Automatic Recognition and Parametrization of Frequency Modulation Techniques in Bass Guitar Recordings

Jakob Abeßer[1], Christian Dittmar[1], and Gerald Schuller[2]

[1]*Semantic Music Technologies Group, Fraunhofer IDMT, Ilmenau, Germany*

[2]*Ilmenau University of Technology, Ilmenau, Germany*

Correspondence should be addressed to Jakob Abeßer (`abr@idmt.fraunhofer.de`)

ABSTRACT

In this paper, we propose a novel method to parametrize and classify different frequency modulation techniques in bass guitar recordings. A parametric spectral estimation technique is applied to refine the fundamental frequency estimates derived from an existing bass transcription algorithm. We apply a two-stage taxonomy of bass playing styles with special focus on the frequency modulation techniques slide, bending, and vibrato. An existing database of isolated note recordings is extended by approx. 900 samples to evaluate the presented algorithm. We achieve comparable classification accuracy values of 85.1% and 81.5% for classification on class-level and subclass-level. Furthermore, two potential application scenarios are outlined.

1. INTRODUCTION

1.1. Motivation

In different music styles, frequency modulation techniques such as pitch-bends, slides, or vibrato are widely used as expressive performance gestures on string instruments like the bass guitar or the electric guitar. These techniques allow musicians to perform more or less subtle micro-tonal variations of the fundamental frequency of each played note and so to expressively vary the melody. Improved automatic music transcription (ATM) methods, which are capable to recognize and parametrize different frequency modulation techniques, are essential for a detailed stylistic description of a musical performance. Potential applications of these methods are artist classification tasks, music education software, and instrument-based audio coding as illustrated in Sect. 1.2.

In this paper, we continue our work from [1], where 10 different bass guitar playing techniques including 5 plucking styles and 5 expression styles have been investigated. We now aim to find a more detailed parametrization of the fundamental frequency trajectories over time that allows to define meaningful sub-classes of the expression styles *vibrato*, *bending* and *slide*. This paper's contribution is a novel approach for automatic music transcription with special focus on lower frequency regions as well as on short-time fluctuations of the fundamental frequency. Throughout the paper, we work on isolated bass guitar recordings with no temporal or spectral overlap of concurrent instruments such as percussion or harmony instruments.

1.2. Application scenarios

In this section, two possible application scenarios for the proposed classification and parametrization method will be detailed.

1.2.1. Instrument-based audio coding

An interesting application for the presented parameter estimation is to use it in a musical instrument coder, i.e., an audio compression technique that is optimized with respect to the source of the audio material. If enough parameters are estimated, then they can be used to reproduce the original instrument sound, or at least a sound with a comparable character. For the sound reproduction a suitable model, using these parameters, is needed. There are many sophisticated models of instrument sound production, but often their parameters cannot be readily extracted or estimated from the sound of the instrument. This is essential for usage with a musical instrument coder. That is an advantage of the parameters we present in this paper, they can be readily extracted directly from the instrument sound. Here it would also be interesting to see how much of the characteristic instrument sound is captured by our parameters. This is best estimated by trying to reproduce the sound using these parameters. Instrument coders could also have other applications. Since the parameters they transmit have a musical meaning, any quantization error results not in external sounding (and hence obvious) artefacts, but rather in a slight change of the character of the sound of the instrument or the playing style, which makes it less obvious, more natural sounding. Also, these parameters could be used to modify an instrument sound in the post-production process, or even on the consumer side, which gives new freedom in customizing musical recordings.

1.2.2. Music pedagogical applications

An alternative application is software for musical education, especially with regards to learning a musical instrument. Children, adolescents and adults must be constantly motivated to practice and complete learning units. Traditional forms of teaching and even current E-Learning systems are often unable to provide this motivation. On the other hand, music-based video games are immensely popular, but they fail to develop skills which are transferable to musical instruments. In the *Songs2See*[1] project [7], a software application for practicing musical skills is developed, striving at the motivation of game playing. The software provides immediate feedback from the real-time transcription of what the student is playing on his instrument. In addition to conventional score notation, a piano roll visualization guides the student. This allows for much more visual feedback regarding fine-grained details like expression, vibrato, pitch and articulation. Therefore, reproducing the extracted parameters on one's instrument can be used as an additional skill to master in a gaming like music exercise.

1.3. Outline

This paper is organized as follows. In Sect. 2, we outline the goals and challenges of this publication. A brief overview over related work is given in Sect. 3. In Sect. 4, we explain our algorithm for spectral analysis and the subsequent parametrization of different frequency modulation techniques. We discuss the obtained results in Sec. 6 and give a conclusion in Sect. 7.

[1] http://www.songs2see.net

2. GOALS & CHALLENGES

In order to classify and parametrize the applied frequency modulation technique, we first need a precise frame-wise estimate $f_0[n]$ of the fundamental frequency over all frames of a given note. Since we focus on the bass guitar, a low frequency range between 41.2Hz up to 220.0Hz needs to be investigated for potential f_0-candidates[2]. One big challenge is the choice of an appropriate spectral estimation method since classic non-parametric techniques such as the FFT suffer from the spectral leakage problem. This problem limits the achievable frequency resolution and is especially impeding in lower frequency bands.

3. PREVIOUS APPROACHES

Automatic music transcription aims at translating real audio recording into a score-like symbolic representation of instrument tracks and their corresponding note events. Thus, it is an important pre-requisite for an semantic analysis of recorded music. In this paper, we focus on the transcription of the bass instrument which is usually considered to be the lowest melody voice in a piece of music. Previously presented algorithms for the transcription of the bass line such as in [8], [6], [13], and [15] are capable to extract basic score parameters such as note onset, note duration, note pitch and note loudness. This representation of an instrument track neglects aspects of timbre, micro-tonal intonation and micro-timing since it leads to note events quantized both in time (when musical time representations are used) and in frequency (since one obtains discrete note pitches). Many publications have studied different performance gestures to extract expressive parameters for the guitar for a realistic instrument synthesis [3], [10], [16]. Järveläinen investigated the perception of two typical frequency modulations in guitar tones – initial pitch glides and vibrato [9]. Furthermore, pitch bends have been studied for clarinet synthesis in [14] and playing techniques such as the vibrato have been extracted from violin recordings using a rule-based approach in [2].

4. NEW APPROACH

[2] This corresponds to the most commonly used bass guitar tuning E2, A2, D3, G3 and a fret range up to the 14th fret position.

Classes		**Sub-classes**	
Slide	**SL**	Slide up	**SLU**
		Slide down	**SLD**
Vibrato	**VI**	Fast vibrato	**VIF**
		Slow vibrato	**VIS**
Bending	**BE**	Semi-tone bending	**BES**
		Quarter-tone bending	**BEQ**
No	**NO**	-	-

Table 1: Taxonomy of frequency modulation techniques.

In this paper, we interpret music transcription as a process including both a *parameter estimation* step and *classification* step. After we estimate the commonly-used note parameters pitch, onset, duration, loudness, we aim to classify the applied *frequency modulation technique*, which can be applied as an expressive gesture after a note is plucked.

4.1. Taxonomy

We use a two-layered taxonomy as illustrated in Tab. 1 to capture the most commonly applied frequency modulation techniques in bass-guitar performances. This taxonomy is an extension of the list of *expression techniques* as presented in [1]. The temporal progression of the fundamental frequency has many degrees of freedom such as the modulation range and the modulation frequency. Three examples are illustrated in Fig. 2. Thus, such a taxonomy can generally be defined in manifold ways.

For each of the frequency modulation techniques vibrato, bending, and slide, we assume that the main degree of freedom can be reduced to one parameter, e.g. the modulation range for bending and slide and the modulation frequency for vibrato. This simplification allows to define two sub-classes for each class as given in Tab. 1. The fourth class **NO** is added for completeness to evaluate the classification of notes with and without applied frequency modulation techniques in the experiments.

4.2. Algorithm

In this section, we explain the different stages of the analysis algorithm illustrated in Fig. 1.

4.2.1. Harmonic spectrum detection

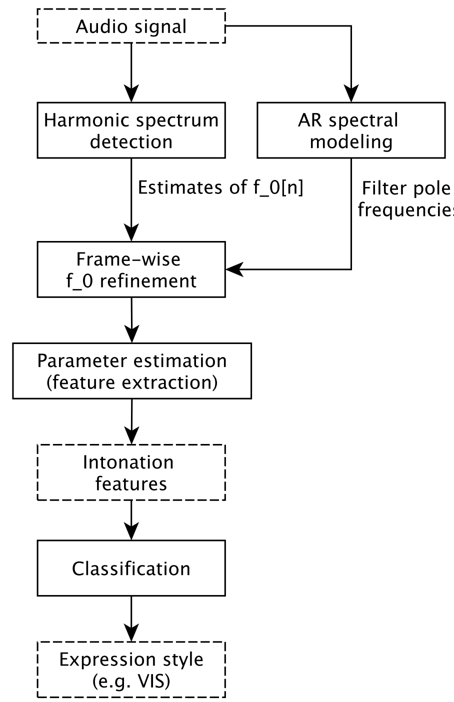
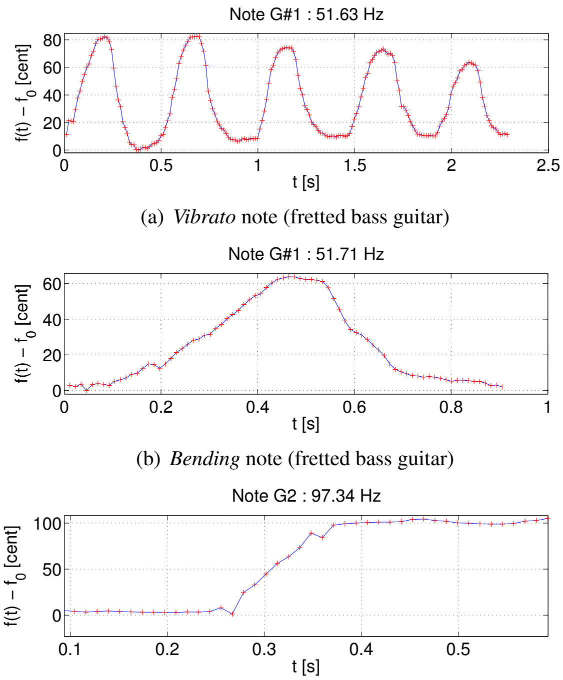


Fig. 1: Proposed algorithm.

In order to identify f_0-candidates of bass notes, the bass transcription as described in [4] is used. The audio signal to be analyzed for bass notes is sub-sampled by factor 32, thus reducing the Nyquist frequency to approximately 700 Hz. It can safely be assumed, that the majority of bass notes and their most important harmonic overtones are captured in that frequency range. A zero-phase low-pass filter is applied before the sub-sampling to avoid aliasing. This well-known filter technique virtually doubles the filter order and has precisely zero phase distortion. The time-frequency transform is an STFT with additional estimation of the instantaneous frequency (IF) spectrogram computed from successive phase spectra via the well-known phase vocoder method [5]. This method computes the evolution of the instantaneous frequency per spectrogram bin, and is thus much more suitable to analyze the low-frequency content of a music signal, where slight frequency deviations may cause erroneous note estimation. In parallel to the frequency transform, a time-domain based onset detection is conducted on the low-pass filtered and sub-sampled signal.



(a) *Vibrato* note (fretted bass guitar)

(b) *Bending* note (fretted bass guitar)

(c) *Slide* note (fretless bass guitar)

Fig. 2: Estimated f_0-course for different modulation techniques.

The detection of candidate note onsets is realized as conventional envelope extraction based on two-wave rectification and low-pass smoothing. It is followed by differentiation of the envelopes and detection of rising slopes exceeding a dynamic threshold criterion. A detailed spectral analysis is deployed in the subsequent stage to overcome the common problem that drum instruments might overlap with the bass notes in the low-frequency range of music signals.

The main issue is that drum instruments sounding simultaneously to a bass note can have a negative influence on the f_0-candidates estimation of this note. To prevent this behavior, a frame-wise criterion for the reliability of the low-frequency spectrum is computed as a combination of the spectral flatness measurement and the peak energy of the spectrum's autocorrelation function (ACF). The motivation for the latter measurement is based on the observation that drum spectra will lead to a less spiky shape of the ACF than harmonic spectra, whose overtones cause salient periodicities along the frequency axis. This property is exploited to retrieve the temporal part of a bass note, where the harmonic spectrum is most salient and least superposed with drum spectra. Inside that part, the strongest peak of the spectrum's autocorrelation function serves as a candidate for the fundamental frequency.

4.2.2. AR spectral modeling

Non-parametric spectral estimation methods such as the Periodogram make no explicit assumption on the type of signal that is analyzed. In order to obtain a sufficiently high frequency resolutions for a precise f_0-detection, large time frames of data samples are needed (to compensate the spectral leakage effect), which is inherent in windowing the signal in the time domain.

It is well known from music acoustics, that notes played on a plucked string instrument such as the bass guitar consist of a very short attack part (approx. 20-40 ms) and a longer decay part of a harmonic structure. In contrast to the percussive nature of its attack part, the decay part of a note can be modeled by a sum of decaying sinusoidal components bearing a nearly perfect harmonic relationship[3]. Parametric estimation techniques can be applied if the signal can be assumed to be generated by a known model. In our case, the power spectral density (PSD) $\Phi(z)$ can be modeled by an auto-regressive (AR) filter such as

$$\Phi_{AR}(z) = \sigma^2 \frac{1}{1+\sum_{k=1}^{p} a_k z^{-k}} \tag{1}$$

with σ^2 denoting the process variance, p denoting the model order, and $\boldsymbol{a} \in \mathbb{R}^{p+1}$ being the filter coefficients. Since auto-regressive processes are closely related to linear prediction (LP), both a *forward* and a *backward prediction error* can be defined to measure the predictive quality of the AR filter. The *least-squares method*[4] is based on a simultaneous least-squares minimization of both prediction errors with respect to all filter coefficients $\boldsymbol{a}$. This method has been shown to outperform related algorithms such as the Yule-Walker method, the Burg algorithm, and the covariance method (See [11], Chapter 8 for more details). The size of the time frames N is only restricted by the model order as $p \leq 2N/3$.

First, the signal is down-sampled to $f_s = 5.5$ kHz. By using overlapping time frames with a block-size of $N = 256$

[3] Since the strings of the bass guitar have a certain amount of stiffness (string diameter is between .45 to 1.05mm), the known phenomenon of *inharmonicity* appears.

[4] A.k.a. *forward-backward method.* or *modified covariance method*

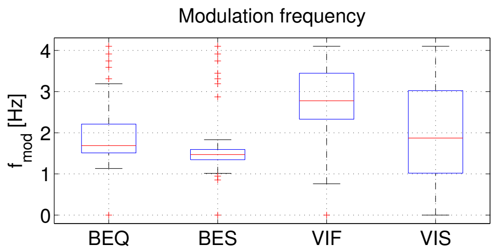


Fig. 3: Modulation frequency values within the dataset for the vibrato and bending classes.

samples (46.4 ms) and a hop-size of $H = 64$ samples (11.6 ms), we apply the spectral estimation algorithm presented in the previous section to compute estimates of the filter coefficients $\hat{\boldsymbol{a}}[n]$. The pole frequencies $\hat{\boldsymbol{f}}_{pole}[n]$ can be computed by finding the roots of the numerator in Eqn. (1) for each frame n.

4.2.3. Frame-wise f_0-refinement

From the harmonic spectrum detection step explained before, we obtain a frame-wise estimate $\hat{f}_0[n^*]$ for the fundamental frequency of a note. Since the resolution is lower for the AR spectral modeling, we extrapolate the values to obtain f_0-estimates $\hat{f}_0[n]$ with the same resolution as the extracted pole frequencies $\hat{\boldsymbol{f}}_{pole}[n]$. For each frame n, we select the closest pole frequency as fundamental frequency value $f_0[n]$.

$$f_0[n] = \hat{f}_{pole,k}[n] \quad (2)$$
$$k = \arg\min_{k^*} \left|\hat{f}_{pole,k^*}[n] - \hat{f}_0[n]\right|$$

4.2.4. Feature extraction

As illustrated in Fig. 4(a) for a *vibrato* note, we propose a temporal segmentation into segments of increasing and decreasing values of $f_0[n]$. This segmentation can be applied to all frequency modulation techniques investigated in this paper. The bending technique typically leads to two segments and the slide technique leads to one segment.

From this segmentation, we compute the number of segments N_{Seg} and the maximum modulation range

$$\Delta_{f,max} = \max f_0[n] - \min f_0[n] \quad (3)$$

as features. Furthermore, based on the first non-zero maximum in the autocorrelation function of $f_0[n]$ over time, we compute an estimate of the mean modulation frequency $\bar{f}_{mod}$ as shown in Fig. 4(b). The box-plots for the distribution of the modulation frequency values of the vibrato and bending subclasses for the samples in the dataset is given in Fig. 3.

To characterize the temporal progression of the applied modulation technique, we compute two vectors $\widehat{\boldsymbol{\Delta}}_f \in \mathbb{R}^{N_{Seg}}$ and $\widehat{\boldsymbol{f}}_{mod} \in \mathbb{R}^{N_{Seg}}$ with estimates of both the modulation range and the modulation frequency for each segment as

$$\widehat{\Delta}_{f,i} = |\max f_0[n] - \min f_0[n]| \quad (4)$$
$$n_{start,i} \leq n \leq n_{end,i}$$

and

$$\widehat{f}_{mod,i} = \frac{1}{2\left(t_{end,i} - t_{start,i}\right)}. \quad (5)$$

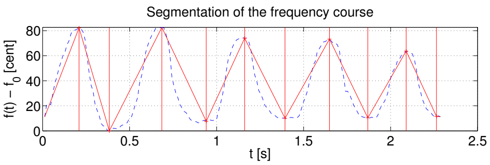
The segment boundaries are denoted by n_{start} and n_{end} in frames and by t_{start} and t_{end} in seconds. We compute the mean, the variance, the kurtosis, and the skewness over both $\widehat{\boldsymbol{\Delta}}_f$ and $\widehat{\boldsymbol{f}}_{mod}$ as features to characterize the course of $f_0[n]$. Even though the three modulation techniques vibrato, bending, and slide lead to different number of segments, using only the above-mentioned statistic descriptors as features allows to extract the same number of features independently of the applied technique.

Finally, to describe the overall pitch progression over the course of a note, we compute the mean fundamental frequency $f_{0,start}$ and $f_{0,end}$ over the first 5% and the last 5% of all note frames and use the difference $\Delta f_0 = f_{0,end} - f_{0,start}$ as feature.

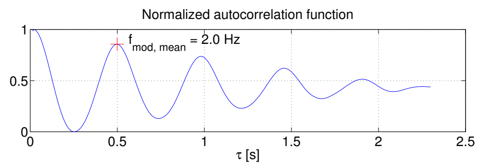
Feature name	Unit	Dim.
Number of segments N_{Seg}		1
Maximum frequency modulation range $\Delta_{f,max}$	cent	1
Mean modulation frequency $\bar{f}_{mod}$	Hz	1
Statistics descriptors of $\widehat{\boldsymbol{\Delta}}_f$		4
Statistics descriptors of $\widehat{\boldsymbol{f}}_{mod}$		4
Progression of the fundamental frequency Δf_0	cent	1

Table 2: Overview over all extracted features.

5. EVALUATION



(a) Proposed temporal segmentation. The blue dashed line illustrates the fundamental frequency course, the vertical red lines indicate the temporal segments, and the diagonal red lines connect the segment-wise minimum and maximum values of the fundamental frequency.



(b) Normalized autocorrelation function. The mean modulation frequency is derived from the first non-zero maximum.

Fig. 4: Proposed temporal segmentation of a *vibrato* note, autocorrelation function with estimated mean modulation frequency.

5.1. Data set

A novel dataset was recorded as an extension to the one presented in [1] and is intended as a public benchmark for the given task[5]. The dataset contains 156 isolated note recordings for each of the 6 new classes **SLU**, **SLD**, **VIF**, **VIS**, **BES**, and **BEQ** (see Tab. 1) with 936 samples altogether. Furthermore, we chose 156 samples for the class **NO** from the existing dataset for evaluation purpose. The plucking technique *finger-style* is used for all samples in this publication. The two sub-classes of the slide technique were recorded on a fretless bass guitar, all others were recorded on one of the fretted bass guitar already used for the dataset in [1], again each one with three different pick-up settings. On a fretted bass guitar, pitch slides result in a stepwise increase of pitch according to the fret positions that are passed during a slide. In contrast, on a fretless bass guitar, one can detect a continuos increase or decrease in frequency. We found this signal characteristic to be of more interest, since pitch slides on other instruments as well as singing voice show similar behavior.

5.2. Procedure

We perform two experiments to evaluate the features discriminative power for the given expression style classes. In the first experiment, we aim to evaluate the classification accuracy on a *class-level* between between the classes **SL**, **VI**, **BE**, and **NO**, which is comparable to the experiment in [1]. For each of the first three classes, we combine the samples of the corresponding sub-classes (see Tab. 1). In the second experiment, we evaluate the classifiers on a *sub-class level*, meaning that we have 7 classes **SLU**, **SLD**, **VIF**, **VIS**, **BES**, **BEQ**, and **NO**.

The feature extraction step described in Sect. 4.2.4 results in 12-dimensional feature vectors. Non-linear classification problems can become linearly solvable if they are transformed into a high-dimensional space. We use the supervised classification technique Support Vector Machines (SVM), which has become state of the art in Music Information Retrieval (MIR) for comparable classification scenarios. SVM is a binary discriminative classifier that attempts to find an optimal decision plane between feature vectors of the different training classes [17]. Usually a linear separation of the classes is not possible, which is why the so called *kernel trick* is applied. The basic idea is to replace the dot product in a high-dimensional space with a kernel function in the original lower-dimensional feature space. We use the Radial Basis Function (RBF), which is the most common kernel type. For both experiments, we apply a 10-fold-cross validation and average the classification results over all folds.

6. RESULTS

As illustrated in Tab. 3, the mean classification accuracy on a *class-level* is 85.7%. The results for the class **SL** are almost 100% and the remaining three classes achieve about 80% of accuracy. We assume that the pitch progression feature introduced in Sect. 4.2.4 allows almost perfect discrimination between slide notes and other notes. A similar distribution can be obtained for the confusion matrix obtained in experiment 2 as illustrated in Tab. 4. Due to similar pitch trajectories, misclassification between the slow vibrato class (**VIS**) and the semi-tone bending class (**BES**) appear naturally. Both attributes fast and slow for the vibrato class remain subjective to a certain degree, thus, we assume that some percentage of misclassification will always remain between **VIS** and **VIF**. In general, the limitation to two subclasses for each class appears to be reasonable since we obtain comparable classification results in both experiments.

[5] http://www.idmt.fraunhofer.de/eng/business areas/dataset_bass_guitar.htm

In [2], 7 different playing techniques were classified for violin notes with error rates between 0% and 13%. Since the authors do not report class-wise error-rates, a comparison of performance is not feasible. Furthermore, the vibrato technique is the only one investigated in both publications. As mentioned in Sect. 5.1, our dataset is published and shall serve as a benchmark for the presented task of bass playing style classification.

	BE	VI	SL	NO
BE	**82.9**	9.6	4.5	3
VI	8.9	**79.9**	5	6.2
SL	0.2	0.6	**98.7**	0.5
NO	5.8	8.7	4.3	**81.2**

Table 3: Confusion matrix for experiment 1 (*class level*). All values are given in percent. Mean accuracy is **85.7**%.

	BEQ	BES	VIF	VIS	SLD	SLU	NO
BEQ	**76.9**	0.8	9.5	6	1.6	0	5.2
BES	7.6	**74.6**	0	6.9	5.3	0.8	4.8
VIF	5.3	0.7	**76.2**	5.1	5	0	7.6
VIS	9.3	1.6	8	**70.5**	5.5	0.8	4.2
SLD	0	1.1	0	0.6	**97.6**	0.6	0.1
SLU	0	0.4	0	0.9	1.3	**96.7**	0.8
NO	2.3	3.9	8.9	2.4	1.8	1	**79.7**

Table 4: Confusion matrix for experiment 2 (*sub-class level*). All values are given in percent. Mean accuracy is **81.7**%.

7. CONCLUSION

In this paper, we presented an algorithm to estimate and characterize the course of the fundamental frequency of a note over time. This allows to extract audio features to classify the applied playing technique on the corresponding instrument. We performed two evaluation experiments based on a two-layered taxonomy of expression styles of the bass guitar. For both the classification of expression styles on a *class-level* and on a *subclass-level*, we obtain high classification accuracy values of 85.7% and 81.7%. Thus, we assume that the presented algorithm is well suited for the requirements of the presented application scenarios audio coding as well as automatic music transcription for music pedagogical applications. Even though the evaluation experiments in this paper were performed only on bass guitar recordings, we assume the presented approach to be applicable to other plucked string instruments such as the guitar.

8. ACKNOWLEDGEMENTS

The Thuringian Ministry of Economy, Employment and Technology supported this research by granting funds of the European Fund for Regional Development to the project *Songs2See*[6], enabling transnational cooperation between Thuringian companies and their partners from other European regions.

9. REFERENCES

[1] J. Abeßer, H. Lukashevich, and G. Schuller. Feature-based extraction of plucking and expression styles of the electric bass guitar. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[2] I. Barbancho, C. de la Bandera, A. M. Barbancho, and L. J. Tardon. Transcription and expressiveness detection system for violin music. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, pages 189–192, 2009.

[3] M. K. C. Erkut, V. Välimäki and M. Laurson. Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar. In *Proc. of the 108th Audio Engineering Society (AES) Convention, Paris, France*, 2000.

[4] C. Dittmar, K. Dressler, and K. Rosenbauer. A toolbox for automatic transcription of polyphonic music. In *Proceedings of the Audio Mostly Conf., Ilmenau, Germany*, 2007.

[5] J. L. Flanagan and R. M. Golden. Phase vocoder. In *Bell System Technical Journal*, pages 1493–1509, 1966.

[6] M. Goto. A real-time music-scene-description system: predominant-f0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication*, 43:311–329, 2004.

[7] S. Grollmisch, E. Cano, and C. Dittmar. Songs2see: Learn to play by playing. In *Proc. of the 41th Audio Engineering Society (AES) Conference: Audio for Games, London, UK*, 2011.

[6] see www.songs2see.net

[8] S. W. Hainsworth and M. D. Macleod. Automatic bass line transcription from polyphonic audio. In *Proceedings of the Int. Computer Music Conf. (ICMC), La Habana, Cuba*, 2001.

[9] H. Järveläinen. *Perception of Attributes in Real and Synthetic String Instrument Sounds*. PhD thesis, Helsinki University of Technology, Helsinki, Finland, 2003.

[10] M. Karjalainen, T. Mäki-Patola, A. Kanerva, A. Huovilainen, and P. Jänis. Virtual air guitar. In *Proc. of the 117th Audio Engineering Society (AES) Convention, San Francisco, USA*, Oct. 2004.

[11] S. L. Marple Jr. *Digital Spectral Analysis*. Prentice-Hall, 1987.

[12] R. M. P. Stoica. *Introduction To Spectral Analysis*. Prentice-Hall, 1997.

[13] M. P. Ryynänen and A. P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32:72–86, 2008.

[14] M. Sterling, X. Dong, and M. Bocko. Pitch bends and tonguing articulation in clarinet physical modeling synthesis. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, pages 289–293, 2009.

[15] E. Tsunoo, N. Ono, and S. Sagayama. Musical bass-line pattern clustering and its application to audio genre classification. In *Proc. of the Int. Society of Music Information Retrieval (ISMIR), Kobe, Japan*, 2009.

[16] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen. Discrete-time modelling of musical instruments. *Reports on Progress in Physics*, 69:1–78, 2006.

[17] V. N. Vapnik. *Statistical learning theory*. Wiley New York, 1998.