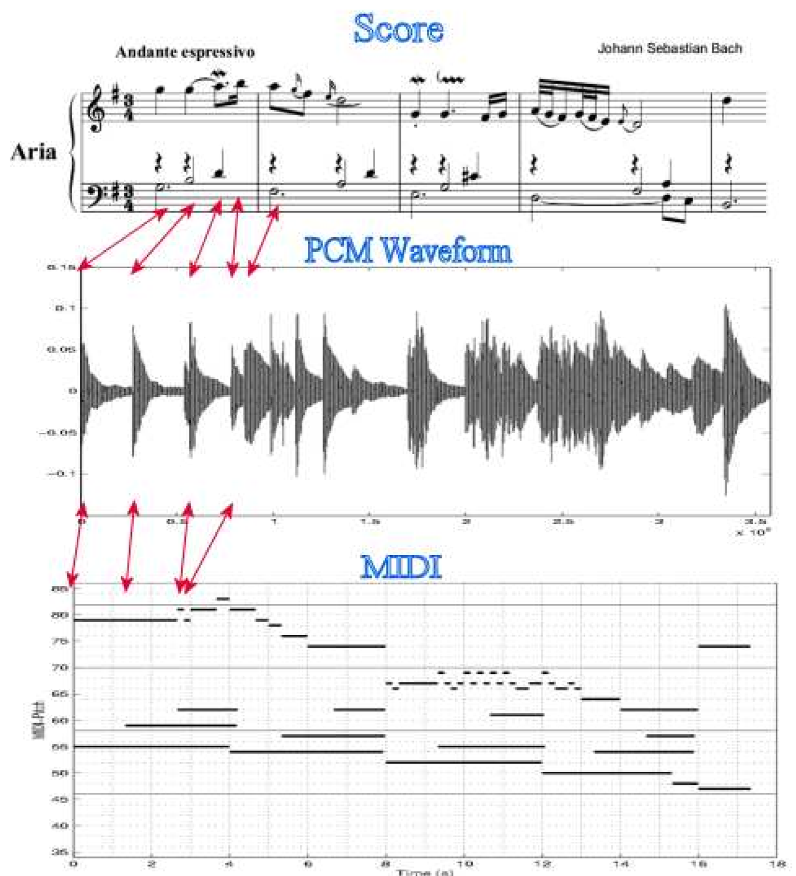
Automatic Synchronization of Music Data in Score-, MIDI- and PCM-Format

Vlora Arifi, Michael Clausen, Frank Kurth, and Meinard Müller
University of Bonn, Germany – Department of Computer Science III
{vlora, clausen, frank, meinard}@cs.bonn.edu, http://www-mmdb.iai.uni-bonn.de

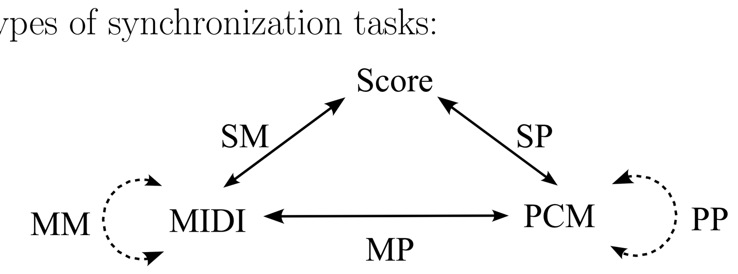
1 Problem Setting: Synchronization

Synchronization task: Given a (time-) position in some representation of a piece of music (e.g., in score or MIDI format), determine the corresponding position within some other representation (e.g., given in PCM-format).

Example: Consider score-, PCM (waveform)-, and MIDI versions of the first $4\frac{1}{3}$ measures of the *Aria con Variazioni* by J.S. Bach, BWV 988:



The red arrows link the corresponding events of the different versions. Based on those three typical music representations, we may consider the following types of synchronization tasks:



In this poster, we shall only consider Score-to-PCM (SP-) synchronization.

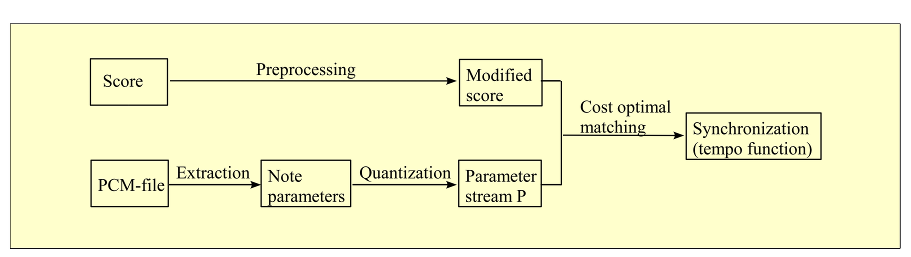
Applications:

- Automatic annotation of a piece of music available in different data formats as a basis for content-based retrieval.
- Usage of link structures to access PCM audio piece accurately after score-based music retrieval.
- Investigation of agogic and tempo studies.
- Automatic tracking of the score positions during a performance.

2 Overview: Synchronization Framework

Synchronization proceeds in three steps:

1. Extract note parameters from the PCM (Section 3).
2. Preprocess score to normalized representation (Section 4).
3. Synchronize extracted note parameters and score (Section 5.3) w.r.t. suitable cost function (Section 5.2).



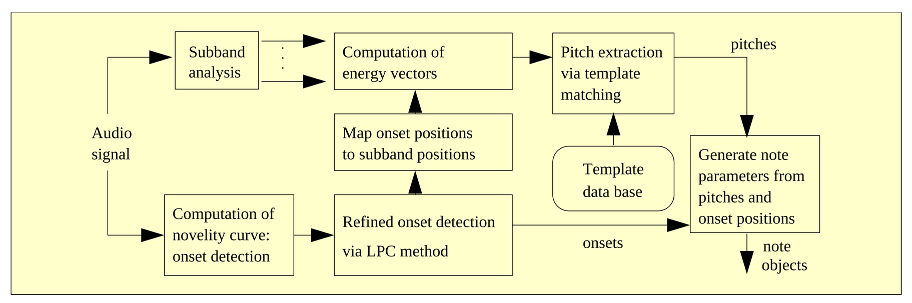
3 Extraction of Note Parameters

Onset detection:

- Track changes of signal's frequency contents over time using novelty curves.
- Refine time-resolution of resulting estimated onset positions using linear prediction (following Foster et al. (1982)).

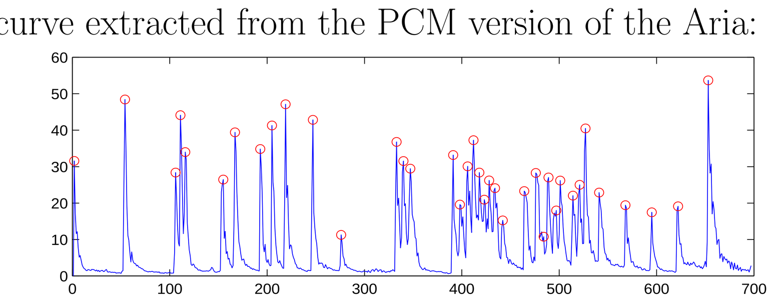
Pitch extraction:

- Subband analysis using tree-structured multirate filterbank adapted to musical scale (at most one note per subband) following Bobrek et al. (1998).
- Establish note positions in subbands using detected onset positions.
- Estimate pitches using template-matching.

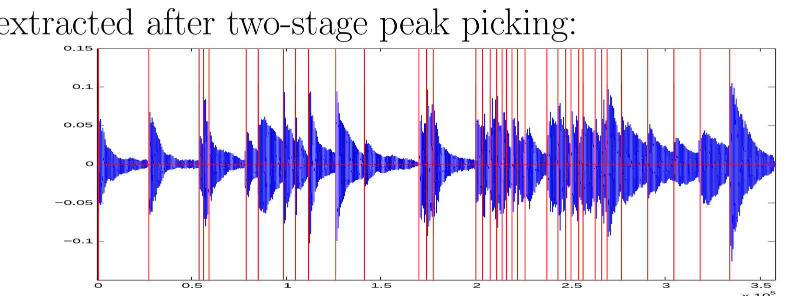


Example: (cont'd)

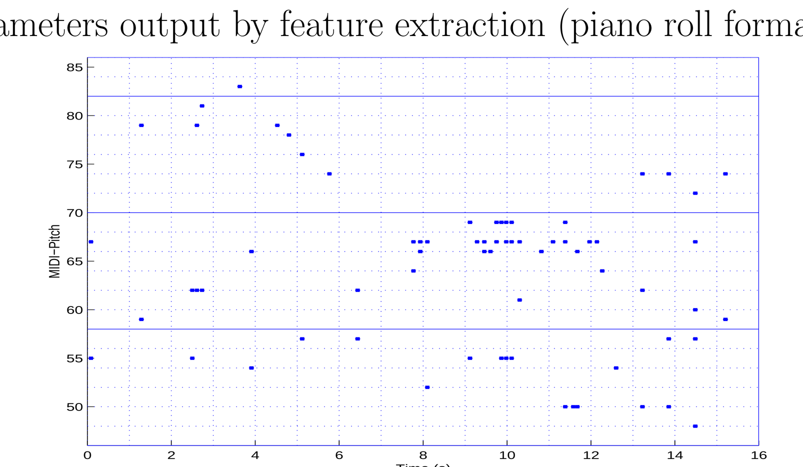
Novelty curve extracted from the PCM version of the Aria:



Attacks extracted after two-stage peak picking:



Note parameters output by feature extraction (piano roll format):

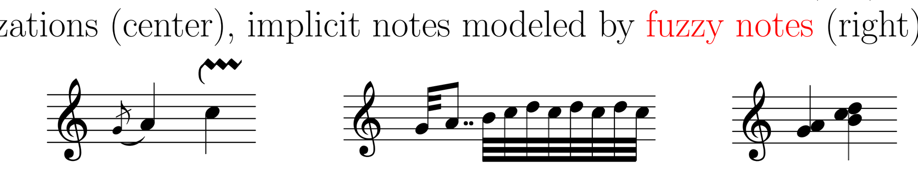


4 Data Modeling

After suitable (time-) quantization, we distinguish two types of score-based note objects:

1. Explicit notes: all time- and pitch parameters are given explicitly
2. Implicit notes: notes with special properties, e.g., trill or arpeggio → different realizations allowed

Example: Two implicit notes, appoggiatura and trill (left), possible realizations (center), implicit notes modeled by fuzzy notes (right).



Fuzzy note = onset time + set of alternative pitches

5 Synchronization Algorithm

5.1 Score-PCM Matches

Assume that score and Δ-quantized extracted PCM-data are given by the sets

$$S = [(s_1, S_{01}, S_{11}), \ldots, (s_s, S_{0s}, S_{1s})]$$

and

$$P_\Delta = [(p_1, P_{01}), \ldots, (p_p, P_{0p})].$$

- s_i: musical onset times,
- p_j: quantized physical onset times,
- S_{0i}, S_{1i}: sets of pitches for the explicit and implicit notes (score),
- P_{0j}: sets of pitches for the (only explicit !) notes (PCM).

Definition 5.1. A Score-PCM match (SP-match) of S and P_Δ is defined to be a partial map μ: $[1:s] \to [1:p]$, which is strictly monotonously increasing on its domain satisfying for all $i \in \text{Domain}(\mu)$: $(S_{0i} \cup S_{1i}) \cap P_{0\mu(i)} \neq \emptyset$.

5.2 Cost Functions For SP-Matches

Definition 5.2. Let $\pi := (\alpha, \beta, \gamma, \delta, \zeta, \Delta) \in \mathbb{R}^6_{\geq 0}$ be a parameter vector. Then the SP-cost of an SP-match μ w.r.t. π between some score S and some Δ-quantized set P_Δ of the corresponding PCM-document is defined as

$$\begin{aligned} C_\pi^{\text{SP}}(\mu|S, P_\Delta) := \alpha \cdot \sum_{(i,j)\in\mu} &\Big(|S_{0i} \setminus P_{0j}| + \lambda(i,j)\Big) \\ +\beta \cdot \sum_{(i,j)\in\mu} &\Big(|P_{0j} \setminus (S_{0i} \cup S_{1i})| + \rho(i,j)\Big) \\ +\gamma \cdot \sum_{k \notin \text{Domain}(\mu)} &\Big(|S_{0k}| + \sigma(k)\Big) \\ +\delta \cdot \sum_{t \notin \text{Image}(\mu)} &|P_{0t}| \\ +\zeta \cdot \sum_{(i,j)\in\mu} &\Big|s_i - p_j \cdot \ell(S)/\ell(P)\Big|. \end{aligned}$$

The single terms account for the following costs:

α-term: non-matched explicit and implicit note objects of the score S ($\lambda(i,j) = 1$ if S has an implicit object at i unmatched by P_Δ at j, $\lambda(i,j) = 0$ otherwise),

β-term: extracted notes not contained in the score; $\rho(i,j) = |P_{0j} \cap S_{1i}| - 1$ if $P_{0j} \cap S_{1i} \neq \emptyset$ and zero otherwise, i.e., for implicit note objects, only one match is free of cost,

γ-term: onset times of the score not belonging to the match μ,

δ-term: notes in P_Δ not having a counterpart in S,

ζ-term: penalizes matches with large relative time deviations.

($\ell(S)$ and $\ell(P)$: differences of last and first musical respectively physical onset times)

Important observation:
If μ is an SP-match then also $\mu' := \mu \setminus \{(i,j)\}$ for some $(i,j) \in \mu$. Hence

$$\begin{aligned} C_\pi^{\text{SP}}(\mu|S, P_\Delta) - C_\pi^{\text{SP}}(\mu'|S, P_\Delta) = \alpha \cdot &\Big(|S_{0i} \setminus P_{0j}| + \lambda(i,j)\Big) \\ +\beta \cdot &\Big(|P_{0j} \setminus (S_{0i} \cup S_{1i})| + \rho(i,j)\Big) \\ -\gamma \cdot &\Big(|S_{0i}| + \sigma(i)\Big) \qquad (1) \\ -\delta \cdot &|P_{0j}| \\ +\zeta \cdot &\Big|s_i - p_j \cdot \ell(S)/\ell(P)\Big|. \end{aligned}$$

5.3 Cost-Optimal SP-Matches

Determine cost-minimizing SP-match using dynamic programming:
Recursively define a matrix $C = (c_{ij})$ with $i \in [0:s]$ and $j \in [0:p]$,

1. Initialize $c_{0j} := c_{i0} := C_\pi^{\text{SP}}(\emptyset|S, P_\Delta)$ for all $i \in [0:s], j \in [0:p]$.
   (costs for the case that there is no match at all between S and P_Δ)
2. For $(i,j) \in [1:s] \times [1:p]$,

$$c_{ij} := \min\{c_{i,j-1}, c_{i-1,j}, c_{i-1,j-1} + d_{ij}^{\text{SP}}\},$$

where

$$d_{ij}^{\text{SP}} := \begin{cases} \text{right hand side of Eq. (1)}, & \text{if } (S_{0i} \cup S_{1i}) \cap P_{0j} \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

By (1), c_{ij} is cost for a cost-minimizing SP-match in $[1:i] \times [1:j] \subset [1:s] \times [1:p]$.
Hence, c_{sp} expresses the minimal cost of a global SP-match.

6 An Example

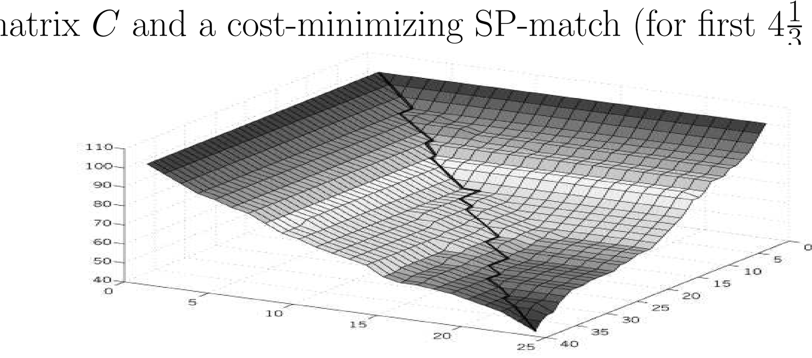
Score-based note objects S for first two measures of the Aria (left) and quantized note parameters P_Δ extracted from a PCM version (right):

S				P_Δ		
i	s_i	S_{0i}	S_{1i}	j	p_j	P_{0j}
1	0	{55, 79}	∅	1	0	{55, 67}
2	1	{59, 79}	∅	2	1.23	{59, 79}
3	2	{62}	{79, 81}	3	2.44	{55, 62}
				4	2.56	{62, 79}
				5	2.68	{62, 81}
4	2.75	{83}	∅	6	3.58	{83}
5	3	{54, 81}	∅	7	3.86	{54, 66}
6	3.5	∅	{78, 79}	8	4.47	{79}
				9	4.75	{78}
7	4	{57}	{74, 76}	10	5.06	{57, 76}
				11	5.71	{74}
8	5	{62}	∅	12	6.39	{57, 62}

Corresponding cost matrix:

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	102	102	102	102	102	102	102	102	102	102	102	102	102
1	102	**98**	98	98	98	98	98	98	98	98	98	98	98
2	102	98	**94**	94	94	94	94	94	94	94	94	94	94
3	102	98	94	**90**	90	90	90	90	90	90	90	90	90
4	102	98	94	90	90	90	**88**	88	88	88	88	88	88
5	102	98	94	90	90	88	88	**85**	85	85	85	85	85
6	102	98	94	90	88	88	88	85	**83**	83	83	83	83
7	102	98	94	90	88	88	88	85	83	83	**79**	79	79
8	102	98	94	90	88	86	86	85	83	83	79	79	**77**

Cost matrix C and a cost-minimizing SP-match (for first $4\frac{1}{3}$ measures):



7 Conclusions

- Tests on a variety of classical polyphonic piano pieces (lengths 10 – 60 secs) played on various instruments yield good results.
- Crucial: Choice of parameter vector π in cost function.