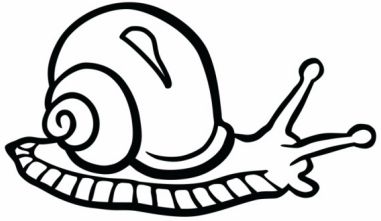
# Психоакустические шкалы

П. А. Холявин

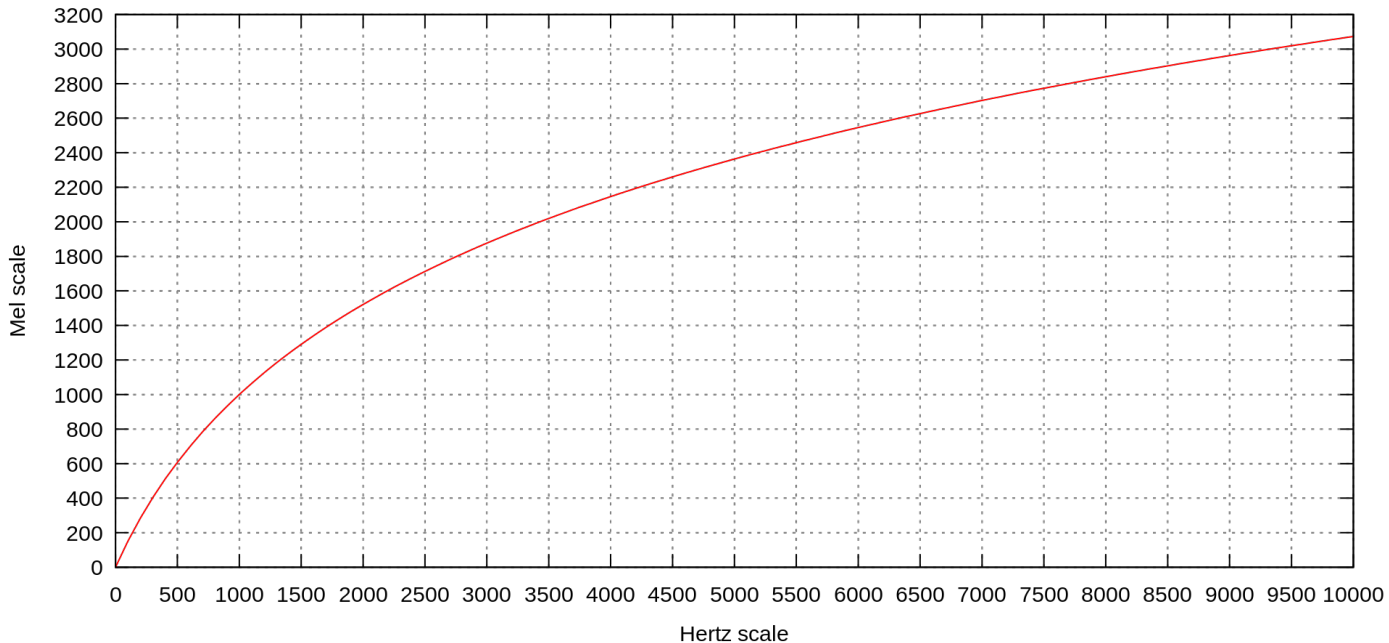
p.kholyavin@spbu.ru

26.09.2024



# Шкала мелов

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



# Шкала барков

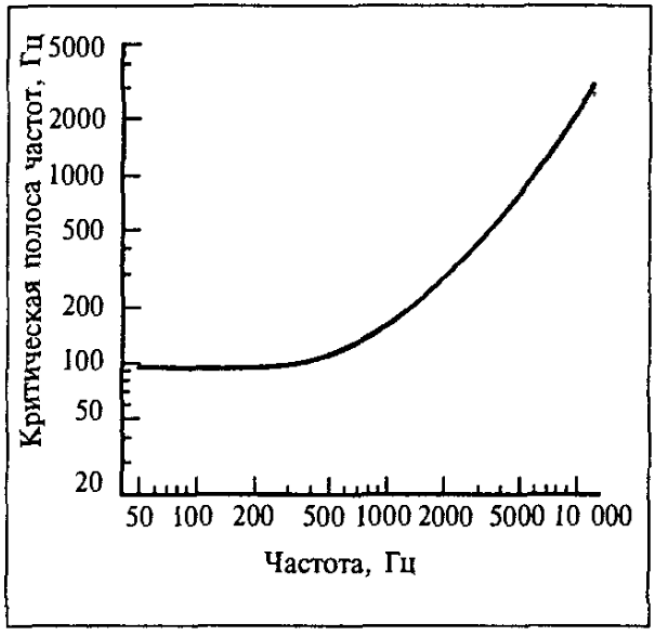
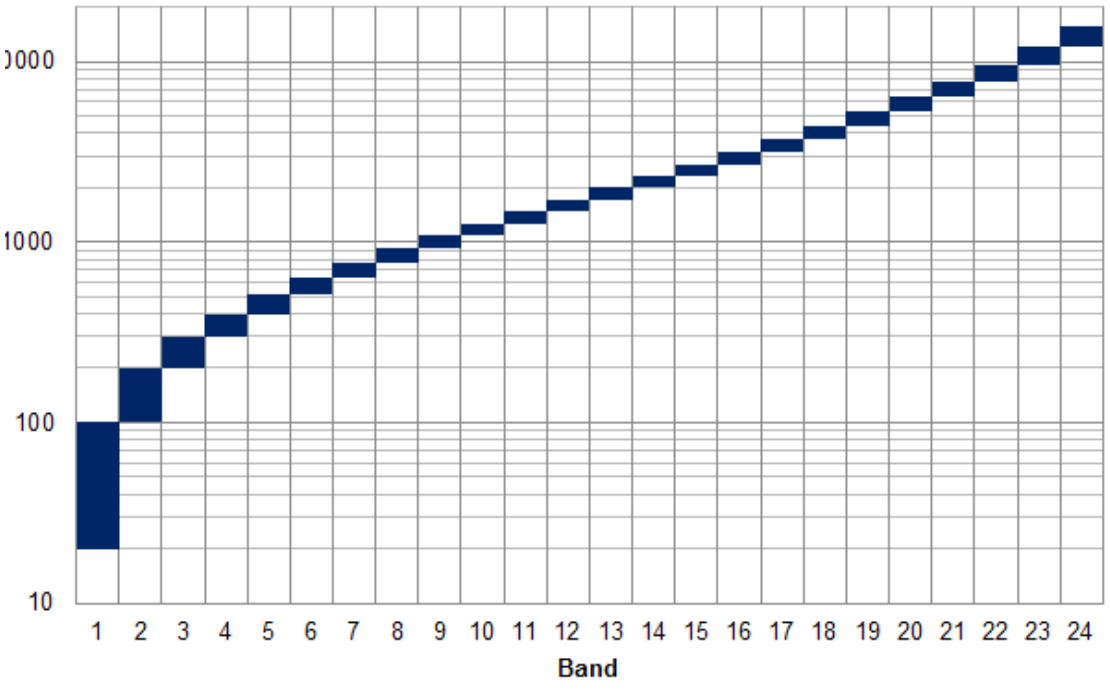



*Рис. 4.14.* Зависимость ширины критической полосы слуха от ее средней частоты

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



# Кривые равной громкости



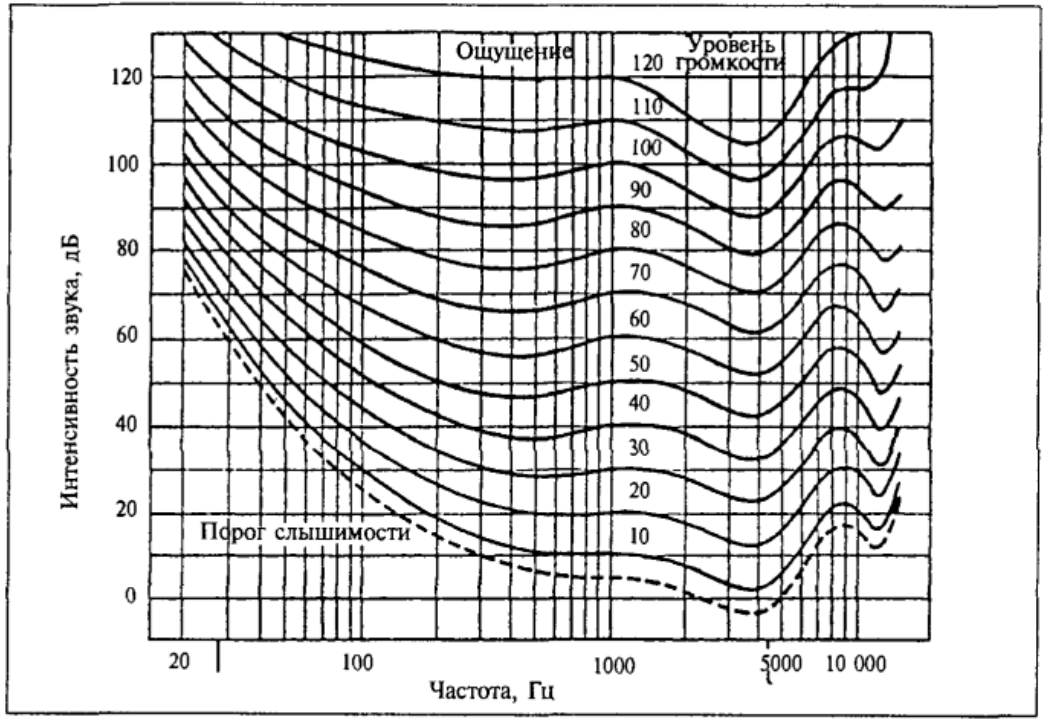



*Рис. 4.11.* Кривые равной громкости для тонов разных частот и интенсивностей

# Громкость

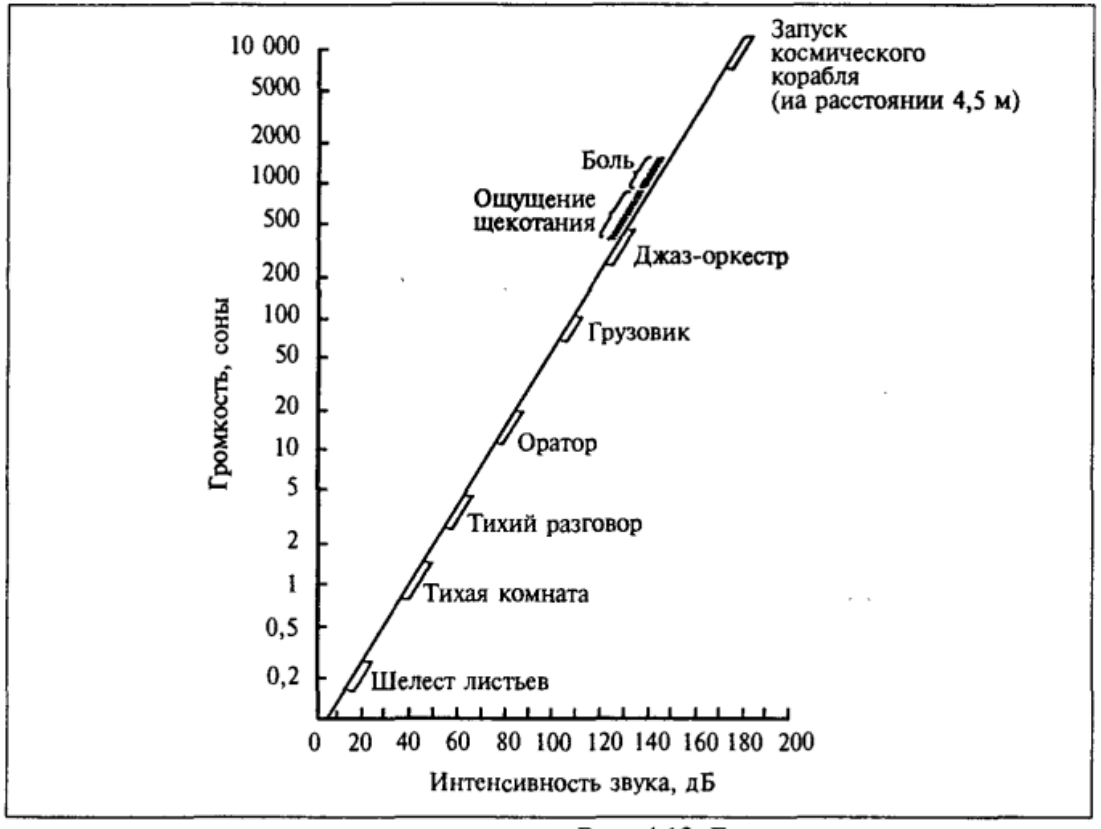



*Рис. 4.12.* Громкость сложных звуков в шкале сонов

# 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.

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

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

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

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

# Спасибо за внимание!