

大規模楽曲信号データ ベースに対する高速検索

北海道大学 大学院情報科学研究科
情報理工学専攻

喜田拓也

BGMで流れている曲がとても気になる



こ、この曲!?

ワタシ気になります!



➡ 音声信号をクエリとした楽曲検索が必要です

音声信号そのままでは大変なので...

クエリの音声データ

変換

音楽指紋

検索

大量楽曲の音楽指紋の索引構造

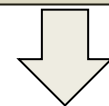
変換

大量楽曲の音声信号データ群

ロビンソン/スピッツ

時間: 4分23秒

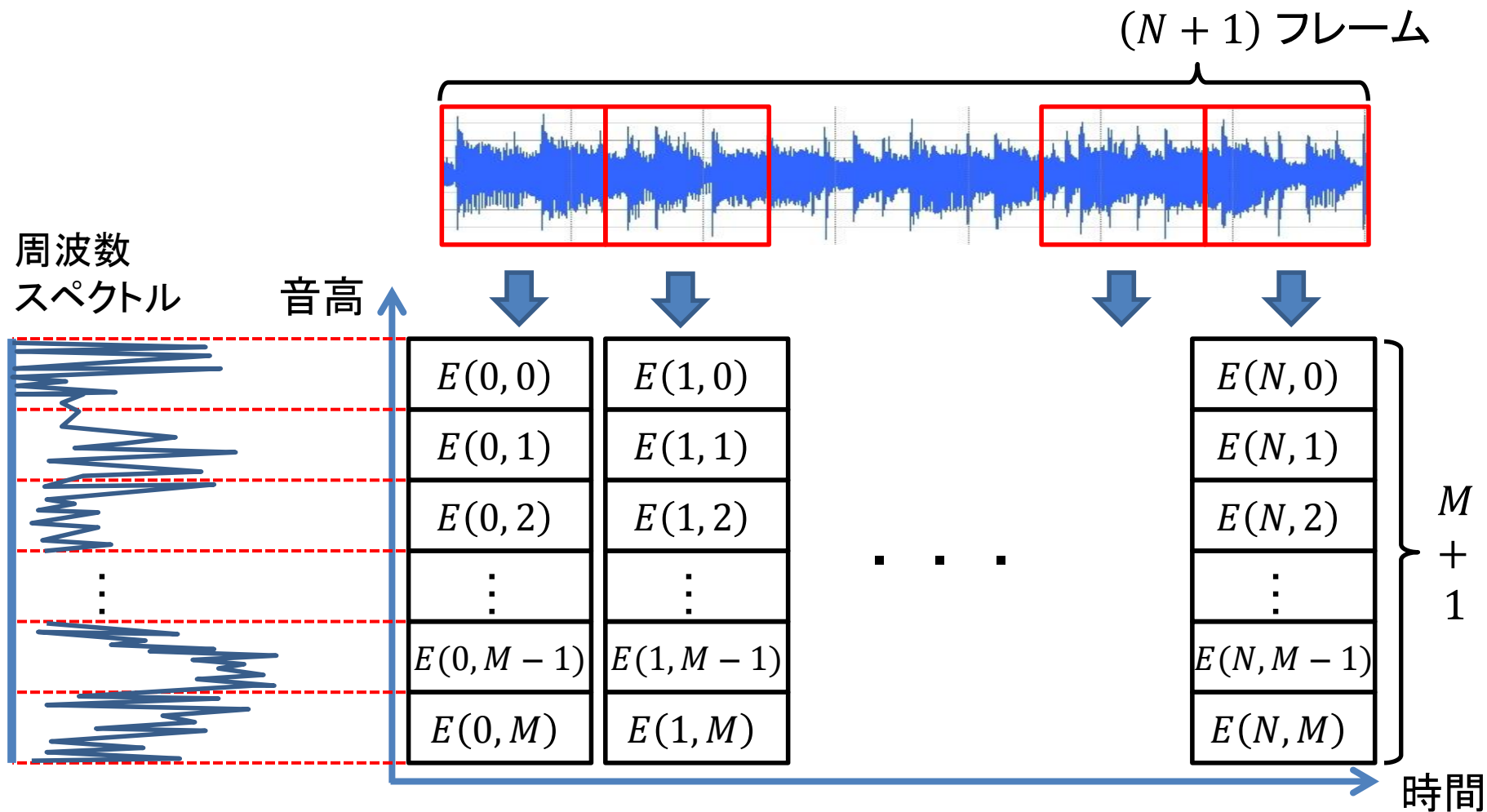
サイズ: 6MB(mp3, 192kbps)



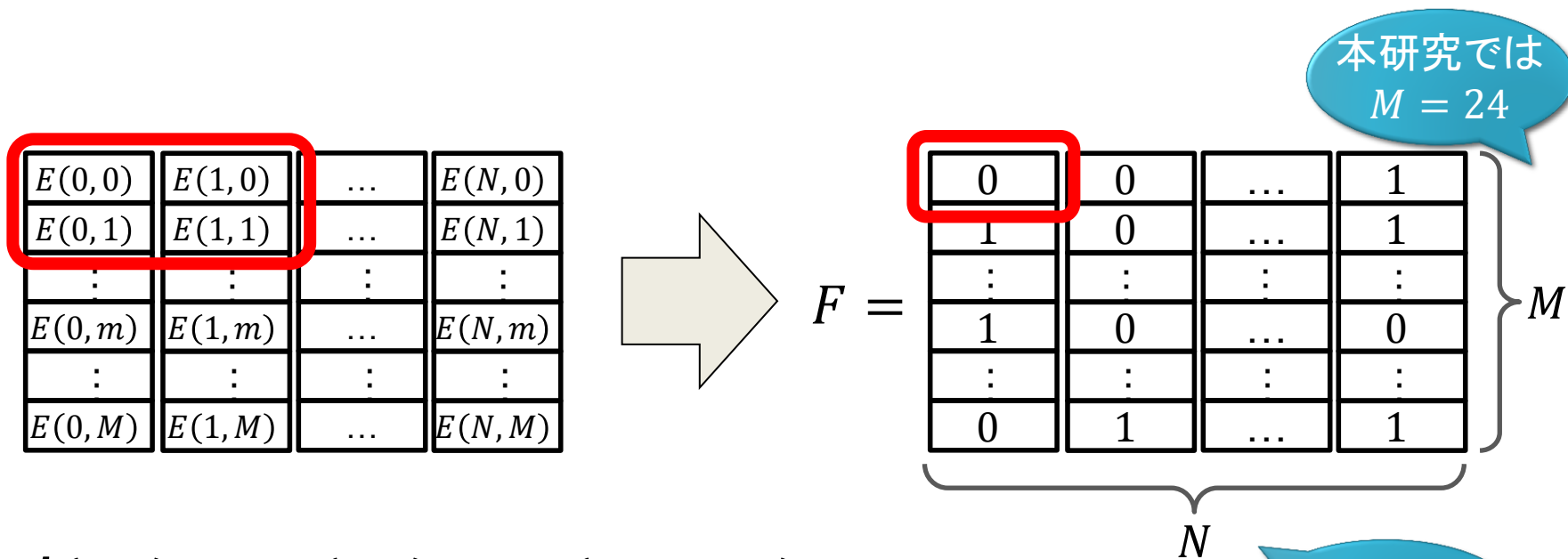
27.8KBのビット列

➡ コンパクトな離散データ系列に変換して検索します

音楽指紋の生成 (スペクトル解析)



音楽指紋の生成 (ビット列への離散化)



$$E'(i, j) = E(i, j) - E(i, j + 1) - (E(i - 1, j) - E(i - 1, j + 1))$$

フレーム数

$$F(i, j) = \begin{cases} 1 & \text{if } E'(i, j) > 0, \\ 0 & \text{if } E'(i, j) \leq 0 \end{cases}$$

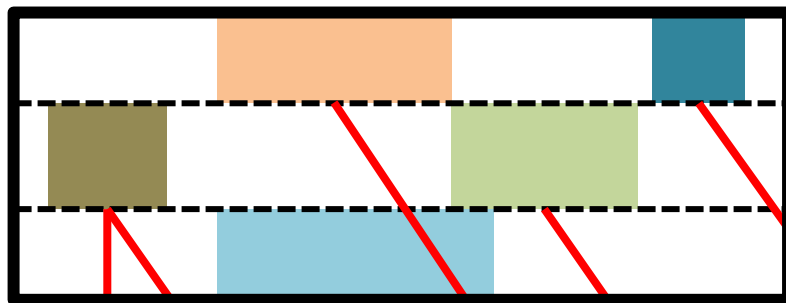
音楽指紋を用いた楽曲検索の課題点

- 開始点が一致するとは限らない
 - クエリが楽曲の一部, あるいは先頭に無音部分があるなど
 - 先頭から順番に比較したのでは不十分
- ノイズによる音楽指紋の変化
 - ある程度の近似検索が必要
- 多次元ベクトルの近似検索
 - NN検索は, 次元数が多くなると非常に困難になる
(めっちゃめっちゃ時間がかかる)
 - 高速化のためには工夫が必要

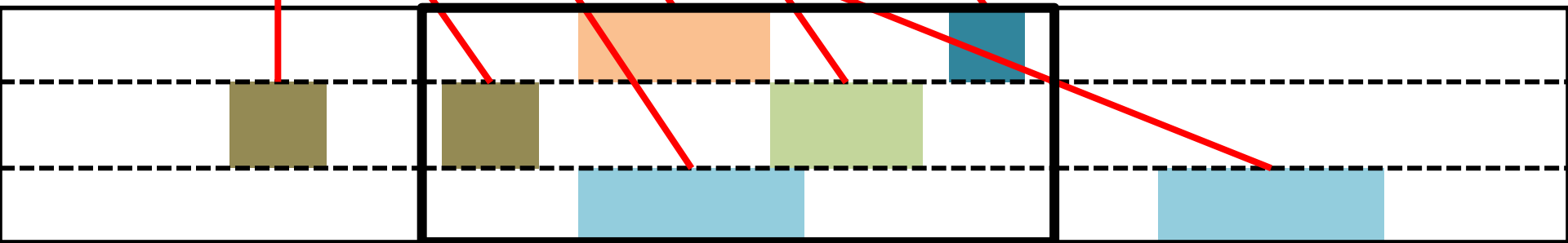
基本アイデア

完全に一致している部分が多い
⇒ 全体的に類似している

クエリの音楽指紋



データベース

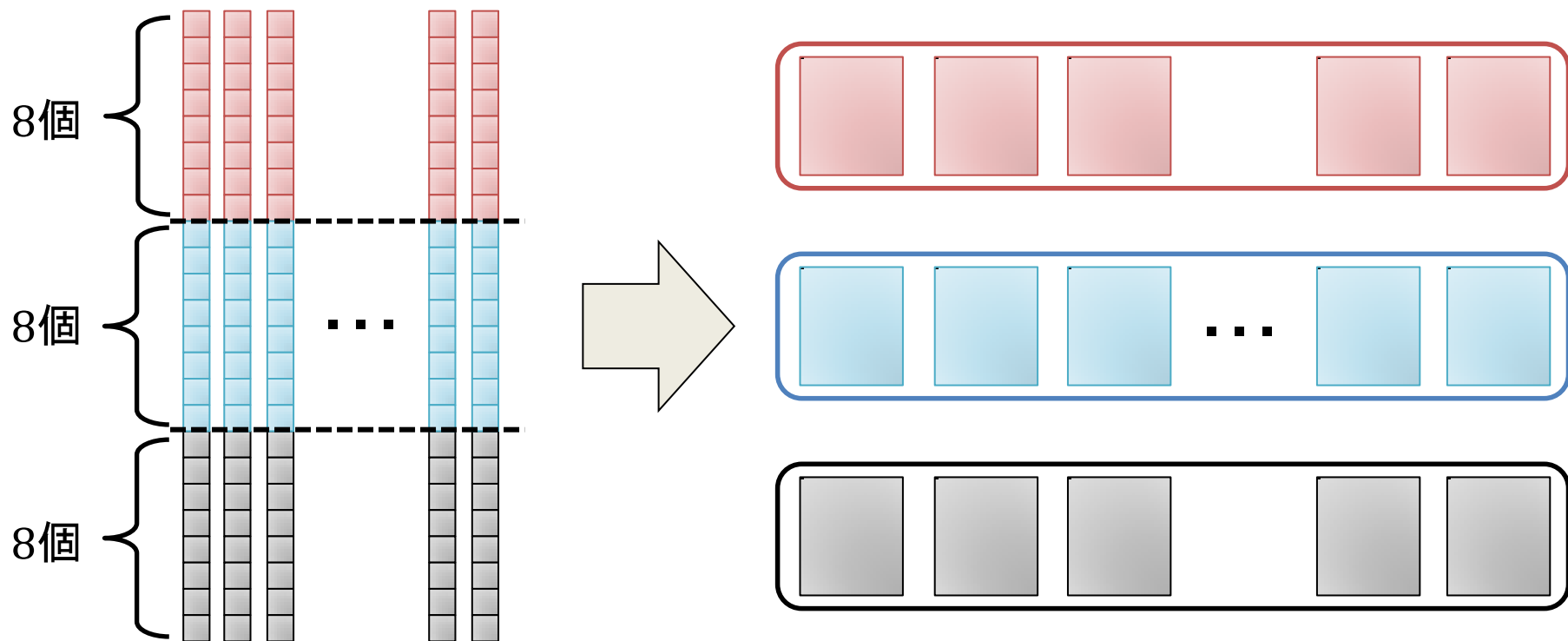


$i - 1$ 番目の曲の
音楽指紋

i 番目の曲の
音楽指紋

$i + 1$ 番目の曲の
音楽指紋

音楽指紋全体を複数の文字列とみなす



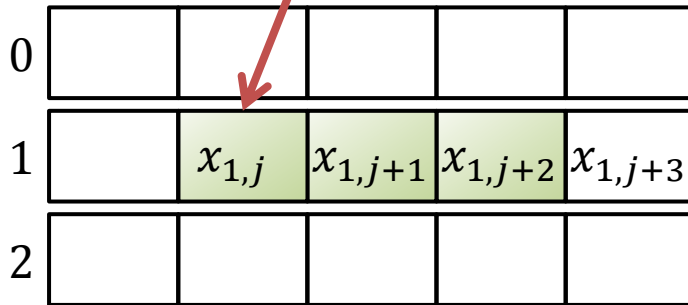
フレームごとに8ビットずつ
3段に分割する

8ビットを1個の文字とみなすと
3本の文字列ができる

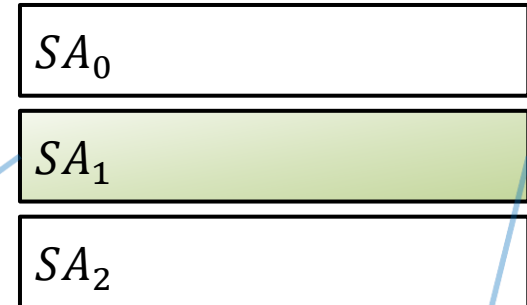
検索方法（小片の厳密一致探索）

ランダムに開始点を選択

クエリ



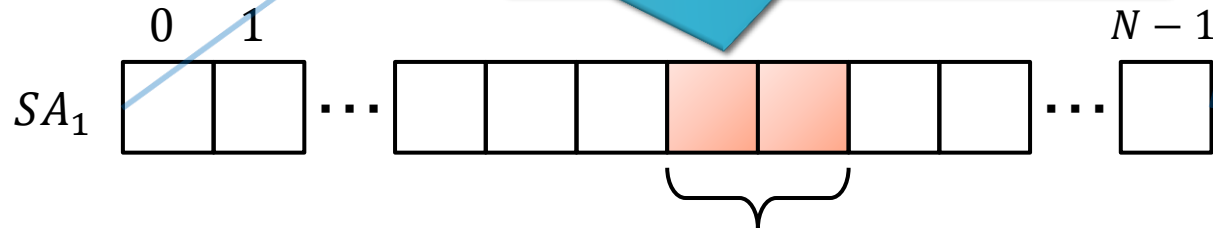
データベースの接尾辞配列



二分探索

$$X = x_{1,j}x_{1,j+1}x_{1,j+2}\dots$$

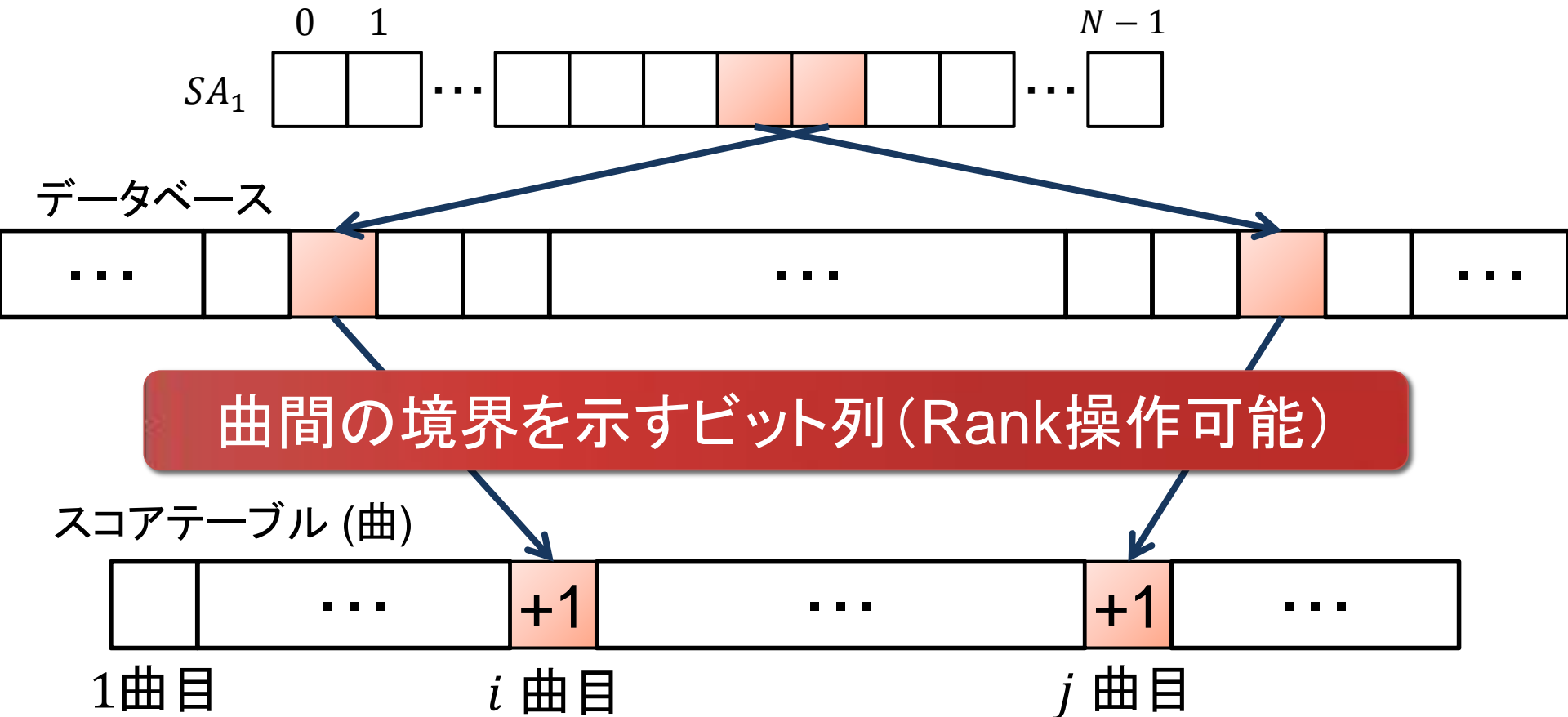
Xを接頭辞として持つ接尾辞



一致する箇所の数 R が閾値 Δ
を下回るまで X を伸ばす

$$R < \Delta$$

検索方法（楽曲のランキング）



ある一定回数, 小片の探索を繰り返し,
最も得点の高い楽曲を最終的に出力する

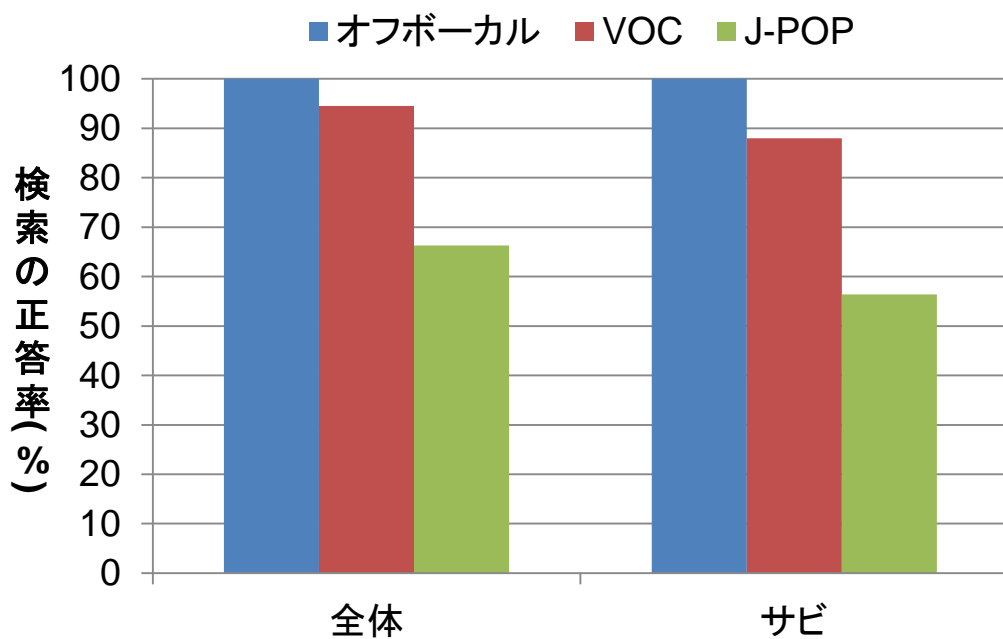
検索精度

データベース(107849曲)

- 7849曲の楽曲データ
 - VOCALOIDオリジナル楽曲, J-POP曲, インストゥルメンタル曲ほか
 - 平均長: 3分49秒
- 10万曲分の擬似楽曲データ
 - 乱数により生成した擬似音楽指紋

クエリ

- データベースに同一の楽曲が存在するが、歌唱者が異なる、もしくはオフボーカルバージョン
 - VOCALOID曲 100曲
 - J-POP曲 78曲(ノイズ多, アレンジ異etc)
 - オフボーカル曲 22曲
- 曲全体, サビ部分のみの2種類
 - 曲全体のクエリの平均長: 4分19秒
 - サビのみのクエリの平均長: 27秒



フレーム長 0.1秒,
閾値 $\Delta = 15$
繰り返し回数 2000回
ランキングの上位10曲を出力
その中にクエリと同じ楽曲があれば成功

データベース107849曲に対して
接尾辞配列...250メガ文字

検索時間

| | 提案手法 | | 既存手法 |
|----------|----------------------------------|---------------------|--------------------------------------|
| 計算機 | Intel Core i5 (2.5GHz) メモリ8GB | | Intel Core i7 (1.73GHz) メモリ4GB |
| データベース | 7849曲 | 107849曲 (擬似データ含) | 8740曲 |
| 検索時間 / 曲 | 0.0176秒 | 0.0351秒 | 0.4~0.6秒 |

Qingmei Xiao, Motoyuki Suzuki, and Kenji Kita. Fast hamming space search for audio fingerprinting systems. In *Proceedings of the 12th International Society on Music Information Retrieval, ISMIR 2011*, pp. 133-138, 2011.