Introduction to **Audio Content Analysis**

module 8.0: intensity

alexander lerch

Georgia Tech Center for Music Technology
College of Design

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

corresponding textbook section

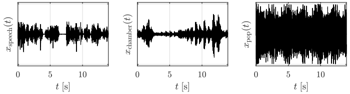
chapter 8

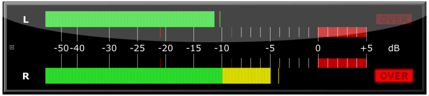
- **lecture content**
  - quick overview: human perception of loudness
  - intensity related features

- **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,...)



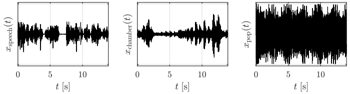


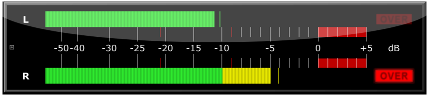
related terms: magnitude • intensity • envelope • level • volume • velocity • loudness

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

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

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

intensity, magnitude & loudness

side note: level computation

Georgia Tech | Center for Music Technology, College of Design

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

- **work-arounds**
  a. add constant ϵ

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

  b. add **if** statement

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

intensity, magnitude & loudness

side note: level computation

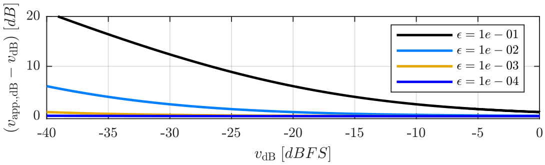
if $v(n) = 0 \quad \Rightarrow$: computation of $\log_{10}(0)$

- **work-arounds**
  - a. add constant ϵ

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

  - b. add **if** statement

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



matlab source: plotLogEpsilon.m

intensity, magnitude & loudness

side note: level computation

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

- **work-arounds**
  - a) add constant ϵ

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

  - b) add **if** statement

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

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

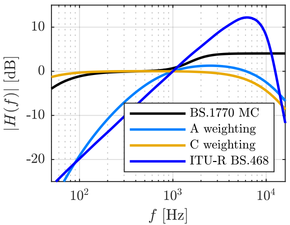
human perception 2/2

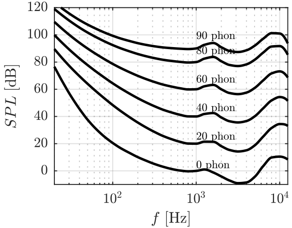
- decibel scale is *not* loudness scale:
  - equal-sized steps on the decibel scale not perceived as equal-sized loudness steps

- perceptual phenomenon loudness depends on
  - frequency
  - cochlear resolution
  - masking effects

intensity, magnitude & loudness

human perception 2/2





matlab source: plotLoudnessWeighting.m

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

$$v_\mathrm{RMS}(n) = \sqrt{\frac{1}{\mathcal{K}}\sum\limits_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} x(i)^2}$$

intensity, magnitude & loudness

features: root mean square 1/2

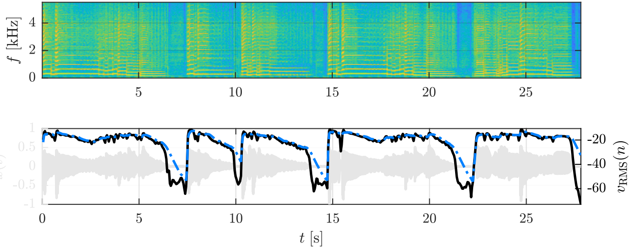
$$v_\mathrm{RMS}(n) = \sqrt{\frac{1}{\mathcal{K}} \sum_{i=i_\mathrm{s}(n)}^{i_\mathrm{e}(n)} x(i)^2}$$

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

intensity, magnitude & loudness

features: root mean square 1/2

$$v_{\mathrm{RMS}}(n) = \sqrt{\frac{1}{\mathcal{K}} \sum_{i=i_{\mathrm{s}}(n)}^{i_{\mathrm{e}}(n)} x(i)^2}$$



matlab source: matlab/plotFeatures.m

intensity, magnitude & loudness

features: root mean square 2/2

Georgia Tech | Center for Music Technology College of Design

common variants (sample processing only):

- reduce computational complexity

$$
\begin{aligned}
v_{\mathrm{RMS}}^2(n) &= \frac{x(i_\mathrm{e}(n))^2 - x(i_\mathrm{s}(n-1))^2}{i_\mathrm{e}(n) - i_\mathrm{s}(n) + 1} + v_{\mathrm{RMS}}^2(n-1) \\
v_{\mathrm{RMS}}(n) &= \sqrt{v_{\mathrm{RMS}}^2(n)}
\end{aligned}
$$

- single pole approximation

$$
\begin{aligned}
v_{\mathrm{tmp}}(i) &= \alpha \cdot v_{\mathrm{tmp}}(i-1) + (1-\alpha) \cdot x(i)^2 \\
v_{\mathrm{RMS}}^{*}(i) &= \sqrt{v_{\mathrm{tmp}}(i)}
\end{aligned}
$$

intensity, magnitude & loudness

features: root mean square 2/2

common variants (sample processing only):

- reduce computational complexity

$$v_{\mathrm{RMS}}^2(n) = \frac{x(i_{\mathrm{e}}(n))^2 - x(i_{\mathrm{s}}(n-1))^2}{i_{\mathrm{e}}(n) - i_{\mathrm{s}}(n) + 1} + v_{\mathrm{RMS}}^2(n-1)$$

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

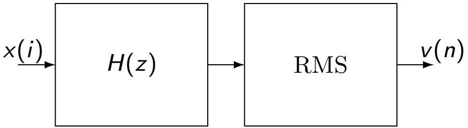
- single pole approximation

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

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

intensity, magnitude & loudness

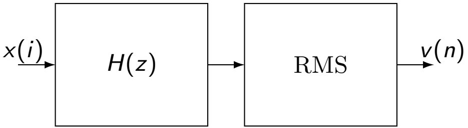
features: weighted root mean square

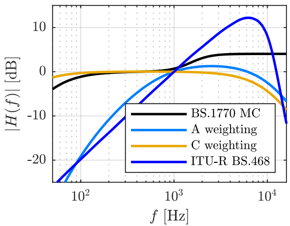


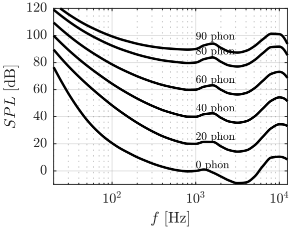
$H(z)$:

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

intensity, magnitude & loudness
features: weighted root mean square







$H(z)$:

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

matlab source: plotLoudnessWeighting.m

intensity, magnitude & loudness

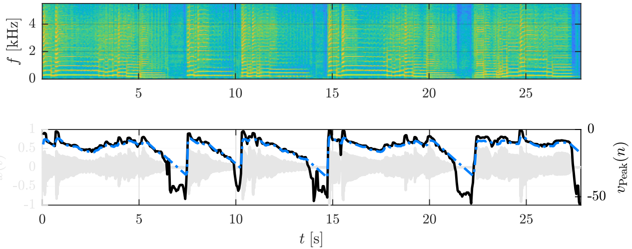
features: peak envelope (max)

$$v_{\mathrm{Peak}}(n) = \max_{i_{\mathrm{s}}(n)\leq i\leq i_{\mathrm{e}}(n)} |x(i)|$$

intensity, magnitude & loudness

features: peak envelope (max)

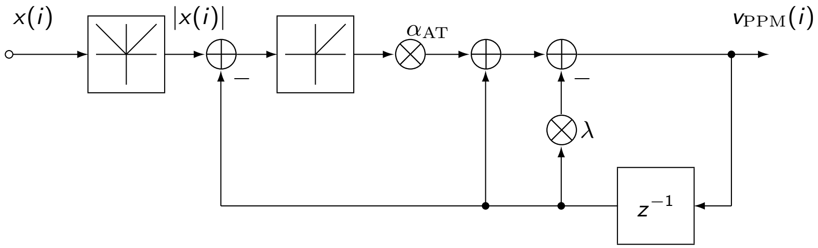
$$v_{\mathrm{Peak}}(n) = \max_{i_{\mathrm{s}}(n) \leq i \leq i_{\mathrm{e}}(n)} |x(i)|$$



matlab source: matlab/plotFeatures.m

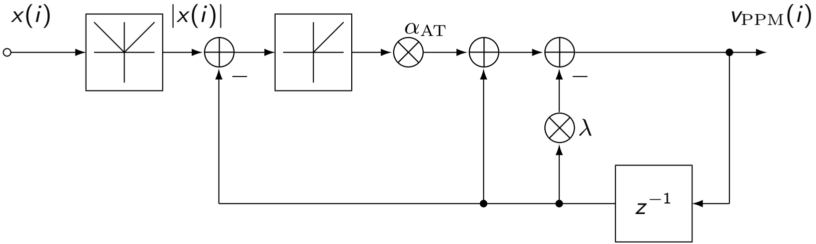
intensity, magnitude & loudness

features: peak envelope (PPM) 1/2



intensity, magnitude & loudness

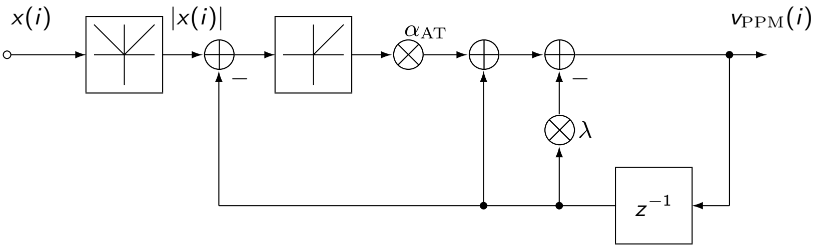
features: peak envelope (PPM) 1/2



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

intensity, magnitude & loudness

features: peak envelope (PPM) 1/2

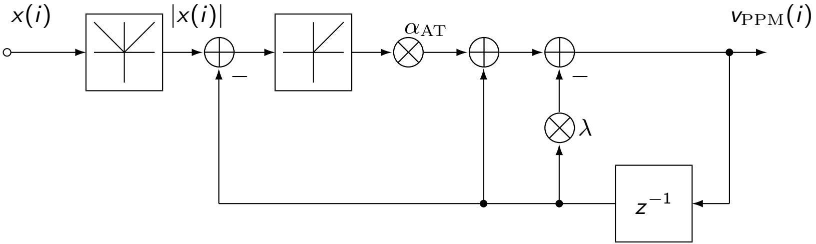


- **release state** $(|x(i)| < v_{\mathrm{PPM}}(i-1) \Rightarrow \lambda = \alpha_{\mathrm{RT}})$

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

intensity, magnitude & loudness

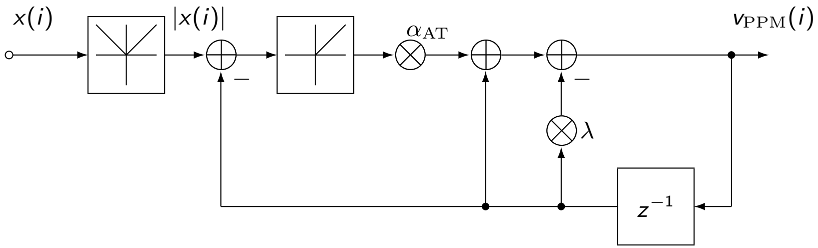
features: peak envelope (PPM) 1/2



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

intensity, magnitude & loudness

features: peak envelope (PPM) 1/2

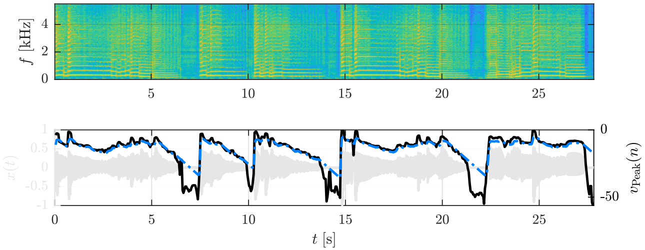


- **attack state** $(|x(i)| \geq v_{\mathrm{PPM}}(i-1) \Rightarrow \lambda = 0)$

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

intensity, magnitude & loudness

features: peak envelope (PPM) 2/2

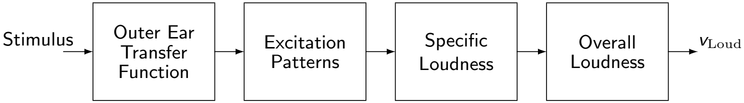


- gold: max per block
- blue: PPM

matlab source: matlab/plotFeatures.m

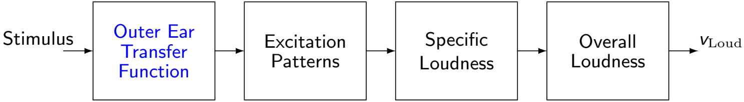
intensity, magnitude & loudness

features: zwicker loudness

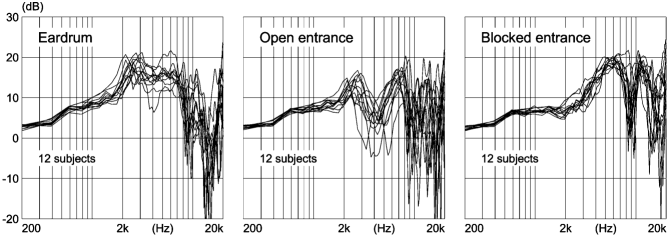


intensity, magnitude & loudness

features: zwicker loudness



- outer ear transfer function[1]



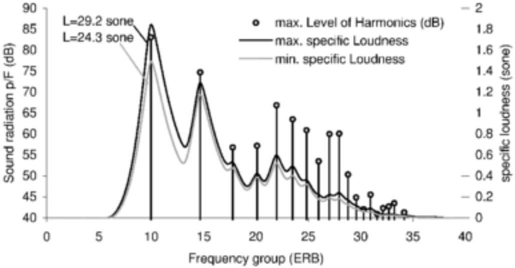
[1] D. Hammershøi and H. Møller, "Methods for Binaural Recording and Reproduction," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 303–311, May 2002.

intensity, magnitude & loudness

features: zwicker loudness

Stimulus → Outer Ear Transfer Function → Excitation Patterns → Specific Loudness → Overall Loudness → v_{Loud}

- excitation patterns[1]



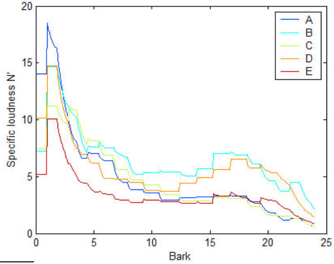
[1] M. Schleske, *Vibrato of the musician*, [Online]. Available: http://www.schleske.de/en/our-research/handbook-violinacoustics/vibrato-of-the-musician.html (visited on 07/29/2015).

intensity, magnitude & loudness

features: zwicker loudness

Stimulus → Outer Ear Transfer Function → Excitation Patterns → Specific Loudness → Overall Loudness → v_{Loud}

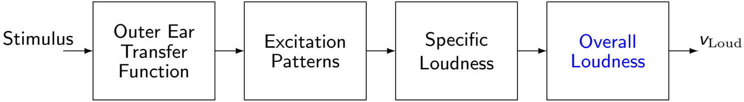
- specific loudness[1]



[1] U. 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

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

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)