

MIDIギターと

NMFによる音響信号処理の統合を用いた

ギター演奏の自動採譜

北原研究室 修士課程 2年

大塚匡紀

1. 研究の背景と目的

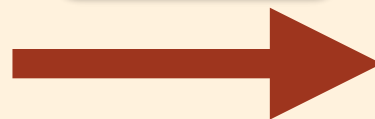
■ 自動採譜とは？

本研究での自動採譜

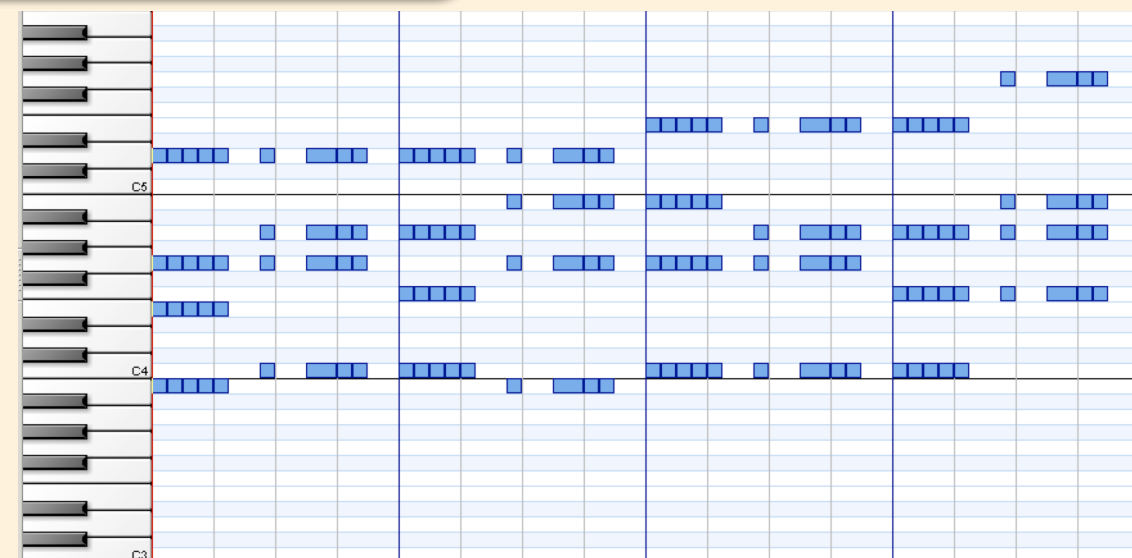
楽器演奏



変換



MIDI データ



音響信号処理の分野において**重要な課題の一つ**

■ MIDI データとは？

EBNFで表すと…

音符 = (発音時刻, 消音時刻, 音高)

音高 = { ド, ド#, レ, レ#, ミ, ファ, ファ#, ソ, ソ#, ラ, ラ#, シ } × オクターブ

楽曲 = (音符*)

可視化

ピアノロール



■ ギター演奏の自動採譜の重要性

ポピュラー音楽におけるギターの役割

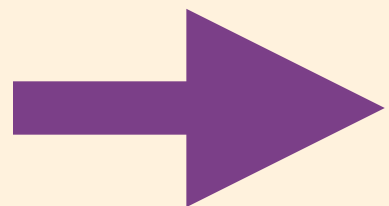
ポピュラー音楽では数多くの楽曲でギター演奏があり、曲の要であることもある

ギターの演奏人口の多さ

他楽器と比較しギター演奏人口は多い

DTMの普及

DTM(コンピュータで音楽を製作すること)が一般的になり、ギター演奏者のためのMIDI入力が必要が拡大



ギター演奏の自動採譜は重要

■ MIDIギター

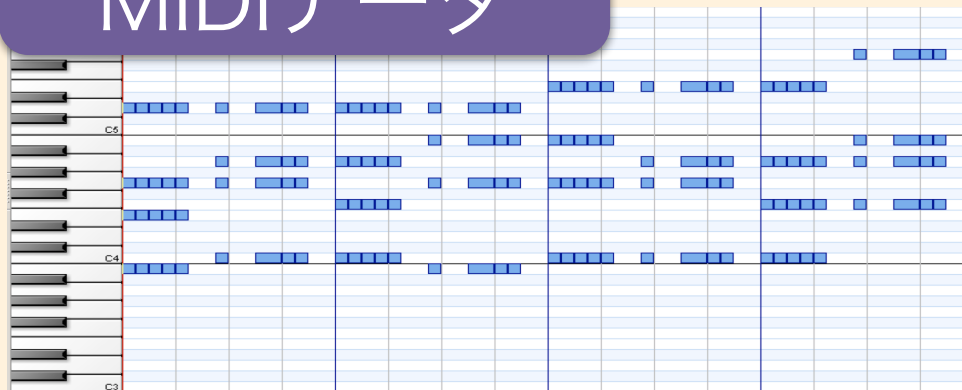
MIDIギター



リアルタイム
に出力

装着

MIDIデータ



問題点

MIDI キーボードなどと比較すると、特定の奏法において出力される演奏情報の正確性に問題がある

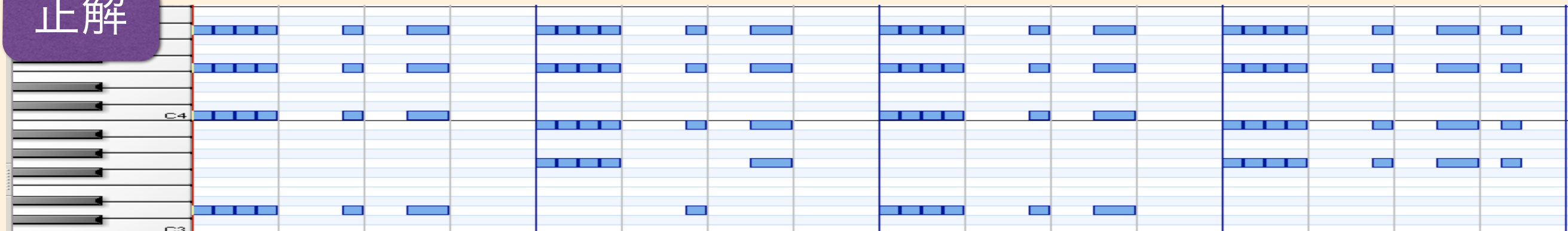
構造



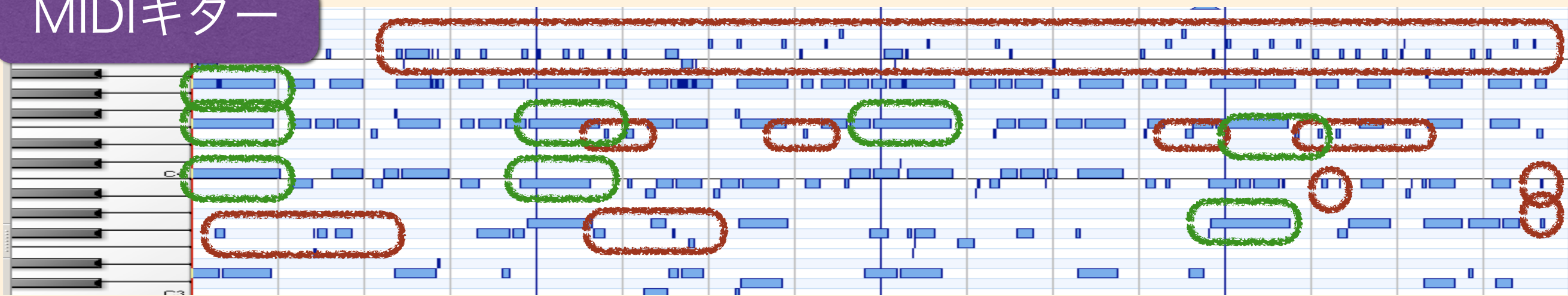
各弦の振動をマグネット
により解析し、弦ごとに
音高を推定する構造

■ MIDI ギターの問題の例

正解



MIDIギター



存在しない音の採譜

連続した音の結合

原因

MIDIギターの弦の振動を解析する構造のため、共鳴してしまった振動を拾ってしまったり、連続した振動を拾えないことによるもの

■ 採譜精度向上のアイデア



+



統合

採譜精度の向上を図る

■ 既存研究

身体的制約を加えた研究

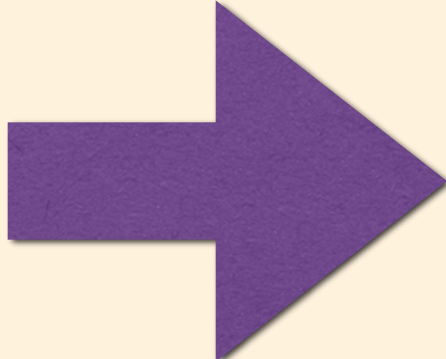
- ・ K. Arimoto, T. Fujishima, and M. Goto, 2006.
- ・ K. Yazawa, K. Itoyama and H. G. Okuno, 2014
- ・ X. Fiss and A. Kwasinski, 2011
- ・ A. M. Barbancho and A. Klapuri, 2012

画像処理を加えた研究

- ・ M. Paleari, N. Huet, A. Schutz and D. Slock, 2008
- ・ T. Yamagami and K. Itou, 2014

トップダウンな制約

演奏動画に依存する



MIDIギターと音響信号処理の併用による高精度化はなされていない

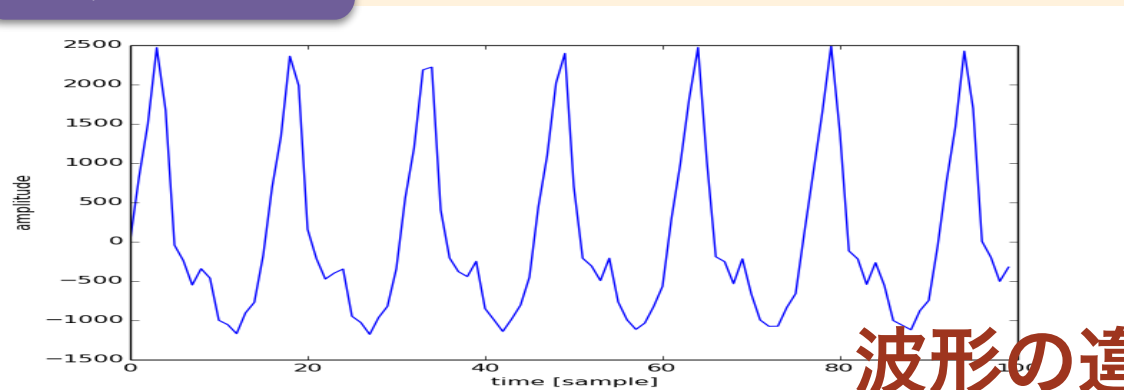
2. NMFによる音響信号処理で の採譜手法

音響信号の仕組み

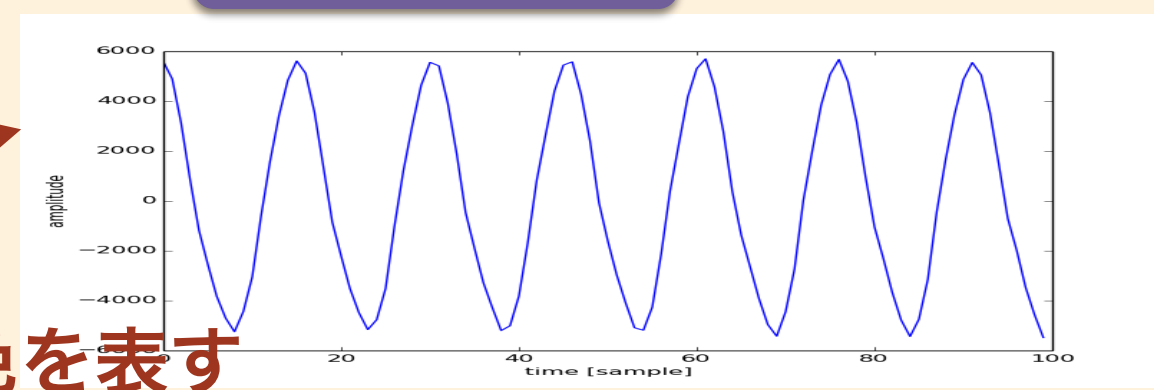
・ 楽音の構造

- ・ 楽音は振動であり、正弦波の組み合わせによって成り立つ
- ・ 楽器ごとに音色が違うのは正弦波の組み合わせの違いによるもの

ギター



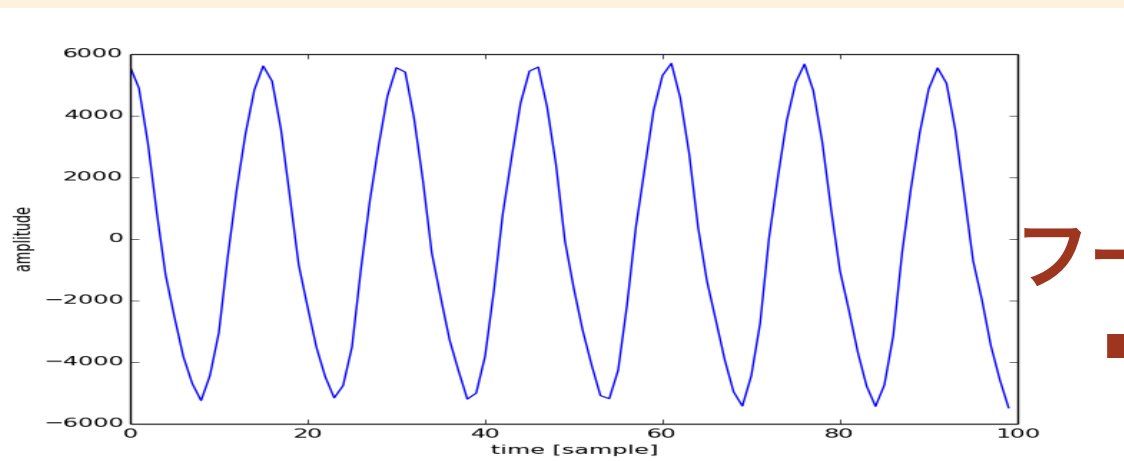
リコーダー



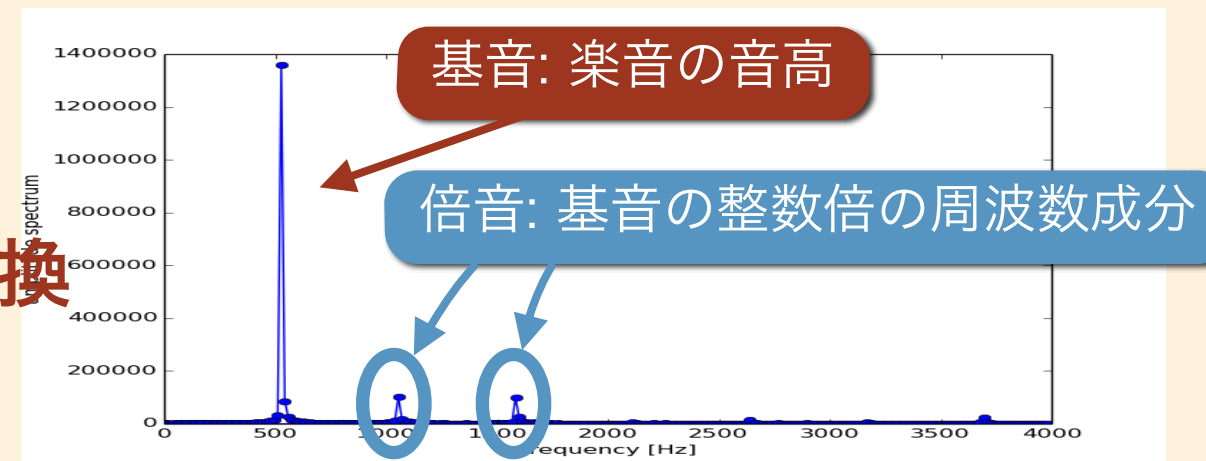
波形の違いが音色を表す

・ スペクトル

- ・ フーリエ変換により、周波数成分を表すスペクトルに変換できる



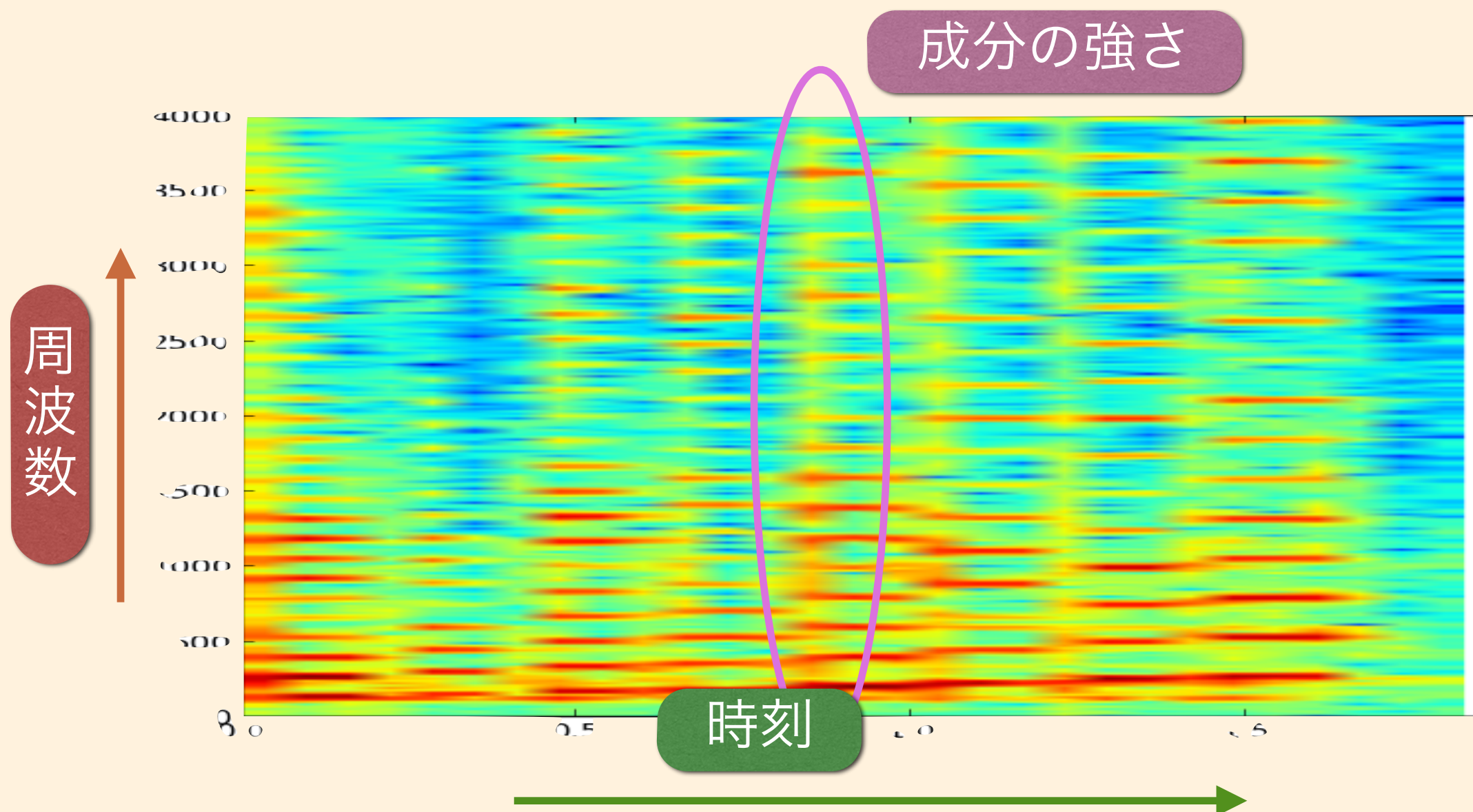
フーリエ変換



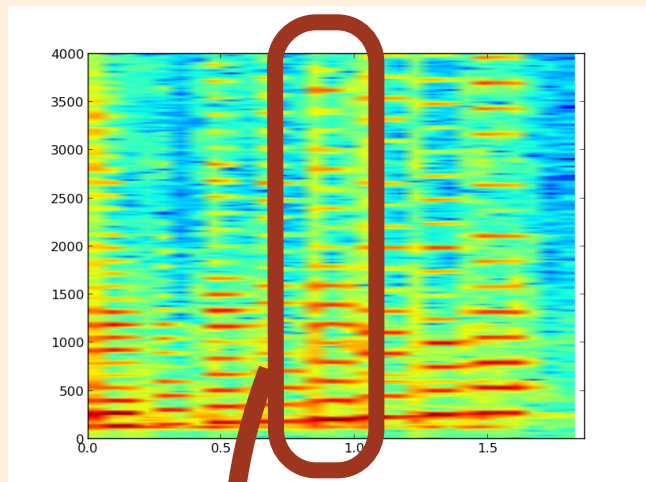
■ 音の可視化

・ 音響信号を可視化するスペクトログラム

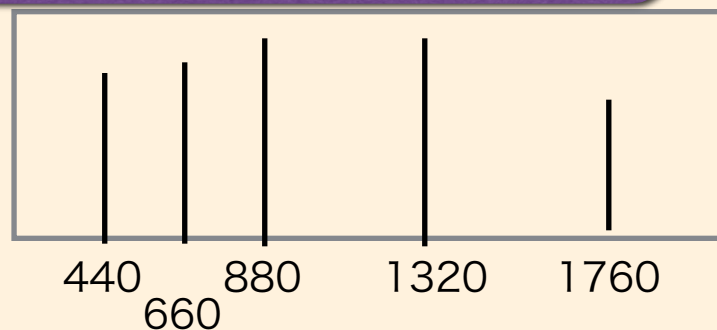
- ・ 音響信号にフーリエ変換をかけた結果を時系列で表したもの
- ・ 縦軸は周波数、横軸は時刻、色の濃淡で成分の強さ



多重音からの周波数推定の難しさ

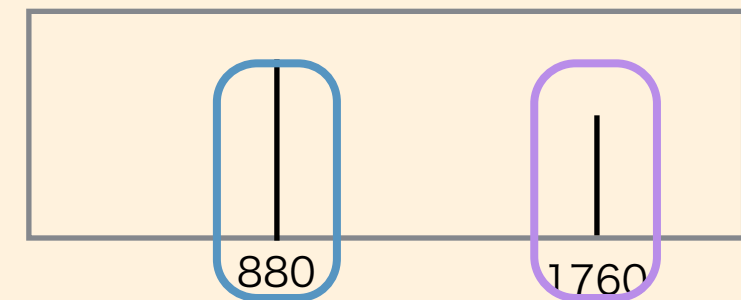
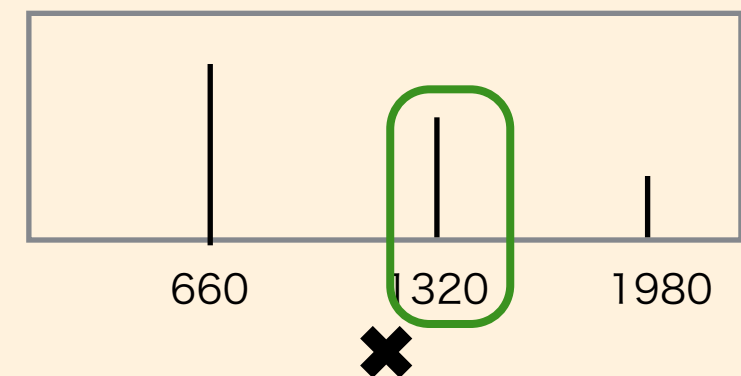
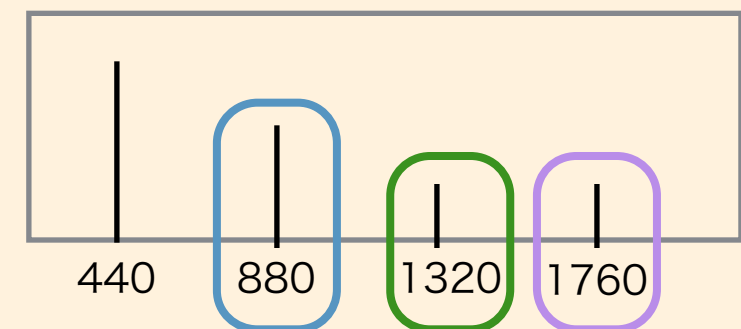


多重音のスペクトル



分解は不良設定問題

倍音により周波数成分が重なる

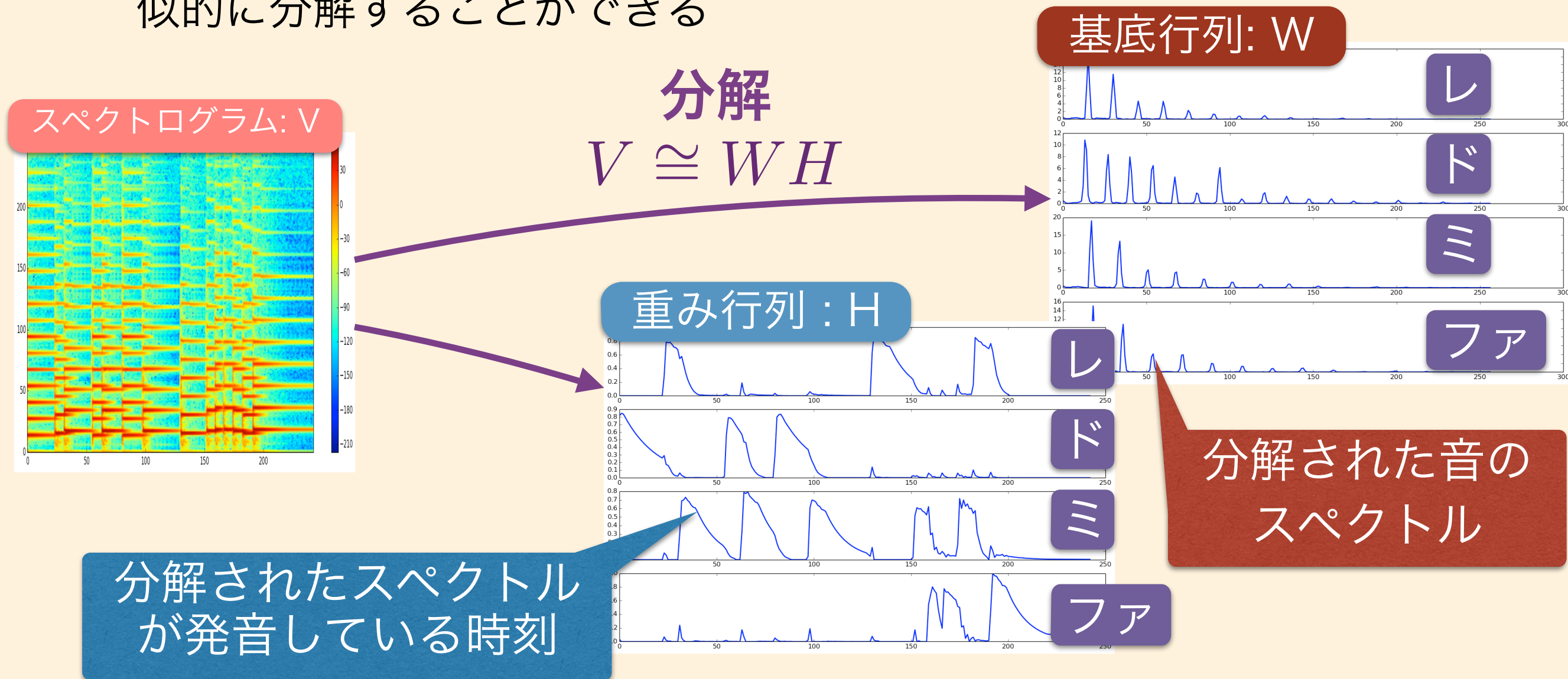


実際の楽音は、個々の音の周波数成分が複雑に重なりあい、各周波数成分がもともとどの音に由来しているのか知ることは困難

■ NMFによる音の分解

- Non-negative Matrix Factorization (NMF)

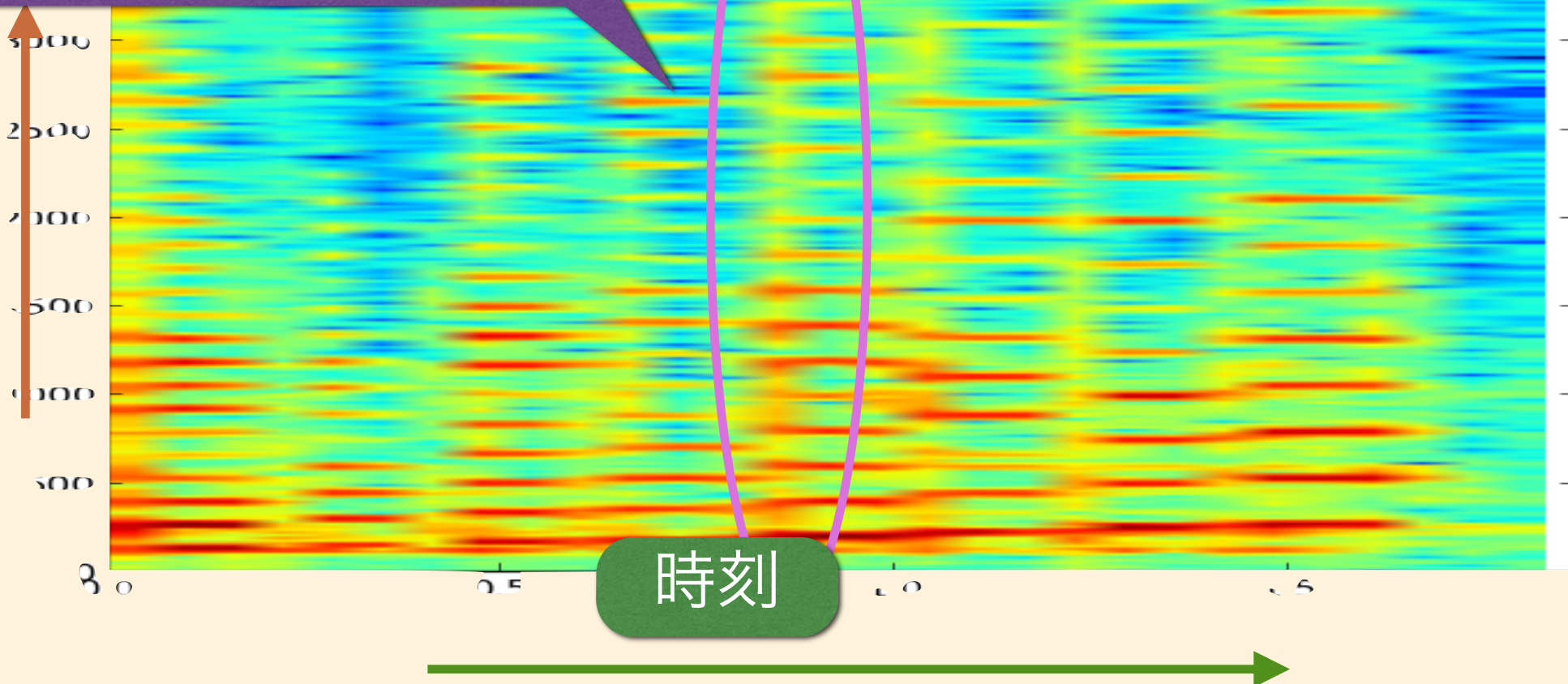
- 非負値行列因子分解のことで、音響信号のスペクトログラムに対して適用すると音高と音色のパターンである基底とそれらの重みに近似的に分解することができる



■ NMFの欠点

スペクトログラムは、
終了した演奏にフーリエ変換を
適用したもの

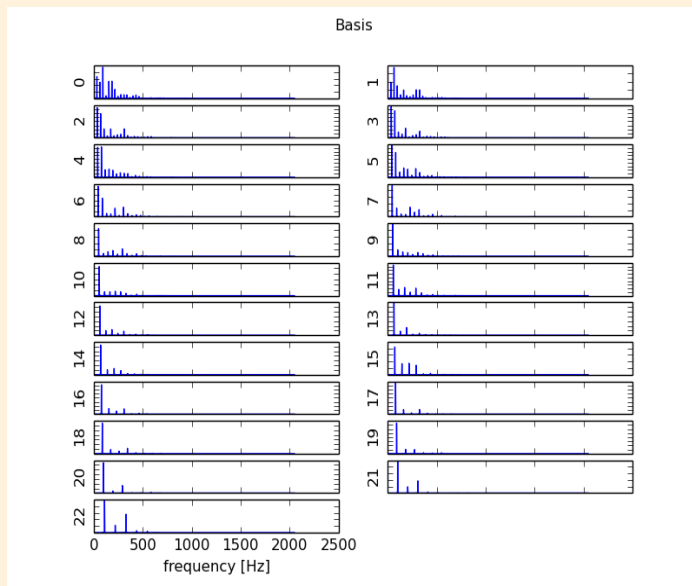
周
波
数



➡ NMFはスペクトログラムに対して適用する
ものであり、すなわち終了した演奏にのみ適用可能

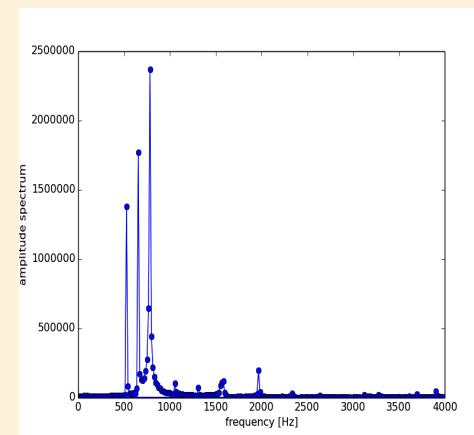
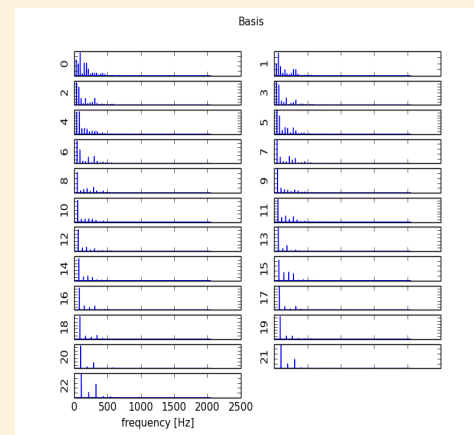
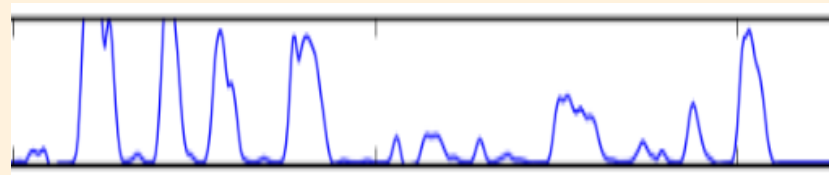
■ NMFを用いたリアルタイム採譜手法

1. 採譜のための事前準備

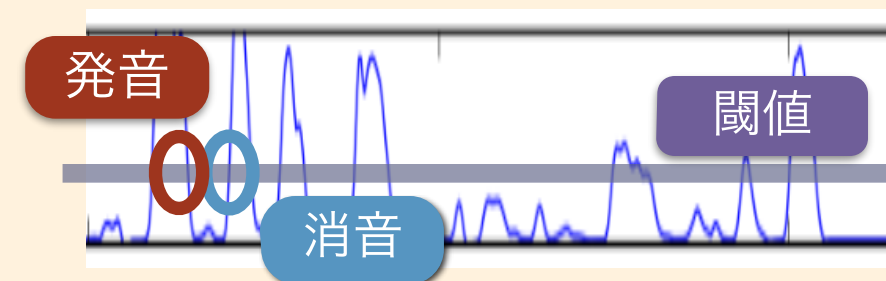


2. 重みベクトルの推定

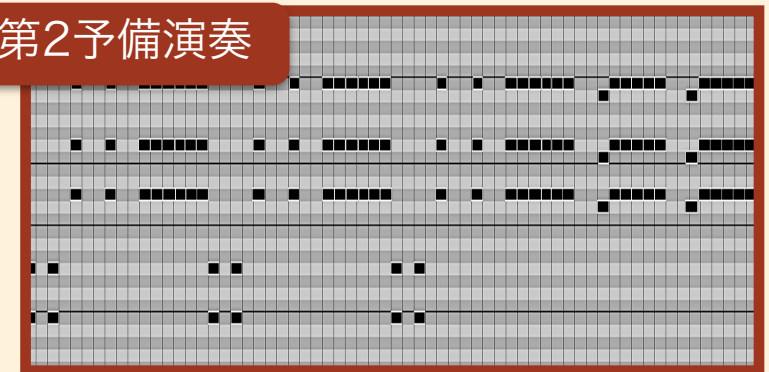
$$h_{k,t} = W_k^{-1} v_{k,t}$$



3. 発音・消音の検出



第2予備演奏



■ Step 1: 採譜のための事前準備

目的: NMFによるリアルタイム採譜を行うための準備

1. 演奏可能な全音の演奏

方法

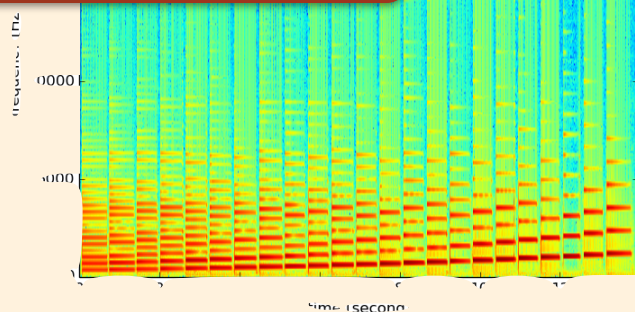
本演奏に用いるものと同じもので
演奏可能な全ての音を正確に演奏

第1予備演奏



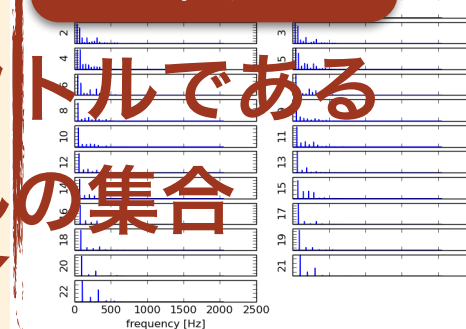
2. 基底行列の推定

第1事前演奏の
スペクトログラム

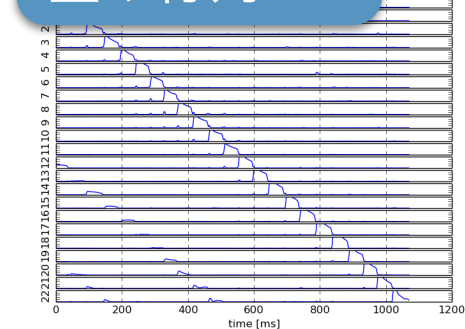


各音のスペクトルである
基底ベクトルの集合

基底行列: W



重み行列: H



$$V \cong WH$$

NMF

仮定

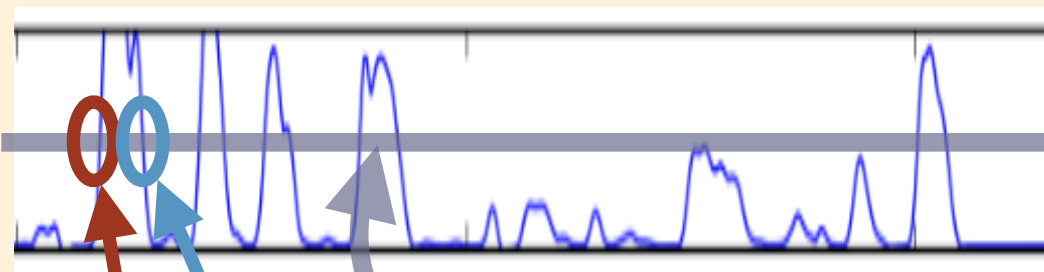
同じギターによる同じ演奏者の
演奏は似たスペクトルを持つ

■ Step 2: 重みベクトルの推定

本演奏の弦k, 時刻t における
基底ベクトル $h_{k,t}$

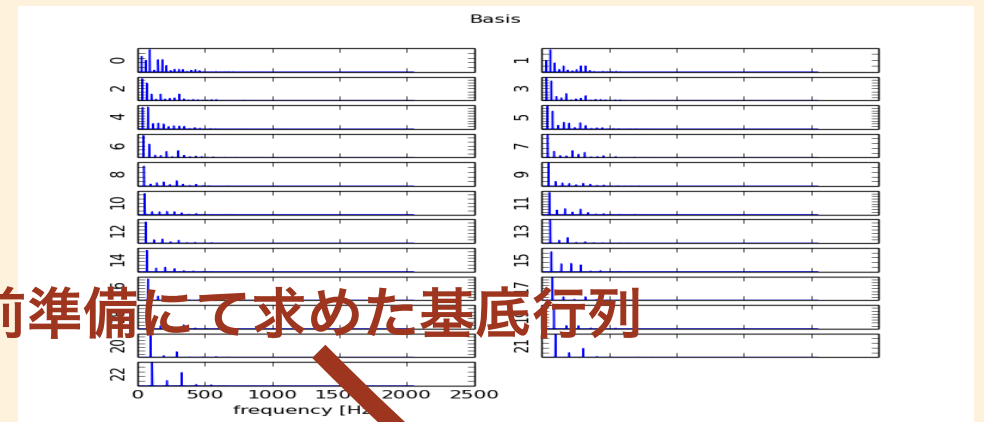
$$h_{k,t} = W_k^{-1} v_{k,t}$$

重みベクトル: $h_{k,t}$



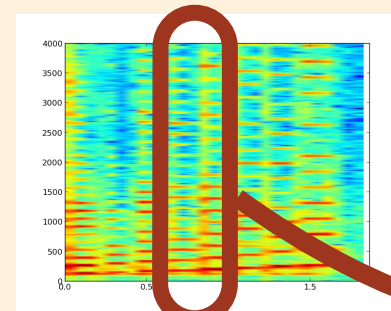
閾値

基底行列: W_k^{-1}

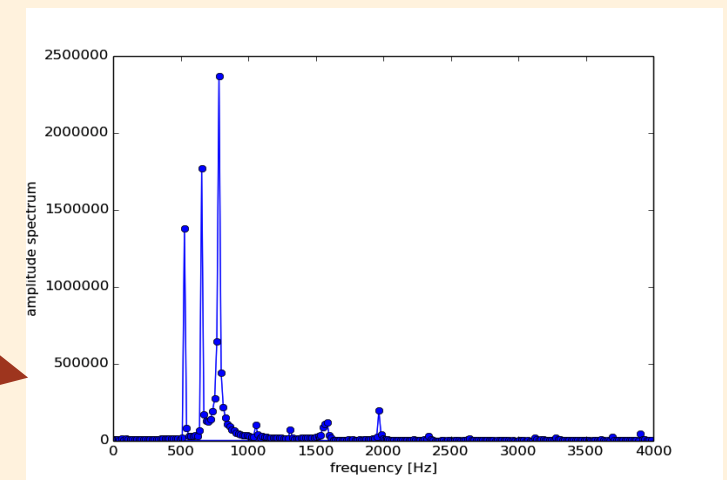


事前準備にて求めた基底行列

本演奏のスペクトル: $v_{k,t}$

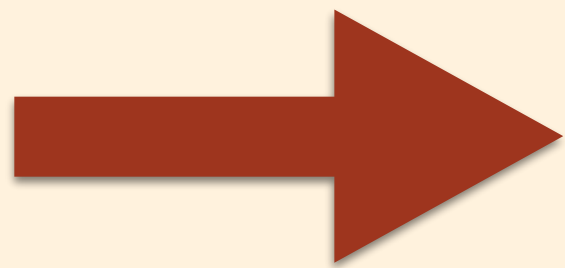


時刻: t



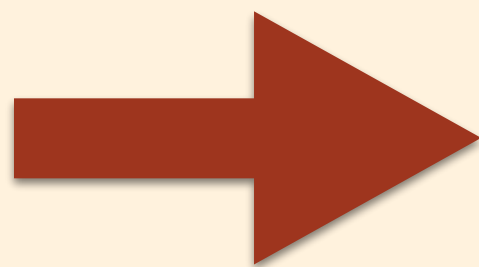
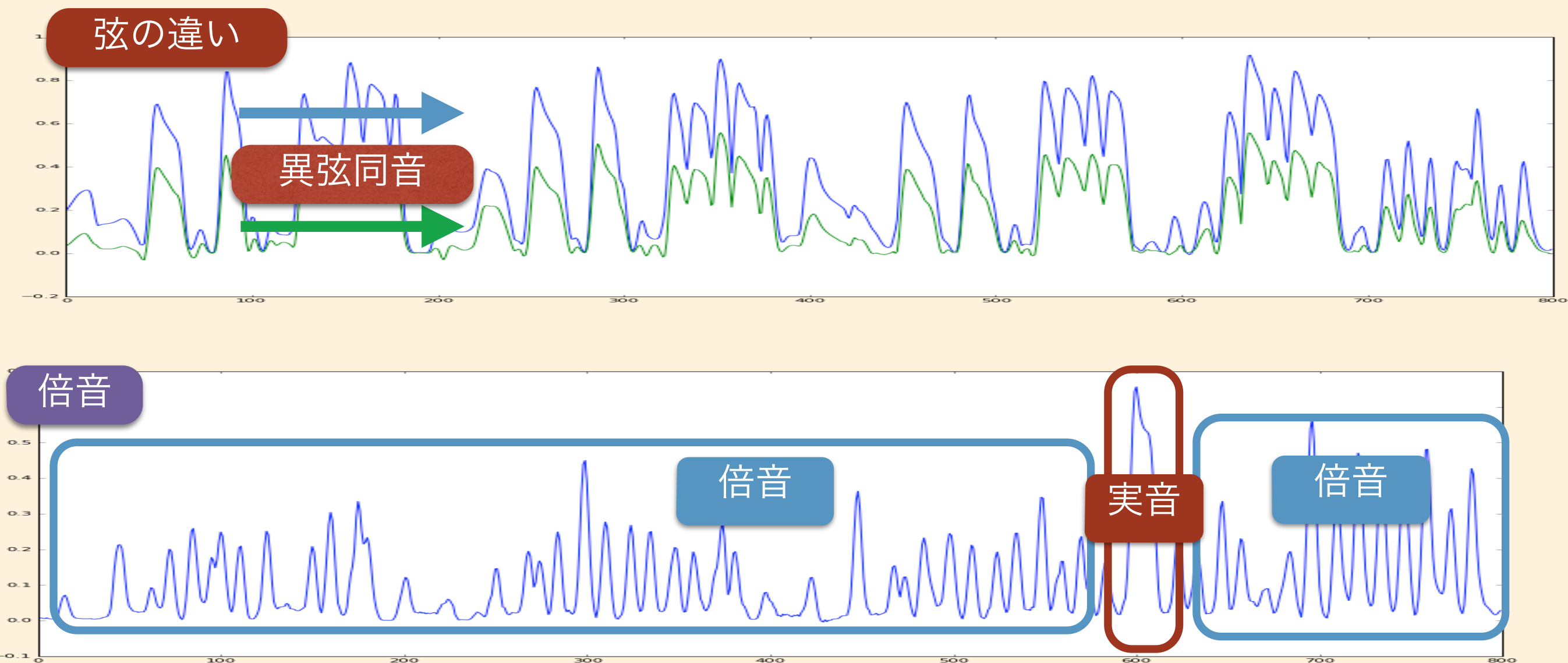
■ Step 3: 発音・消音の検出

仮に単純な閾値処理では...



発音・消音の閾値調整が鍵となる

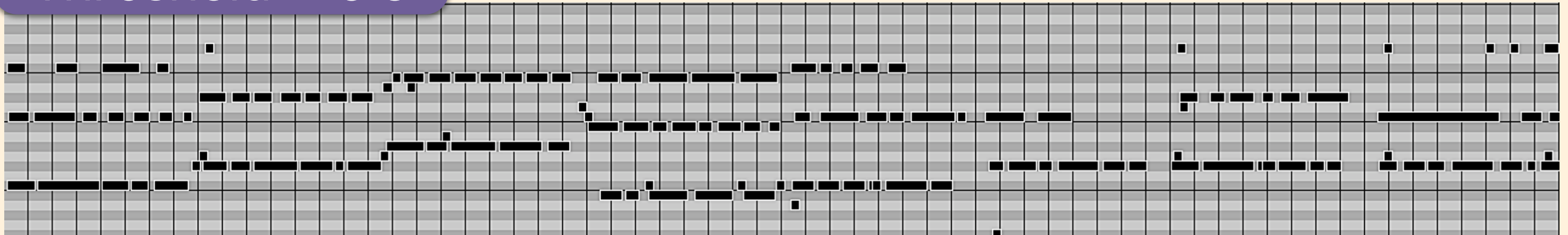
■ 閾値調整に絡む様々な要因



- ・ これらの要因を満たす閾値を一意に決定するのが難しい
- ・ 様々なパラメータに依存してしまい、汎用的でなくなる

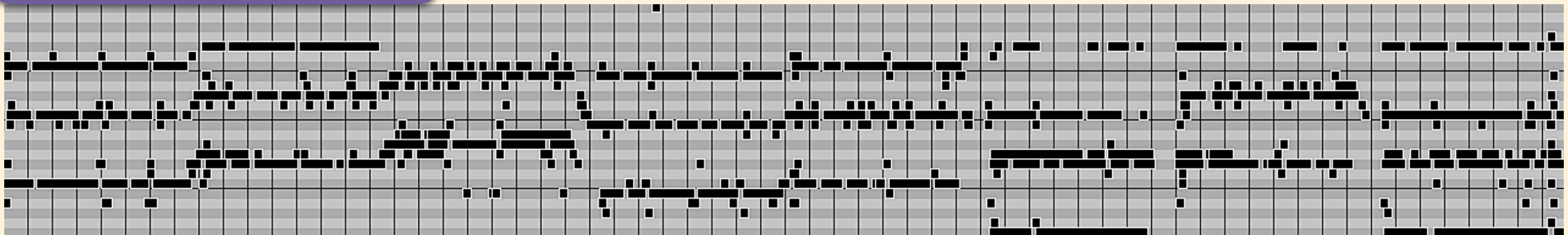
■ 閾値設定の依存例

Threshold = 0.5



小さな違いが大きな違いを生む

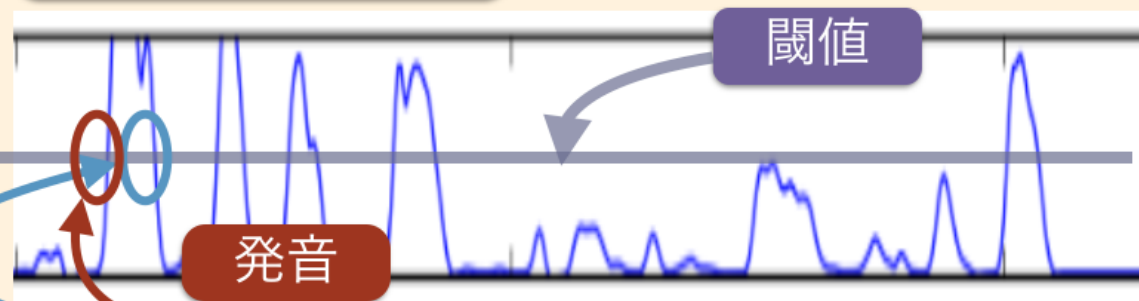
Threshold = 0.2



■ ニューラルネットワークを用いたパラメータ調整

閾値調整

重みベクトル: $h_{k,t}$



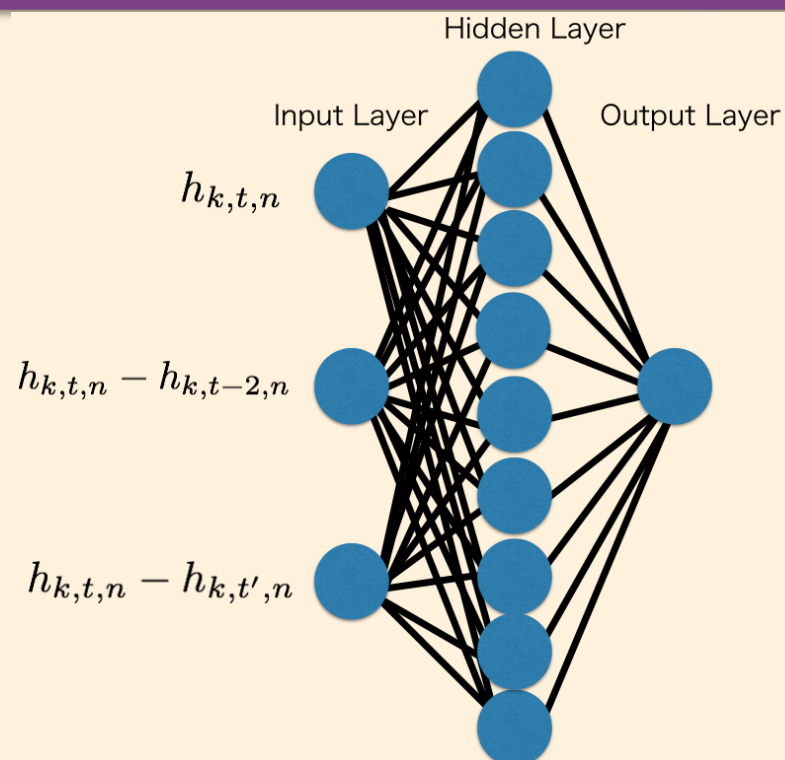
調整

消音 h_t の値が閾値を上回った時刻に、発音を検出

発音状態かつ閾値を下回った時刻に、消音を検出

ニューラルネットワーク

ニューラルネットワークで
学習することと同価



第2予備演奏

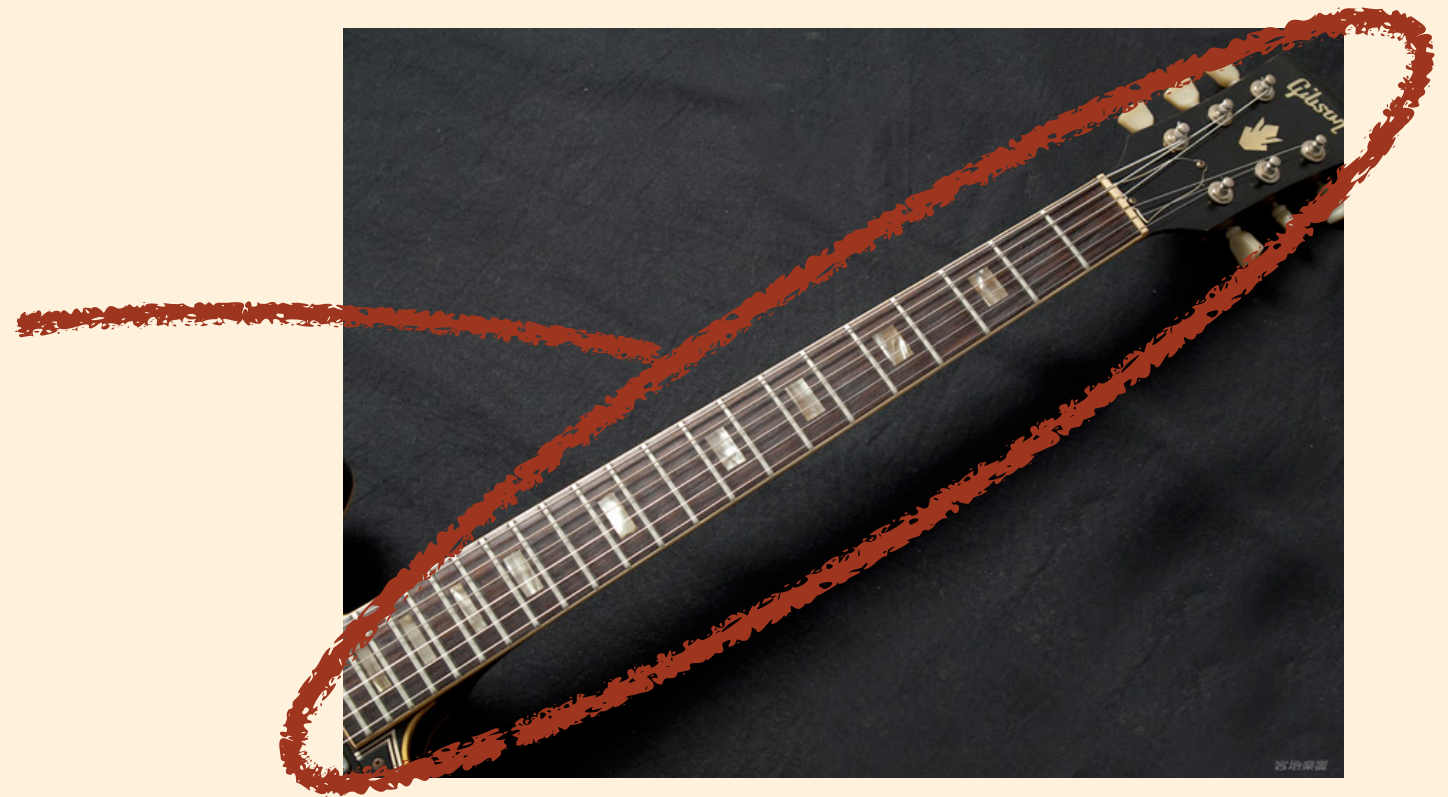
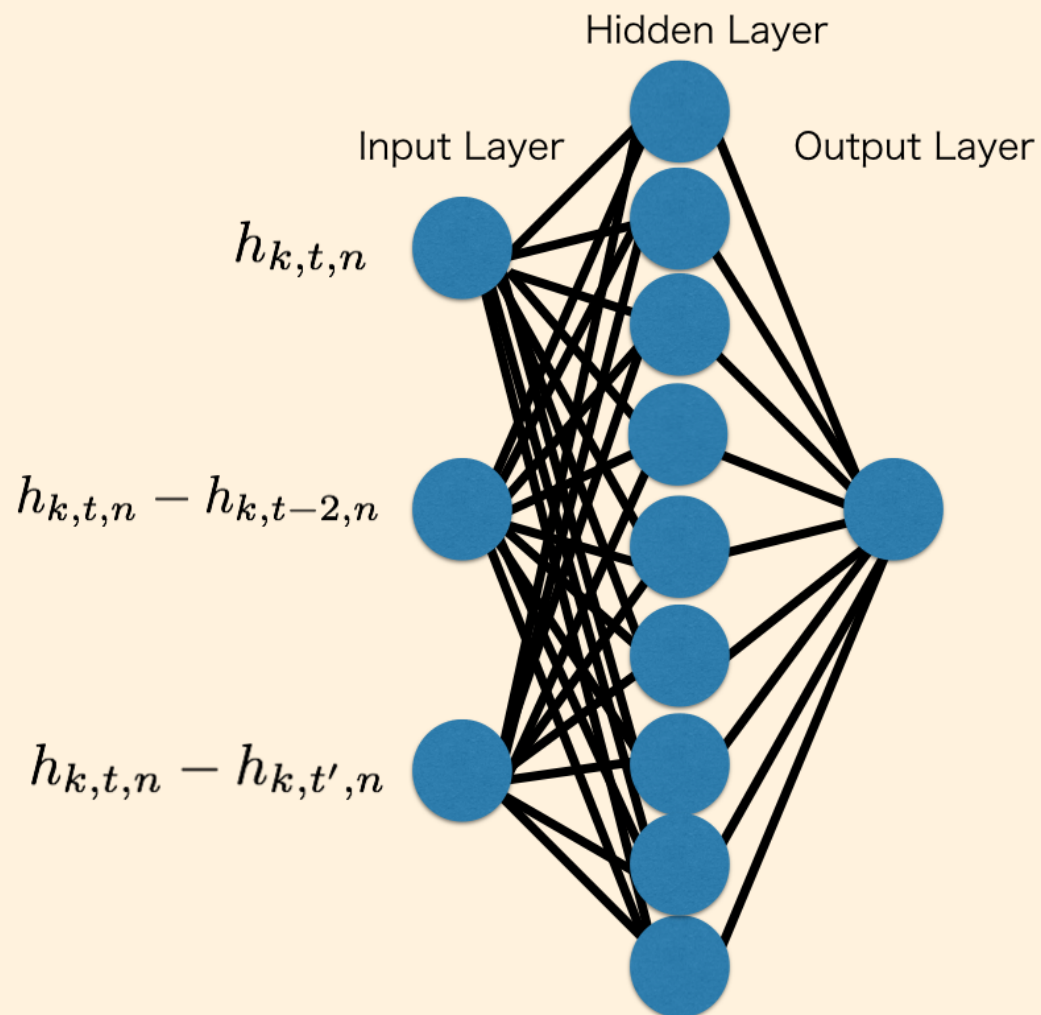


条件

- ・ 実際に採譜するフレーズに近い
- ・ 予め正解の発音時刻が把握できる

■ 弦ごとのニューラルネットワークによる学習

ニューラルネットワーク



弦ごとにニューラルネットワークを用意し学習を行う

発音学習のための入力ベクトル

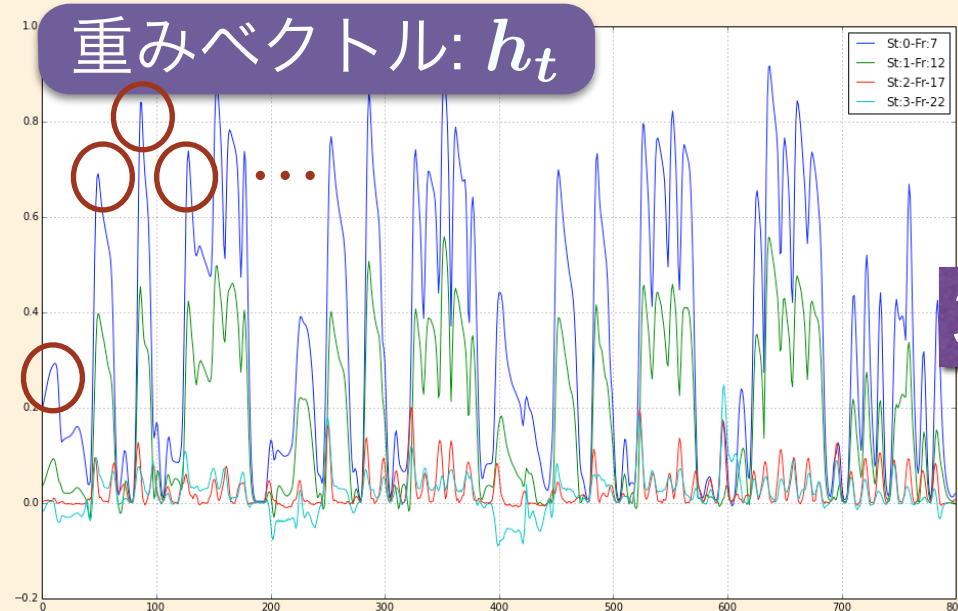
入力ベクトル: $\mathbf{x}_{k,t,n}$

弦 k の時刻 t における 各フレット n の音から

$\mathbf{x}_{k,t,n}$ を作成

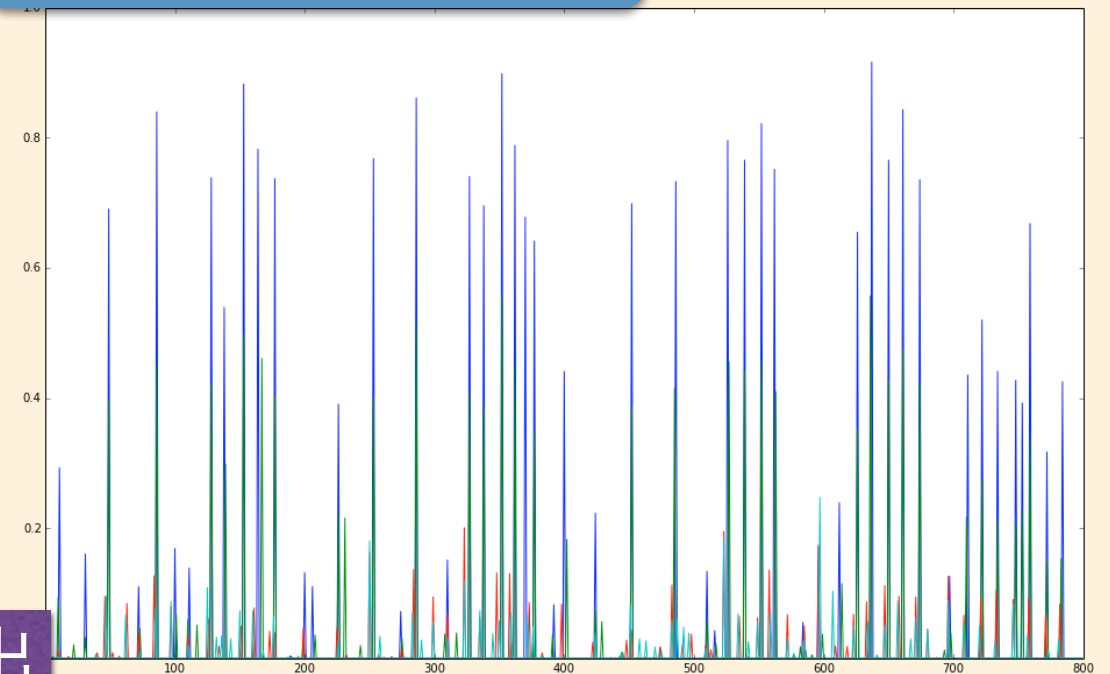
$$\mathbf{x}_{k,t,n} = (h_{k,t,n}, h_{k,t,n} - h_{k,t-2,n}, h_{k,t,n} - h_{k,t',n})$$

$h_{k,t}$ の各要素 $h_{k,t,n}$ がピークの時

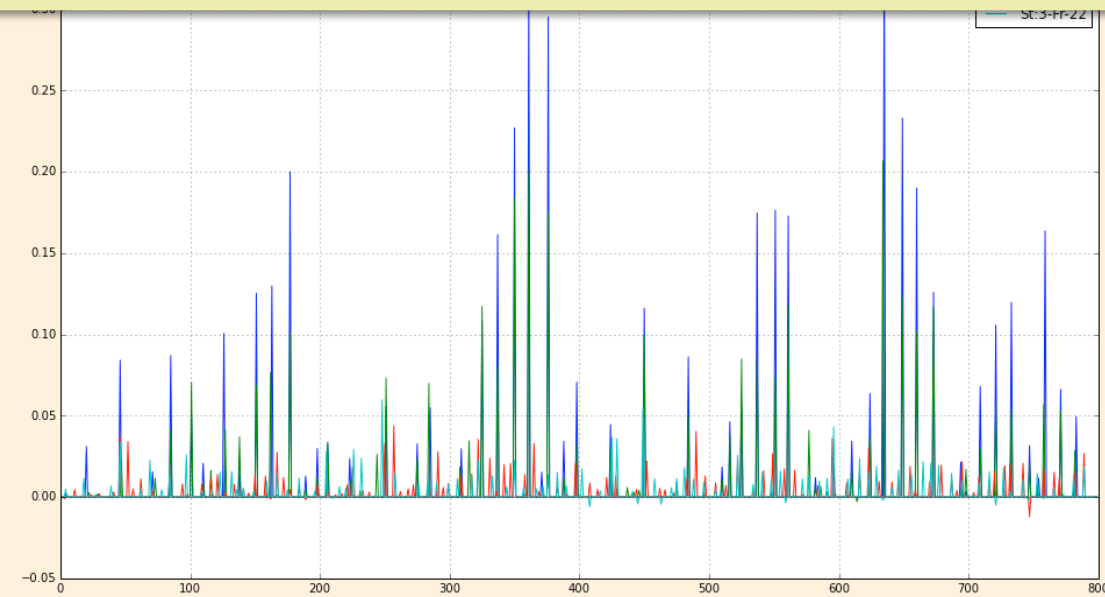


3つの特徴を抽出

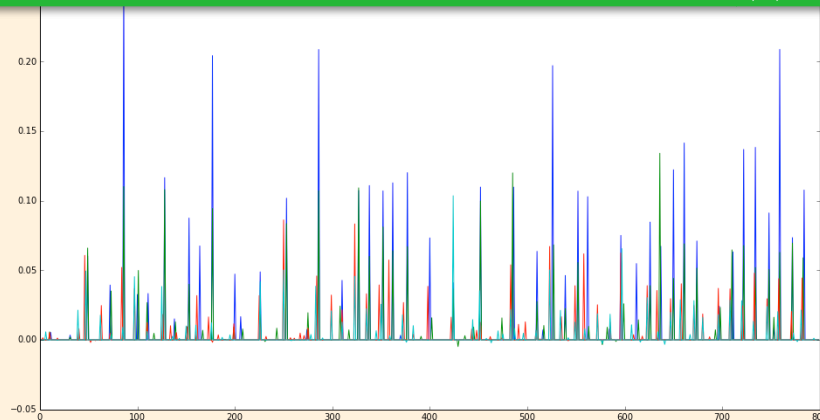
ピークの重み: $h_{k,t,n}$



ピークと2フレームとの差: $h_{k,t,n} - h_{k,t-2,n}$



ピークと直前の谷との差: $h_{k,t,n} - h_{k,t',n}$



■ 発音学習のための出力

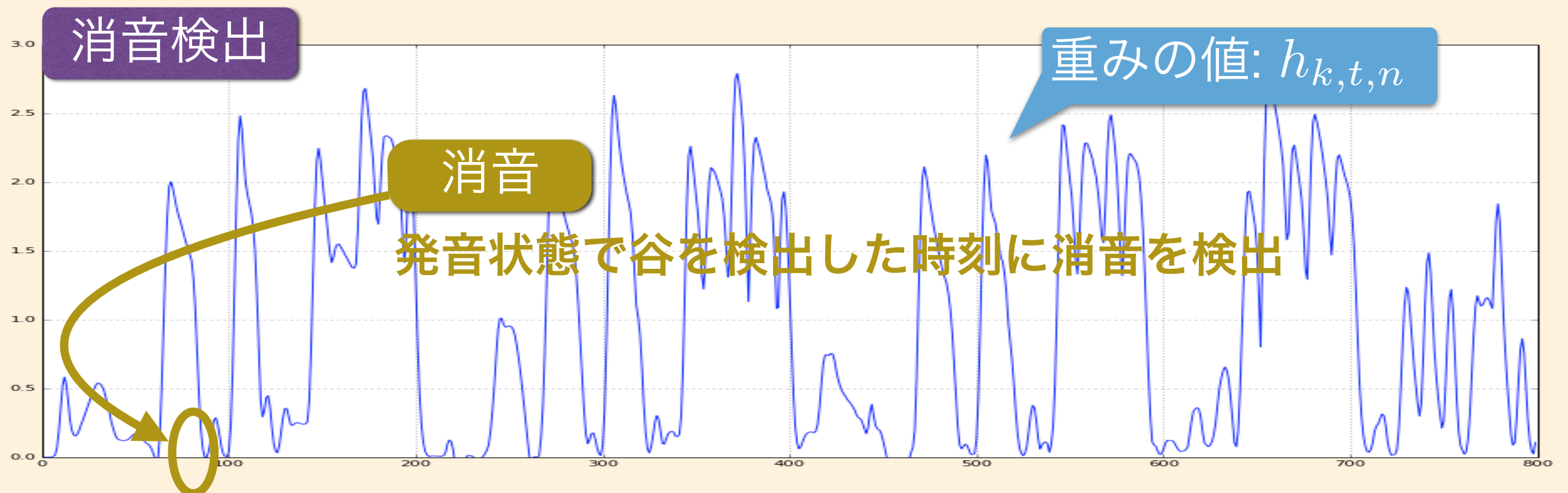
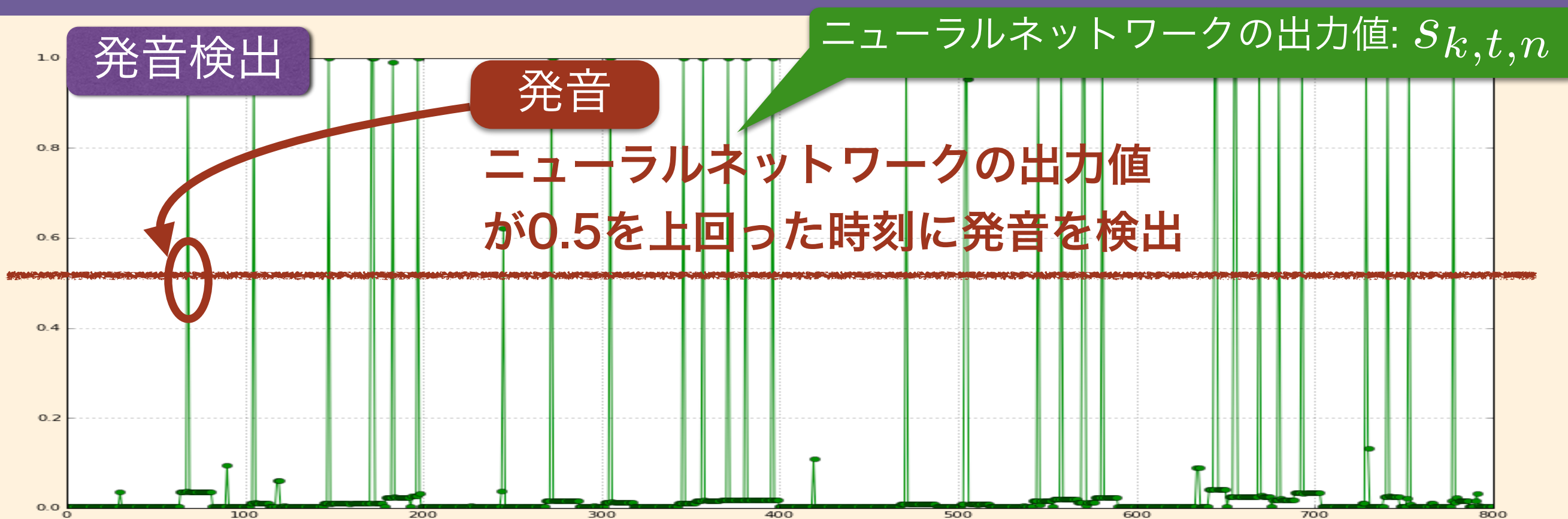
出力値: $s_{k,t,n}$



第2予備演奏に対して予め発音時刻を設定し、以下の様な値を設定：

$$s_{k,t,n} = \begin{cases} 1 & (t \text{ がフレット } n \text{ に対して発音時刻の時}) \\ 0 & (t \text{ がフレット } n \text{ に対して発音時刻でない時}) \end{cases}$$

発音・消音検出



評価実験

・ 目的

- ・ ニューラルネットワークによる発音検出手法(提案手法)の有効性の検証

・ 実験方法

- ・ 4小節のフレーズ計79個に対して各々の採譜手法を適用し、音高と発音時刻を再現率、適合率とF値を用いて評価
- ・ 単純な閾値処理によるをBaseline手法とし、閾値は 0 ~ 1.5の間で最も結果が良かったものを採用

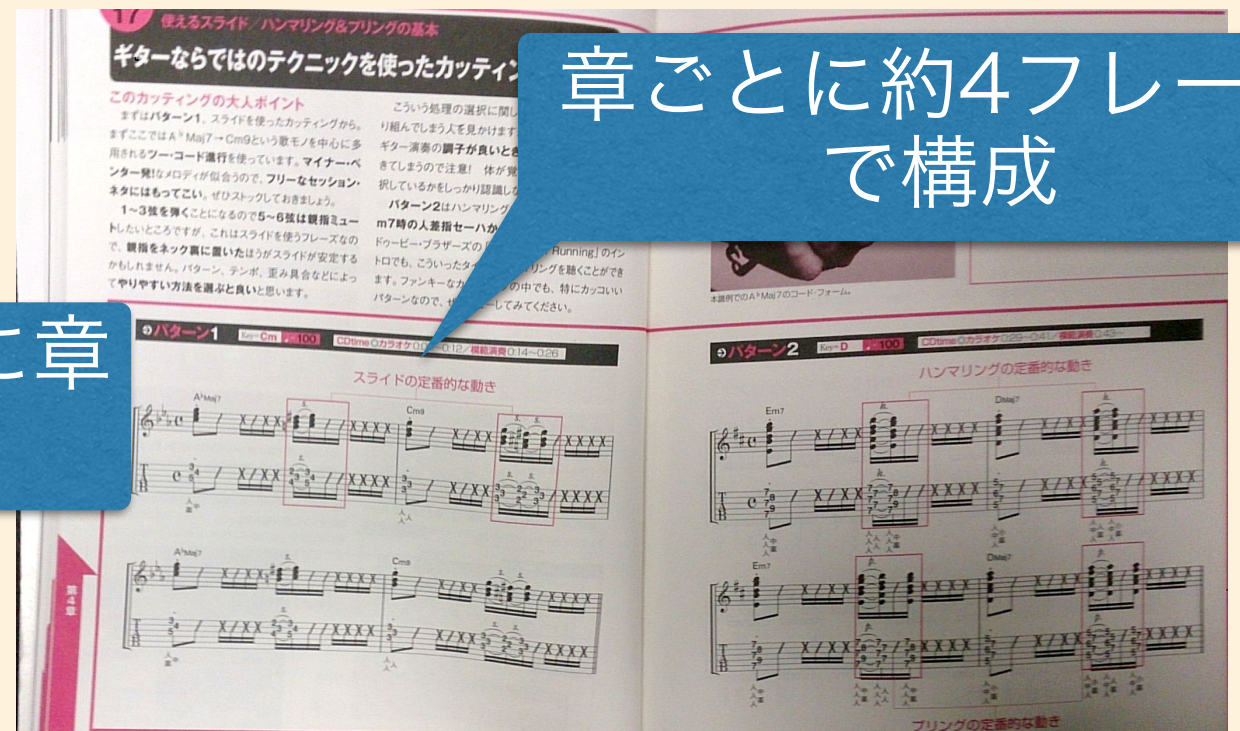
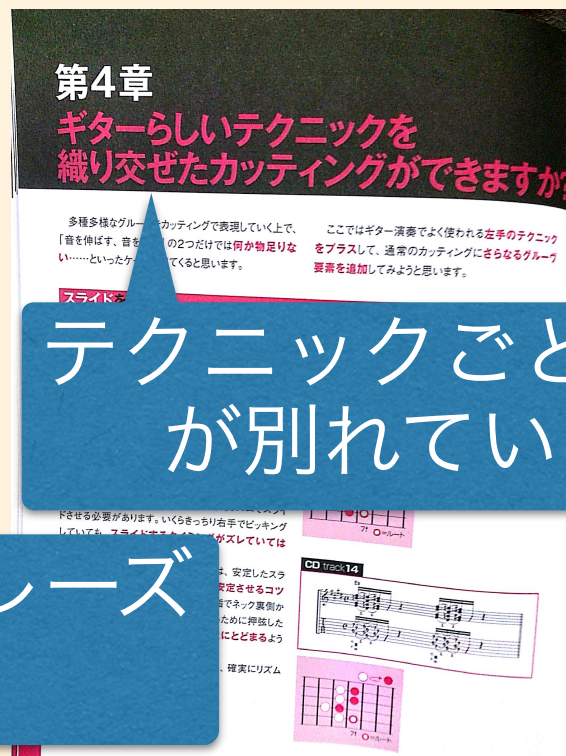
内容

第4章
ギターらしいテクニックを
織り交ぜたカッティングができますか?

テクニックごとに章
が別れている

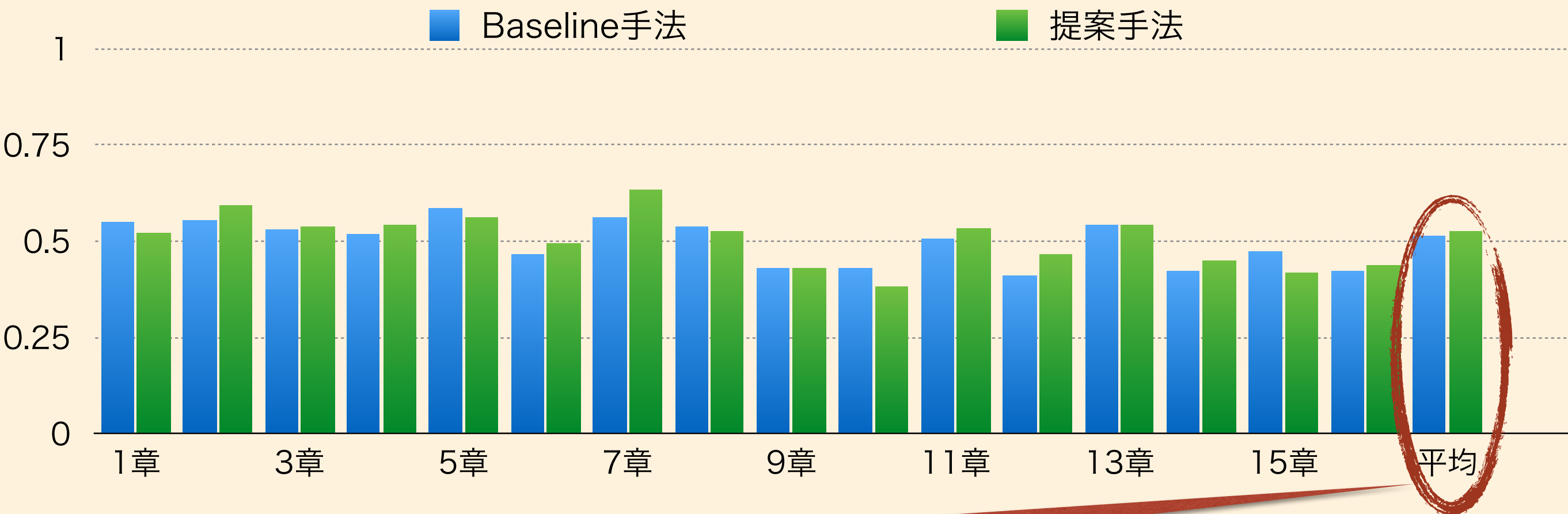
この本の全フレーズ
が対象

章ごとに約4フレーズ
で構成



実験結果

章ごとのF値の平均の結果



Baseline手法: 0.516

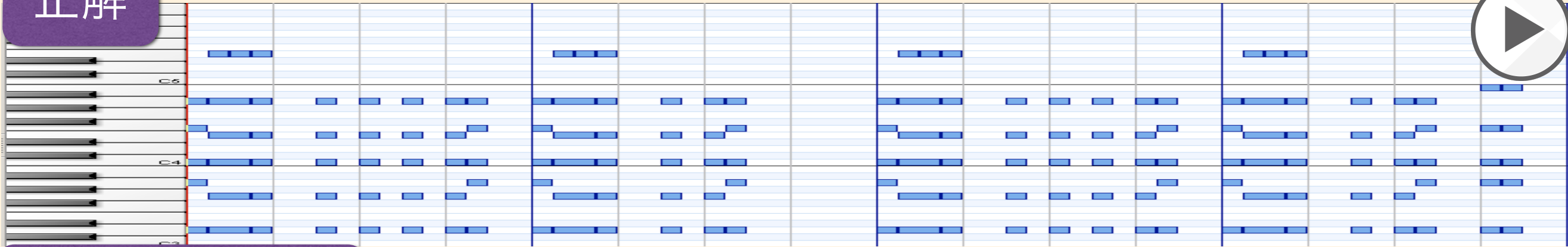
提案手法: 0.526

僅かながらに提案手法が上回る

提案手法の結果がBaseline手法における最も良かった閾値の結果を僅かに上回る

■ 平均的な採譜結果例(Track 47-2)

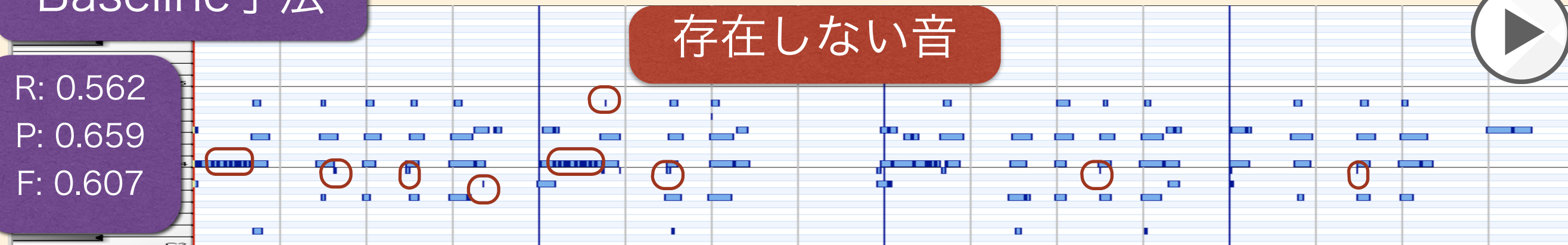
正解



Baseline手法

R: 0.562
P: 0.659
F: 0.607

存在しない音



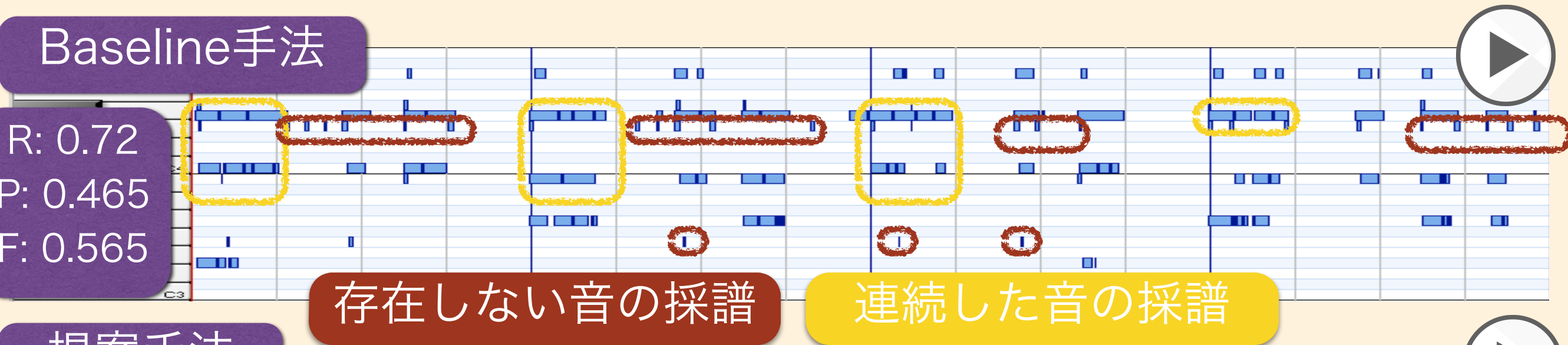
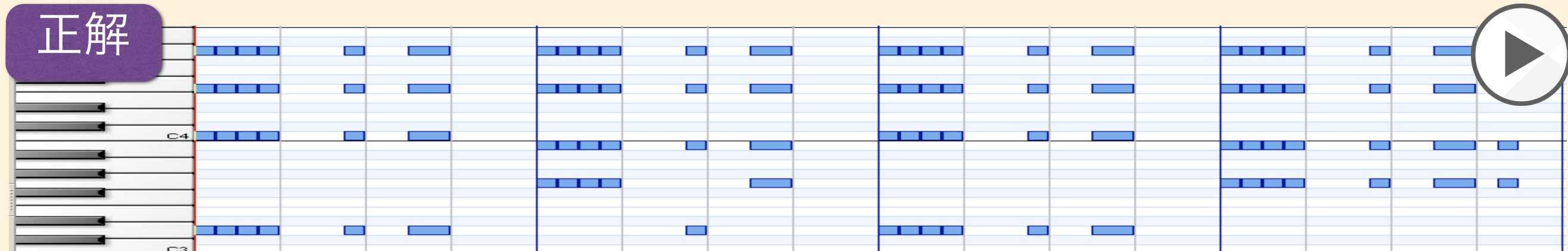
提案手法

R: 0.556
P: 0.692
F: 0.617



提案手法の方が存在しない音が減り、適合率が向上

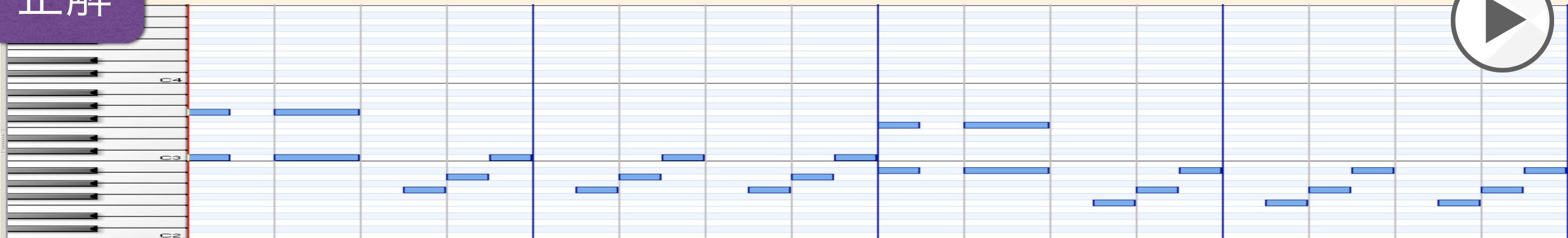
■ 良かった採譜結果例(Track 09-1)



提案手法の方が存在しない音が減り、適合率が向上

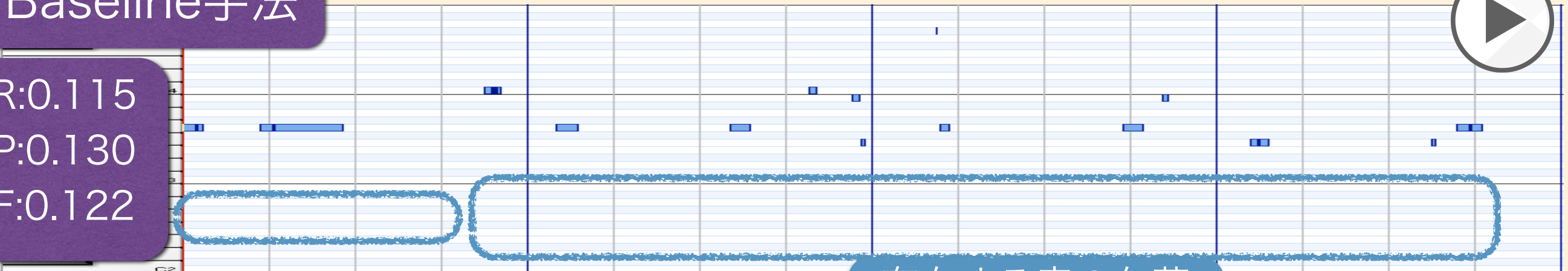
■ 悪かった採譜結果例(Track 70-2)

正解



Baseline手法

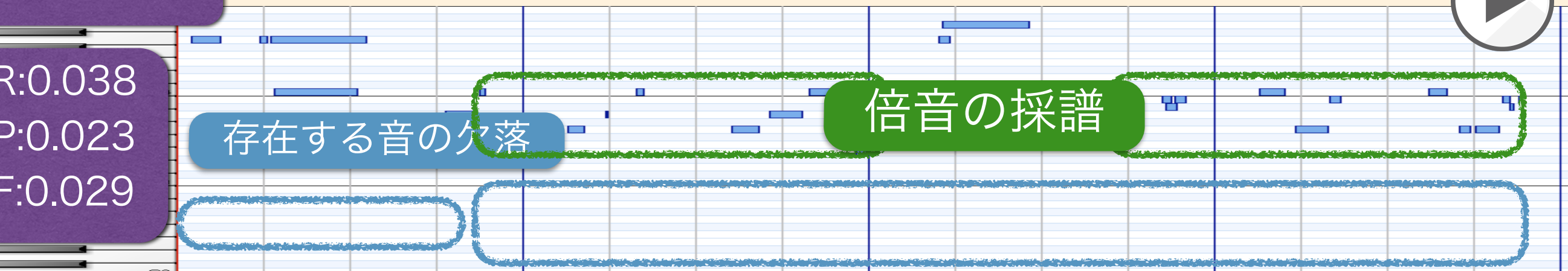
R:0.115
P:0.130
F:0.122



存在する音の欠落

提案手法

R:0.038
P:0.023
F:0.029



存在する音の欠落

倍音の採譜

低音弦の存在する音の欠落が目立ち、低精度の採譜結果になった

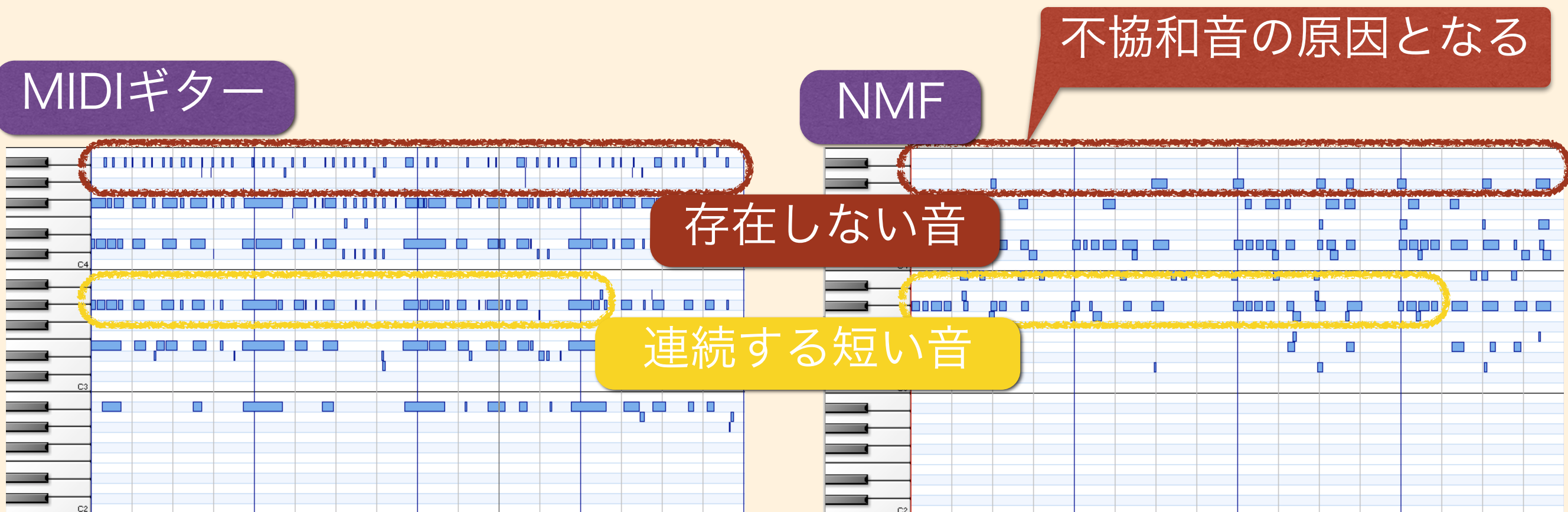
3. 統合による採譜手法

■ 解決すべきMIDIギターの問題点

MIDIギターの問題点

1. 存在する音の欠落
2. 存在しない音の採譜
3. 連続する短い音の融合

■ 誤り傾向の違い



共通した音を中心に採譜するような方針で各々の誤り傾向と長所短所を考慮した統合により問題点を解決

■ 統合の流れ

ギター演奏

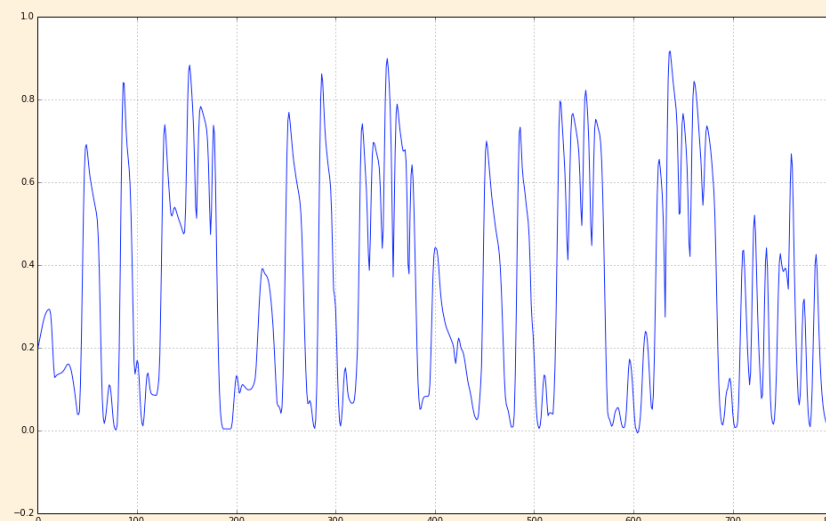


MIDIギター

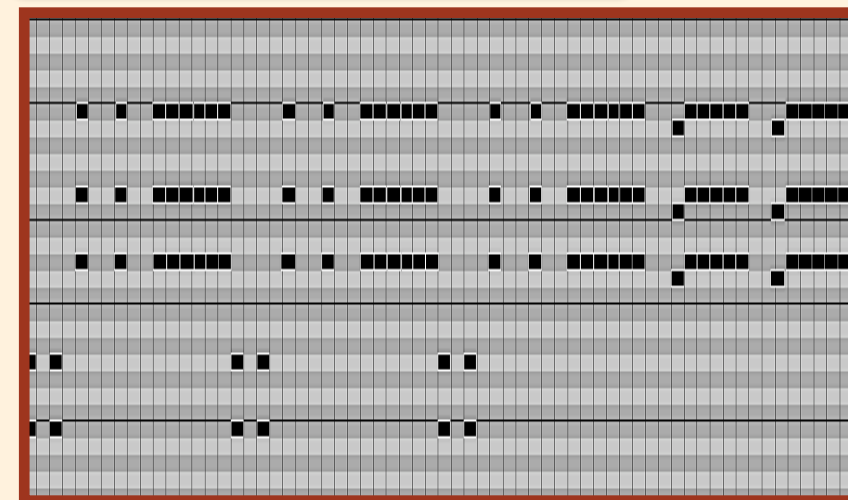


同一の演奏に対して
独立した処理

音響信号処理手法



統合による採譜



各々の処理の傾向を
考慮した統合

■ MIDIギターとNMFとの統合

3つの統合手法

時間の都合上この手法のみを紹介

1. 発音スコアの積による統合

- ・ 共通した音を残す方針で統合

2. 発音スコアの和による統合

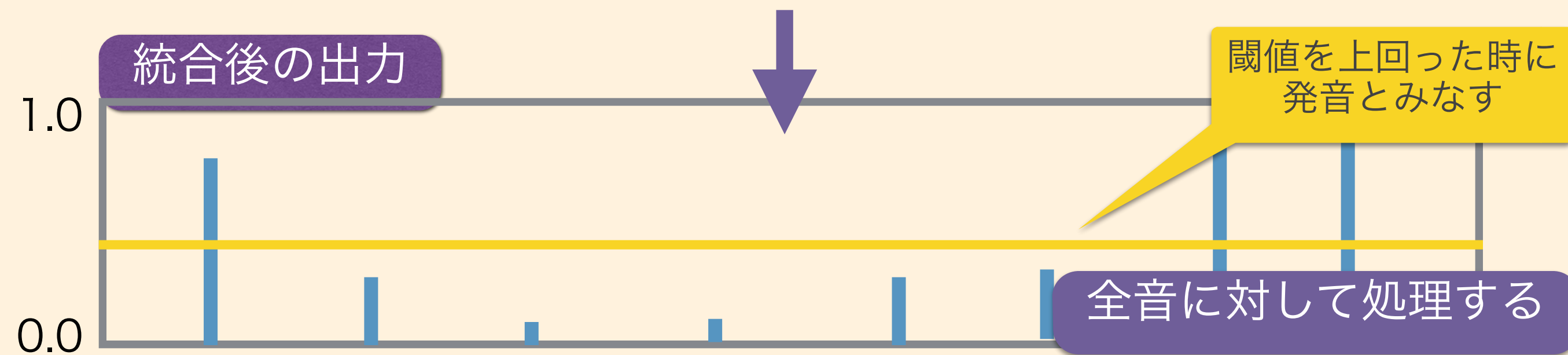
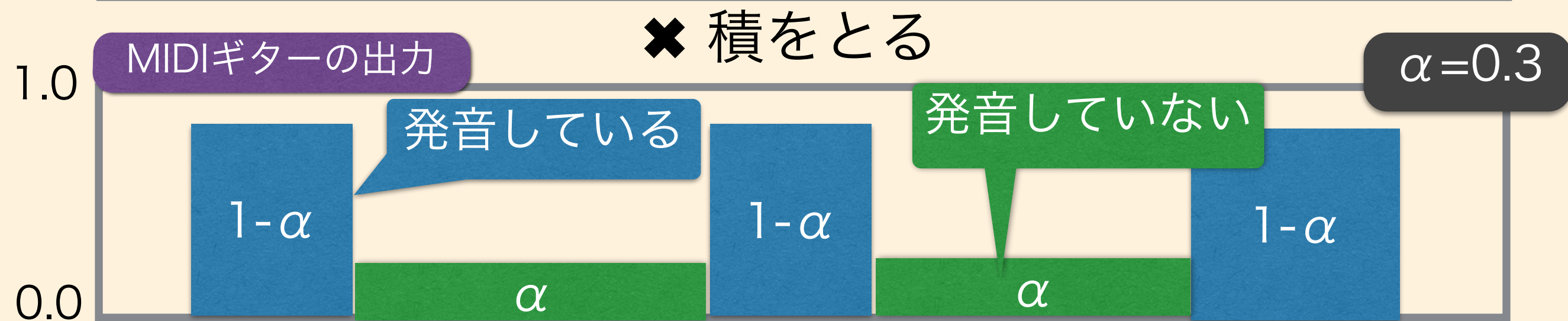
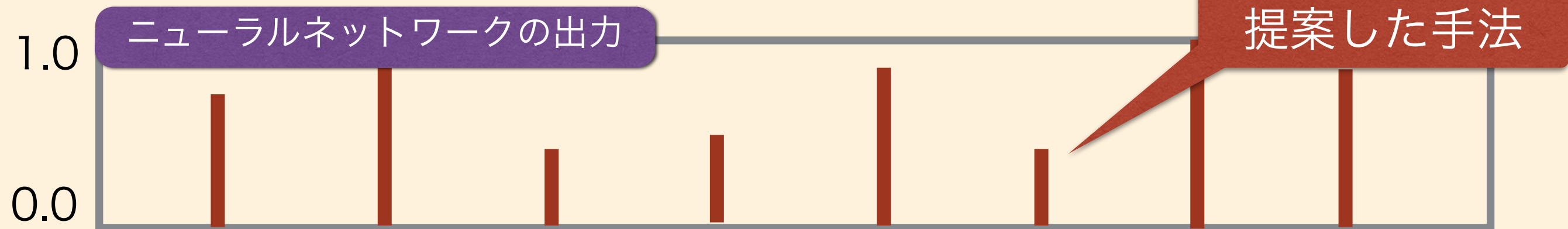
- ・ 共通した音もどちらかしか発音していない音も残す方針で統合

3. ニューラルネットワークを用いた統合

- ・ 音響信号処理で提案したニューラルネットワークにMIDIギターのノードを追加

■ 積による統合

ある音の各々採譜手法の出力



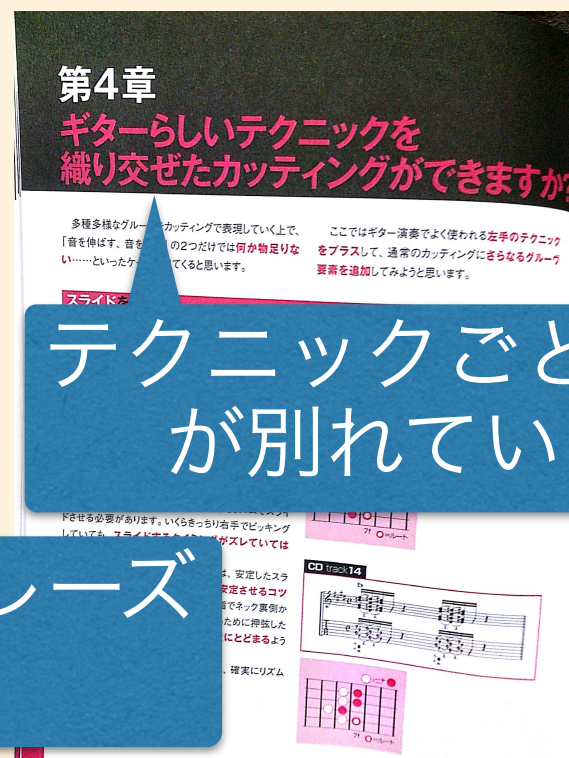
■ 評価実験

- ・ 目的

- ・ 積による統合を用いた採譜の有効性の検証

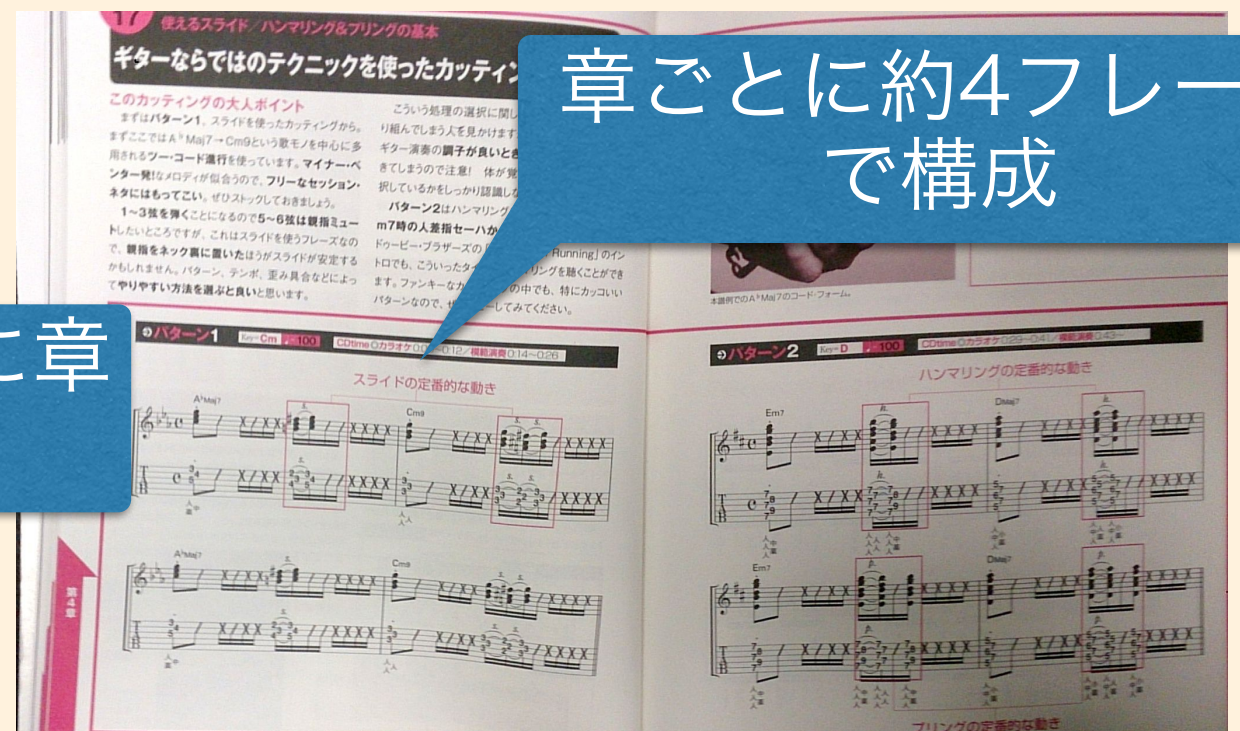
- ・ 実験方法

- ・ 4小節のフレーズ計79個に対して各々の採譜手法を適用し、音高と発音時刻を再現率、適合率とF値を用いて評価



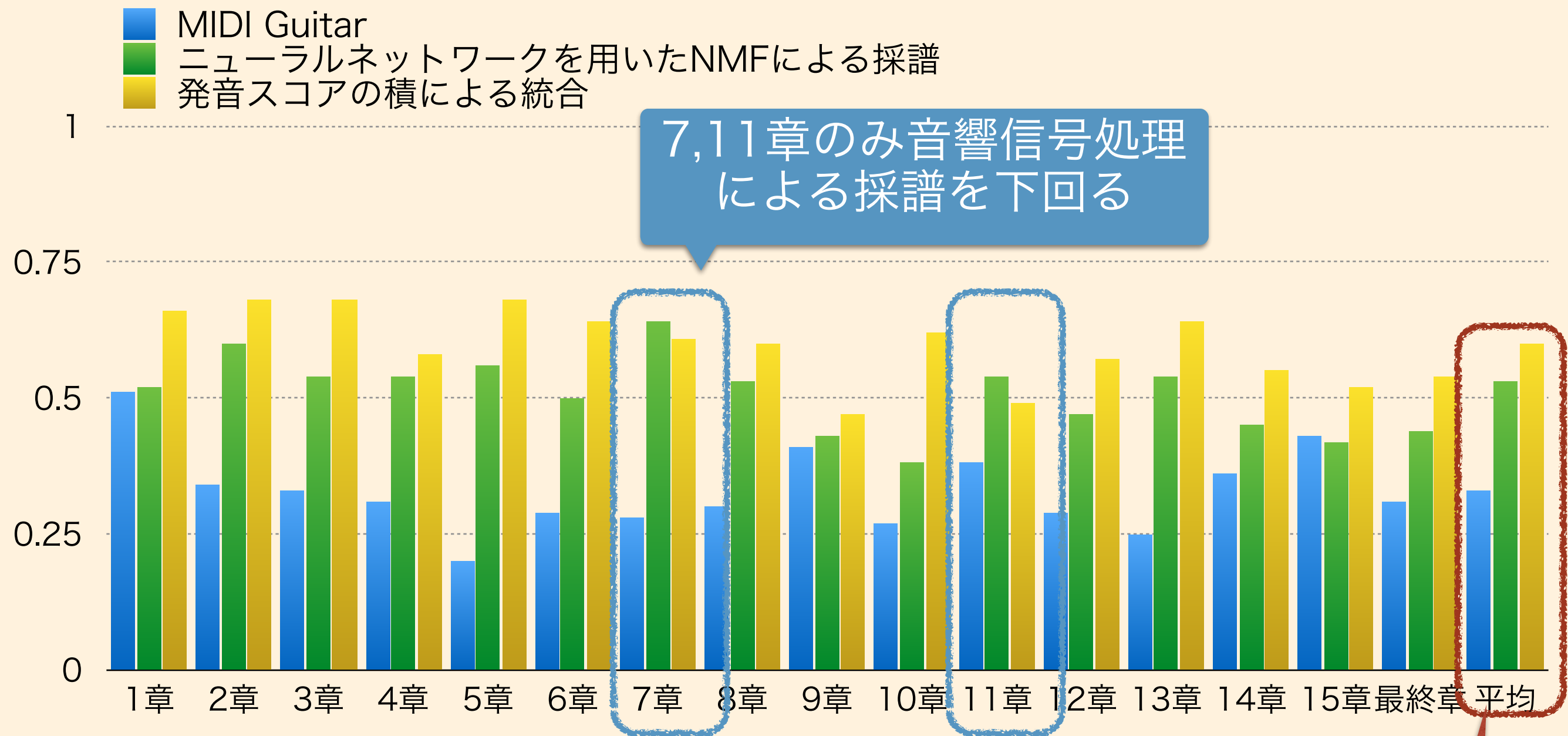
テクニックごとに章
が別れている

この本の全フレーズ
が対象



章ごとに約4フレーズ
で構成

■ 発音スコアの積による統合

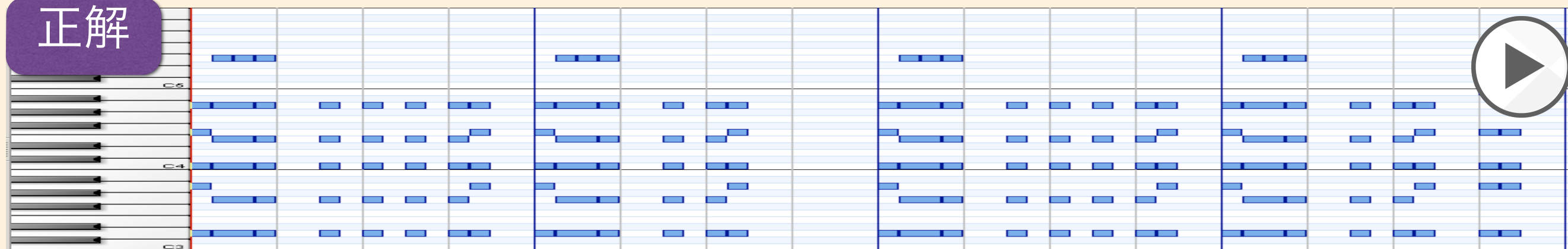


7,11章を除いて、統合による採譜が
高いF値となった

MIDIギターのF値: 0.33
統合後のF値: 0.60

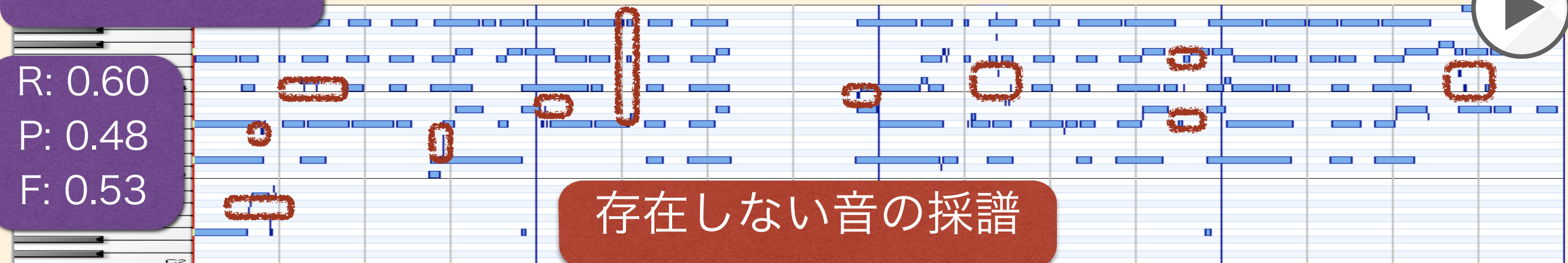
■ 平均的な採譜結果例(Track 47-2)

正解



MIDIギター

R: 0.60
P: 0.48
F: 0.53



統合による採譜

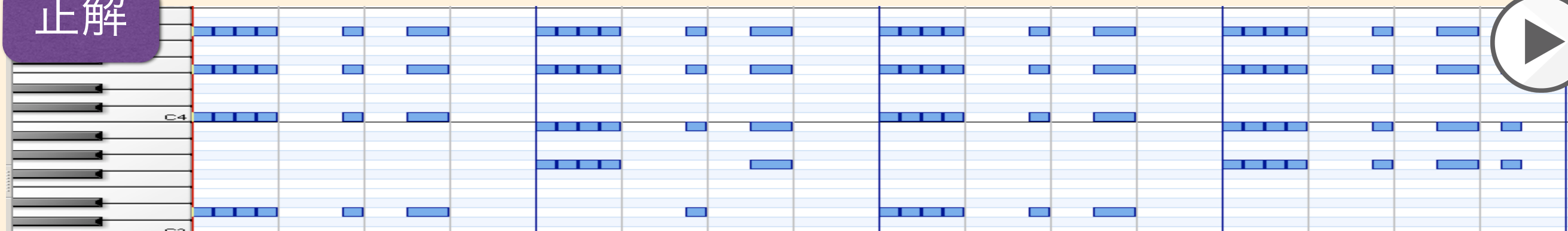
R: 0.67
P: 0.83
F: 0.74



提案手法は、音の欠落により再現率は低下したが、適合率は向上した

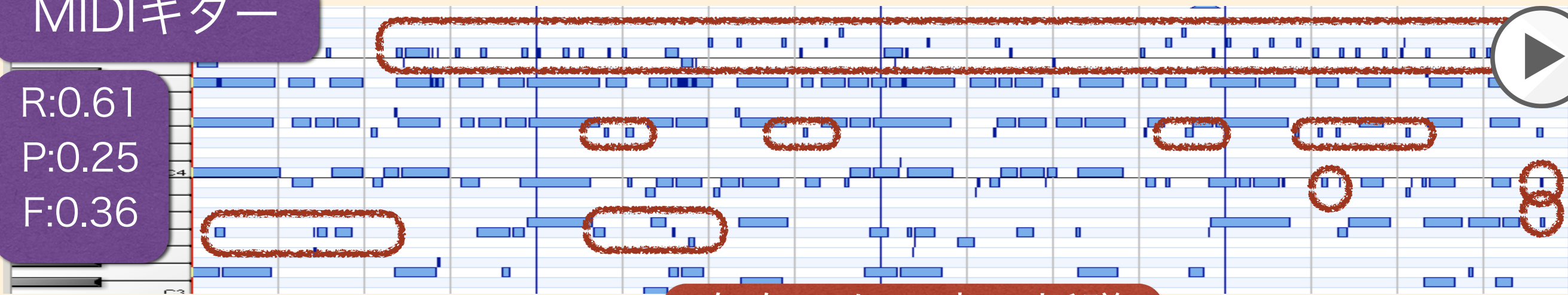
■ 良い採譜結果例(Track 09-2)

正解



MIDIギター

R:0.61
P:0.25
F:0.36



存在しない音の採譜

統合による採譜

R:0.80
P:0.78
F:0.79

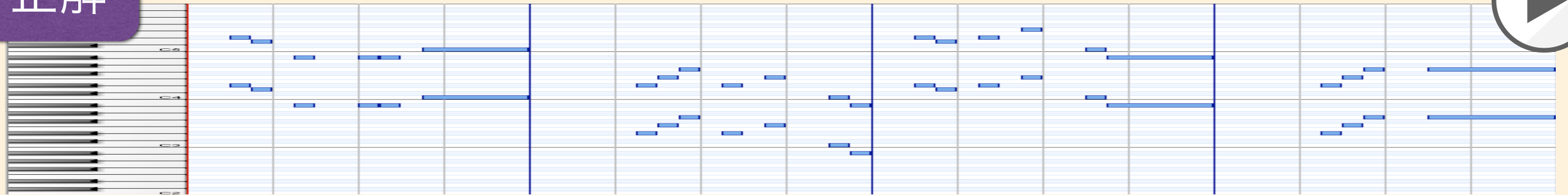


連続する音の採譜

提案手法の方が存在しない音が削減され、適合率が向上

■ 悪かった採譜結果例(Track 69-2)

正解



MIDIギター

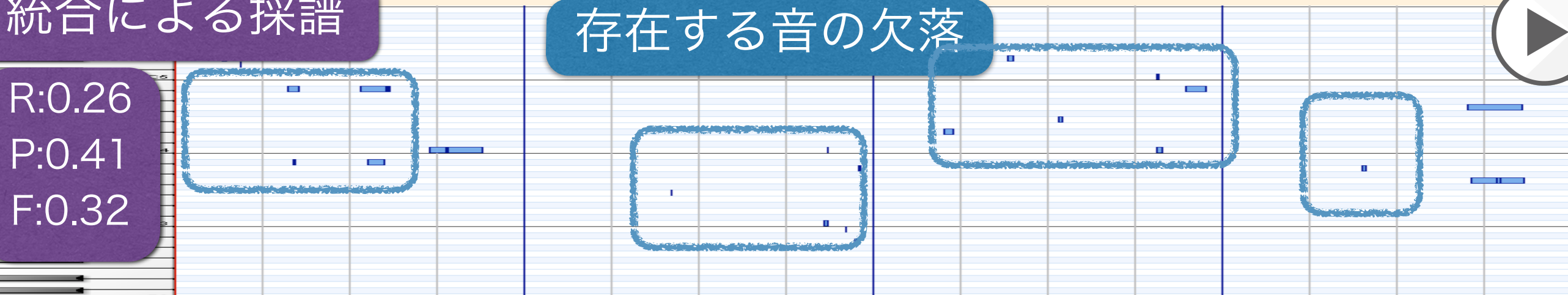
R:0.78
P:0.26
F:0.39



統合による採譜

R:0.26
P:0.41
F:0.32

存在する音の欠落



存在する音が大幅に欠落してしまったため、再現率が大幅に低下

■ まとめ

NMFによる音響信号処理

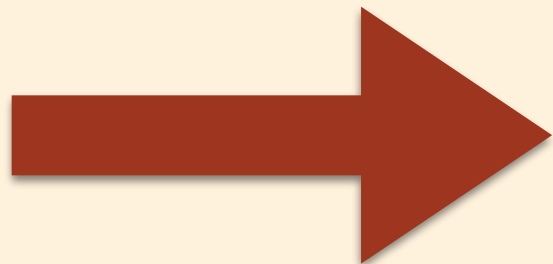
- ・ NMFを用いたオンラインアルゴリズムによるギター演奏の採譜手法を提案し、発音検出における閾値の調整の問題に取り組んだ



問題を解決し、F値の平均を0.526と示した

統合による採譜

- ・ MIDIギターと音響信号処理による採譜を統合することによって、MIDIギターの3つの問題点を解決した



3つ問題点を解決し、MIDIギターのF値: 0.33
に対し、統合結果はF値: 0.60となった

■ 今後の課題

リアルタイムで動作するシステムへの実装

- ・ 統合による採譜のアルゴリズムをリアルタイムで動作するシステムに実装する

効果的な第2事前演奏の条件の選定

- ・ 最も効果的な事前演奏の条件が不明なため、様々なフレーズによる学習を試し条件を求める

音響信号処理に用いるNMFの変更

- ・ 最も基本的なNMFを用いたが、音響信号処理に適したNMFが研究されているため、それらを用いて採譜を行う

■ 業績一覧

- ・ 電子情報通信学会 総合大会 (2014-03)
- ・ 音学シンポジウム2014 (2014-05)
- ・ The 12th IEEE International Conference on Signal Processing (2014-10)
- ・ The 15th International Society for Music Information Retrieval Conference (2014-10)
- ・ 情報処理学会 全国大会 (2015-3) (予定)