OCR-BASED POST-PROCESSING OF OMR FOR THE RECOVERY OF TRANSPOSING INSTRUMENTS IN COMPLEX ORCHESTRAL SCORES

Verena Thomas   **Christian Wagner**   **Michael Clausen**

Computer Science III, University of Bonn

`{thomas,wagnerc,clausen}@iai.uni-bonn.de`

ABSTRACT

Given a scanned score page, Optical Music Recognition (OMR) attempts to reconstruct all contained music information. However, the available OMR systems lack the ability to recognize transposition information contained in complex orchestral scores. [1] An additional unsolved OMR problem is the handling of orchestral scores using compressed notation. [2] Here, the information of which instrument has to play which staff is crucial for a correct interpretation of the score. But this mapping is lost along the pages of the score during the OMR process. In this paper, we present a method for retrieving the instrumentation and transposition information of orchestral scores. In our approach, we combine the results of Optical Character Recognition (OCR) and OMR to regain the information available through text annotations of the score. In addition, a method to reconstruct the instrument and transposition information for staves where text annotations were omitted or not recognized is presented. In an evaluation we analyze the impact of transposition information on the quality of score-audio synchronizations of orchestral music. The results show that the knowledge of transposing instruments improves the synchronization accuracy and that our method helps in regaining this knowledge.

1. INTRODUCTION

A conductor reading an orchestral score can easily recognize which instrument is notated in which staff of a system.

---

[1] For transposing instruments the written notes are several semitones higher/lower than the sounding notes.

[2] In our context, the notion of compressed score is used to describe a score, where after the first system staves of instruments not playing are temporarily removed from a system.

For this to be possible, a set of common conventions of typesetting scores was developed. Examples are the introduction of all instruments playing in a piece of music by labeling the staves of the first system, a fixed instrument order or the usage of braces and accolades to cluster instruments [12]. In case of compressed scores, in addition to the labeling of the first system, subsequent systems are annotated with instrument text labels as well (see Figure 1). However, these are typically annotated by abbreviations instead of full instrument names. In several scores, labels are omitted when a system does not differ structurally from the preceding system.

Figure 1. Extracts from Franz Liszt: *Eine Sinfonie nach Dantes Divina Commedia* using compressed notation (Publisher: *Breitkopf & Härtel*).

The PROBADO project [3] aims at developing a digital library system offering new presentation methods for large collections of music documents (i.e., scans of sheet music and digitized music CDs). Similar to a conductor following the score while listening to a performance, the PROBADO system highlights the measure in the score matching the currently audible part of an audio track. One prerequisite for this type of presentation is a mapping/synchronization of pixel areas in the score scans to time intervals in the audio track (see Section 3). The first step in calculating this mapping is the reconstruction of the musical information contained in the score scans using OMR. [4] However, for orchestral scores the existing OMR systems lack the ability to reconstruct all information given in the score. Orchestral scores contain instrumentation information which

---

[3] `http://www.probado.de`

[4] We apply *SharpEye2* (`http://www.visiv.co.uk`)

might be important, e.g., for extracting the score of a single instrument. In addition, the instrument text labels also mark transposing instruments in the score. Their ignorance results in shifts of single voices with respect to the rest of the voices in the score. In [14] the impact of typical OMR errors on the results of score-audio synchronizations was analyzed. It turned out that lacking transposition information has to be classified as the most influential OMR error. This suggests that the transposition information contained in the score should be reconstructed. Unfortunately, at the current state no OMR system known to us offers the extraction of transposition information as well as a correct instrument labeling [5] of the score. *SharpEye* provides some text recognition. However, the recognitions are not analyzed with respect to instrument names and are not mapped to the according staves, let alone propagated to the following (unlabeled) systems. The OMR system *PhotoScore Ultimate 6* [6] offers instrument labeling to some extent. The included OCR engine recognizes the instrument texts (often including transpositions) and maps them to the staves. In addition, the recognized instrument text labels from the first system are propagated to the following systems. However, the used method seems to be rather simple. *PhotoScore* maps the instrument text labels extracted from the first system to the following systems line by line. Unfortunately, text labels from these systems and structural differences in case of compressed scores are ignored. Therefore, particularly for compressed scores, incorrect instrument labelings are created. Another OMR system dealing with instrument labels is *capella-scan*. [7] The observed abilities of *capella-scan* to create and propagate instrument and transposition labels are comparable to those of *PhotoScore*. In both OMR systems, the recognized transposition text labels—even if correctly recognized—do not seem to be transformed into transposition labels that are considered during the creation of a symbolic representation, such as MIDI or MusicXML.

In OMR research two crucial questions exist: Firstly, which music format is processed? Each format (e.g., handwritten score, medieval score) calls for specialized reconstruction methods. Secondly, what is the application scenario? The intended application strongly influences the required OMR accuracy. On the one hand, there are situations where an exact reconstruction of the score is crucial. In this context, OMR systems that allow for manual corrections of the recognition results were proposed (e.g., [5]). Learning mechanisms integrated into those systems then use the user feedback to gradually improve the OMR accuracy. On the other hand, some applications demand OMR pro-

cesses providing a sufficient quality without requiring user interactions. In our scenario, we are interested in processing a large data collection with as little user interaction as possible. The generated OMR results are only required for score-audio synchronization, which is robust with respect to missing notes and incorrect note durations. Therefore, in this situation accuracy loss in favor of automation is desirable.

Although a great deal of research on OMR has been conducted (see, e.g., [1]), the special challenges of orchestral scores have not yet been addressed. However, the extraction of instrumentation and transposition information has to be considered a crucial part of OMR for orchestral scores regardless of whether the goal is an exact digital reconstruction of the score (e.g., for score-informed voice separation) or a rough representation intended for further processing.

In this contribution we present a method to reconstruct the missing instrument and transposition labels in orchestral scores. We combine OCR and OMR to regain information from text labels in the score. Subsequently, instrument and transposition labels for staves lacking text annotations are reconstructed using music-related constraints and properties.

In Section 2 we will describe our instrument and transposition label reconstruction method. In Section 3 we will give a short description of the applied score-audio synchronization technique. Afterwards, the results of our evaluation using a set of 11 orchestral pieces are presented and discussed. We close this paper with a summary and an outlook on possible future work in Section 4.

2. METHOD

We present our method to reconstruct the instrument and transposition labels in staves of orchestral scores. Basically, the algorithm can be subdivided into three parts: In the first part of the process (Subsection 2.1) the text areas on the score scans are identified and processed by an OCR software. Subsequently, the recognition results are transformed into instrument labels and matched to the corresponding staves. After this step, all staves, where textual information was given in the score and recognized by the OCR software, possess an instrument label. But in orchestral scores, after the first system, instrument text labels are often omitted. Therefore, in the second step of the algorithm (Subsection 2.2) missing labels are reconstructed by propagating existing labels. Afterwards, each staff has an instrument label associated with it. In the final step of the algorithm the transposition labels that were found in the first system are propagated through the score (Subsection 2.3).

We impose some assumptions on the scores processed with our method:

- The first system contains all instrument names that occur in the piece.

---

[5] In contrast to the text actually placed on the score—which we call *instrument text label*—*instrument labels* are language independent. All known abbreviations or names of the same instrument are mapped to the same label. These labels are used to identify which instrument is meant to play in a staff.

[6] `http://www.sibelius.com/products/photoscore/ultimate.html`

[7] `http://www.whc.de/capella-scan.cfm`

- The instrument order established in the first system is not changed in subsequent systems.
- A maximum of two staves share a common instrument text label.
- When first introduced, full instrument names are used.
- For compressed scores, text labels are given if the instrumentation changed compared to the preceding system.

For most orchestral scores these assumptions are met.

We will now provide a detailed account of the three steps of the instrument and transposition labeling algorithm. For an even more extensive description we refer to [15].

2.1 OCR-based instrument labeling

In this part of the reconstruction, we analyze textual information given on the score sheets to create instrument and transposition labels.

First, given a scanned score image, the contained connected components (CCs) of black pixels are determined [11, 15]. Afterwards, CCs that definitely do not contain letters are discarded. Using a sweep line algorithm [3] horizontally neighboring CCs are then merged to form words. Subsequently, the thereby determined image areas are used as input for the *ABBYY FineReader 10* OCR software. [8]

At this point, we have a list of OCR recognitions and their positions on the score scans. To achieve a proper instrument labeling two additional steps are required. First, the recognized text is compared to an instrument library. The library contains names and abbreviations for typical orchestral instruments in German, English, French, and Italian. Using the Levenshtein distance [8], the library entries with the longest word count that are the most similar to the recognitions are identified and used as instrument labels in the according text areas. Secondly, using the staff position information available in *SharpEye*, the identified instrument labels are mapped to the according staves of the score.

In the majority of cases, transposition information is available from text labels like "clarinet in A" (see Figure 2). To detect transpositions we therefore search for occurrences of text labels containing the keyword "in" followed by a valid transposition.

2.2 Instrument label reconstruction

This section constitutes the main part of the proposed method. We will use the labeling from the previous section as initialization of an iterative process to reconstruct the labeling for all staves. Given the score of a piece of music, we define the sequence of all systems $M = (M_0, \ldots, M_m)$ and the set of all instrument labels $\mathcal{I}$ of system M_0 that were reconstructed in Section 2.1. With $S = [1:N]$ we enumerate all the staves in M and let $S_a \subset S$ denote the staff numbers corresponding to M_a. Furthermore, we create a ma-

trix $\pi \in [0,1]^{S \times \mathcal{I}}$, where $\pi(i, I)$ will be interpreted as the "plausibility" of staff i having the instrument label I. The submatrix $\pi_a \in [0,1]^{S_a \times \mathcal{I}}$ corresponds to M_a. We initialize π with the instrument labels determined in Section 2.1. As plausibility values, the Levenshtein distances between the instrument labels and the original instrument text on the score sheets are applied. Note that due to this initialization, several instruments might be mapped to one staff (e.g., for the text label "viola and violoncello"). Afterwards, the plausibility matrix $\pi^0 := \pi$ is iteratively updated using an update method that can be subdivided into three steps

$$\pi^{k+1} = IOC \circ IP \circ POP(\pi^k).$$

We will now explain these three steps of the update process in chronological order.

2.2.1 Propagation of plausibilities (POP)

In this step we will propagate the plausibilities from system M_a to system M_b, for several $a < b$ specified below. To perform a plausibility propagation, we fist calculate the set $C_{a,b} \equiv C_{a,b}(\pi_a, \pi_b)$ consisting of all triples $(i, j, I) \in S_a \times S_b \times \mathcal{I}$ whose joint plausibility $\pi_a(i, I) \cdot \pi_b(j, I)$ is positive. We then reduce $C_{a,b}$ by removing all crossings. A crossing between two triples (i, j, I) and (k, ℓ, K) with $i < k$ occurs if $j > \ell$. In case of a crossing, the triple with smaller joint plausibility is removed. The resulting set will be denoted by $C'_{a,b}$. By projecting the elements of $C'_{a,b}$ onto the first two components, $(i, j, I) \mapsto (i, j)$, we end up with the set $C^\times_{a,b} \equiv C^\times_{a,b}(\pi_a, \pi_b)$. To deal with uninitialized systems and full scores, we add the pairs $(0,0)$ and $(|S_a|+1, |S_b|+1)$ to $C^\times_{a,b}$. After sorting $C^\times_{a,b}$ lexicographically, we perform the following update process $\uparrow(\pi_b|\pi_a)$ for π_b given π_a:

1. For the smallest element $(i, j) \in C^\times_{a,b}$ search the minimal $t \geq 1$ such that $(i+t, j+t) \in C^\times_{a,b}$.
2. If no such t exists, goto 5.
3. Compute P_{ij} consisting of all $(i+s, j+s) \in S_a \times S_b \setminus C^\times_{a,b}$ such that $s \in [1:t-1]$ and staff $i+s$ and staff $j+s$ share the same clef label.
4. For all $(\ell, I) \in S_b \times \mathcal{I}$ update π_b as follows:
   $\pi_b(\ell, I) = \max(\{\pi_b(\ell, I)\} \cup \{\pi_a(k, I) \mid (k, \ell) \in P_{ij}\})$.
5. Update $C^\times_{a,b}$ by removing (i, j).
6. If $|C^\times_{a,b}| > 1$, goto 1.

Using this local update instruction, we define $POP(\pi^k)$ in two steps. First we calculate $\widetilde{\pi}^k_b := \uparrow(\pi^k_b|\pi^k_0)$ for all $b \in [1:m]$ and then $POP(\pi^k_b) := \uparrow(\widetilde{\pi}^k_b|POP(\pi^k_{b-1}))$ is recursively computed. We redefine $\pi^k := POP(\pi^k)$.

2.2.2 Applying instrument properties (IP)

In this step, we extract knowledge from the plausibility matrix to reconstruct missing instrument labels and to fortify already existing plausibility entries. We define some staff-related properties $E_1, \ldots, E_p$ as subsets of S where $i \in E_j$

---

[8] http://finereader.abbyy.com

means that staff i has property E_j (e.g., staff i has treble clef or staff i is the first/last staff in the system). Similarly, we define properties $F_1, \ldots, F_q \subset \cup_{a=0}^m S_a \times S_a$ between two staves of the same system (e.g., staff i is in the same brace as staff j). We now use these staff related properties and π to deduce instrument related properties.

For each instrument I we calculate the probability distribution P_I on $\mathbb{E} := \{E_1, \ldots, E_p\}$ given π:

$$P_I(E|\pi) = \frac{\sum_{i \in E} w_i \cdot \pi(i, I)}{\sum_{E' \in \mathbb{E}} \sum_{i \in E'} w_i \cdot \pi(i, I)},$$

where $w_i = \frac{3}{4}$ for staves i in S_0 and $w_i = \frac{1}{4}$ otherwise. For $(I, F) \in \mathcal{I} \times \mathbb{F}$ with $\mathbb{F} := \{F_1, \ldots, F_q\}$ we compute the probability distribution $P_{I,F}$ on $\mathcal{I}$ given π:[9]

$$P_{I,F}(J|\pi) := \frac{\sum_{(i,j) \in F} w_i \sqrt{\pi(i, I) \cdot \pi(j, J)}}{\sum_{J' \in \mathcal{I}} \sum_{(i,j) \in F} w_i \sqrt{\pi(i, I) \cdot \pi(j, J')}}.$$

Using these global instrument properties, we now define the plausibility increase

$$\pi_\Delta(I, i) := \sum_{E \in \mathbb{E}: i \in E} w_E P_I(E|\pi) + \\ \sum_{j \in S, J \in \mathcal{I}} \sum_{F \in \mathbb{F}: (i,j) \in F} w_F \sqrt{\pi(j, J) P_{I,F}(J|\pi)},$$

where w_E, w_F are suitable property weights. Using π_Δ, we define $IP(\pi^k) := \mathrm{N}(\pi^k + \pi_\Delta^k)$, where for a non-zero matrix X, $\mathrm{N}(X) := X/\max_{ij} |x_{ij}|$. We redefine $\pi^k := IP(\pi^k)$.

2.2.3 Exploiting the instrument order constraint (IOC)

A common convention for score notation is that the instrument order established in the first system is not altered in subsequent systems. Therefore, we use the instrument labels of S_0 to penalize systems where the instrument order established by S_0 is violated.

Given M_0 and a system M_a, $a > 0$, we extract the sequences $\mathcal{I}_0 = (I_1, \ldots, I_{|S_0|})$ and $\mathcal{I}_a = (J_1, \ldots, J_{|S_a|})$ of most plausible instrument labels. Afterwards we calculate the set L_{0a} of all pairs $(i, j) \in S_0 \times S_a$ with $I_i = J_j$ for which a pair $(k, \ell) \in S_0 \times S_a$ exists with $I_k = J_\ell$ such that (i, j, I_i) and (k, ℓ, I_k) constitute a crossing (Subsection 2.2.1). The plausibility decrease $\pi_{\nabla,a}(j, J_j) := \lambda \sum_{i:(i,j) \in L_{0a}} \pi_a(i, I_i)$ with suitable parameter $\lambda > 0$ is calculated for all $a \in [1:m]$. Finally, the plausibility update using the instrument order constraint is given by $IOC(\pi^k) := \mathrm{N}(\pi^k - \pi_\nabla^k)$, where $\pi_\nabla^k = \left(\pi_{\nabla,0}^k, \ldots, \pi_{\nabla,m}^k\right)$.

2.3 Transposition propagation

During the OCR-based reconstruction of the instrument labels, the available transposition information is also transformed into transposition labels and subsequently mapped

---

[9] We chose two different probability distributions to account for the differences between the two sets of properties $\mathbb{E}$ and $\mathbb{F}$.

to the according staves. After the reconstruction process described in the previous subsection has terminated, the transposition labels from the first system are propagated through the whole score. For each staff in S_0 holding a transposition label, the occurrences of its instrument label in the rest of the score are determined. The concerned staves will then be assigned with the transposition label from S_0.

In the context of our evaluation in Section 3 we used this method to propagate manually corrected transposition labels in the first system to the whole score.

We are aware of the fact that some orchestral scores contain transposition information next to arbitrary staves. However, extracting those short text labels (e.g., "in A") is a new challenge and is left to be analyzed.

3. EVALUATION

As the need for an algorithm that reconstructs the transposition information contained in musical notations arose from our application scenario, we will evaluate the impact of our method with respect to the task of score-audio synchronization. In Subsection 3.1 we provide a short overview of the technique of score-audio synchronization. Afterwards, we give a detailed account on the performed evaluations (Subsection 3.2).

3.1 Score-audio synchronization

The goal of music synchronization in general is the calculation of a mapping between each position in one representation of a piece of music to the musically matching position in another representation of the same piece of music. For score-audio synchronization tasks the given input documents are score scans and audio tracks.

In the first step of the synchronization both music documents are transformed into a common representation which then allows for a direct comparison. We chose to use the well-established chroma-based features. For details on the calculation of chroma features from audio recordings we refer to [2,9]. To extract chroma features from score scans the given sheets are first analyzed with an OMR system to reconstruct the musical information. After storing the recognition results in a MIDI file, the chroma features are calculated similarly as for the audio recordings.

In the next step a similarity matrix is calculated from the two feature sequences. Finally, by applying multiscale dynamic time warping [10, 13] a minimal path through this matrix is calculated. The synchronization between the music documents is then encoded by this path.

3.2 Experiments

For our evaluation, we employ the beat annotations from the RWC Music Library [6] as ground truth. We extracted the measure starting points from these files to generate a reference synchronization on the measure level. As test data

we selected the 11 orchestral pieces which contain at least one transposing instrument (see Table 1). In addition, the respective orchestral scores were collected and processed with *SharpEye* (data sources: IMSLP [10] and Bavarian State Library [11]). For four of the pieces we found scores that use a compressed notation. Obviously, the labeling task is harder for those scores than for scores using a full notation. To perform the synchronization experiments, we took audio excerpts of roughly two minutes length and the according score clippings.

Label	Work	Publisher
C1	Haydn: Symphony no. 94 in G major, 1st mvmt.	Kalmus
C2	Tchaikovsky: Symphony no. 6 in B major, 4th mvmt.	Dover Publications
C3	Mozart: *Le Nozze di Figaro*: Overture	Bärenreiter
C4	Wagner: *Tristan und Isolde*: Prelude	Dover Publications
F1	Beethoven: Symphony no. 5 in C minor, 1st mvmt.	Breitkopf & Härtel
F2	Brahms: Horn Trio in Eb major, 2nd mvmt.	Peters
F3	Brahms: Clarinet Quintet in B minor, 3rd mvmt.	Breitkopf & Härtel
F4	Mozart: Symphony no. 40 in G minor, 1st mvmt.	Bärenreiter
F5	Mozart: Clarinet Quintet in A major, 1st mvmt.	Breitkopf & Härtel
F6	Mozart: Violin Concerto no. 5 in A major, 1st mvmt.	Bärenreiter
F7	Strauss: "An der schönen Blauen Donau"	Dover Publications

Table 1. Overview of the test data. The scores of C1–C4 use compressed and the scores of F1–F7 use full notation.

Before presenting the synchronization results, we want to briefly comment on the accuracy of the instrument labeling results of the proposed method. For our test data there were a total of 464 instrument text labels given in the score. In addition, 87 transposition text labels were found. Our evaluation method could correctly reconstruct 88% of the instrument and 77% of the transposition labels (see Table 2). The error sources are diverse (e.g., OCR misrecognitions, unconsidered instrument abbreviations) and some will be discussed after the presentation of the synchronization results.

	Instrument labels		%	Transposition labels		%
	total	errors		total	errors	
Compressed	401	53	87	75	17	77
Full	63	1	98	12	3	75
Total	**464**	**54**	**88**	**87**	**20**	**77**

Table 2. Percentage of wrongly reconstructed text labels.

For each piece of music we calculated four synchronizations. In the first case, we used the MIDI created from the *SharpEye* recognition data (OMR) to create the score-audio synchronization. In the other cases we manipulated the OMR recognition before performing the synchronization. In the second case, we manually annotated the missing transposition labels in the scores (OMR*). In the third case, we applied the label reconstruction method described in Section 2 (OMR+LR).[12] In the last case, we manually corrected the transposition labels in the first system before the transposition propagation is performed (OMR+LR*). Table 3 shows the evaluation results for all of the mentioned

---

settings. The numbers state the mean and standard deviations from the ground truth. Comparing the results of OMR

Label	OMR		OMR*		OMR+LR		OMR+LR*	
	mean	std	mean	std	mean	std	mean	std
C1	456	1016	283	441	456	1016	283	441
C2	434	502	385	378	424	505	425	503
C3	247	349	128	178	134	183	181	247
C4	1005	980	889	884	889	884	889	884
Av	**536**	**712**	**421**	**470**	**476**	**647**	**445**	**519**
F1	462	700	265	391	284	493	265	391
F2	390	672	110	125	110	125	110	125
F3	266	803	124	84	124	84	124	84
F4	93	88	93	86	93	88	93	86
F5	243	383	65	53	65	53	65	53
F6	79	81	69	66	69	66	69	66
F7	451	658	310	492	310	492	310	492
Av	**243**	**405**	**148**	**185**	**151**	**200**	**148**	**185**

Table 3. Overview of the deviation of the different synchronization results from the ground truth (in ms).

and OMR*, it becomes evident that knowing all transposition labels results in a significant improvement of the synchronization results.

For six pieces—one of which has a compressed score—our method could correctly reconstruct all transposition labels (C4, F2, F3 and F5–F7, see column OMR+LR). For the remaining pieces, other than C1 and F4, the method improved the synchronization results compared to not applying any post-processing. By annotating the transposition labels in the first system manually before propagating them through the score (OMR+LR*) the results became equal to OMR* for all full scores and the compressed score C1. Although, manual interaction was still required, only annotating the first system constitutes a significant improvement compared to annotating all systems of an orchestral piece manually. For C2 and C3 a correct reconstruction of the transposition labels was not possible. In addition, using the propagation of the transposition labels from the first system results in a degradation of the synchronization compared to OMR+LR (due to instrument labeling errors).

We will now discuss the labeling results for some scores in more detail. For two pieces the transposition text labels given in the score were not recognized. In C1 the score notation uses an unusual setting of the transposition text labels (see Figure 2). The text labeling in C1 results in the recognition of three separate text labels ("in", "G"and "Sol") instead of one text label (e.g., "in G"). Therefore, our method could not reconstruct the transposition labeling. In F4 the alignment of the transposition text labels would allow for a successful recognition but the OCR engine produced results such as "i n Sol" or "inSiw". In both of these examples the keyword "in" with a subsequent space was not available. Although for all other pieces the transposition labels in the first system were correct, some instrument labeling errors occurred which sometimes influenced the transposition labeling of subsequent systems in a negative manner. Some of these errors result from incorrect OCR recognitions (e.g., recognition of "FI." instead of "Fl." (flute) results in a map-

ping to "Fg." (Fagott, German for bassoon)). Furthermore, some text labels are wrongly interpreted as instrument text labels and thereby produce wrong instrument labels. An interesting mix-up occurred for C3. Here, Italian text labels are used and both the clarinet and the trumpet are part of the instrumentation. However, in Italian the trumpet is called "clarino" which is abbreviated by "Cl.". But, in English this abbreviation is used for the clarinet.

Figure 2. Examples of missed transposition text labels.

We also performed an evaluation of the impact of other OMR errors (clefs, accidentals, pitches, durations) on the prospective synchronization results (see Table 4). In accordance with the results in [14], correcting the OMR data almost consistently resulted in an improvement. However, the accuracy increase is less pronounced than for transpositions.

Label	OMR		OMR*		OMR+LR		OMR+LR*	
	mean	std	mean	std	mean	std	mean	std
C4	1018	967	936	856	936	856	936	856
Av	**486**	**517**	**426**	**405**	**436**	**449**	**445**	**457**
F1	342	528	151	169	172	219	151	169
Av	**269**	**471**	**131**	**144**	**134**	**151**	**131**	**144**

Table 4. Synchronization results for corrected OMR data. The averages are calculated for C1–C4 and F1–F7, respectively.

4. CONCLUSIONS AND FUTURE WORK

We presented a method for the reconstruction of instrument and transposition labels from orchestral scores. Our method reconstructs instrument labels based on an OCR recognition and propagates those labels to staves where no instrument text labels existed in the score. We tested our method in the context of score-audio synchronization. The evaluation showed both the need for the reconstruction of transposition labels to improve the synchronization results and the ability of our method to achieve this.

At the moment our method is being integrated into the preprocessing workflow of the PROBADO application (see [4]). We hope to thereby reduce the manual annotation effort required to administer large music databases.

To make the reconstruction more robust—especially for compressed scores and with respect to the imposed assumptions—we suggest several ideas. We found that although *ABBYY FineReader* produces a very high recognition rate for words ($> 97\%$), the recognition of instrument abbreviations was often inferior to other OCR engines. Therefore, a promising idea is the combination of several OCR engines to make the initial OCR-based instrument labeling

more reliable. Our method takes advantage of some conventions for music notation while currently ignoring several others. We assume that, e.g., key signatures, braces, and instrument groups form powerful tools w.r.t. the task of instrument labeling. However, *SharpEye* does not recognize those features reliably and prevents their reasonable usage. We therefore suggest to reconstruct them by, e.g., combining several OMR engines as proposed in [7] and to subsequently integrate them into the proposed method.

5. REFERENCES

[1] D. Bainbridge and T. Bell. The Challenge of Optical Music Recognition. *Computers and the Humanities*, 35(2):95–121, 2001.

[2] M.A. Bartsch and G.H. Wakefield. Audio Thumbnailing of Popular Music Using Chroma-Based Representations. *IEEE Transactions on Multimedia*, 7(1):96 – 104, 2005.

[3] J.L. Bentley and T.A. Ottmann. Algorithms for Reporting and Counting Geometric Intersections. *IEEE Transactions on Computers*, 100(9):643–647, 1979.

[4] D. Damm, C. Fremerey, V. Thomas, M. Clausen, F. Kurth, and M. Müller. A Digital Library Framework for Heterogeneous Music Collections—from Document Acquisition to Cross-Modal Interaction. *International Journal on Digital Libraries: Special Issue on Music Digital Libraries (to appear)*, 2011.

[5] M. Droettboom and I. Fujinaga. Interpreting the semantics of music notation using an extensible and object-oriented system. In *Proc. Python Conference*, 2001.

[6] M. Goto. AIST Annotation for the RWC Music Database. In *Proc. ISMIR*, 2006.

[7] I. Knopke and D. Byrd. Towards MusicDiff: A foundation for improved optical music recognition using multiple recognizers. In *Proc. ISMIR*, 2007.

[8] V. I. Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.

[9] M. Müller. *Information Retrieval for Music and Motion*. Springer, Berlin, 2007.

[10] M. Müller, H. Mattes, and F. Kurth. An Efficient Multiscale Approach to Audio Synchronization. In *Proc. ISMIR*, 2006.

[11] A. Rosenfeld and J. L. Pfaltz. Sequential Operations in Digital Picture Processing. *Journal of the ACM*, 13:471–494, 1966.

[12] S. Sadie, editor. *The New Grove Dictionary of Music and Musicians (second edition)*. Macmillan, London, 2001.

[13] S. Salvadore and P. Chan. FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space. In *3rd Workshop on Mining Temporal and Sequential Data*, 2004.

[14] V. Thomas, C. Fremerey, S. Ewert, and M. Clausen. Notenschrift-Audio Synchronisation komplexer Orchesterwerke mittels Klavierauszug. In *Proc. DAGA*, 2010.

[15] C. Wagner. OCR based postprocessing of OMR results in complex orchestral scores – Which (transposing) instrument corresponds to which staff ? Diploma thesis, University of Bonn, 2011.