Fundamentals of Music Processing

Chapter 3: Music Synchronization

Meinard Müller
International Audio Laboratories Erlangen

www.music-processing.de

Springer

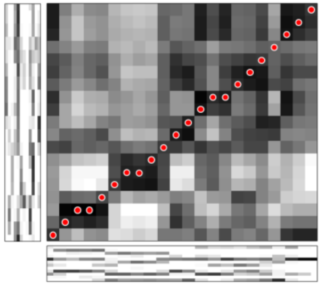
AUDIO LABS

Book: Fundamentals of Music Processing

Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
483 p., 249 illus., 30 illus. in color, hardcover
ISBN: 978-3-319-21944-8
Springer, 2015

Accompanying website:
www.music-processing.de

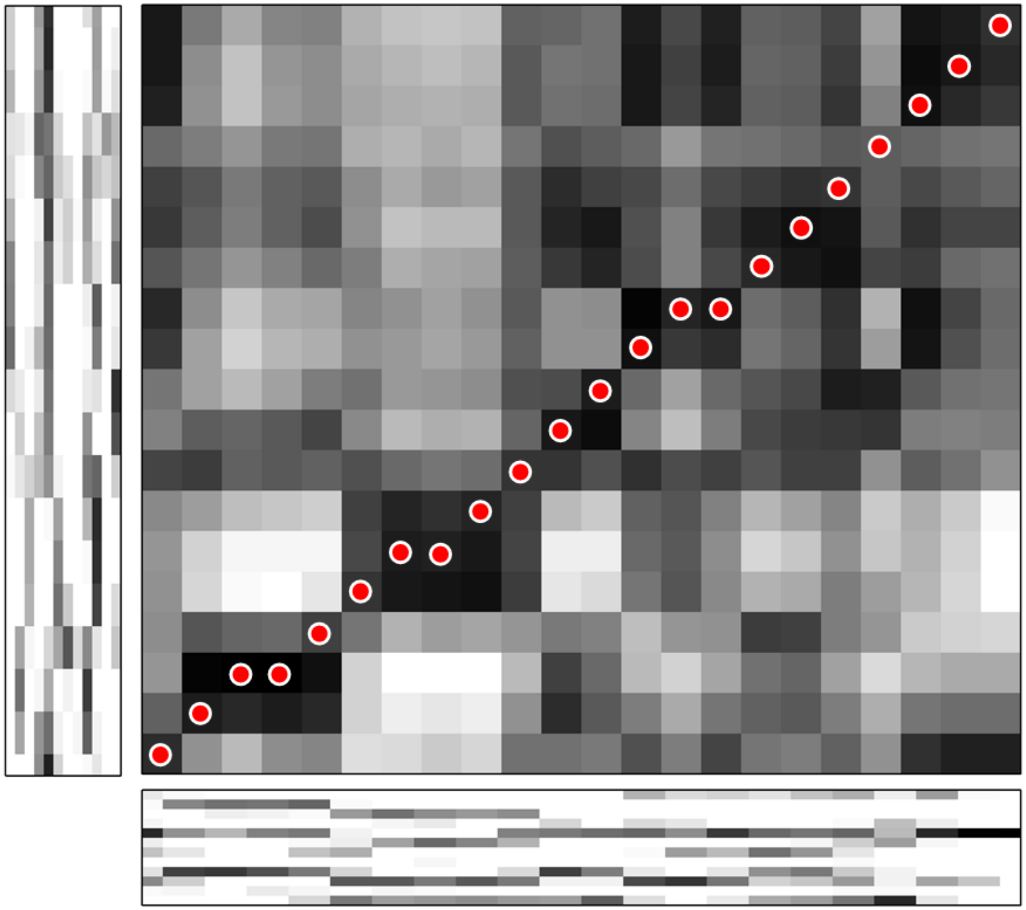
Chapter 3: Music Synchronization



As a first music processing task, we study in Chapter 3 the problem of music synchronization. The objective is to temporally align compatible representations of the same piece of music. Considering this scenario, we explain the need for musically informed audio features. In particular, we introduce the concept of chroma-based music features, which capture properties that are related to harmony and melody. Furthermore, we study an alignment technique known as dynamic time warping (DTW), a concept that is applicable for the analysis of general time series. For its efficient computation, we discuss an algorithm based on dynamic programming—a widely used method for solving a complex problem by breaking it down into a collection of simpler subproblems.

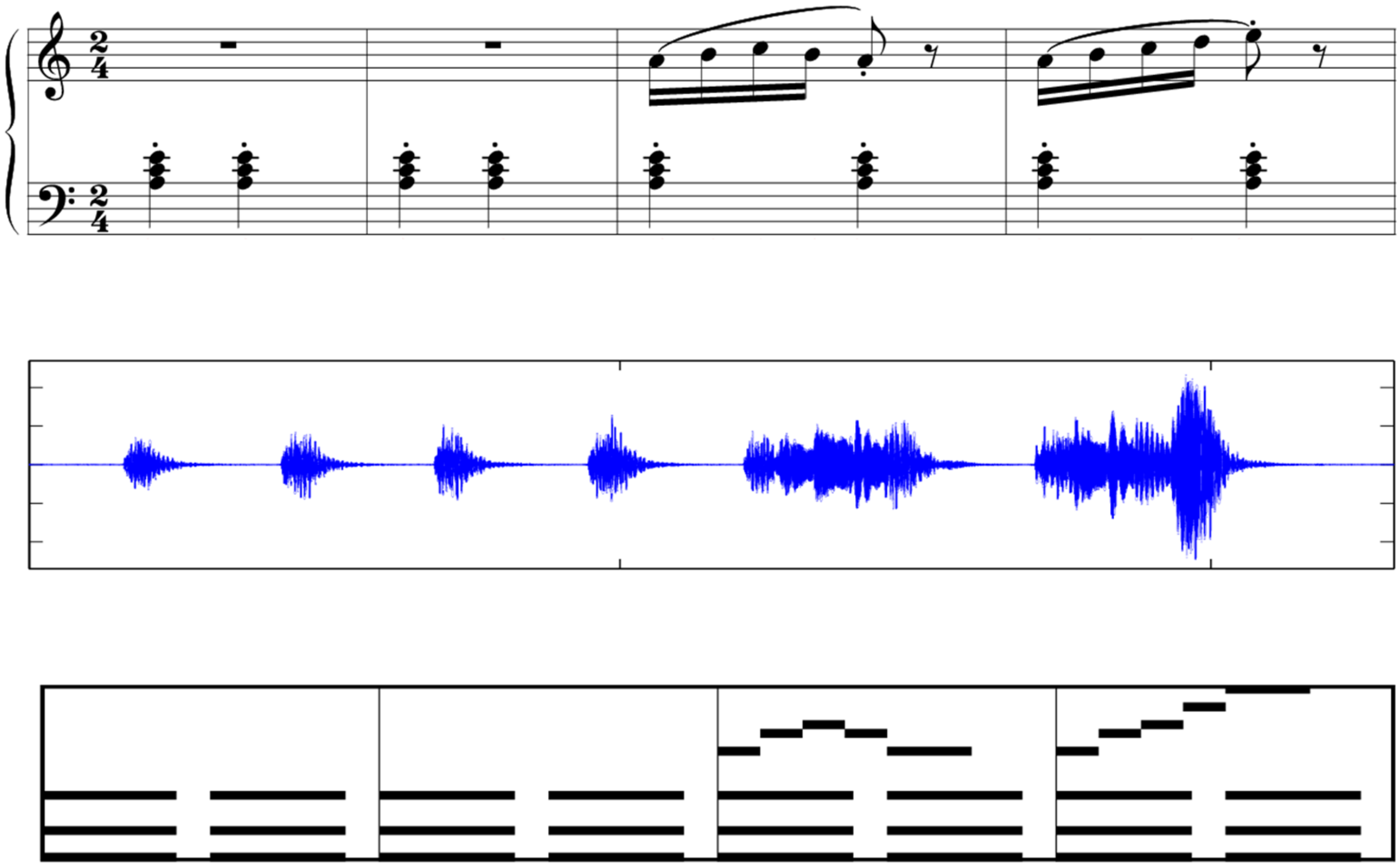
3 Music Synchronization

Teaser



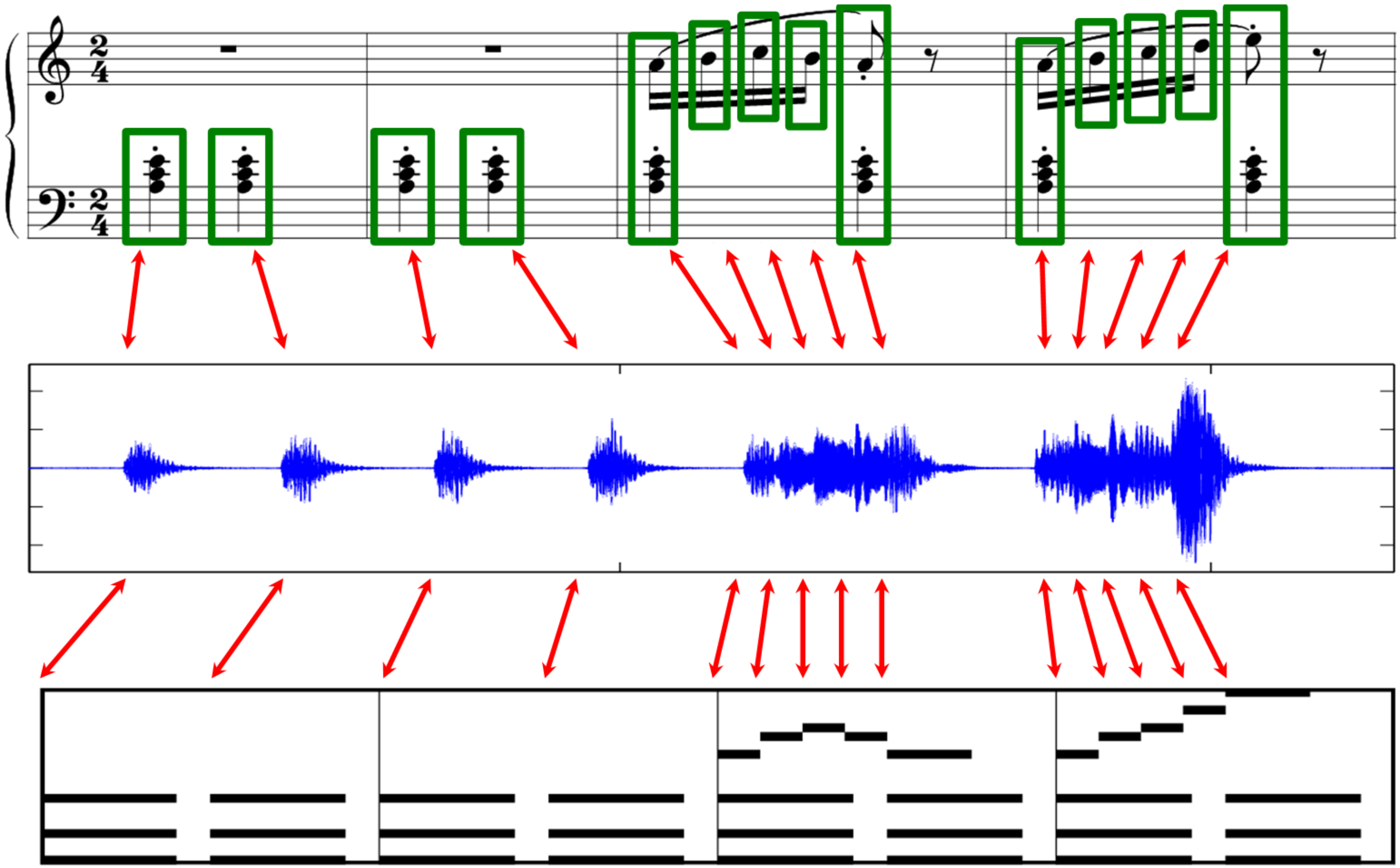
3 Music Synchronization

Fig. 3.1



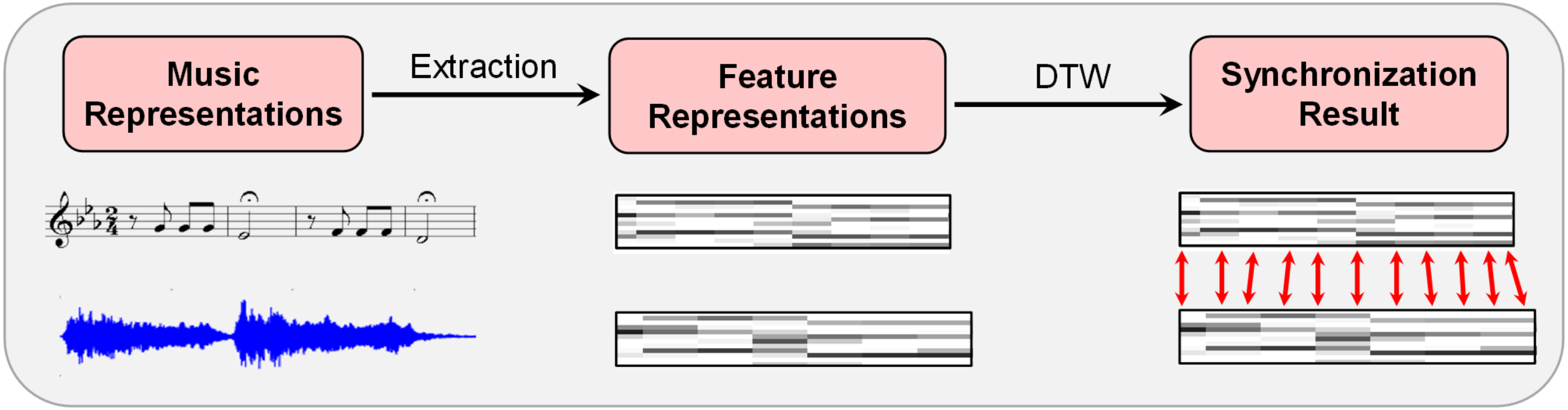
3 Music Synchronization

Fig. 3.1



3 Music Synchronization

Fig. 3.2



3.1 Audio Features

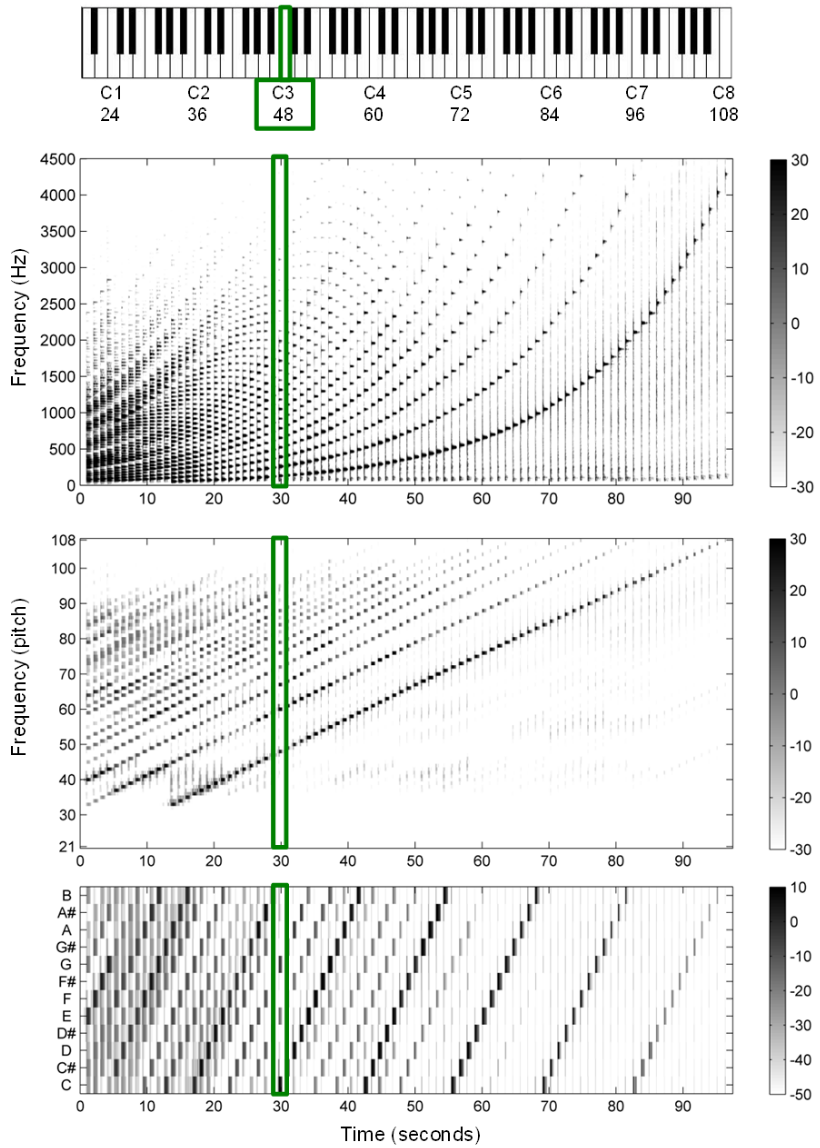
Fig. 3.3

Chromatic scale played on a real piano

Magnitude Spectrogram

Pitch-based log-frequency Spectrogram

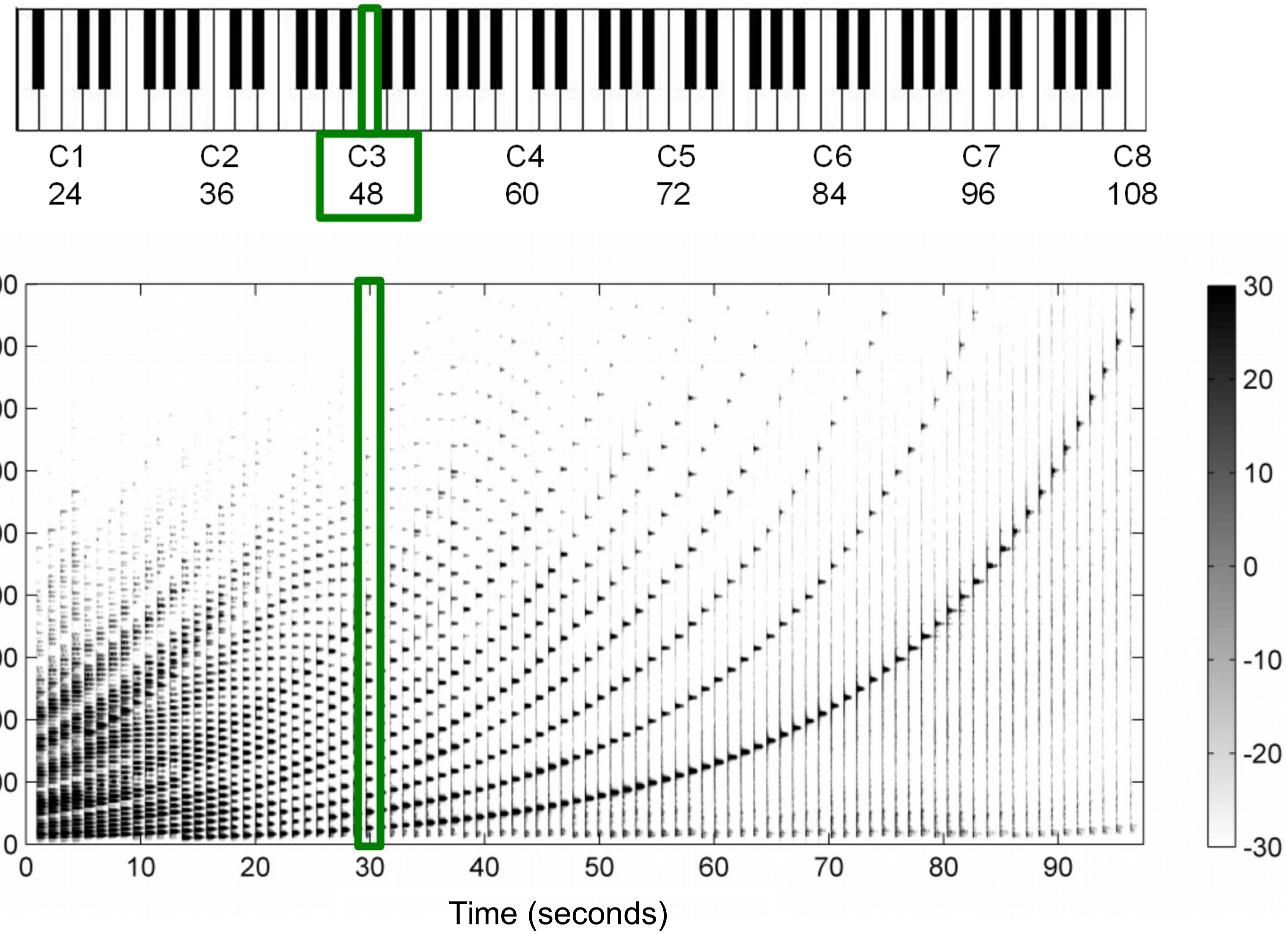
Chrogmagram



3.1 Audio Features

Fig. 3.3

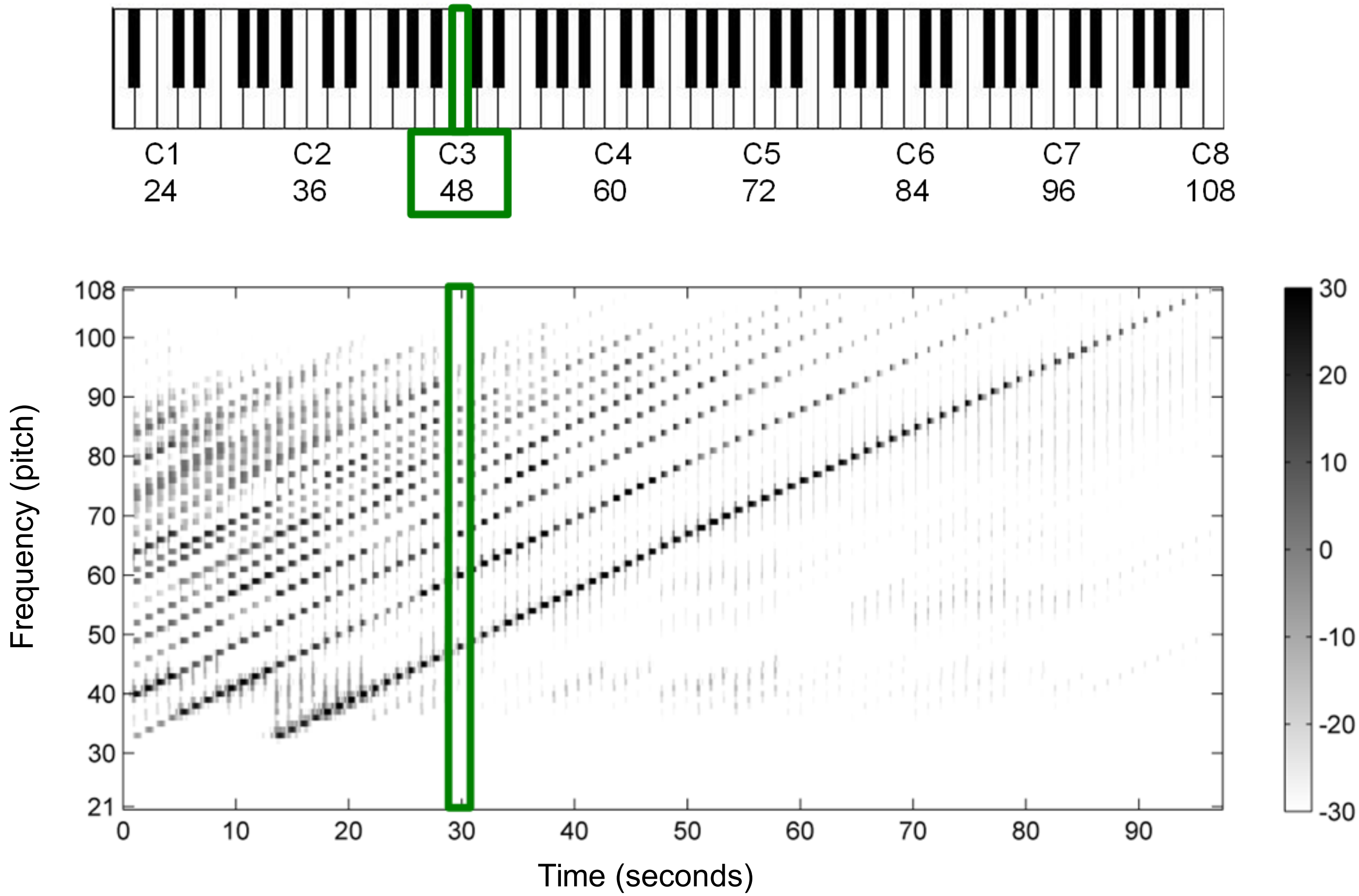
$$\mathcal{X}(n,k) := \sum_{\ell=0}^{N-1} x(\ell + nH) w(\ell) \exp(-2\pi i k \ell / N)$$



3.1 Audio Features

$$F_{\mathrm{pitch}}(p) = 2^{(p-69)/12} \cdot 440$$

$$P(p) := \{k : F_{\mathrm{pitch}}(p-0.5) \leq F_{\mathrm{coef}}(k) < F_{\mathrm{pitch}}(p+0.5)\}$$

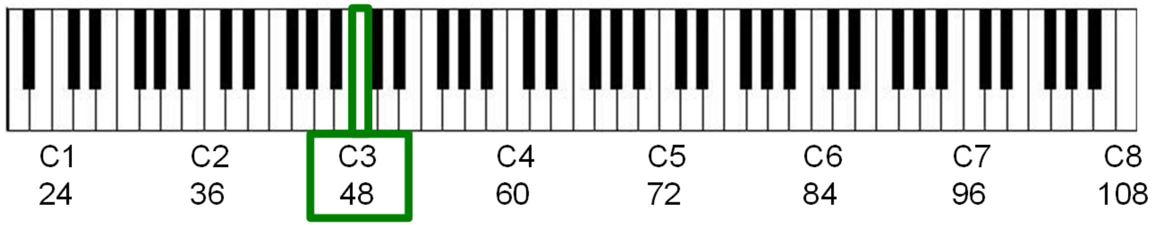


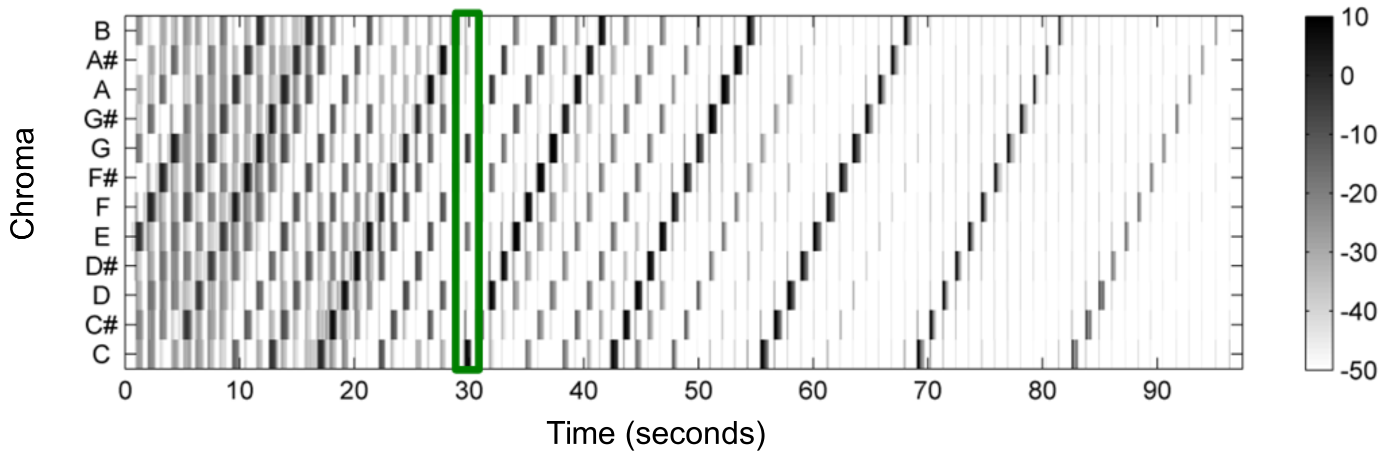
$$\mathcal{Y}_{\mathrm{LF}}(n,p) := \sum_{k \in P(p)} |\mathcal{X}(n,k)|^2$$

3.1 Audio Features

Fig. 3.3

$$\mathcal{C}(n,c) := \sum_{\{p\in[0:127]\,:\,p \bmod 12 = c\}} \mathcal{Y}_{\mathrm{LF}}(n,p)$$





3.1 Audio Features

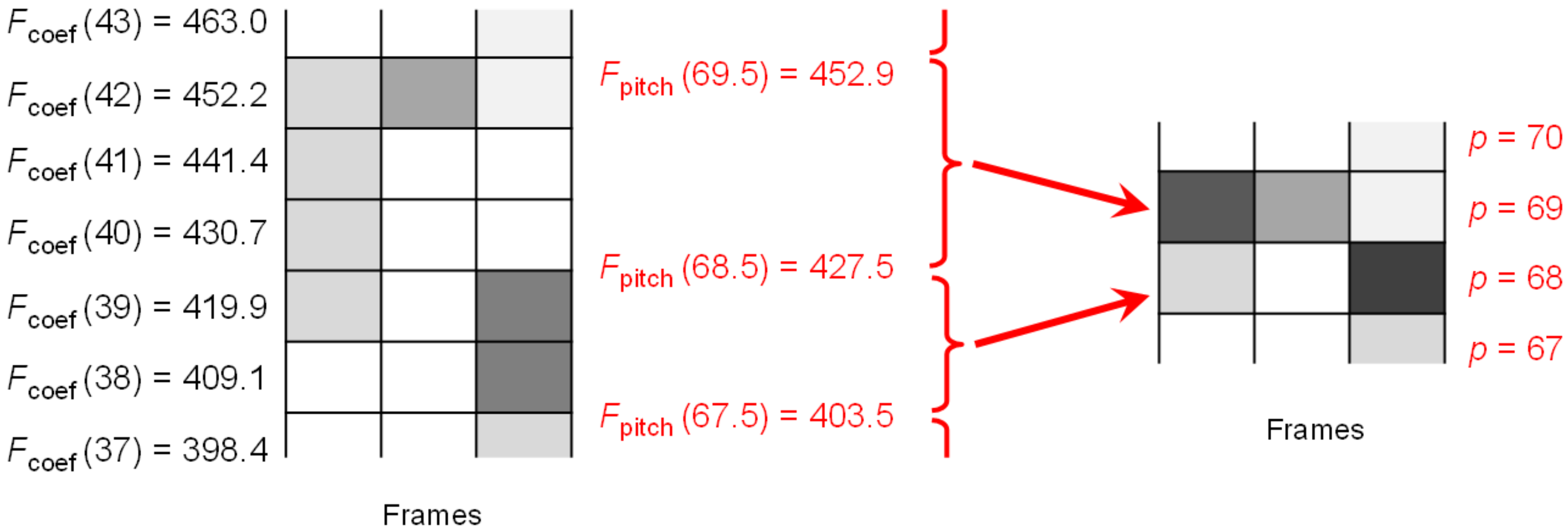
$$P(p) := \{k : F_{\mathrm{pitch}}(p-0.5) \leq F_{\mathrm{coef}}(k) < F_{\mathrm{pitch}}(p+0.5)\}$$

Note	p	$F_{\mathrm{pitch}}(p)$	$F_{\mathrm{pitch}}(p-0.5)$	$F_{\mathrm{pitch}}(p+0.5)$	$\mathrm{BW}(p)$
C4	60	261.63	254.18	269.29	15.11
C$^\sharp$4	61	277.18	269.29	285.30	16.01
D4	62	293.66	285.30	302.27	16.97
D$^\sharp$4	63	311.13	302.27	320.24	17.97
E4	64	329.63	320.24	339.29	19.04
F4	65	349.23	339.29	359.46	20.18
F$^\sharp$4	66	369.99	359.46	380.84	21.37
G4	67	392.00	380.84	403.48	22.65
G$^\sharp$4	68	415.30	403.48	427.47	23.99
A4	69	440.00	427.47	452.89	25.41
A$^\sharp$4	70	466.16	452.89	479.82	26.93
B4	71	493.88	479.82	508.36	28.53
C5	72	523.25	508.36	538.58	30.23

3.1 Audio Features

Fig. 3.4

$$\mathcal{Y}_{\mathrm{LF}}(n,p) := \sum_{k \in P(p)} |\mathcal{X}(n,k)|^2$$

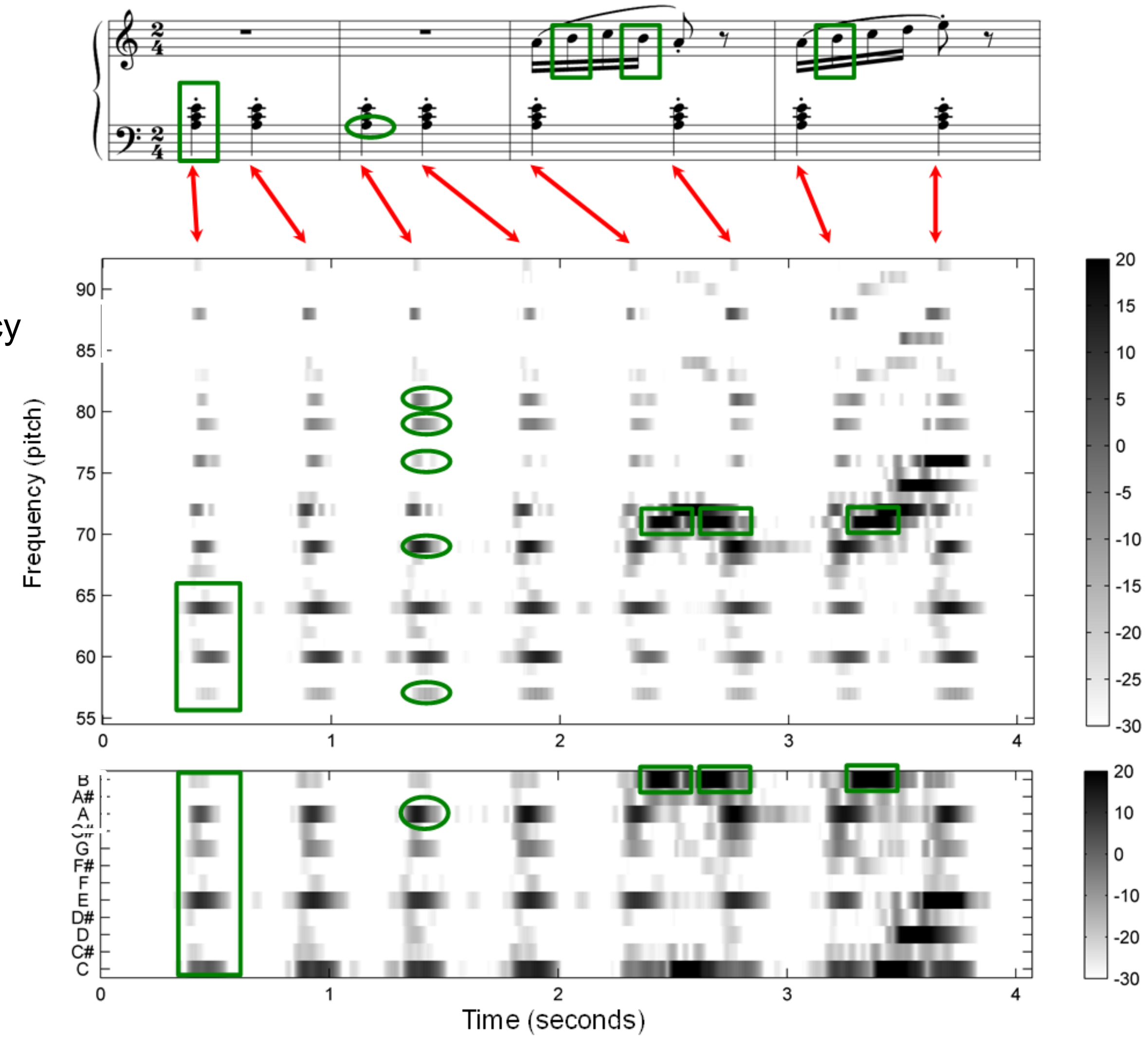


3.1 Audio Features

Fig. 3.5

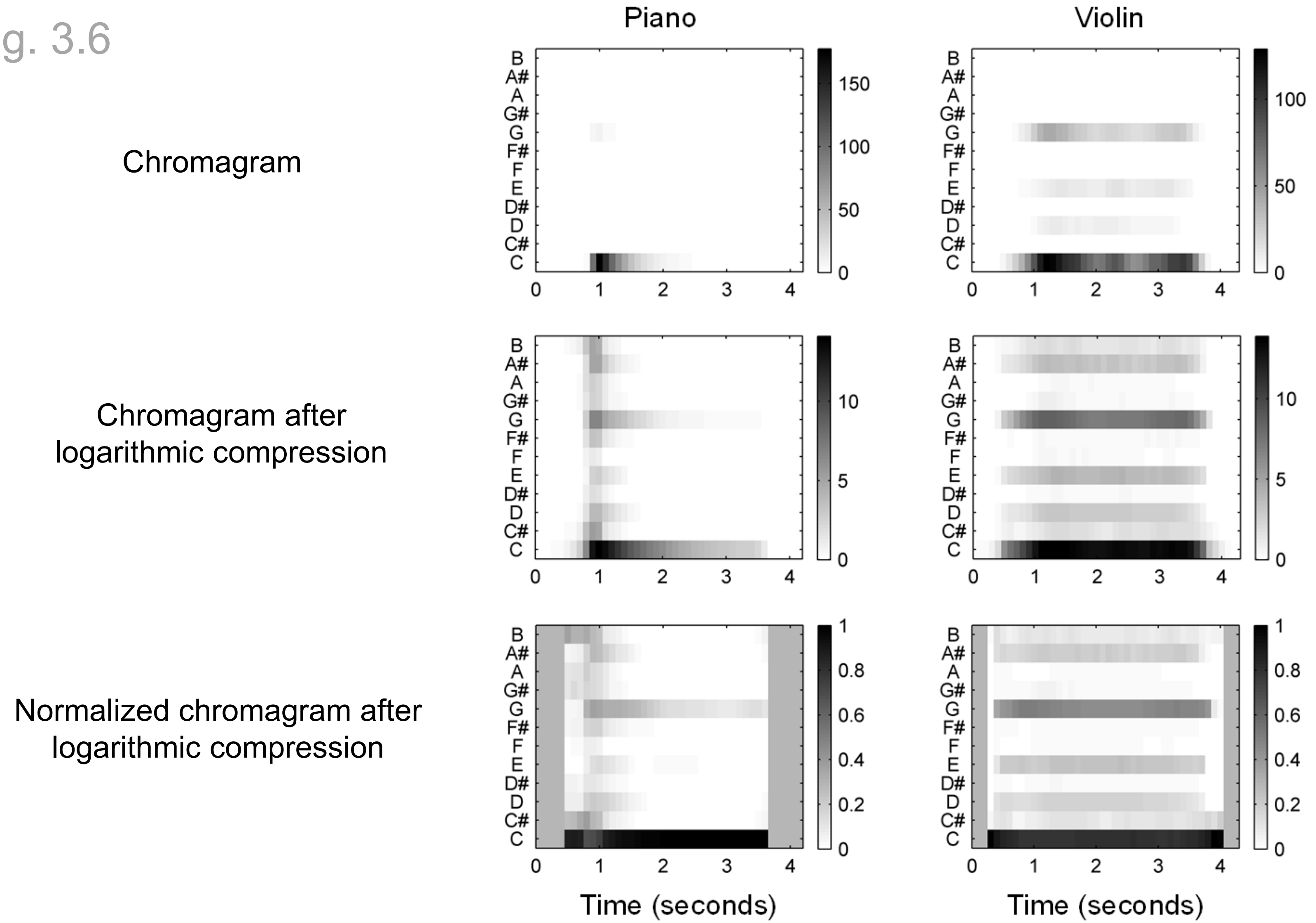
Pitch-based log-frequency Spectrogram

Chromagram



3.1 Audio Features

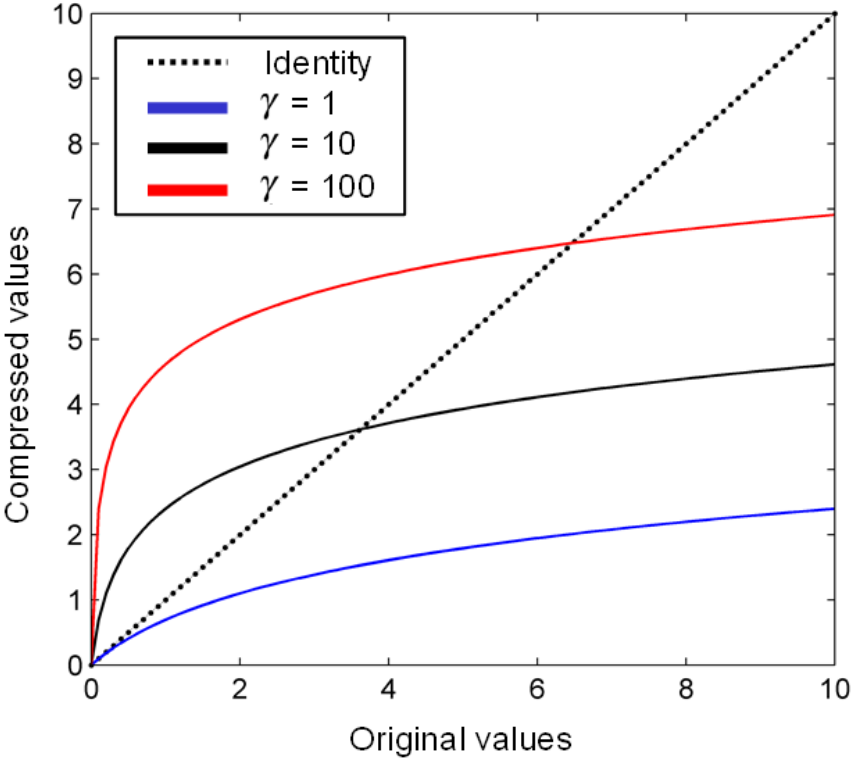
Fig. 3.6

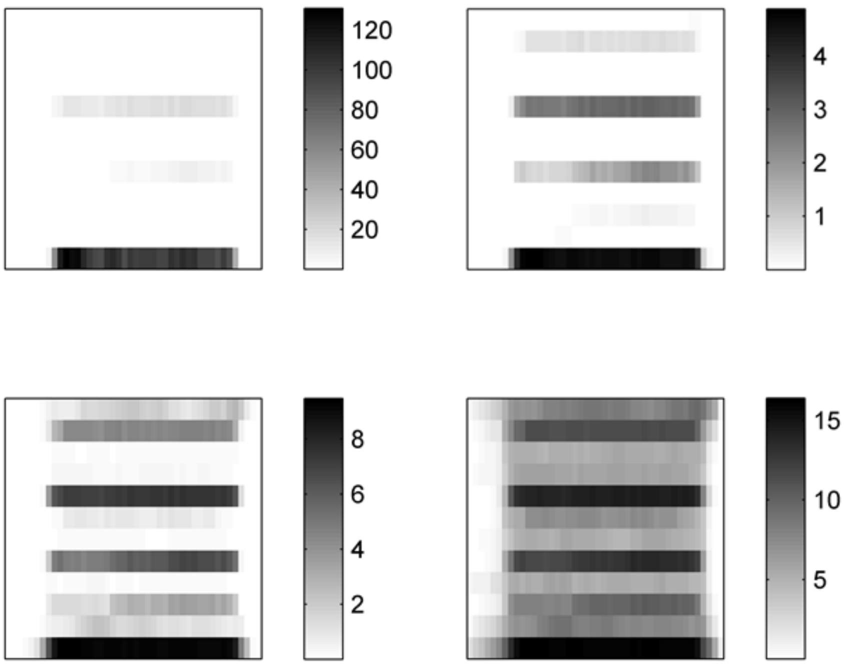


3.1 Audio Features

Fig. 3.7

Logarithmic compression





3.1 Audio Features

Fig. 3.8

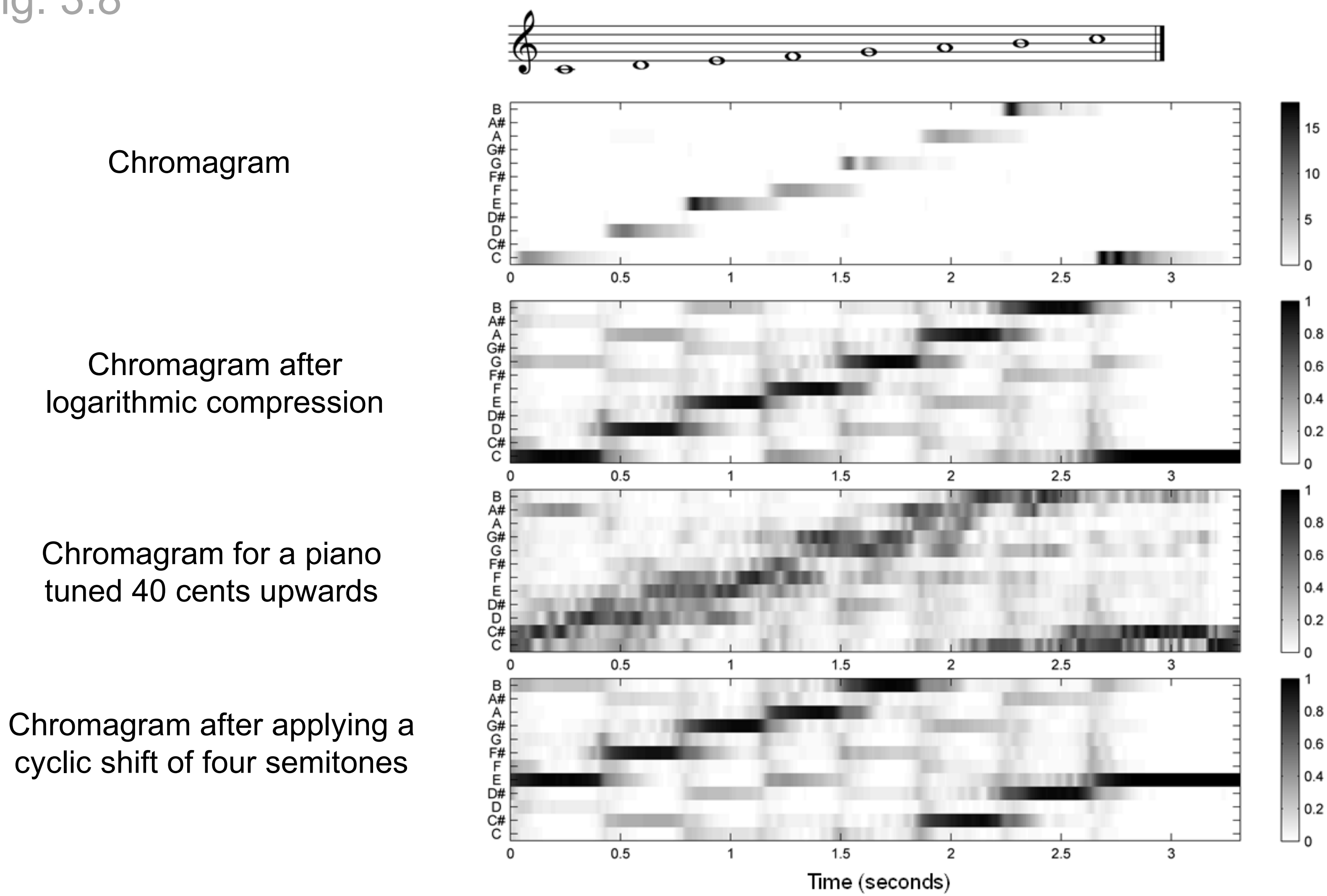
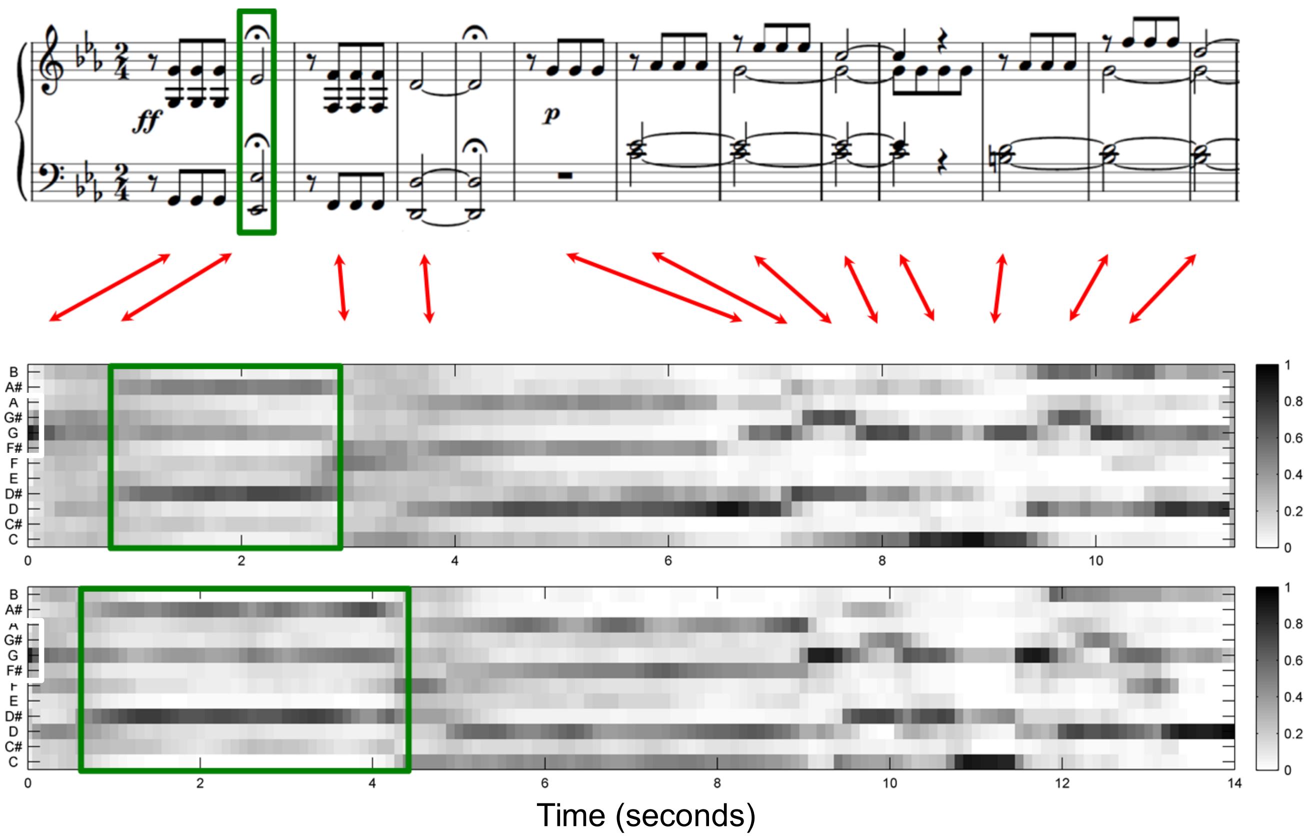


Fig. 3.9

Beethoven's Fifth Symphony

Orchestra performance

Piano performance



Compressed and normalized chromagrams

3.1 Audio Features

Fig. 3.9

Beethoven's Fifth Symphony

Orchestra performance

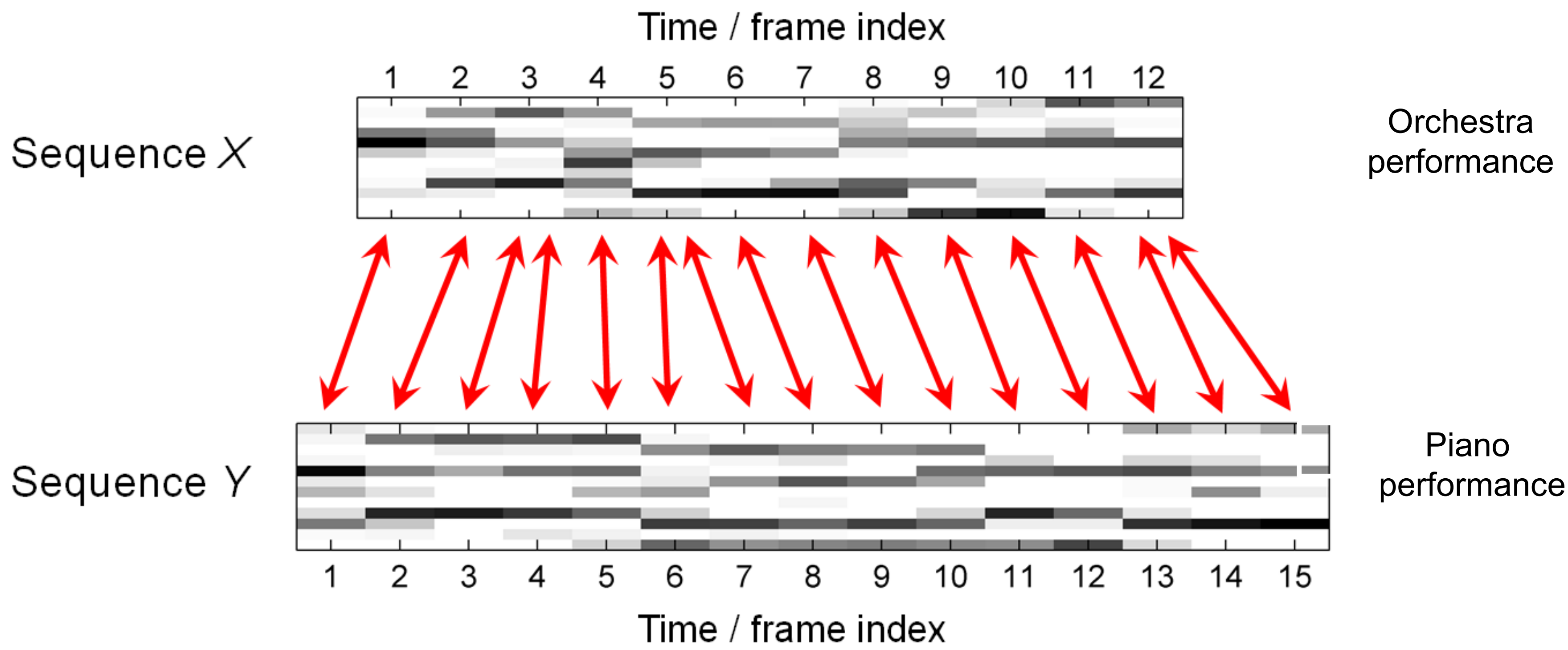
Piano performance

Time (seconds)

Enhanced compressed and normalized chromagrams

3.2 Dynamic Time Warping

Fig. 3.10



3.2 Dynamic Time Warping

Fig. 3.11

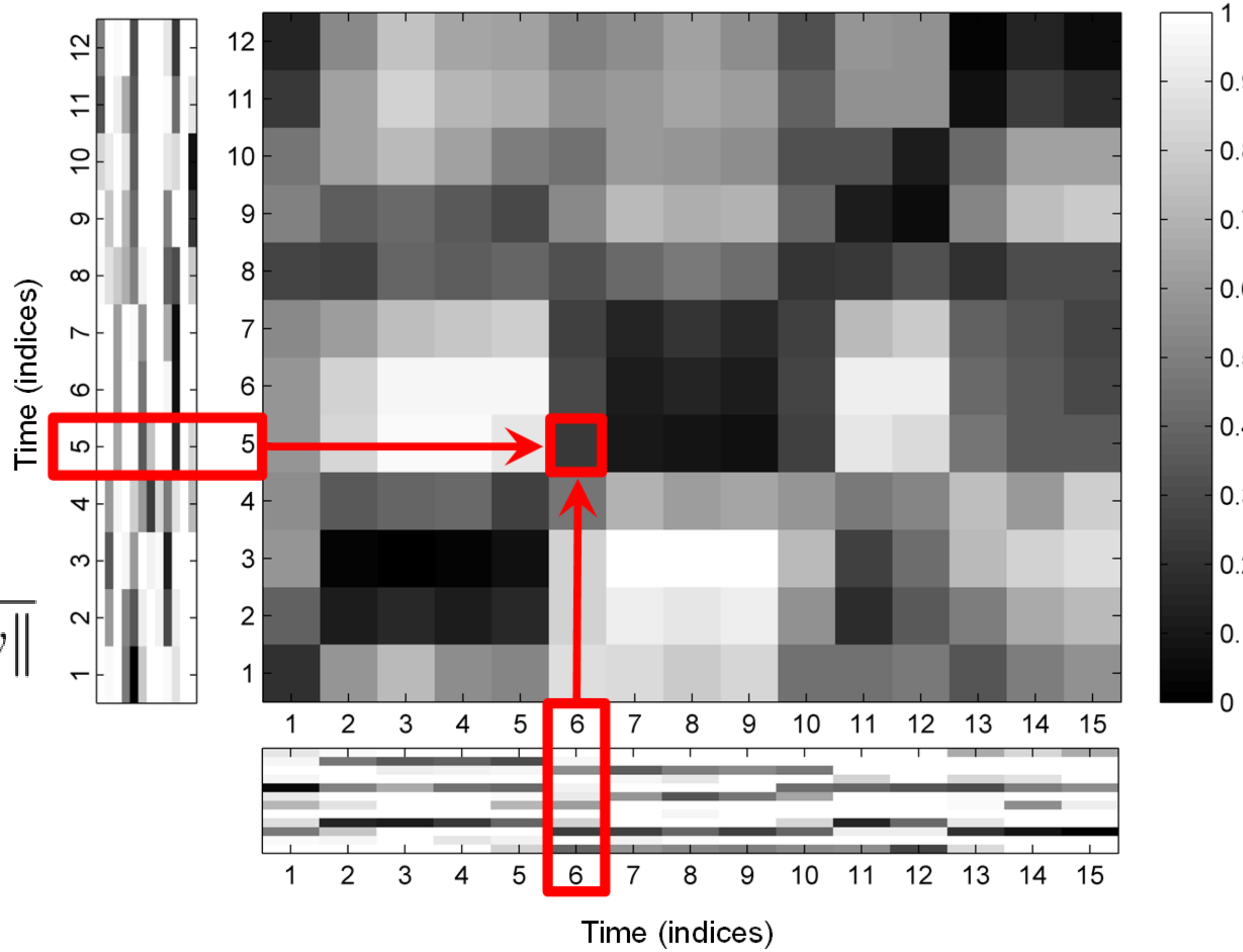
Local cost measure

$$\mathbf{C}(n,m) := c(x_n, y_m)$$

Cosine distance

$$c(x,y) := 1 - \frac{\langle x|y\rangle}{\|x\| \cdot \|y\|}$$

Cost matrix



3.2 Dynamic Time Warping

Fig. 3.11

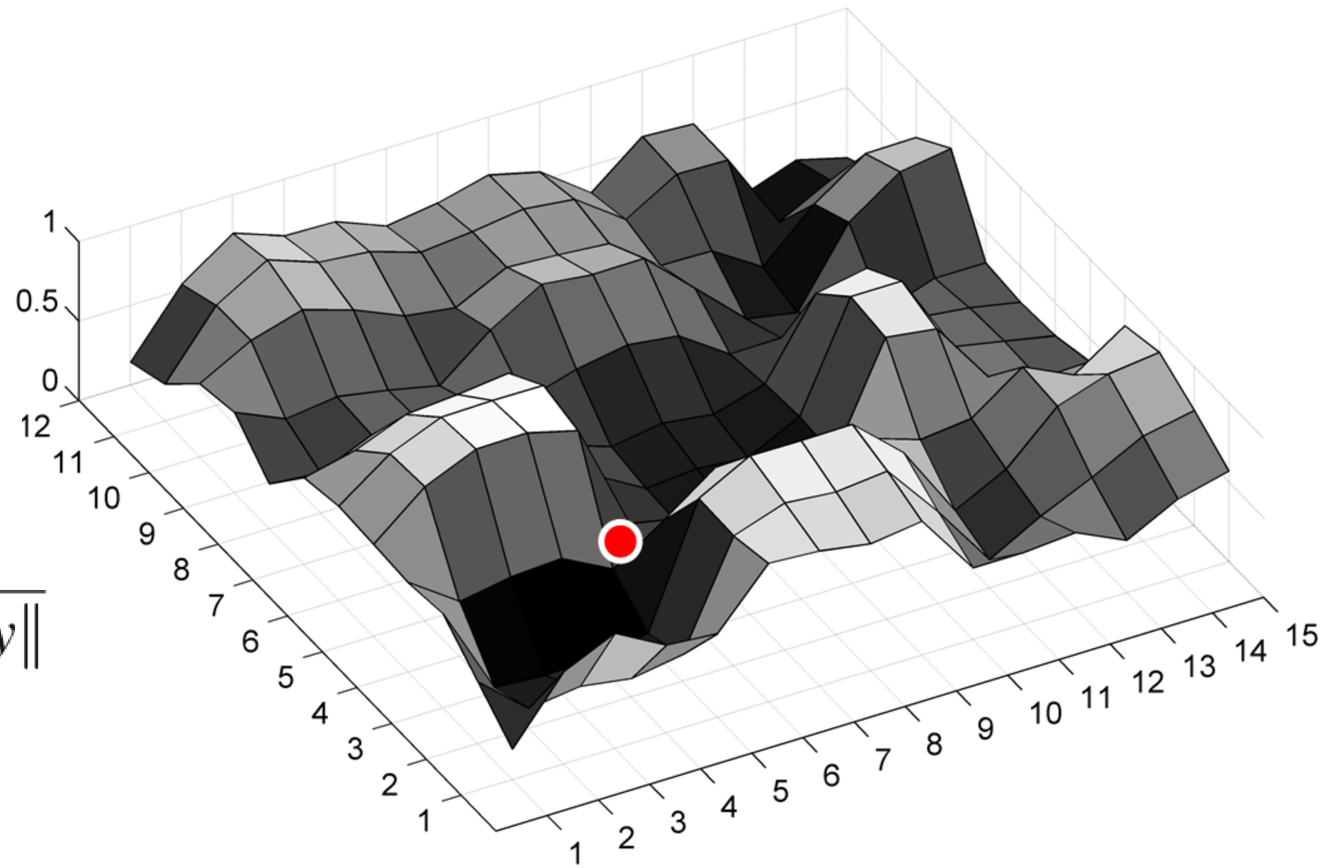
Cost matrix

Local cost measure

$$\mathbf{C}(n,m) := c(x_n, y_m)$$

Cosine distance

$$c(x,y) := 1 - \frac{\langle x|y \rangle}{\|x\| \cdot \|y\|}$$

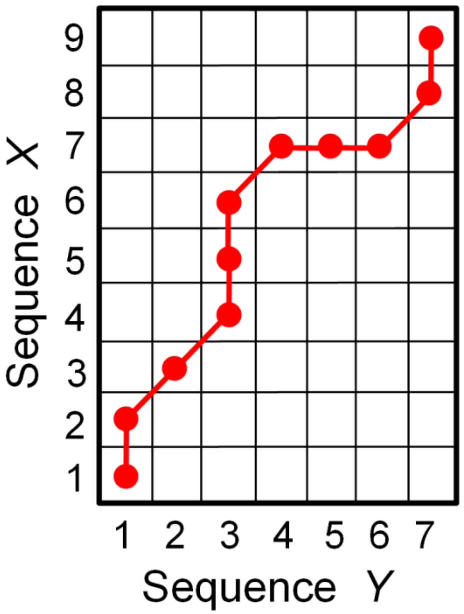


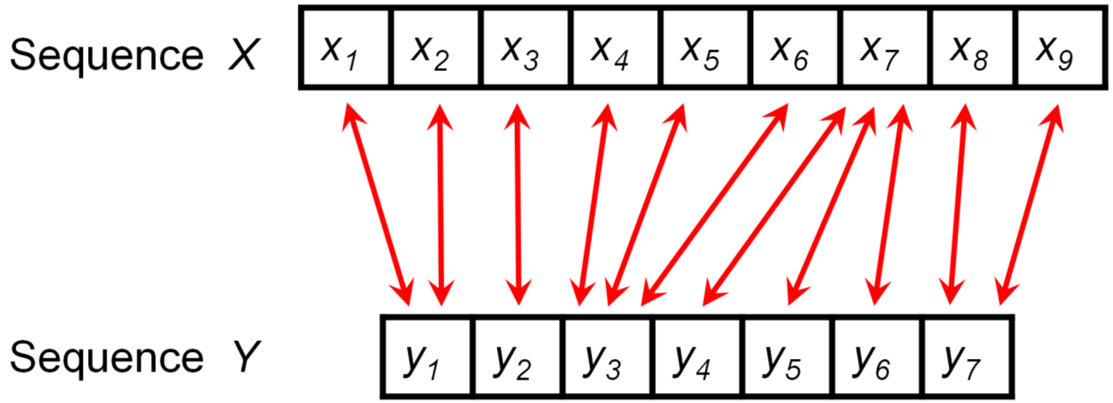
3.2 Dynamic Time Warping

Fig. 3.12

Warping path: $P = (p_1, \ldots, p_L)$

$$p_\ell = (n_\ell, m_\ell) \in [1:N] \times [1:M] \text{ for } \ell \in [1:L]$$





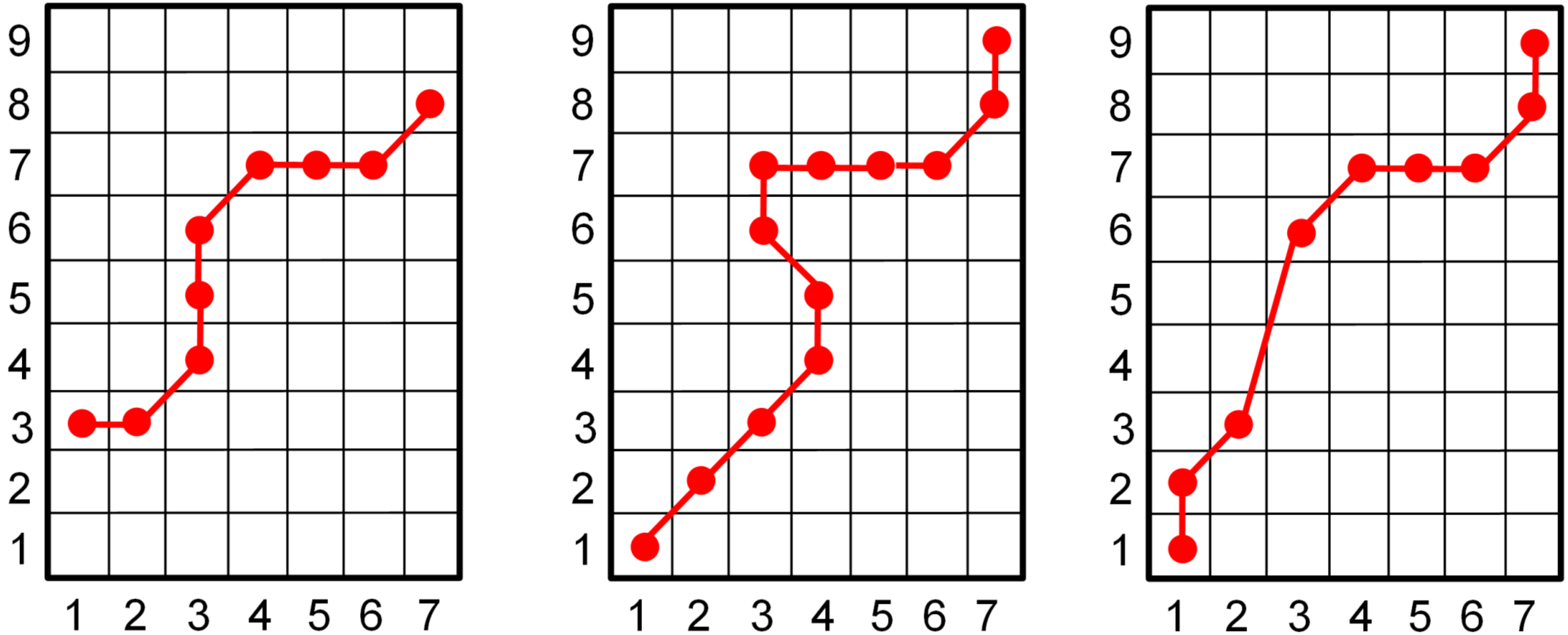
Boundary condition: $p_1 = (1,1)$ and $p_L = (N,M)$.

Monotonicity condition: $n_1 \leq n_2 \leq \ldots \leq n_L$ and $m_1 \leq m_2 \leq \ldots \leq m_L$.

Step size condition: $p_{\ell+1} - p_\ell \in \{(1,0),(0,1),(1,1)\}$ for $\ell \in [1:L-1]$.

3.2 Dynamic Time Warping

Fig. 3.13



Boundary condition: $p_1 = (1,1)$ and $p_L = (N,M)$.

Monotonicity condition: $n_1 \leq n_2 \leq \ldots \leq n_L$ and $m_1 \leq m_2 \leq \ldots \leq m_L$.

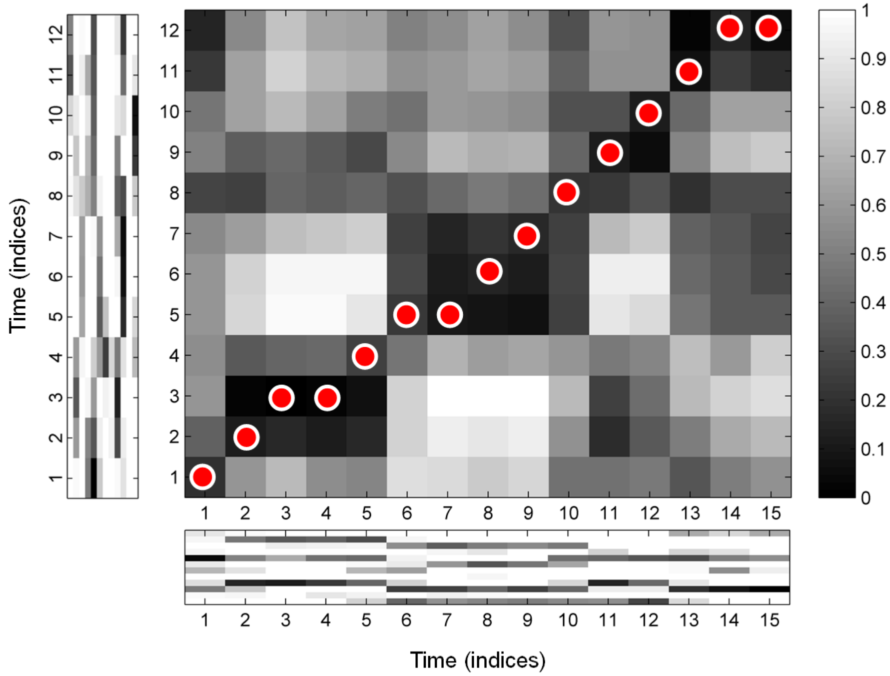
Step size condition: $p_{\ell+1} - p_\ell \in \{(1,0),(0,1),(1,1)\}$ for $\ell \in [1:L-1]$.

3.2 Dynamic Time Warping

Fig. 3.14

Total cost of a warping path:

$$c_P(X,Y) := \sum_{\ell=1}^{L} c(x_{n_\ell}, y_{m_\ell}) = \sum_{\ell=1}^{L} \mathbf{C}(n_\ell, m_\ell).$$



DTW distance:

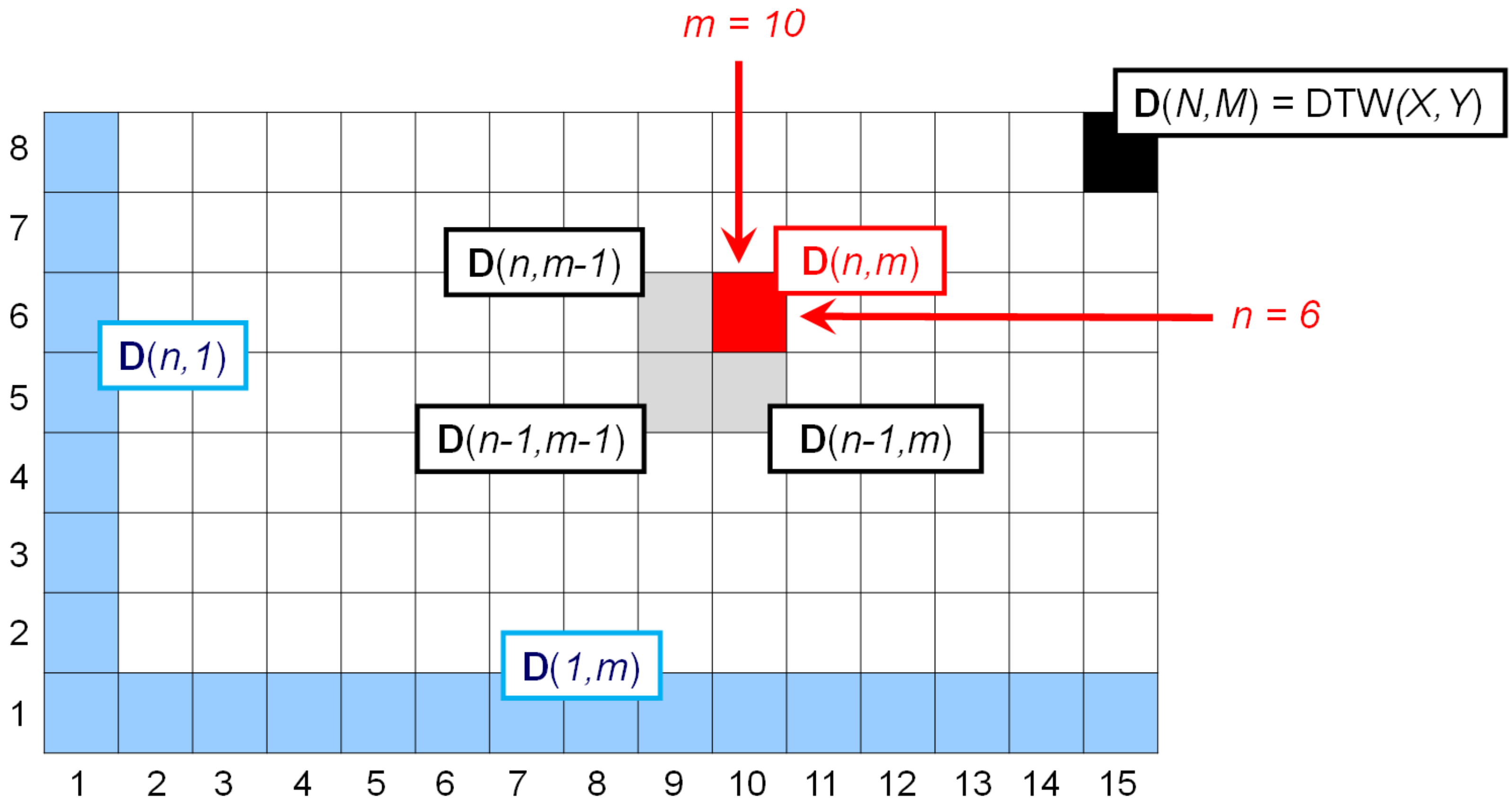
$$\begin{aligned}\mathrm{DTW}(X,Y) &:= c_{P^*}(X,Y)\\ &= \min\{c_P(X,Y) \mid P \text{ is an } (N,M)\text{-warping path}\}.\end{aligned}$$

3.2 Dynamic Time Warping

Fig. 3.15

DTW prefix:

$$\mathbf{D}(n,m) := \mathrm{DTW}(X(1:n), Y(1:m)).$$



3.2 Dynamic Time Warping

Fig. 3.15

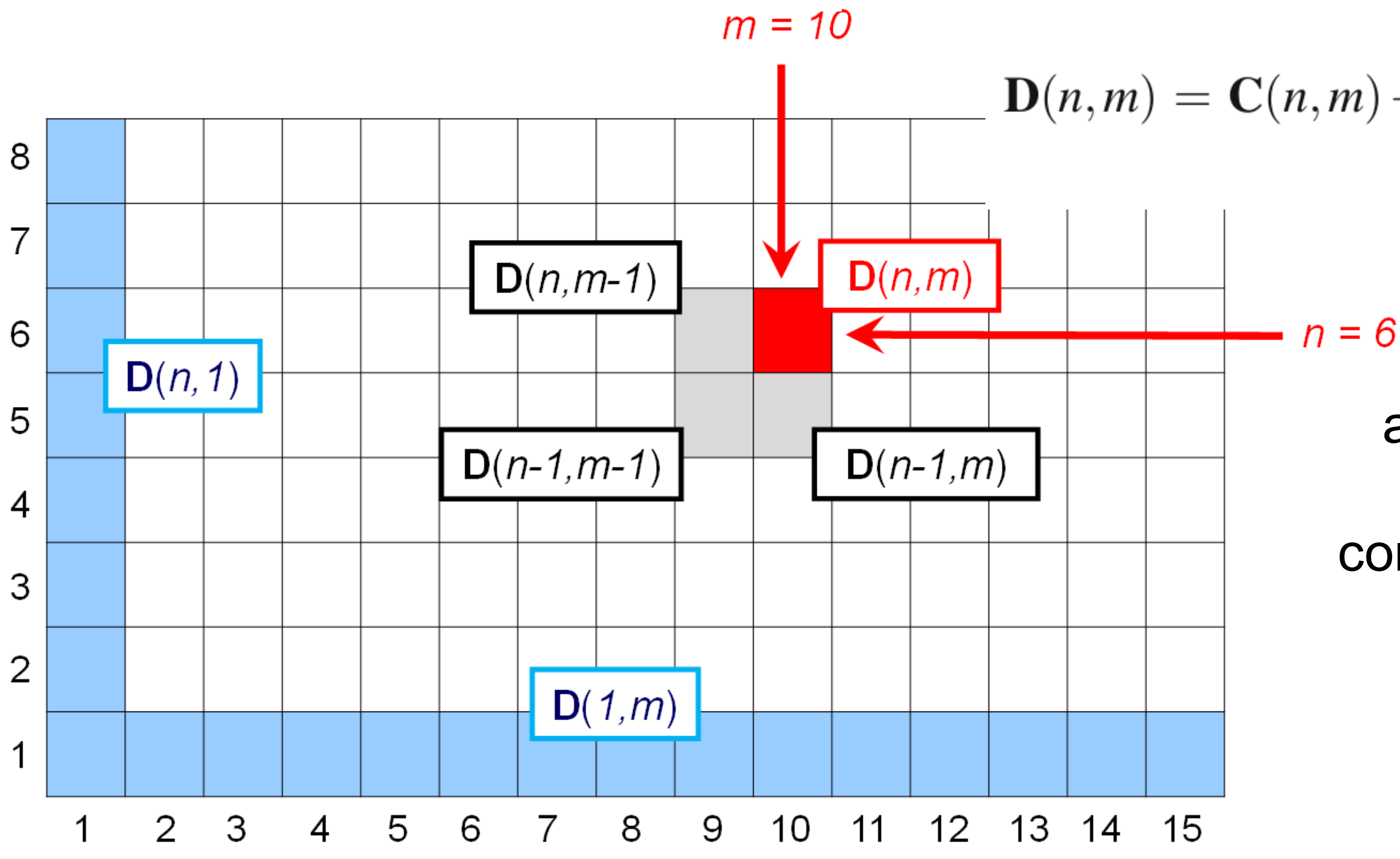
DTW prefix:

$$\mathbf{D}(n,m) := \mathrm{DTW}(X(1:n), Y(1:m)).$$

$$\mathbf{D}(n,1) = \sum_{k=1}^{n} \mathbf{C}(k,1) \quad \text{for} \quad n \in [1:N],$$

$$\mathbf{D}(1,m) = \sum_{k=1}^{m} \mathbf{C}(1,k) \quad \text{for} \quad m \in [1:M],$$

$$\mathbf{D}(n,m) = \mathbf{C}(n,m) + \min \begin{cases} \mathbf{D}(n-1,m-1) \\ \mathbf{D}(n-1,m) \\ \mathbf{D}(n,m-1) \end{cases}$$



accumulated cost matrix D can be computed recursively

3.2 Dynamic Time Warping

Table 3.2

Algorithm: DTW

Input: Cost matrix **C** of size $N \times M$
Output: Accumulated cost matrix **D**
Optimal warping path P^*

Procedure: Initialize $(N \times M)$ matrix **D** by $\mathbf{D}(n,1) = \sum_{k=1}^{n} \mathbf{C}(k,1)$ for $n \in [1:N]$ and $\mathbf{D}(1,m) = \sum_{k=1}^{m} \mathbf{C}(1,k)$ for $m \in [1:M]$. Then compute in a nested loop for $n = 2,\ldots,N$ and $m = 2,\ldots,M$:

$$\mathbf{D}(n,m) = \mathbf{C}(n,m) + \min\{\mathbf{D}(n-1,m-1), \mathbf{D}(n-1,m), \mathbf{D}(n,m-1)\}.$$

Set $\ell = 1$ and $q_\ell = (N,M)$. Then repeat the following steps until $q_\ell = (1,1)$:

Increase ℓ by one and let $(n,m) = q_{\ell-1}$.
If $n = 1$, then $q_\ell = (1, m-1)$,
else if $m = 1$, then $q_\ell = (n-1, m)$,
else $q_\ell = \operatorname{argmin}\{\mathbf{D}(n-1,m-1), \mathbf{D}(n-1,m), \mathbf{D}(n,m-1)\}$.
(If 'argmin' is not unique, take lexicographically smallest cell.)

Set $L = \ell$ and return $P^* = (q_L, q_{L-1}, \ldots, q_1)$ as well as **D**.

cost matrix C

accumulated cost matrix D

backtracking the optimal path

$$q_{\ell+1} = (1, m-1) \quad \text{if } n = 1,$$

$$q_{\ell+1} = (n-1, m) \quad \text{if } m = 1,$$

$$q_{\ell+1} = \operatorname{argmin} \begin{cases} \mathbf{D}(n-1,m-1), \\ \mathbf{D}(n-1,m), \\ \mathbf{D}(n,m-1) \end{cases}$$

3.2 Dynamic Time Warping

Fig. 3.16

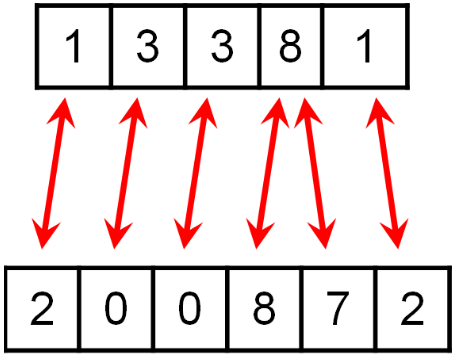
Cost matrix

	2	0	0	8	7	2
1	1	1	1	7	6	1
8	6	8	8	0	1	6
3	1	3	3	5	4	1
3	1	3	3	5	4	1
1	1	1	1	7	6	1

Accumulated cost matrix and optimal warping path

	2	0	0	8	7	2
1	10	10	11	14	13	9
8	9	11	13	7	8	14
3	3	5	7	10	12	13
3	2	4	5	8	12	13
1	1	2	3	10	16	17

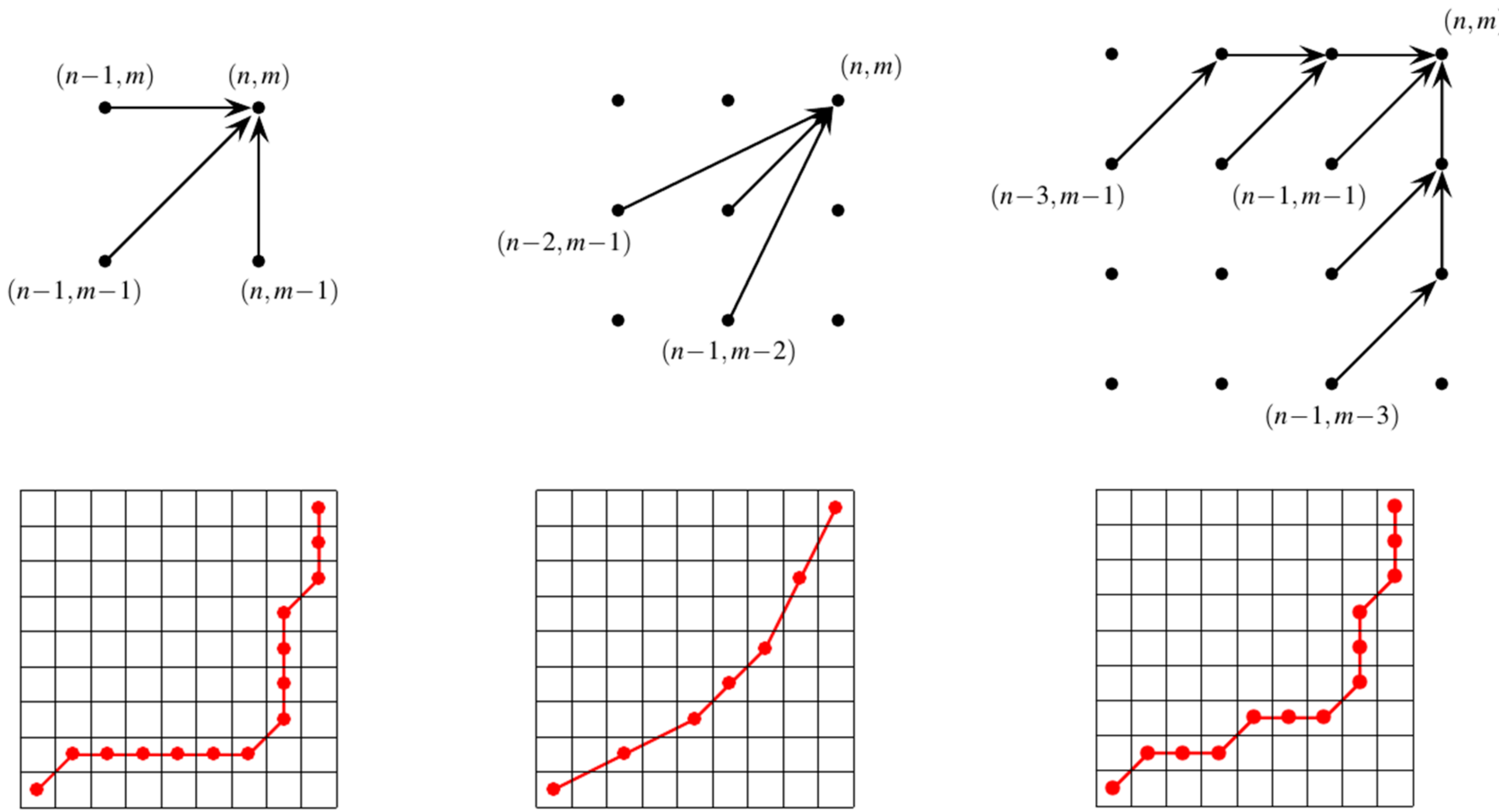
Resulting alingment



3.2 Dynamic Time Warping

Fig. 3.17

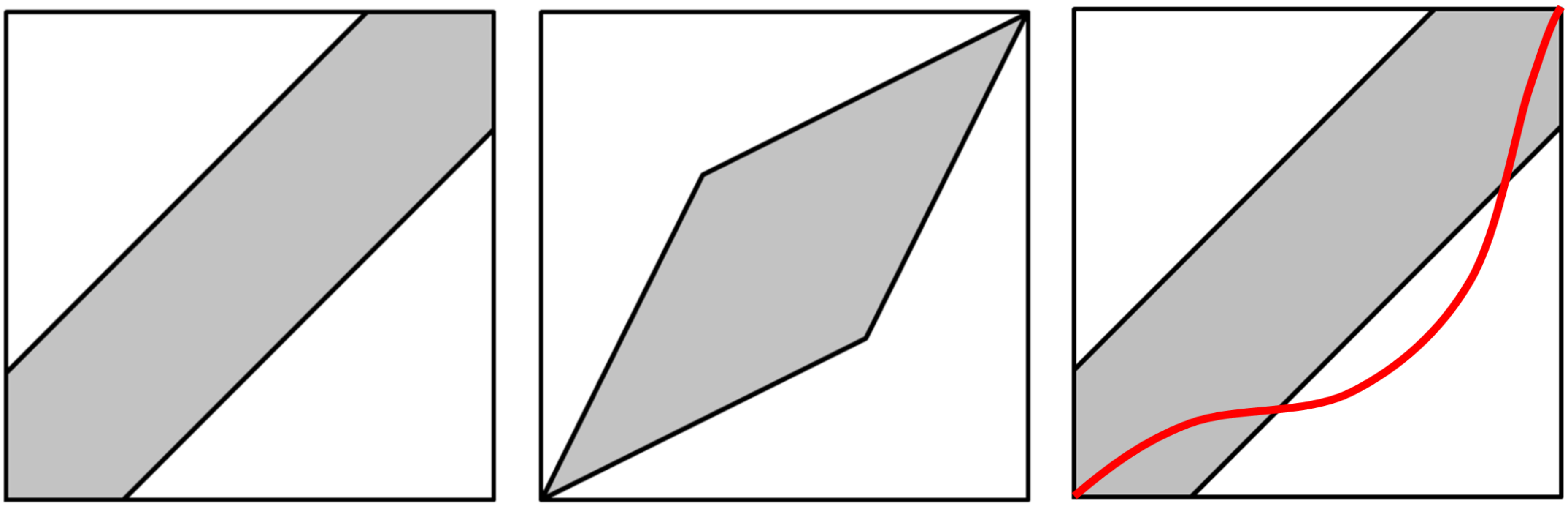
DTW Variants: Step Size Condition



3.2 Dynamic Time Warping

Fig. 3.18

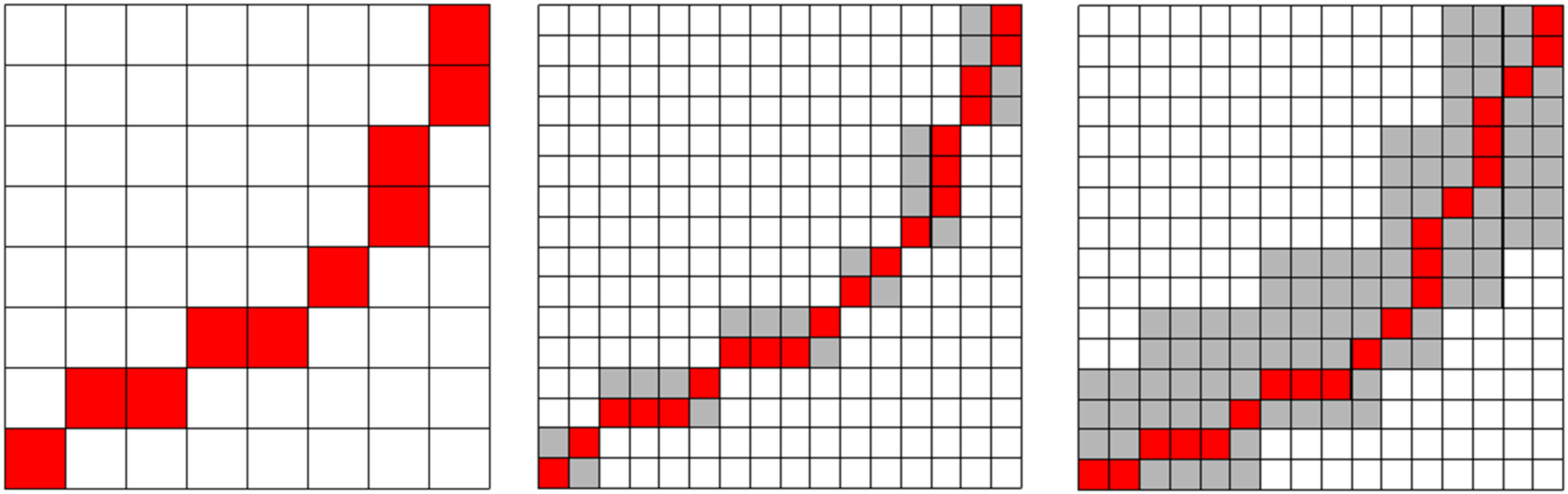
DTW Variants: Global Constraints



3.2 Dynamic Time Warping

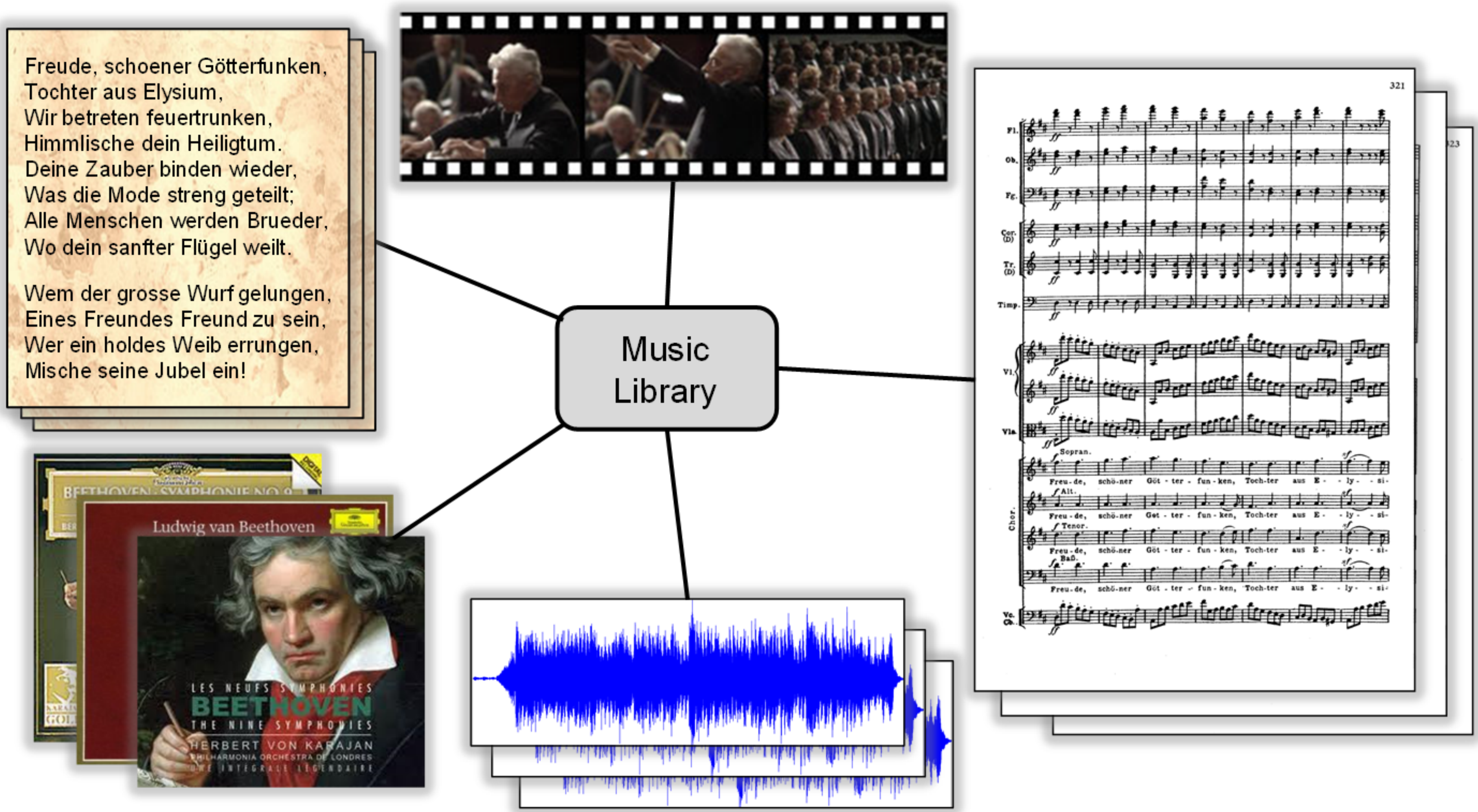
Fig. 3.19

DTW Variants: Multiscale DTW



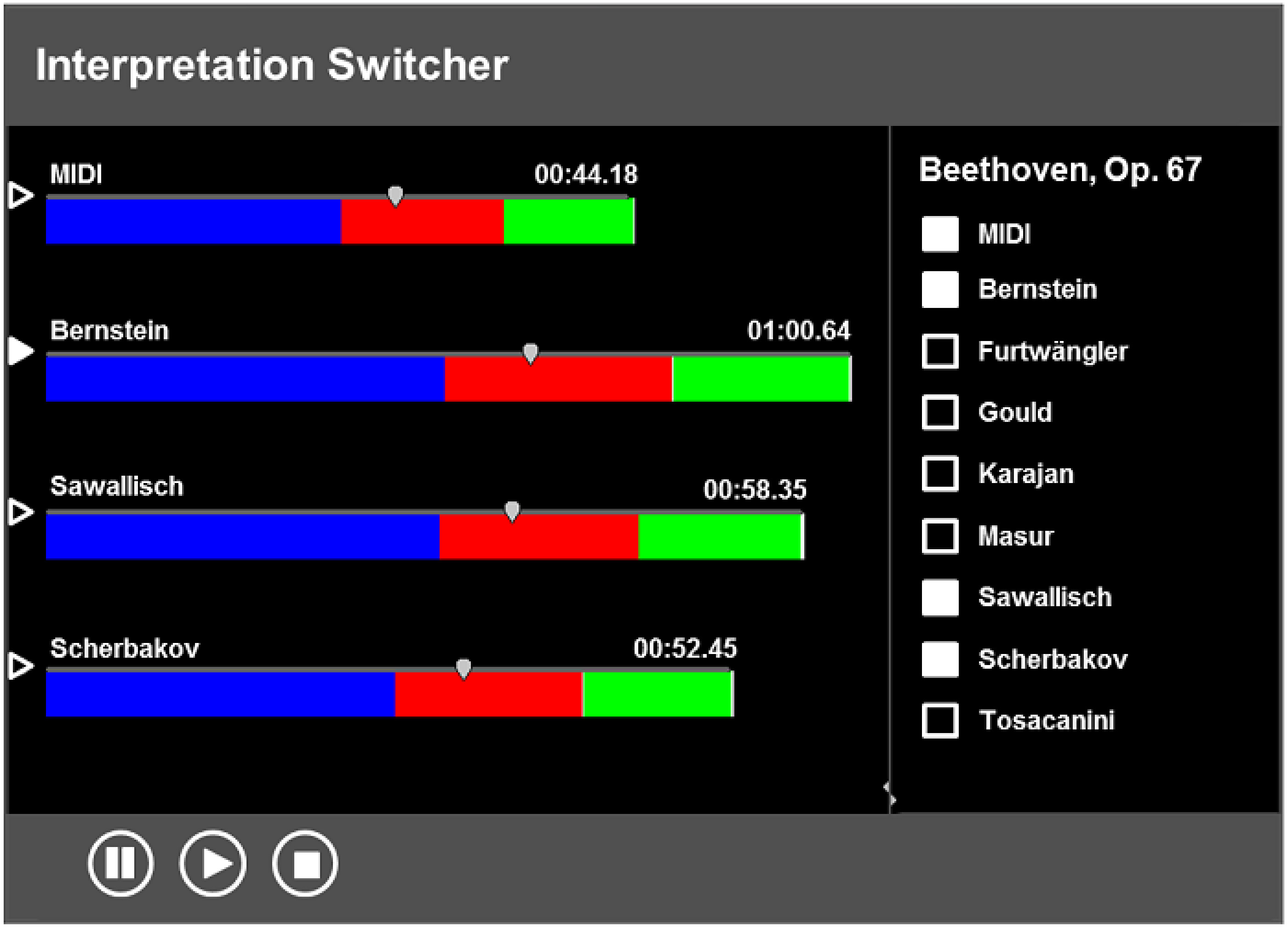
3.3 Applications

Fig. 3.20



3.3 Applications

Fig. 3.21

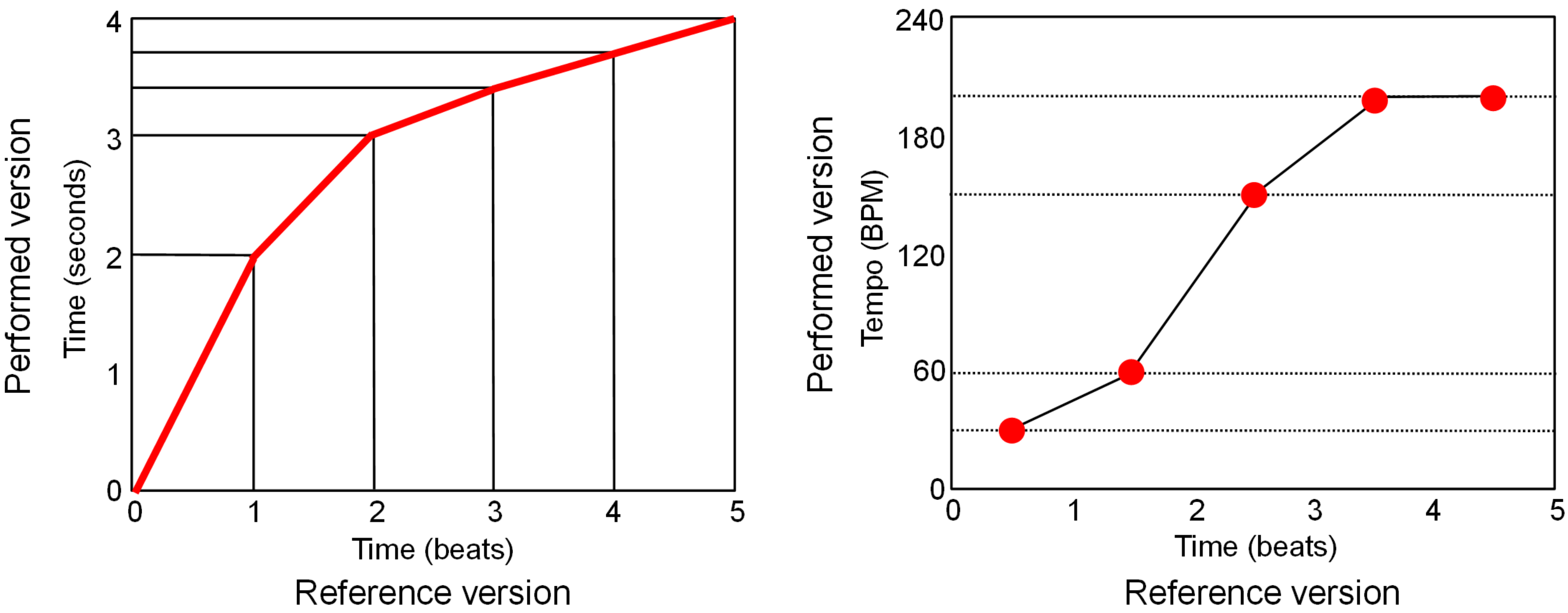


3.3 Applications

Fig. 3.22

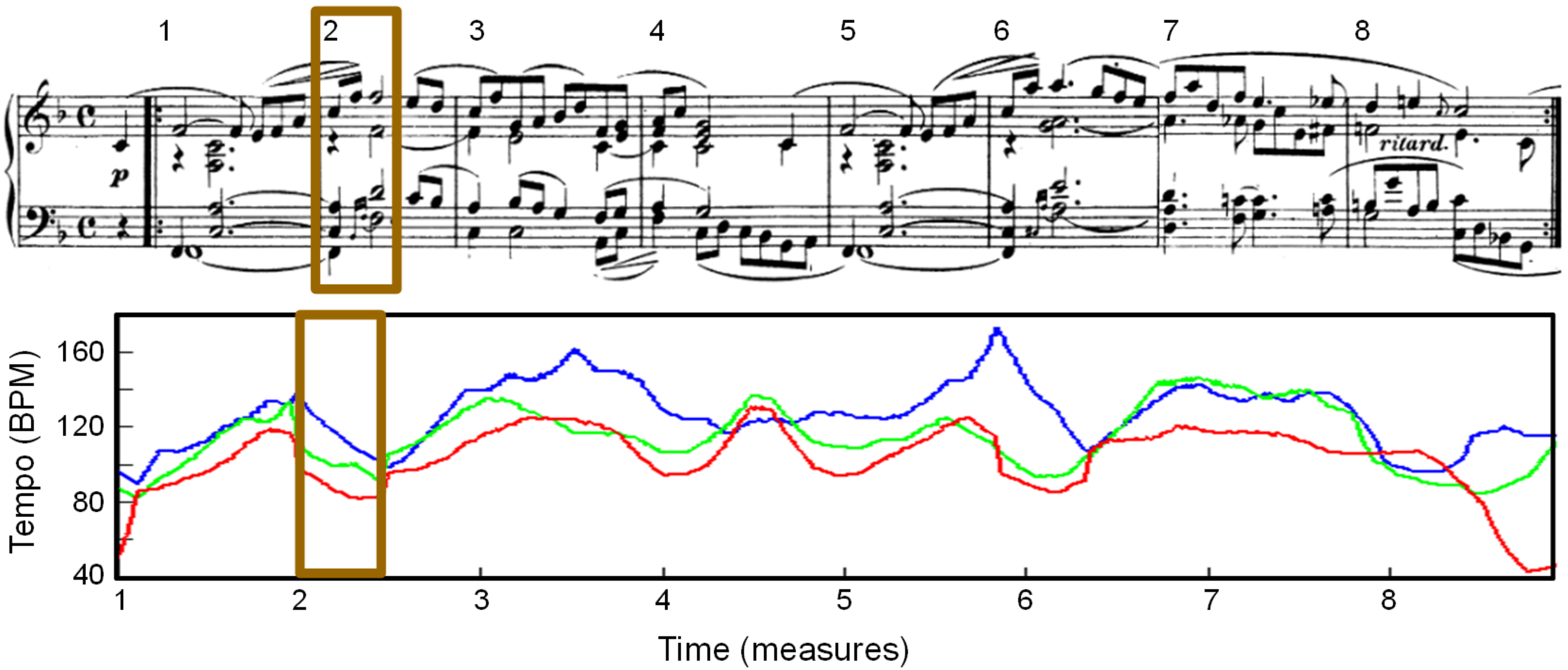
3.3 Applications

Fig. 3.23



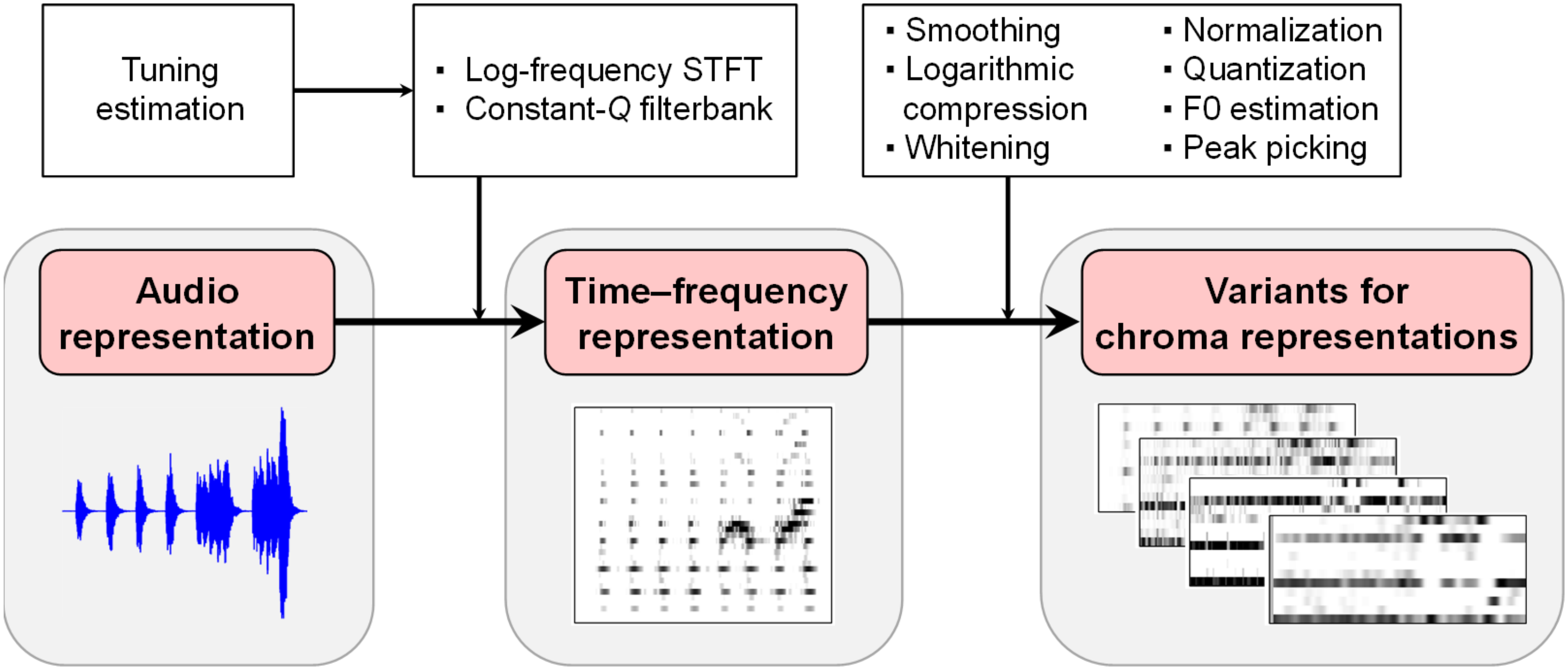
3.3 Applications

Fig. 3.24



3.4 Further Notes

Fig. 3.25



3.4 Further Notes

Fig. 3.26