NON-NEGATIVE MATRIX FACTORIZATION WITH SELECTIVE SPARSITY CONSTRAINTS FOR TRANSCRIPTION OF BELL CHIMING RECORDINGS

Matija Marolt
University of Ljubljana
Faculty of Computer and Information Science
matija.marolt@fri.uni-lj.si

ABSTRACT

The paper presents a method for automatic transcription of recordings of bell chiming performances. Bell chiming is a Slovenian folk music tradition involving performers playing tunes on church bells by holding the clapper and striking the rim of a stationary bell. The tunes played consist of repeated rhythmic patterns into which various changes are included. Because the sounds of bells are inharmonic and their tuning not known in advance, we propose a two step approach to transcription. First, by analyzing the covariance matrix of the time-frequency representation of a recording, we estimate the number of bells and their approximate spectra using prior knowledge of church bell acoustics and bell chiming performance rules. We then propose a non-negative matrix factorization algorithm with selective sparsity constraints that learns the basis vectors that approximate the previously estimated bell spectra. The algorithm also adapts the number of basis vectors during learning. We show how to apply the proposed method to bell chiming transcription and present results on a set of field recordings.

1. INTRODUCTION

Bell chiming is a Slovenian folk music tradition that still exists in its original context today. It takes place in the church tower and its original role is strongly connected to Christian religious contexts. Bell chiming combines the signaling, ritual, and musical functions, because it is most often used to call the faithful to mass in a musical way, and at the same time to mark important church holidays. This is how the difference between conventional bell ringing and bell chiming as a more solemn form of playing the bells is established [1].

Slovenian-style bell chiming is performed by the musicians holding the clapper and striking the rim of the stationary bell at regular intervals. The sound is thus not produced by a swinging bell hitting the clapper, but by the clapper, typically held close to the rim, hitting the bell's rim. This gives musicians more control in altering the rhythm, speed, dynamics and accents of individual strikes, as well as leaving out strikes if desired. In the so called "Flying" tunes, one of the bells (usually the largest) is swung with a rope or electronically, and all the other bells, which are stationary, are played by striking the clapper. As a rule, each musician is responsible for playing one bell, and should strike the bell only with its clapper (touching the bell's rim with hands or other tools is not allowed). Another important rule in bell chiming is that two tones can never be played at the same time, but exceptions do occur.

Bell-chiming tunes contrast one another in the method of playing, the number of bells used, and their rhythmic and metric structure. Tunes themselves consist of repeated rhythmic patterns into which various changes, typically dynamic and agogic are included to enliven the performance. Since musicians perform in groups, without the group's consent, only small changes are possible within the time limits allocated to the bell chimer for performing his role. These changes are usually expressed as double strikes, triplets, or pauses [1].

Pioneering work in analysis of bell chiming practices was made by Ivan Mercina in the late 19th and early 20th century, who introduced a numerical notation system and published a repertoire of 243 bell chiming tunes. His work is carried on by researchers of the Institute of Ethnomusicology of the Scientific Research Centre of the Slovenian Academy of Sciences and Arts, who are still actively researching bell chiming practices. Their digital archive of Slovenian folk music and dances Ethnomuse [2] holds a large collection of bell chiming recordings, collected from the 1950s onwards. Only parts of the archive have been manually transcribed and annotated.

In this paper, we present a method for automatic transcription of bell chiming recordings. Automatic music transcription is a difficult problem to solve, although methods are improving constantly; Klapuri and Davy provide an extensive overview of the current state of the art [3]. Unsupervised learning techniques have been used by several authors to perform transcription. Abdallah and Plumbley [4] used sparse coding for transcription of synthesized harpsichord music, while Virtanen used it to transcribe drums [5]. A number of authors use variants of

SMC 2009, July 23-25, Porto, Portugal

non-negative matrix factorization to transcribe polyphonic music [6-9]. Their methods, however, were devised for music composed of harmonic sounds and are thus difficult to apply to our domain, because the sounds of bells are inharmonic and their tuning not known in advance.

To solve this problem, we propose a two step approach to bell chiming transcription. We first present an algorithm that analyzes a bell chiming recording and estimates the number of church bells and their approximate spectra by using prior knowledge of church bell acoustics and bell chiming performance rules. We then show how non-negative matrix factorization (NMF) can be used for transcription by introducing two extensions to the standard NMF learning algorithm: selective sparsity constraints that take prior knowledge of approximate bell spectra into account, and adaptation of the number of basis vectors during NMF learning.

2. ESTIMATING THE NUMBER OF BELLS AND THEIR SPECTRA

The shape or profile of a bell determines the relative frequencies of its vibrations. The conventional western shape of bells, which stems from the middle ages, tends to give the bell a single dominant pitch. Figure 1 shows the magnitude spectrum of a bell, whose dominant pitch lies at 412 Hz. The names of significant partials of the bell are also shown. These partials are usually the strongest, although (as can be seen) many others exist. The dominant pitch of the bell is defined by relations of three of its significant partials: nominal, superquint and octave nominal [10]. These form a near harmonic series with ratios 2/2, 3/2 and 4/2 resulting in a perceived virtual pitch at approximately half the nominal frequency. Most of the other partials, including the strongest for this bell (tierce), do not belong to this harmonic series.

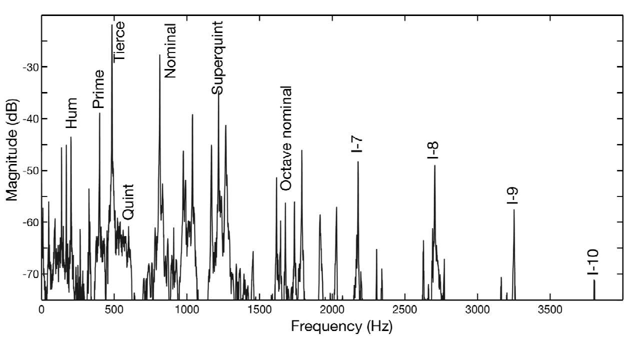

Figure 1. Magnitude spectrum of a bell

An extensive analysis of acoustics of church bells of western shape was made by Hibbert [10]. He showed that a relationship exists between the position of significant partials above the nominal and the ratio of octave nominal to nominal frequency. Hence, we can quite accurately infer the frequencies of these partials if we know the frequencies of the nominal and octave nominal. Frequencies of partials below the nominal do not exhibit strong relationships with the nominal. To assess their positions relative to the nominal, we analyzed a set of 318 church bells and used the means and standard deviations of partial frequencies in this collection.

We estimate the number of bells in a given recording and their spectra by analyzing the covariance matrix **C** of frequency bins of the time-frequency magnitude spectrogram. The elements of the matrix are calculated as:

$$c_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}\left(f_{ik}-\mu_i\right)\left(f_{jk}-\mu_j\right), \quad (1)$$

where n is the length (number of frames) of the spectrogram, f_{ij} its elements and μ_i the average magnitude of the i-th frequency bin. We exploit the fact that in a bell chiming performance, bells are usually struck many times, but typically not at the same time. Therefore, the amplitude envelope of a bell's partial will be correlated to amplitude envelopes of other partials of the bell, but not to amplitude envelopes of partials of other bells. Groups of bells are usually tuned so that some of their strong partials overlap, but then these will also be at least partially correlated with non-overlapping partials. Figure 2 shows two rows of the covariance matrix of a bell chiming performance including three bells. The top row is placed at the tierce frequency (488 Hz) of the bell from Figure 1 (B1). The bottom row is placed at the nominal frequency of B1 (824 Hz), which coincides with the superquint of another bell with nominal frequency 548 Hz (B2).

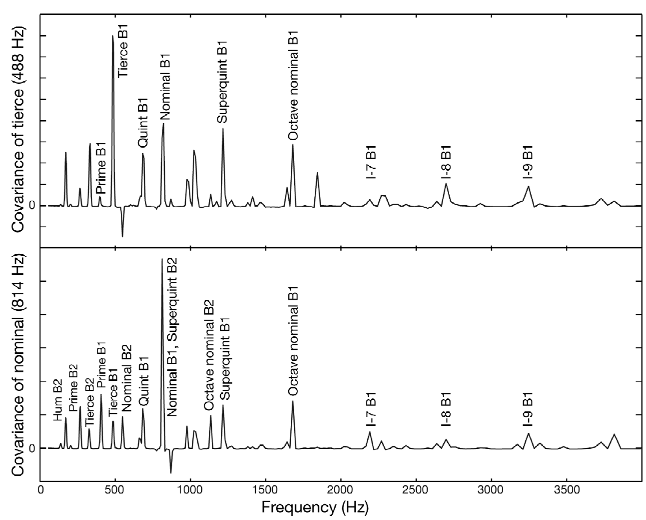

Figure 2. Two rows of the covariance matrix.

The top row clearly shows that the amplitude envelope of the tierce partial of B1 correlates well with amplitude envelopes of other partials of B1. Even though two more bells are present in the recording, amplitude envelopes of their partials are quite different and therefore not correlated with B1's tierce. The bottom row shows that even though partials belonging to two bells share the frequency of 814Hz, partial series of both bells can be discerned from this row of the covariance matrix.

These findings led us to the following algorithm for estimation of the number of bells and their approximate spectra from the covariance matrix:

1. given a time-frequency representation **F** of an audio signal, we calculate its covariance matrix **C**, which contains covariances of amplitude envelopes of all frequency bins;
2. for each row $\mathbf{c}_i$ of the covariance matrix, we find all pairs of peaks (c_{ij} , c_{ik}) that may form a nominal - octave nominal pair of partials. Our analysis of bells showed that the octave nominal typically lies in the range of 1050 to 1350 cents above the nominal, so all pairs within this range are taken into consideration;
3. for each pair of peaks (c_{ij} , c_{ik}), we construct a spectral template $B_{ijk}=\{ (\mu_p, \sigma_p), p=1,...,l \}$ of a bell using the Hibbert's model [10] for calculating the frequencies of partials above the nominal and our own analysis of bells for partials below the nominal. The template contains estimates of frequencies of significant bell partials and their standard deviations;
4. for each template B_{ijk} we calculate its correlation to the row $\mathbf{c}_i$ of the covariance matrix. If it exceeds a threshold T_1, we include the template in the set of all bell templates $\mathfrak{B}$.

To test the bell finding algorithm, we collected a set of 22 bell chiming recordings performed on three to five church bells and manually labeled the nominal frequencies of bells used in performances. We calculated the Constant-Q magnitude spectrogram of each recording by using a maximum window size of 125ms, a step size of 31.25 ms and 20 cent spacing between adjacent frequency bins. To flatten the spectral energy distribution, we scaled the bark scale sub-bands inversely proportional to their variance, as suggested by Klapuri [11]. This especially enhances the amplitudes of higher frequency partials, which is helpful for finding the correct number and spectra of bells with the algorithm described previously, as well as for subsequent NMF learning, which tends to be more sensitive to high-energy observations. Comparable methods for weighting the spectrum were also used by other authors. Virtanen [12] used a weighted cost function in which the observations were weighted so that the quantitative significance of the signal within each critical band was equal to its contribution to the total loudness. Similarly, Vincent [7] used perceptual weights to improve the transcription of low energy notes.

The whitened power spectrum was used as input to the previously described algorithm. We evaluated the algorithm by calculating the precision and recall scores describing how the nominal frequencies of the found bell templates match the manually annotated nominal frequencies of bells. The mean precision-recall curve for all 22 recordings, calculated by varying the threshold T_1, is shown in Figure 3. For our further experiments, we chose to set the threshold T_1 at 0.35. In this way, a high recall value of 0.97 was obtained, meaning that virtually all bells in all recordings were correctly identified, while precision of 0.68 yielded an average of two false positives (superfluous bell templates) per recording.

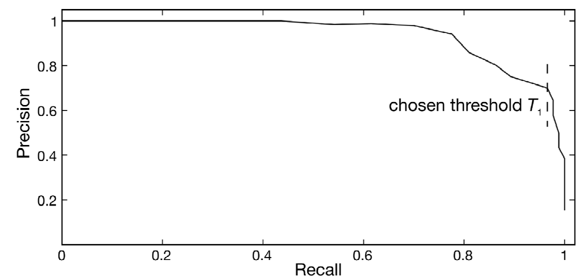

Figure 3. Precision-recall curve of the bell finding algorithm

3. TRANSCRIBING BELL-CHIMING RECORDINGS

When non-negative matrix factorization (NMF) is applied to transcription of polyphonic music, a time-frequency transform is first used to transform the time-domain audio signal into time-frequency space, thus obtaining a time-frequency representation **A**. NMF approximates **A** by two non-negative matrices **W** and **H**, so that:

$$\mathbf{A} \cong \mathbf{WH} \,, \qquad (2)$$

where **W** is a matrix of basis vectors and **H** a matrix of coefficients. For music applications, the columns of **W** corresponds most naturally to individual music events (spectra of bells in our case), while the rows of **H** explain how amplitudes of these events change over time. Several efficient implementations of NMF exist in the literature. Our experiments are based on a recent algorithm introduced by Kim and Park [13] that allows sparsity constraints on matrices **W** or **H**.

The naive approach to applying non-negative matrix factorization to transcription of music signals simply by factorizing the magnitude or power spectrum has several shortcomings. As music events overlap in time, there is no guarantee that NMF will separate individual music events into separate basis vectors. A single vector may end up containing partials of several events or only a subset of partials of an individual event. We also have to estimate the number of basis vectors in advance; using too few basis vectors will result in vectors containing several events, on the other hand, too many vectors may result in fragmentation of events over several vectors. Authors have addressed these issues in the past by constraining basis vectors to predefined harmonic templates. Niedermayer [9] used a preset number of fixed basis vectors learned from recordings of individual piano notes and only adapted the matrix of coefficients **H** during learning. Raczynski et al. [8] used a preset number of basis vectors corresponding to individual piano notes, initialized and constrained the vectors to non-zero values only for frequency bins corresponding to perfectly harmonic partial series and learned the weights of these harmonics, as well as the matrix of coefficients. The idea was extended by Vincent [7], who learned the weights of predefined hamonic narrowband partial series belonging to different notes. Vincent also allowed for inharmonicity of the partial series.

The described approaches all assume that the transcribed music signal is composed of a number of individual events (notes) that are composed of (almost) harmonic partials series and thus spectral templates of these events can be constructed in advance. While this works well for piano music, on which all of the mentioned approaches were tested, it cannot be directly applied to our domain. The sound of church bells is inharmonic. We do not know what frequencies the bells are tuned to and even when we do, frequencies of significant bell partials can only be coarsely approximated. Our analysis of bell partials below the nominal showed that their frequencies can vary by as much as 150 cents from an estimated average. In addition, we simply have no data to model all of the bell partials (see unlabelled partials in Figure 1).

We therefore devised an NMF learning algorithm with selective sparsity constraints that uses the spectral templates found by the bell finding algorithm presented in section 2 to initialize and guide the learning process, so that the learned basis vectors approximate the actual bell spectra and as a result, the matrix of coefficients describes the amplitude envelopes of individual bells.

3.1. NMF Learning with Selective Sparsity Constraints

As described previously, non-negative matrix factorization can be used to factor the spectrum into two non-negative matrices **W** and **H**, so that columns **W** corresponds to individual music events, while rows of **H** explain how the amplitudes of these events change over time. We wish to use the set of bell templates $\mathfrak{B}$ obtained by the algorithm described in the section 2 to guide NMF learning, so that the basis vectors of **W** will approximate the found bell templates, while at the same time allowing the algorithm to find the best fit to the actual bell spectra.

To this end, we introduce selective sparsity constraints into NMF learning. Our algorithm is derived from Kim and Park's SNMF/L learning algorithm, which is based on alternating non-negativity constrained least squares and active set method [13]. The algorithm already supports sparsity constraints by imposing L_1-norm based constraints on **H** or **W**. Factorization is calculated by solving:

$$\min_{\mathbf{W},\mathbf{H}} \frac{1}{2}\left(\left\|\mathbf{A}-\mathbf{W}\mathbf{H}\right\|_F^2 + \zeta \left\|\mathbf{H}\right\|_F^2 + \alpha \sum_{i=1}^{m} \left\|\mathbf{w}_i^T\right\|_1^2 \right), \quad \mathbf{W},\mathbf{H} \geq 0, \quad (3)$$

where $\mathbf{w}_i^T$ is the i-*th* row vector of **W**, m the number of rows of **W**, $\zeta \geq 0$ a parameter that suppresses the growth of **H**, while $\alpha \geq 0$ balances the trade-off between accuracy of approximation and sparsity of **W**. Eq. (3) is minimized by iteratively solving two sub-problems using an active set based fast algorithm for non-negativity constrained least squares with multiple right hand side vectors:

$$\min_{\mathbf{W}} \left\| \begin{bmatrix} \mathbf{H}^T \\ \sqrt{\alpha}\mathbf{e}_{1\times k} \end{bmatrix} W^T - \begin{bmatrix} \mathbf{A}^T \\ \mathbf{0}_{1\times m} \end{bmatrix} \right\|_F^2, \ \mathbf{W} \geq 0, \quad (4)$$

and

$$\min_{\mathbf{H}} \left\| \begin{bmatrix} \mathbf{W} \\ \sqrt{\zeta}\mathbf{I}_k \end{bmatrix} \mathbf{H} - \begin{bmatrix} \mathbf{A} \\ \mathbf{0}_{k\times n} \end{bmatrix} \right\|_F^2, \ \mathbf{H} \geq 0, \quad (5)$$

where k is the number of basis vectors, m the number of rows of **W**, n the number of columns of **H**, $\mathbf{e}_{1\times k}\in\mathbb{R}^{1\times k}$ a vector of ones, $\mathbf{0}_{k\times n}\in\mathbb{R}^{k\times n}$ a matrix of zeros and $\mathbf{I}_k$ a $k\times k$ identity matrix. Minimization of equation (4) involves L_1-norm minimization of each row of **W**, thus imposing sparsity on **W**. The strength of this constraint is controlled by the parameter α.

To introduce prior knowledge into NMF learning, we propose a modification of the above approach that selectively enables sparsity constraints only for parts of **W** where no partials are expected. The goal is to constrain **W** to approximate the spectral templates derived from the covariance matrix, while still allowing NMF to learn the best match to the actual bell spectra. Learning improves the estimated partial frequencies, adds partials not included in spectral templates and estimates partial amplitudes.

The bell finding algorithm described in section 2 can estimate the number of church bells in a given recording, as well as their approximate spectral templates ($B_i\in\mathfrak{B}$, i=1,...,k) by analyzing the covariance matrix of the time-frequency representation. The templates are represented as a set of partial frequencies and their standard deviations: $B_i=\{ (\mu_{ip}, \sigma_{ip}), p=1,...,l \}$. We can thus construct a selectivity matrix **V** as:

$$\mathbf{V} = [v_{ij}]_{m\times k}, \ v_{ij} = \begin{cases} 1, \ \exists p : \left| f_i - \mu_{jp} \right| < T_2 \sigma_{jp}, p = 1,...,l \\ 0, \ elsewhere \end{cases}, \quad (6)$$

where f_i is the center frequency of i-th row of the time-frequency representation **A** and T_2 a threshold determining the amount of allowed deviation of a partial from its estimate in $\mathfrak{B}$. The number of columns in the selectivity matrix is equal to the number of bell templates found. Each column of the matrix corresponds to one template and contains ones in places where we expect partials to occur in the time-frequency representation (according to the corresponding template), and zeros elsewhere. The matrix can be used to selectively apply sparsity constraints only for components of **W** where no partials are expected (**V** contains zeros), and leave components where partials should occur unconstrained. To incorporate selective sparsity constraints, we modify the first part of the SNMF/L learning iteration (Equation (4)) to obtain the SSNMF/L algorithm:

$$\min_{\mathbf{w}_i^T} \left\| \begin{bmatrix} \mathbf{H}^T \\ \sqrt{\alpha}\left(1-(\mathbf{v}_i^T)^T\right) \end{bmatrix} \mathbf{w}_i^T - \begin{bmatrix} \mathbf{a}_i^T \\ 0 \end{bmatrix} \right\|_F^2, \mathbf{w}_i^T \geq 0, i = 1,...,m, \quad (7)$$

where $\mathbf{w}_i^T, \mathbf{v}_i^T$ and $\mathbf{a}_i^T$ are i-*th* rows of matrices **W**, **V** and **A** respectively. The matrix **V** acts as a selector that enables or disables sparsity constraints with regard to corresponding bell templates. Equation (7) should be minimized for each row of **W**; for efficiency, we can group together all rows with the same values of $\mathbf{v}_i^T$ and perform non-negativity constrained least squares calculation once for each group.

3.2. Adapting the number of basis vectors during learning

To initialize NMF learning, we must decide on the number of basis vectors to use. We set this number to the number of bells found by the algorithm presented in section 2. The algorithm is tuned to correctly find most of the bells in a recording, with an average of two additional false positives (bells not present in the recording). These false positives are problematic, as partials may be incorrectly attributed to the false-positive vectors during learning. We observed that this usually happens with only a small number of partials, so that the false-positive vectors are a poor match to their corresponding spectral templates. We therefore introduce an additional step to the SSNMF/L learning algorithm, which removes the basis vectors that do not match the templates. The entire learning algorithm is as follows:

1. calculate a set of bell templates $\mathfrak{B}$ using the algorithm described in section 2;
2. set the number of NMF basis vectors *k* to the number of bell templates found and initialize SSNMF/L learning;
3. repeat minimization of equations (7) and (5) until the change in the normalized Karush-Kuhn-Tucker residual Δ, as defined in [13], falls below a set threshold ε_1;
4. remove all basis vectors that do not match any of the bell templates in $\mathfrak{B}$;
5. repeat points 3. and 4. until the change in Δ falls below a threshold ε_2.

We tested the algorithm on the same set of manually annotated bell chiming recordings used for evaluation of the bell finding algorithm. We also used the same time-frequency representation as previously described. The values of other parameters were set to: $\alpha = 0.0005n$, $T_2 = 2$, ε_1=0.01, ε_2=10^{-9}, $\zeta = 0.01$. For comparison, Table 1 displays results achieved with the standard NMF algorithm without constraints, Kim and Park's SNMF/L with sparsity constraints on **W**, the proposed SSNMF/L with selective sparsity constraints on **W** and finally SSNMF/L with adaptation of the number of basis vectors. The listed precision and recall scores show how well the found basis vectors correspond to the actual bells. Gradual improvement is attained when sparsity and selective sparsity constraints are introduced, while a major boost is achieved by adapting the number of basis vectors during learning. Adaptation is especially helpful in boosting precision, because it removes irrelevant basis vectors, which in turn also boosts recall by correctly distributing the partials amongst the remaining basis vectors.

	precision	recall
standard NMF	0.61	0.83
SNMF/L	0.65	0.82
SSNMF/L without adaptation	0.66	0.86
SSNMF/L with adaptation of the number of basis vectors	0.89	0.89

Table 1. Evaluation of basis vectors learned by different NMF methods

3.3. Transcribing bell chiming recordings

The presented SSNMF/L algorithm factors the time-frequency representation of a bell chiming recording into two non-negative matrices **W** and **H**, so that **W** corresponds to the spectra of bells, while **H** explains how the amplitudes of bells change over time. The transcription algorithm we propose is straightforward. Onsets are first detected with the complex domain algorithm devised by Bello et al. [14]. The matrix **H** is filtered with a fourth order low-pass Butterworth filter with cutoff frequency at 0.25π to smooth the amplitude changes and remove small irregularities. Then, for each onset found, we analyze the 150 ms section of the filtered matrix **H** around the onset and assign the bell with the strongest increase in **H** to the onset.

	num bells	bells found prec. / rec.	onset precision/recall	transcription precision/recall
r1	3	0.75 / 1	0.94 / 0.96	0.93 / 0.95
r2	3	1 / 1	0.91 / 0.96	0.78 / 0.83
r3	4	1 / 1	0.88 / 1	0.79 / 0.90
r4	4	0.75 / 0.75	0.99 / 1	0.71 / 0.72
r5	5	1 / 1	0.87 / 0.92	0.66 / 0.69
r6	5	0.8 / 0.8	0.75 / 0.98	0.45 / 0.59

Table 2. Evaluation of transcriptions of six bell chiming recordings

To test the algorithm, we manually transcribed 20 second excerpts of 6 bell chiming recordings containing three to five bells and evaluated the number of correctly transcribed bells. Table 2 lists results for the 6 recordings. Its columns contain: the number of bells in each recording (2), precision and recall of the match between the basis vectors and the actual bells (3), precision and recall of onset detection (4) and precision and recall of transcription (5).

When all bells are correctly represented by the basis vectors, the average recall is around 0.84. We are satisfied with this result, because we are transcribing real field recordings, which are affected by factors such as poor microphone placement, weather conditions, bell tower acoustics etc. Even though onset detection itself works very well, most of the errors are still made due to indistinctive bell onsets, which may occur because of the aforementioned factors, long bell decay times or change of dynamics by performers. Weak onsets make it hard to determine which bell actually sounded at a given onset, resulting in ignored onsets or incorrectly labeled bells. Because of long bell decay times, which cause most bells to sound throughout a performance, as well as bell tower acoustics, partials may be attenuated or amplified during a performance, which may also lead to false positives or incorrect bell labeling. Accuracy drops sharply when bells are not accurately identified, as wrong bells are assigned to the found onsets.

4. CONCLUSION

The proposed approach to transcription of bell chiming recordings is a good first step into making this part of Slovenian cultural heritage more accessible to interested researchers. There is plenty of room for improvement, especially with the transcription algorithm, but we also plan to extend our researches into automatic extraction of bell chiming patterns, as well as the development of a retrieval system for queries based on bell chiming patterns and recording excerpts.

Acknowledgments. This work was supported in part by the Slovenian Government-Founded R&D project EthnoCatalogue: creating semantic descriptions of Slovene folk song and music.

5. REFERENCES

[1] M. Kovačič "New Contexts, Esthetics, and Transfer in Bell-chiming Tradition," *Slovene studies,* vol. 29, pp. 19-34, 2007.

[2] M. Marolt*, et al.*, "Ethnomuse: Archiving Folk Music and Dance Culture," in *Eurocon 2009*, St. Petersburg, Russia, 2009.

[3] A. Klapuri and M. Davy, Eds., *Signal Processing Methods for Music Transcription*. New York: Springer, 2006, p.^pp. Pages.

[4] S. A. Abdallah and M. D. Plumbley, "Unsupervised analysis of polyphonic music by sparse coding," *Neural Networks, IEEE Transactions on,* vol. 17, pp. 179-196, 2006.

[5] T. Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 15, pp. 1066-1074, 2007.

[6] 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.*, 2003, pp. 177-180.

[7] E. Vincent*, et al.*, "Harmonic and inharmonic Nonnegative Matrix Factorization for Polyphonic Pitch transcription," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 109-112.

[8] S. A. Raczynski*, et al.*, "Multipitch Analysis with Harmonic Nonnegative Matrix Approximation," in *ISMIR 2007, 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007, pp. 381-386.

[9] B. Niedermayer, "Non-negative Matrix Division for the Automatic Transcription of Polyphonic Music," in *ISMIR 2008, 9th International Conference on Music Information Retrieval*, Philadelphia, USA, 2008, pp. 544-545.

[10] W. A. Hibbert, "The Quantification of Strike Pitch and Pitch Shifts in Church Bells," Ph.D., Faculty of Mathematics, Computing and Technology, The Open University, Milton Keynes, UK, 2008.

[11] A. Klapuri, "A perceptually motivated multiple-FO estimation method for polyphonic music signals," presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,, New Paltz, USA, 2005.

[12] T. Virtanen, "Separation of sound sources by convolutive sparse coding," in *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, Jeju, Korea, 2004.

[13] H. Kim and H. Park, "Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method," *SIAM Journal on Matrix Analysis and Applications,* vol. 30, pp. 713-730, 2008.

[14] J. P. Bello*, et al.*, "On the use of phase and energy for musical onset detection in the complex domain," *Signal Processing Letters, IEEE,* vol. 11, pp. 553-556, 2004.