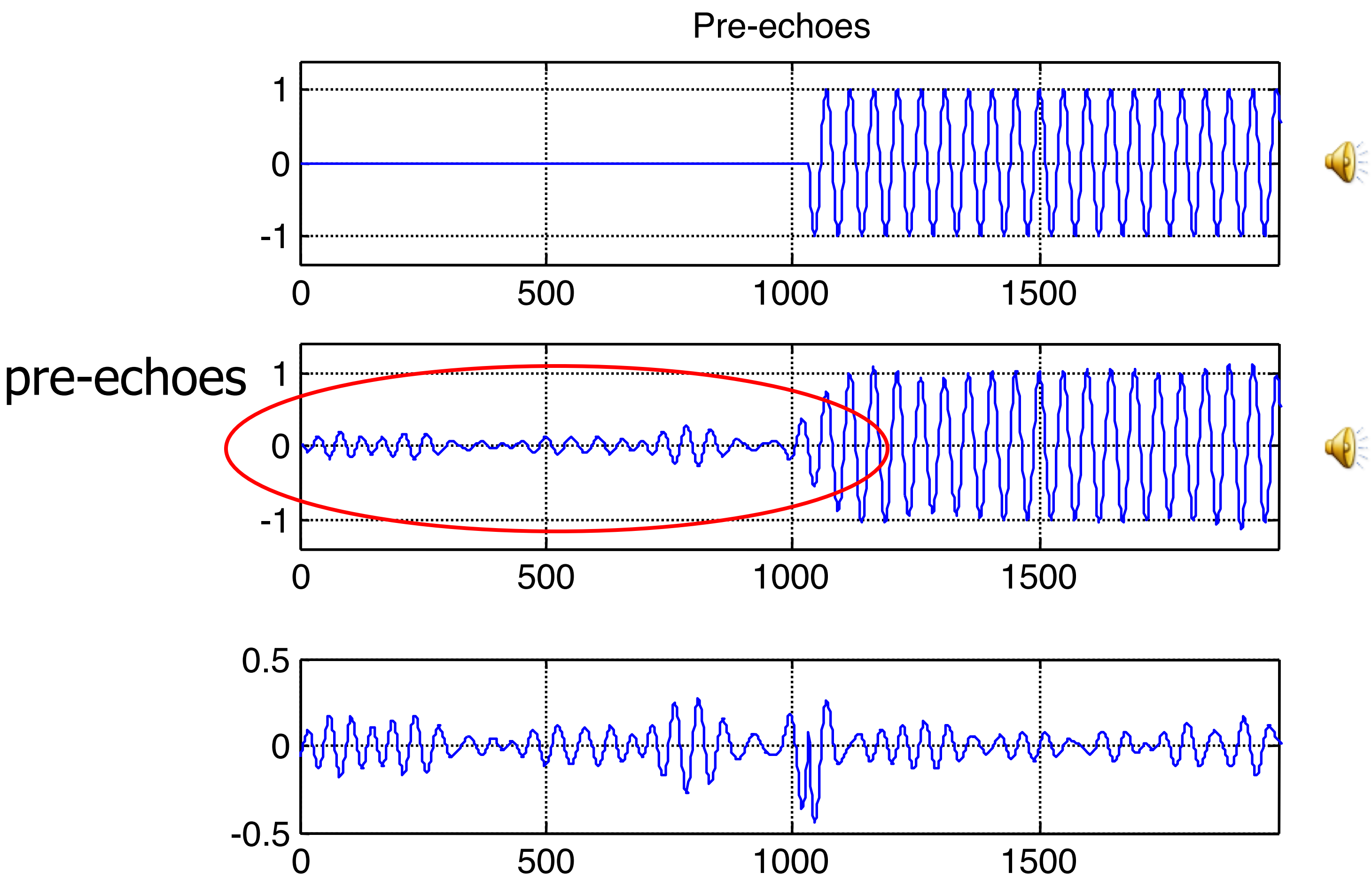


More realistic situation for audio coding:

- Quantisation noise is distributed uniformly across a segment
- Position of masker in segment is very relevant

TU Delft

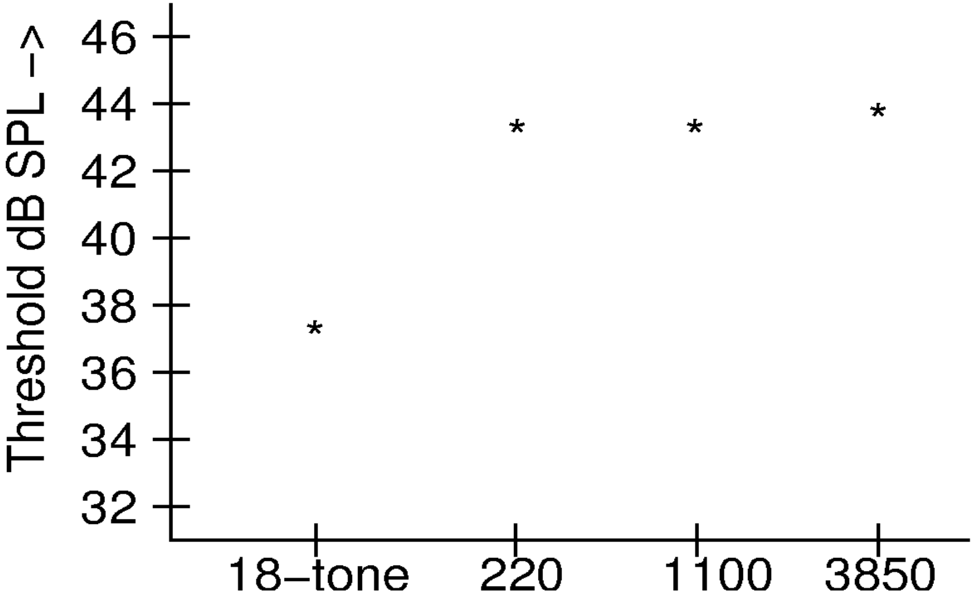
Pre-Echoes



TU Delft

Other factors that influence masking

Detection of complex signals Buus et al. 1986



Multiband energy detector model (Green and Swets, 1966):

- Spectral integration of "detectability" information
- Each doubling of the number of components leads to a 1.5 dB reduction in threshold
- Trading detectability across frequency

TU Delft

Concluding Remarks

- Although the auditory system is complicated, usefull quantitative perceptual models can be made
- It is impossible to include all psycho-acoustical effects in a perceptual model
- Often heuristic modifications to the model or encoder are used
- The quality of an audio coder can depend strongly on the quality of the perceptual model