



Introduction to **Audio Content Analysis**

module 8.0: intensity

alexander lerch

introduction

overview

corresponding textbook section

chapter 8

■ lecture content

- quick overview: human perception of loudness
- intensity related features

■ learning objectives

- discuss level and loudness
- list and describe typical intensity related low level features



introduction

overview

corresponding textbook section

chapter 8

■ lecture content

- quick overview: human perception of loudness
- intensity related features

■ learning objectives

- discuss level and loudness
- list and describe typical intensity related low level features

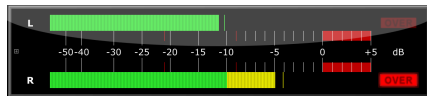
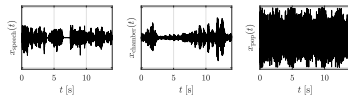


intensity, magnitude & loudness

introduction

■ intensity-related descriptors **commonly used**

- waveform view
- level monitoring (PPM, VU, ...)



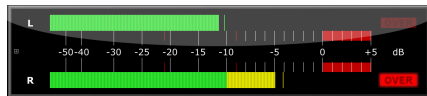
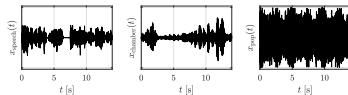
related terms: magnitude • intensity • envelope • level • volume • velocity • loudness

intensity, magnitude & loudness

introduction

■ intensity-related descriptors **commonly used**

- waveform view
- level monitoring (PPM, VU, ...)



related terms: magnitude • intensity • envelope • level • volume • velocity • loudness

intensity, magnitude & loudness

human perception 1/2

perception has non-linear relation to magnitude/RMS:

- model: logarithmic relation

$$v_{\text{dB}}(n) = 20 \cdot \log_{10} \left(\frac{v(n)}{v_0} \right)$$

- v_0 : reference constant (0 dB point)
 - ▶ digital: $v_0 = 1 \Rightarrow \text{dBFS}$
 - ▶ sound pressure $v_0 = 20 \cdot 10^{-6} \Rightarrow \text{dBSPL}$
- scaling factor: $1 \text{ dB} \approx \text{JNDL}$ for sound pressure level

intensity, magnitude & loudness

side note: level computation

if $v(n) = 0 \Rightarrow$: computation of $\log_{10}(0)$

■ work-arounds

a add constant ϵ

$$v_{\text{dB}}(n) = 20 \cdot \log_{10}(v(n) + \epsilon)$$

b add if statement

$$v_{\text{trunc}}(n) = \begin{cases} v(n), & \text{if } v(n) \geq \epsilon \\ \epsilon, & \text{otherwise} \end{cases}$$

intensity, magnitude & loudness

side note: level computation

if $v(n) = 0 \Rightarrow$: computation of $\log_{10}(0)$

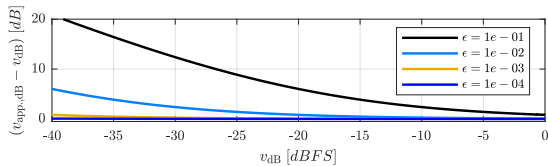
work-arounds

a add constant ϵ

$$v_{\text{dB}}(n) = 20 \cdot \log_{10}(v(n) + \epsilon)$$

b add if statement

$$v_{\text{trunc}}(n) = \begin{cases} v(n), & \text{if } v(n) \geq \epsilon \\ \epsilon, & \text{otherwise} \end{cases}$$



intensity, magnitude & loudness

side note: level computation

if $v(n) = 0 \Rightarrow$: computation of $\log_{10}(0)$

■ work-arounds

a add constant ϵ

$$v_{\text{dB}}(n) = 20 \cdot \log_{10}(v(n) + \epsilon)$$

b add **if** statement

$$v_{\text{trunc}}(n) = \begin{cases} v(n), & \text{if } v(n) \geq \epsilon \\ \epsilon, & \text{otherwise} \end{cases}$$

intensity, magnitude & loudness

human perception 2/2

- decibel scale is *not* loudness scale:
 - equal-sized steps on the decibel scale not perceived as equal-sized loudness steps
- perceptual phenomenon loudness depends on
 - frequency
 - cochlear resolution
 - masking effects

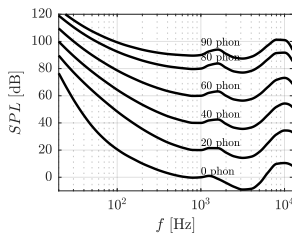
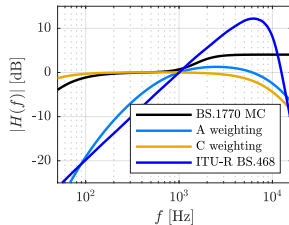
intensity, magnitude & loudness

human perception 2/2

- decibel scale is *not* loudness scale:
 - equal-sized steps on the decibel scale not perceived as equal-sized loudness steps
- perceptual phenomenon loudness depends on
 - frequency
 - cochlear resolution
 - masking effects

intensity, magnitude & loudness

human perception 2/2



intensity, magnitude & loudness

dynamics in music

■ score:

- only several rough dynamic steps, e.g.:
pp, p, mf, f, ff
- comparably vague instructions on volume modifications, e.g.:
crescendo, decrescendo, sf
- dynamics influenced by
 - ▶ instrumentation
 - ▶ timbre
 - ▶ number of voices
 - ▶ context and musical tension

■ MIDI:

- 128 velocity steps
- no standardized relation to magnitude, power, ...

intensity, magnitude & loudness

dynamics in music

■ score:

- only several rough dynamic steps, e.g.:
pp, p, mf, f, ff
- comparably vague instructions on volume modifications, e.g.:
crescendo, decrescendo, sf
- dynamics influenced by
 - ▶ instrumentation
 - ▶ timbre
 - ▶ number of voices
 - ▶ context and musical tension

■ MIDI:

- 128 velocity steps
- no standardized relation to magnitude, power, ...

intensity, magnitude & loudness

features: root mean square 1/2

$$v_{\text{RMS}}(n) = \sqrt{\frac{1}{\mathcal{K}} \sum_{i=i_s(n)}^{i_e(n)} x(i)^2}$$

intensity, magnitude & loudness

features: root mean square 1/2

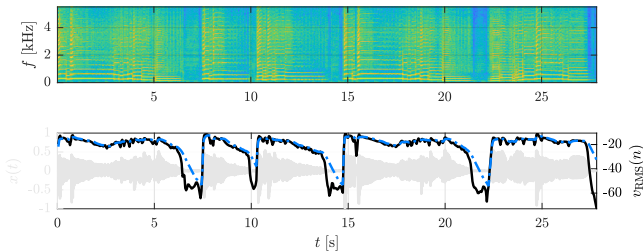
$$v_{\text{RMS}}(n) = \sqrt{\frac{1}{\mathcal{K}} \sum_{i=i_s(n)}^{i_e(n)} x(i)^2}$$

- value of this feature for the hypothetical prototype signals
 - silence
 - sinusoidal (Amplitude A)

intensity, magnitude & loudness

features: root mean square 1/2

$$v_{\text{RMS}}(n) = \sqrt{\frac{1}{K} \sum_{i=i_s(n)}^{i_e(n)} x(i)^2}$$



intensity, magnitude & loudness

features: root mean square 2/2

common variants (sample processing only):

- reduce computational complexity

$$v_{\text{RMS}}^2(n) = \frac{x(i_e(n))^2 - x(i_s(n-1))^2}{i_e(n) - i_s(n) + 1} + v_{\text{RMS}}^2(n-1)$$

$$v_{\text{RMS}}(n) = \sqrt{v_{\text{RMS}}^2(n)}$$

- single pole approximation

$$v_{\text{tmp}}(i) = \alpha \cdot v_{\text{tmp}}(i-1) + (1 - \alpha) \cdot x(i)^2$$

$$v_{\text{RMS}}^*(i) = \sqrt{v_{\text{tmp}}(i)}$$

intensity, magnitude & loudness

features: root mean square 2/2

common variants (sample processing only):

- reduce computational complexity

$$v_{\text{RMS}}^2(n) = \frac{x(i_e(n))^2 - x(i_s(n-1))^2}{i_e(n) - i_s(n) + 1} + v_{\text{RMS}}^2(n-1)$$

$$v_{\text{RMS}}(n) = \sqrt{v_{\text{RMS}}^2(n)}$$

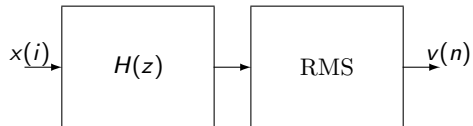
- single pole approximation

$$v_{\text{tmp}}(i) = \alpha \cdot v_{\text{tmp}}(i-1) + (1 - \alpha) \cdot x(i)^2$$

$$v_{\text{RMS}}^*(i) = \sqrt{v_{\text{tmp}}(i)}$$

intensity, magnitude & loudness

features: weighted root mean square

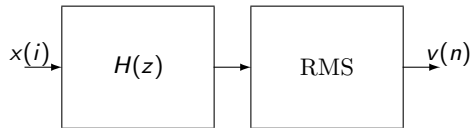


$H(z)$:

- A, B, C weighting
- RLB (BS.1770)
- ...

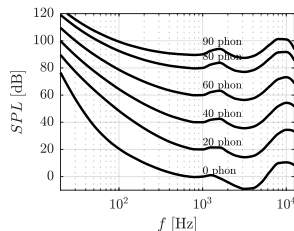
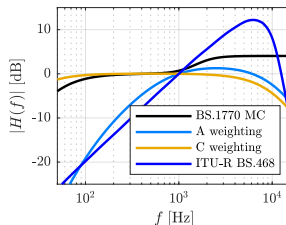
intensity, magnitude & loudness

features: weighted root mean square



$H(z)$:

- A, B, C weighting
- RLB (BS.1770)
- ...



intensity, magnitude & loudness

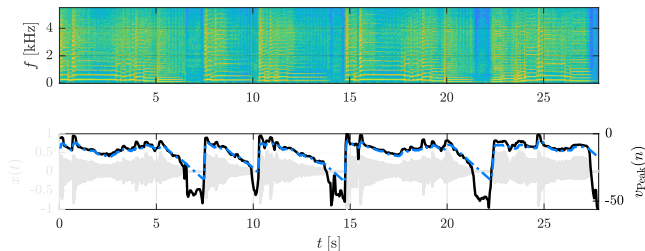
features: peak envelope (max)

$$v_{\text{Peak}}(n) = \max_{i_s(n) \leq i \leq i_e(n)} |x(i)|$$

intensity, magnitude & loudness

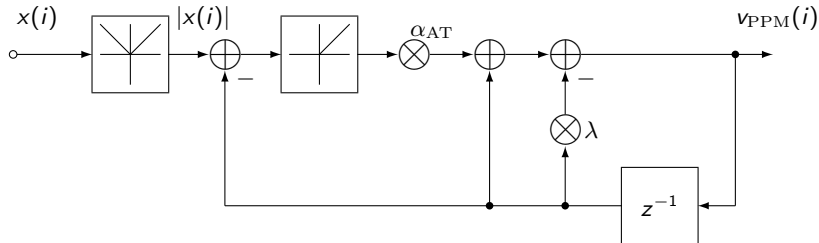
features: peak envelope (max)

$$v_{\text{Peak}}(n) = \max_{i_s(n) \leq i \leq i_e(n)} |x(i)|$$



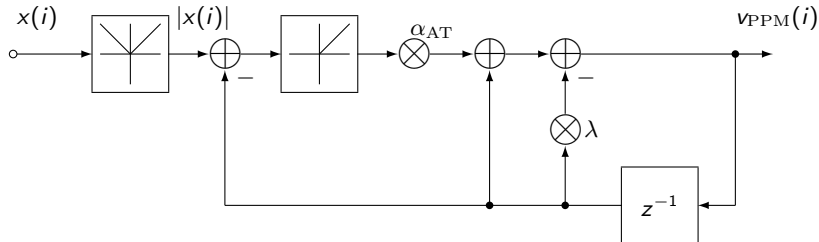
intensity, magnitude & loudness

features: peak envelope (PPM) 1/2



intensity, magnitude & loudness

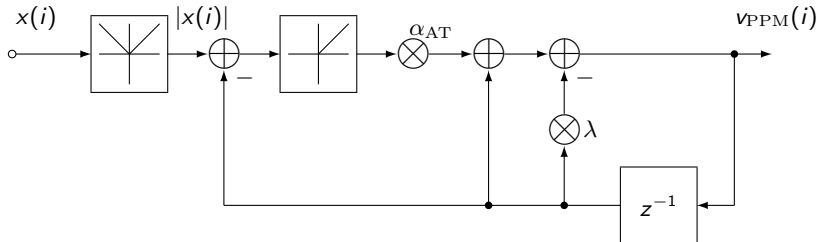
features: peak envelope (PPM) 1/2



■ **release state** ($|x(i)| < v_{PPM}(i-1) \Rightarrow \lambda = \alpha_{RT}$)

intensity, magnitude & loudness

features: peak envelope (PPM) 1/2

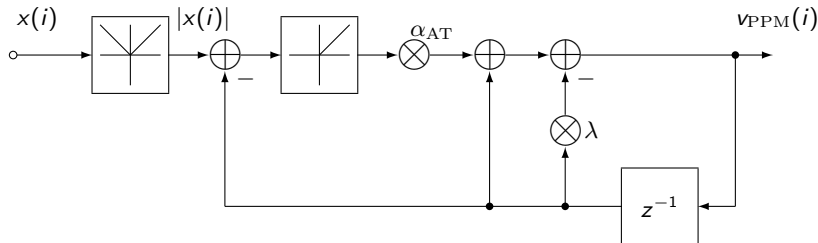


■ **release state** ($|x(i)| < v_{PPM}(i-1) \Rightarrow \lambda = \alpha_{RT}$)

$$\begin{aligned} v_{PPM}(i) &= v_{PPM}(i-1) - \alpha_{RT} \cdot v_{PPM}(i-1) \\ &= (1 - \alpha_{RT}) \cdot v_{PPM}(i-1) \end{aligned}$$

intensity, magnitude & loudness

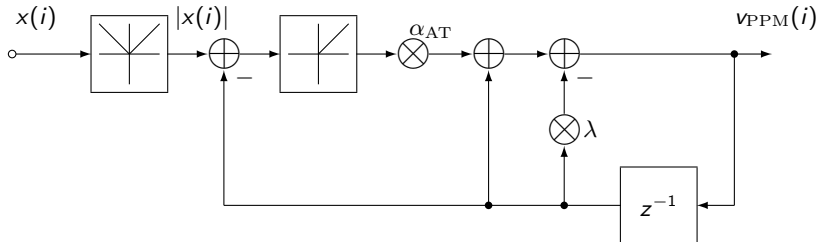
features: peak envelope (PPM) 1/2



- **attack state** ($|x(i)| \geq v_{PPM}(i-1) \Rightarrow \lambda = 0$)

intensity, magnitude & loudness

features: peak envelope (PPM) 1/2

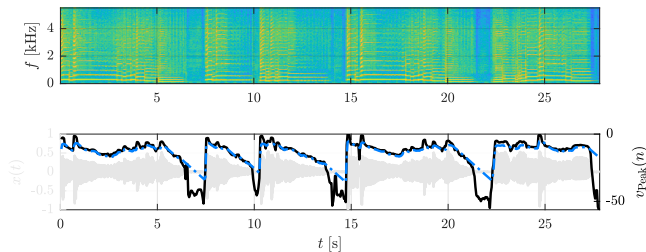


■ **attack state** ($|x(i)| \geq v_{\text{PPM}}(i-1) \Rightarrow \lambda = 0$)

$$\begin{aligned} v_{\text{PPM}}(i) &= \alpha_{\text{AT}} \cdot (|x(i)| - v_{\text{PPM}}(i-1)) + v_{\text{PPM}}(i-1) \\ &= \alpha_{\text{AT}} \cdot |x(i)| + (1 - \alpha_{\text{AT}}) \cdot v_{\text{PPM}}(i-1) \end{aligned}$$

intensity, magnitude & loudness

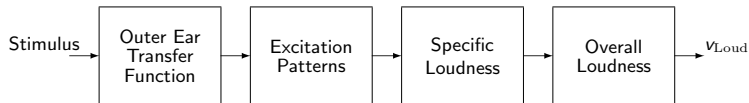
features: peak envelope (PPM) 2/2



- gold: max per block
- blue: PPM

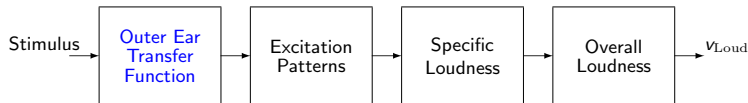
intensity, magnitude & loudness

features: zwicker loudness

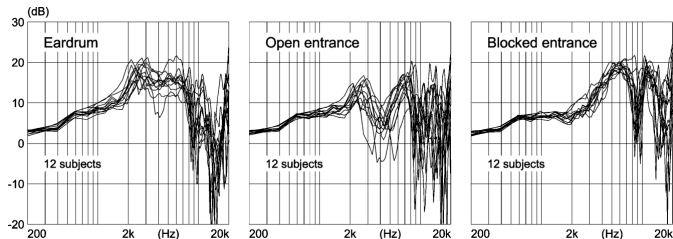


intensity, magnitude & loudness

features: zwicker loudness



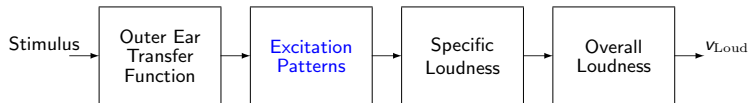
■ outer ear transfer function¹



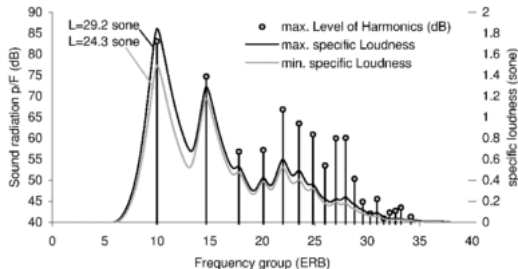
¹D. Hammershøi and H. Møller, "Methods for Binaural Recording and Reproduction," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 303–311, May 2002.

intensity, magnitude & loudness

features: zwicker loudness



■ excitation patterns¹

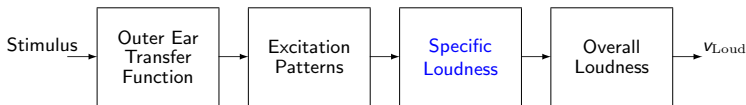


¹M. Schleske, *Vibrato of the musician*, [Online]. Available:

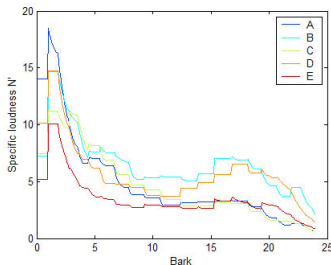
<http://www.schleske.de/en/our-research/handbook-violinacoustics/vibrato-of-the-musician.html> (visited on 07/29/2015).

intensity, magnitude & loudness

features: zwicker loudness



■ specific loudness¹

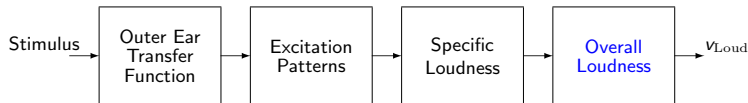


¹U. of Salford, *Customised metrics*, [Online]. Available:

<https://www.salford.ac.uk/computing-science-engineering/research/acoustics/psychoacoustics/sound-quality-making-products-sound-better/sound-quality-testing/customised-metrics> (visited on 07/29/2015).

intensity, magnitude & loudness

features: zwicker loudness



■ overall loudness

$$v_{\text{loud}} = \sum_{\forall i} z_i$$

intensity, magnitude & loudness

derived features

- number or ratio of pauses
- dynamic range
- other statistical features from (RMS) histogram
- . . .

summary

lecture content

■ loudness perception

- nonlinear relation to magnitude or power
- depends also on frequency, level, and signal (masking)

■ typical features

- derived from envelope (peak, RMS, weighted RMS)
- derived from histogram (range, mode)

