ROBOTIC PIANO PLAYER MAKING PIANOS TALK

Winfried Ritsch
Institute for Electronic Music
and Acoustics Graz
ritsch@iem.at

ABSTRACT

The overall vision of a piano which can talk, a piano that produces understandable speech playing notes with a robotic piano player has been developed as artwork over the last decade. After successfully transcribing recorded ambient sound for piano and ensembles, the outcome of this mapping was applied by the composer Peter Ablinger in his artwork, which explores the auditory perception in the tradition of artistic phenomenalists [1] . For this vision a robotic piano player has been developed to play the result from the mapping of voice recordings, by reconstructing the key features of the analyzed spectrum stream, so that a voice can be imagined and roughly recognized. This paper is a report on the artistic research, mentioning different solutions. The output as artworks will be referenced.

1. INTRODUCTION

The basic idea from Peter Ablinger was to create a kind of phonorealism [1], which can be compared to photo-realist painting in visual arts. There are also other aspects than the technical aspect of the ”Quadraturen” for his series of works in this area to be considered, which has been covered by musicologists more precisely [2] and not described here. For his ideas and first experiments with half-tone filters, the aesthetic principles leas him to use the analysis data of recorded sounds as a base material for his compositions, leading to the one of the the most challenging disciplines, using voices as reconstruction of voices in instrumental domains, like the piano domain.

On the other side, a piano is a machine producing velocity depended sound on key press, which is rich on overtones and also includes the noise of the attack. So it seemed at a first glance, that it is impossible to reconstruct a voice with a piano.

Understanding human voice is an essential capability of man and trained from childhood. It is what we are able to recognize easily and fast, even out of noisy audio material and surroundings. So discrimination of unwanted sound on reception of voices is very well implemented in the psychoacoustic processing in the brain. On the other side we are most sensible in the recognition of small differences in speech. So speech reconstruction needs to focus on the key features of speech very precisely, but is robust to introduce additional spectral noise, such as mechanics-noise and attack spectra and we are able to mask them during speech recognition.

Looking deeper in this aspect, we recognized, there is only a need to trigger the key features for the voice reception, which are the impression of vowels and some unvoiced pitches. Like on opera singing, the unvoiced parts can be easily overheard without losing the context, they are masked by volume, superseded by music and can be imagined and reconstructed to understandable words and text during reception. The reconstructed voice is recognized more and more clearly with repetition showing, that the human brain can learn to understand a piano talking.

It is also the ambiguity between the recognition of the sound as piano and as understandable speech, which is intended and makes this work unique. To get over this border to both sides during reception as voice or piano is an important aspect for these compositions.

To render the mapping on an real piano implies the use of a computer controllable player-piano. Since non of the available player-pianos could play this, the Autopianoplayer was developed, to drive the pianos to their limits, inspiring new art work.

2. QUADRATUREN - MAPPING

The mapping of recorded sound in other domains has been developed for the idea of phonorealism in the series of Quadraturen (”Squarings”) at the Institute of Electronic music during first studies in 1996 from Peter Ablinger and implemented by Thomas Musil. Squaring refers to raster recorded sound in time and frequency. This lead to his cycle of works for symphony orchestras, sound installation to compositions for computer-controlled player piano.

> Actually however, my main concern is not the literal reproduction itself, but precisely this borderzone between abstract musical structure and the sudden shift into recognition - the relationship between musical qualities and “phonorealism“: the observation of ”reality” via “music”. (Peter Ablinger [3])

He came up with the idea of the two dimensional raster of the recording in time and frequencies, which he recognized

[1] phenomenalist in the sense of the work done by Alvin Lucier, who himself described him as phenomenalist, which is not a scientist, but uses phenomenas as a material for his art.

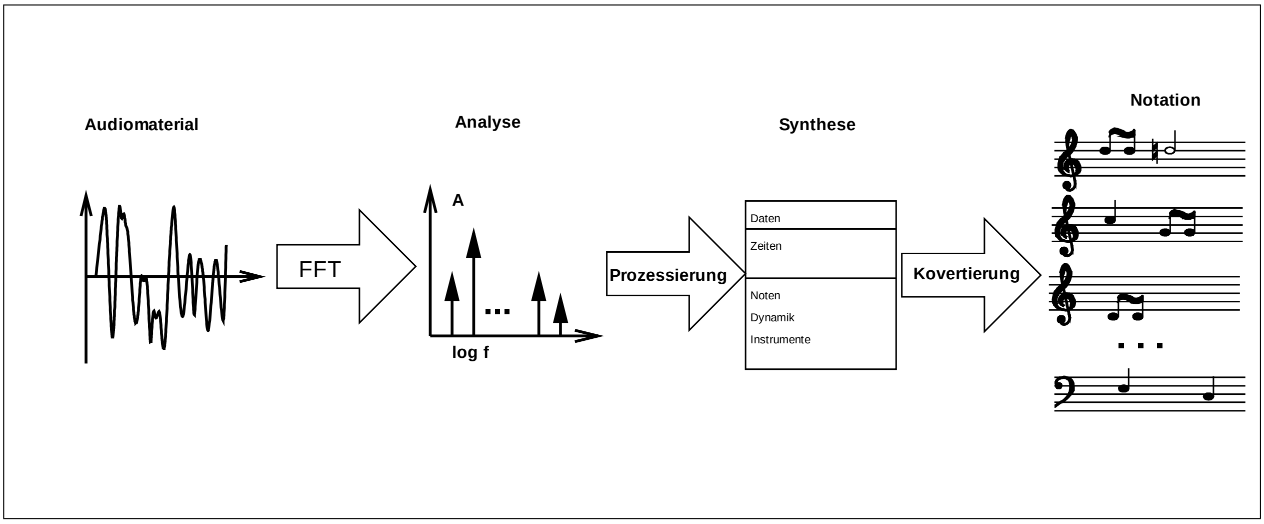

Figure 2. Analysis path.

as squares. Technical speaking it is the quantization in time grains holding the spectral power within each frequency ranges, for instrumental reproduction mostly half-tones.

2.1 Analysis

To get sufficient results, several experiments with different analysis methods has been done over the years of development. Starting with FFT, where the power of the bins are mapped to semitones like half tone filters, trying constant-Q transform[4] and wavelet analysis[5] . They have shown to work all properly with different qualities:

The plain FFT tends to wrong results in distribution in lower octaves, especially when tested with a sinus sweep as input. So for lower frequencies additional larger FFTs has been implemented, realizing a banded analysis but rising problems with the resolution of time. So a kind of constant-Q was used instead, leading to problems in mapping to equal time slices.

Wavelet analysis did not result in better transformations compared to FFT and constant-Q and was not used in later versions of the mapping software, since it was harder to handle of the data and lost the plausibility.

But overall, the algorithm of transformation, had not much influence in the result of perception of a voice via piano, especially using a real piano, instead of piano-simulators or synthesizer in comparison to influence of the extraction of the key features. As an effect, mostly the simpler to use FFT algorithm has been used, applied on different frequency bands. Also the preprocessing of the audio data before applying the frequency transformation had an great impact, and also was easy to accomplish by the composer listening to the spectral distribution. Using constant-Q was chosen on later pieces mostly for lower octaves in combination with the FFT algorithm. This was chosen by composer for different purposes.

The process for art pieces like "Deus cantando" or "A letter from Schnberg" the composer himself selected the recordings and did a preprocessing with common sound editors, to get a sound composition which is usable for the mapping process. Segmenting this audio data in pieces to work with different parameters for different parts and frequency bands like octaves. Therefore it is hard to judge the analysis step of the process for the artistic purpose.

2.2 Extraction of Key Features

The more sophisticated part was the extraction of notes from the analysis, to be played by the piano for the reconstruction of the targeted voice in respect of the best recognition. Also the more significant notes has to be selected in favor for a playable piece. All this can be controlled by parameters for composition. One of this is the number of parallel played notes in different bands. This was not only very useful for precise reconstruction, but also for the aesthetic and style of the artistic expression, which goes beyond the technical aspects.

Like in audio compression algorithms, several rules has to be implemented to accomplish the suppression of not relevant notes. The most important rule is the masking of successive notes for the same key, selecting the most significant one over time. Also the note neighborhood has to be taken into account, for the decision of the selection of the most significant note. This was needed not only for the piano to be played not beyond its limits, but also reducing unnecessary noises of not relevant notes. Unnecessary key presses can be in rising the noise floor a deal breaker. The exclusion of this notes is also controllable by parameters to vary the composition.

As a next rule, the processing and choosing of the note length in combination with concatenation of successive notes became skillful. It could be accomplished that he spectral elements of a vowel, which are present over longer time in the analysis data, can be used for reconstruction reducing onset noises. Here also predecessor notes with lower velocity are taken into account. Since it seems, that during the reconstruction of the voice in the brain, also delayed spectra elements are effective. This was percept by hearing to it, but was not proven and could be a subject for further research.

The mapping of the power of the semitones bins to velocity of the piano has shown to be an important aspect for

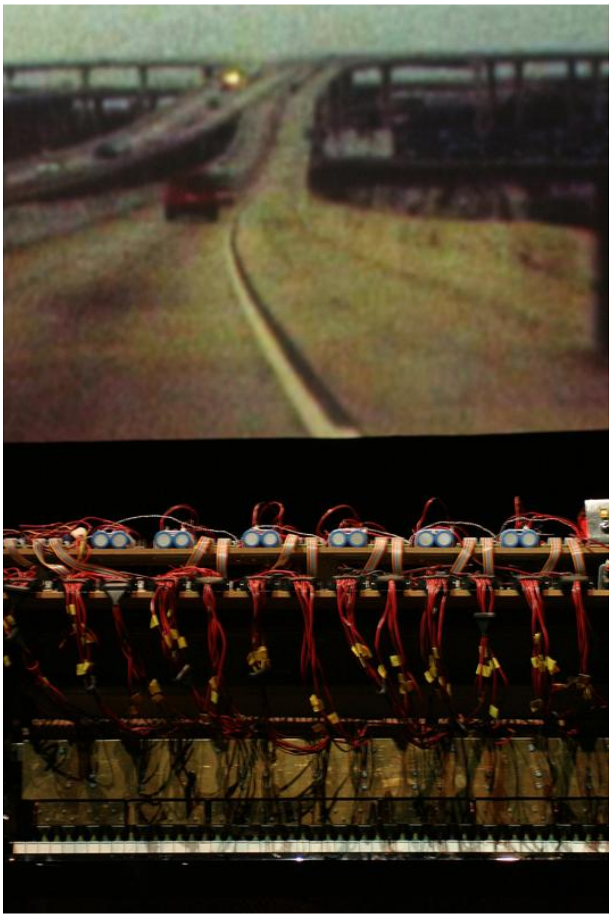

Figure 1. Playing the noise of a road with automatic piano player for the opera "Stadtoper" of Peter Ablinger performed 2004 in Graz.

the reception. This is mapping is also accomplished in an second step within the interpretation of the notes by the robotic piano player. Since interpretations also depends on different situations of playback, this has to be modified on stage.

Other rules like the avoidance of neighborhoods in notes, or the amount of parallel notes over the frequency range, the portamento within a analysis segment, the isolation of notes representing fundamental pitch was added for compositional purposes and cannot be evaluated only from a technical standpoint to enhance the perception of voices. It is also hard to describe them here precisely, since they are part of the artwork applied by the composer after the software development.

So it came out that the interlock between software and composition was very high and a general mapping software could not be finished nor released and is changed very often for new pieces, since parts of the composition is coded in this mapping.

2.3 Rendering

For the reproduction targets files are produced or a special version for real-time rendering is used. Files can be MIDI standard files [2] , for more precise data message-files interpreted by Pure Data [6] and also music-xml files for scores are foreseen.

Real-time output for the piano player, since some of the art works needs the real-time operation of the mapping described above, can be achieved by a special version of the software. This version uses a limited set of steps and therefore parameters, excluding those which introduces to much delay. Especially some steps has to be skipped in note the processing and a lower frequency resolution are needed for a faster analysis.

These leads to software tools, which must be handled by the composer, for doing new pieces and studies without technician support.

The rendering during the composition process is mostly done by software simulators, which does not really reflect the limits of a real piano. They differ in repetition rates at different velocities and mostly produce less noise in their attacks. In a second step the Autopianoplayer is targeted for corrections, because of the previously listed limits. On some pieces, also the different pianos with different behavior has to be considered for a correct performance.

3. AUTOKLAVIERSPIELER

A massive frame with 88 electromechanical finger, which are moved by solenoids, is mounted on a keyboard. controlled by micro-controllers, which are driven over a dedicated computer, the Autoklavierspieler can be controlled over Network, MIDI files and real time generated music and has been constructed at the Atelier Algorythmics [3] . As a reference some player pianos has been analyzed and the idea of a robot piano player sitting in front of a piano, has been taken from the idea of Trimpins player piano [7] . This fits also the performance purposes, since pianos are widely spread and hard to transport.

When the first piece of this serie "Zeit im Bild 2" has to be performed in 2003 on a festival, we tried to play it on serveral different player pianos, but failed. The Yamaha Diskklavier could only play 16 keys in parallel and only with 4 different parallel velocities. Marantz player could not play different velocities and the "Bsendorfer Computerklavier", precessor of the Bsendorfer Ceus system invented 2006, with his Zilog controllers always crashed after few seconds if played as fast as we needed and more than 32 keys in parallel.

A major problem to all of them, seemed to be, that they do not have enough power, especially power supply, to play that many parallel keys at high velocities. Since quite all of the western music literature for pianos can be played on these, nobody thought on the need of this for our application. Also the repetition rate was mostly to slow and repetitions at various velocities was not clean enough.

Since Trimpins automata was not available any more for being adapted, we have been forced to develop our own robot piano player, specialized for this purpose. It had also the advantage taking the needed aesthetics of performance

[2] The Music Instrument Device Interface -file standard, was the facto standard exchange for notation software imports

[3] Atelier Algorythmics Graz: `http://algo.mur.at/`

Figure 3. Millitron Autoklavierspieler 2010.

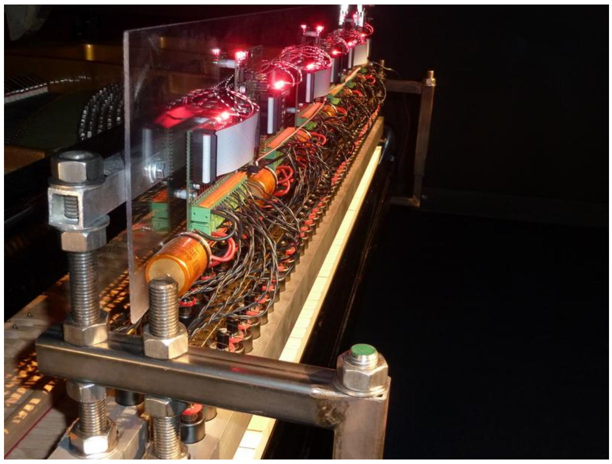

Figure 4. Rhea Autoklavierspieler 2010.

into account, but the disadvantage to do a lot of work. Also it should be affordable and usable over a longer period playing unattended.

The first version of the Autoklavierspieler, a robotic piano player named **Kantor**, powered by 1,5 kW and a weight of 120kg, was constructed 2003. As also an artistic research on robotic electromechanical instruments for extreme performances, the main target was the realization of algorithmic compositions for 88 finger, focused, but not only, on the work of Peter Ablinger. The initial project was interpreting audio recordings on the piano for the series ”Quadraturen III”.

With the need of dialogs in the compositions and more performances in Europe, an additional new Autoklavierspieler was needed, especially for opera production ”Stadtoper” from Peter Ablinger. **Millitron**, as this one was named, could be optimized from previous experiences. It has half the weight, a dedicated micro-controller board, the `algopic` and `algofet` [4] , it played more precise, was better usable, especially for quiet pieces and therefore had a better ”piano-forte“ dynamics. Exchanging the use of a hold circuit, not using PWM Modulation during the hold phase of a key, eliminated the high frequency noise from the solenoids.

Rhea was developed with the focus on even faster repetition, better dynamics for pianissimo, easy transportation and fast setup. It also was a first prototype series for a performance for 12 robotic piano player [5] . The new electronics developed for this, should enable much finer calibration and better adoption on old imprecise pianos and an easier control over Ethernet. Furthermore Rhea should be a first series for reproduction as open hardware, enabling others to build and handle the robot pianoplayer.

[4] Based on PIC16F877 and two stage FET solenoid driver this circuits wa also used in a lot of other artwork, see [8]

[5] See Maschinenhalle performance with Bernhard Lang and Christine Gaigg `http://algo.mur.at/`

3.1 Finger Strike

The idea was, that every key is playable individually, driven by a force envelope which can be adjusted to any piano, also elderly imprecise ones. It has shown that a punch of up to 3 kg per finger is needed for full velocity. The repetition rates are restricted by the mechanics of the piano and can go as low as 50 ms on pianos with good repetition mechanics. Grand pianos seems to be general faster and have better dynamics.

In the first phase of the strike the maximum force accelerates the finger, where the acceleration is more important than the actual force at the end of the strike, since velocity correspond to sound intensity. If pushed soft, like pianissimo notes, also the attack shape has to be adapted since a bouncing effects has been detected. With the new electronics it is also possible to press the key without a pluck sound of the hammer. After the attack phase, the hold phase is initiated. At this phase the key should not be pressed to the limit to enable a faster repetition. The release phase is enhanced with the reset spring of the solenoid. All of this leads to a complex system, where parameter depends on each other and these parameters has to be calibrated for each key by software.

3.2 Electronics

The electronics for Rhea uses the micro controllers boards Escher, with Ethernet and serial interfaces driving FET amplifiers for the custom made solenoids. Escher has been developed with the Autopianoplayer in mind.

3.2.1 Escher Board

The open hardware Escher is based on the motor control DSP-microcontroller dsPIC33F708MC from Microchip, a 16 bit controller with 160MIPS and dsp-unit and also hosts an Ethernet controller. Escher boards provides more than 48 I/O pins including 12 channel hardware PWM driven by 20kHz with 10-12 bit.

Each Rhea needs 3 Escher, where one master receives OSC [6] commands over Ethernet, driving the piano-player

[6] Open Sound Control over Ethernet UDP

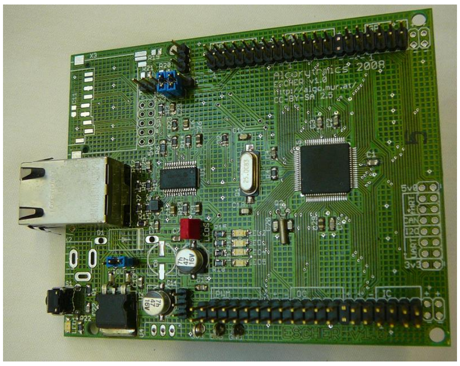

Figure 5. Escher Controller.

and forwards OSC messages to the 2 slave boards via fast serial connection. So each player piano has one IP address and status can be accessed during playing.

3.2.2 EscherFET

To drive the solenoids each Escher drives two EscherFET boards with each 16 channel of FET amplifiers.

This enables the individual control of the solenoids up to 100 MHz switching frequency, and each of 4 A constant and up to 20 A peak load, if the power supply is strong enough. After using high frequency PWM the speed of the piano could be improved from 80 ms up to 50 ms repetition rate on fast pianos. It was crucial to get a fine control over amplitude of each key for a better reconstruction of the formant spectra.

3.2.3 Solenoids

The custom designed solenoids for Rhea have been construed for optimum power on the needed 10 mm stroke length at maximum 30 V and has been manufactured for this purpose. They originate also a simpler mechanical construction and therefore reduce the overall weight. Driving them by 16 times of the power, of the allowed 100 % duty cycle, up to 20 ms increases the acceleration enough and they can be operated up to 70 degree Celsius temperature.

Another critical task was to silence them, especially on the stroke to the key. This was done via Kashmir felt proven to work for pianos for centuries. It showed that the felt reduces a lot more playing noise then industrial offered damper, so the mechanical finger also was made of non magnetic material brass with felt as finger tips.

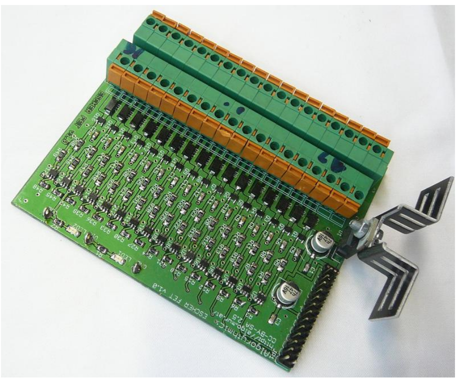

Figure 6. Escher Fet Amplifier Board.

3.3 Software

The overall system was designed for realizing performances playing within ensembles, synchronized playing to video projections, installations with automatic operations and simple interfaces and standalone solo performances. Also a fast set-up and individual calibration to different pianos and grand pianos has been implemented.

Making the piano speak needs three stages: A stage for the composition phase, which is done mostly off line, the performing software implementing the art works, integration in other environments and the micro-controllers firmware as standalone software, representing the robot player. In figure7 the structure and parts of the software system is shown.

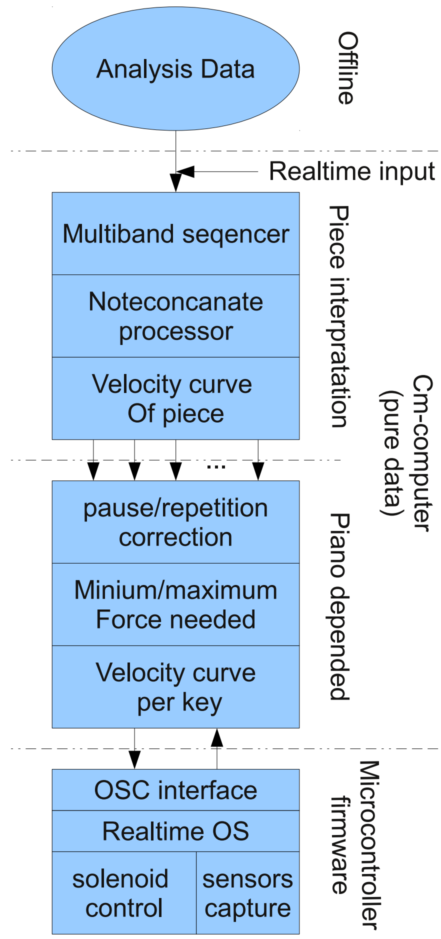

Figure 7. Software Structure.

3.3.1 Composition Tools

For the composition a set of programs, which are operated on command line interpreter was developed at the IEM for the art work of Peter Ablinger, mostly by Thomas Musil.

3.3.2 The Player

For playing with the robot piano player the programming language Pure Data was used to enhance fast prototyping and adapt and integrate easily to other projects and guarantee a fast extension to new needs. This software can also be used as a framework for usage in other projects and runs under Linux. Following task has to be matched:

3.4 Escher Realtime OS

The firmware of the player Kantor and Millitron was written in assembler language, to get proper performance on the 16F877 micro-controller. With Escher, also C with inclusion of has been used in combination with assembler code for time critical tasks.

The firmware was realized as a small real-time OS, which has to glue the OSC commands to parallel running and independent hardware control threads.

4. OUTCOME

Like mentioned at the beginning, the motivation for the robot piano player, was transcription of recordings in the domain of pianos, following a special aesthetic principle, so the outcome can be evaluated on the pieces and their performances at various festivals and in installations. Here the two major works as examples for the speaking piano from Peter Ablinger are described.

4.1 Audioanalyse / Die Auflsung / Freud in England / Le Grain de la Voix

For computer-controlled piano and video text, done in 2006. Maybe the only recording of Freud from 1938, where he already suffered on tongue cancer and explained why he immigrated to England. Therefore the beginning of the piece was underlaid with withe noise where over time the the voice appears and afterwards is more and more thinned out so that the voice ends monophonic. Here clearly the border recognizing voices can be heard depending on the trained ear for this material. With the projection of the text behind the piano the recognition of the heard speech is enhanced. The piece was commissioned commissioned by the MAK Center Wien.

4.2 Deus Cantando

This piece was commissioned for the opening of "World Venice Forum 2009" for the demand of an international environmental courtyard, where the declaration written by the Dalai Lama was spoken by a young boy from Berlin. Here the focus was the text, produced with a multi-band analysis. This piece is presented since March 2011 in the standard exhibition of the ars electronica center in Linz.

5. CONCLUSIONS AND FURTHER OUTLOOK

After successful proven with mentioned artworks, that voices can be understood played on pianos, the aesthetics of music done with this technique was the important step in this field. What first was foretold never will work, that a piano can talk and was first only a vision, became a successful artistic concept and can be surely enhanced furthermore.

Figure 8. Installation at Ars Electronica Center Museum 2011.

Acknowledgments

I want to thank all who helped developing this pieces of hardware especially Thomas Musil for his analysis program, Peter Ablinger for artwork and the many helpers implementing the robot piano player and hardware beside their work.

6. REFERENCES

[1] P. Ablinger, *Phonorealism*. `http://ablinger.mur.at/phonorealism.html`, accessed 30.3.2011, online article.

[2] C. Mello, *Mimesis und musikalische Konstruktion*. Shaker Verlag, Aachen 2010, 2011, ch. Zwischen Abbild und Selbstreferenzialitaet: Mimesis und Rauschen bei Peter Ablinger, dissertation.

[3] A. Peter, *Peter Ablinger - Quadraturen*. `http://ablinger.mur.at/docu11.html`, accessed 30.3.2011, online article.

[4] J. C. Brown, "Calculation of a constant q spectral transform," in *J. Acoust. Soc. Am., 89(1)*, 1991, pp. 425–434.

[5] P. G. R. Kronland-Martinet and S. Ystad, "Modelling of natural sounds by time–frequency and wavelet representations," in *Org. Sound, 2(3)*, 1997, pp. 179–191.

[6] M. Puckette, "Pure data," in *Proceedings, International Computer Music Conference.*, San Francisco, 1996, pp. 224–227.

[7] Wikipedia, "Trimpin — wikipedia, the free encyclopedia," 2011, [Online; accessed 27-April-2011]. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Trimpin&oldid=421091526

[8] W. Ritsch, "Algopic - algofet," accessed 30.3.2011, online article. [Online]. Available: http://algo.mur.at/projects/microcontroller/algopic/algopic