Psychoacoustic Model for Audio Compression

MUMT 621 – Presentation 2

Eto Sun

Volume Testing

min

max

* Audio from: ak+q - ikasu [Youtube]

Motivation

- lossy audio compression
  - remove components in frequency domain
  - encode signal with compact codes
- What does a lossy algorithm really do?
  - listen between lossless and lossy music

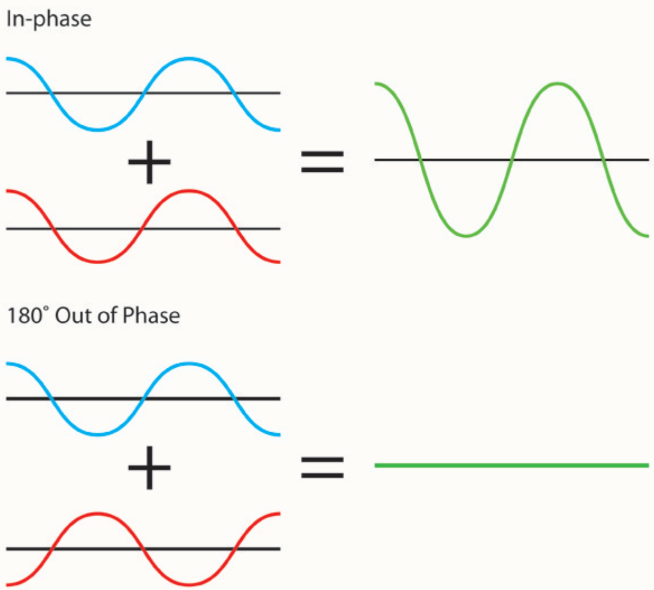
A

B

* Audio from: YOASOBI - Blue (「群青」English Ver.) Youtube

Motivation

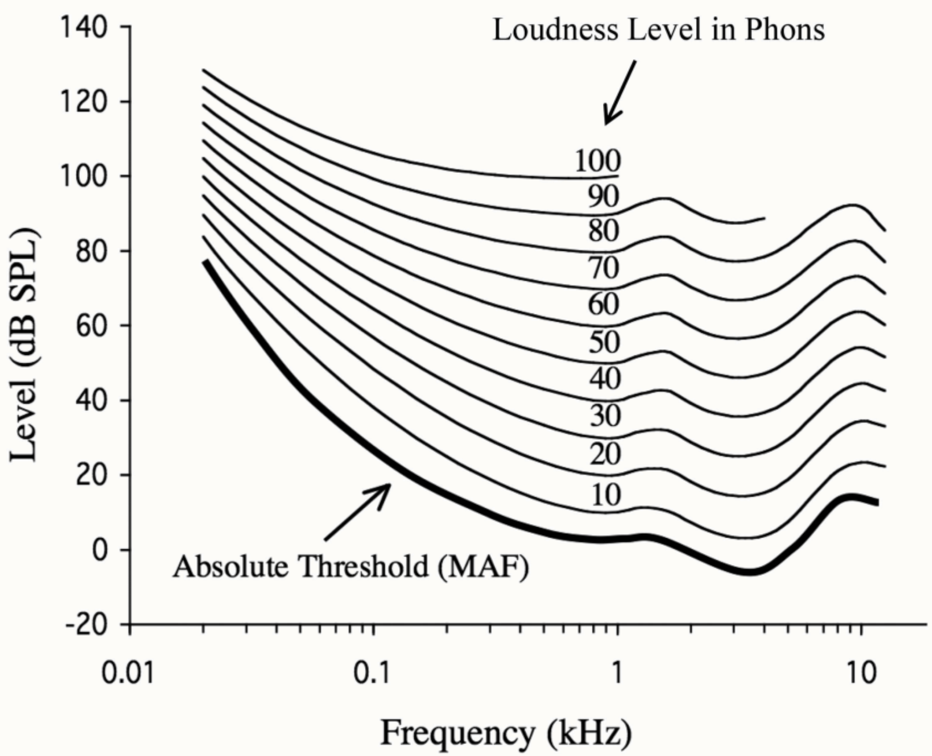
- **phase cancellation** between two versions
- find out which components the algorithm remove



* Figure from: https://music.stackexchange.com/a/66744

Absolute Hearing Threshold

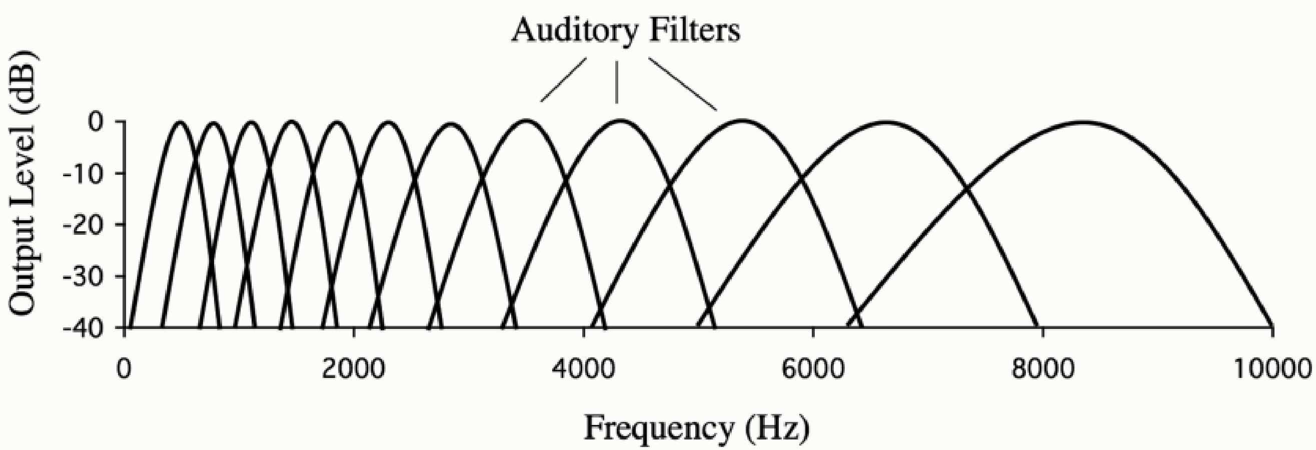
- the minimum required energy of a pure tone which can be detected by a listener



(Plack 2018)

Critical Bands

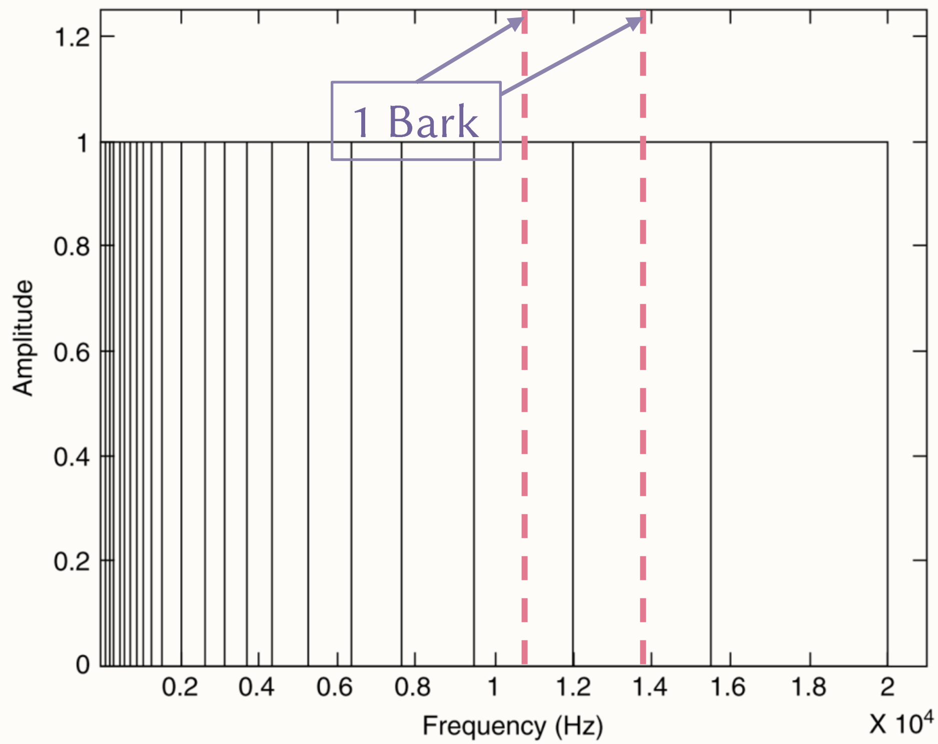
- our ears percept different frequency bands along the basilar membrane
- like **overlapping bandpass filters**
- narrow bandwidth for low frequency, wide for high frequency



(Plack 2018)

Critical Bands

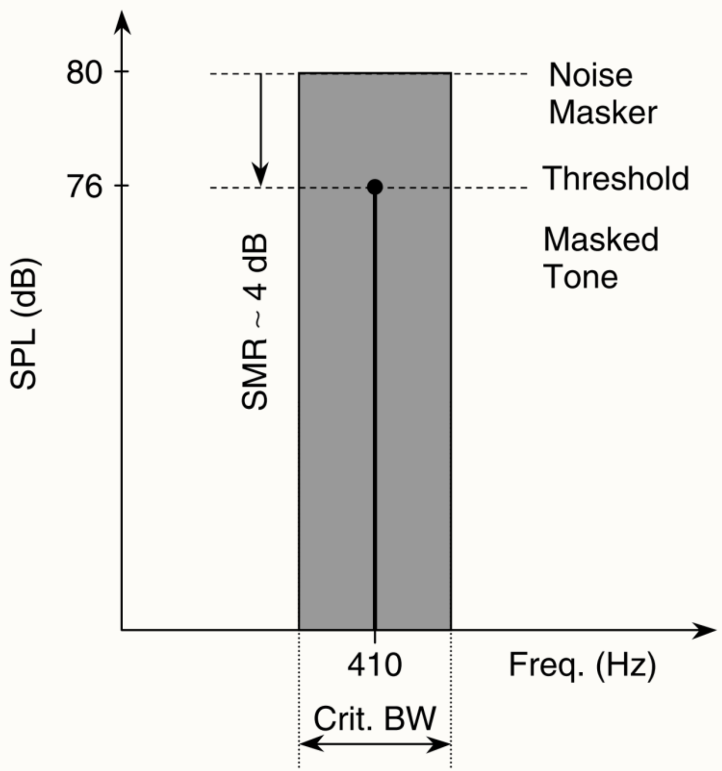
- critical bands: discriminate adjacent tones
- the distance of 1 critical band is referred to as one Bark

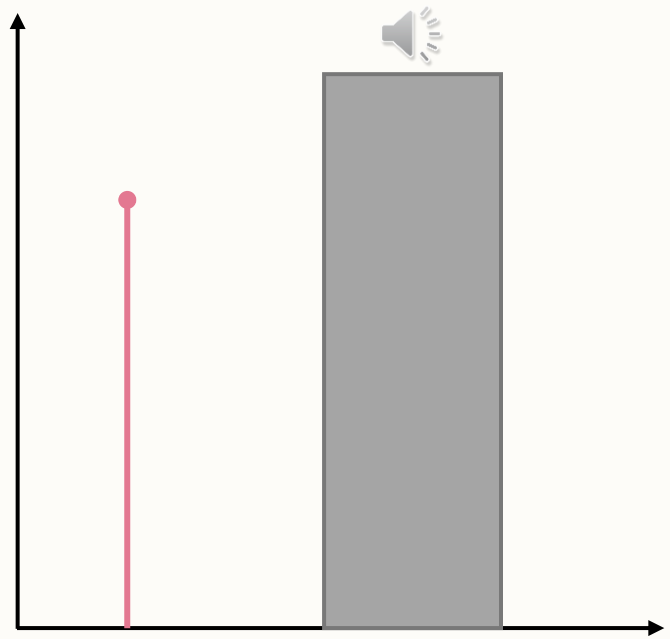


(Spanias, Painter, and Atto 2007)

Masking in Frequency Domain

- a narrow-band **noise** can **mask** a **tone** within the same critical band

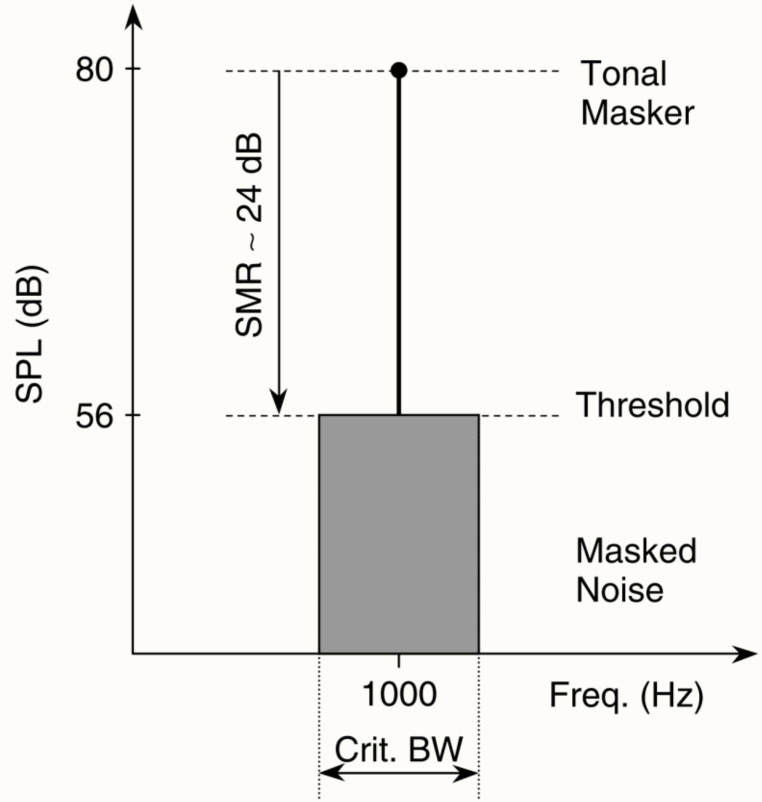




(Herre and Dick 2019; Spanias, Painter, and Atto 2007)

Masking in Frequency Domain

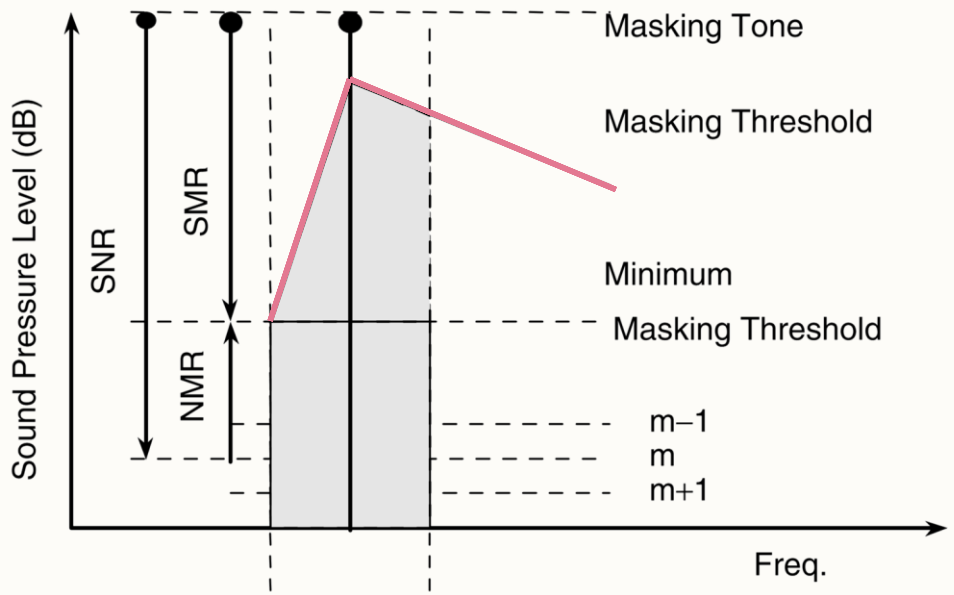
- a **tone** can **mask** a narrow-band **noise** within the same critical band
- **asymmetry**: harder for tone masking noise than noise masking tone



(Herre and Dick 2019; Spanias, Painter, and Atto 2007)

Frequency Spread of Masking

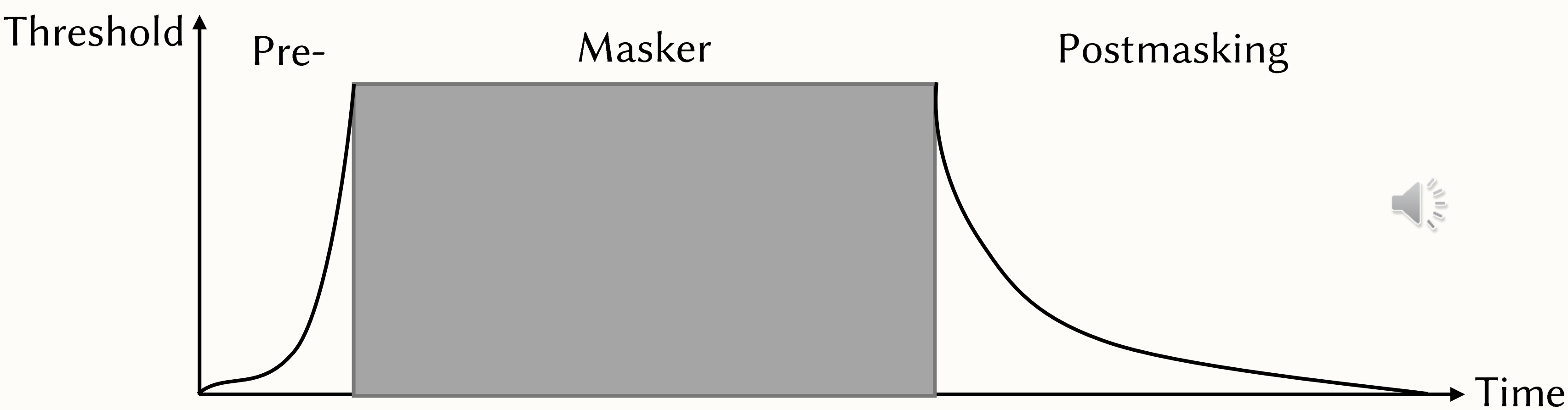
- one masker could effect on adjacent critical bands ⟶ **spread** of masking
- individual masking thresholds are combined to a global masking threshold



(Spanias, Painter, and Atto 2007)

Masking in Time Domain

- for masker and maskee in same critical band
- a strong sound could mask a weak sound before and after it
  - forward (post-) masking: less than 200 ms
  - backward (pre-) masking: less than 20 ms



(Herre and Dick 2019; Spanias, Painter, and Atto 2007)

Perceptual Entropy & Bit Allocation

- signal quantization
  - represent analog signal to digital values
  - less bits per sample $\longrightarrow$ more noise
    - 16 bits: signal-to-noise ratio (SNR) = 96 dB
    - 8 bits: SNR = 48 dB
    - 4 bits: SNR = 24 dB

- What if all quantization noise be masked?

* Audio source: Schubert - Symphony No. 8 In B Minor, D.759

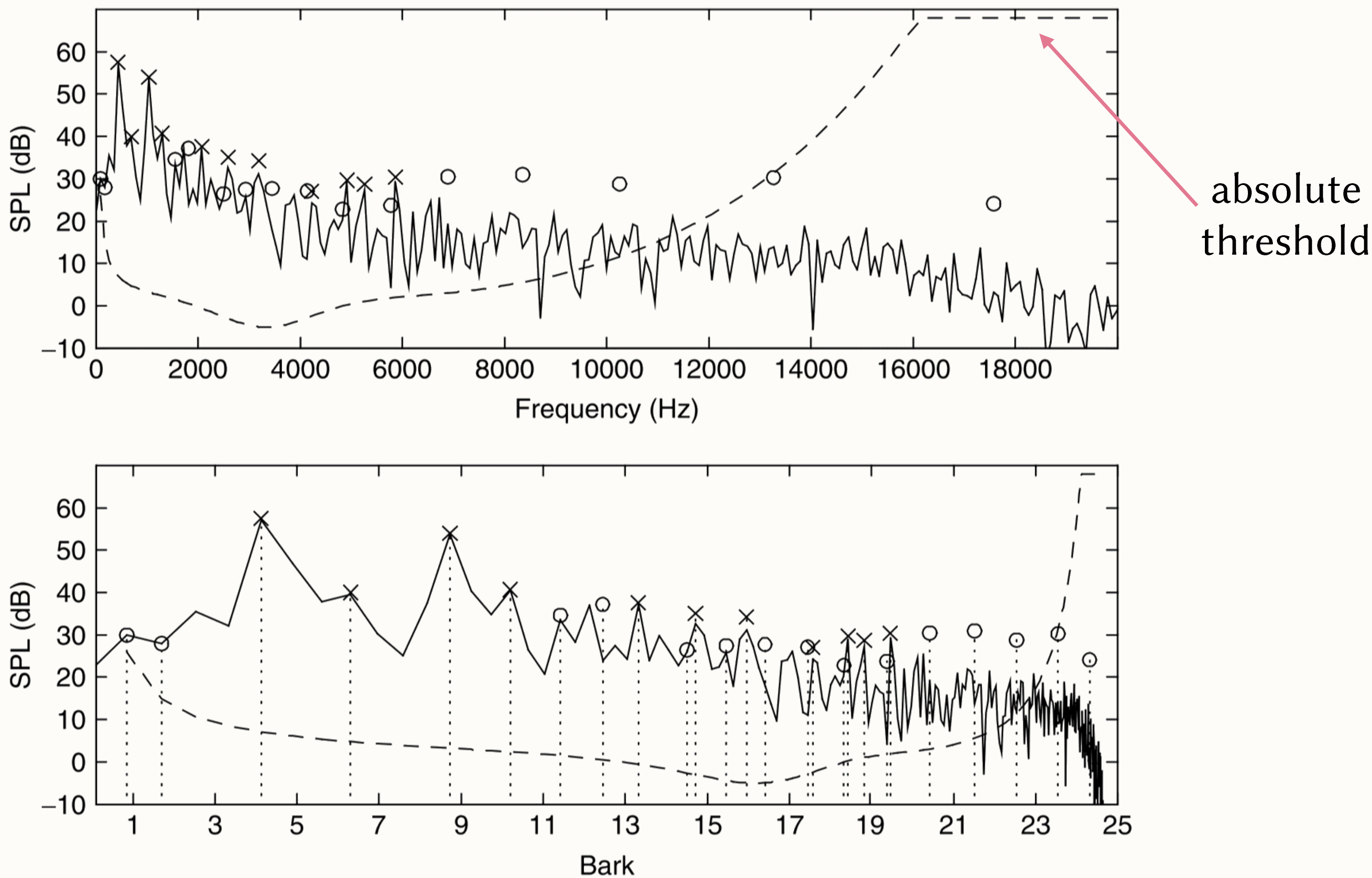
Perceptual Entropy & Bit Allocation

- perceptual entropy
  - estimate the number of bits required to quantize the spectrum without introducing perceptible noise

- perceptual bit allocation
  - low masking threshold $\longrightarrow$
    low noise threshold required $\longrightarrow$
    more bits allocated for this critical band

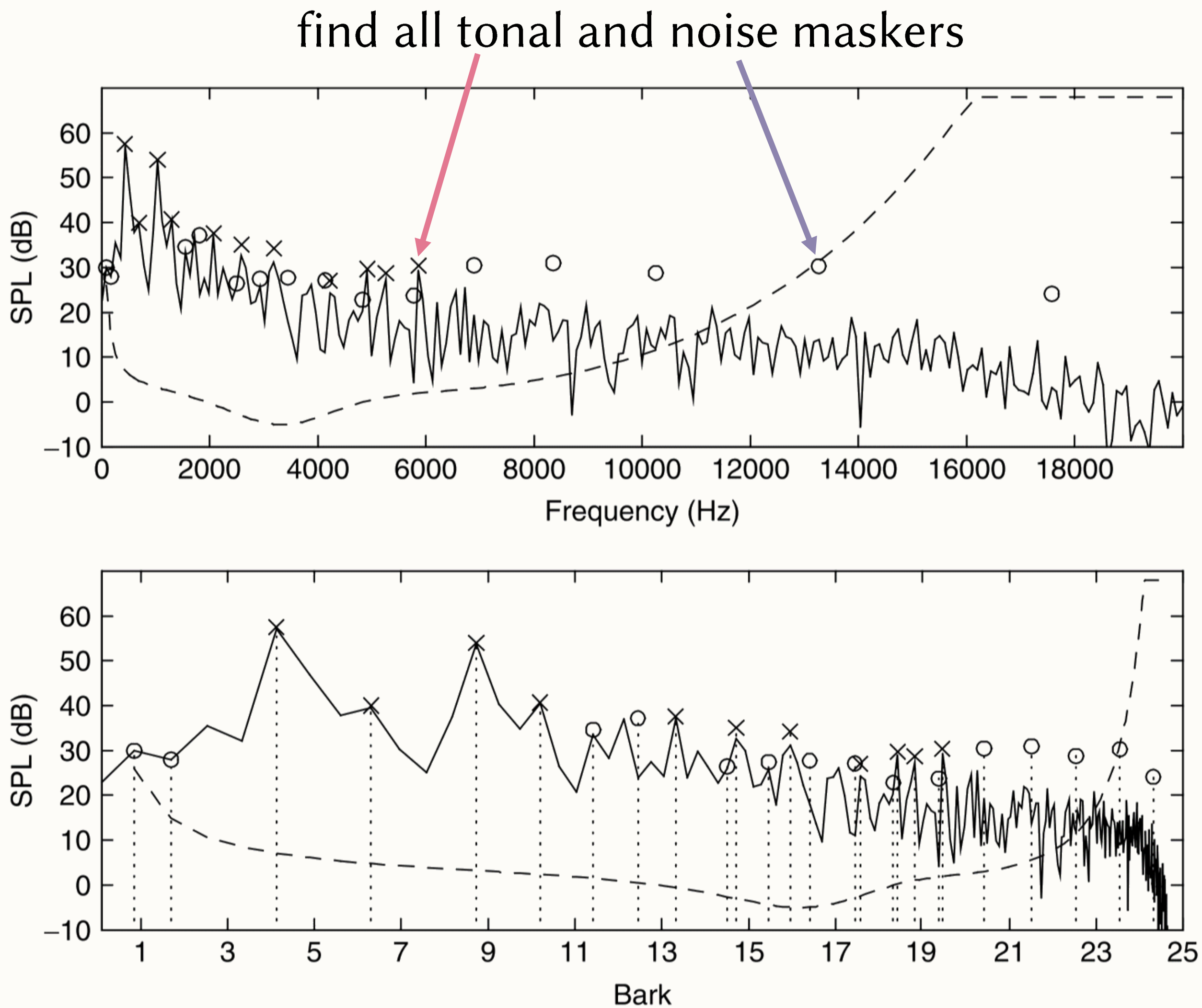
Example Model: MPEG-1, Layer 1

estimate the power spectral density (PSD) and SPL normalization



(Spanias, Painter, and Atto 2007)

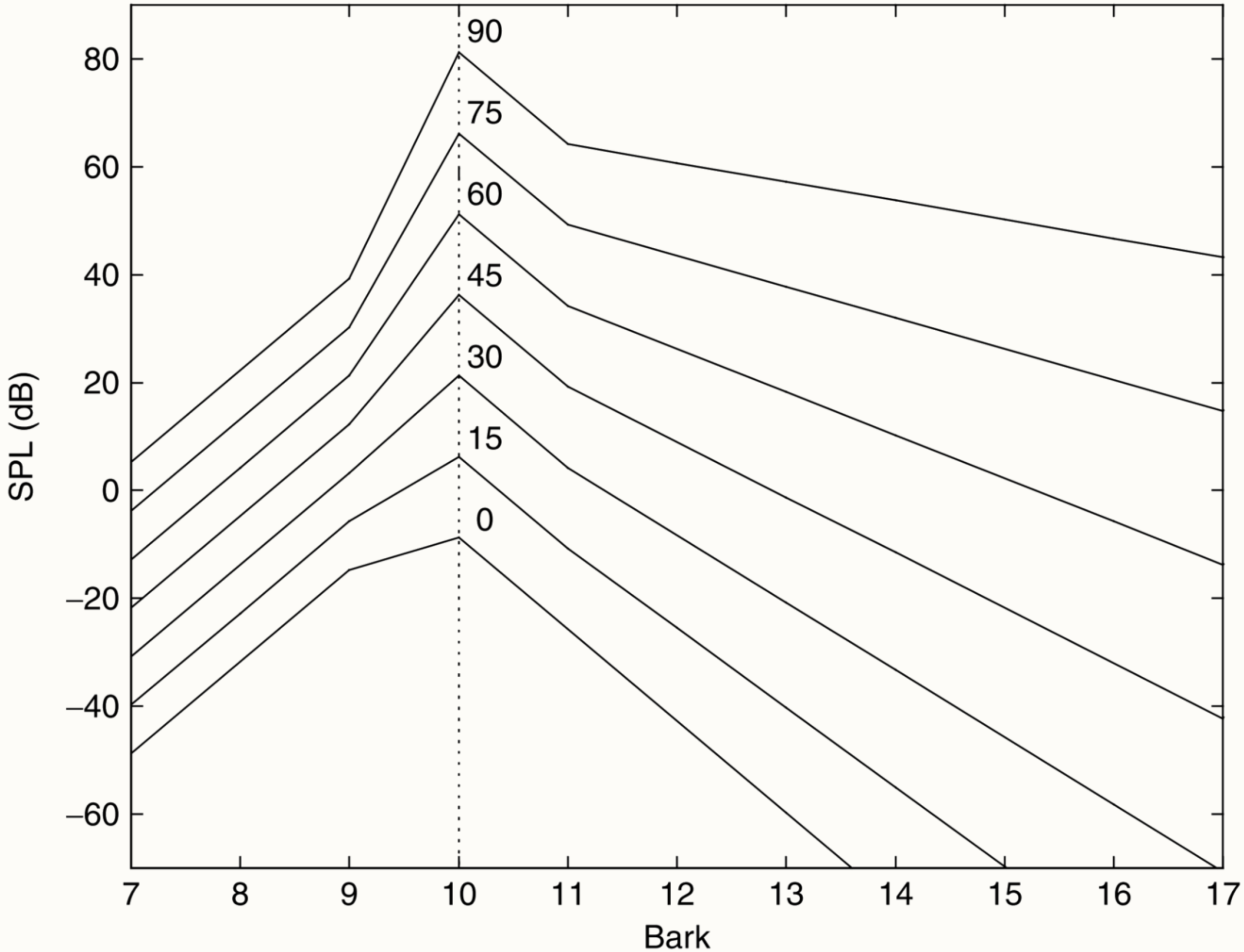
Example Model: MPEG-1, Layer 1



(Spanias, Painter, and Atto 2007)

Example Model: MPEG-1, Layer 1

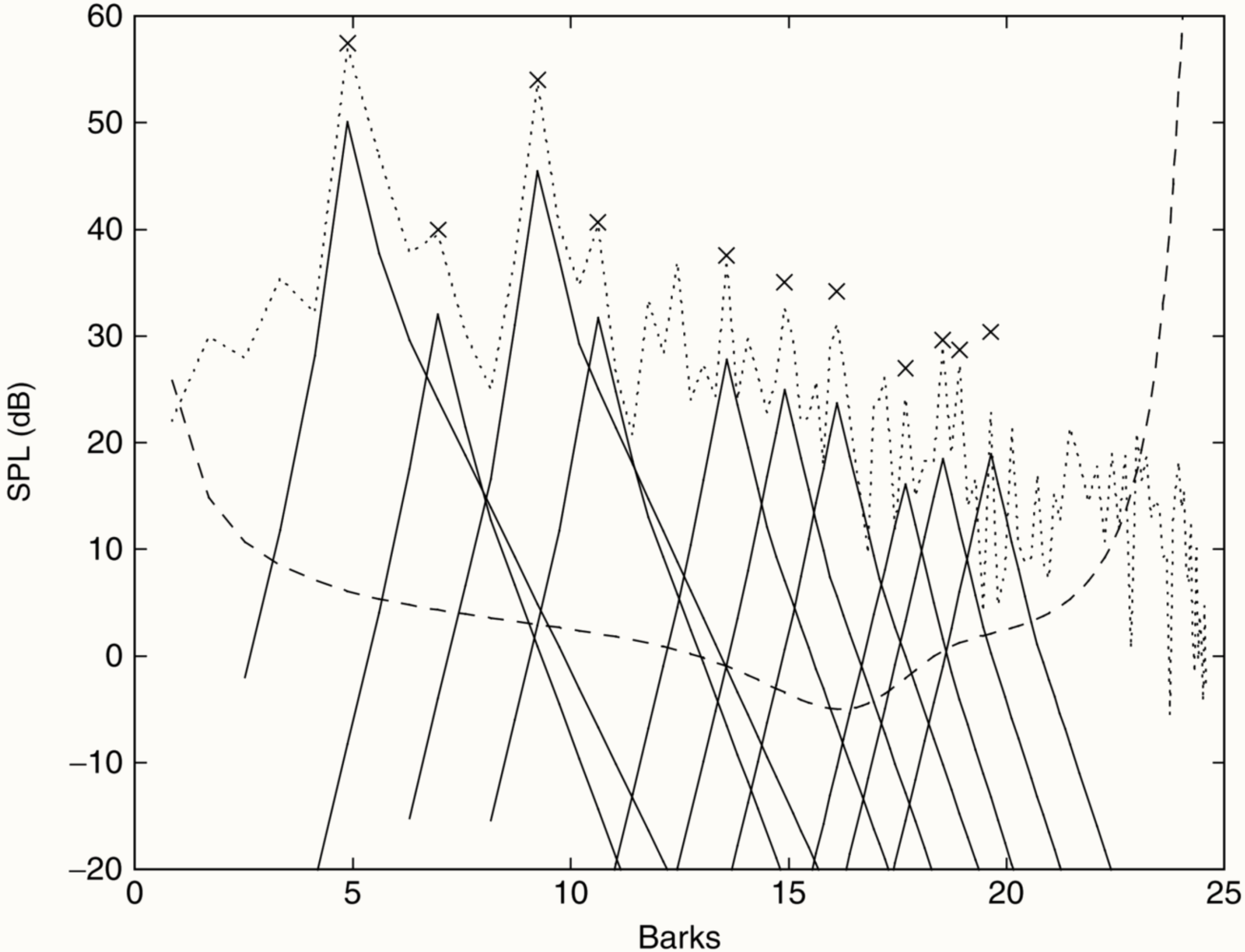
spreading functions in different level



(Spanias, Painter, and Atto 2007)

Example Model: MPEG-1, Layer 1

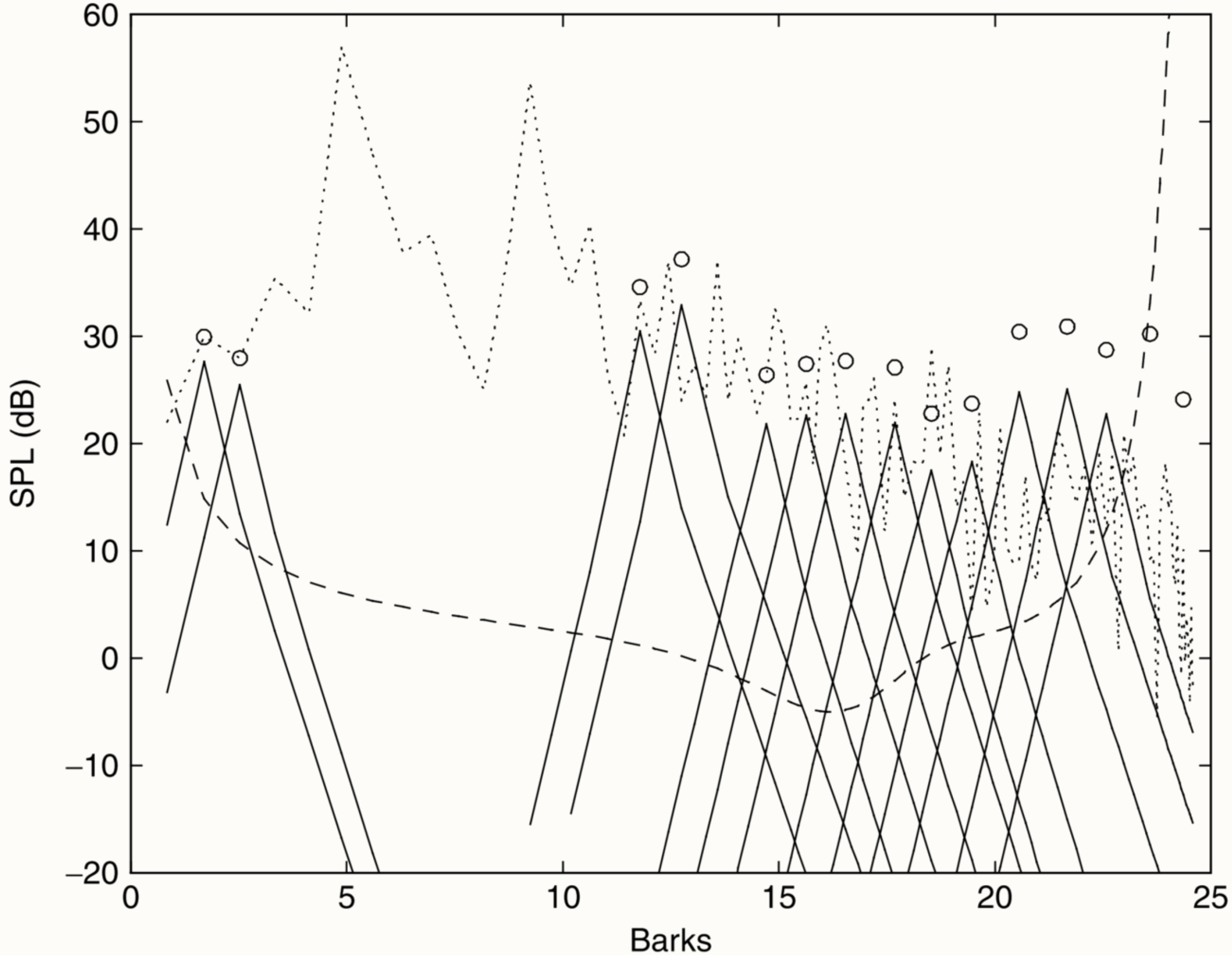
apply spreading functions for tonal maskers



(Spanias, Painter, and Atto 2007)

Example Model: MPEG-1, Layer 1

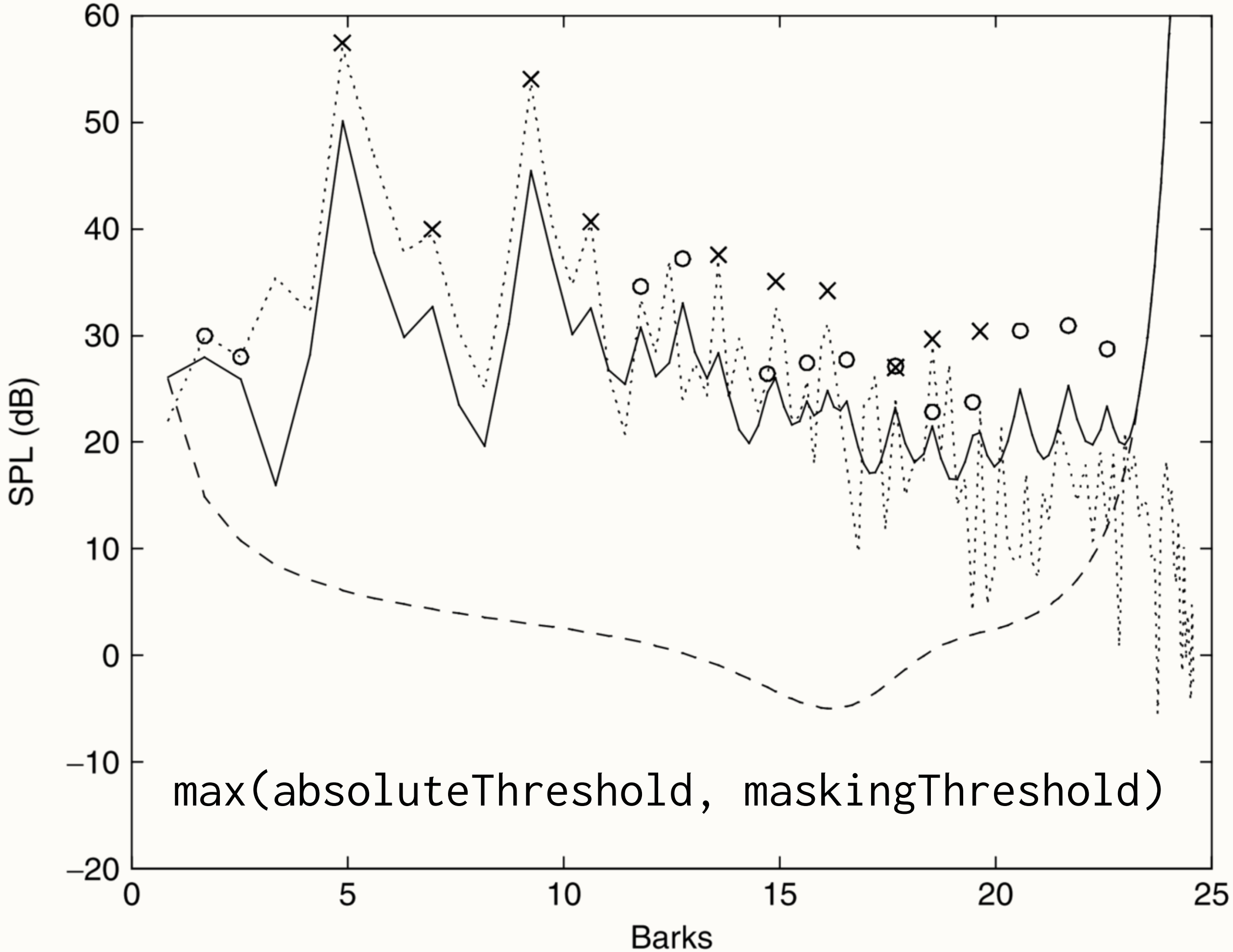
apply spreading functions for noise maskers



(Spanias, Painter, and Atto 2007)

Example Model: MPEG-1, Layer 1

obtain the global threshold



(Spanias, Painter, and Atto 2007)

Summary

- hearing threshold
- critical bands
- masking
  - frequency domain (simultaneous)
  - time domain (non-simultaneous)
- compression in quantization
  - perceptual entropy
  - perceptual bit allocation
- example: MPEG-1, layer 1 perceptual model

References

Herre, Jürgen, and Sascha Dick. 2019. “Psychoacoustic Models for Perceptual Audio Coding—A Tutorial Review.” *Applied Sciences* 9 (14): 2854.

Plack, Christopher J. 2018. “Loudness and Intensity Coding.” In *The Sense of Hearing*, Third edition, 110–27. New York: Routledge.

Spanias, Andreas, Ted Painter, and Venkatraman Atto. 2007. “Psychoacoustic Principles.” In *Audio Signal Processing and Coding*, 113–44. Hoboken, New Jersey: Wiley-Interscience.

Annotated Bibliography

MUMT 621 – Assignment 2

Eto Sun

Herre, Jürgen, and Sascha Dick. 2019. "Psychoacoustic Models for Perceptual Audio Coding—A Tutorial Review." *Applied Sciences* 9 (14): 2854.
This paper reviews and provides a tutorial overview of commonly used psychoacoustic models for perceptual audio coding. Introducing the spectral and temporal masking effects, models for perceptual audio coding, models for multichannel audio coding, and models for parametric audio coding. In addition, some recent models are reviewed and discussed. The researchers describe different masking effects with illustrations, but no illustrations for classic models (such as the MPEG-1 model). This paper hides the mathematical details for perceptual coding and explains them in sequential order, thus is easy to understand. Moreover, a detailed list of abbreviations is placed at the end of the article for reference.

Plack, Christopher J. 2018. "Loudness and Intensity Coding." In *The Sense of Hearing*, Third edition, 110–27. New York: Routledge.
This chapter of the book introduces the analysis process and the representation of physical sound intensity in the auditory system. The author provides an overview of hearing range, loudness, representation of sound intensity in the auditory system, and intensity comparisons across frequency and across time. For psychoacoustic models, the absolute threshold is introduced. The loudness pattern of noise with different bandwidths is related to critical bands, which is utilized in various perceptual compression.

Spanias, Andreas, Ted Painter, and Venkatraman Atto. 2007. "Psychoacoustic Principles." In *Audio Signal Processing and Coding*, 113–44. Hoboken, New Jersey: Wiley-Interscience.
This chapter of the book covers the fundamental principles for building psychoacoustic models, including the absolute threshold of hearing, the Bark scale, the critical bands, two kinds of masking effects, the perceptual entropy, and the perceptual bit allocation. A detailed step-by-step procedure for MPEG-1 psychoacoustic model 1 is also provided. Compared to Herre and Dick (2019)'s article, this article describes the preliminary knowledge for psychoacoustic models in great details, with both mathematical representations and illustrations.