Instrument-centered Parameter Estimation and Sound Synthesis

Jakob Abeßer

Semantic Music Technologies Group
Fraunhofer IDMT

Ilmenau, 07.04.2016

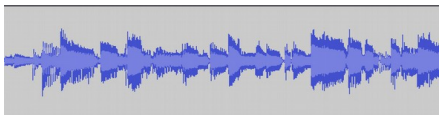
Fraunhofer IDMT

Outline

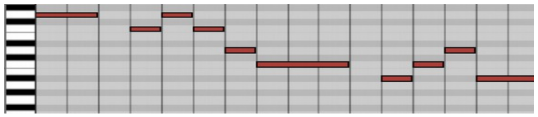
- **Motivation**
- Parameter Estimation / Transcription
- Sound Synthesis & Audio Coding
- Summary

Motivation

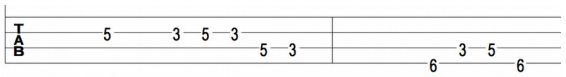
- **Automatic Music Transcription**



Transcription



Piano roll



Tablature

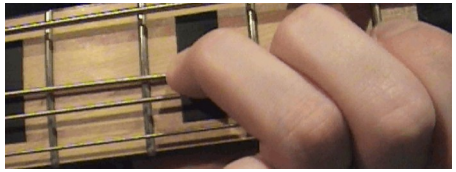
  - Audio signal → sequence of note events
  - Note parameters
    - **Score-level** (onset, offset, pitch & loudness)
    - **Instrument-level** (playing techniques, string & fret number)

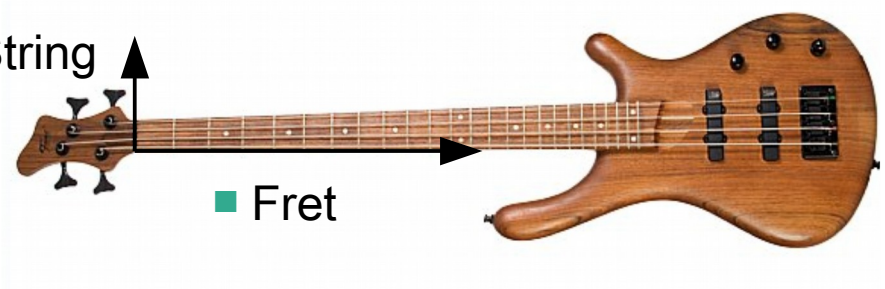
Motivation

- Automatic Music Transcription algorithms reach a **glass ceiling**
  - Analysis of polyphonic mixtures is challenging
- Analysis of musical performances requires a **high level of detail**
  - Playing techniques on the instrument
  - Note articulation (legato – staccato)
  - Dynamics
- **Instrument characteristics & sound production mechanisms** must be taken into account!

Motivation

- **Electric bass guitar**
  - 4 - 6 strings → f_0 ~ 40 - 400 Hz
- Sound production → **2 consecutive physical gestures**
  - **Plucking style** & **Expression style**
- Fretboard position
  - **String number** & **Fret number**
  - Ambiguity between pitch & fretboard position



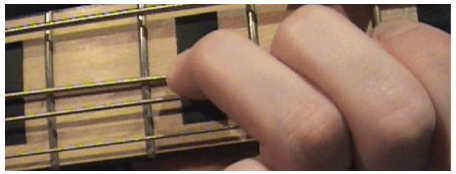


Fraunhofer IDMT

Motivation

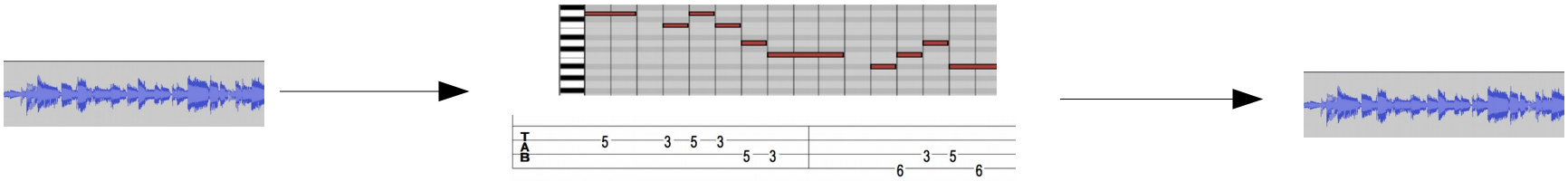
- **Audio examples (isolated bass guitar notes)**
  - Plucking Styles
    1. **Finger style**
    2. **Muted** (damped sound)
    3. **Picked** (brighter sound)
    4. **Slap-Thumb** (metal-like sound)
    5. **Slap-Pluck** (metal-like sound)

  - Expression Styles
    6. **Harmonics** (flageolet tones, higher)
    7. **Dead-notes** (strong damping, percussive sound)
    8. **Bending** & **Vibrato** (pitch modulation)



Motivation

- **Audio Synthesis / Audio Coding**
  - Physical modeling of electric bass guitar
  - Parametric audio coder



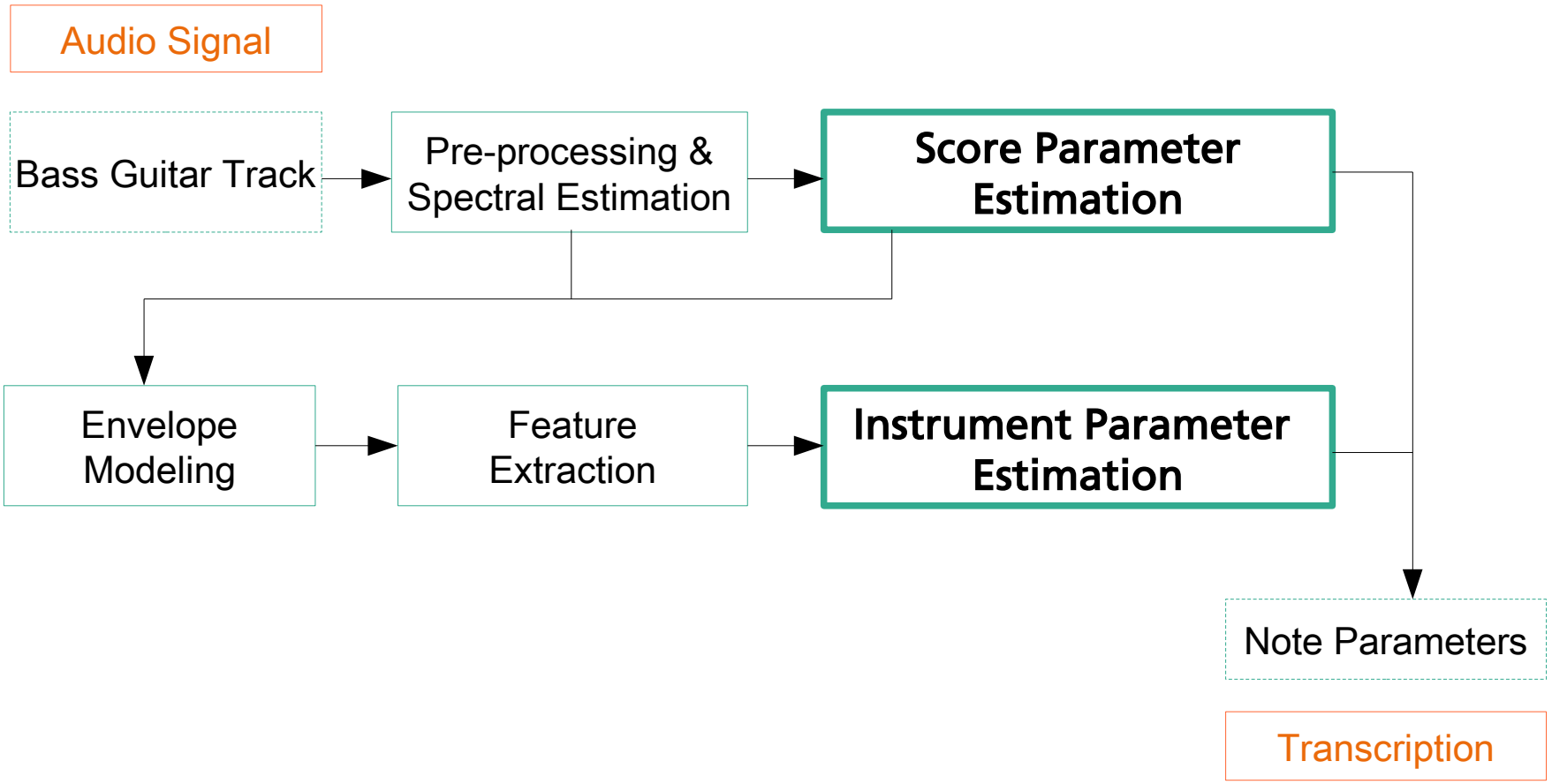
Outline

- Motivation
- **Parameter Estimation / Transcription**
- Sound Synthesis & Audio Coding
- Summary

Parameter Estimation → Related Work

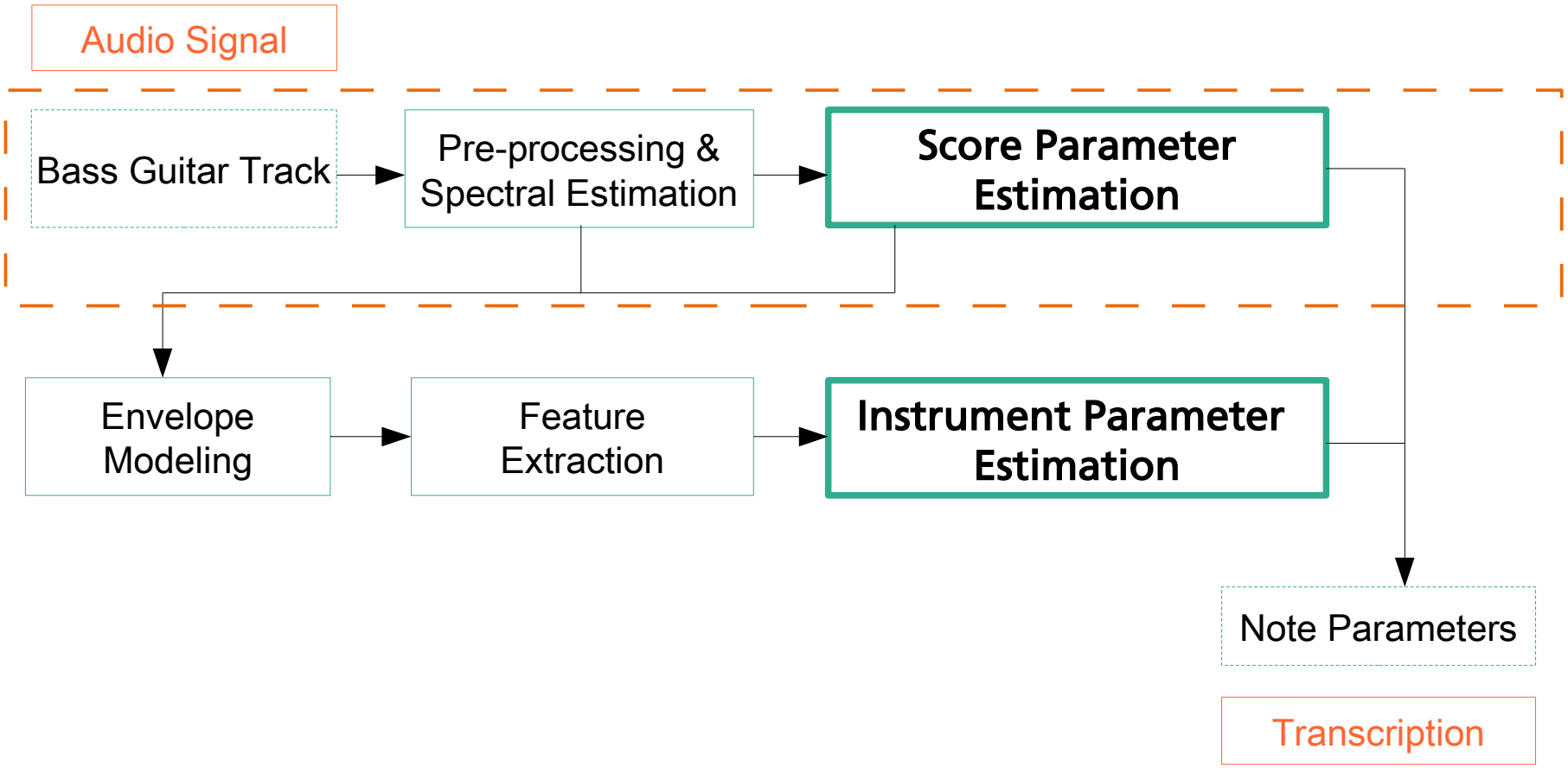
- Instrument-centered transcription algorithms for violin, cello, electric guitar, piano (use constraints: pitch range, magnitude templates, string number)
- Different data acquisition methods (audio, video, sensors, motion capturing)
- Estimation of playing techniques
  - String damping
  - Vibrato, bending, slides
- Estimation of fretboard positions / chord voicings

Music Transcription → Proposed Approach



Fraunhofer IDMT

Music Transcription → Proposed Approach



Fraunhofer IDMT

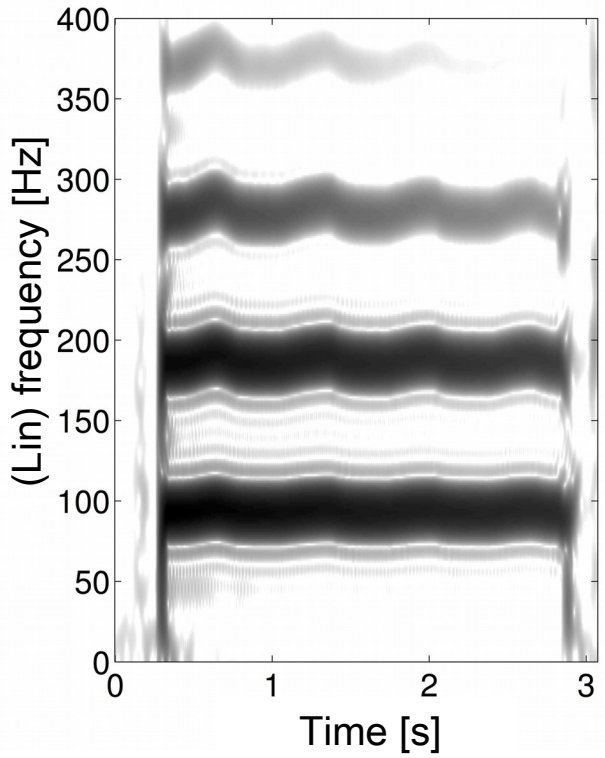
Music Transcription → Proposed Approach

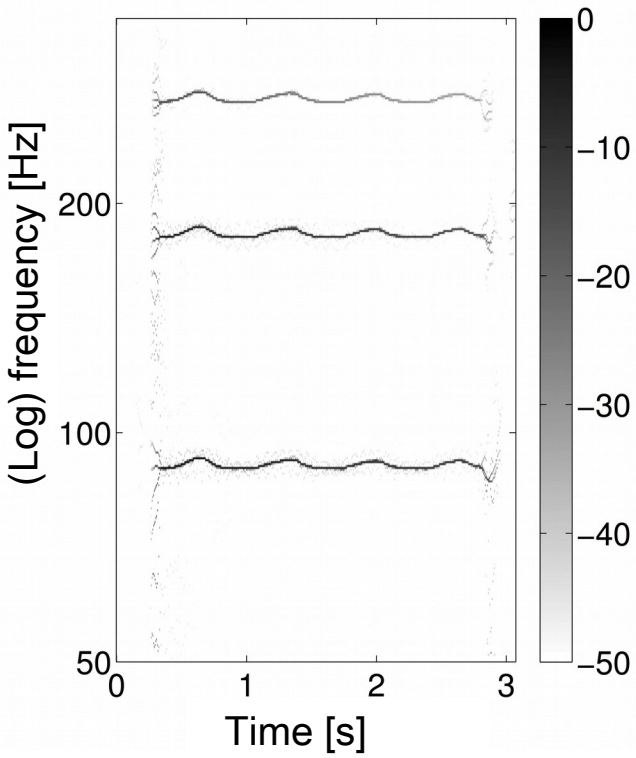
- **Pre-processing**
  - Down-sampling to 5.5 kHz, stereo → mono
- **Spectral Estimation**
  - **Short-time Fourier Transform (STFT)**
    - Linear frequency axis
    - Spectral leakage → limited frequency resolution
  - **Reassigned instantaneous frequency (IF) spectrogram**
    - Logarithmic frequency axis → 120 bins / octave
    - Sharper peaks → better frequency resolution!

Music Transcription → Proposed Approach

- **Example** (Bass guitar note with vibrato)

STFT



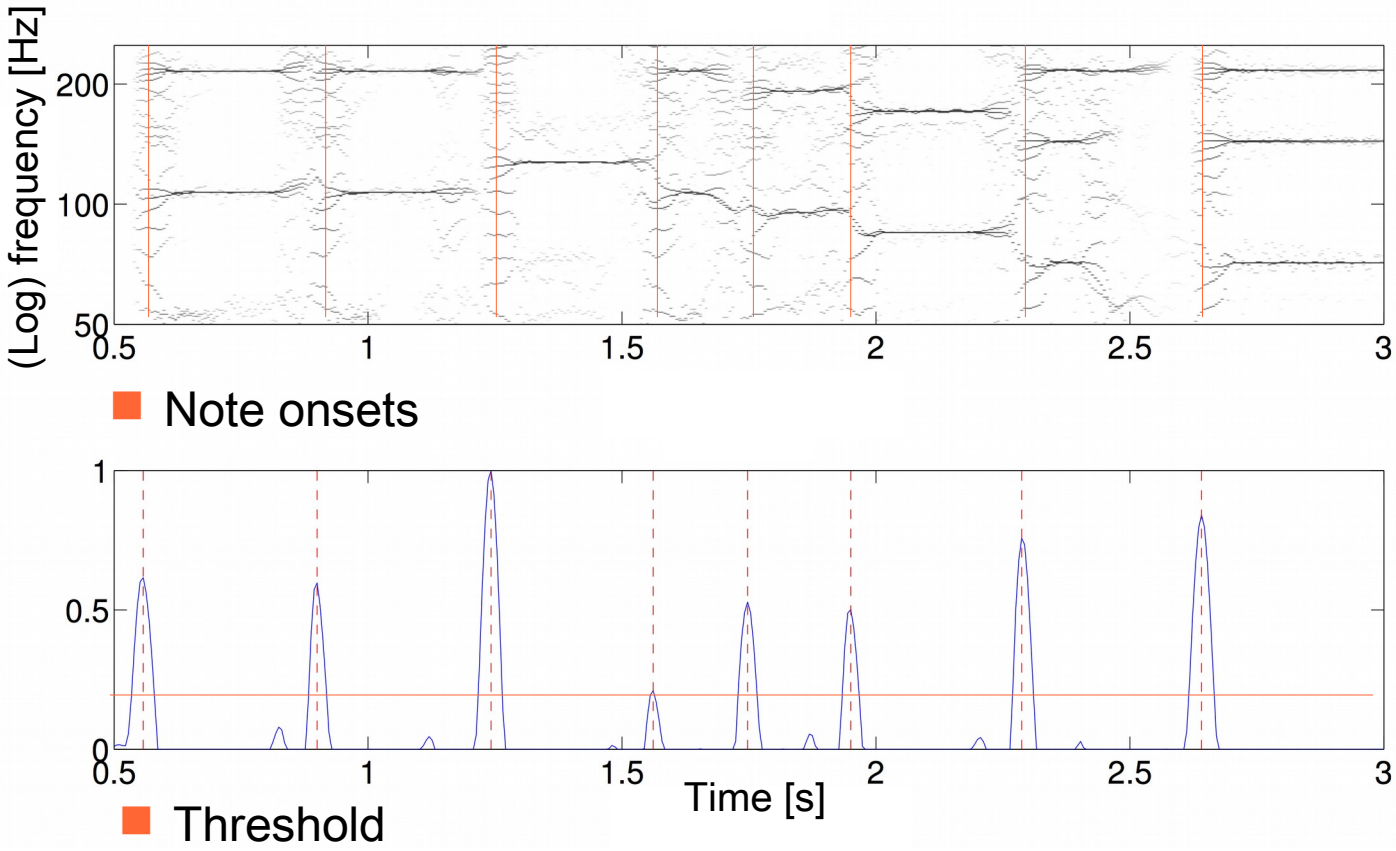


Reassigned IF

Fraunhofer IDMT

Music Transcription → Proposed Approach

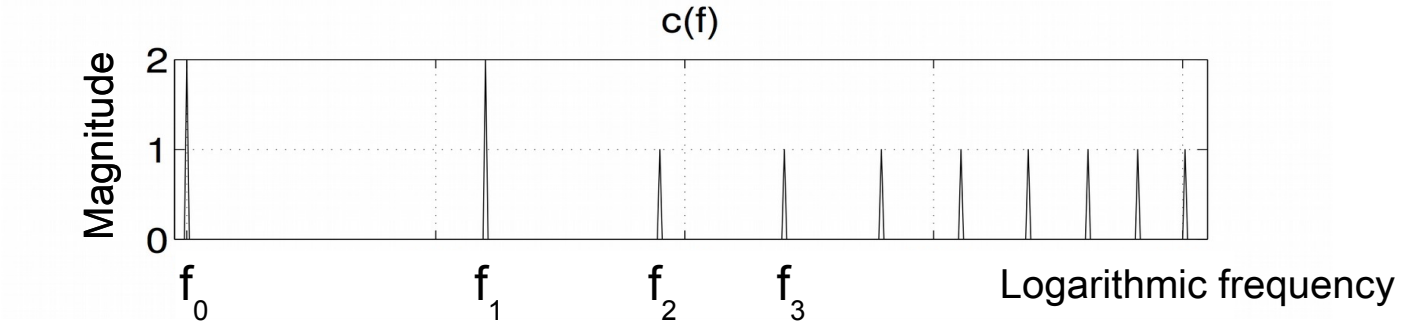
- **Score Parameter Estimation**
  - IF spectrogram
  - Onset detection functions



Fraunhofer IDMT

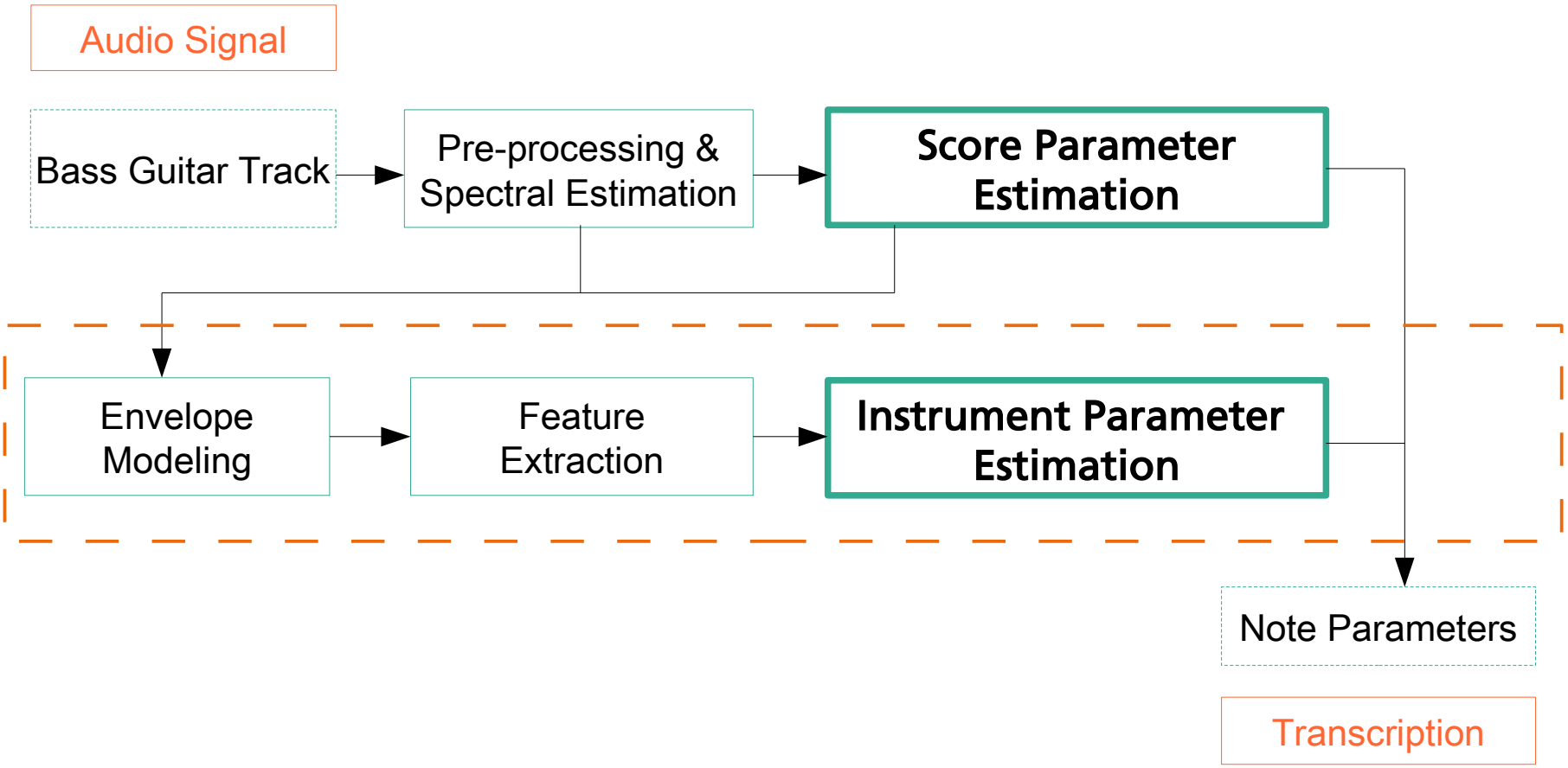
Music Transcription → Proposed Approach

- **Pitch detection**
  - **Harmonic spectral template** $f_k \approx f_0(k+1)\sqrt{1+\beta(k+1)^2}$



  - Maximize **cross-correlation** (**monophony** assumption)
- **Fundamental frequency (f_0) tracking**
  - **Frame-wise tracking** (**continuity** assumption)

Music Transcription → Proposed Approach



Fraunhofer IDMT

Music Transcription → Proposed Approach

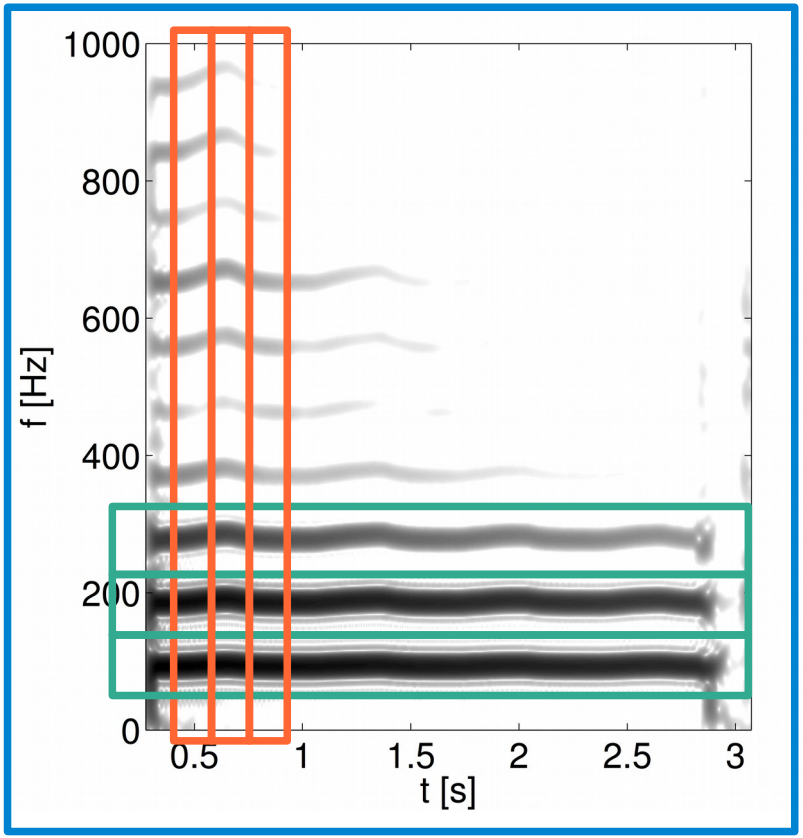
- **Envelope Modeling**
  - Simple **parametric model**
    - Spectrum = Sum of **harmonic components** (f_0 + overtones) with **time-varying magnitudes**
    - **Quasi-harmonic relationship**

$$f_k \approx f_0(k+1)\sqrt{1+\beta(k+1)^2}$$

    - Wideband attack transients are not modeled
  - **Frame-wise** estimation of harmonic magnitudes using linear interpolation

Music Transcription → Proposed Approach

- **Feature extraction**
  - **Frame-wise**
  - **Note-wise**
  - **Envelope-wise**
- **Goal**
  - Describe sound (timbre)
  - Automatic classification of playing techniques & string number



Music Transcription → Proposed Approach

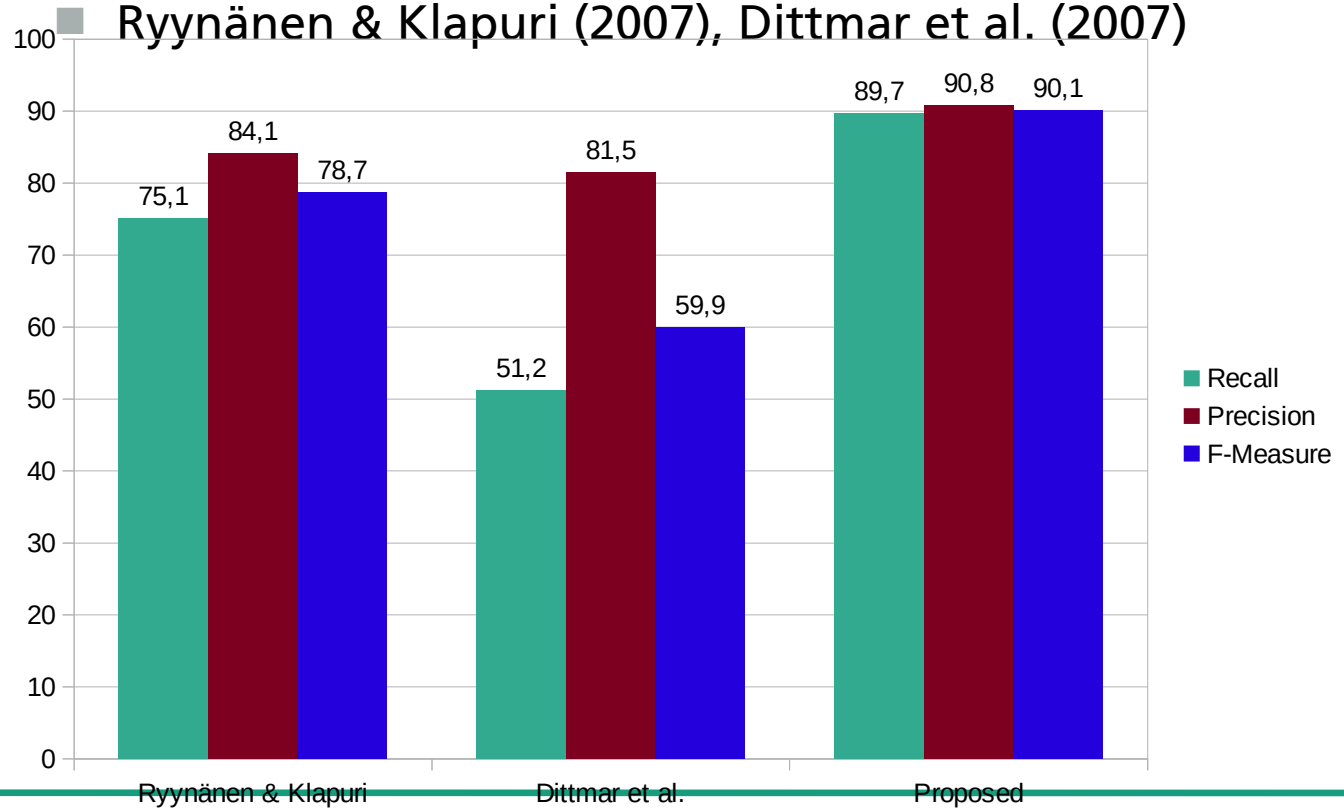
- **Examples (features)**
  - **Frame-wise** features
    - Magnitude & frequency relationships (overtones)
    - Noisiness & subharmonic components
  - **Envelope-wise** features
    - Modulation frequency
    - Number of modulation periods…
    - Modulation tendency

Music Transcription → Proposed Approach

- **Classification** of **Plucking Style & Expression Style & String Number**
  - Machine learning approach (e.g., Support Vector Machine – SVM)
- **Estimation** of **Fret Number**
  - Derived from string number & pitch & string tuning
- **Context-based Error Correction**
  - Most note pitches cannot be played on all strings!

Music Transcription → Evaluation

- Estimation of **score-level parameters (onset & pitch) → 17 bass lines**



Fraunhofer IDMT

Music Transcription → Evaluation

- Estimation of **instrument-level parameters**

Parameter (# Classes)	**Isolated note samples**	**Isolated bass lines**
Plucking Style (5)	96 %	85 %
Expression Style (6)	95 %	73 %
String Number (4)	88 %	92 %

- **Problem**
  - Bass lines have more variations w.r.t. playing styles & note duration

Outline

- Motivation
- Parameter Estimation / Transcription
- **Sound Synthesis & Audio Coding**
- Summary

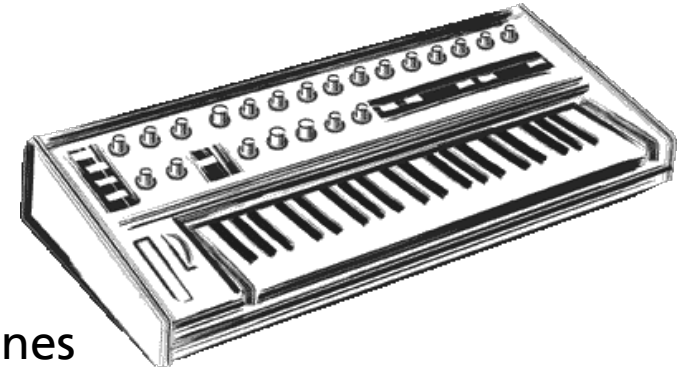
Sound Synthesis → Goals

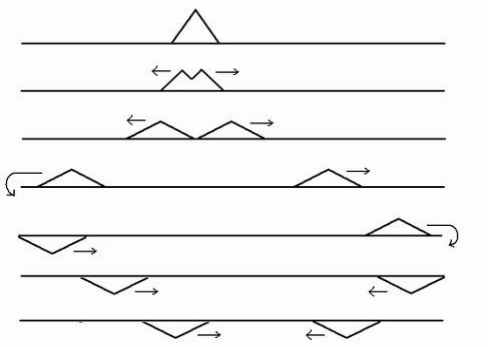
- **Audio Synthesis**
  - Model the **sound production** of the electric bass guitar
  - **Incorporate instrument-level parameters** (string number, playing techniques)
  - Incorporate knowledge about sound production mechanisms

- **Parametric Audio Coding**
  - Represent **audio track** as **sequence of note events & parameters**
  - Which **acoustic parameters** are important for natural synthesis?
  - Strong data rate reduction

Sound Synthesis → Related Work

- Sample-based synthesis
- Additive / subtractive / FM synthesis
- Physical modeling of vibrating systems
  - 1D → strings, 2D → plates, membranes
  - Karplus-Strong algorithm (Karplus & Strong 1983, Smith 1992)
  - Models for guitar, piano, drums, brass





Sound Synthesis → Goals

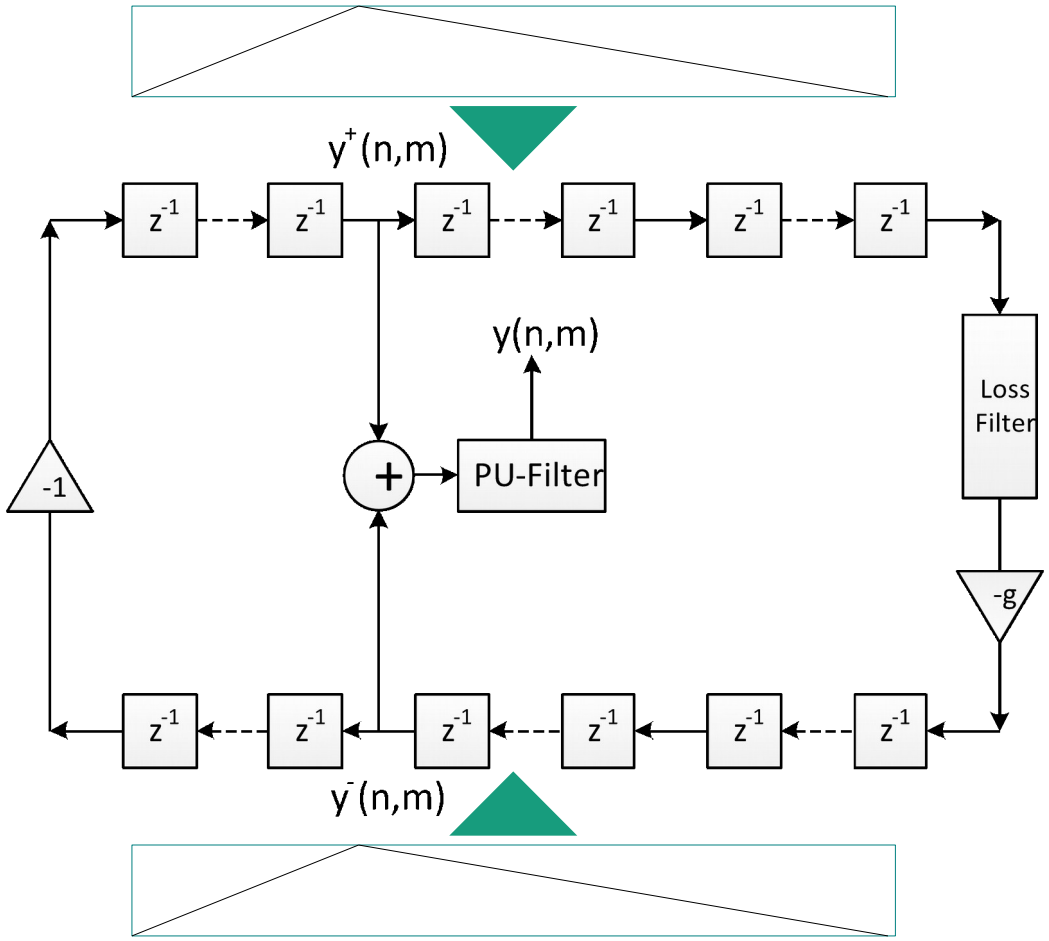
- Extend existing physical modeling algorithms
  - Various **excitation functions** (different plucking styles)
  - **String damping** (muted, harmonics, dead-notes)
  - **String-fretboard collision** (slap techniques)
  - **Inharmonicity** (thick bass guitar strings)

Sound Synthesis

Video: Superposition of two waves

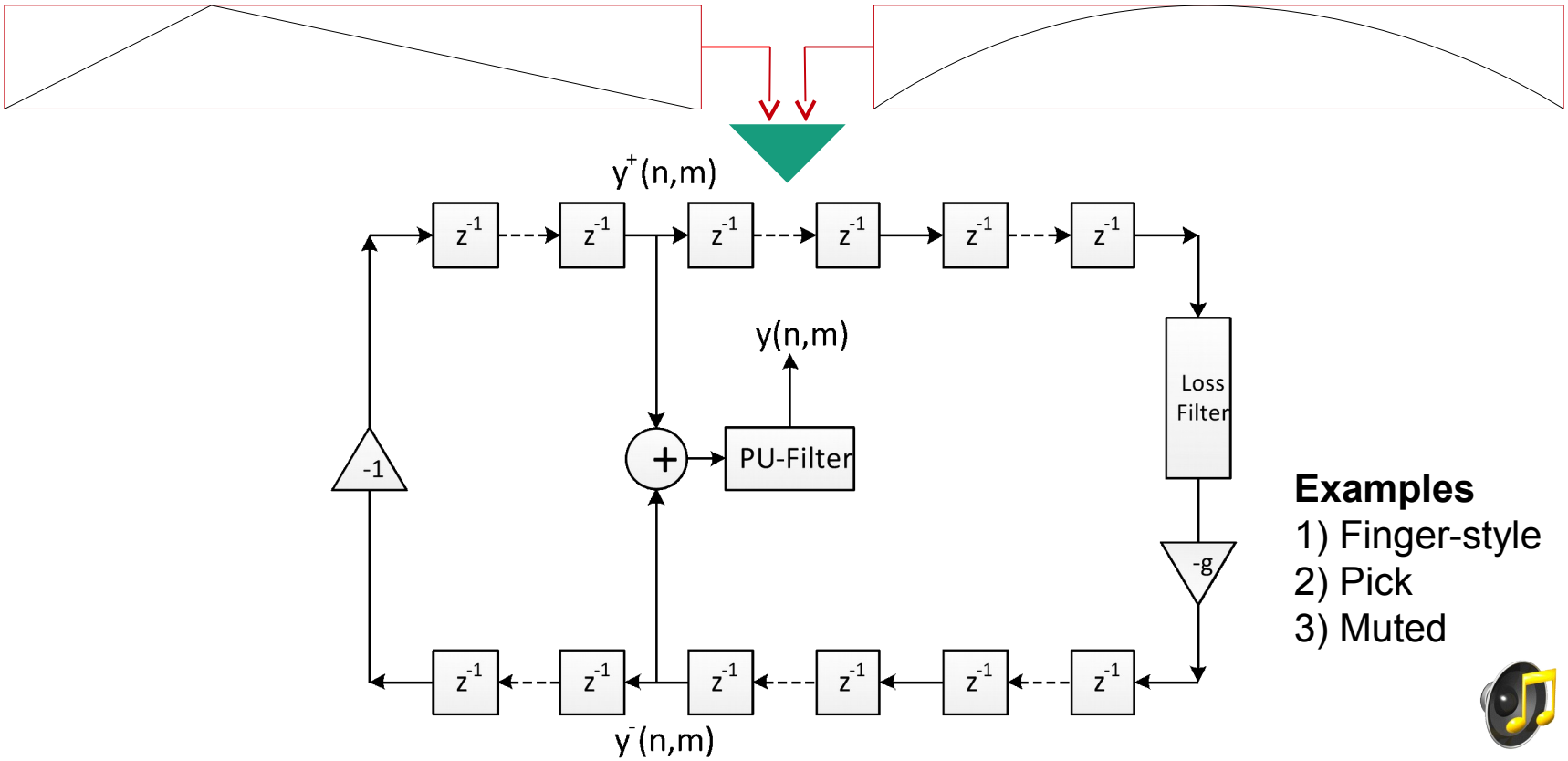
Sound Synthesis

Basic Model:



Sound Synthesis

Excitation function relate to plucking styles:



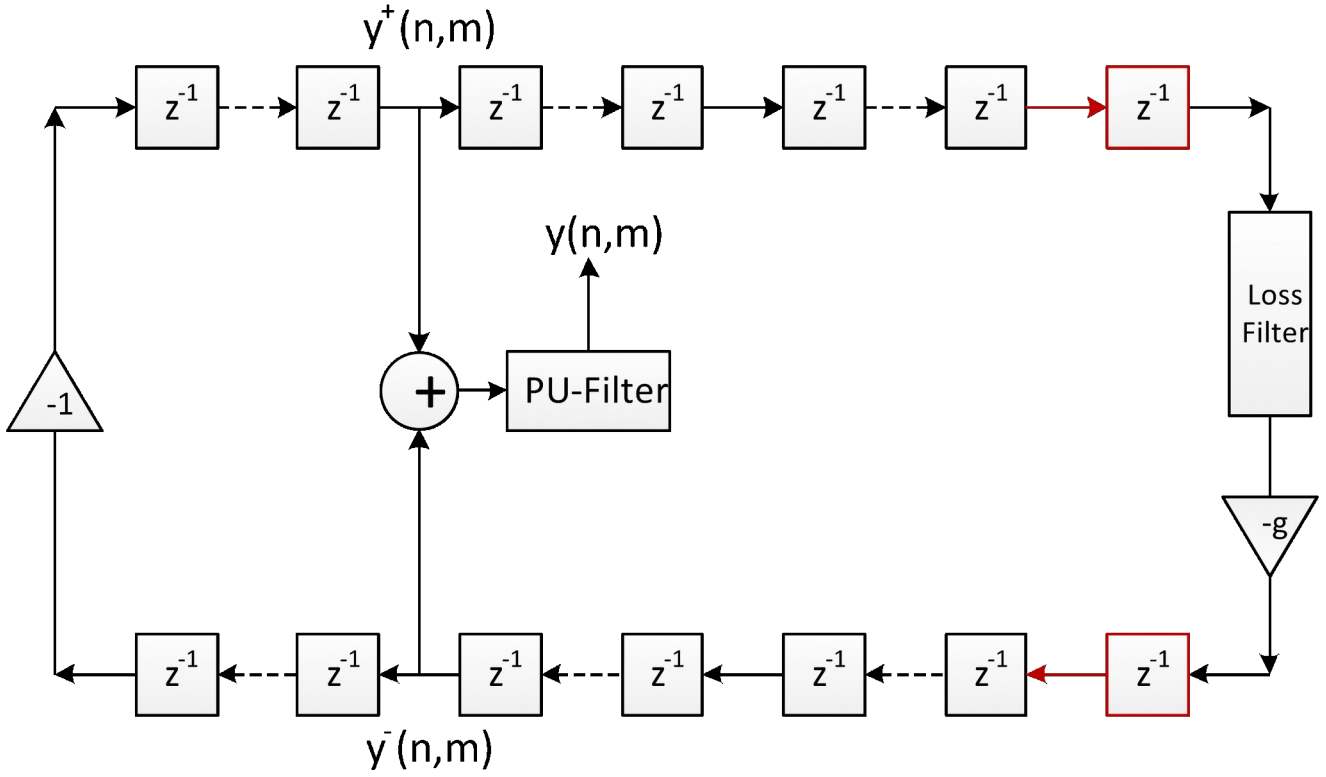
Examples

1) Finger-style
2) Pick
3) Muted

Fraunhofer IDMT

Sound Synthesis

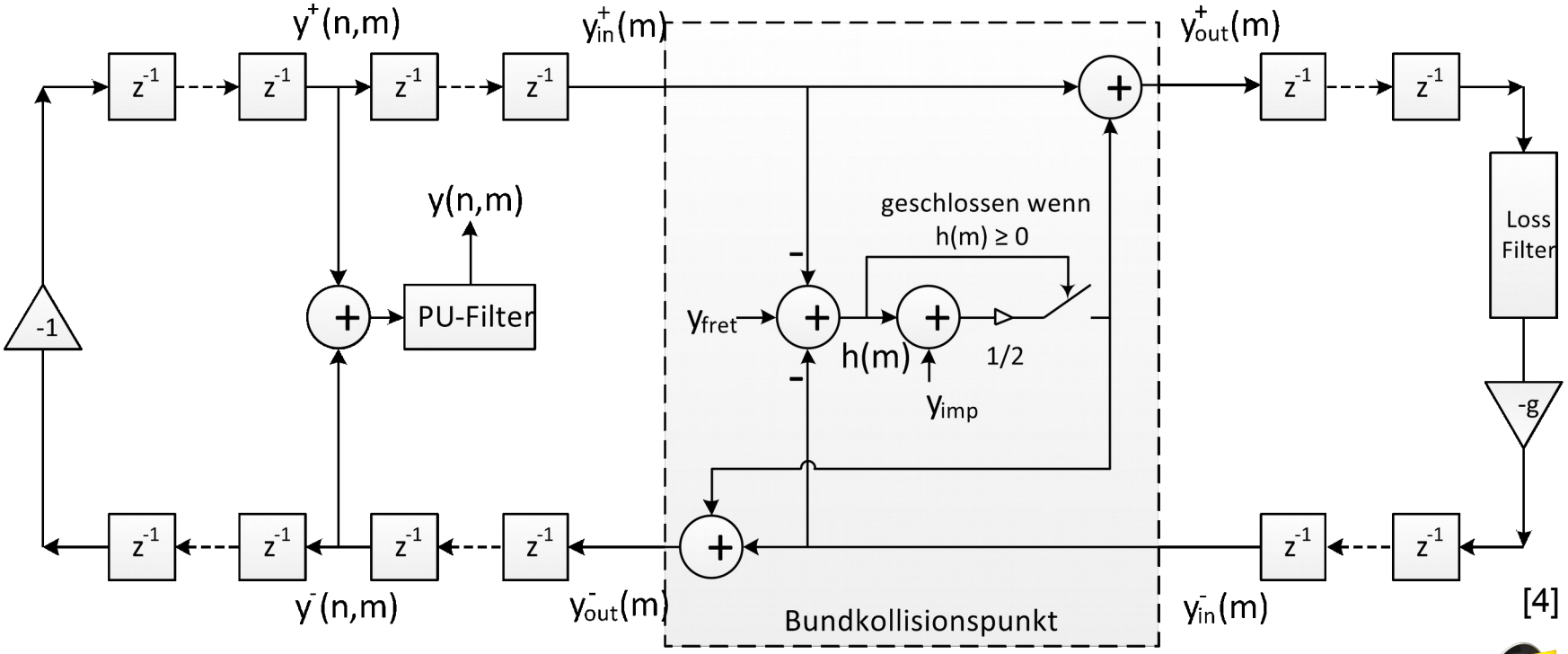
Frequeny Modulations:



Examples
1) Normal
2) Bending
3) Vibrato

Sound Synthesis

Collision between strings & fretboard (slap styles):



Examples 1) Slap-Thumb 2) Slap-Pluck

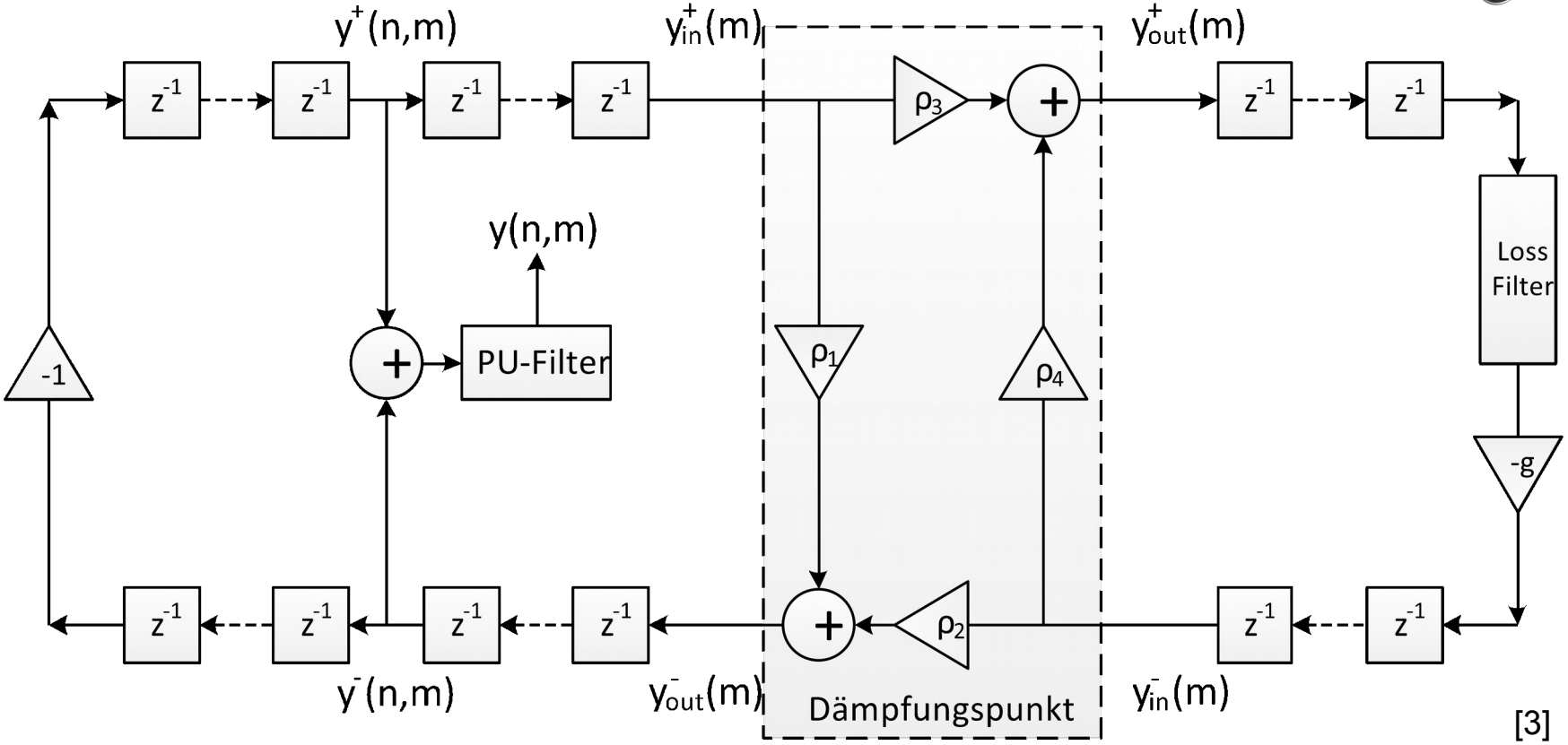
Sound Synthesis

Video: String Vibration & Collision with fretboard

Sound Synthesis

Damping points:

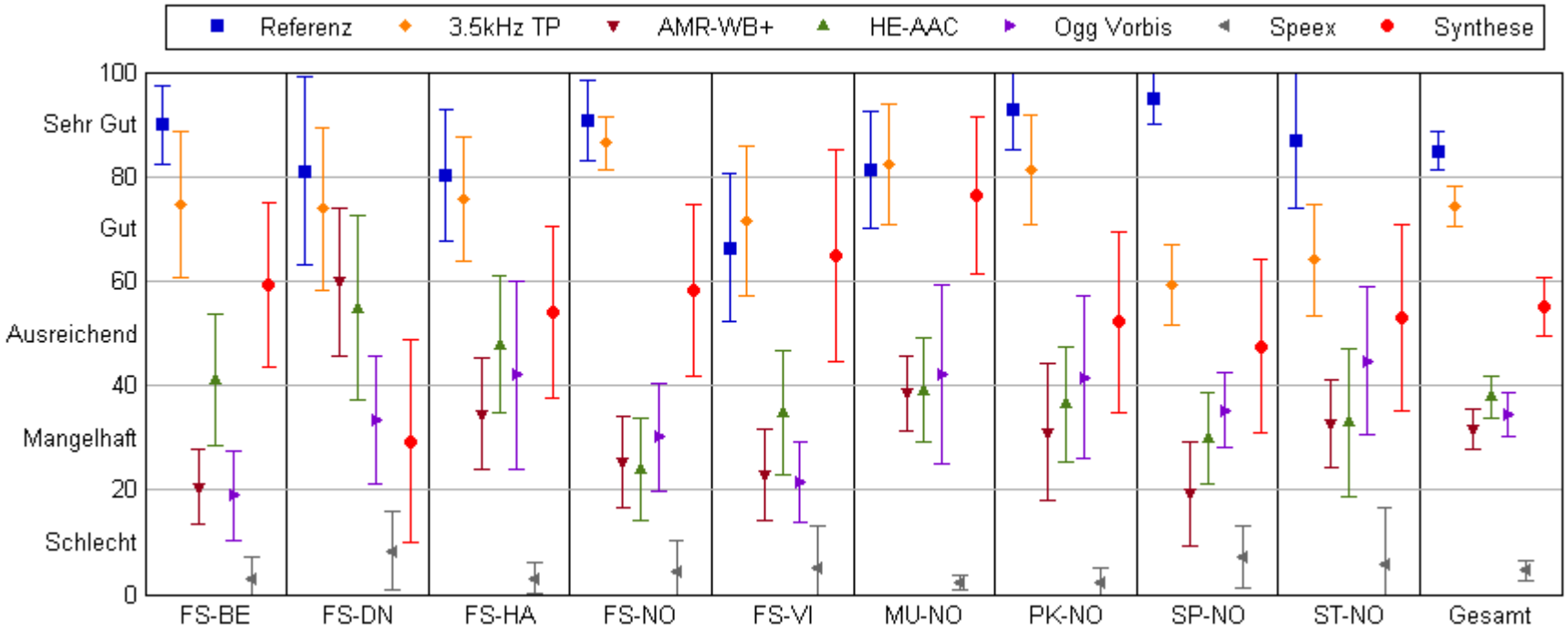
Examples: 1) Dead-notes 2) Harmonics



[3]

Audio Coding - Evaluation

MUSHRA listening test



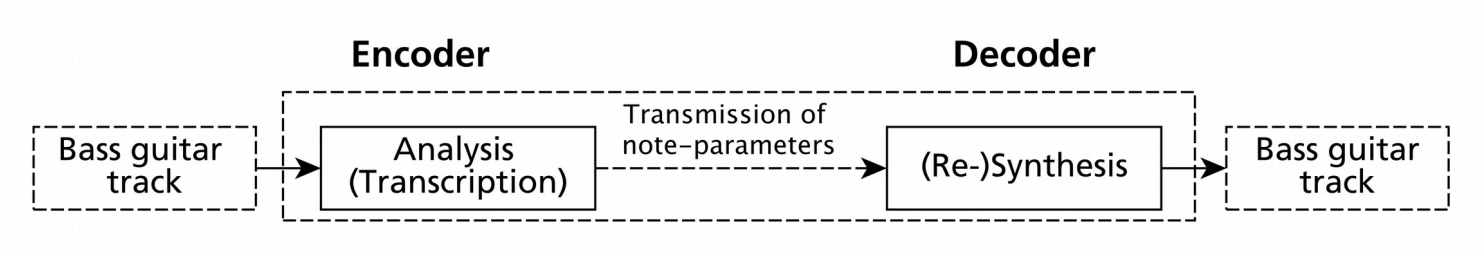
Audio Coding Algorithm	AMR-WB+	HE-AAC	Ogg Vorbis	Speex	Synthesis
Average Bitrate	ca. 6 kbit/s	14,15 kbit/s	17,24 kbit/s	7,14 kbit/s	0,25 kbit/s

Sound Synthesis

- **Audio Examples**
  1. Plucking Style: picked, Expression style: normal, vibrato
  2. Plucking style: slap-pluck, slap-thumb
  3. Plucking style: finger-style, expression style: normal, harmonics

Audio Coding → Proposed Approach

- **Parametric Instrument Coder**



- **Parameters**
  - Note onset, duration, fundamental frequency, loudness
  - Plucking style, expression style, string number, fret number
  - Modulation frequency & extend (f_0)
- **Average bit-rate** = 225.4 bit/s

Audio Coding → Proposed Approach

- **Codec Parameters**

Parameter	Range	Resolution	Quantization bits (steps)
Loudness $\mathcal{L}$	$[0, 127\,\mathrm{dB}]$	$0.1\,\mathrm{dB}$	11 (2048)
Plucking Style $\mathcal{S}_\mathrm{P}$	$[1, 5]$	1	3 (8)
Expression Style $\mathcal{S}_\mathrm{E}$	$[1, 6]$	1	3 (8)
String Number $\mathcal{N}_\mathrm{S}$	$[1, 4]$	1	2 (4)
Fret Number $\mathcal{N}_\mathrm{F}$	$[0, 24]$	1	5 (32)
Onset $\mathcal{O}$	$[0, 30\,\mathrm{s}]$	$0.01\,\mathrm{s}$	12 (4096)
Duration $\mathcal{D}$	$[0, 20\,\mathrm{s}]$	$0.01\,\mathrm{s}$	11 (2048)
Magnitude decay α_t	$[0, 127\,\mathrm{dB/s}]$	$1\,\mathrm{dB/s}$	7 (128)
Fundamental frequency f_0	$[41.2\,\mathrm{Hz}, 382.0\,\mathrm{Hz}]$	$0.1\,\mathrm{Hz}$	12 (4096)
Modulation frequency $\chi_{\mathrm{f,mod}}$	$[0, 12\,\mathrm{Hz}]$	$0.1\,\mathrm{Hz}$	7 (128)
Modulation lift $\chi_{\mathrm{mod,lift}}$	$[0, 500\,\mathrm{cent}]$	$1\,\mathrm{cent}$	9 (512)
		$\sum$	82 bit/Note

Audio Coding & Sound Synthesis → Evaluation

- **MUSHRA listening tests**
  - **Perceptual quality** of synthesized bass lines
    - **Higher ratings** for new synthesis algorithm compared to AMR-WB+, HE-AAC, Ogg Vorbis algorithms (with lowest bit rate settings)
  - **Perceptual improvements** by **tuning the synthesis algorithm** (inharmonicity, note decay parameters)
    - Only **small improvements** compared to un-tuned version
  - **Importance of plucking & expression styles** for perceptual quality
    - **Higher importances** for correct **plucking style** synthesis

Summary

- **Instrument-centered parametrization & synthesis algorithms**
  - Rich set of expressive parameters (playing styles, geometric position on the instrument, dynamics etc.)
  - Robust extraction require isolated instrument tracks
  - Re-synthesis allows for transmission / encoding with low bitrates
  - Future challenges
    - Adaptation to specific instrument models
    - Parameter estimation from polyphonic mixtures
    - Adaptation of approach to other instrument types

Thank you for your attention!

Jakob Abeßer

Semantic Music Technologies Group
Fraunhofer IDMT

Ilmenau, 07.04.2016

Fraunhofer IDMT