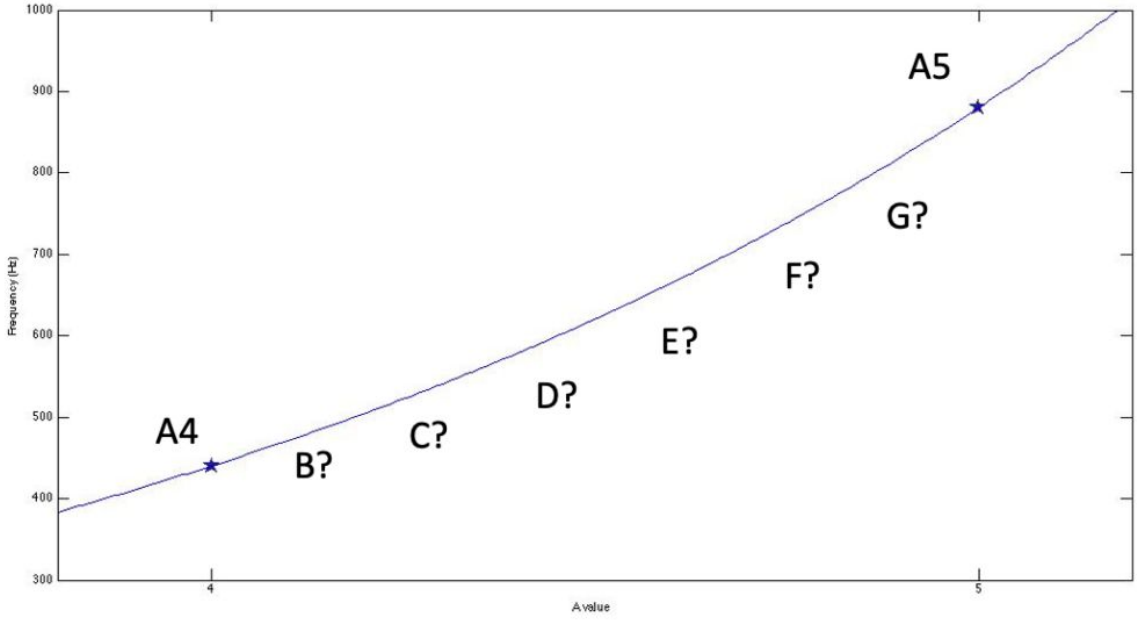
Психоакустические шкалы

П. А. Холявин

p.kholyavin@spbu.ru

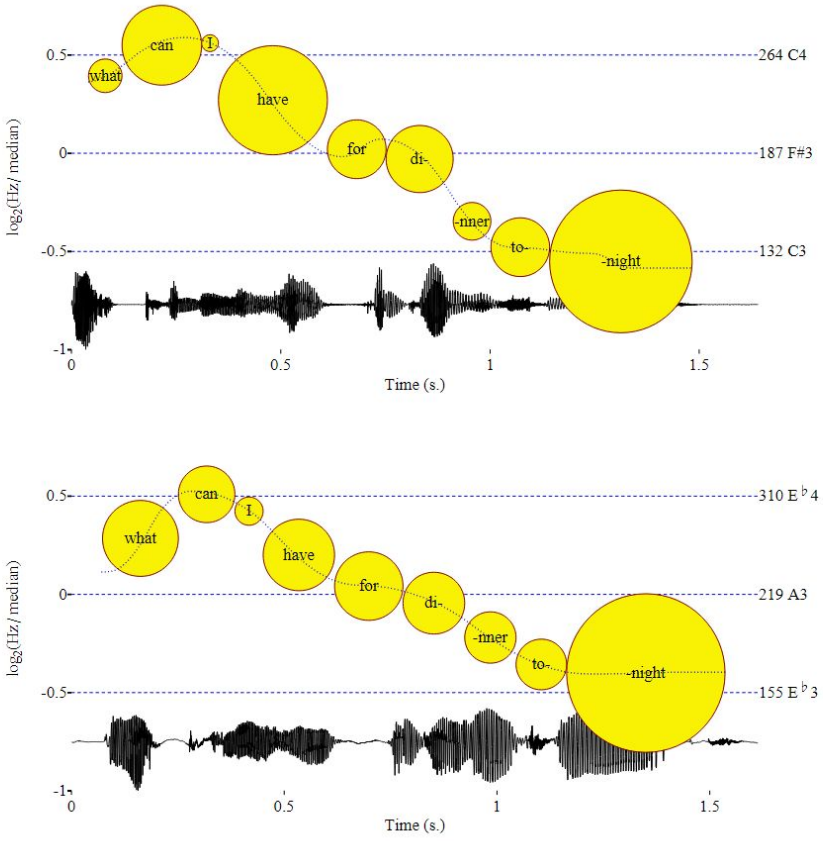
Шкала полутонов



$$st = 12 \cdot \log_2 \frac{f_2}{f_1}$$

Note	A4	A#4	B4	C5	C#5	D5	D#5	E5	F5	F#5	G5	G#5	A5
Pitch (Hz)	440.0	466.2	493.9	523.3	554.4	587.3	622.3	659.3	698.5	740.0	766.0	830.6	880.0

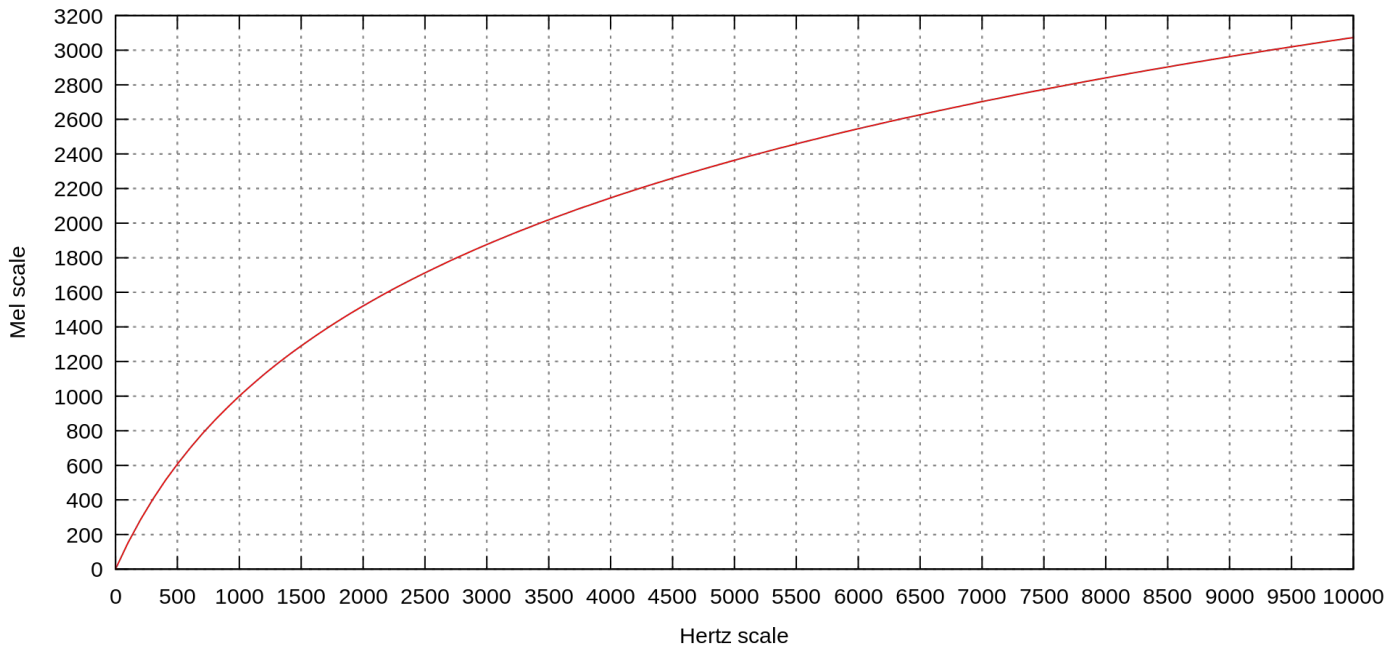
Октавно-медианная шкала



$$ome = log_2(\frac{Hz}{median})$$

Шкала мелов

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$



Шкала барков

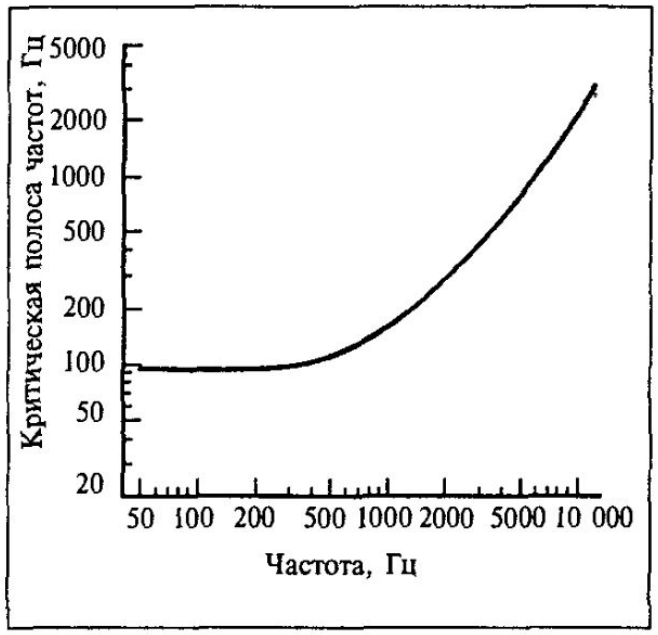
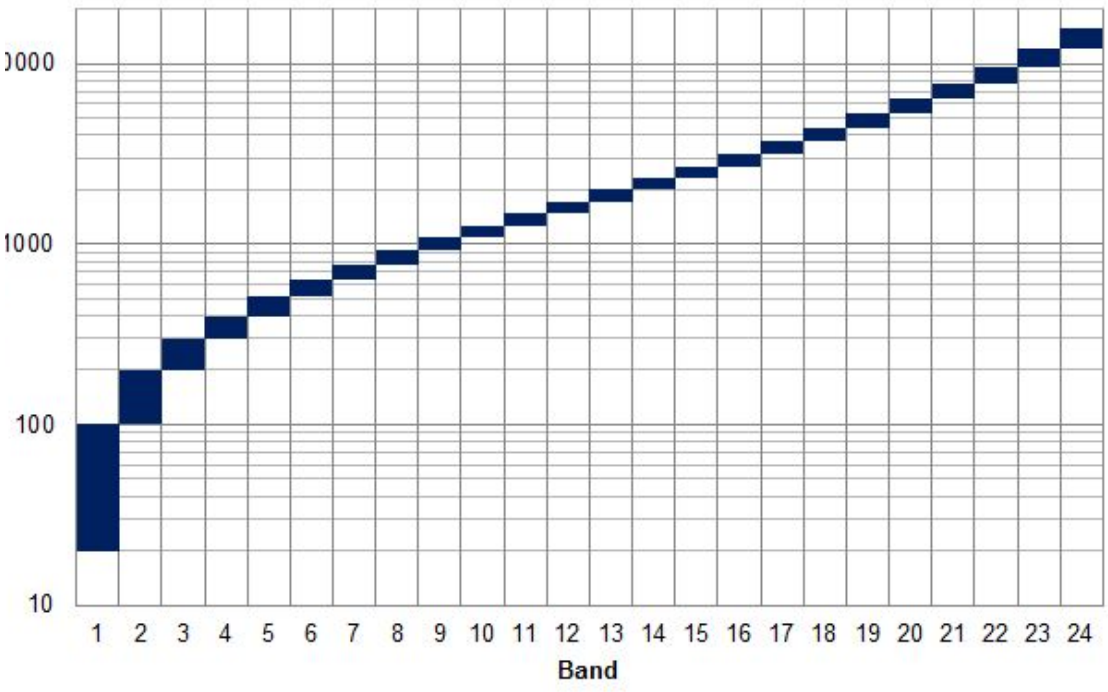


Рис. 4.14. Зависимость ширины критической полосы слуха от ее средней частоты

$$\text{Bark} = 13\arctan(0.00076f) + 3.5\arctan((f/7500)^2)$$



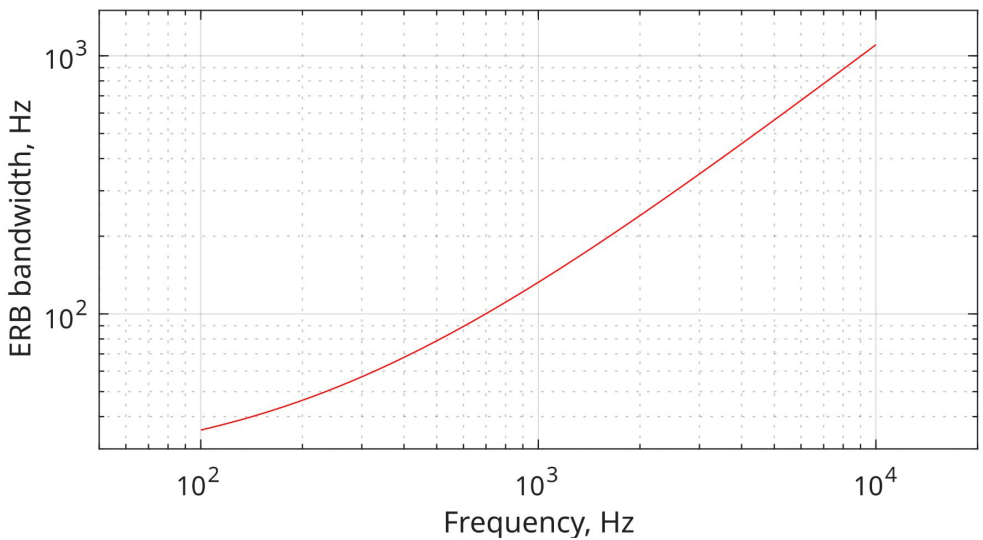
Шкала эрбов

$$\mathrm{ERB}(F) = 6.23 \cdot F^2 + 93.39 \cdot F + 28.52$$

$$\mathrm{ERB}(f) = 24.7\ \mathrm{Hz} \cdot \left(\frac{4.37 \cdot f}{1000\ \mathrm{Hz}} + 1 \right)$$

$$\mathrm{ERBS}(f) = 11.17 \cdot \ln\left(\frac{f + 0.312}{f + 14.675} \right) + 43.0$$

$$\mathrm{ERBS}(f) = 21.4 \cdot \log_{10}(1 + 0.00437 \cdot f)$$



Кривые равной громкости

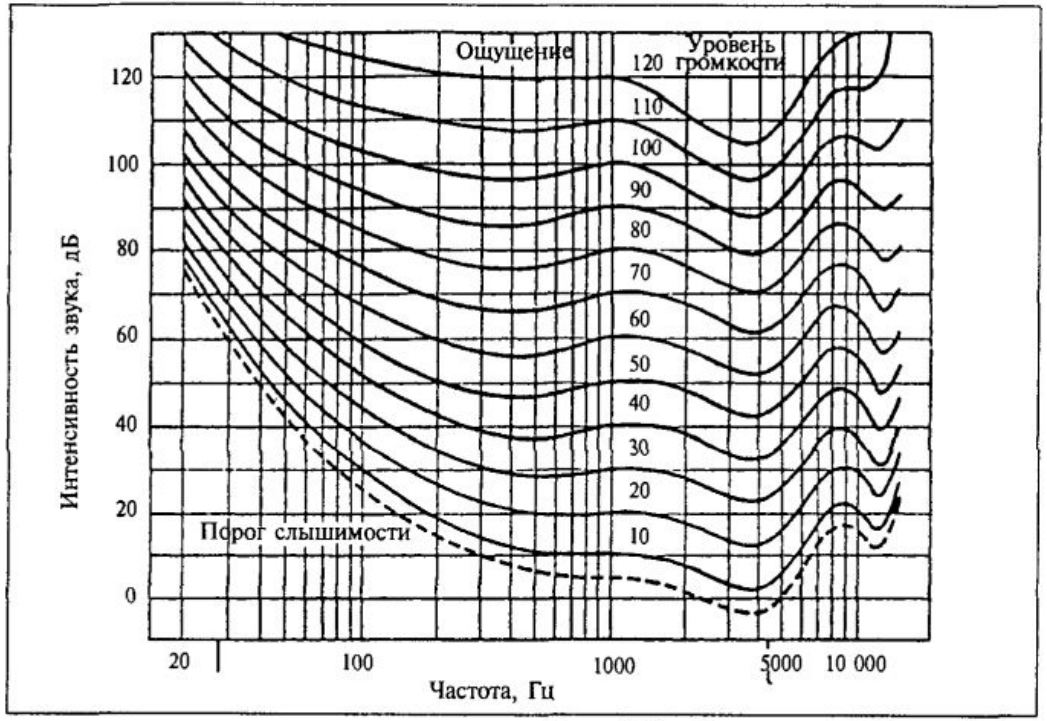


Рис. 4.11. Кривые равной громкости для тонов разных частот и интенсивностей

Громкость

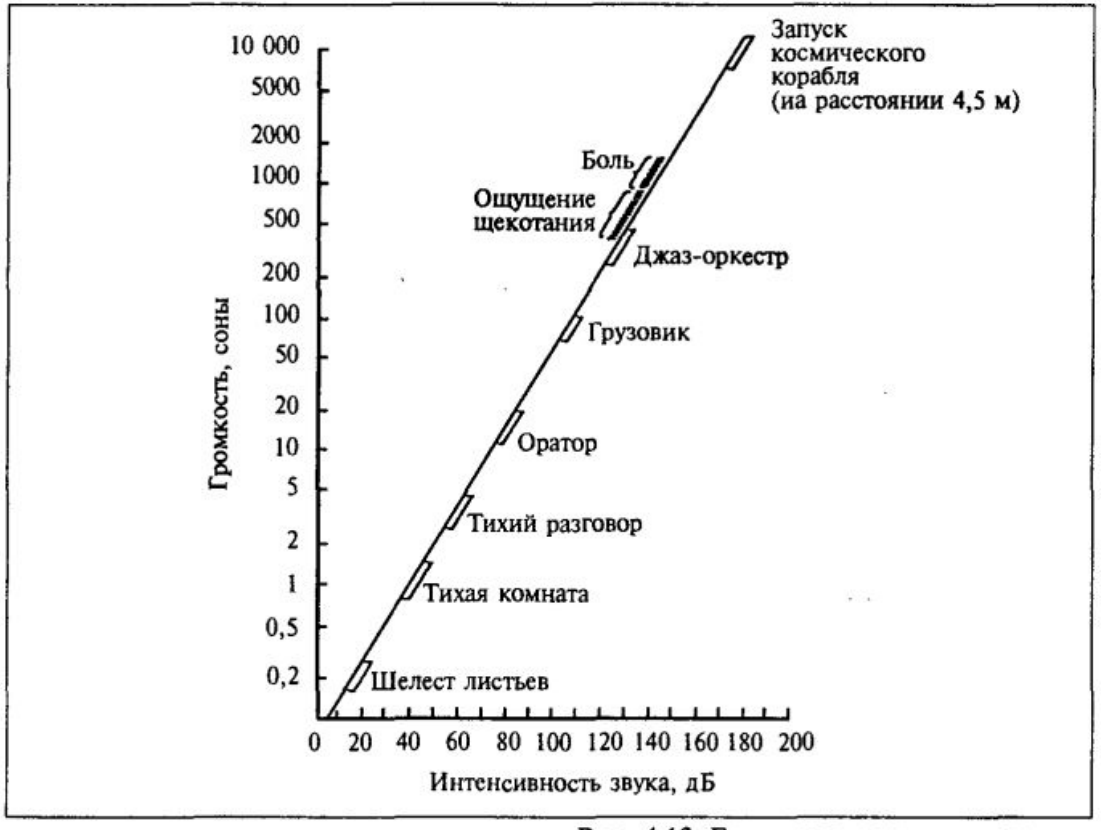


Рис. 4.12. Громкость сложных звуков в шкале сонов

Громкость по алгоритму GeMAPS

Loudness is used here as a more perceptually relevant [62] alternative to the signal energy. In order to approximate humans' non-linear perception of sound, an auditory spectrum as is applied in the Perceptual Linear Prediction (PLP) technique [63] is adopted. A non-linear Mel-band spectrum is constructed by applying 26 triangular filters distributed equidistant on the Mel-frequency scale from 20–8000 Hz to a power spectrum computed from a 25 ms frame. An auditory weighting with an equal loudness curve as used by [63] and originally adopted from [64] is performed. Next, a cubic root amplitude compression is performed for each band b of the equal loudness weighted Mel-band power spectrum [63]. resulting in a spectrum which is referred to as *auditory spectrum*. Loudness is then computed as the sum over all bands of the auditory spectrum.

The function $E(\omega)$ is an approximation to the nonequal sensitivity of human hearing at different frequencies (Robinson and Dadson, 1956) and simulates the sensitivity of hearing at about the 40-dB level. Our particular approximation is adopted from Makhoul and Cosell (1976) and is given by

$$E(\omega) = [(\omega^2 + 56.8\times 10^6)\omega^4]/[(\omega^2 + 6.3\times 10^6)^2 \times (\omega^2 + 0.38\times 10^9)]. \quad (7)$$

cy. For moderate sound levels, this approximation is reasonably good up to 5000 Hz. For applications requiring a higher Nyquist frequency, an additional term representing a rather steep (about -18 dB/oct) decrease of the sensitivity of hearing for frequencies higher than 5000 Hz might be found useful. Equation (7) would then become

$$E(\omega) = [(\omega^2 + 56.8\times 10^6)\omega^4]/[(\omega^2 + 6.3\times 10^6)^2 \times (\omega^2 + 0.38\times 10^9)(\omega^6 + 9.58\times 10^{26})]. \quad (7')$$

F. Eyben et al. (2015), The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing

H. Hermansky (1990), Perceptual linear predictive (PLP) analysis of speech // JASA

Спасибо за внимание!