



Fundamentals of Music Processing

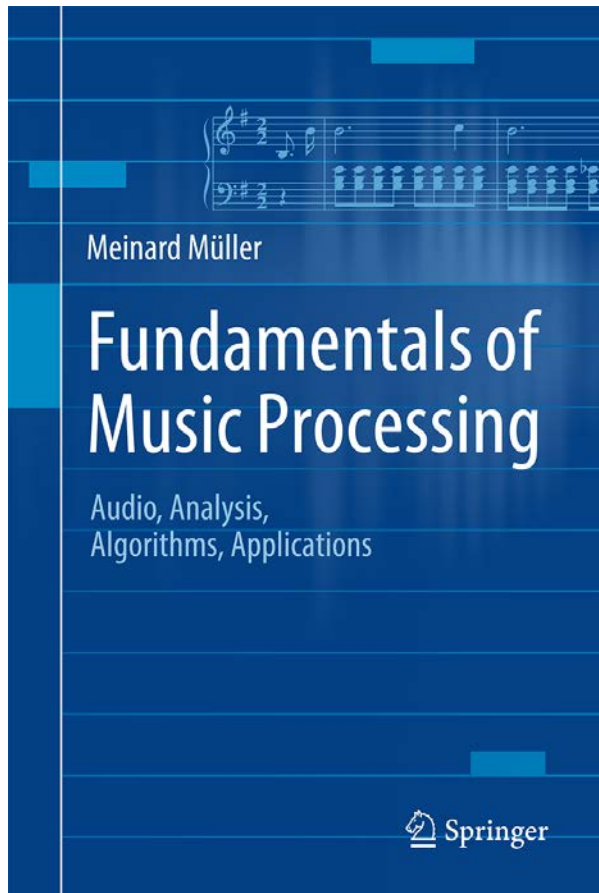
Chapter 7: Content-Based Audio Retrieval

Meinard Müller

International Audio Laboratories Erlangen

www.music-processing.de

Book: Fundamentals of Music Processing

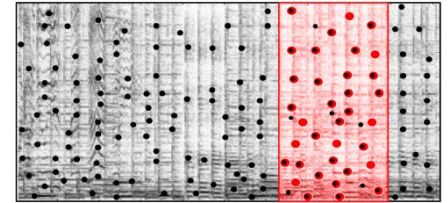


Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
483 p., 249 illus., 30 illus. in color, hardcover
ISBN: 978-3-319-21944-8
Springer, 2015

Accompanying website:
www.music-processing.de

Chapter 7: Content-Based Audio Retrieval

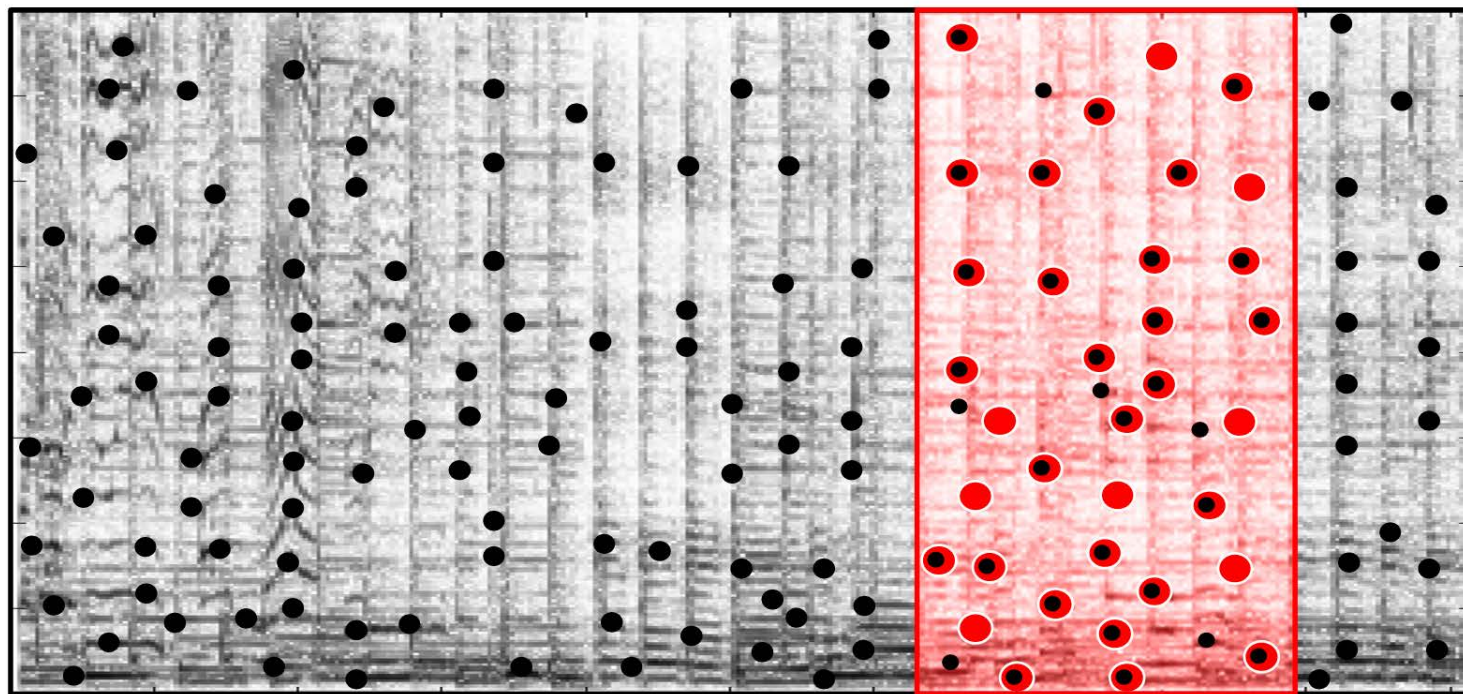
- 7.1 Audio Identification
- 7.2 Audio Matching
- 7.3 Version Identification
- 7.4 Further Notes



One important topic in information retrieval is concerned with the development of search engines that enable users to explore music collections in a flexible and intuitive way. In Chapter 7, we discuss audio retrieval strategies that follow the query-by-example paradigm: given an audio query, the task is to retrieve all documents that are somehow similar or related to the query. Starting with audio identification, a technique used in many commercial applications such as Shazam, we study various retrieval strategies to handle different degrees of similarity. Furthermore, considering efficiency issues, we discuss fundamental indexing techniques based on inverted lists—a concept originally used in text retrieval.

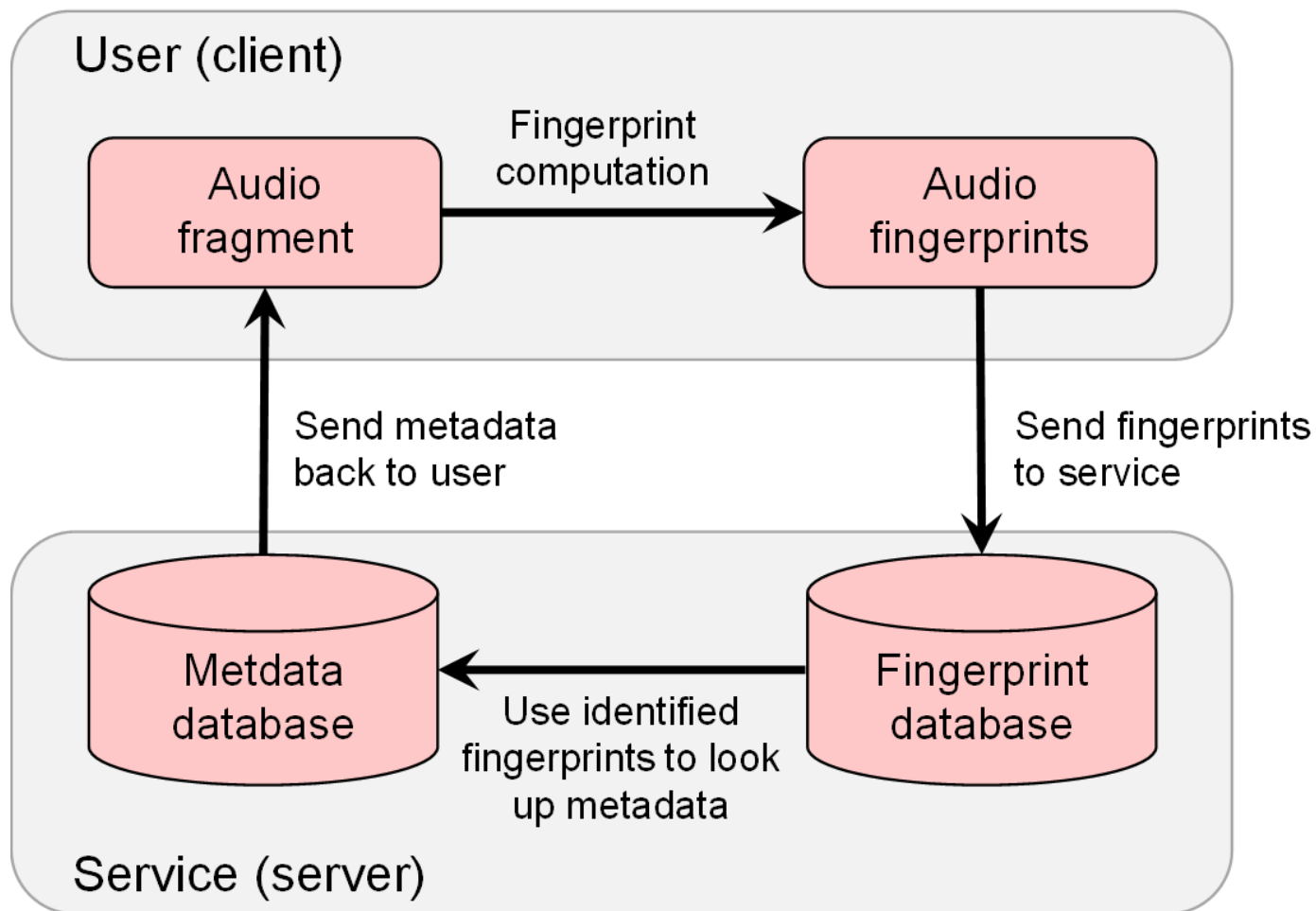
7 Content-Based Audio Retrieval

Teaser



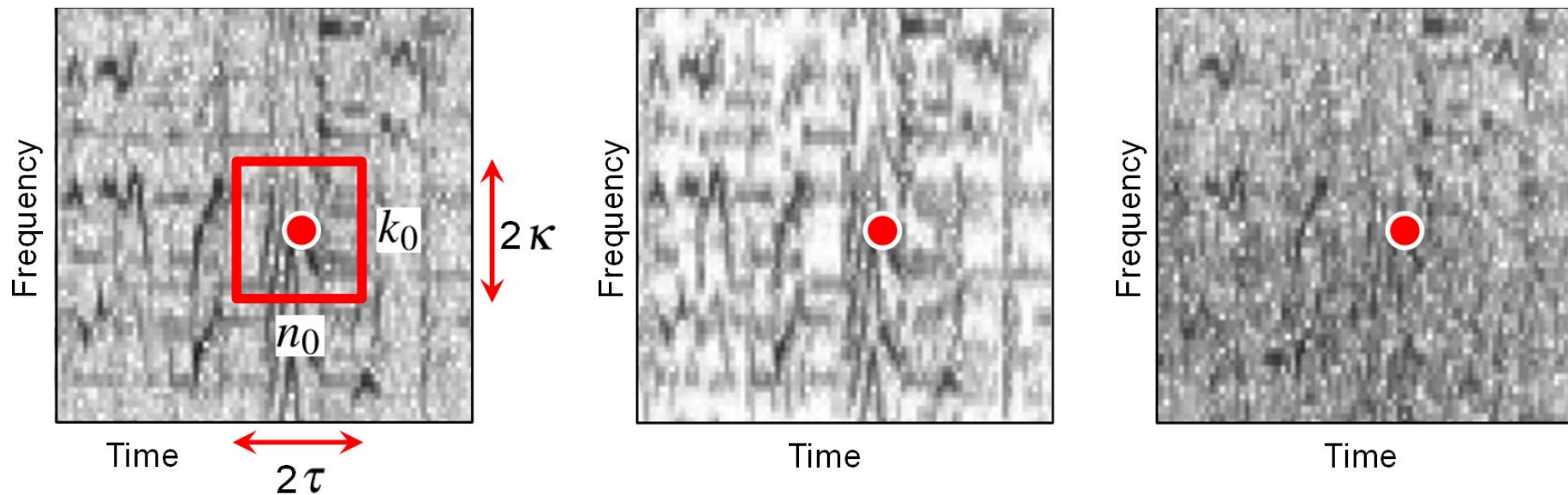
7.1 Audio Identification

Fig. 7.2



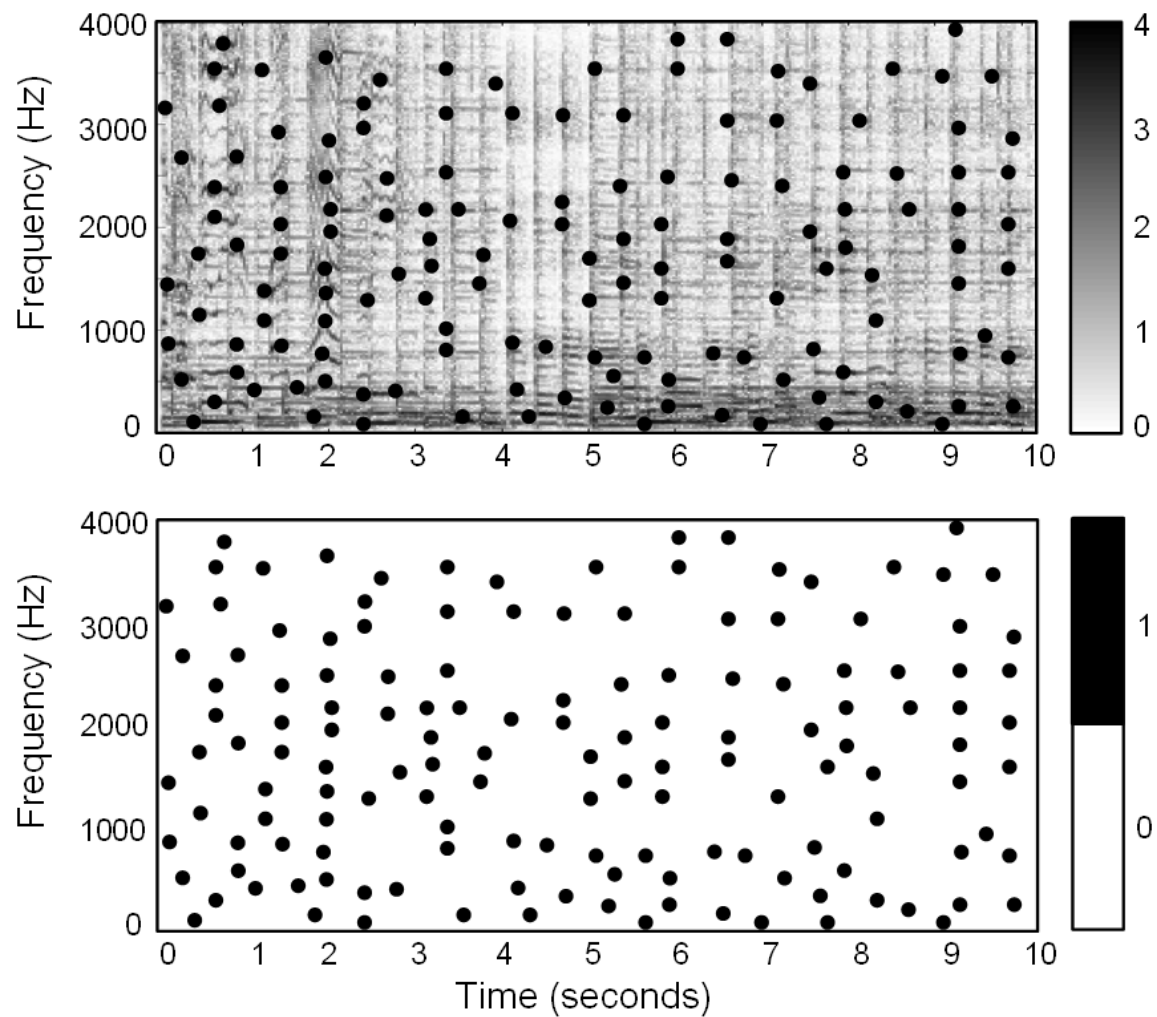
7.1 Audio Identification

Fig. 7.3



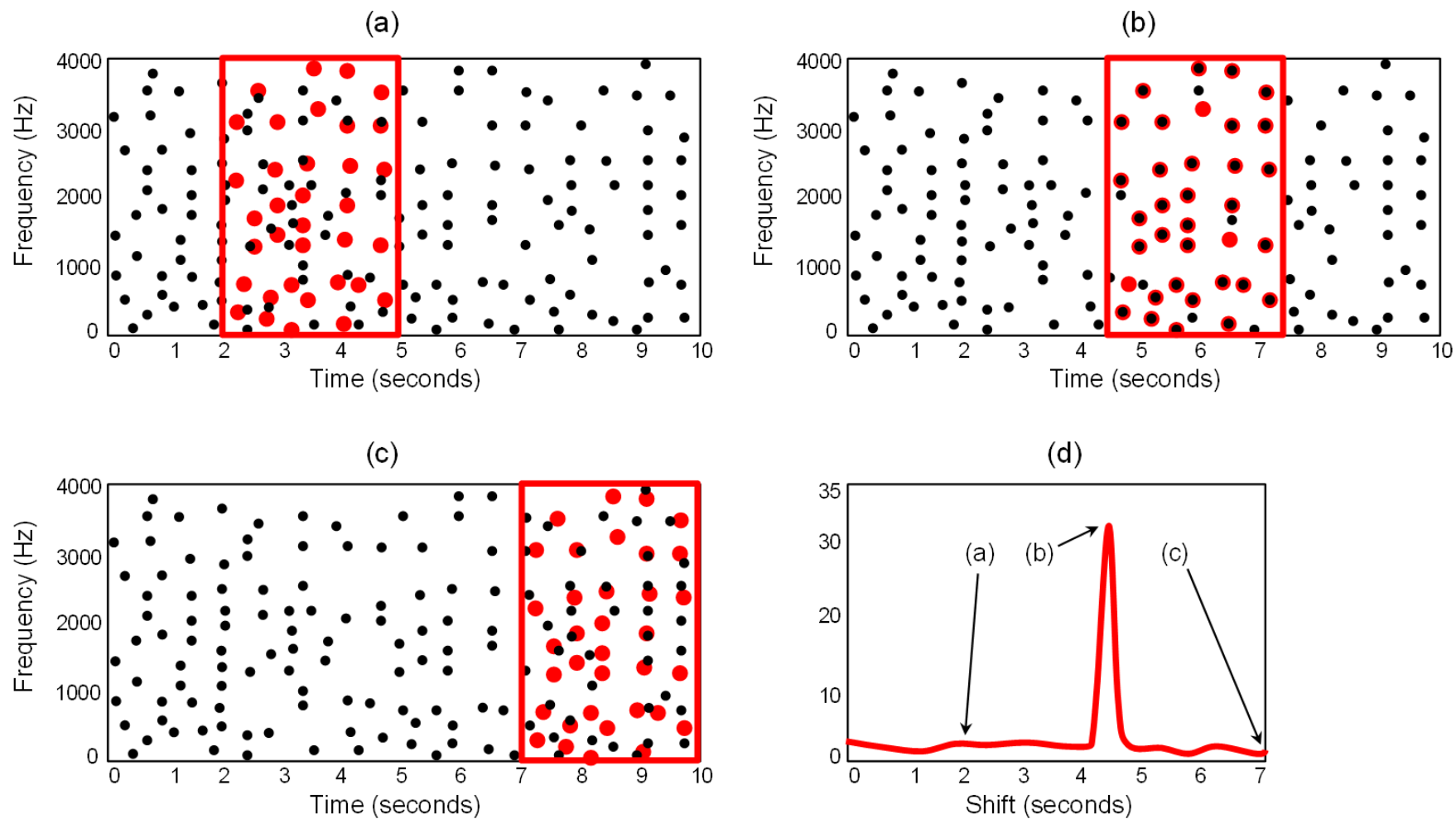
7.1 Audio Identification

Fig. 7.4



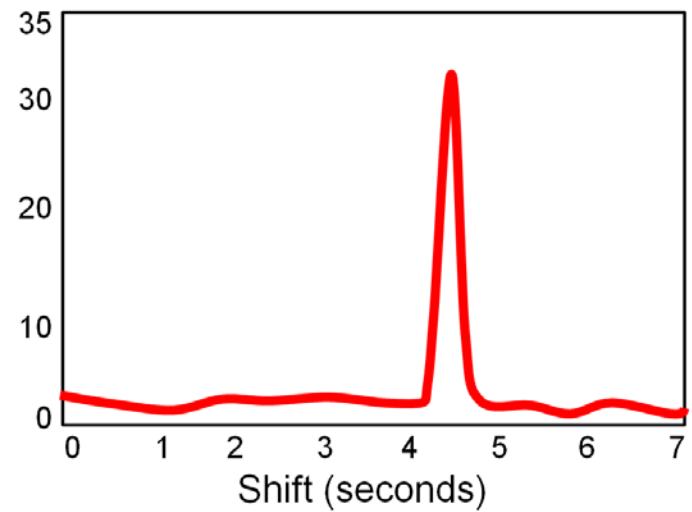
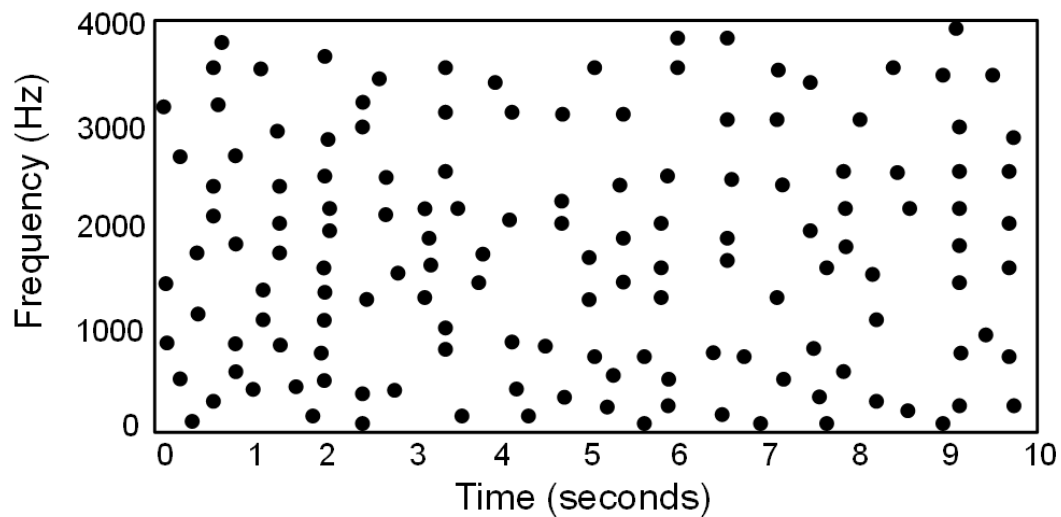
7.1 Audio Identification

Fig. 7.5



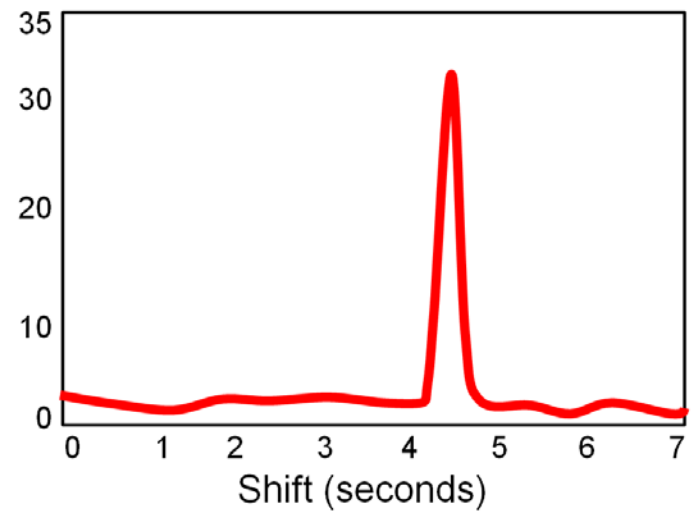
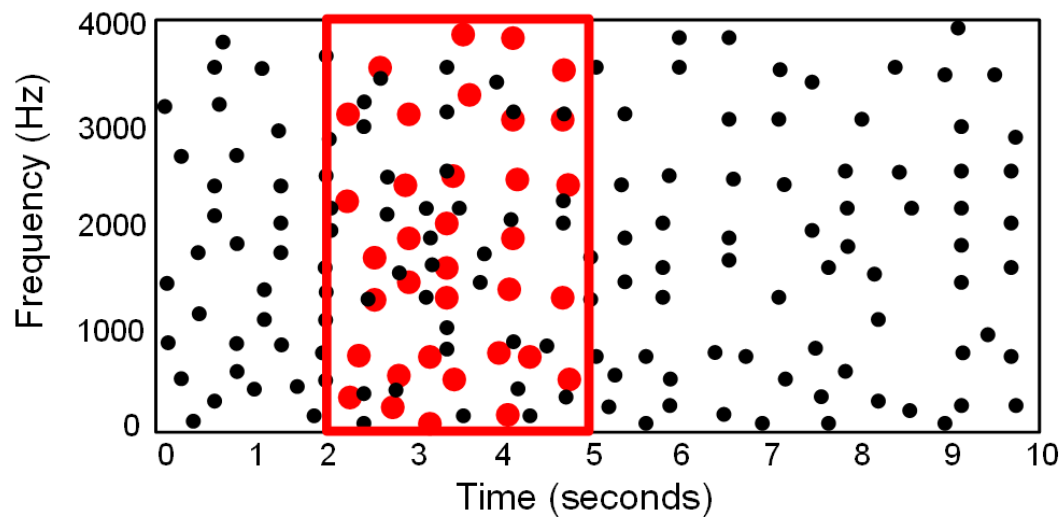
7.1 Audio Identification

Fig. 7.5



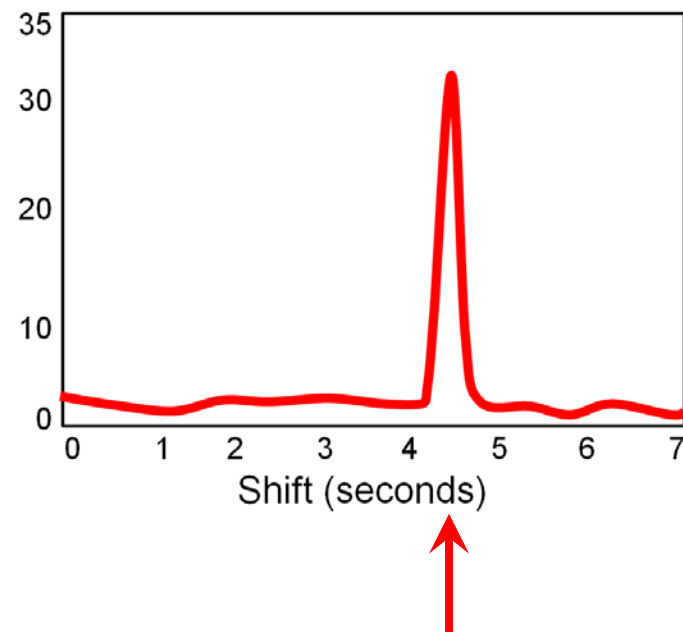
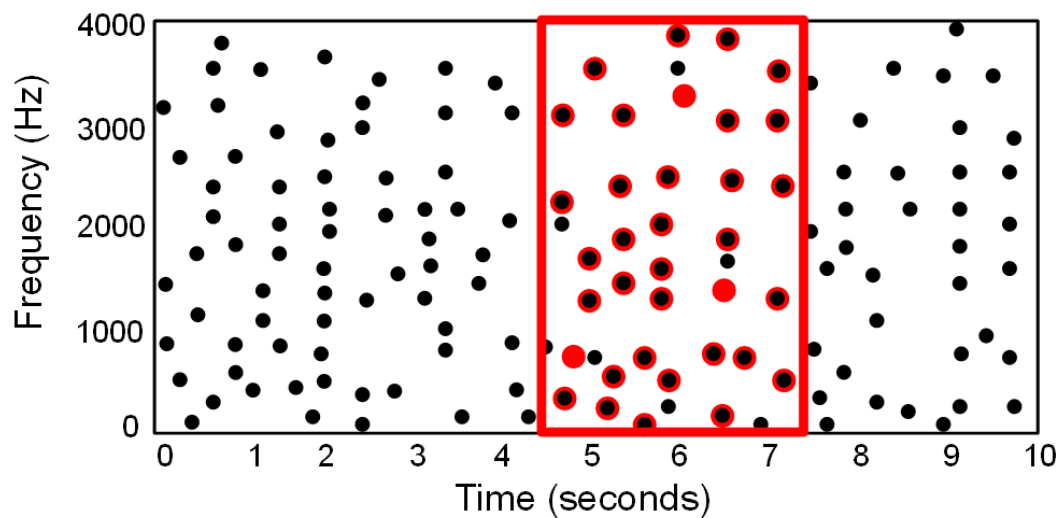
7.1 Audio Identification

Fig. 7.5



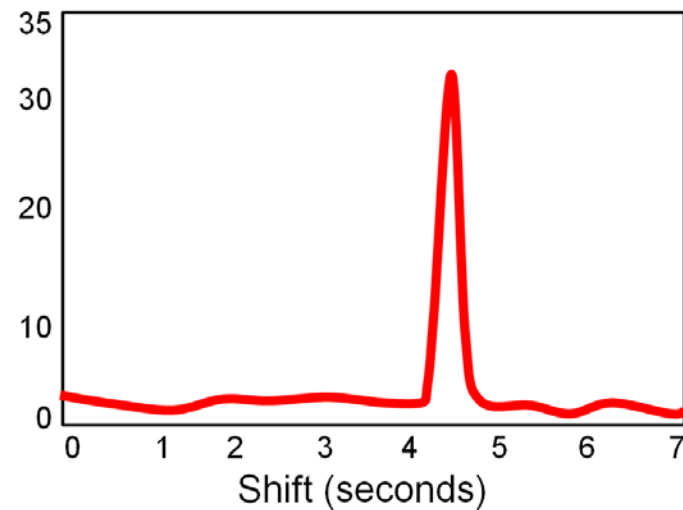
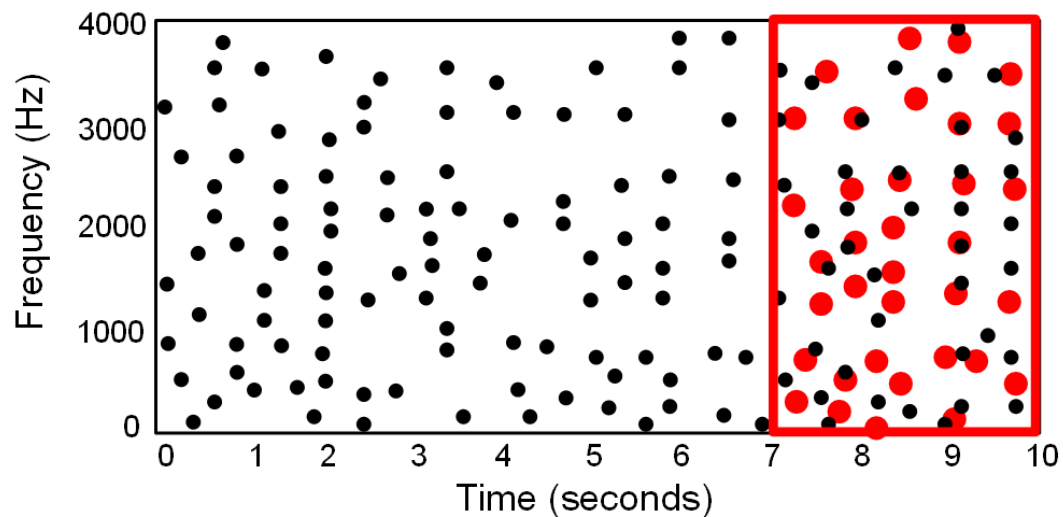
7.1 Audio Identification

Fig. 7.5



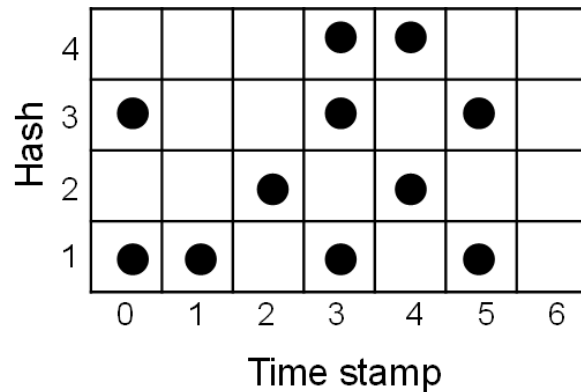
7.1 Audio Identification

Fig. 7.5



7.1 Audio Identification

Fig. 7.6

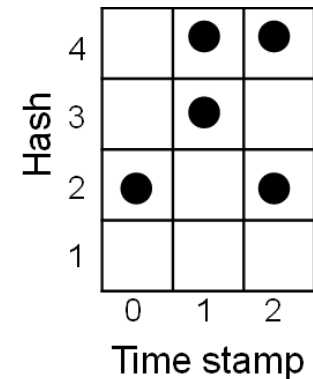


$$L(4) = (3,4)$$

$$L(3) = (0,3,5)$$

$$L(2) = (2,4)$$

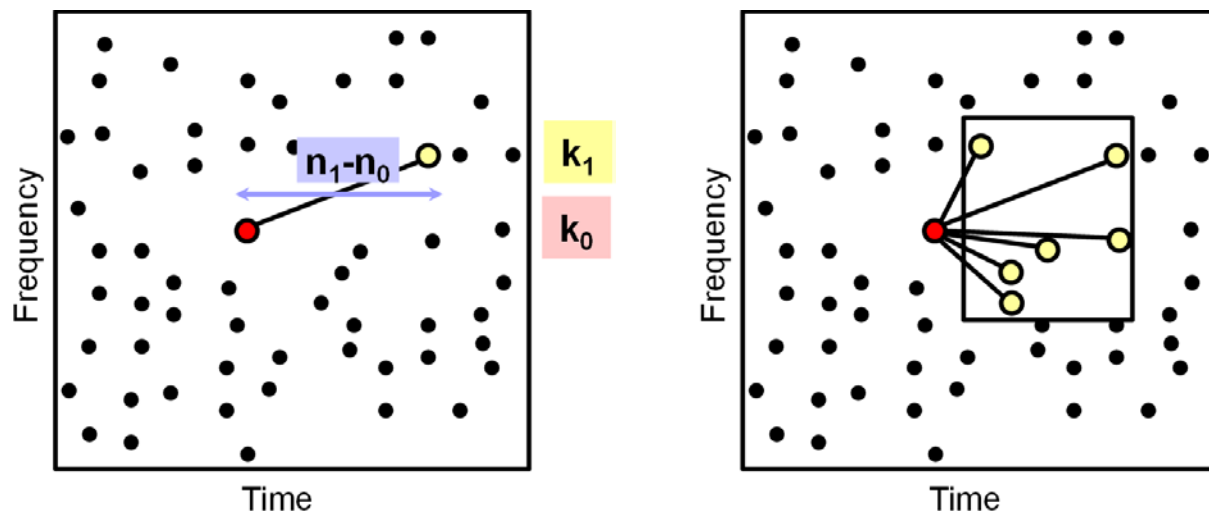
$$L(1) = (0,1,3,5)$$



Query (<i>n,h</i>)	<i>L(h) - n</i>	Indicator functions									
		...	-1	0	1	2	3	4	5	6	...
(0,2)	(2,4)	0	0	0	0	1	0	1	0	0	0
(1,3)	(-1,2,4)	0	1	0	0	1	0	1	0	0	0
(1,4)	(2,3)	0	0	0	0	1	1	0	0	0	0
(2,2)	(0,2)	0	0	1	0	1	0	0	0	0	0
(2,4)	(1,2)	0	0	0	1	1	0	0	0	0	0
Matching function		0	1	1	1	5	1	2	0	0	0

7.1 Audio Identification

Fig. 7.7



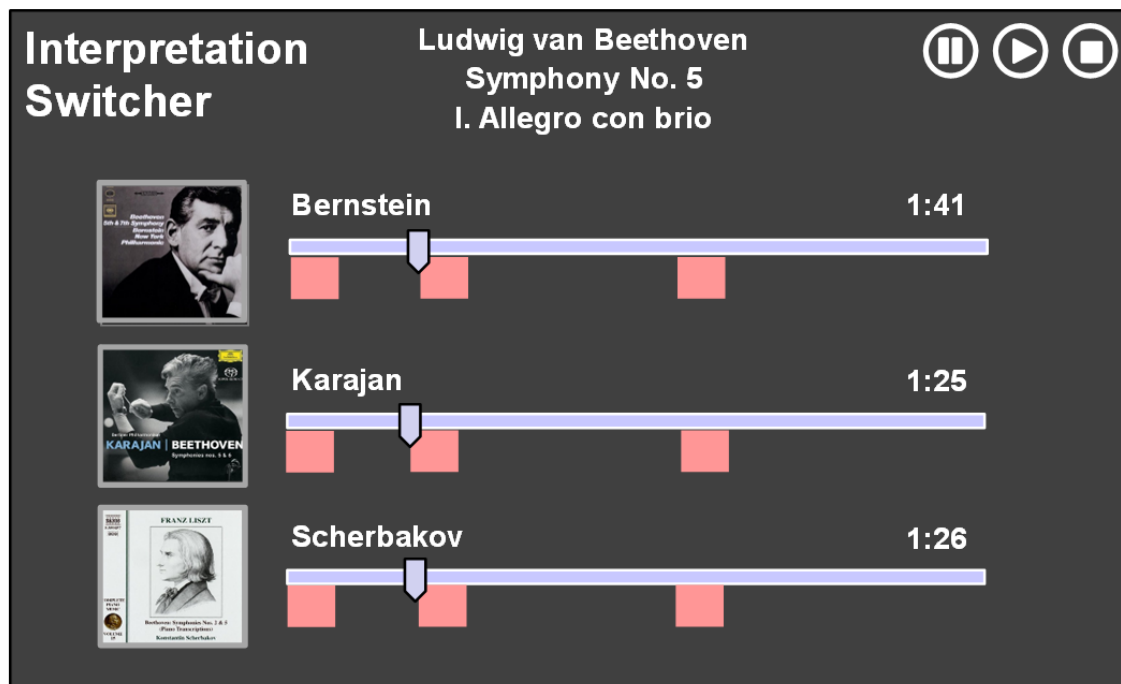
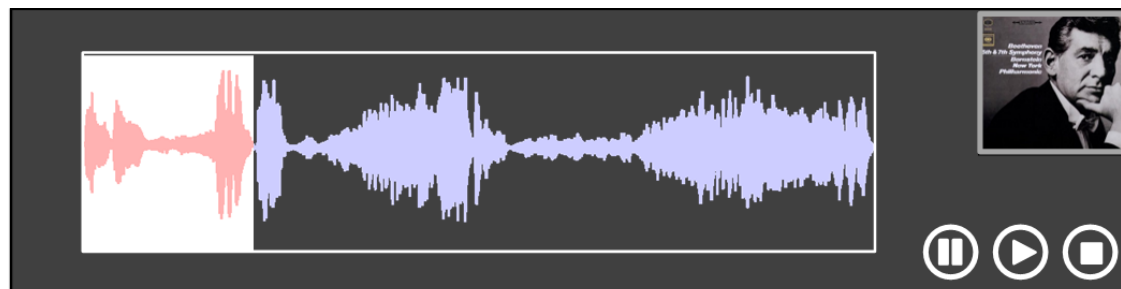
Hash:

Consists of two frequency stamps and a time stamp difference:

$$\left(k_0, k_1, n_1 - n_0 \right)$$

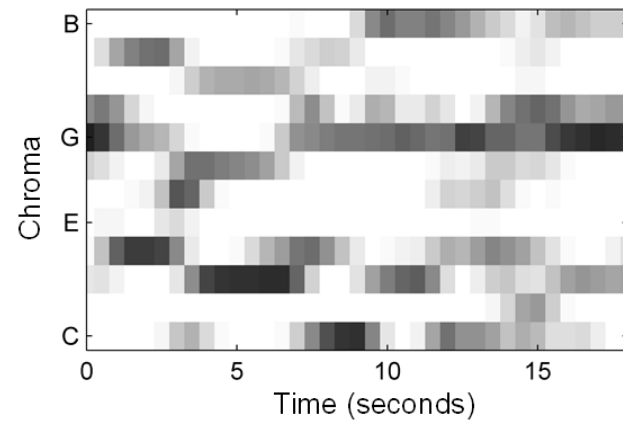
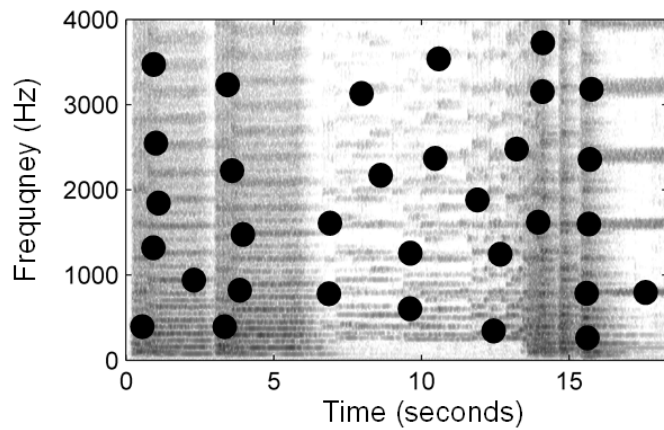
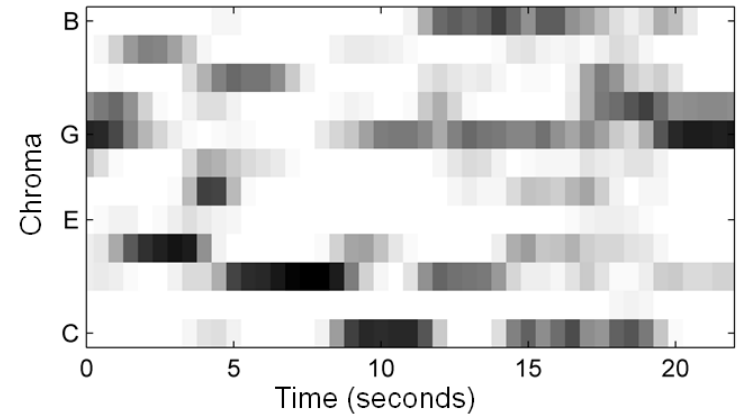
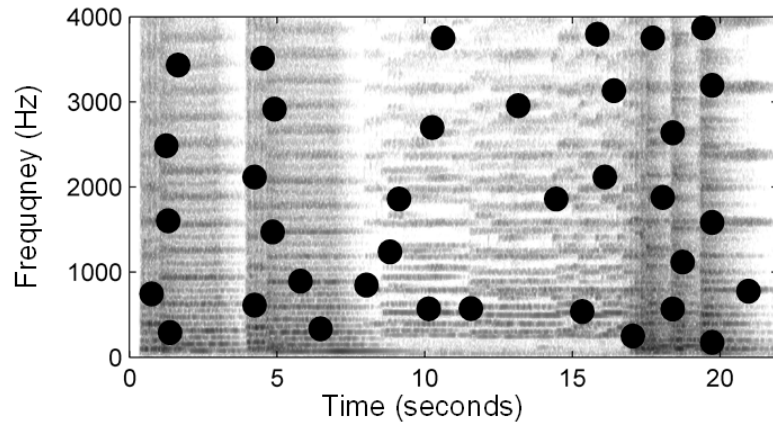
7.2 Audio Matching

Fig. 7.8



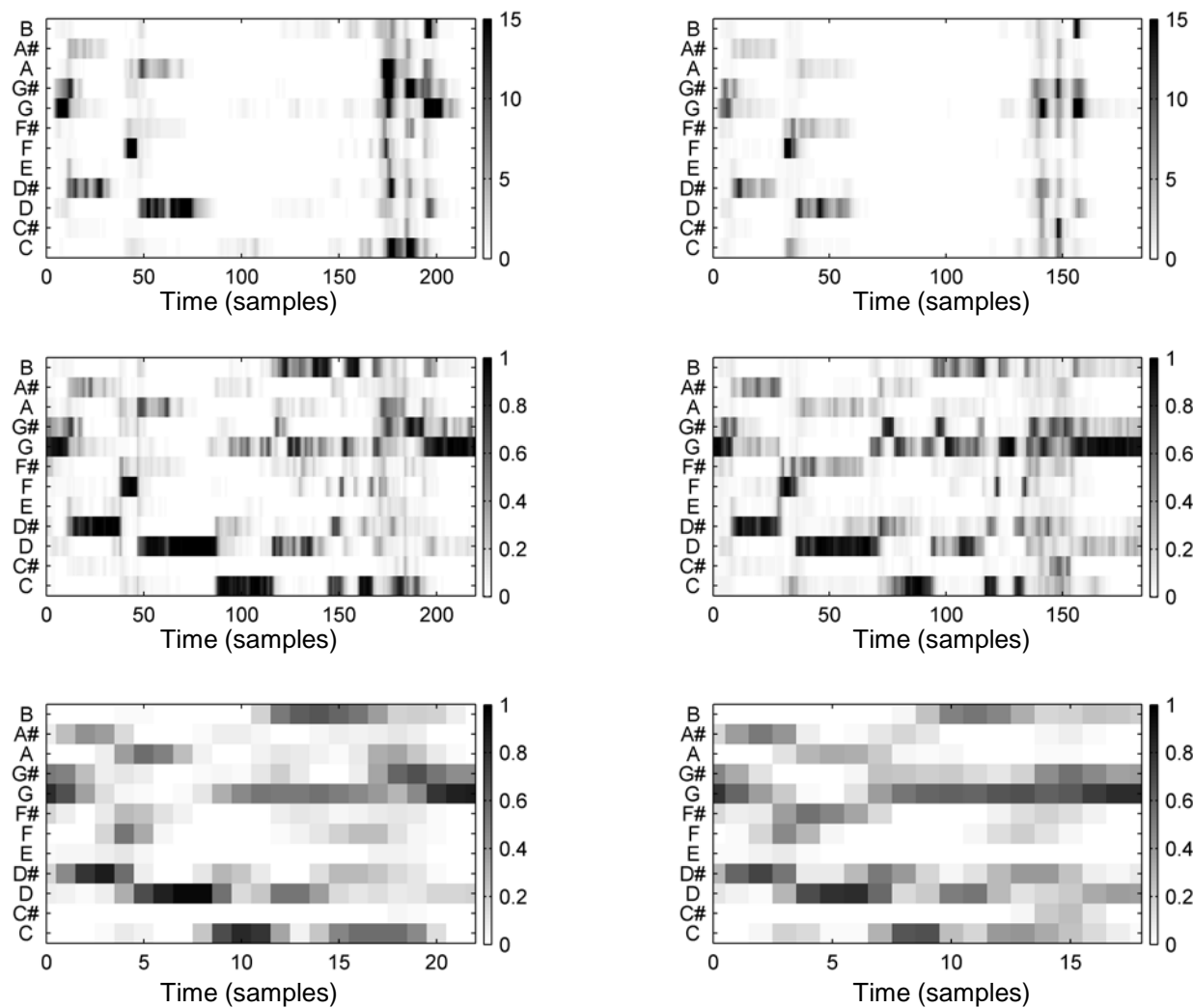
7.2 Audio Matching

Fig. 7.9



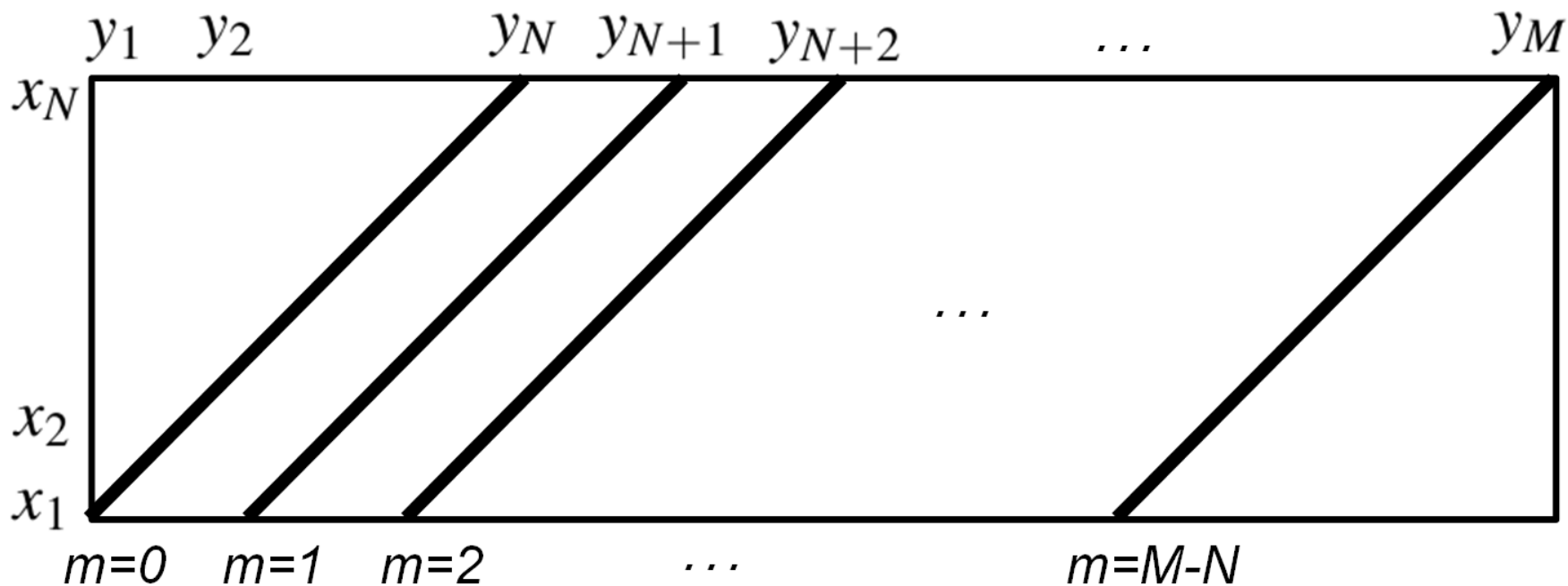
7.2 Audio Matching

Fig. 7.10



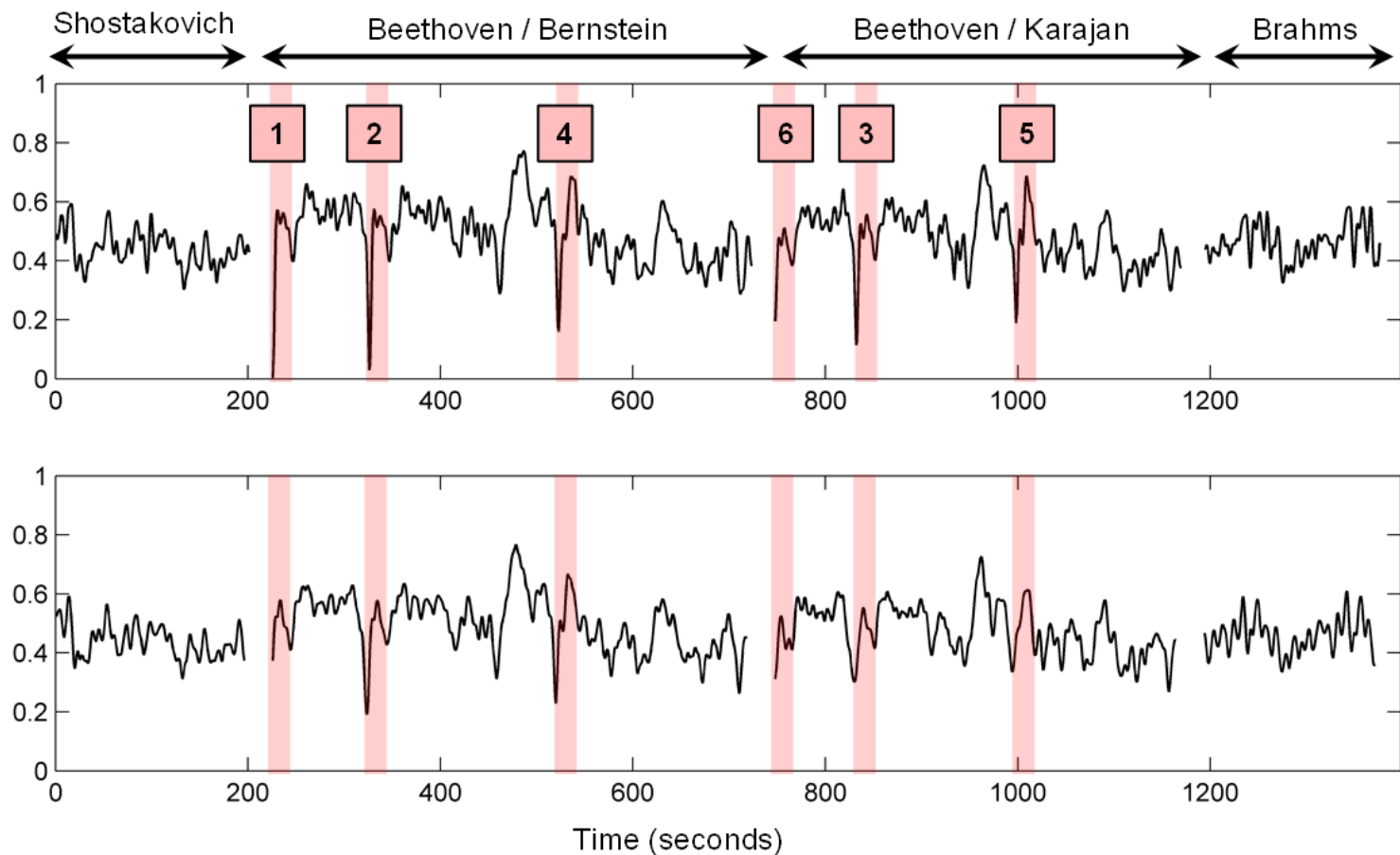
7.2 Audio Matching

Fig. 7.11



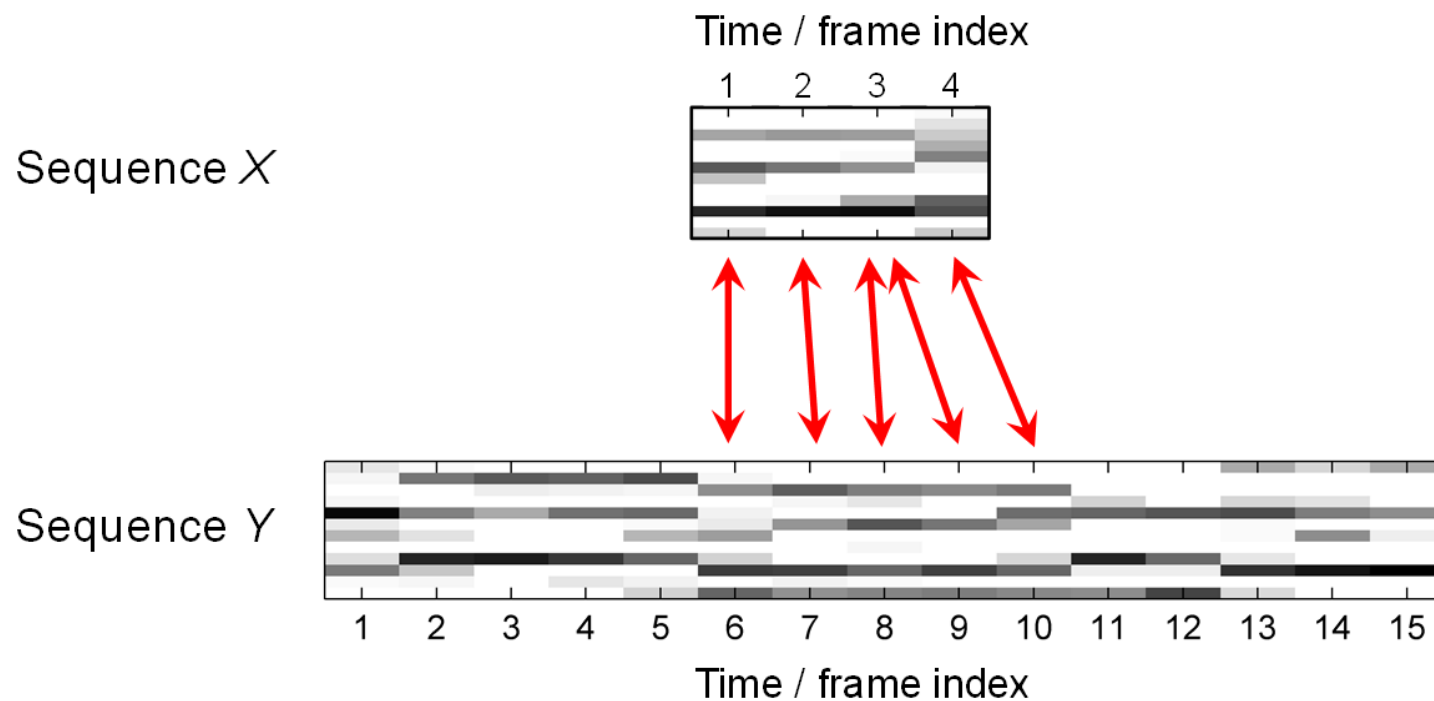
7.2 Audio Matching

Fig. 7.12



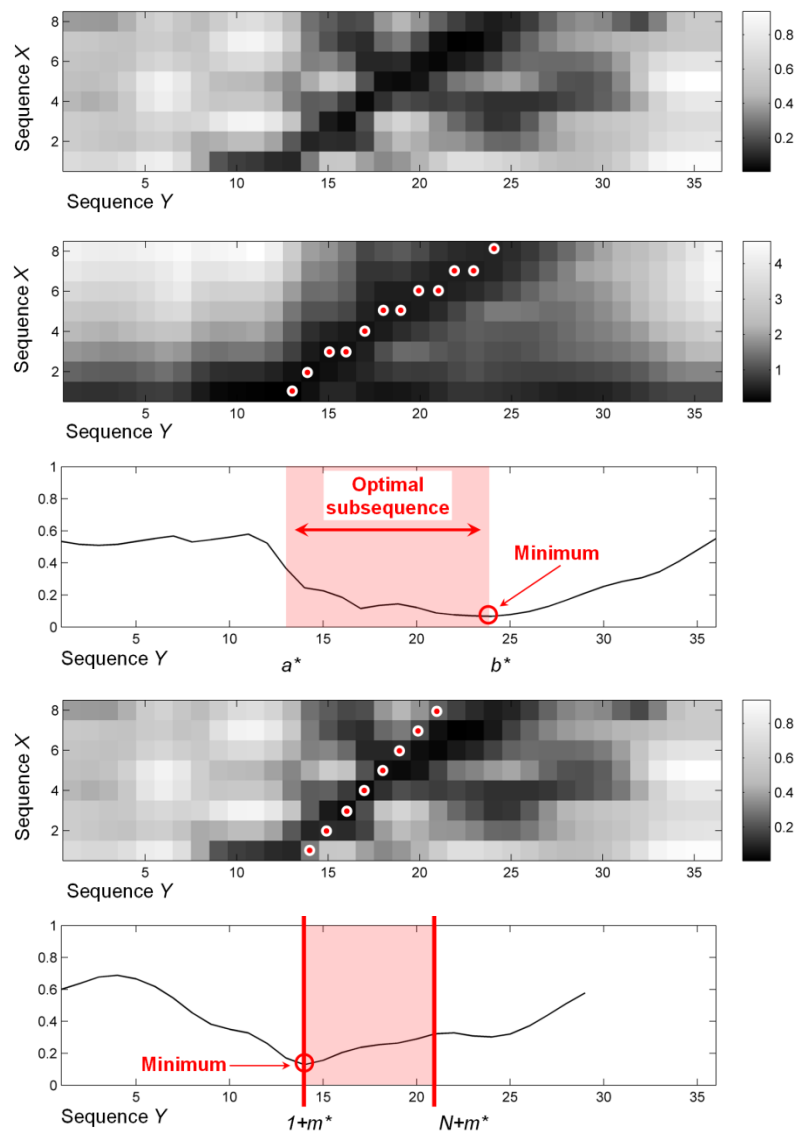
7.2 Audio Matching

Fig. 7.13



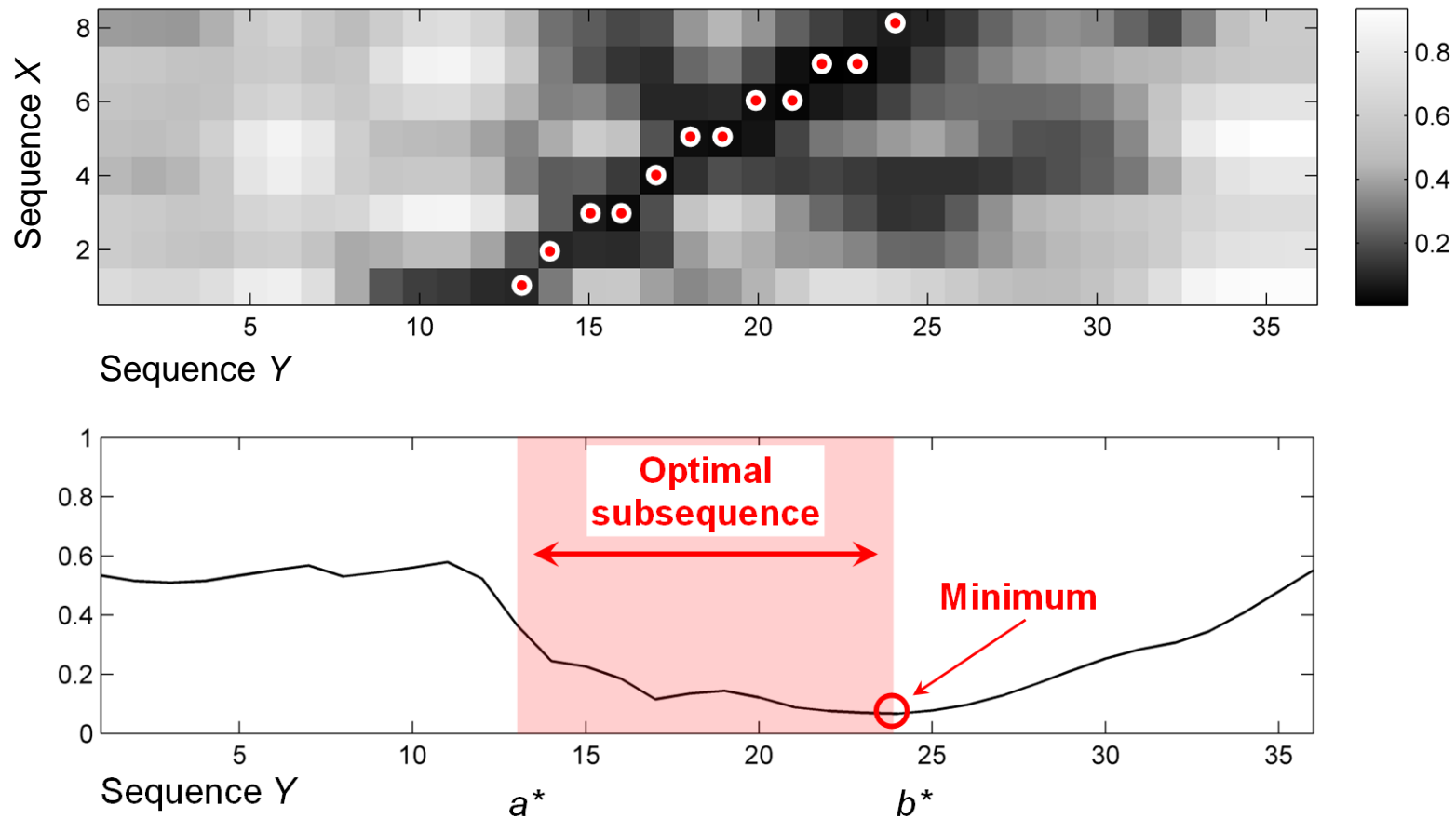
7.2 Audio Matching

Fig. 7.14



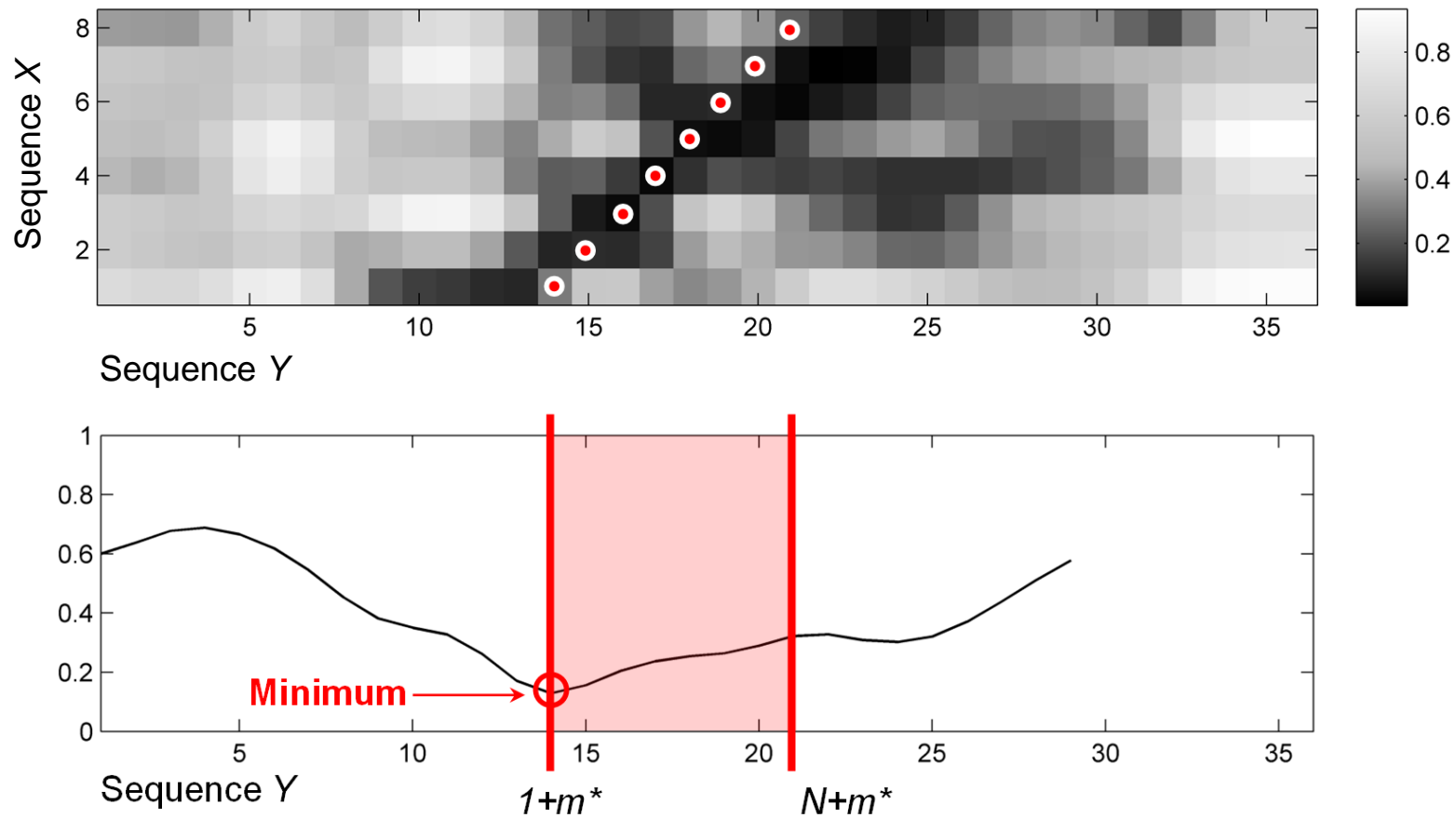
7.2 Audio Matching

Fig. 7.14



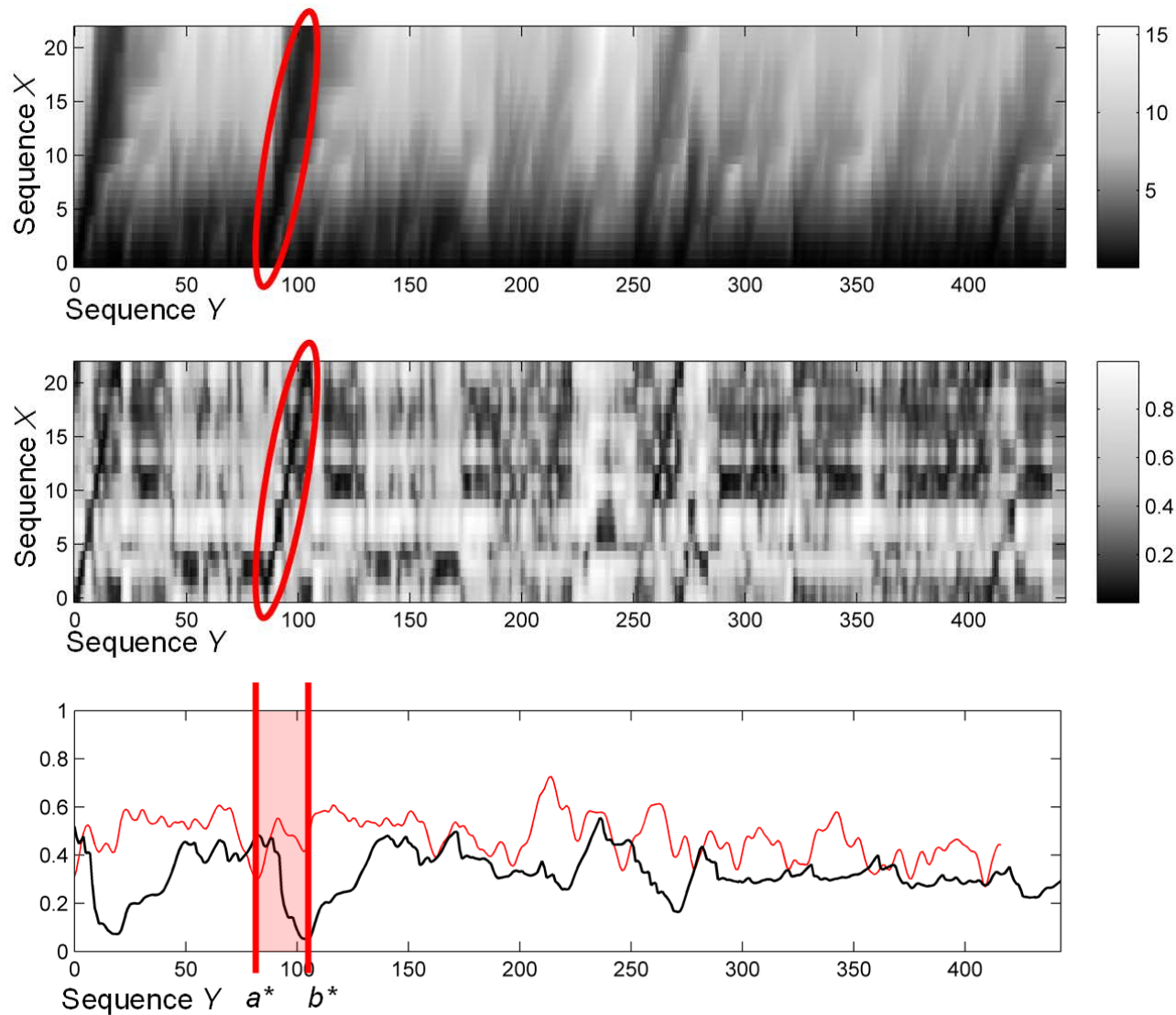
7.2 Audio Matching

Fig. 7.14



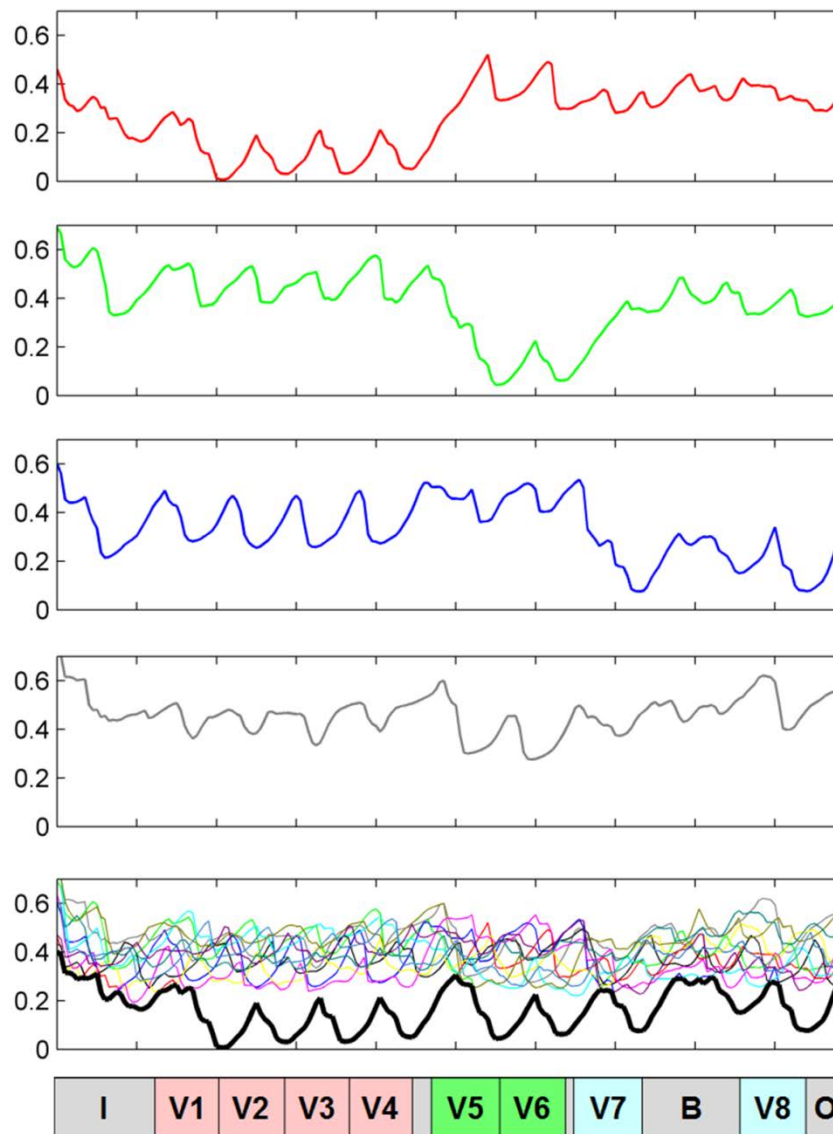
7.2 Audio Matching

Fig. 7.15



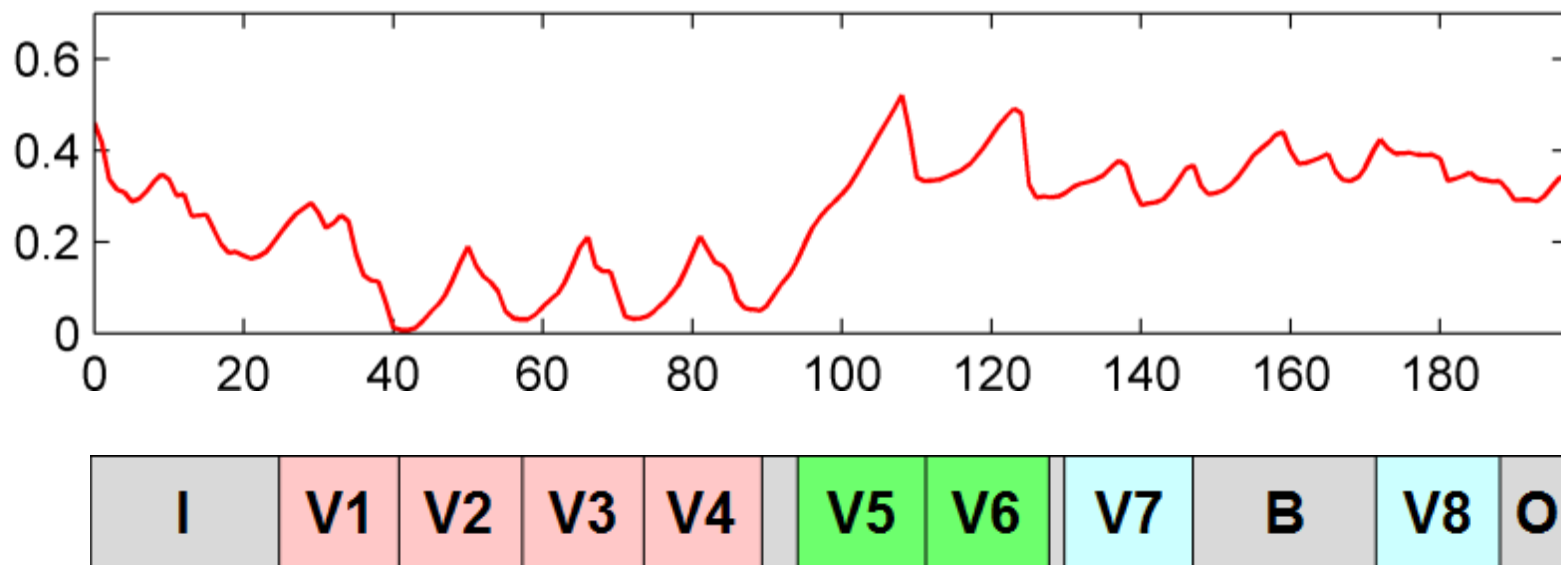
7.2 Audio Matching

Fig. 7.16



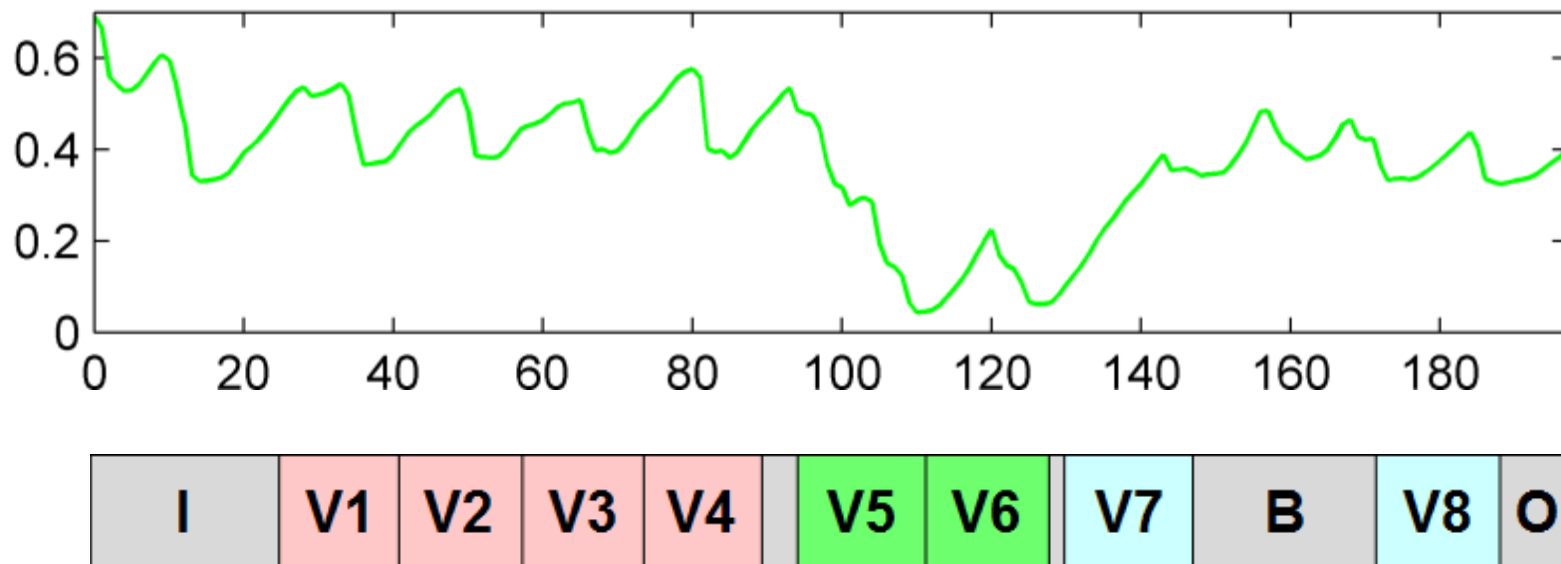
7.2 Audio Matching

Fig. 7.16



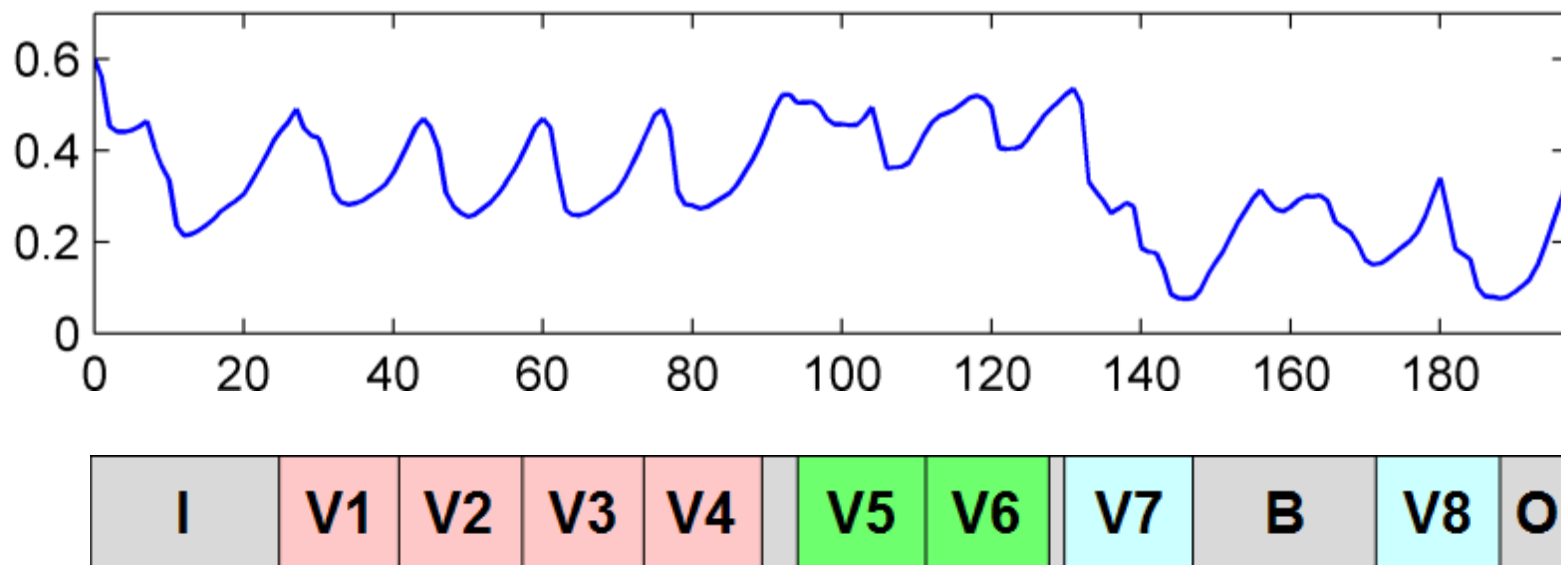
7.2 Audio Matching

Fig. 7.16



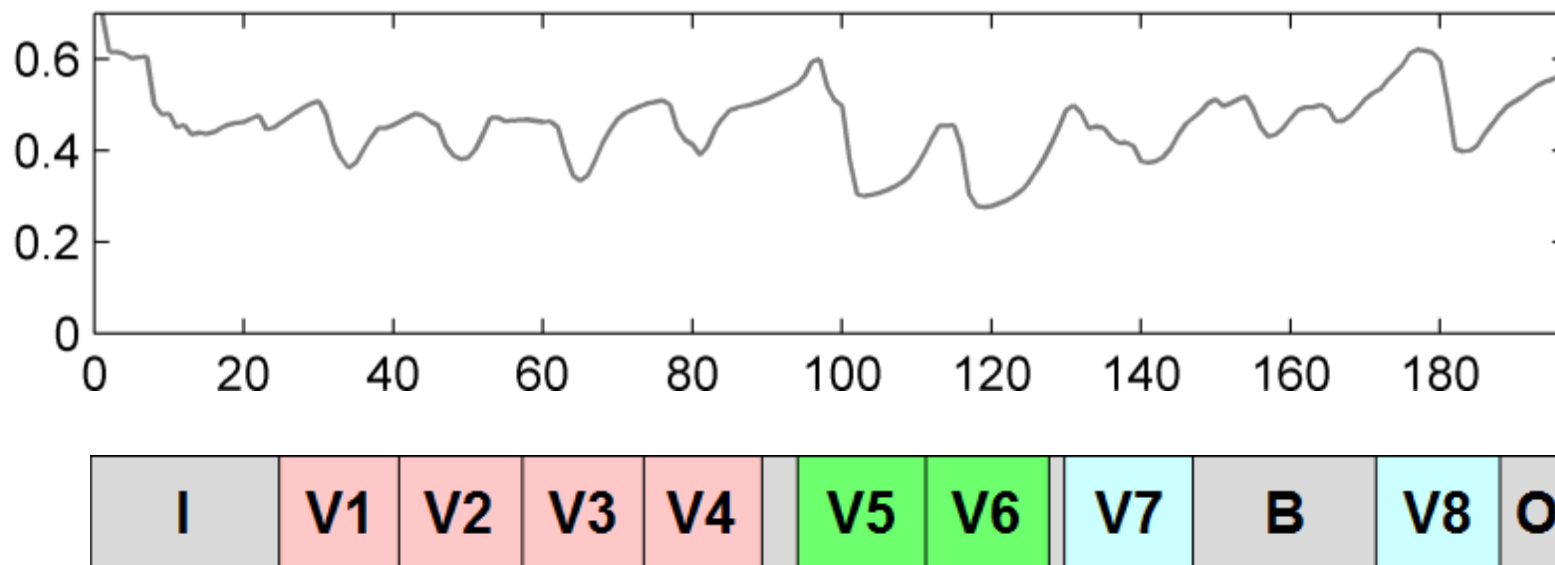
7.2 Audio Matching

Fig. 7.16



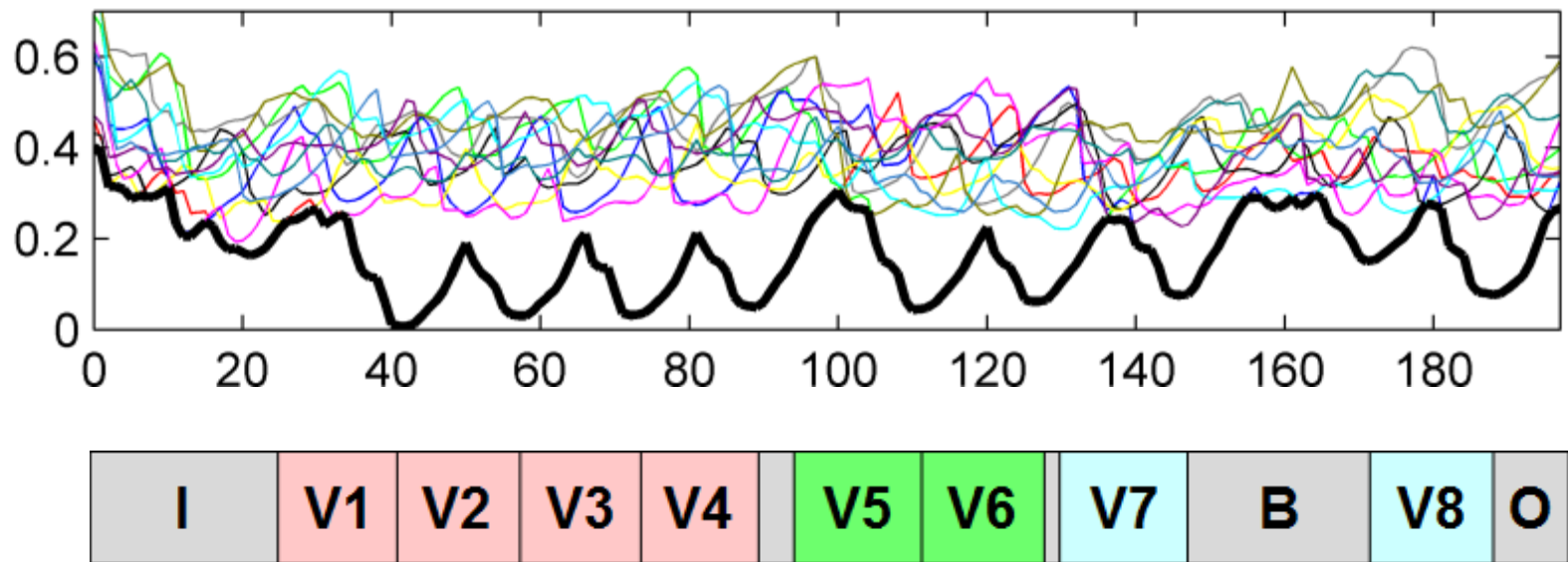
7.2 Audio Matching

Fig. 7.16



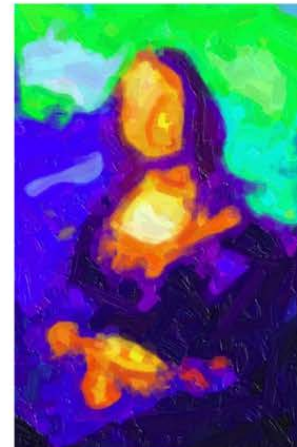
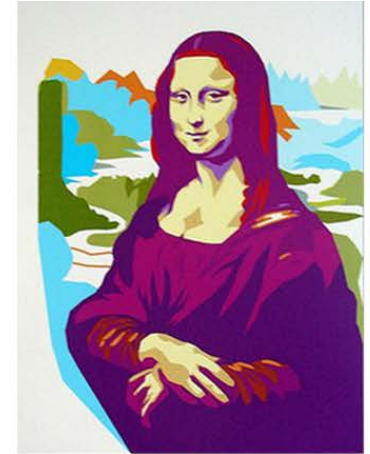
7.2 Audio Matching

Fig. 7.16



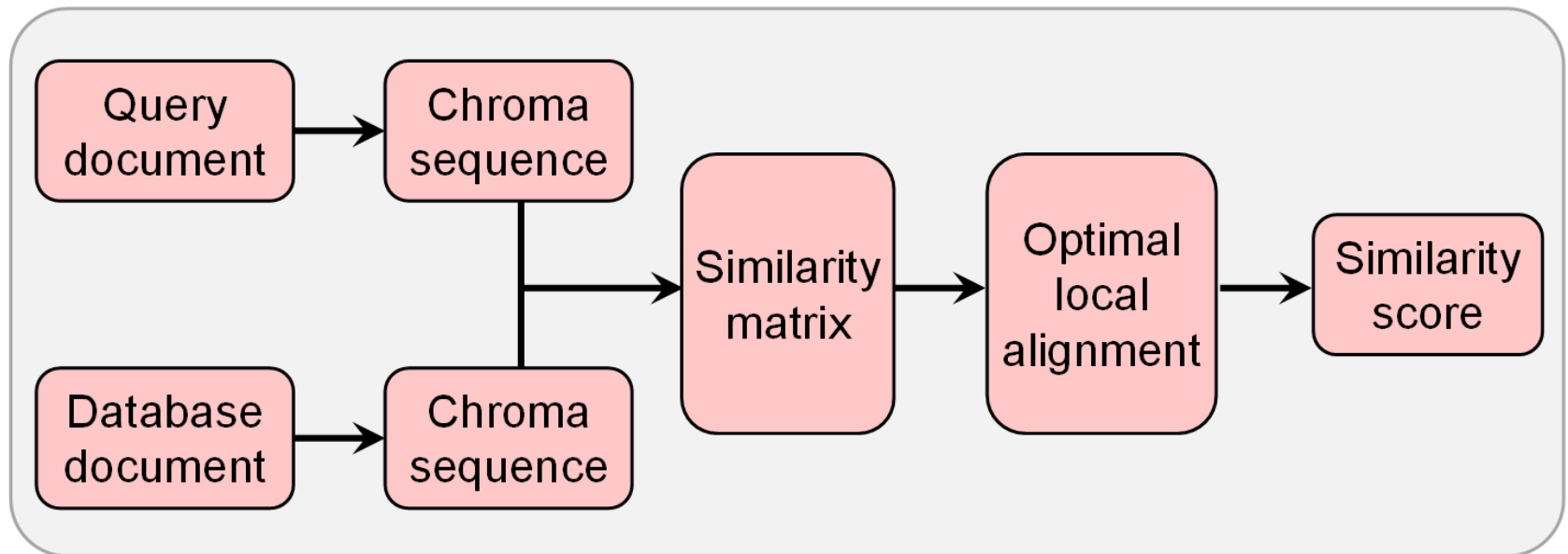
7.3 Version Identification

Fig. 7.17



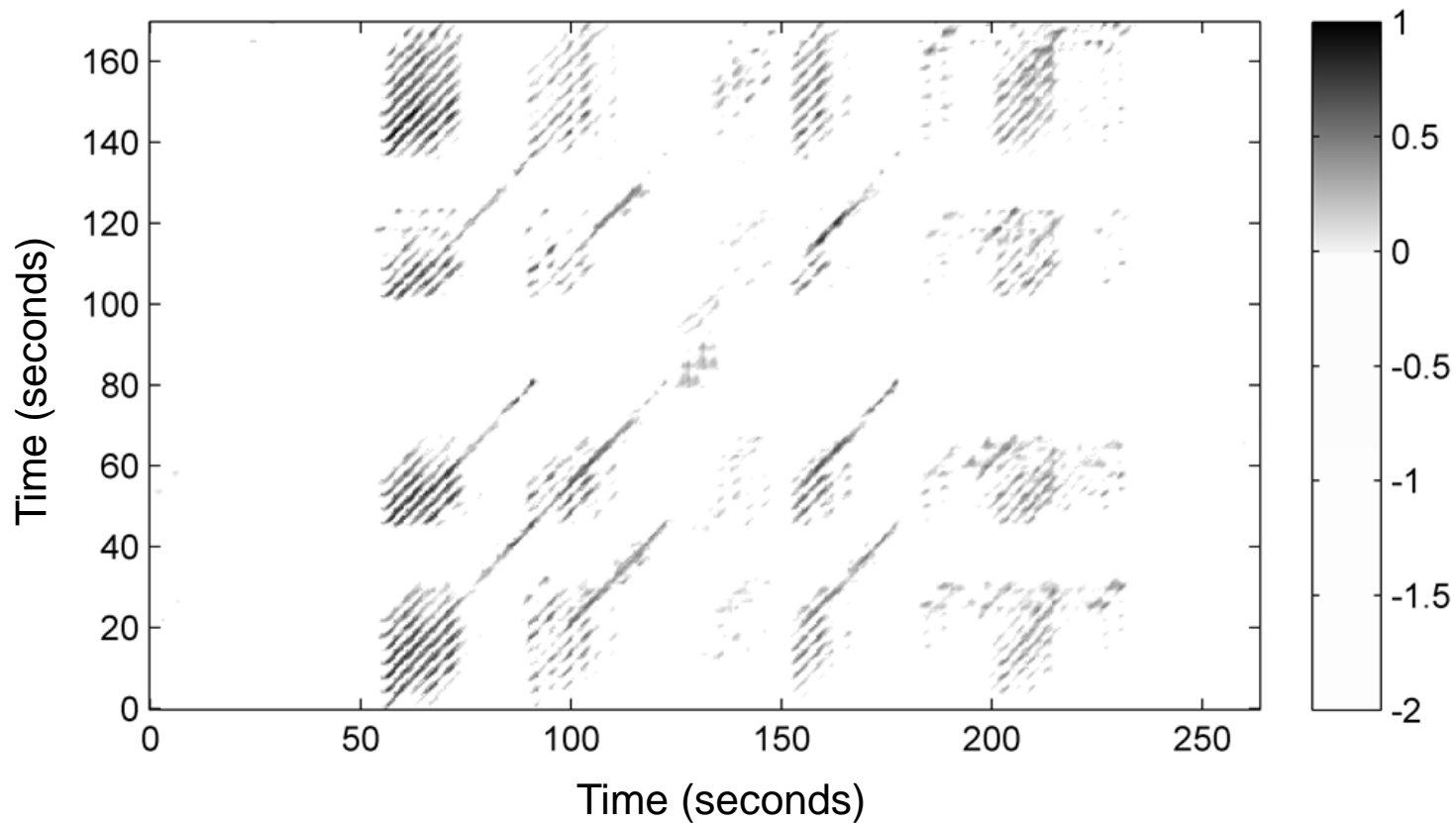
7.3 Version Identification

Fig. 7.18



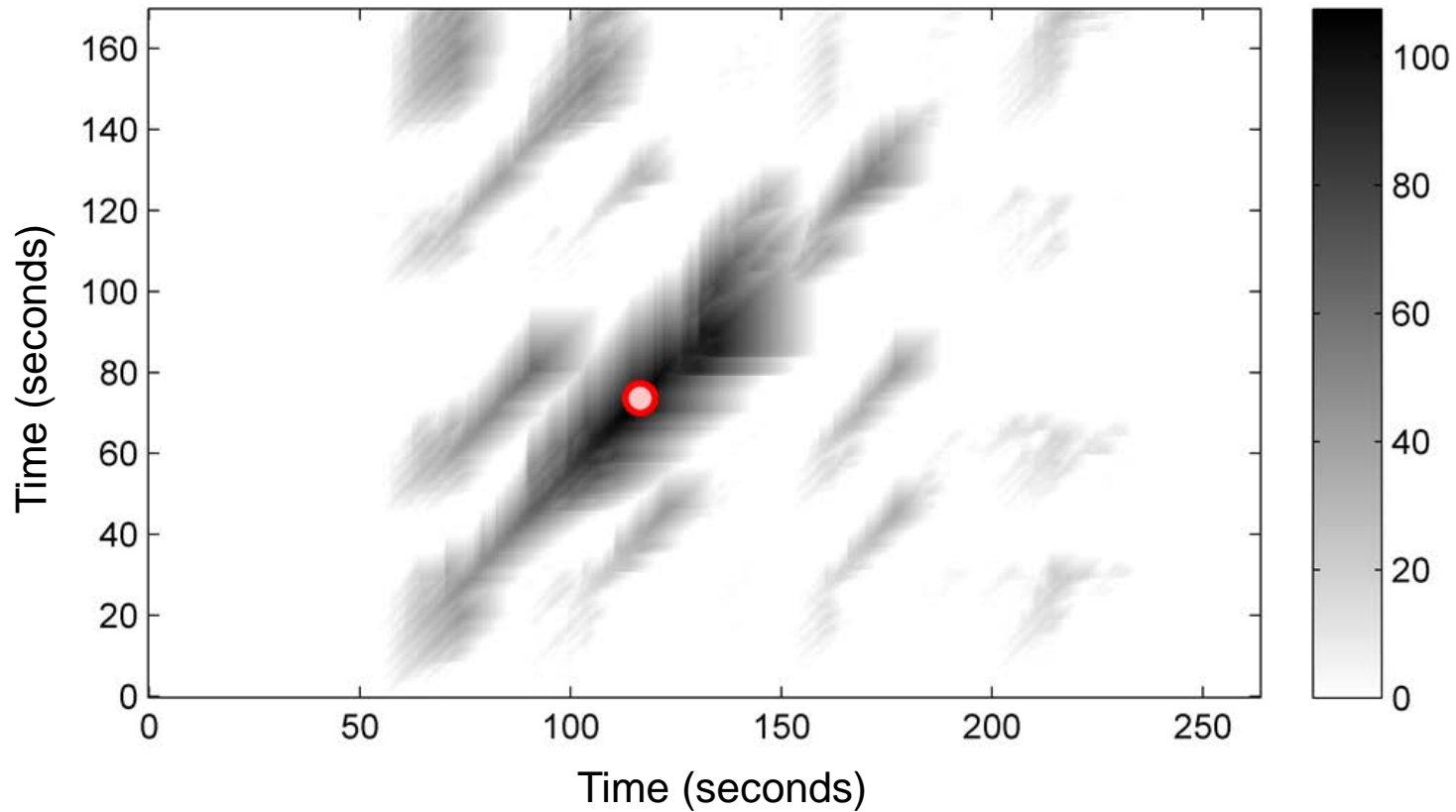
7.3 Version Identification

Fig. 7.19



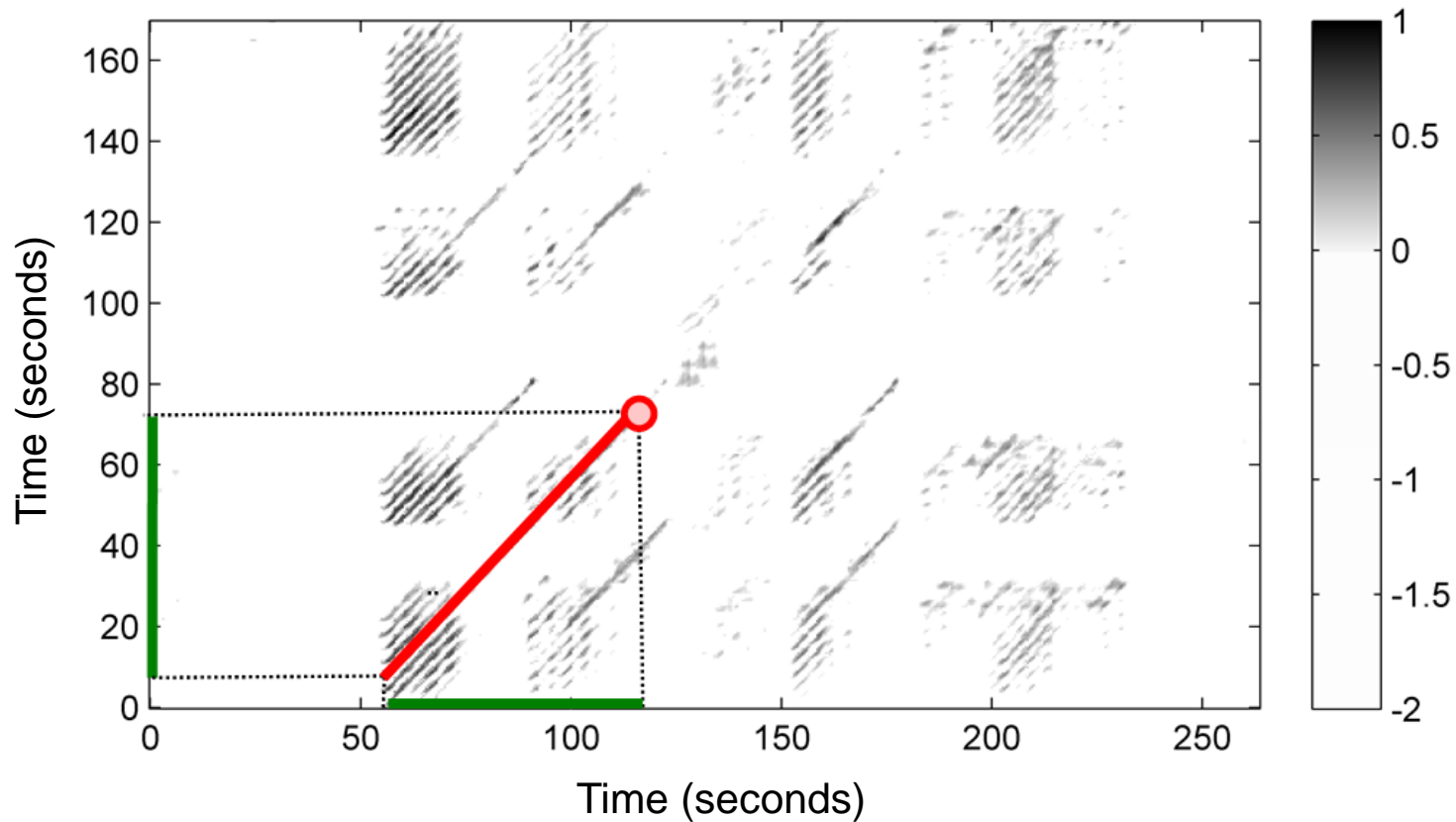
7.3 Version Identification

Fig. 7.19



7.3 Version Identification

Fig. 7.19



7.3 Version Identification

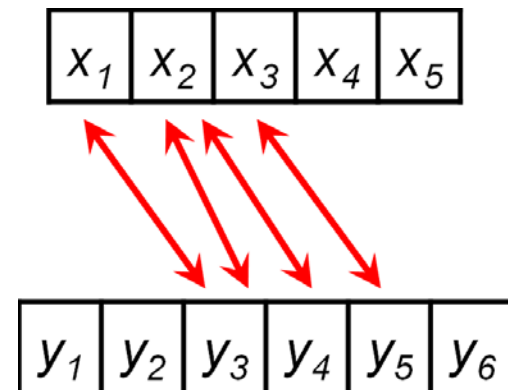
Fig. 7.20

x_5	-2	-2	1	-2	1	0
x_4	-2	1	-2	1	-2	-2
x_3	0	1	-2	-2	1	-2
x_2	0	-2	1	2	-2	1
x_1	1	-2	1	1	0	-2

y_1	y_2	y_3	y_4	y_5	y_6
-------	-------	-------	-------	-------	-------

x_5	0	1	4	2	4	4
x_4	0	3	1	3	3	3
x_3	1	2	0	2	5	3
x_2	1	0	2	4	2	3
x_1	1	0	1	2	2	0

y_1	y_2	y_3	y_4	y_5	y_6
-------	-------	-------	-------	-------	-------

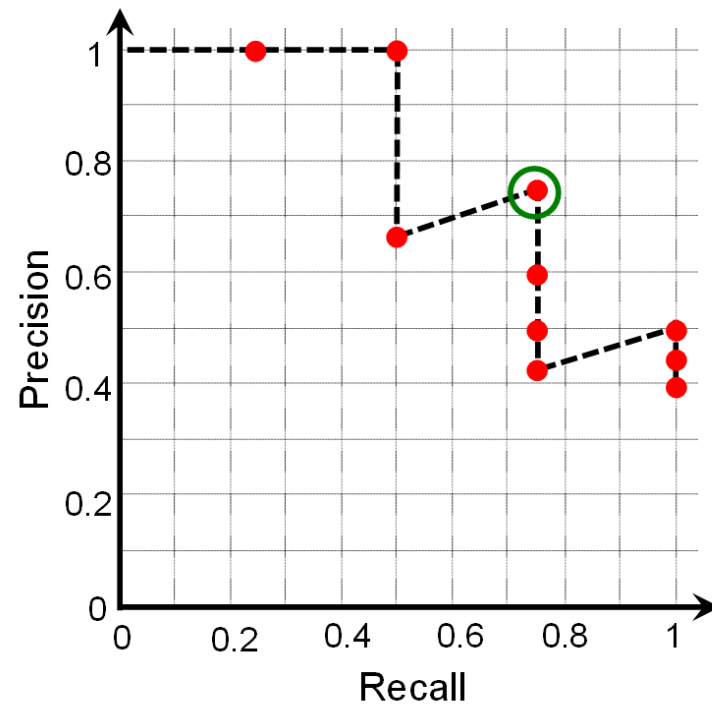


7.3 Version Identification

Fig. 7.21

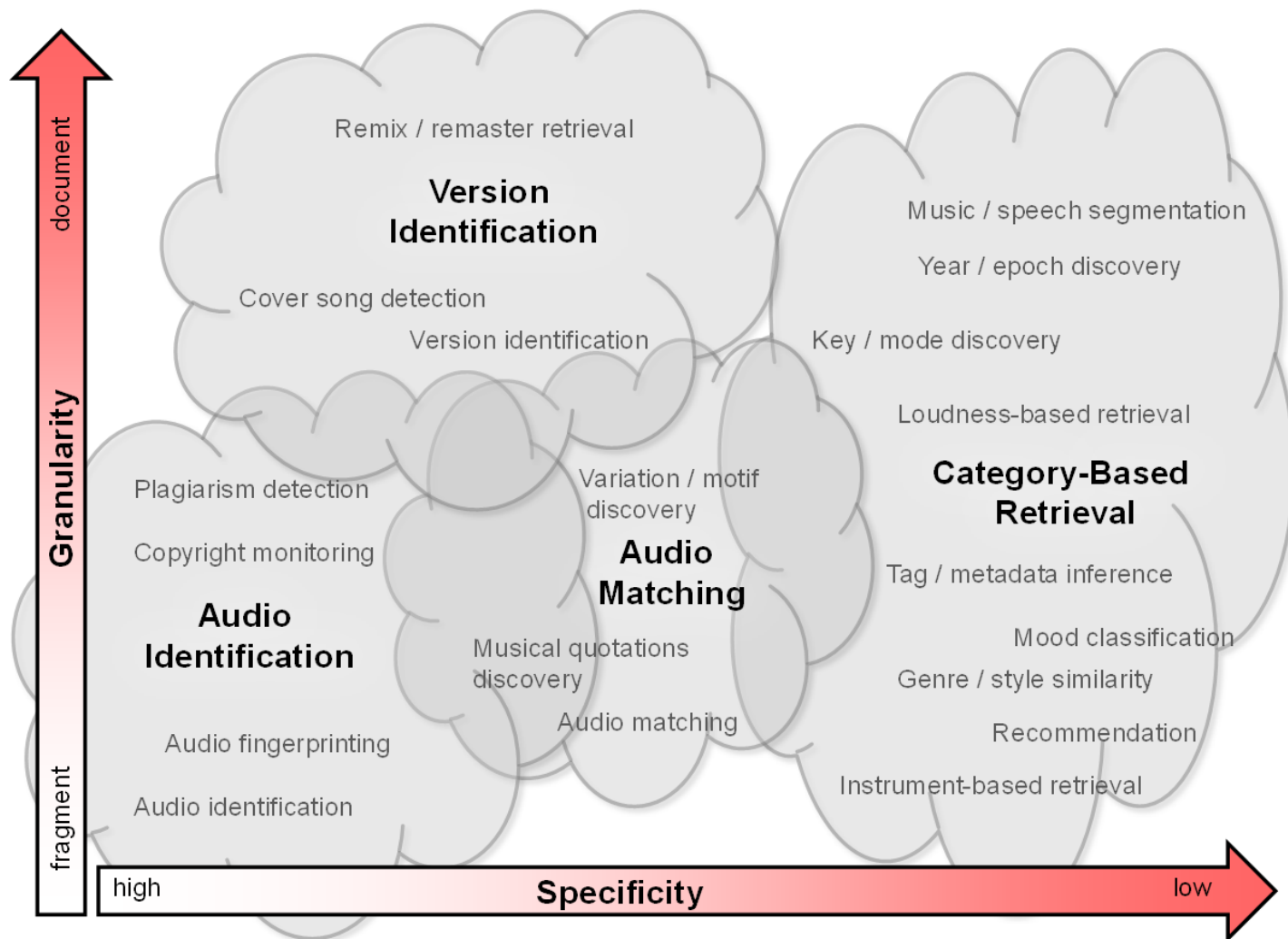
ID	Score	Rank
1	8	10
2	52	2
3	22	5
4	10	9
5	12	7
6	34	3
7	11	8
8	27	4
9	72	1
10	18	6

Rank	ID	Rel.	P(r)	R(r)
1	9	+	1.00	0.25
2	2	+	1.00	0.50
3	6	-	0.67	0.50
4	8	+	0.75	0.75
5	3	-	0.60	0.75
6	10	-	0.50	0.75
7	5	-	0.43	0.75
8	7	+	0.50	1.00
9	4	-	0.44	1.00
10	1	-	0.40	1.00



7.4 Further Notes

Fig. 7.22



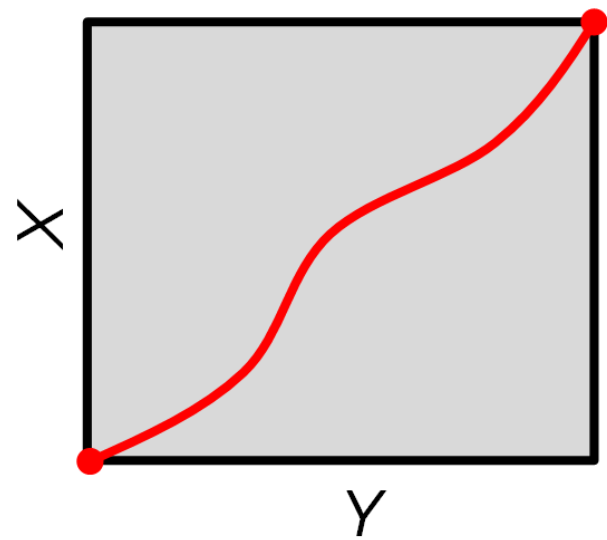
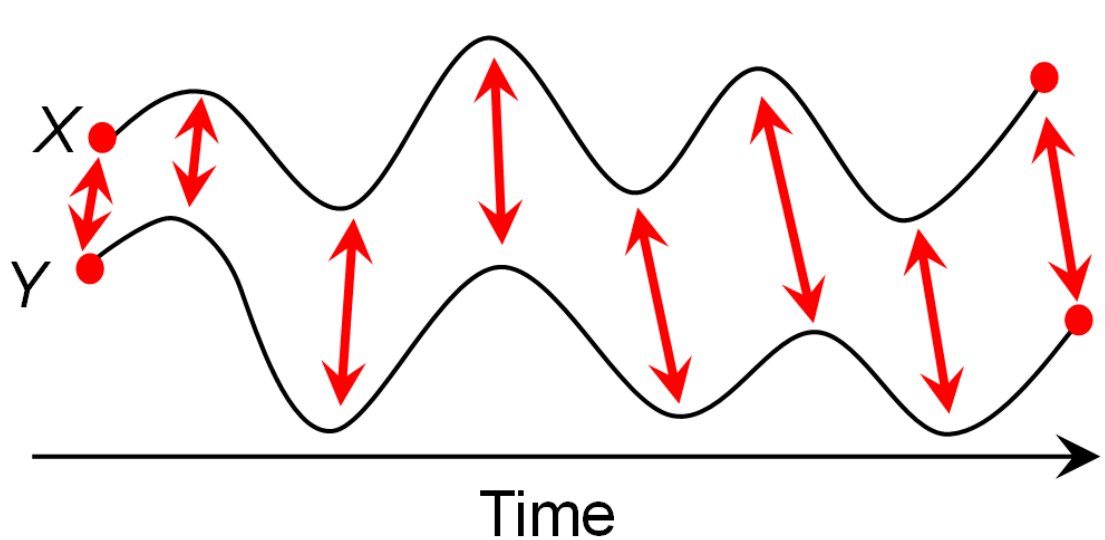
7.4 Further Notes

Table 7.1

Retrieval task	Audio identification	Audio matching	Version identification
Identification	Specific audio recording	Different interpretations	Different versions
Query	Short fragment (5–10 seconds)	Audio clip (10–40 seconds)	Entire recording
Retrieval level	Fragment	Fragment	Document
Specificity	High	Medium	Medium / low
Features	Spectral peaks (abstract)	Chroma (harmony)	Chroma (harmony)

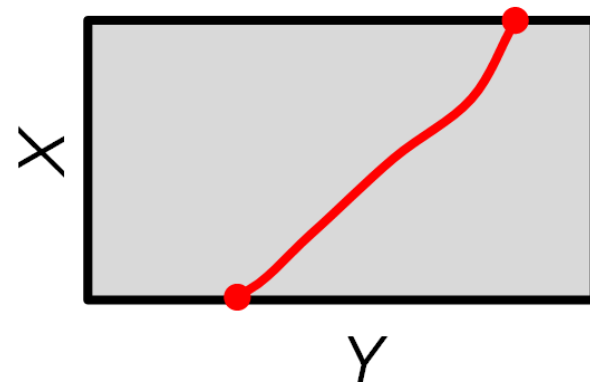
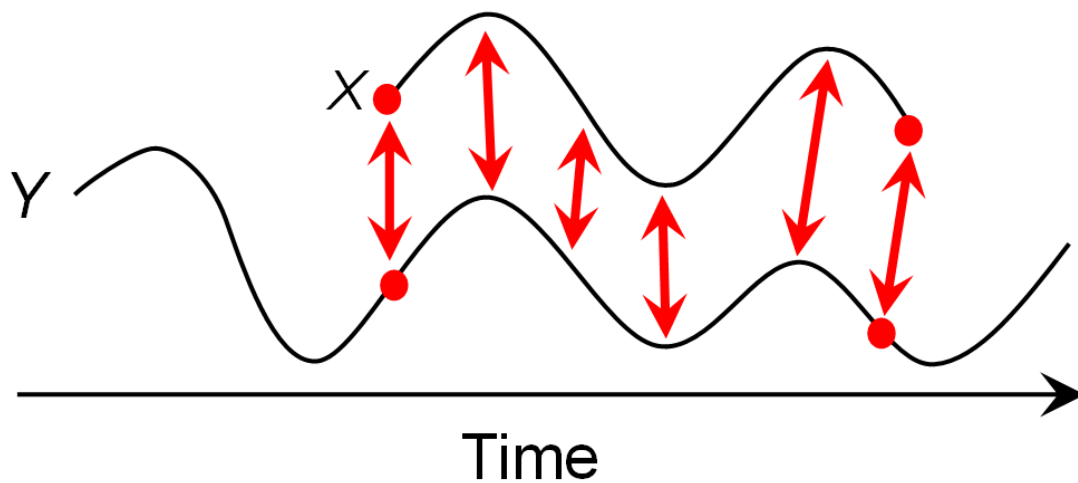
7.4 Further Notes

Fig. 7.23



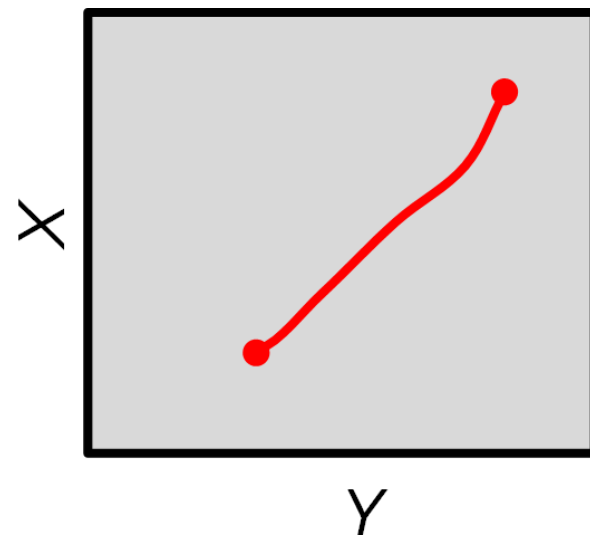
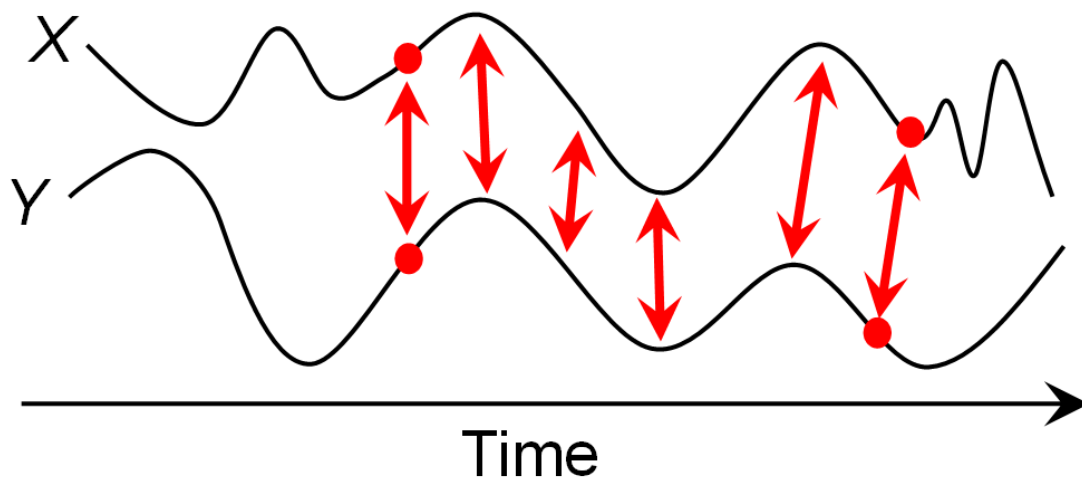
7.4 Further Notes

Fig. 7.23



7.4 Further Notes

Fig. 7.23



7.4 Further Notes

Fig. 7.23

