Automatic Transcription of Bass Guitar Tracks applied for Music Genre Classification and Sound Synthesis

Dipl.-Ing. Jakob Abeßer

Wissenschaftliche Aussprache

Ilmenau, 18.09.2014

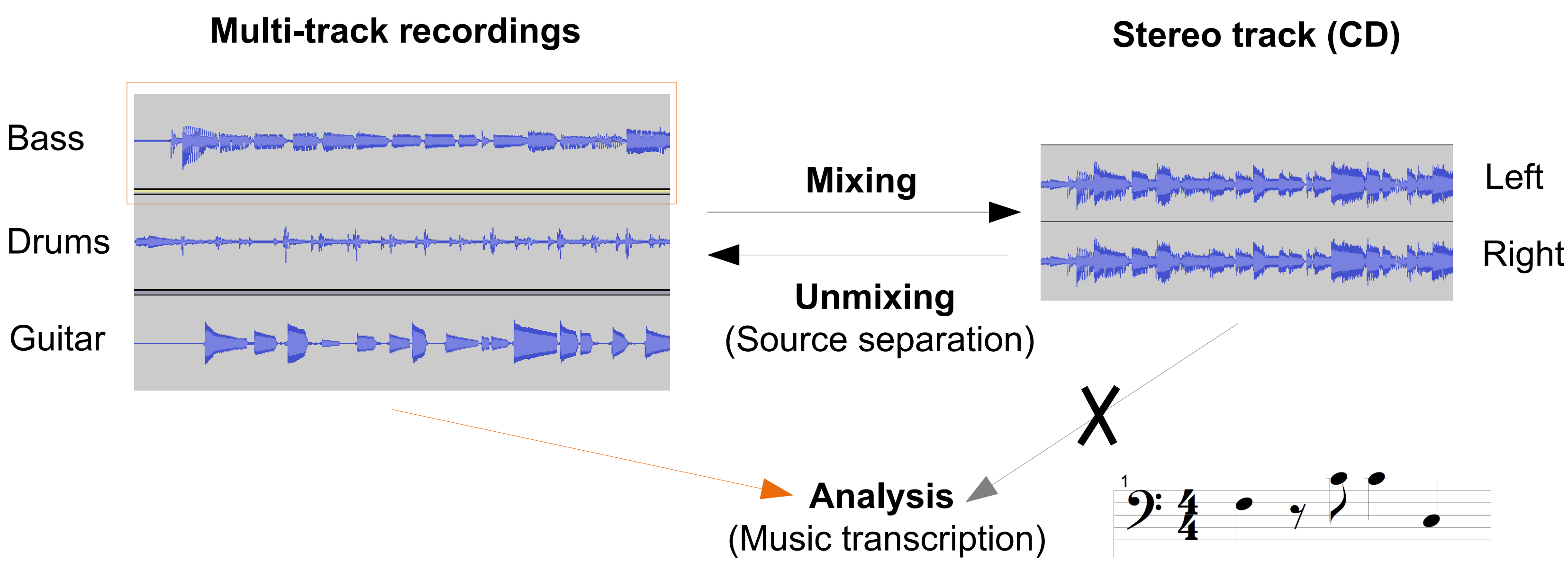
Fraunhofer
IDMT

Outline

- **Motivation**
- Thesis Structure
  - Music Transcription
  - Music Genre Classification
  - Audio Coding & Sound Synthesis
- Summary

Motivation

- Music Information Retrieval (MIR)
- Music recording = **mixtures** of overlapping instrument tracks

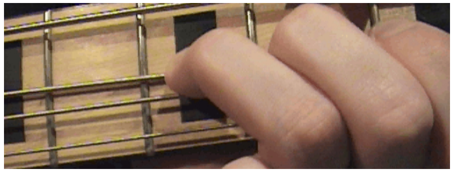


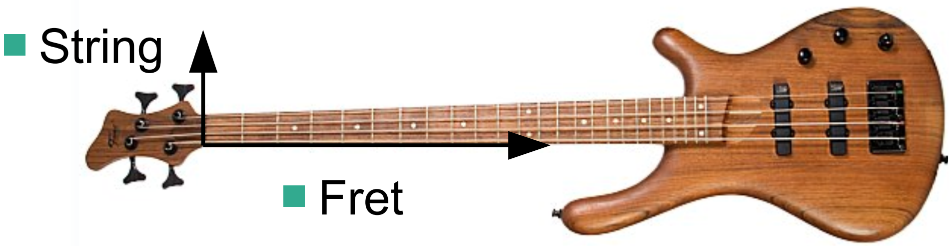
- In this PhD thesis → focus on i**solated electric bass guitar** tracks

Fraunhofer IDMT

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

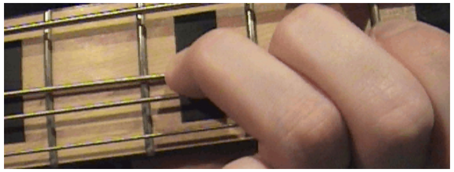




Motivation

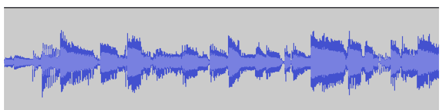
- **Audio examples (isolated bass guitar notes)**
  - Plucking Styles
    1. **Finger style**
    2. **Muted** (damped sound)
    3. **Picked** (brighter sound)
    4. **Slap-Thumb** (metal-like sound)
    5. **Slap-Pluck** (metal-like sound)

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

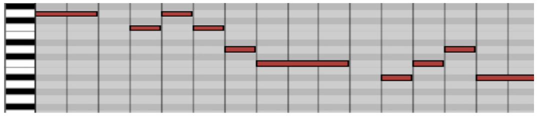


Motivation

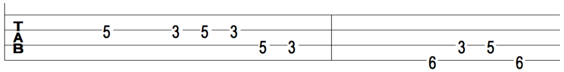
- **Automatic Music Transcription**



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)

Fraunhofer IDMT

Motivation

- **Application Scenarios**
  - **Music Genre Classification**
    - Quantify musical properties of **repetitive bass patterns**
    - Genre classification via Machine Learning

Bass Pattern

???

Minimal Techno

Bossa Nova

Swing

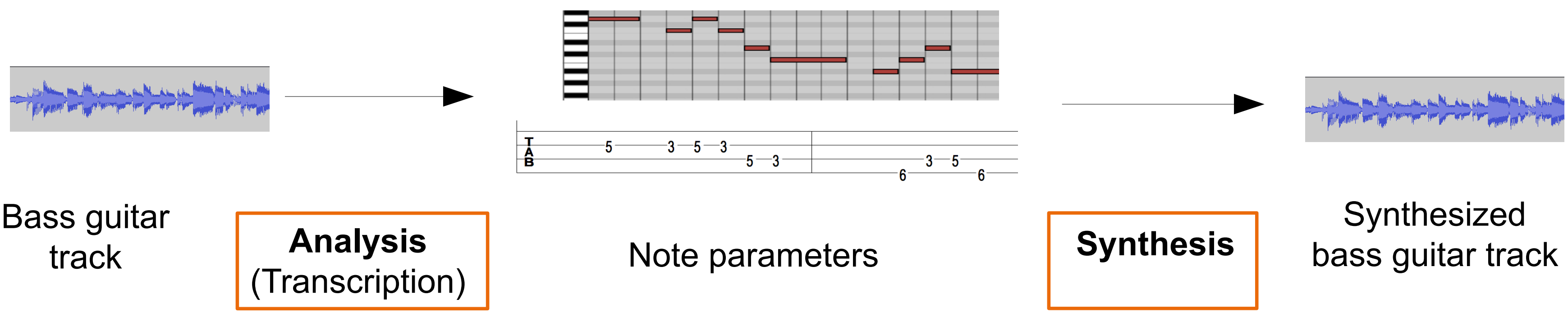
Funk

Blues

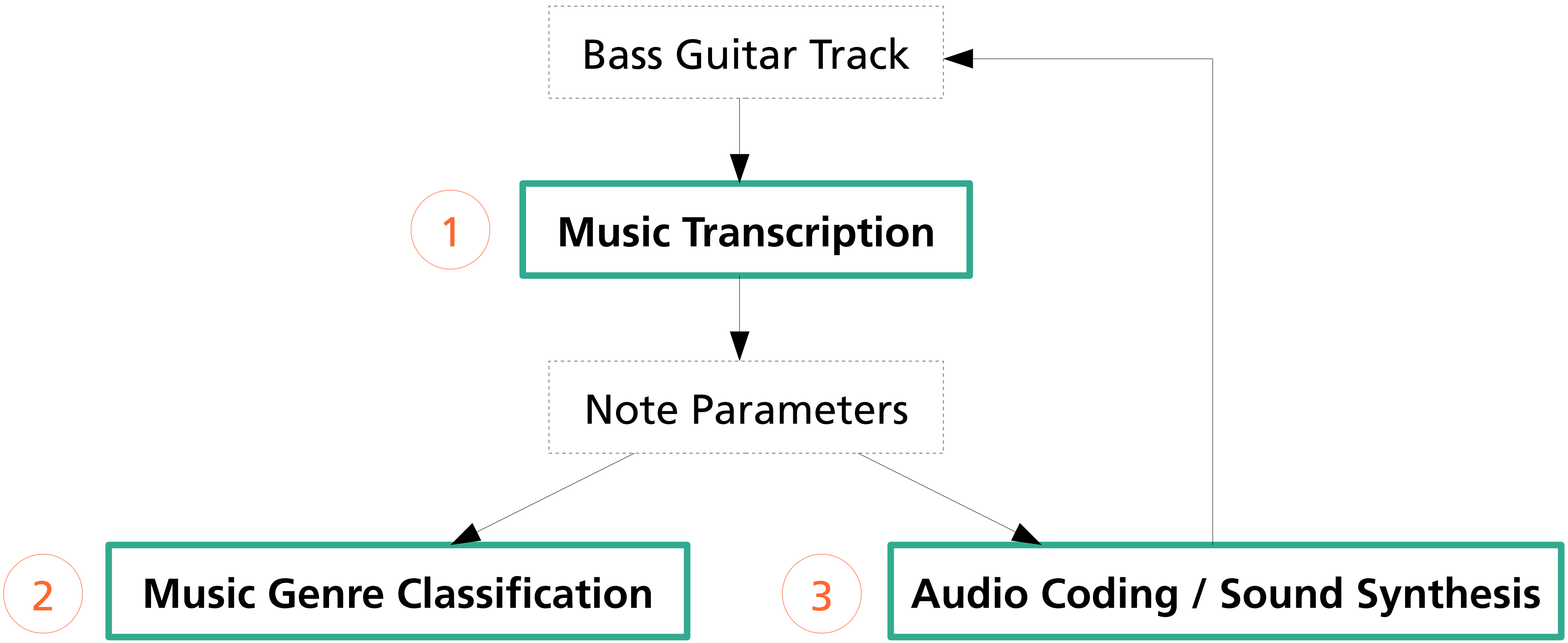
Music Genre

Motivation

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



Thesis Structure



Outline

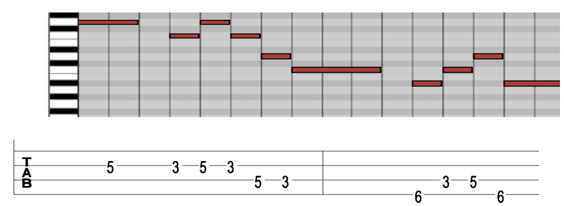
- Motivation
- Thesis Structure
  - **Music Transcription**
  - Music Genre Classification
  - Audio Coding & Sound Synthesis
- Summary

Music Transcription → Goals

- **Instrument-centered** transcription algorithm for **bass guitar**
  - **Score-level & instrument-level** parameters
- Focus on **isolated instrument tracks**
- Algorithm **adaptation & optimization** to the user's instrument
  - Application scenario: **music education / instrument tutoring**

Analysis
(Transcription)

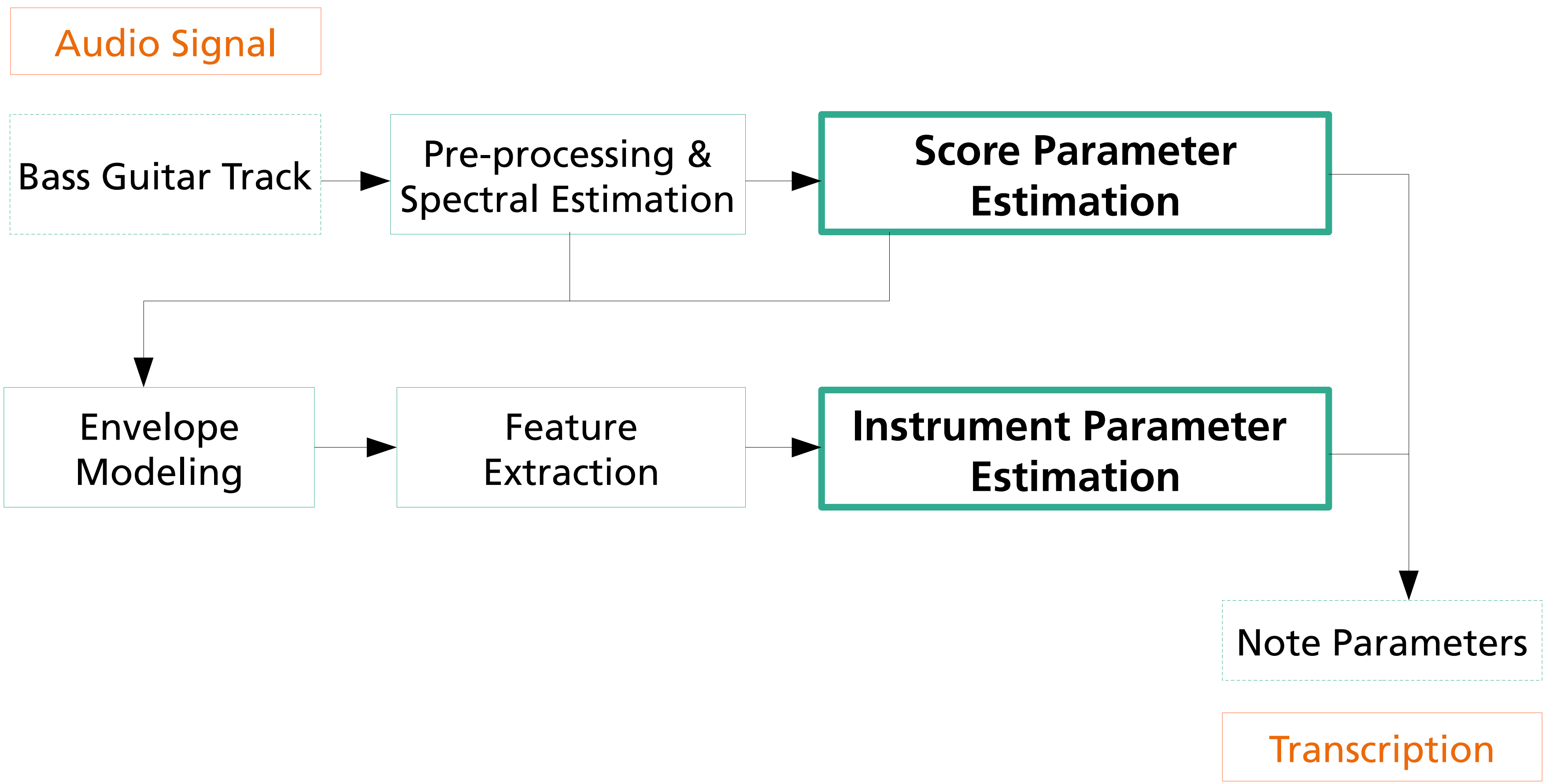
Performance Assessment
(Feedback)



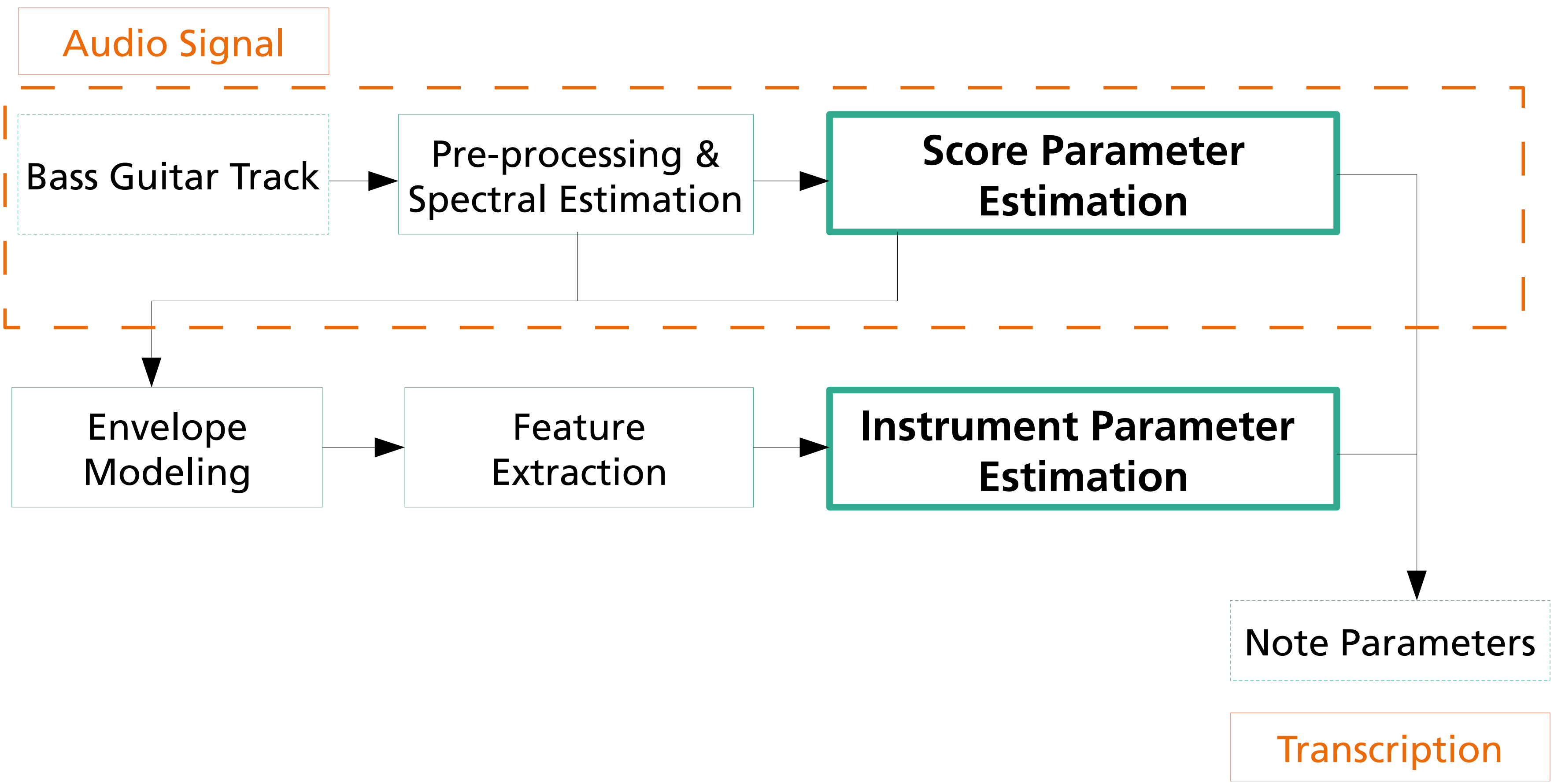
Music Transcription → Related Work

- **General-purpose** bass transcription algorithms [Hainsworth 2001], [Goto 2004], [Dittmar 2007], [Ryynänen 2007], [Salamon 2009]
- Common **presumptions** concerning the **bass line**
- **General processing steps**
  - Down-sampling / source separation
  - Spectral estimation (STFT, IF, CQT)
  - Note event detection
  - f_0-tracking & note modeling
- **Evaluation**
  - RWC (best F-measure: 0.59) & non-published datasets
  - No work with focus on electric bass guitar

Music Transcription → Proposed Approach



Music Transcription → Proposed Approach



Fraunhofer IDMT

Music Transcription → Proposed Approach

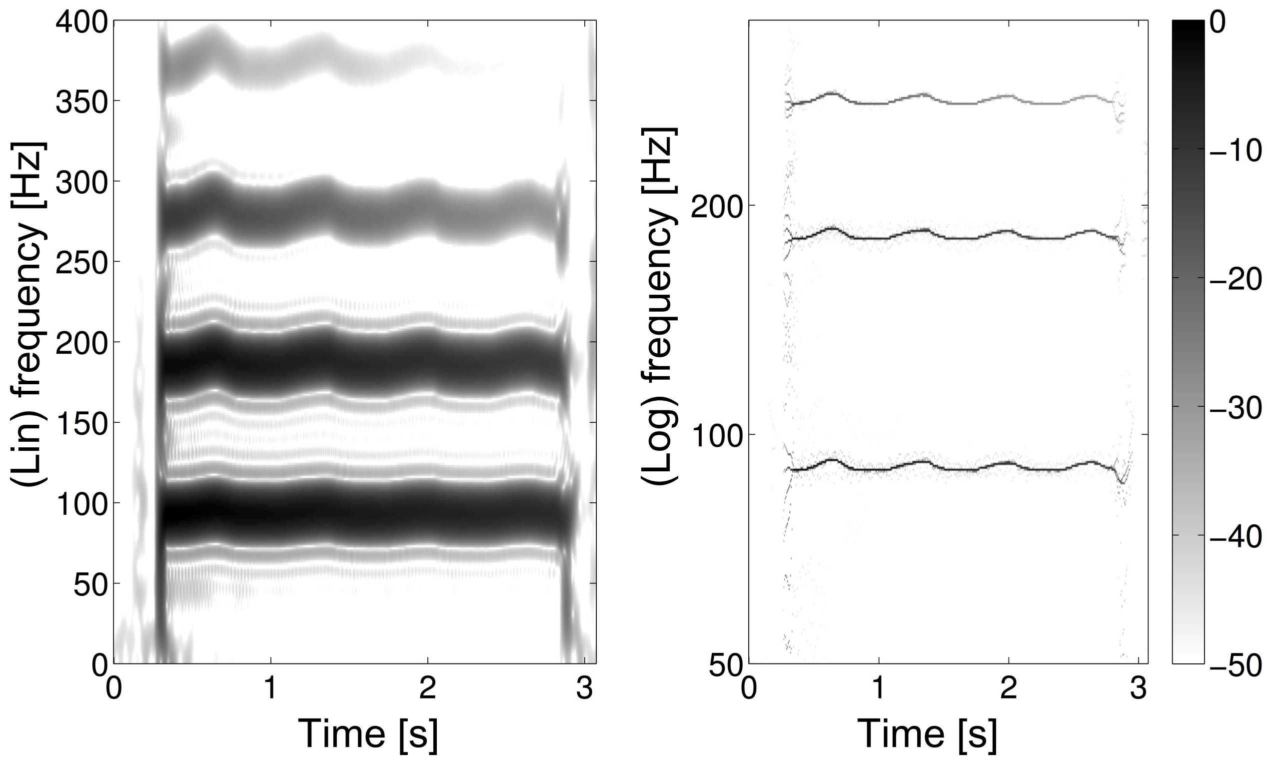
- **Pre-processing**
  - Down-sampling to 5.5 kHz, stereo → mono
- **Spectral Estimation**
  - **Short-time Fourier Transform (STFT)**
    - Linear frequency axis
    - Spectral leakage → limited frequency resolution
  - **Reassigned instantaneous frequency (IF) spectrogram**
    - Logarithmic frequency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

Fraunhofer IDMT

Music Transcription → Proposed Approach

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

- STFT

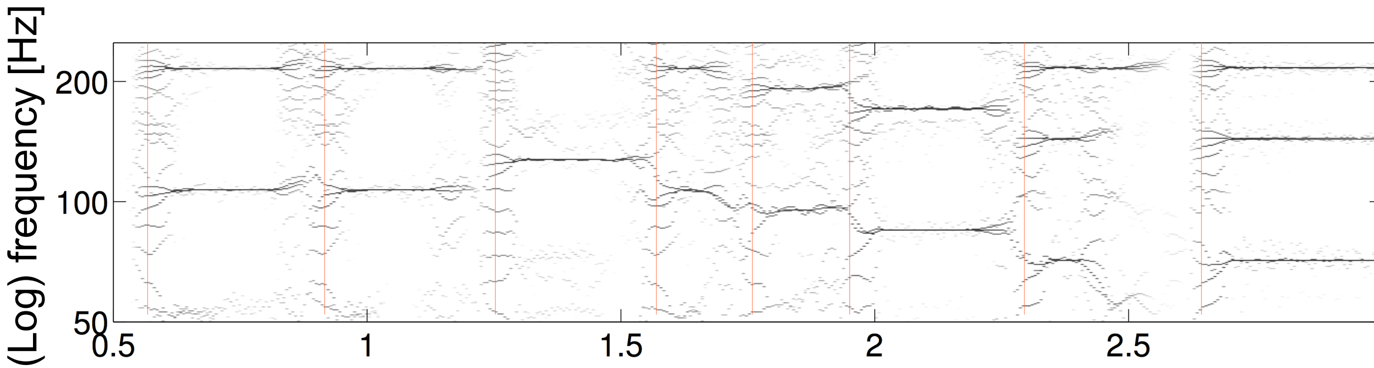


- Reassigned IF

Music Transcription → Proposed Approach

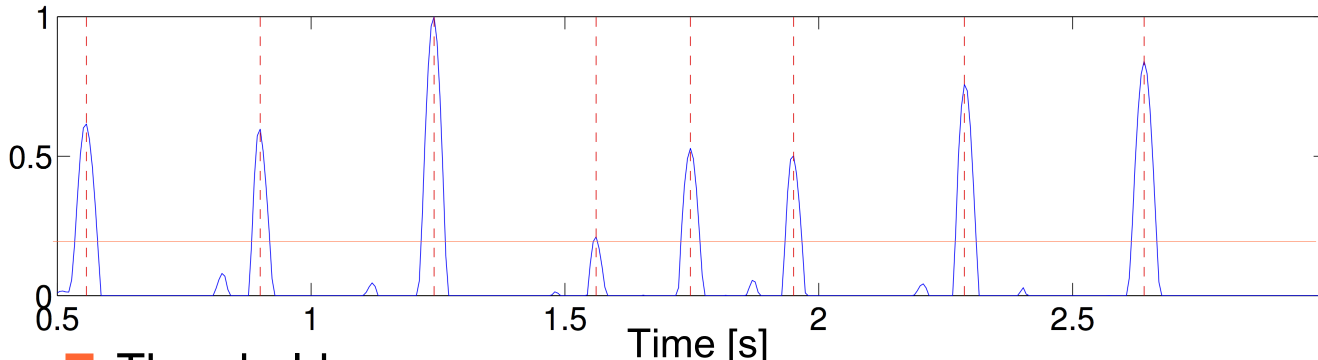
- **Score Parameter Estimation**

- IF spectrogram



- Note onsets

- Onset detection functions

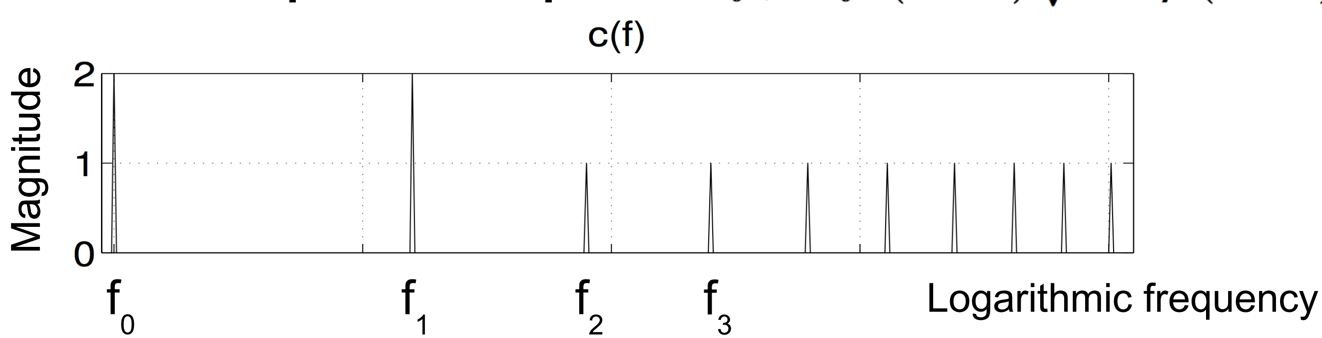


- Threshold

Fraunhofer IDMT

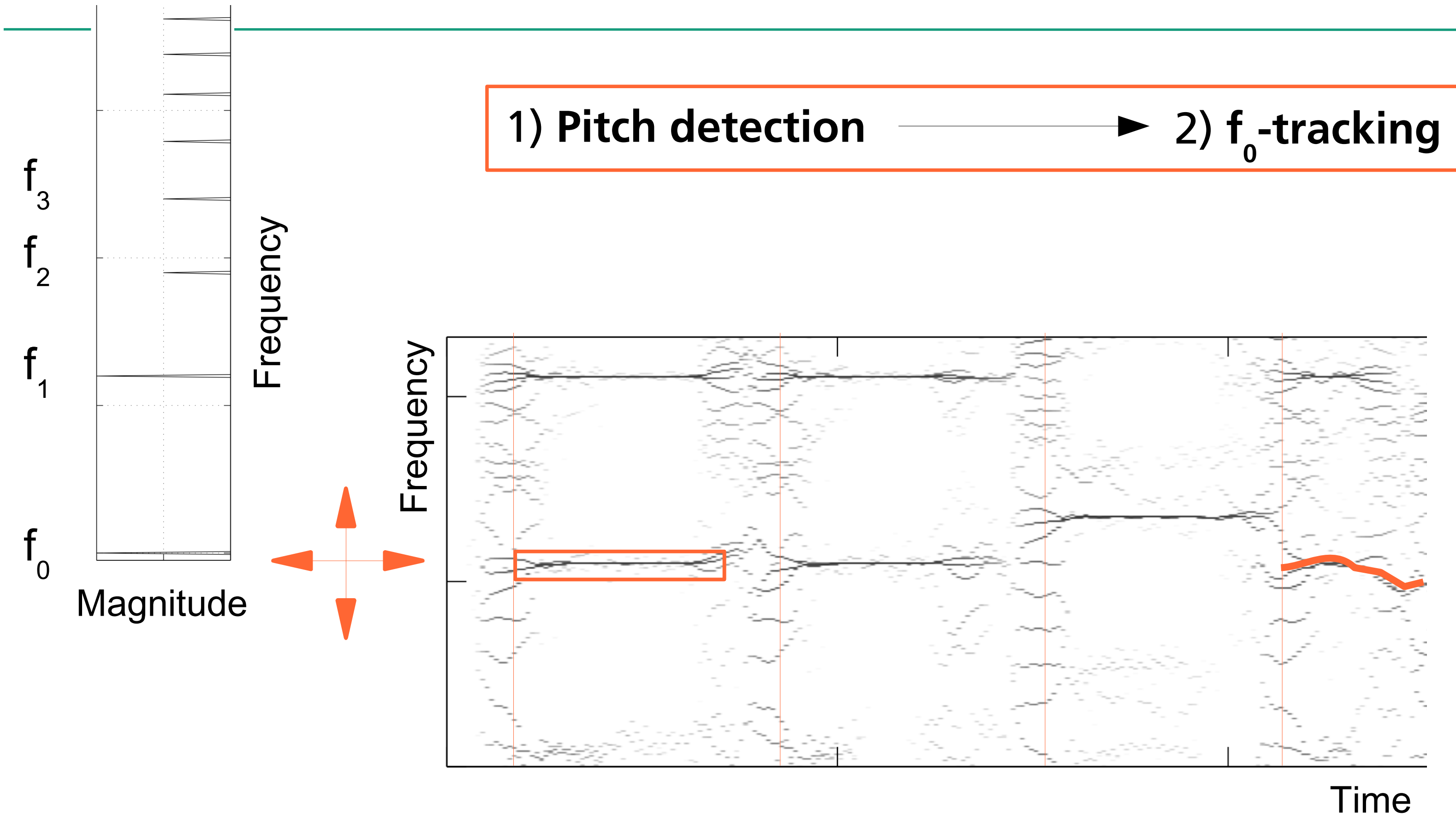
Music Transcription → Proposed Approach

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



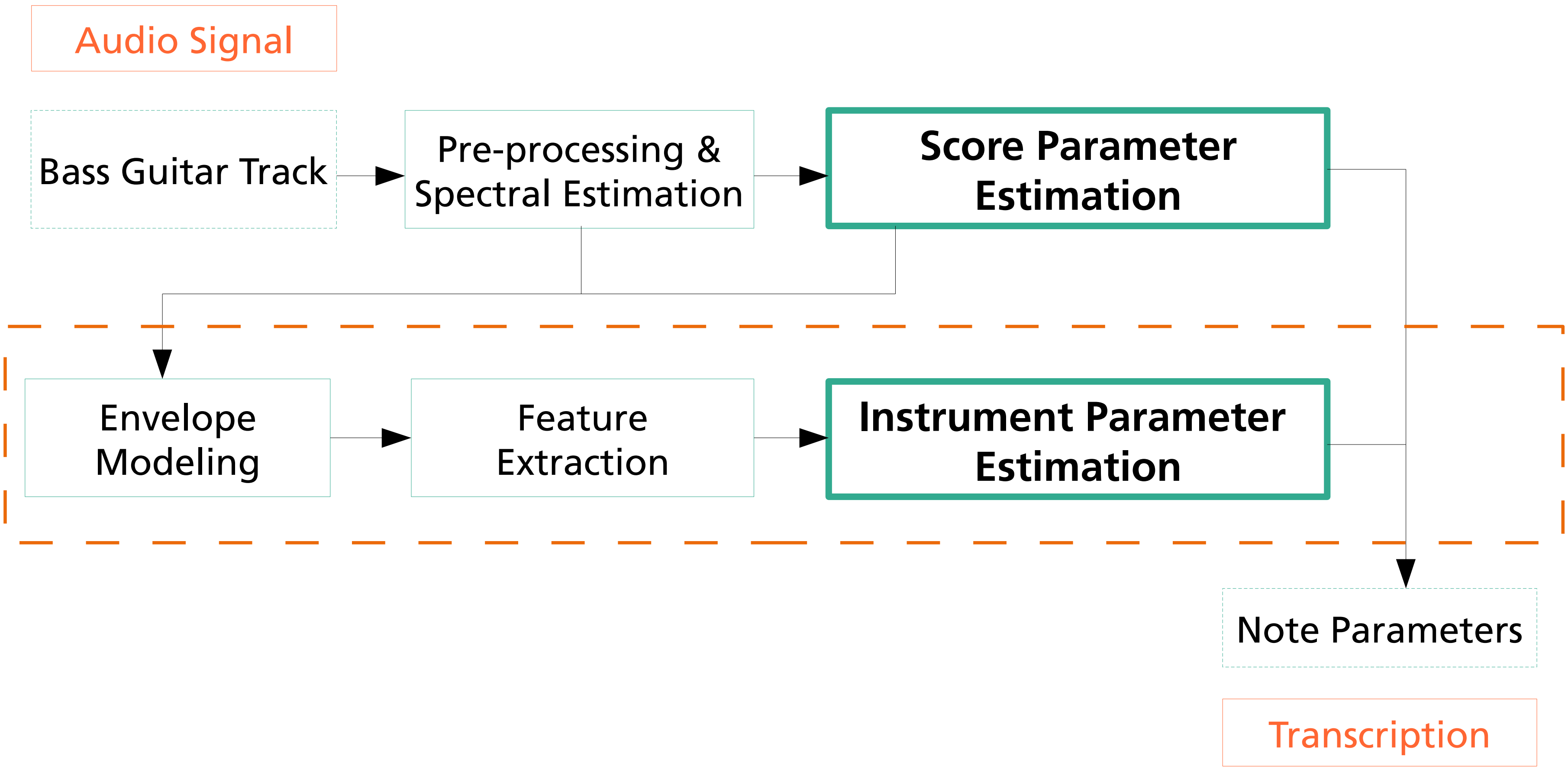
  - Maximize **cross-correlation** (**monophony** assumption)
- **Fundamental frequency (f_0) tracking**
  - **Frame-wise tracking** (**continuity** assumption)

Music Transcription → Proposed Approach



- Spectral template
- IF Spectrogram

Music Transcription → Proposed Approach



Music Transcription → Proposed Approach

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

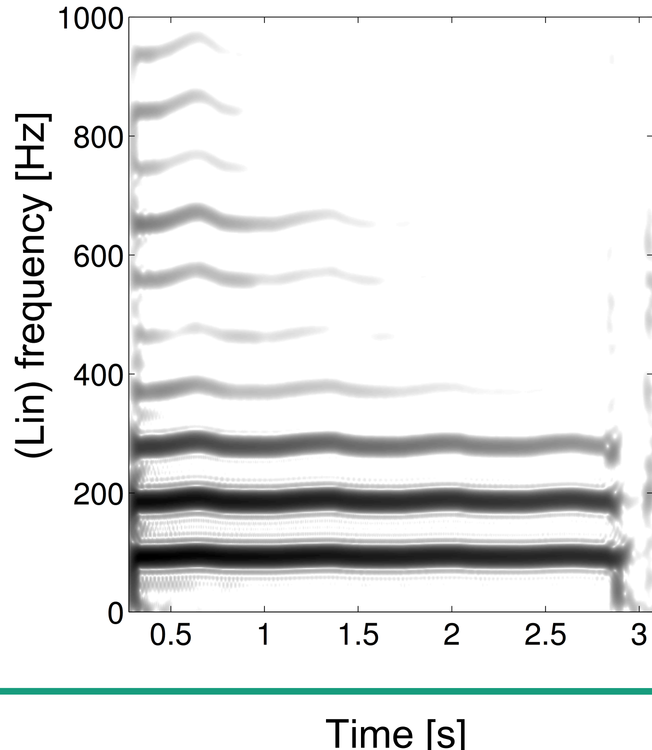
$$f_k \approx f_0(k+1)\sqrt{1+\beta(k+1)^2}$$

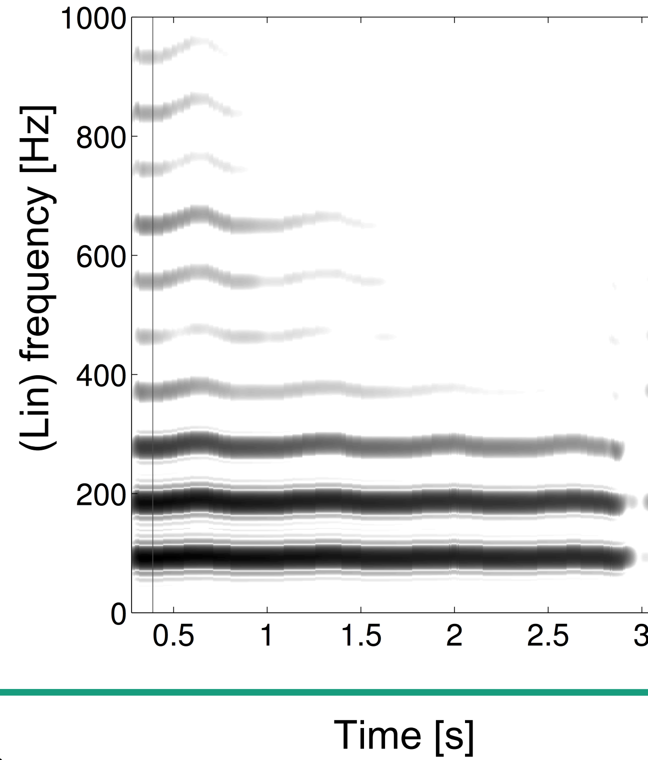
    - Wideband attack transients are not modeled
  - **Frame-wise** parameter estimation

Music Transcription → Proposed Approach

- **Example** (bass note with vibrato)

  - STFT magnitude spectrogram
  - Modeled spectrogram

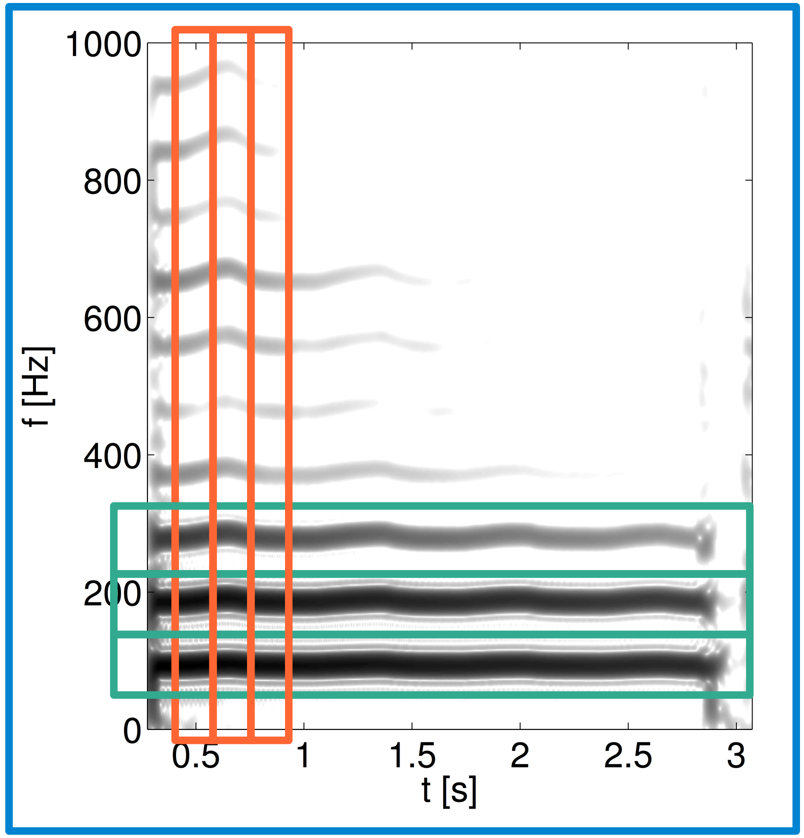




Fraunhofer IDMT

Music Transcription → Proposed Approach

- **Feature extraction**
  - **Frame-wise**
  - **Note-wise**
  - **Envelope-wise**
- **Goal**
  - Describe sound (timbre)



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 of f_0

Music Transcription → Proposed Approach

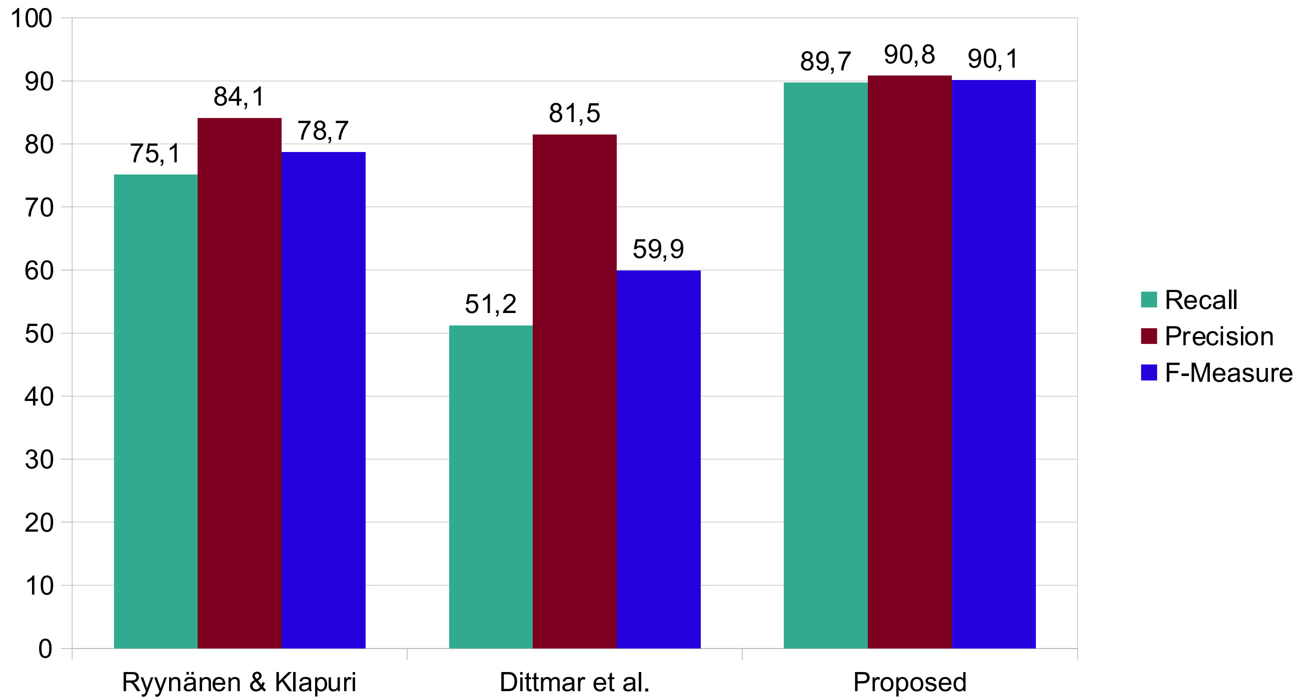
- **Classification** of **Plucking Style & Expression Style & String Number**
  - Machine learning approach (e.g., Support Vector Machine – SVM)
- **Estimation** of **Fret Number**
  - Derived from string number & pitch & string tuning
- **Context-based Error Correction**
  - Most note pitches cannot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

Music Transcription → Evaluation

- **Published Datasets**
  - **DS 1 → IDMT-SMT-BASS** (2010)
    - Isolated bass guitar notes, ~ 4300 notes
  - **DS 2 → IDMT-SMT-BASS-SINGLE-TRACKS** (2013)
    - 17 bass-lines recorded with bass guitar, ~1000 notes
  - Cover all 11 playing techniques & various fretboard positions

Music Transcription → Evaluation

- Estimation of **score-level parameters (onset & pitch) → DS 2**
  - Ryynänen & Klapuri (2007), Dittmar et al. (2007)



Fraunhofer IDMT

Music Transcription → Evaluation

- Estimation of **instrument-level parameters**

Parameter (# Classes)	**Isolated Notes (DS 1)**	**Isolated Bass Guitar Tracks (DS 2)**
Plucking Style (5)	93,3 %	63,6 %
Expression Style (5)	95,6 %	44,2 %
String Number (4)	93,0 %	75,2 %

- **Problem**
  - DS2 contains shorter notes & more variations w.r.t. playing styles

Music Transcription → Summary

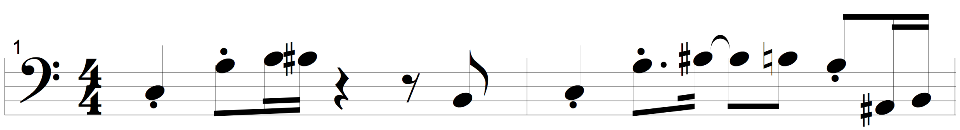
- **Main contributions**
  - Algorithm for **transcription** of **isolated bass guitar tracks**
    - Estimation of score-level & instrument-level parameters
  - **2 novel datasets** with bass guitar recordings (notes & bass lines)
    - Published as evaluation benchmarks to the scientific community

Outline

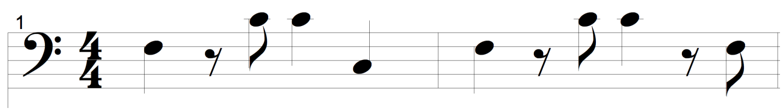
- Motivation
- Thesis Structure
  - Music Transcription
  - **Music Genre Classification**
  - Audio Coding & Sound Synthesis
- Summary

Music Genre Classification → Goals

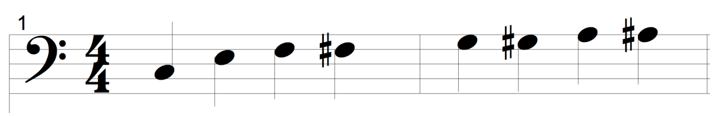
- **Music Genre Classification**
  - Bass line → repetition (**bass pattern**) & variation
  - Genre classification possible **only** using bass patterns?
  - **Examples**



→ **Funk**



→ **Bossa Nova**



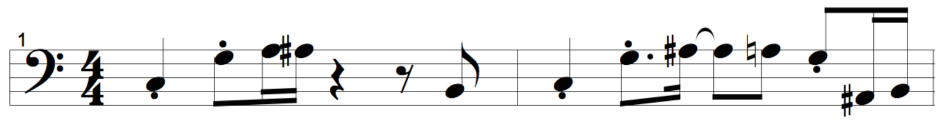
→ **Swing**

Music Genre Classification → Proposed Approach

- **Analysis Steps**
  - Convert bass pattern into different **representations** (rhythm, tonality)
  - Extract **score-based features** (rhythm, tonality, structure)
  - **Classify** the music genre
    - based on **bass pattern similarity**
    - based on **statistical pattern recognition** (classification)
    - based on **decision trees** (rule-based classification)

Music Genre Classification → Proposed Approach

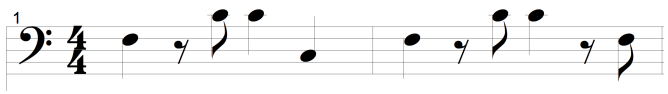
- **Example (funk bass pattern, tonal note representations)**



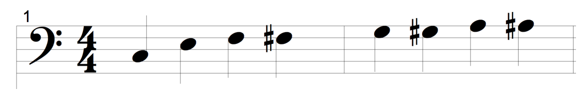
Note number	i	1	2	3	4	5	6	7	8	9	10	11	12
Note name		C_3	G_3	A_3	$A\sharp_3$	B_2	C_3	G_3	$A\sharp_3$	A_3	G_3	$A\sharp_2$	B_2
Absolute pitch	$\mathcal{P}(i)$	48	55	57	58	47	48	55	58	57	55	46	47
Pitch class	$\mathcal{P}_{12}(i)$	0	7	9	10	11	0	7	10	9	7	10	11
Relative pitch	$\mathcal{P}_{\Delta}(i)$	7	2	1	-11	1	7	3	-1	-2	-9	1	-

Music Genre Classification → Proposed Approach

- **Example (tonal features)**
  - Ambitus (pitch range)

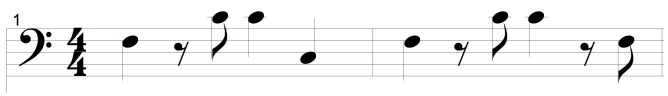


F = 12

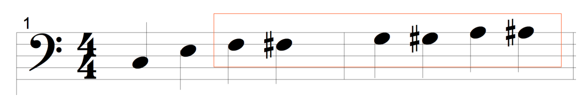


F = 10

  - Ratio of chromatic note transititions



F = 0



F = 5/7

- Bossa nova
- Swing

Fraunhofer IDMT

Music Genre Classification → Proposed Approach

- **Classification**
  - Bass patterns of the same genre → same class label
  - Pattern Similarity
    - Compare bass patterns using distance measures
  - Statistical Pattern Recognition & Decision Trees
    - Train classification algorithm
    - Bass patterns are represented as features

Music Genre Classification → Evaluation

- **Novel Dataset**
  - **IDMT-SMT-BASS-GENRE-MIDI**
    - 520 bass patterns & 13 music genres

Blues	Nineties Rock
Bossa Nova	Reggae
Funk	Salsa & Mambo
Forró	Seventies Rock
Hip-Hop	Swing
Minimal Techno	Zouglou
Motown	

Music Genre Classification → Results

- Best results (64.8 % mean class accuracy)

	BLU	BOS	FOR	FUN	HIP	MIN	MOT	REG	RON	ROS	SAL	SWI	ZOU
BLU	**70**	0	0	5	0	0	10	7.5	5	0	2.5	0	0
BOS	0	**67.5**	12.5	0	0	5	2.5	0	0	2.5	0	0	10
FOR	2.5	10	**77.5**	0	0	0	2.5	0	0	0	5	0	2.5
FUN	2.5	0	0	**77.5**	0	2.5	10	0	0	2.5	2.5	0	2.5
HIP	7.5	2.5	0	5	**50**	0	0	10	0	2.5	12.5	2.5	7.5
MIN	5	0	0	0	0	**60**	7.5	0	2.5	5	10	0	10
MOT	5	0	0	0	0	0	**72.5**	2.5	2.5	10	2.5	0	5
REG	10	2.5	0	2.5	2.5	0	2.5	**55**	0	0	7.5	0	17.5
RON	17.5	7.5	0	0	0	0	5	17.5	**35**	5	10	0	2.5
ROS	10	0	0	5	0	0	5	7.5	12.5	**45**	10	0	5
SAL	5	0	0	5	0	0	0	2.5	2.5	2.5	**65**	0	17.5
SWI	0	0	0	0	0	0	2.5	0	2.5	0	2.5	**92.5**	0
ZOU	2.5	0	0	2.5	2.5	0	5	2.5	0	5	5	0	**75**

Music Genre Classification → Summary

- **Main contributions**
  - **Score-based audio features** to describe tonal, rhythm, and structural properties of bass patterns
  - Novel **genre classification strategy** based on **bass pattern similarity**
  - **Novel dataset** with 520 bass patterns from 13 music genres

Outline

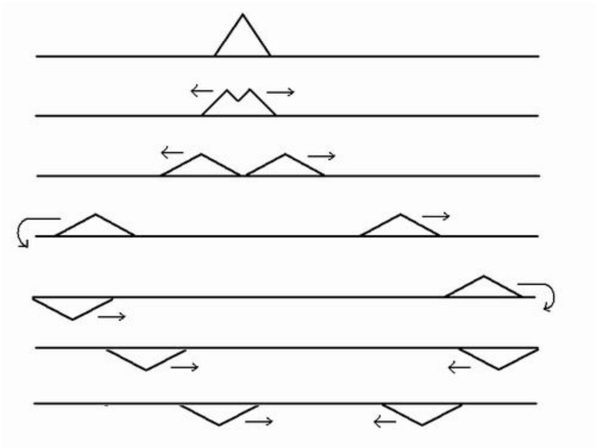
- Motivation
- Thesis Structure
  - Music Transcription
  - Music Genre Classification
  - **Audio Coding & Sound Synthesis**
- Summary

Audio Coding & Sound Synthesis → Goals

- **Audio Synthesis**
  - Model the **sound production** of the electric bass guitar
  - **Incorporate instrument-level parameters** (string number, playing techniques)
- **Parametric Audio Coding**
  - Represent **audio track** as **sequence of note events & parameters**
  - Which **acoustic parameters** are important for natural synthesis?
  - Strong data rate reduction

Audio Coding & Sound Synthesis → Proposed Approach

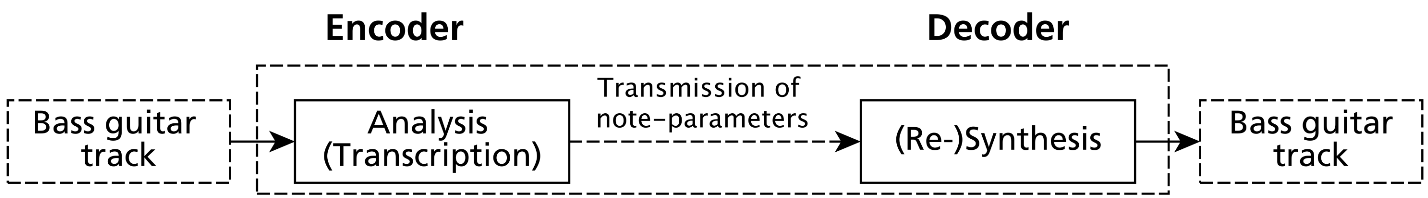
- Existing **physical modeling** algorithm for string instruments [Smith 1992]
  - Superposition of two traveling waves
  - Reflections at string ends



- Model extensions
  - Various **excitation functions** (different plucking styles)
  - **String damping** (muted, harmonics, dead-notes)
  - **String-fretboard collision** (slap techniques)
  - **Inharmonicity** (thick bass guitar strings)

Audio Coding & Sound Synthesis → 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

Fraunhofer IDMT

Audio Coding & Sound Synthesis → Proposed Approach

- **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 & 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

Audio Coding & Sound Synthesis → Summary

- **Main contributions**
  - **Extension** of existing **physical modeling algorithm** for bass guitar synthesis
    - **Incorporation** of various aspects of **sound production**
  - **Parametric audio coding scheme** for **isolated bass guitar recordings**
  - **Listening tests** to investigate
    - **Perceptual quality** of **proposed synthesis algorithm**
    - **Relevance** of different **acoustic parameters**

Outline

- Motivation
- Thesis Structure
    - Music Transcription
    - Music Genre Classification
    - Audio Coding & Sound Synthesis
- **Summary**

Summary

- **Publications**
  - 8 conference papers
  - 1 journal article
  - 2 book chapters

Summary

- **Limitations**
  - Score generation instead of piano roll
  - Extraction of bass pattern from bass line
  - Evaluation of transcription of non-isolated bass guitar tracks
  - Influence of amplifier & audio effects
  - Evaluation of genre classification including transcription errors
  - Incorporate instrument-level parameters for genre classification
  - What is missing to avoid “synthetic sound“?

Summary

- **Thank you!**
  - Reviewers
  - Family
  - Friends
  - Colleagues / Students

References

- **[Hainsworth 2001]** Stephen W. Hainsworth and Malcolm D. Macleod, "*Automatic bass line transcription from polyphonic music*", in Proceedings of the International Computer Music Conference (ICMC), pp. 431–434, La Habana, Cuba, 2001.
- **[Goto 2004]** Masataka Goto, "*A Real-Time Music-Scene-Description System - Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals*", Speech Communication, vol. 43, no. 4, pp. 311–329, Sept. 2004.
- **[Dittmar 2007]** Christian Dittmar, Karin Dressler, and Katja Rosenbauer, "*A Toolbox for Automatic Transcription of Polyphonic Music*", Proceeding of the Audio Mostly Conference, pp. 58–65, 2007.
- **[Ryynänen 2007]** Matti Ryynänen and Anssi Klapuri, "*Automatic Bass Line Transcription from Streaming Polyphonic Audio*", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'), pp. 1437–1440, Honolulu, Hawaii, USA, 2007
- **[Salamon 2009]** Justin Salamon and Emilia Gómez, "*A Chroma-Based Salience Function for Melody and Bass Line Estimation From Music Audio Signals*", in Proceedings of the 6th Sound and Music Computing Conference (SMC), pp. 23–25, Porto, Portugal, 2009.
- **[Smith 1992]** Julius O. Smith, "*Physical Modeling Using Digital Waveguides*", Computer Music Journal, vol. 16, no. 4, pp. 74, 1992

Thank you! Questions?

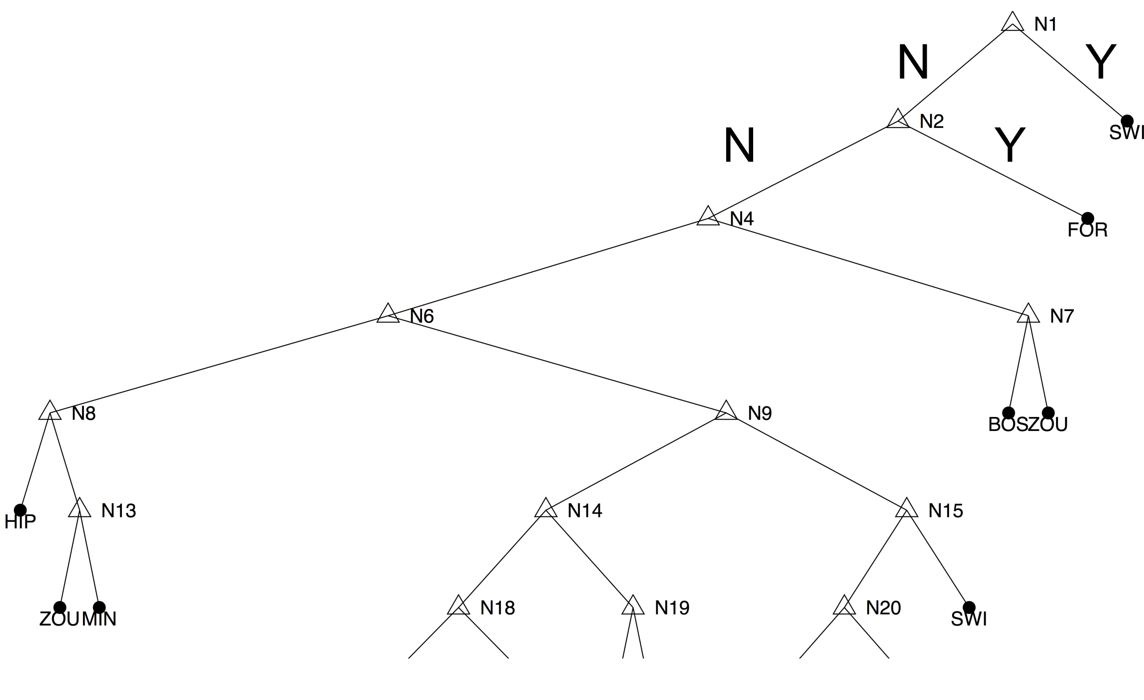
Dipl.-Ing. Jakob Abeßer

Wissenschaftliche Aussprache

Ilmenau, 18.09.2014

ADDITIONAL SLIDES

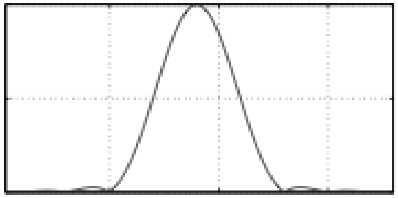
Music Genre Classification – Decision Tree



- N1: p(D=1/4) > 0.95
- N2: Tempo < 202 bpm

Music Transcription → New Approach

- **Envelope Modeling**
  - Simple **parametric model**



$$X(f,t) \approx \sum_{k=0}^{N_{\mathrm{Harm}}} a_k(t)\, h_X\left(f - f_k(t)\right)$$

    - $f_k \approx f_0(k+1)\sqrt{1+\beta(k+1)^2}$
    - Wide-band characteristics are not modeled
  - **Frame-wise parameter estimation** using Expectation-Maximization (EM) algorithm → more robust than simple partial tracking
  - Estimated **model parameters**
    - Harmonic magnitudes, fundamental frequency
    - Inharmonicity coefficient

Fraunhofer IDMT

Music Transcription → Evaluation

- Estimation of **instrument-level parameters** (bass lines, confusion matrices)

PS

PS (correct) \ PS (classified)	FS	MU	PK	SP	ST
FS	37.6	62.4	0.0	0.0	0.0
MU	0.0	100.0	0.0	0.0	0.0
PK	0.0	39.9	60.1	0.0	0.0
SP	0.0	0.0	0.0	81.6	18.4
ST	0.0	0.0	0.0	61.1	38.9

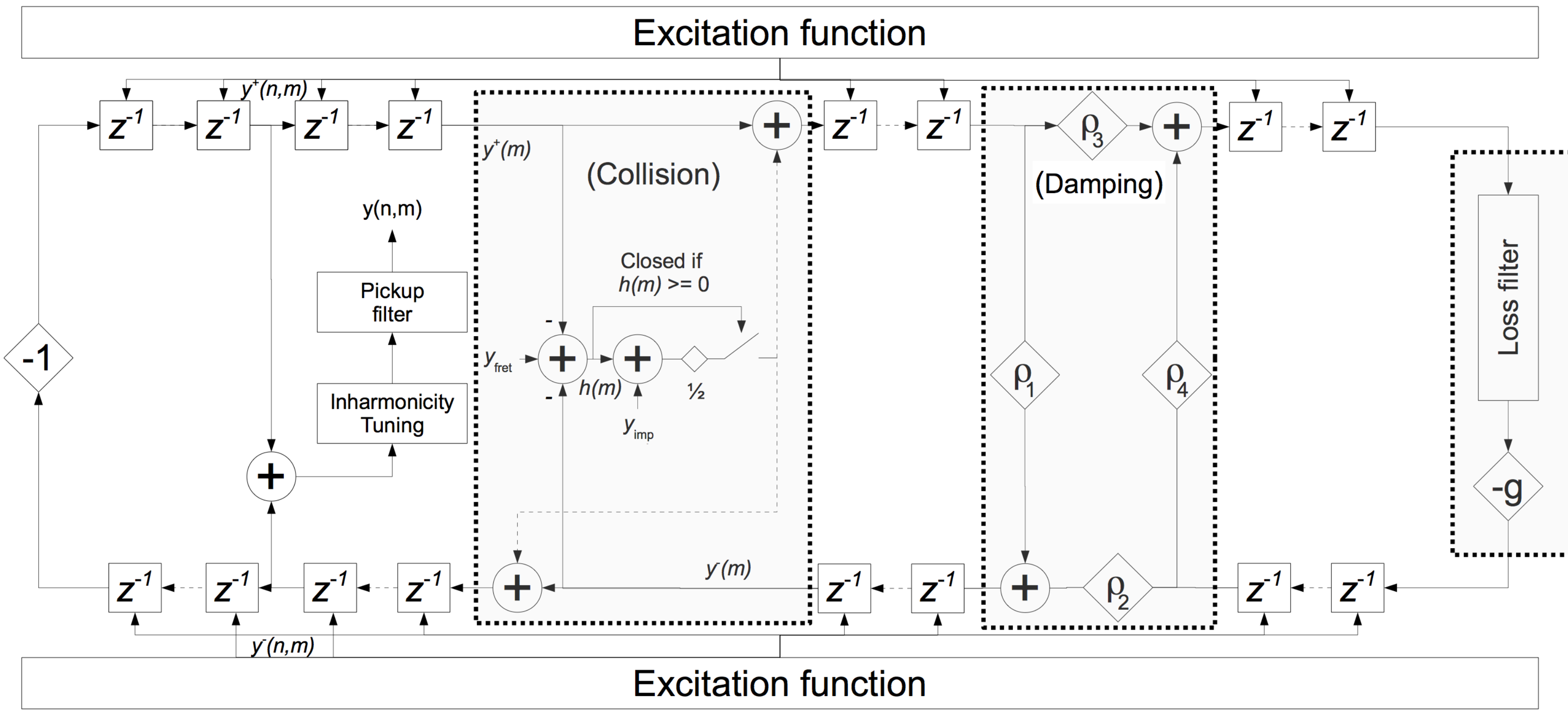
ES

ES (correct) \ ES (classified)	NO	BE	VI	HA	DN	SL
NO	0.3	54.9	17.2	2.3	0.3	25.0
BE	0.0	87.5	12.5	0.0	0.0	0.0
VI	0.0	41.9	51.6	0.0	0.0	6.5
HA	16.1	19.4	9.7	48.4	0.0	6.5
DN	11.1	0.0	11.1	55.6	22.2	0.0
SL	0.0	35.0	10.0	0.0	0.0	55.0

SN

SN (correct) \ SN (classified)	1	2	3	4
1	98.8	1.2	0.0	0.0
2	15.2	68.9	15.6	0.4
3	1.0	19.4	66.0	13.6
4	0.0	0.0	33.0	67.0

Audio Coding & Sound Synthesis → New Approach



Fraunhofer IDMT