
Automatic Music Monitoring and Boundary Detection for Broadcast Using Audio Watermarking

Taiga Nakamura
taiga@jp.ibm.com

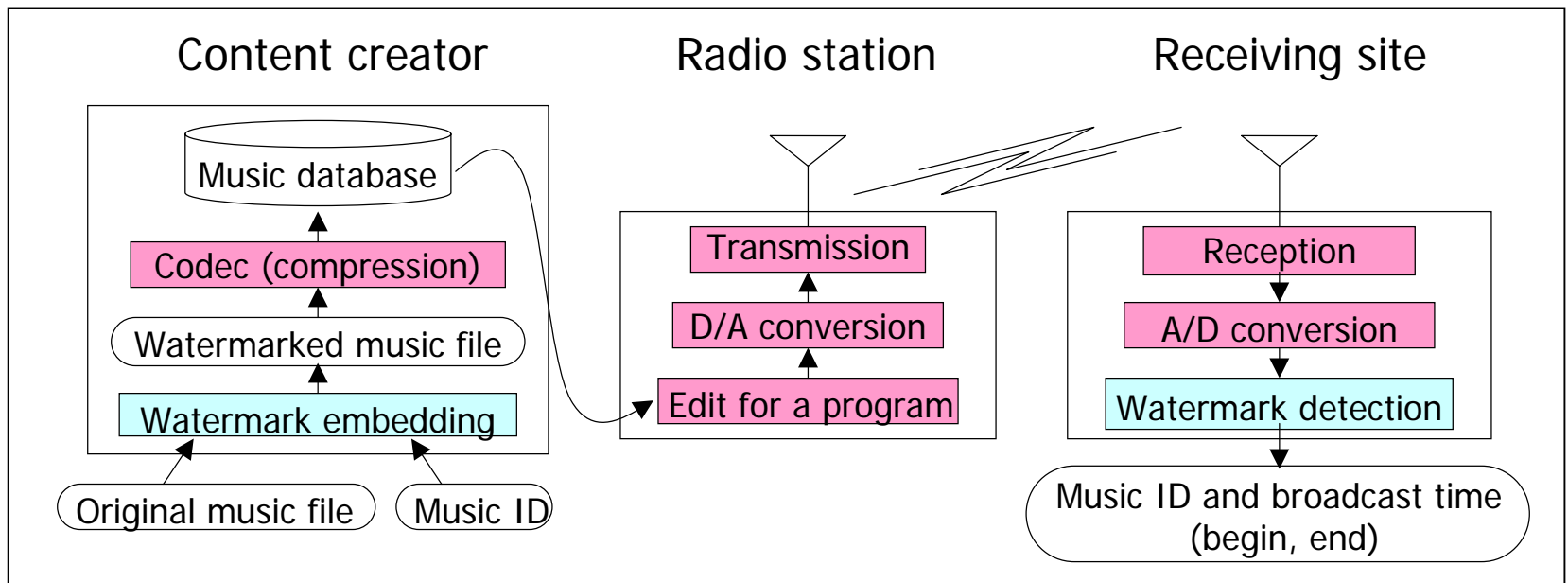
Ryuki Tachibana
ryuki@jp.ibm.com

Seiji Kobayashi
kobayas@jp.ibm.com

IBM Japan, Tokyo Research Laboratory
Jan. 21, 2002

Objectives

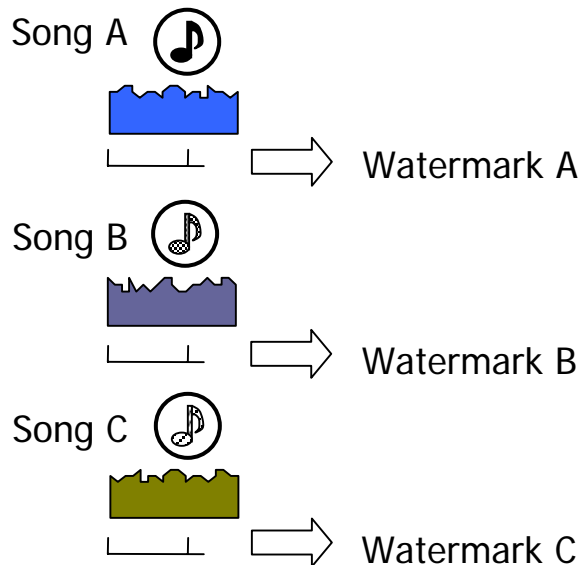
- ⊙ Broadcast monitoring = means of enabling automatic detection of the songs that have been on the air
 - What song? **(Correctness of music identification)**
 - At what time, and for how long? **(Time resolution)**
- ⊙ Broadcast monitoring as an application of watermarking



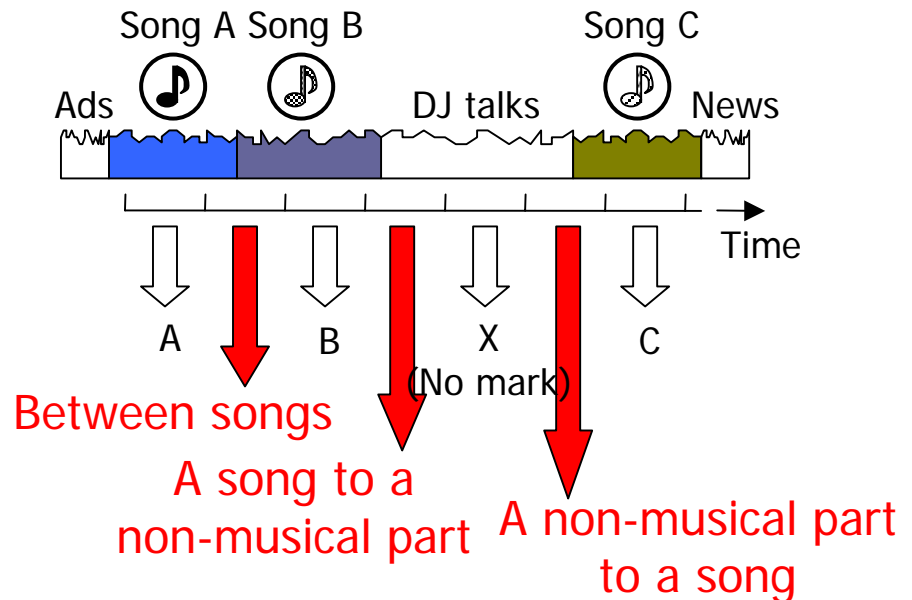
Difficulties

- ⦿ Music is on the air either in sequence or between other programs (news, sports, weather forecasts,...)
- ⦿ Watermark signal is interrupted or interfered with at the boundary of songs
 - Leads to higher error rates. and therefore a bad time resolution

Watermark detection directly from a music file

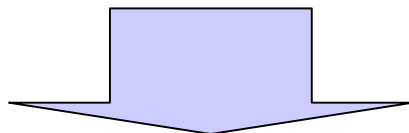


Watermark detection in the broadcast monitoring



Problem to Solve = Boundary Detection

- ⦿ How to detect all broadcast music correctly?
- ⦿ How to determine when the broadcast of each song started and ended as accurately as possible?

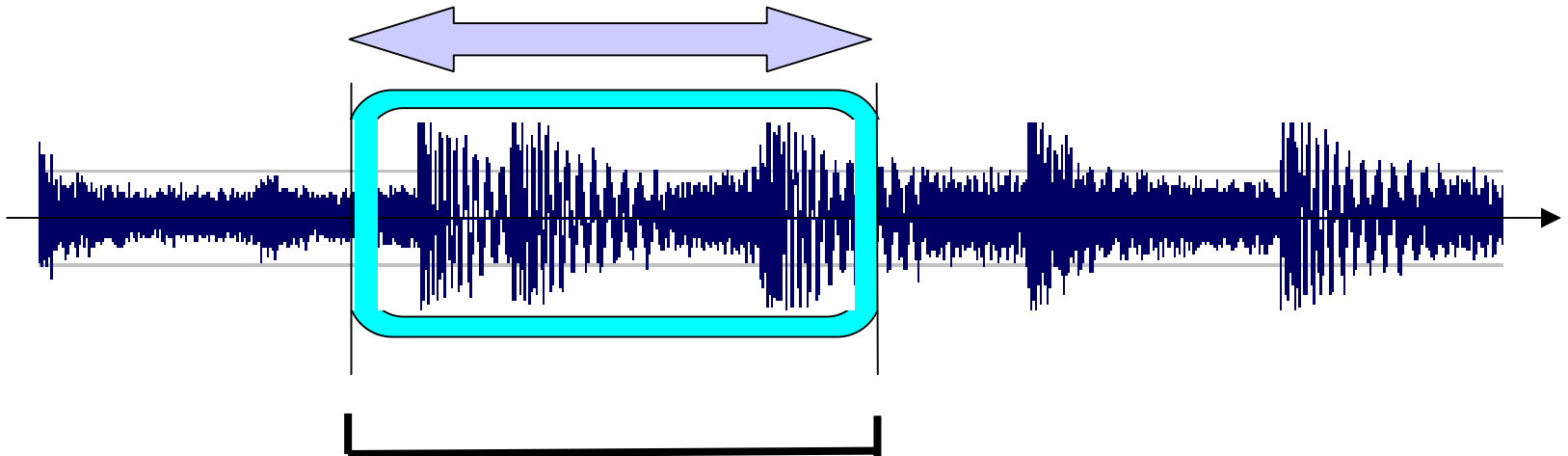


- ⦿ How to deal with the content transition, determining the existence of content boundaries and avoiding outputs of wrong detection results?

Detection Window

- ⦿ Audio watermark is extracted from an music segment that has certain amount of time length.
- ⦿ The time length of the “detection window” represents a minimum (or nominal) length required for correct watermark detection.

“Detection window” = content segment for a single detection process

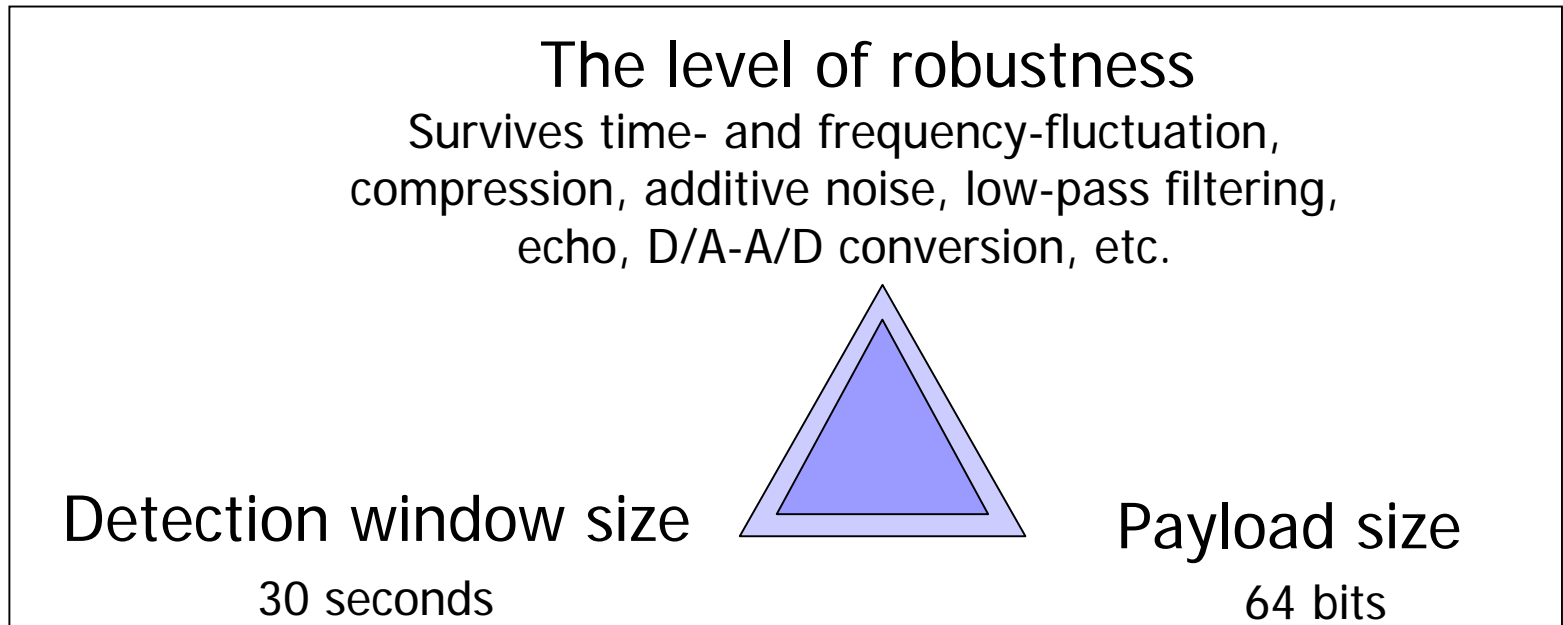


Basic Algorithm

⊙ We use the algorithm based on [Tachibana01]

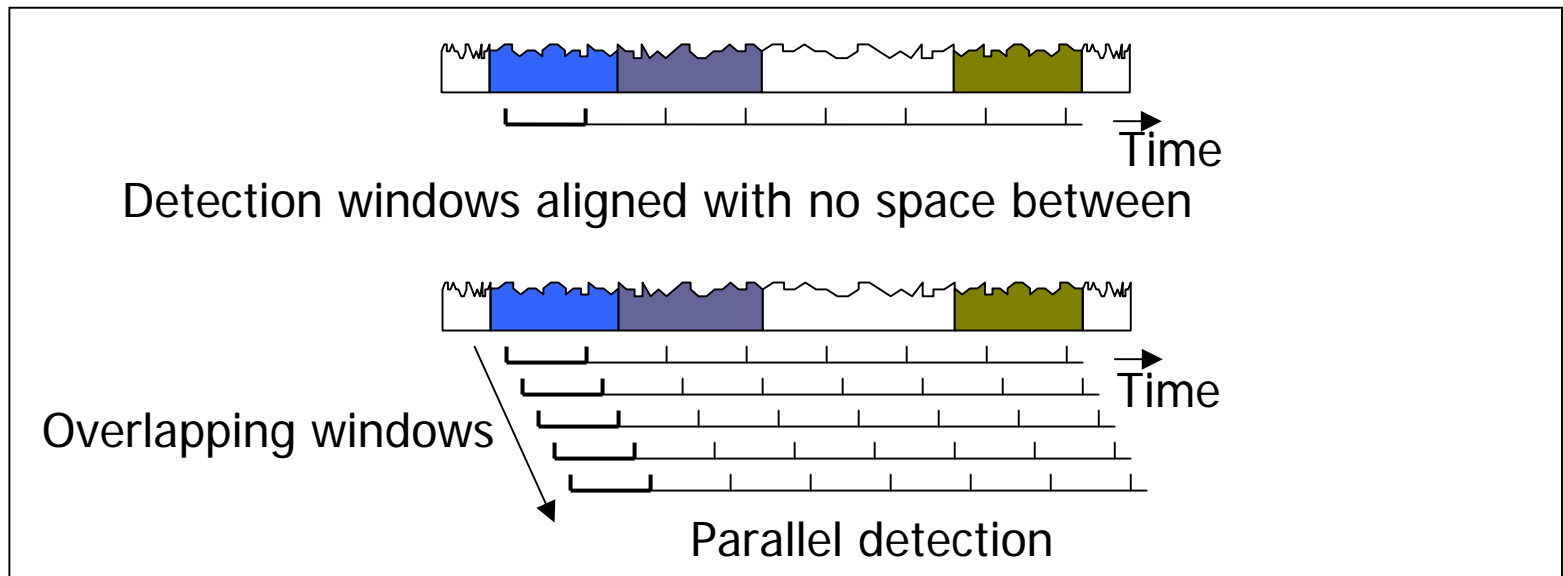
- Ryuki Tachibana, Shuichi Shimizu, Seiji Kobayashi, Taiga Nakamura, "An audio watermarking method robust against time- and frequency-fluctuation," *Security and Watermarking of Multimedia Contents III*, 2001.

⊙ Three basic properties to describe algorithms



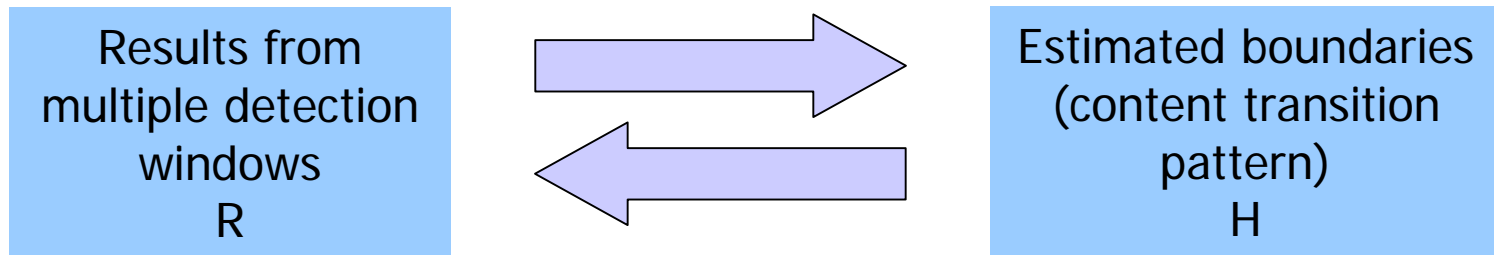
Overlapping Detection Windows

- ⦿ If the detection windows are aligned with no space between them, the time resolution is limited to the length of the detection window in the best case.
- ⦿ To improve time resolution, use multiple detection sets that have their origins slightly offset from each other.
 - If the overlap ratio between neighboring windows is α , the boundary will be included in at least $\lceil 1/(1-\alpha) \rceil$ windows.



Method of Boundary Detection

- ⦿ Filter incorrect results at the boundaries, to prevent them from being displayed as final outputs.
- ⦿ Determine the most probable pattern of content transition for given detection windows from multiple windows.

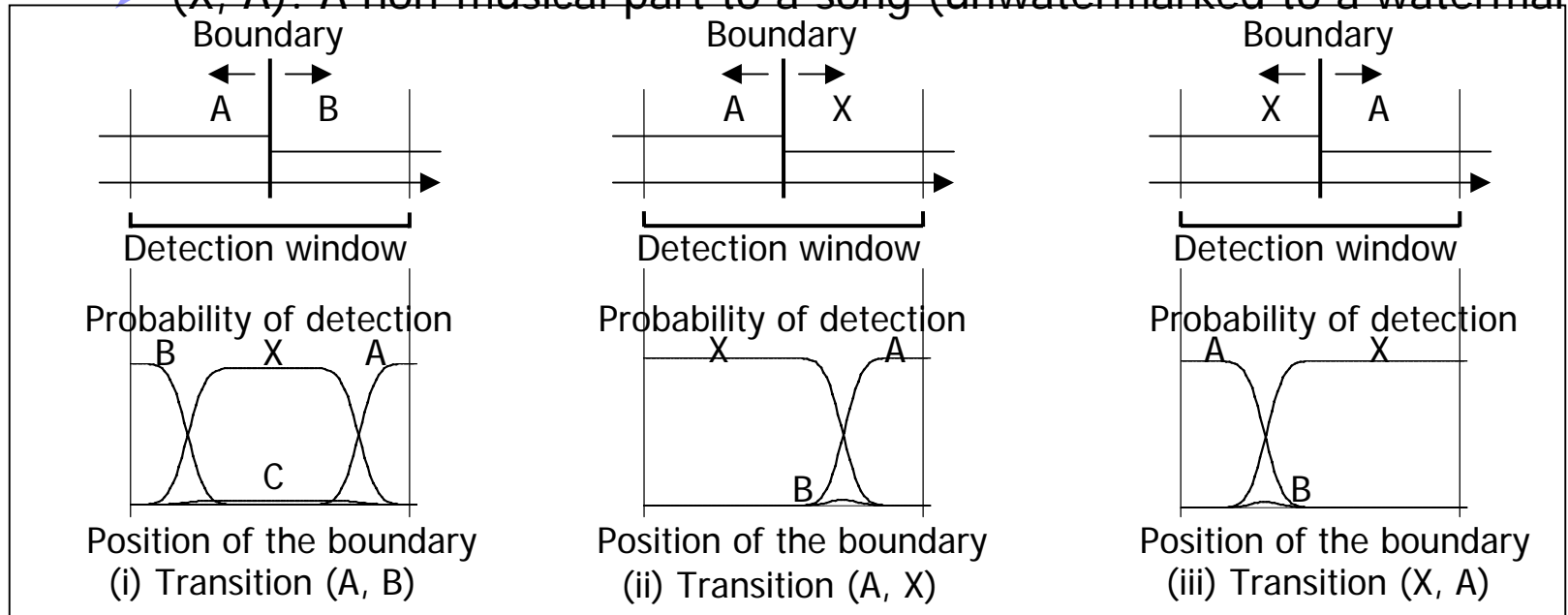


$$\begin{aligned} H_{selected} &= \operatorname{argmax}_H (\operatorname{Prob}(H | R_{given})) \\ &= \operatorname{argmax}_H \left(\frac{\operatorname{Prob}(R_{given} | H) \operatorname{Prob}(H)}{\operatorname{Prob}(R_{given})} \right) \end{aligned}$$

Assumption for Evaluating Probabilities

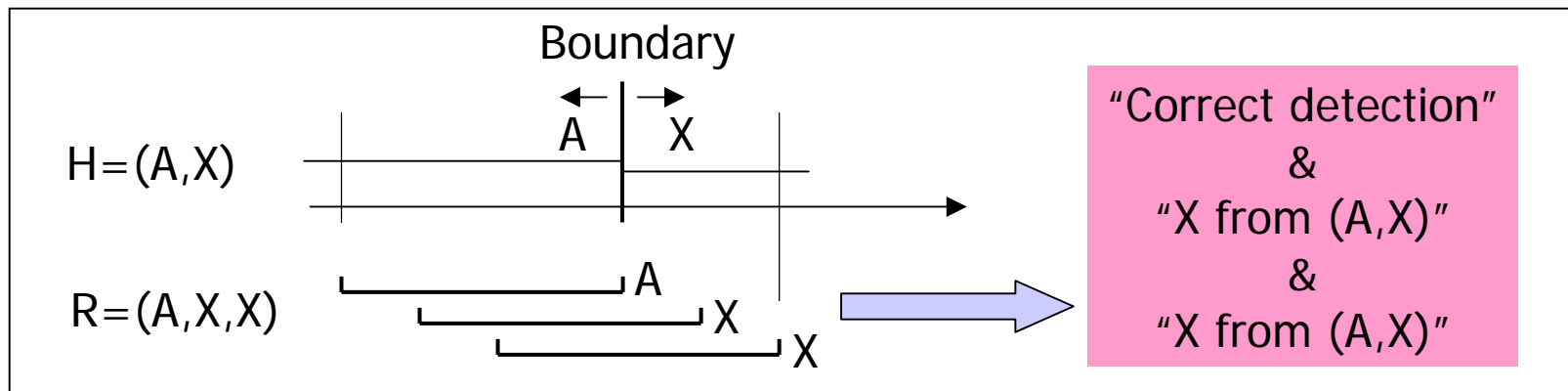
Assumption of qualitative characteristics of three basic content transition patterns:

- The effect of content transition: the degradation of the average SNR
- (A, B): One song to another song (a watermark to another watermark)
- (A, X): A song to a non-musical part (a watermark to unwatermarked)
- (X, A): A non-musical part to a song (unwatermarked to a watermark)



Example

- ⊙ If $R=(A,X,X)$, the most probable transition pattern is $H=(A,X)$.



Set of possible detection results
(the case of 3 overlapping windows)

(X,X,X) , (A,A,A) , (X,X,A) ,
 (A,A,X) , (A,A,B) , (X,A,X) ,
 (X,A,A) , (X,A,B) , **(A,X,X)** ,
 (A,X,A) , (A,X,B) , (A,B,X) ,
 (A,B,A) , (A,B,B) , (A,B,C)

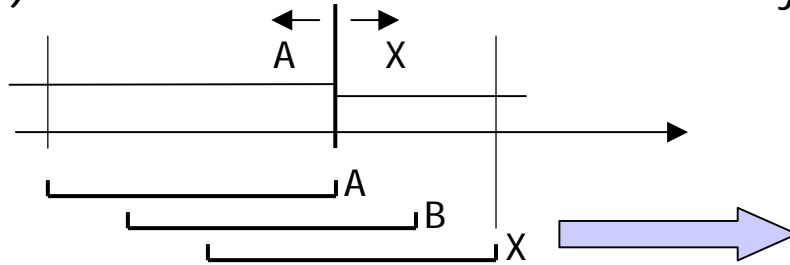
Set of possible content transition patterns

(X) , (A) , (B) , (C) , (D) ,
 (X,A) , (X,B) , (X,C) , (X,D) , **(A,X)** , ...,
 (C,X) , (C,A) , (C,B) , (C,D) , (D,X) , (D,A) ,
 (D,B) , (D,C) , (D,E) ,
 (X,A,X) , (X,A,B) , (X,A,C) , (X,A,D) , ...,
 (D,C,B) , (D,C,D) , (D,C,E) , (D,E,F) ,
 (X,A,X,A) , (X,A,X,B) , (X,A,X,C) , ...,
 (D,X,C,X) , (D,X,D,X) , (D,X,E,X)

Example: More Complicated Case

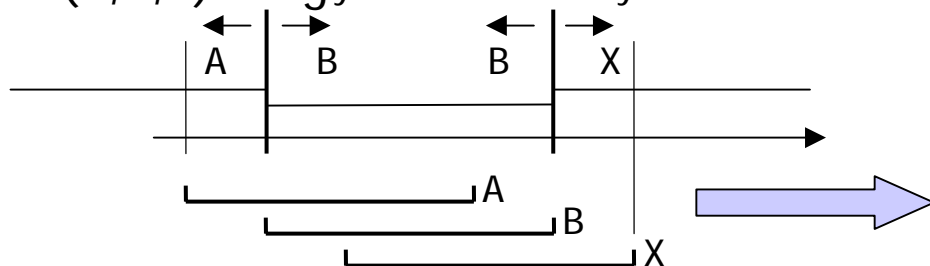
- ⊙ If $R=(A,B,X)$, the most probable transition pattern depends on the bit error rate occurring at the boundary.

- $H(A,X)$: Low bit error rate at the boundary



"Correct detection"
&
"Bit error"
&
"X from (A,X)"

- $H(A,B,X)$: High bit error rate at the boundary



"A from (A,B)"
&
"Correct detection"
&
"X from (B,X)"

Example of Derived Rules (cont.)

⊙ Rules applied for 3 detection windows

Observed detection pattern R	Selected content transition pattern H (Low bit error at the boundary)	Selected content transition pattern H (High bit error rate at the boundary)
R=(X,X,X)	H=(X)	H=(X)
R=(A,A,A)	H=(A)	H=(A)
R=(X,X,A)	H=(X,A)	H=(X,A)
R=(A,A,X)	H=(A,X)	H=(A,X)
R=(A,A,B)	H=(A,B)	H=(A,B)
R=(X,A,X)	H=(X,A,X)	H=(X,A,X)
R=(X,A,A)	H=(X,A)	H=(X,A)
R=(X,A,B)	H=(X,B)	H=(X,A,B)
R=(A,X,X)	H=(A,X)	H=(A,X)
R=(A,X,A)	H=(A)	H=(A)
R=(A,X,B)	H=(A,B)	H=(A,B)
R=(A,B,X)	H=(A,X)	H=(A,B,X)
R=(A,B,A)	H=(A)	H=(A,B,A)
R=(A,B,B)	H=(A,B)	H=(A,B)
R=(A,B,C)	H=(A,C)	H=(A,B,C)

***Sorry, this table was incorrectly reversed in the manuscript!**

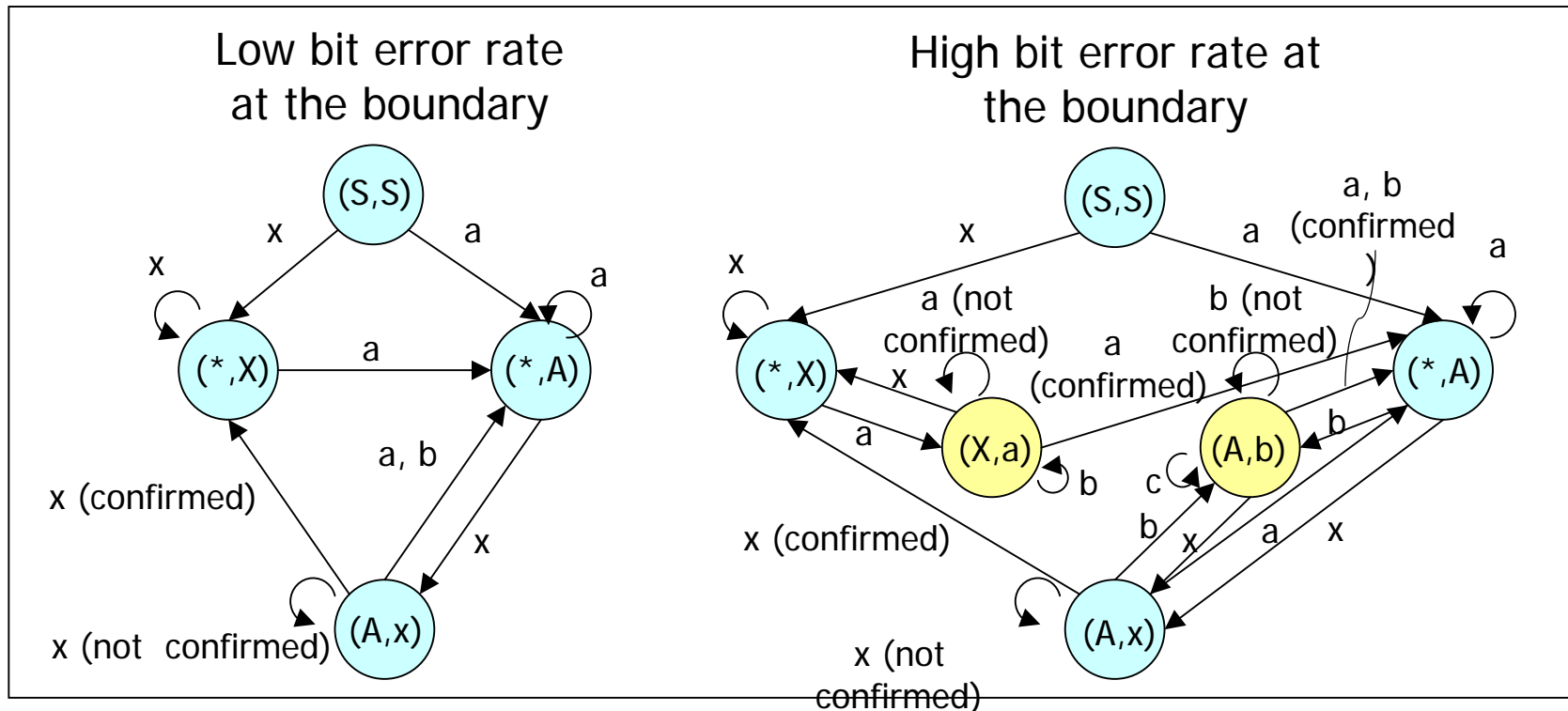
Finite State Model

- ⊙ Representation of decision rules using a finite state model as a general expression of the rule sets
- ⊙ Convenient for implementing real-time processing
 - Detection results can be evaluated successively.
 - The output can be displayed as soon as a certain content is confirmed to be broadcast
- ⊙ Keep track of three variables during the detection process according to status of the detection results
 - current: the detection result of the most current detection window
 - prev: the second most recent detection result from one or more detection windows that returned the same results
 - pprev: the third most recent detection result from one or more detection windows that returned the same results

Finite State Model (cont.)

Derived model: (pprev, prev)

- The next operation is determined according to the value of current.
- Capital characters represent the values that are confirmed to really exist and therefore can be output, while lower-case characters represent the unconfirmed values.



Experimental Results

- ⊙ In early 2001, we conducted an experiment on the real FM radio broadcast.
 - Embed ISRC (International Standard Recording Code) information (5 ASCII + 7 digits) and some additional digits as 64-bit watermarks
 - MPEG-2 AAC 128 kbps compression after embedding
 - Detection window of 30 seconds
 - Overlapping ratio of 90%
 - 2 music samples, each embedded with one of 10 different ISRCs
 - 12 music samples, each embedded with one ISRC

Content transition pattern	Correct detection ratio	Average time resolution	Worst time resolution
(X,A)	100%	-3.9 sec	-13.0 sec
(A,B)	100%	+1.5 sec	+3.0 sec

Summary

- ⊙ By using multiple results from overlapping detection windows, we can determine the most probable content transition pattern.
- ⊙ Two rule sets are obtained depending on how often bit errors occur at the boundary
- ⊙ The derived rules can be represented as a finite state model, which is useful for real time monitoring.
- ⊙ We could obtain accurate identification and good time resolution in the experiment in real FM-broadcast environment.