Direct and surrogate sensing for the Gyil african xylophone

Shawn Trail
Dept. of Computer Science
University of Victoria
Victoria, Canada
shawn@intrinsicaudiovisual.com

Tiago Fernandes Tavares
Dept. of Computer Science
University of Victoria
Victoria, Canada
tiagoft@gmail.com

Dan Godlovitch
School Of Earth And Ocean Sciences
University of Victoria
Victoria, Canada
dgodlovitch@gmail.com

George Tzanetakis
Dept. of Computer Science
University of Victoria
Victoria, Canaad
gtzan@cs.uvic.ca

ABSTRACT

The *Gyil* is a pentatonic African wooden xylophone with 14-15 keys. The work described in this paper has been motivated by three applications: computer analysis of *Gyil* performance, live improvised electro-acoustic music incorporating the *Gyil*, and hybrid sampling and physical modeling. In all three of these cases, detailed information about what is played on the *Gyil* needs to be digitally captured in real-time. We describe a direct sensing apparatus that can be used to achieve this. It is based on contact microphones and is informed by the specific characteristics of the *Gyil*. An alternative approach based on indirect acquisition is to apply polyphonic transcription on the signal acquired by a microphone without requiring the instrument to be modified. The direct sensing apparatus we have developed can be used to acquire ground truth for evaluating different approaches to polyphonic transcription and help create a "surrogate" sensor. Some initial results comparing different strategies to polyphonic transcription are presented.

Keywords

hyperinstruments, indirect acquisition, surrogate sensors, computational ethnomusicology, physical modeling, performance analysis

1. INTRODUCTION

The ability to interface existing musical instruments to computer systems opens up a variety of interesting possibilities. The term 'hyperinstrument' has been used to refer to an acoustic musical instrument that can be played conventionally which has been augmented with various sensors to transmit information about what is played to a computer syste expanding the sonic palette of the acoustic instrument. The most common use of hyperinstruments has been in the context of live electro-acoustic music performance where they combine the wide variety of control possibilities of digital instruments such as MIDI keyboards with the expressive richness of acoustic instruments. Another interesting application of hyperinstruments is in the context of performance analysis. The most common example is the use of (typically expensive) acoustic pianos fitted with a robotic actuation system on the keys that can capture the exact details of the player actions and replicate them. That system allows the exact nuances of a particular piano player to be captured, and when played back on the same acoustic piano will sound identical to the original performance. The captured information can be used to analyze specific characteristics of the music performance such as how timing of different sections varies among different performers. The majority of hyperinstruments that have been proposed in the literature have been modifications of Western musical instruments. The focus of this paper is extending the *Gyil* which is a traditional wooden african xylophone with digital sensing capabilities. The conventional *Gyil* has 14 or 15 wooden bars tuned to a pentatonic scale that are mounted on a wooden frame. Hollowed-out gourds are hung from the frame blow the wooden bars to act as resonators. The *Gyil* sound is characterized by a buzzing resonance, due to the preparation of the gourds, which have holes drilled and them that are papered over with spider silk egg casings. The egg casings act as vibrating membranes, but have irregular forms due to their material properties and the hole shape.

The use of computers to support research in ethnomusicology has been termed "Computational Ethnomusicology". Most music in the world is orally transmitted and traditionally was analyzed based on manual transcriptions either based on common music notation or using invented conventions specific to the culture studied. In the same way that direct sensing technology opened up new possibilities in the study of piano performance, we believe that direct sensing can provide valuable information in the study of traditional music. Music is defined as much by the process of creation as by recorded artifacts. Capturing information about the musician's actions can aid in understanding the process of music creation. This is particularly important in orally-transmitted music cultures that are slowly hybridizing or disappearing due to the onslaught of western pop music culture. We hope that by interfacing the *Gyil* with the computer we will be able to understand more about how it is played, enable the creation of electro-acoustic compositions and performances that integrate it, and better understand the physics of its sound production. In addition we hope to spark collaborations with masters of the tradition where

NIME'12, May 21 – 23, 2012, University of Michigan, Ann Arbor.

electro-acoustic mediums are largely unexplored due to the typically limited access to the associated technologies.

The paper is organized as follows: The next section describes related work and provides context for our system. Section 3 describes the sensing apparatus we have developed specifically for the *Gyil*. In Section 4 we show how this sensing technology can be used together with a physical model of the gourd resonators to create an acoustically excited hybrid model that retains the wooden bars but models the gourds virtually. Section 5 describes our experiments in trying to use sound source separation algorithms that have been tailored specifically to the *Gyil* to perform real-time causal transcription potentially bypassing the need for direct sensing. The direct sensing apparatus is used to train this indirect "surrogate" sensor by automatically providing ground truth to evaluate how good the transcription is. An underlying theme of our work is that we can achieve better results in both sensor design and audio analysis by tailoring them specifically for the *Gyil*. The paper concludes by discussing future work. Audio and video examples of the work described in this paper can be found at `http://opihi.cs.uvic.ca/nime2012gyil`.

2. RELATED WORK

Hyperinstruments are acoustic instruments that have been augmented with sensing hardware to capture performance information [1]. The majority of existing hyperinstruments have been standard western instruments such as guitars, keyboards, piano and strings. A similar emphasis on western music occurs in musicological and music information retrieval research. It has motivated research in computational ethnomusicology which is defined as the use of computers to assist ethnomusicological research [2, 3]. In world music, the use of hyper-instruments has been explored in the context of North Indian music [4] and digital sensors have been used in the development of Gamelan Electrica [5], a new electronic set of instruments based on Balinese performance practice. One interesting motivation behind the design of Gamelan Electrica is a reduction in physical size and weight, simplifying transportation. This concern also motivated us to investigate replacing the gourd resonators by simulating them digitally using physical modeling. The hybrid use of microphones to capture sound excitation and simulatation to model the needed resonances has been proposed in context of percussion instruments and termed acoustically excited physical models [6, 7]. Past work in the physical modeling of pitched percussion instruments has mostly focused on the vibraphone and marimba [8].

Indirect acquisition refers to the process of extracting performance information by processing audio of the performance captured by a microphone rather than using direct sensors. It has been motivated by some of the disadvantages of hyperinstruments such as the need for modifications to the original instrument and the difficulty of replication [9, 10]. In general it requires sophisticated audio signal processing and sometimes machine learning techniques to extract the required information. An interesting connection between direct and indirect acquisition is the concept of a surrogate sensor. The idea is to utilize direct sensors to train and evaluate an algorithm that takes as input the audio signal from a microphone and outputs the same control information as the direct sensor [11]. When the trained surrogate sensor(s) exhibits satisfactory performance it can replace the direct sensor(s).

The goal of automatic source separation is to separate the individual sound sources that comprise a mixture of sounds from a single channel recording of that mixture. Such systems can be used as front ends to automatic transcription which has the related but different goal of detecting when and for how long different sources are activated without requiring actual separation of the signals. A large family of source separation/music transcription algorithms are based on various forms of factorization in which the time-frequency representation of the mixture is decomposed into a weighted linear combination from a dictionary consisting of basis functions that correspond to the time-frequency representations of the individual sound sources [12]. The majority of existing approaches are not causal (require all of the input signal) and are designed for more general types of sound sources. Therefore they are not appropriate for real-time processing. In this paper we investigate the potential of such techniques in a real-time context and with a dictionary specifically tailored to the *Gyil*. One possible criterion for evaluating how good is a certain aproximation is the least mean squares [13]. Since sound source activities are always positive, the non-negativity criterion may be incorporated in the calculation of the approximation, leading to the Non-Negative Least Squares (NNLSQ) algorithm [13]. The NNLSQ algorithm has been used in the context of transcription (that is, the detection of symbol note data) by Niedermayer [14], and, as it represents a form of dictionary-based factorization, it may also be used to perform some forms of sound source separation. It forms the basis of our transcription system.

3. SYSTEM DESCRIPTION

The *Gyil* (pronounced Jee-Lee) is a central instrument native to the Lobi culture in the Black Volta River Basin on the boarders of Burkina Faso, Cote d'Ivoire, and Ghana [15] [1]. It is a wooden xylophone whose actual number of bars vary between 11 and 22, but is typically 14. The bars are arranged in a single row, made of Legaa wood (the region's indigenous hardwood), and use gourds (relative in size to the bar they are under) as resonators. The instrument resembles the marimba sonically and is graduated in size left to right- bigger/lower pitch to smaller/higher pitch, respectively. Its tuning is technically pentatonic, and spans three to four octaves (specifically accommodating vocal range). Because West Africans don't have a collectively established fixed frequency to designate as a systematic tuning convention, *e.g.* A=440, instruments are tuned at the discretion of the instrument maker and each has its own unique tuning, even if meant to be a pair. In general, African equidistant tuning is based on the recognition of steps that resemble one another. When a key does not conform yet is tolerated, it is considered dissonant. This idea of dissonance in the Lobi context will actually change the scale system from tetratonic to pentatonic, and is perceived as something that should be avoided. This is an important consideration when it comes to physical modeling, because the convenience of uniform distribution is not an afforded luxury, therefore each instrument and each bar on that instrument must be considered individually, along with its corresponding gourd and the sympathetic resonances from each unique neighboring gourd and the subsequent individual character of each bar.

The direct sensing we have prototyped relies on a combination of standard digital pro-audio equipment: Firewire AD/DA multi-channel audio interface, personal computer with appropriate audio processing capabilities, necessary software, playback speakers and low cost (less than 1 cent when bought in bulk) piezo transducers being used as contact microphones (hot glued directly to the bars) to ac-

[1] `http://www.mandaramusic.com/writings/Oct92percnotesdagari.html`

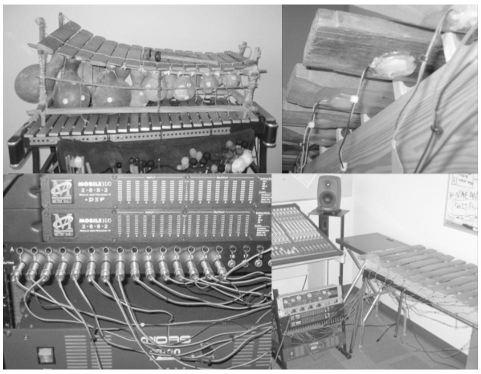

Figure 1: Top left: Gyil on top of vibraphone for showing scale, Top right: piezo sensors under bars, Bottom Left: 2882 input, Bottom Right: overview

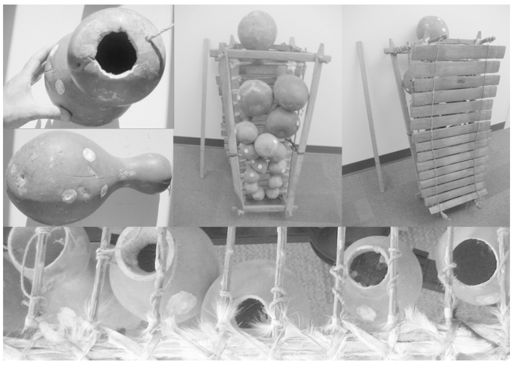

Figure 2: bass gourd shown (14" deep, 10" dia.- body, and 4" dia.- mouth) and frame construction.

quire a high-resolution direct digital audio signal. The acquired signals have a very low signal-to-noise-ratio and capture well the acoustic sound. A similar approach does not work very well when used with metal vibraphone bars, as metal bars are much more sensitive to the damping effect of attaching a piezo, which attenuates the higher frequency partials, and shortens the decay time. It does work with other wood pitched percussion instruments, such as the chromatic, western concert xylophone and Marimba. However, digital audio input using the method we have described is impractical, and cost-prohibitive (unless custom multiplexing is used) requiring more than 37 channels for any instrument over the standard 3 octave range.

Our system is designed so that each bar has a piezo-electrical sensor[2] glued onto it on the player's side near the bridge, but not uniform as the underside of the *Gyil* bars are irregular. This tends to be the most resonant location for the piezos, with the best replication of the bars sound. Each sensor is connected to a 1/4" jack input into a Metric Halo 2882 audio interface. The 2882 interface has eight 1/4" inputs with digitally controlled gain. Our system utilizes two 2882's linked optically allowing for 16 direct, identical channels of audio input, ideal for our *Gyil* which has 15 bars. The cabling was fabricated by ourselves using low cost materials and the analog circuit is completely passive. We have utilized Ableton Live 8 and Max/MSP 5 with a recent laptop computer to process the incoming 15 channels of audio without any issues. Each bar is assigned its own channel in Live and can be processed individually as needed using native devices, third party plug-ins, or custom Max4Live devices. Currently we monitor on on a stereo pair of Genelec 8030A speakers via a MIDAS Verona 320 mixing console, but any monitoring configuration can be implemented. Figure 1 shows the enhanced instrument.

4. HYBRID ACOUSTIC/PHYSICAL MODEL

The *Gyil* differs from the marimba or the xylophone primarily in the use of gourds that are hung on the body of the instrument, below the wooden bars as seen in Figure 2. They act as resonators amplifying the sound of the wooden bars. The characteristic buzzing sound of the *Gyil* is the result of the preparation of the gourds, which have holes drilled in them, which are papered over with spider silk egg casings. The egg casings act as vibrating membranes, but are irregular in their form, due to the shape of the holes, and their material properties. In this section we describe a physical model for the *Gyil* gourd. It has been motivated by the desire to create sound synthesis techniques based on physical modeling for the *Gyil* as well as a way to simplify the transportation of the instrument as the gourds are large in volume, awkward in shape, and fragile. The direct sensing apparatus described in the previous section can be easily packed as it consists of wires, contact, microphones and an external multi-channel sound card. The wooden bars can be easily folded and are quite robust. In addition it lowers the cost of making an instrument. The developed physical model is designed to act as an electronic replacement for the gourd while preserving the unique timbre of the instrument. By having separate audio input through the contact microphones for each bar, as described in the previous section, each bar can be fed into separate gourd models resulting in a more realistic model of the actual instrument.

The proposed system consists of a model for the gourd itself, and a model for the membrane. We make the assumption that the feedback from the gourd to the wooden bar is negligible and therefore can be viewed as a source-filter system. In order to make measurements of the acoustic properties of a single gourd, we removed the bars from the *Gyil* frame. Investigations were carried by exciting the gourd with a sine and triangle waves of fixed frequency and varying amplitude which led to the wave wrapping model of the membrane described below. It was found that there is a clear threshold over which there is significant high frequency distortion which we speculate is caused when the egg casing membranes start deforming instead of simply vibrating. The gourds are considered as simple resonant bodies, which can be modeled by resonant band pass filters [16]. The frequency of the filter is inversely proportional to the volume of the gourd, and the resonance or Q-factor is inversely proportional to the size of the opening at the top. For that reason, larger gourds are chosen to be associated with the lower pitched notes. The range of gourd sizes may either be selected by hand, or a scaling constant, $0 < s_f < 1$, may be used, so that the center frequency of the bandpass filter representing the smallest gourd, f_0, is selected and the frequency of the nth gourd is given by $s_f^{n-1} f_0$.

The membranes are modeled using wave folders, which have the property to wrap the input signal around two predefined limits. Mathematically, for an input signal with amplitude x_{in}, and specified reflection amplitude A, the

[2] Diameter: 20mm; Center Disc: 15mm wide; Profile: approx. 0.22mm thick; Resonant Frequency: 3.5 ± 0.5 KHz; Resonant Resistance: 500 ohms max.; Capacitance: 30 000pF ± 30% at 100Hz; Input Voltage: 30Vp-p max.; Operating Temperature: -20 to +50°C; Storage Temperature: -20 to positive 60°C. http://contactmicrophones.com/

output signal y is given by:

$$x_{out} = \begin{cases} A & -cA < x_{in} < A \\ x_{in} - A & x > A \\ cA - x_{in} & x_{in} < -cA \end{cases} \quad (1)$$

where $0 \leq c \leq 1$ is a parameter which we may set to adjust the asymmetry of the folding, that is, the difference between the absolute values of the high-level and the low-level limits, as can be seen in Figure 3. The choice of wave folding over a clipping function reflects the need for a greater degree of harmonic content to be generated than clipping produces, when processing a dynamic signal.

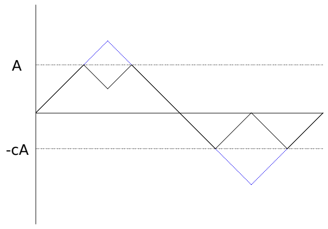

Figure 3: A simple wave folder with adjustable symmetry used to model membranes on the gourds.

The resonant filter yields signal to a number of membranes, which are wave folders as described by Expression 1. To mimic the different sizes of the membranes, and the slight variations in material, the wave folders have different parameter settings for A and c (as defined in Expression 1). To simplify the implementation of the model, and to make it more transparent, an adjustable scaling parameter $0 < s_w < 1$ is specified so that, for a model with N membranes, the level at which signal is folded in the nth membrane is given by $s_w^n A$, thus reducing the number of adjustable parameters for the wave folders from N to 2.

Last, the output of the wave folders is fed back into the resonant filter, with a controllable gain. The whole system may be visualized in Figure 4.

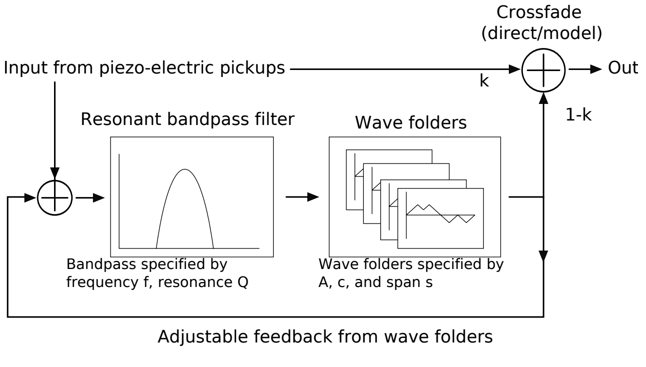

Figure 4: Diagram of signal flow for one gourd.

The signal yielded by the physical model is the output of the resonant filter, which will be very close to the sound of the vibrating bar for low amplitudes and will have a considerable amount of distortion for high amplitudes.

The gourd has the effect of amplifying the two harmonics present in the bar, and introducing a significant amount of spectral content between 1kHz and 7kHz, as can be seen by comparing the top and middle plots in Figure 5. Our gourd model driven by the output of the piezo-electric direct sensor succeeds in introducing high frequency content in a similar manner to the *Gyil* gourd, as seen in the lowest spectral plot in Figure 5. It can be seen that the *Gyil* gourd produces a more broadband spectrum than our model, which is characterized by well-defined spectral peaks. The broadband spectrum introduced by the gourd suggests a greater degree of non-linearity in the physical system than in our first-order model. We plan to address this issue in future revisions of the model.

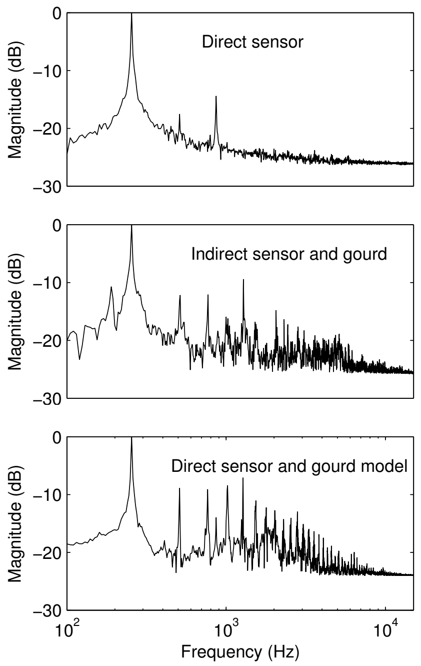

Figure 5: Spectral plots for the sound of a Gyil: without Gourd (top), with physical Gourd (middle), and with virtual Gourd (bottom)

The proposed model benefits from the direct sensors. In the physical instrument, larger gourds are related to the sound of lower pitched notes with higher intensity, and the same holds for smaller gourds and higher pitched notes. If the sound signal from the *Gyil* is simply obtained through a single microphone, this pitch-dependent coupling is not possible. By using a different channel for sensing each bar, it is possible to route signals through different instances of the model described above, hence obtaining a sound whose timbre is closer to that of the natural *Gyil* sound.

5. SOUND SOURCE SEPARATION AND AUTOMATIC TRANSCRIPTION

This section discusses the application of digital signal processing techniques for the purpose of obtaining the same data yielded by the multi-channel direct sensors, but using as input only a single channel. This single channel can either be the summation of the sensor signals or the sound data acquired from a regular microphone that is not directly attached to the instrument. The motivation is to attempt to extract the same information without requiring any modifications to the actual instrument. For this purpose we decided to investigate sound source separation algorithms. In order to obtain satisfactory performance we tailored the approach to the *Gyil.* In order to evaluate transcription algorithms it is necessary to have some form of ground truth of what the right answer should be. In most of the existing literature this is obtained through symbolic representations of the music that is then synthesized and passed through the source separation system. In the music tradition we are interested there is no concept of a score and therefore evaluation of a transcription system would require time-consuming manual creation of ground truth. By utilizing the direct sensing apparatus described above we can effectively collect unlimited amounts of ground truth and training data simply by playing the instrument.

The techniques described below rely on a factorization algorithm called Non-Negative Least Squares (NNLSQ), as proposed in [13], which aims to obtain the best description (considering the least square error) of a certain signal using a non-negative combination of a set of pre-defined basis functions. The signal representation used for the purpose of this detection is the logarithm of the magnitude spectrum, which was obtained by dividing the input signal in frames of 12 ms (with a 6 ms overlap), multiplying each frame by a hanning window, zero-padding it to twice its length and, for each frame x, calculating the spectrogram as $y = \log_{10}(1 + \|\mathrm{DFT}(x)\|)$. The resulting spectra are trimmed in order to eliminate frequency components outside the spectra of the instrument's notes (and reduce the computational load). The basis vectors are obtained by averaging a few frames of the spectrogram of an execution of each isolated note. The NNLSQ algorithm is, then, executed over each frame of the analyzed piece's spectrogram, yielding the activation level of each note. This approach is similar to the one proposed by [14]. This information may be used either for sound source separation of for automatic transcription, as described below.

The dataset used in this experiment consists of an audio recording approximately 2 minutes long. It was simultaneously recorded using the direct sensors (as separate tracks) and a microphone placed in front of the instrument. After the recording, the data from the direct sensors was artificially mixed, simulating the process of analog mixing that is part of the first scenario described above. Hence, there were three different synchronized tracks: a multi-channel track where each bar of the instrument is explicitly recorded in one different channel, a single-channel track containing data obtained from the direct sensors and a single-channel track obtained from a microphone recording. All experiments in this section take in account using direct (contact) and indirect (microphone) sensors both to acquire the basis functions and to test the performance of the analyzed methods.

5.1 Sound Source Separation

In this section, an NNLSQ-based technique for sound source separation is evaluated. This technique relies on the assumption that polyphonic audio signals that result from the mixing of several different sources are, in terms of physical measures, the sum of the signals corresponding to each individual source. Also, the human perception derived from listening to that sound is essentially the superposition of the sensations triggered when listening to each individual source. Therefore, a reasonable mathematical model for the phenomenon of sound source identification is:

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{A}. \qquad (2)$$

In that model, $\boldsymbol{B}$ is a set of basis functions, that is, a set of vector representations of the sources to be identified, $\boldsymbol{X}$ is the representation of several measures of the phenomenon in the same domain as $\boldsymbol{B}$, and $\boldsymbol{A}$ is a set of weight coefficients that represent how much each source defined in $\boldsymbol{B}$ is active in each measurement. Evaluation of the technique is based on a resynthesis schema, as follows. Since the spectral representation related to each base vector is known, it is possible to re-synthesize the expected audio for each channel. That is done by summing the magnitude spectral representation of each basis, weighted by the activation level of that note, using the phase information from the input signal and then calculating an inverse DFT. In the experiments, the input signal was separated and then remixed. The final result was compared to the input using the evaluation method described by Vincent [17]. The Signal-to-Distortion ratio (SDR) is reported, as it is the only meaningful evaluation metric for one input, one output systems. The results are reported in Table 1.

Table 1: Signal-to-Distortion Ratio for the source separation techniques.

	Basis source	
Input	Microphone	Contact
Microphone	0.0408	0.0049
Contact	0.0857	0.1558

As it may be seen, the use of sound-source separation techniques is not as effective as the use of individual sensor data regardless of the basis vectors used. This means that the use of multiple channels of direct sensors is interesting for the purpose of obtaining audio data. The next section describes an algorithm for obtaining control data, namely, the onsets and pitches that are played.

5.2 Automatic Transcription

Although obtaining audio data from the mono signal, as described above, may be difficult with today's techniques, it might only be desired to simply obtain the control data that would be provided by the direct sensors. Experiments were conducted aiming to determine the limits under which control data may be reliably obtained from these single-channel mixes. The ground truth transcription data was obtained by automatically detecting onsets using straightforward energy thresholding in the multi-channel track (offsets were ignored, as the *Gyil* does not have gestures related to offsets). This method of obtaining ground truth data is called surrogate sensing [11], and by using it a large amount of annotated data can be acquired in real-time.

The method used to obtain symbol data in this paper relies on obtaining basis vectors and then running the NNLSQ algorithm over the testing data. In order to obtain discrete control data, the activation levels are, yielded to a rule-based decision algorithm that works as follows. First, all activation levels below a certain threshold a are set to zero. After that, all values whose activation level difference are below another threshold b are set to zero. When a non-zero value for the activation value is found, an adaptive threshold value is set to that level multiplied by an overshoot factor c. The adaptive threshold decays linearly at a known rate d, and all values activation levels below it are ignored. Finally, the system deals with polyphony by assuming that a certain activation level only denotes an onset if it is greater than a ratio g of the sum of all activation values for that frame. After this process a list of events, described by onset and pitch, is yielded. The events in the ground truth list and in the obtained list are matched using the automatic algorithm described in [18]. An event is considered correct if its pitch is equal to the pitch of the matched event and its onset is within a 100 ms range of the ground truth. The values reported are the Recall (R, number of correct events divided by the total number of events in the ground truth), Precision (P, number of correct events divided by the total number of yielded events) and the F-Measure ($F = 2RP/(R + P)$). These are standard metrics used for evaluating transcription and originating from information retrieval. Table 2 show these coefficients for both analyzed pieces.

Table 2: Event detection accuracy (%), using different sensors to acquire basis.

	Basis from direct sensor		
Input Source	Recall	Precision	F-Measure
Microphone	36.39	33.63	34.95
Contact	81.01	44.91	57.79
	Basis from indirect sensor		
Input Source	Recall	Precision	F-Measure
Microphone	59.81	47.97	53.24
Contact	20.89	74.16	32.59

As can be seen, results degrade when using different sensors to calculate the basis vectors than the sensors used to acquire the signal in which the detection will be performed on. That is because the microphone signal is considerably noisier - due to normal ambient noise when playing, as well as sounds from the *Gyil* frame - and the basis vectors obtained from microphone recordings take these model imperfections into account. It is also important to note that the precision obtained when analyzing direct sensors was always greater, which is easily explainable by the absence of ambient noise in the recording. The results, however, indicate that there is significant room for improvements, and for greater accuracy requirements it is necessary to use the multi-channel direct sensor data, as current real-time techniques for polyphonic transcription have limited reliability. The direct sensing has been essential in enabling us to evaluate these different approaches. It is also important to note that the results degrade drastically when using basis functions that are not obtained from *Gyil* recordings.

6. CONCLUSIONS AND FUTURE WORK

We have described a direct sensing apparatus specifically designed for the African wooden xylophone called the *Gyil*. An array of piezo-electric pickups mounts on the wooden bars enables the capture of detailed performance data. The raw audio data produced can be used to drive a physical model of the gourds. We also describe how the direct sensors can be used to obtain ground truth information for evaluation audio transcription approaches tailored to the particular instrument. There are many directions for future work. We hope to study the variations in technique among different players in the context of performance analysis. The physical model can be improved by more detailed modeling of the gourd resonators as well as including coupling between neighboring gourds. We also plan to create a full physical model in which the wooden bar excitation is simulated. Based on our existing algorithms, indirect acquisition is not sufficiently accurate to obtain performance data. We plan to investigate several variations of factorization methods. The surrogate sensor approach to obtaining ground truth will be essential in continuously improving our algorithms. Finally we plan to include the developed hyperinstruments in live performances of electro-acoustic music.

Acknowledgments

Tiago Fernandes Tavares thanks CAPES - Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior - Proc. 2300-11-7 - for funding. The support of the National Science and Research Council of Canada is gratefully acknowledged. We thank the late Kakraba Lobi, a great maestro and advocate of the *Gyil*, who built the two instruments used in this work.

7. REFERENCES

[1] T. Machover. Hyperinstruments: A composer's approach to the evolution of intelligent musical instruments. *Organized Sound*, pages 67–76, 1991.

[2] G. Tzanetakis, A. Kapur, W.A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of interdisciplinary music studies*, 1(2):1–24, 2007.

[3] T. Lidy, C.N. Silla Jr, O. Cornelis, F. Gouyon, A. Rauber, C.A.A. Kaestner, and A.L. Koerich. On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-western and ethnic music collections. *Signal Processing*, 90(4):1032–1048, 2010.

[4] A. Kapur, P. Davidson, P.R. Cook, P. Driessen, and W.A. Schloss. Digitizing north indian performance. In *Proc. of the Int. Computer Music Conf.*, pages 556–563. Citeseer, 2004.

[5] L. S. Pardue, A. Boch, M. Boch, C. Southworth, and A. Rigopulos. Gamelan elektrika: An electronic balinese gamelan. In *Proc. Int. Conf. on New Interfaces for Musical Expression (NIME)*, 2011.

[6] R.M. Aimi. *Hybrid percussion: extending physical instruments using sampled acoustics.* PhD thesis, Massachusetts Institute of Technology, 2007.

[7] A.R. Tindale. A hybrid method for extended percussive gesture. In *Proc. New Interfaces for Musical Expression*, pages 392–393. ACM, 2007.

[8] V. Doutaut, D. Matignon, and A. Chaigne. Numerical simulations of xylophones. ii. time-domain modeling of the resonator and of the radiated sound pressure. *Journal of the Acoustical Society of America*, 104(3):1633–1647, 1998.

[9] C. Traube and J.O. Smith. Estimating the plucking point on a guitar string. In *Proc. Digital Audio Effects*, 2000.

[10] A.R. Tindale, A. Kapur, W.A. Schloss, and G. Tzanetakis. Indirect acquisition of percussion gestures using timbre recognition. In *Proc. Conf. on Interdisciplinary Musicology (CIM)*, 2005.

[11] A. Tindale, A. Kapur, and G. Tzanetakis. Training surrogate sensors in musical gesture acquisition systems. *Multimedia, IEEE Transactions on*, 13(1):50 – 59, Feb. 2011.

[12] P. Smaragdis and J.C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 177–180. IEEE, 2003.

[13] R. J. Hanson and C. L. Lawson. *Solving least squares problems.* Philadelphia, 1995.

[14] Bernhard Niedermayer. Non-negative matrix division for the automatic transcription of polyphonic music. In *Proc. of the ISMIR*, 2008.

[15] J.H. Kwabena Nketia. *The music of African.* Norton, 1974.

[16] Olson, H. F. *Music, Physics and Engineering.* Dover Publications Inc., 2 edition, 1967.

[17] C. Fevotte E. Vincent and R. Gribonval. Performance measurement in blind audio source separation. *IEEE Trans. Audio, Speech and Language Processing*, 14(4):1462 – 1469, 2006.

[18] Tavares, T. F., Barbedo, J. G. A., and Lopes, A. Towards the evaluation of automatic transcription of music. In *Anais do VI Congresso de Engeharia de Áudio*, pages 47–51, may 2008.