

特徴量について

Deep-people #3

- 音響データ
 - 音は波で表現できる
 - 音はフーリエ変換により、異なる周波数を持つ複数の波に分解できる
 - 音の時間的变化を捉えるために、短時間フーリエ変換を用いてスペクトログラムを作ることができる
- 楽譜データ
 - 楽譜は音楽演奏を意味的に解釈して符号化したもの
 - 様々なファイル形式（MIDI, MusicXML, ABC）, 表現形式（ピアノロール, MIDIトークン etc）が存在

- 深層学習の前段階として、「深層じゃない」学習を知ろう
 - $y = f(X)$ の X を f に合わせた形で作る方法
 - 特に、音響データに関する方法を中心に解説（音響だと特徴抽出は避けられないため）
- 機械学習の前処理となる特徴量について
- 特徴量

今日の内容

- 多くをこちらからとっています

[https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/04-bookFMP/02-slides/
Mueller_FMP_Chapter1.pdf](https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/04-bookFMP/02-slides/Mueller_FMP_Chapter1.pdf)

[https://www.microsoft.com/en-us/research/uploads/prod/2021/10/Tutorial-on-AI-Music-Composition-
@ACM-MM-2021.pdf](https://www.microsoft.com/en-us/research/uploads/prod/2021/10/Tutorial-on-AI-Music-Composition-@ACM-MM-2021.pdf)

[https://mac.kaist.ac.kr/~juhan/gct634/Slides\[week1-3\]%20audio%20data%20representations.pdf](https://mac.kaist.ac.kr/~juhan/gct634/Slides[week1-3]%20audio%20data%20representations.pdf)

- 特に引用なく載せた図はこちらのいずれかからピックアップしています

特徴量とは（一般）

5

対象に関する特徴を数値化したもの

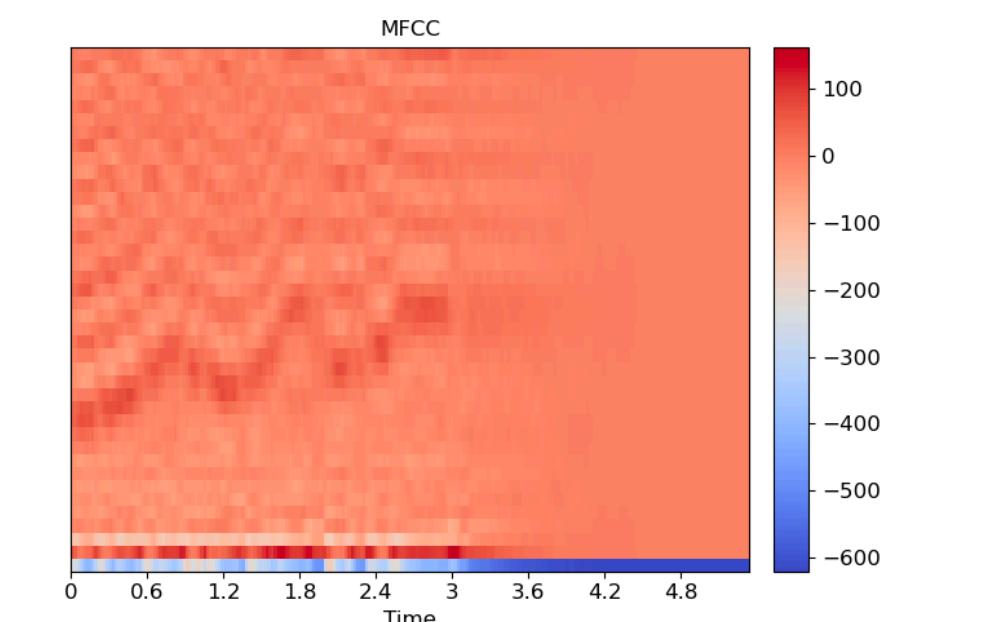
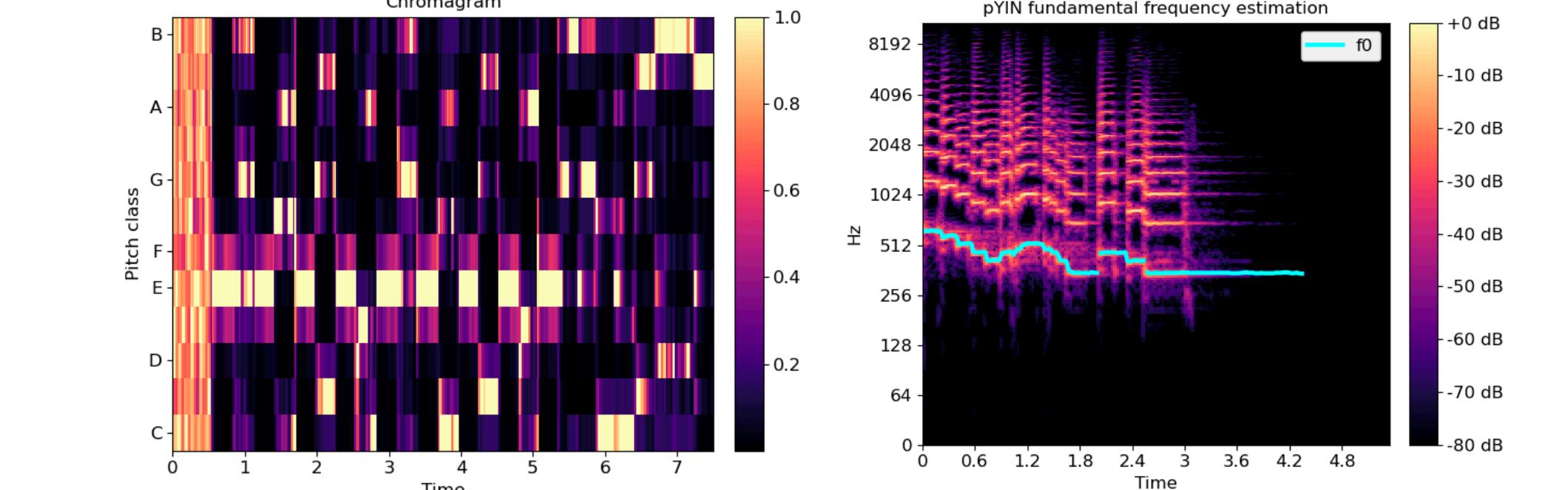
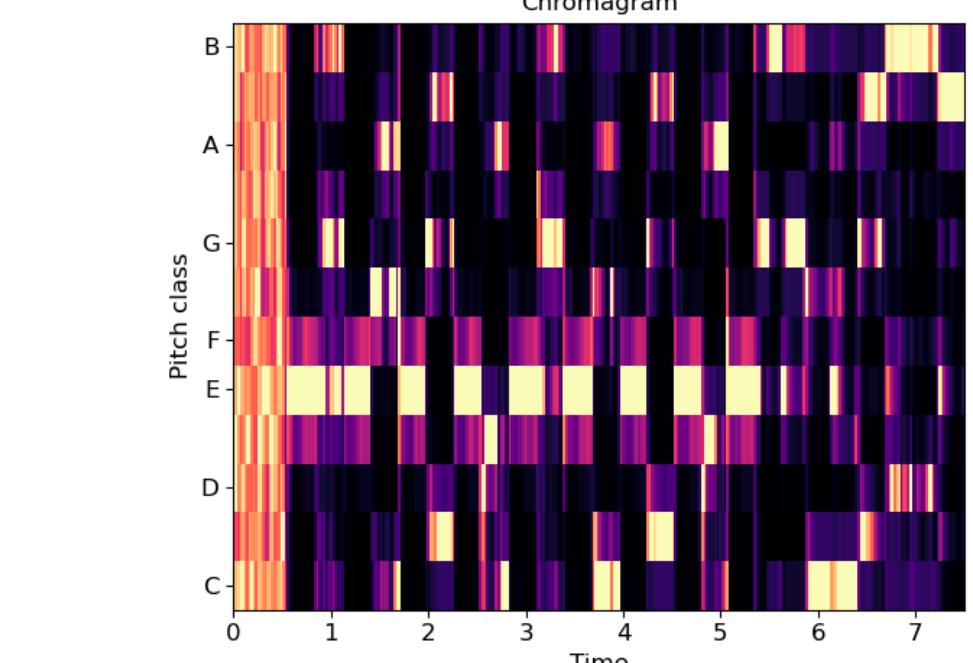
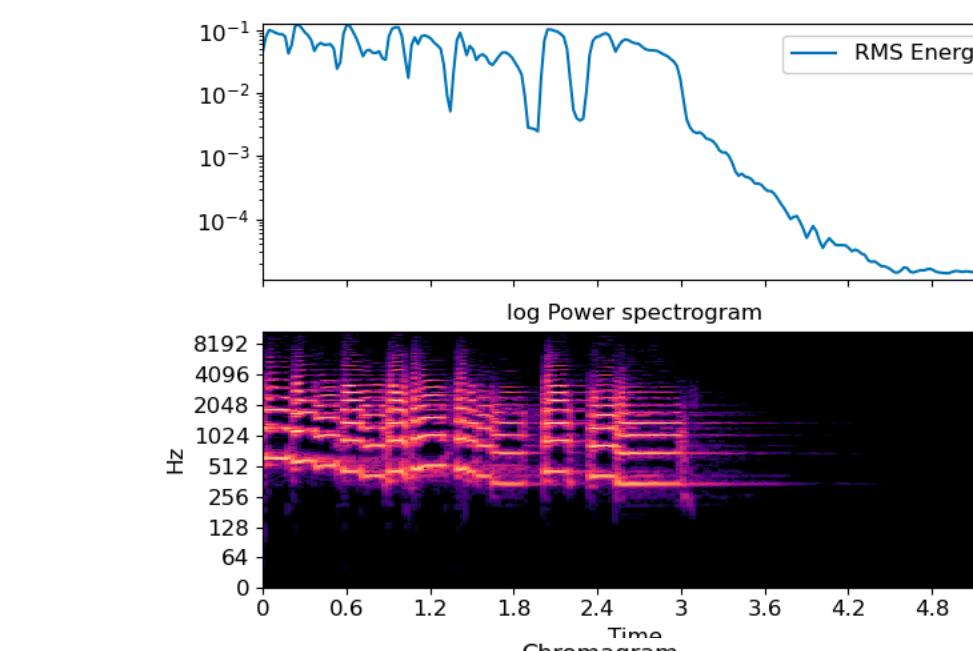
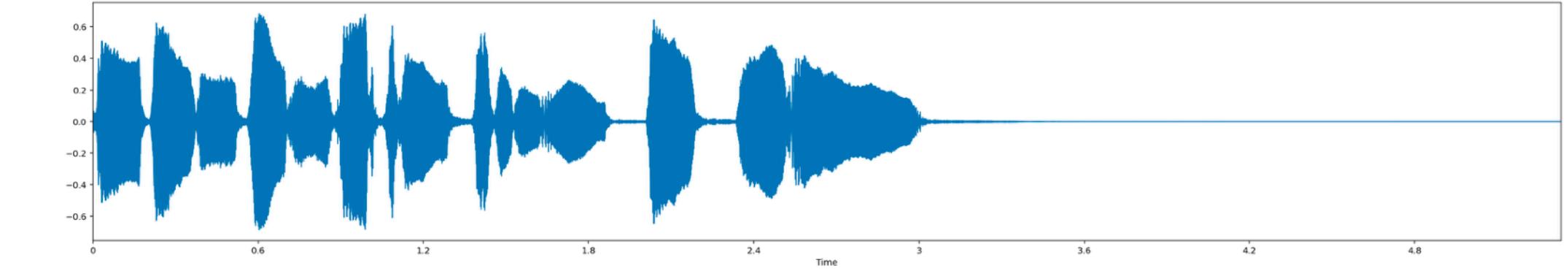
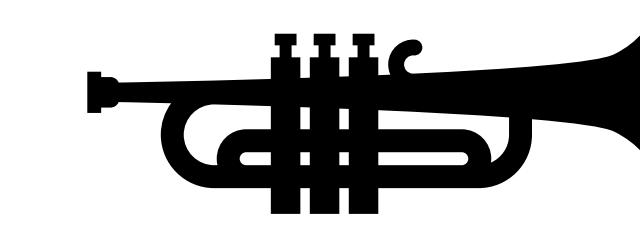
- ex: 麺料理
 - 麺の量, どんぶりの大きさ, 汁の量, 具の量 etc...
 - 特徴量は対象の性質の分析や, 他との弁別を行う上で大切
- 特徴量の計算は特徴抽出 ; feature-extractionとも



“音響”特徴量

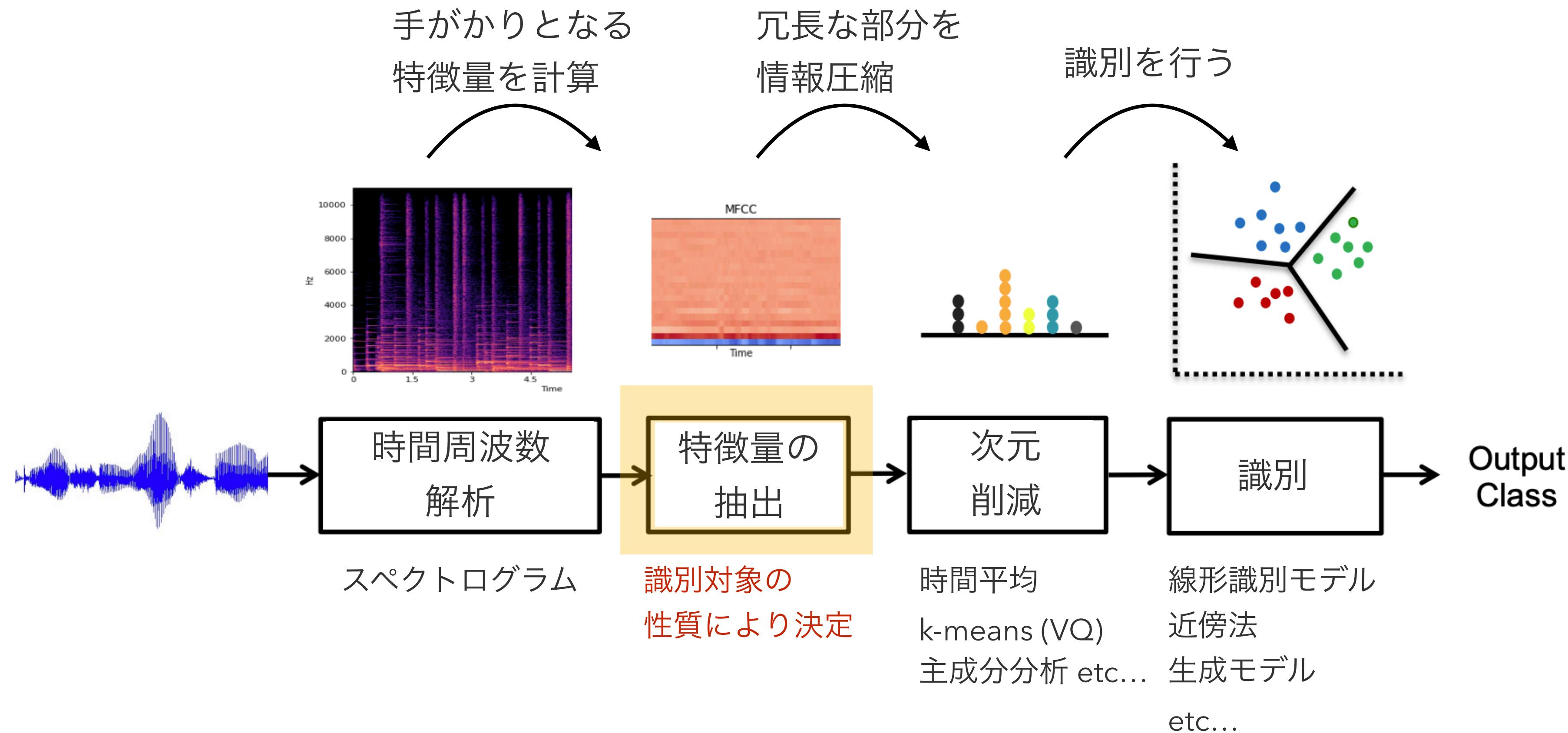
音に関する特徴を数値化したもの

- ex: トランペットの音
- 音量; RMSE
- ピッチ; f_0
- 音色; MFCC
- ピッチクラス; クロマベクトル



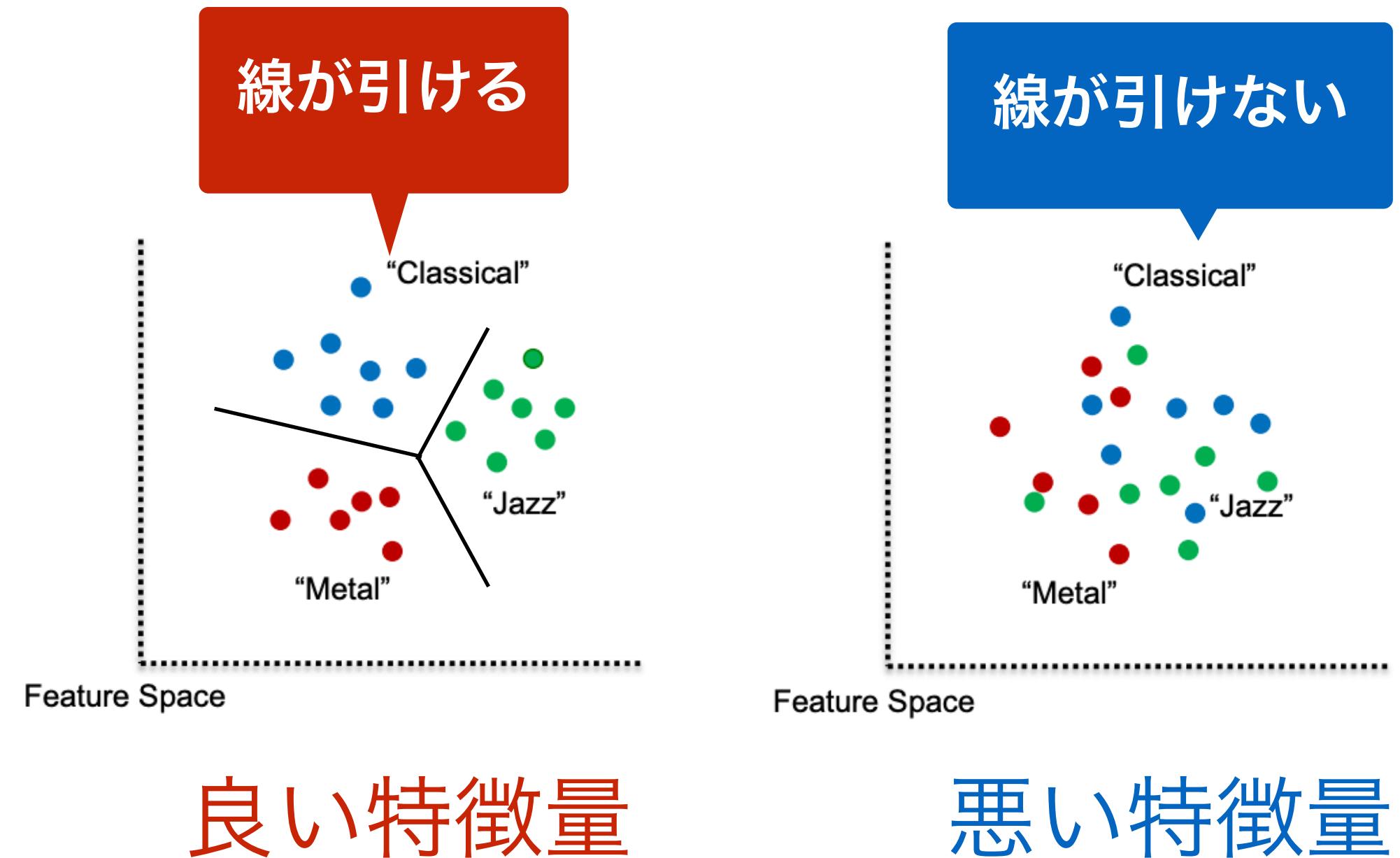
機械学習による識別のステップ

特徴量の計算は識別の手がかりとするために必須



良い識別には良い特徴量が不可欠

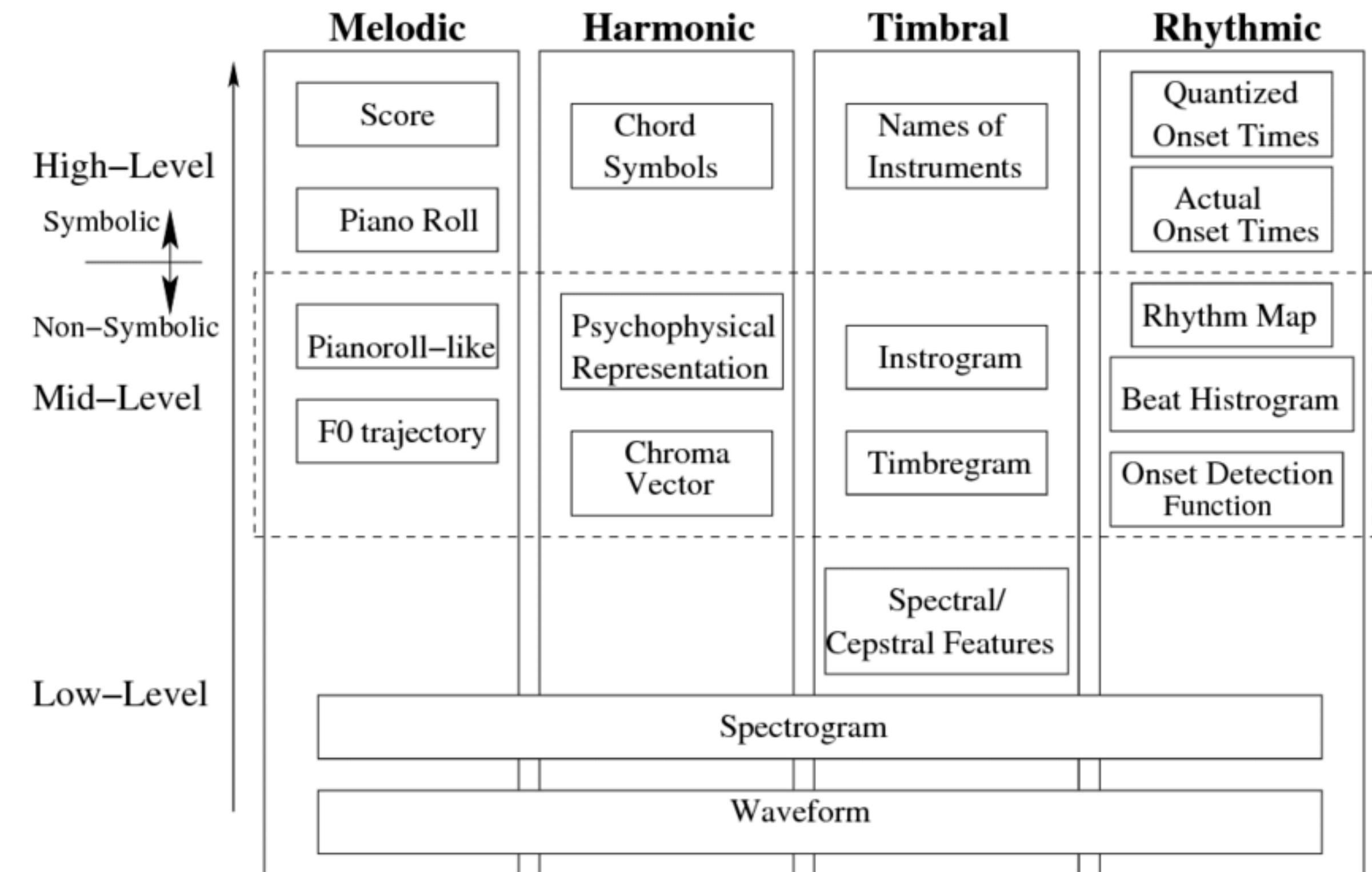
- 識別はいわば「マッピングへの線引き」
 - マッピングは識別対象をくっきり分割するようになされるべき
 - 特徴量への変換はマッピングに相当。ここで失敗すると、よい分割線（識別平面）を引けなくなる
- ex. アンチパターン
 - ロングトーンの演奏からの楽器識別にF0（ピッチの音高系列）のみを使う
 - 音高の推定にMFCC（音色に関する特徴量）のみを使う



対象を観察して、対象に合った特徴量を使おう！

いろいろな音響特徴量

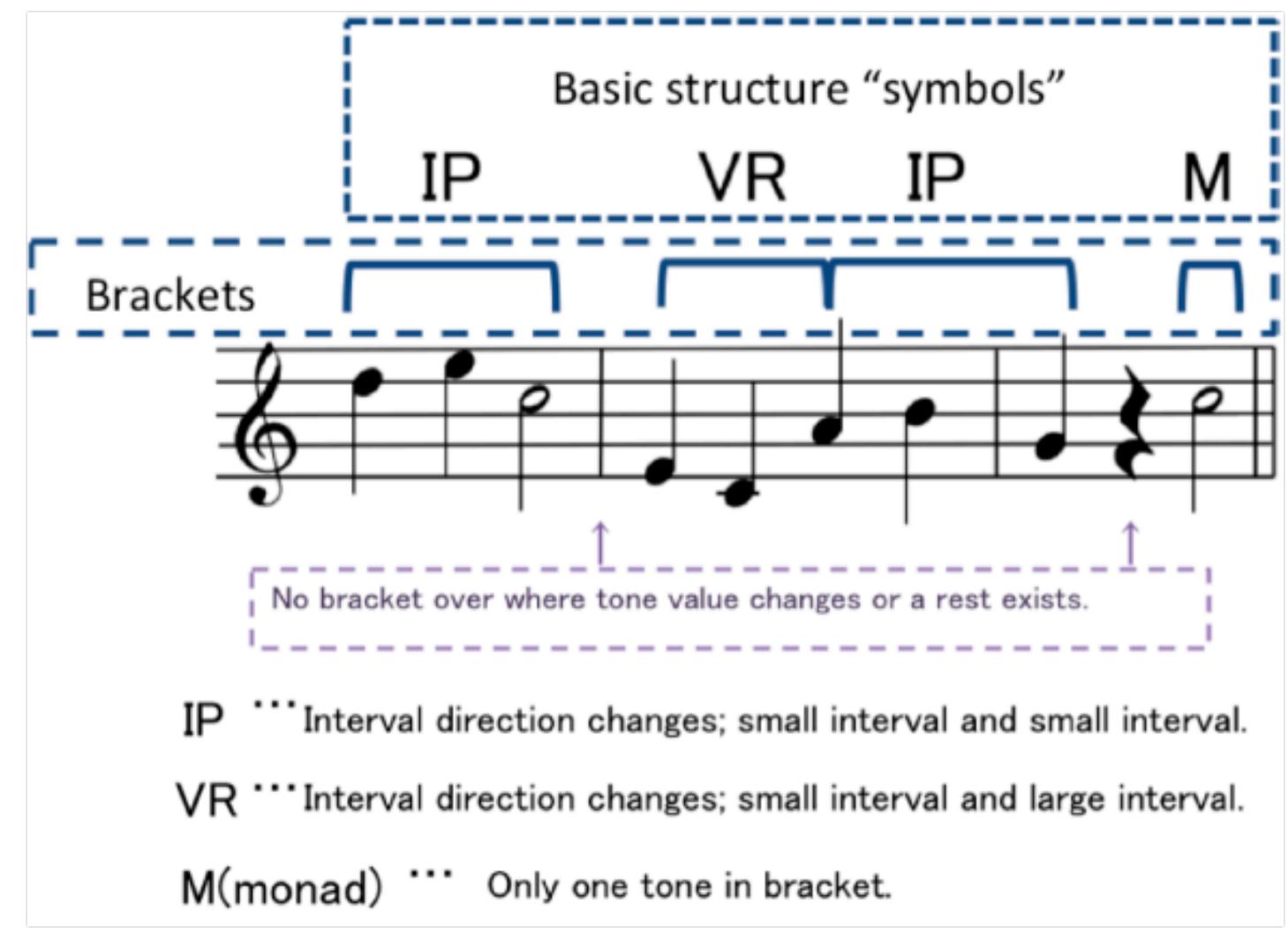
- **low-level**
 - 波形・スペクトログラム等プリミティブなものや、そこから計算される特徴量
- **mid-level**
 - 計算可能かつ、より何かの目的に特化したもの
- **high-level**
 - 人の解釈に基づく特徴量
 - 教師ラベル（識別対象）と同値になることも



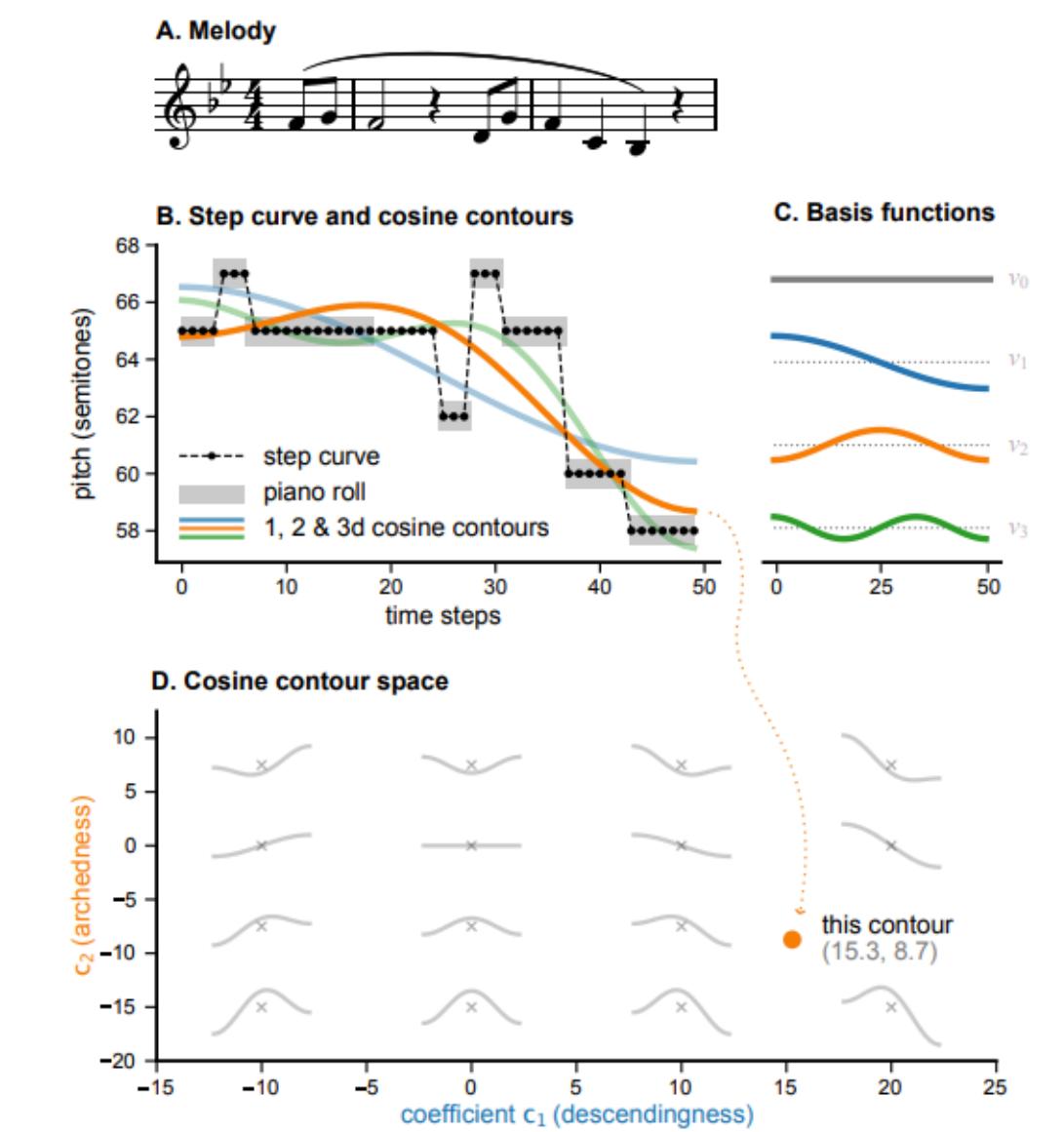
Mid-level Representations of Musical Audio Signals for Music Information Retrieval, T. Kitahara. 2010
Application of Multiway Methods for Dimensionality Reduction to Music, A. Ramaseshan. 2013.

楽譜の特徴量は？

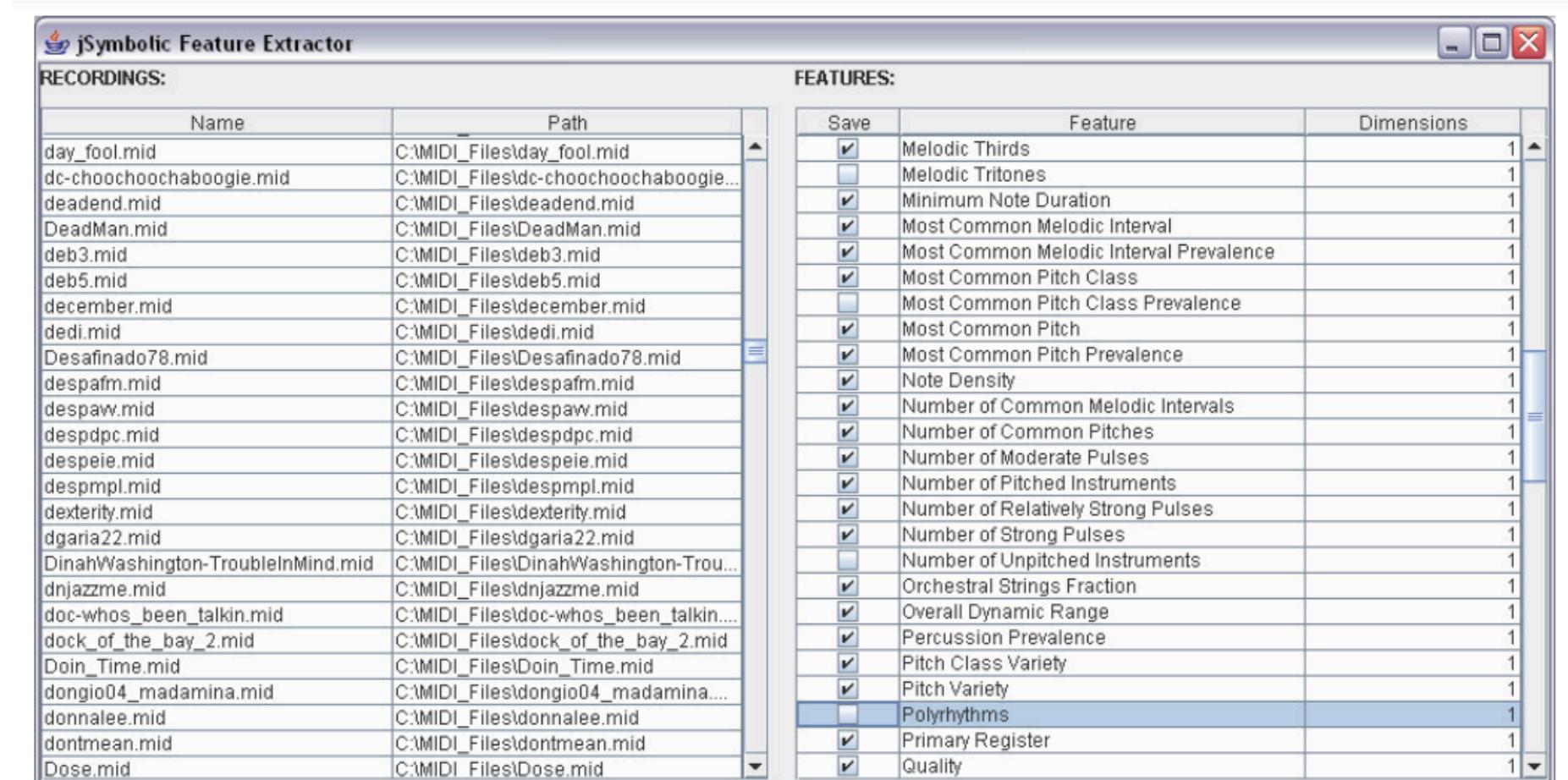
- 音響に比べて決定的なものはない？（印象）
- 有名な特徴抽出方法
 - 統計量の計算に基づく方法；jSymbolic
<http://jmir.sourceforge.net/jSymbolic.html>
 - music21でも使える
- メロディの概形化を行うもの
 - 暗意実現モデルによるシンボル化
 - Cosine-contour [Cornelissen 21]
- (詳しい方いたら補足お願いします)



暗意実現モデルによるメロディの分析
(<https://gttm.jp/hamanaka/music-analysis/>)



Cosine-contour
(B. Cornelissen et al. COSINE CONTOURS: A
MULTIPURPOSE REPRESENTATION FOR
MELODIES, ISMIR2021)



jSymbolicによる特徴抽出

2. 音響データにおける 有名特徴量の紹介

主要な特徴量

- RMS -> 音量
- MFCC -> 音色
- スペクトル統計量
- F0 -> 音高
- クロマベクトル -> 調性
- 他にもたくさんあるが割愛

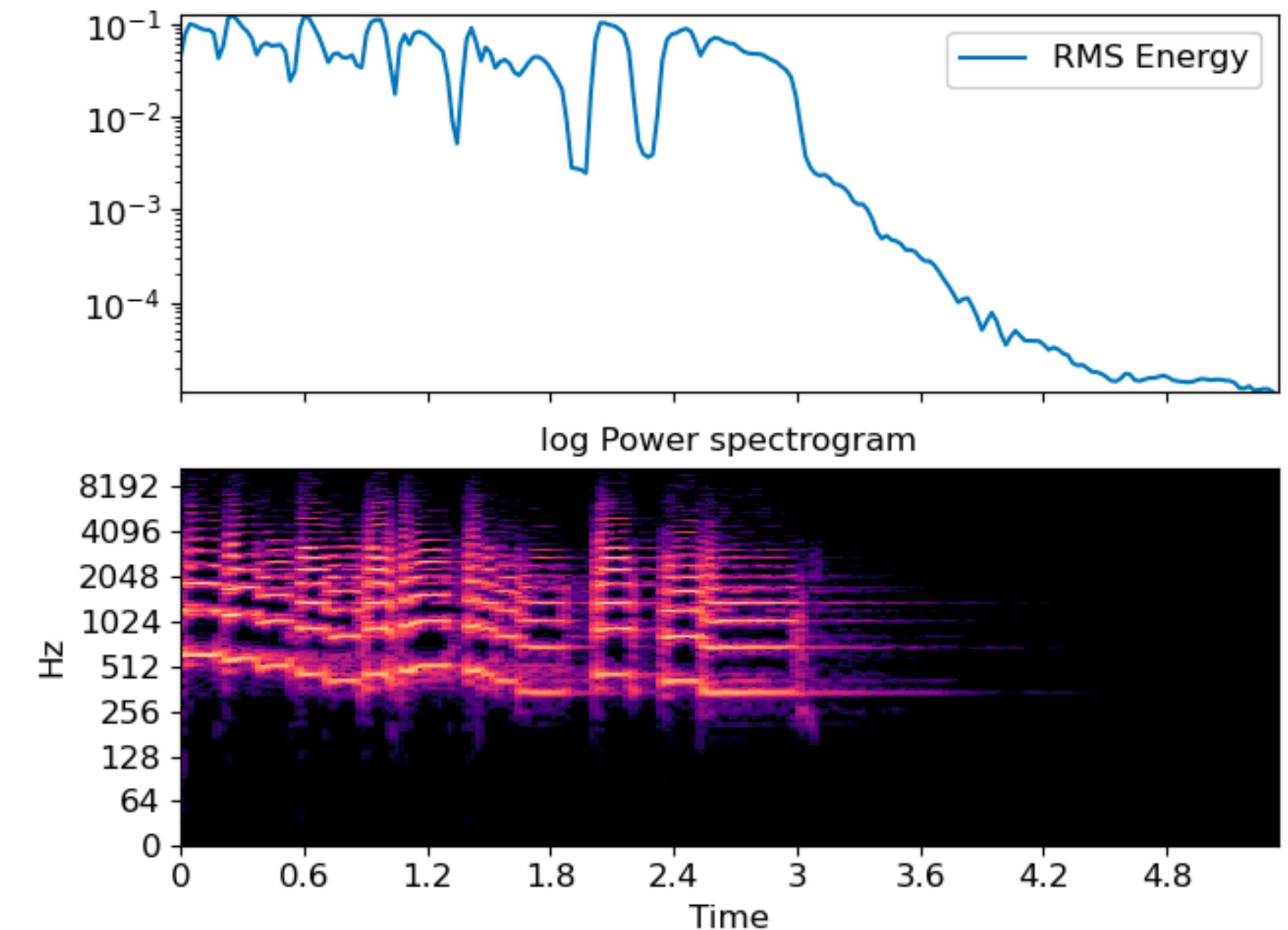
Root-Mean-Square (RMS)

13

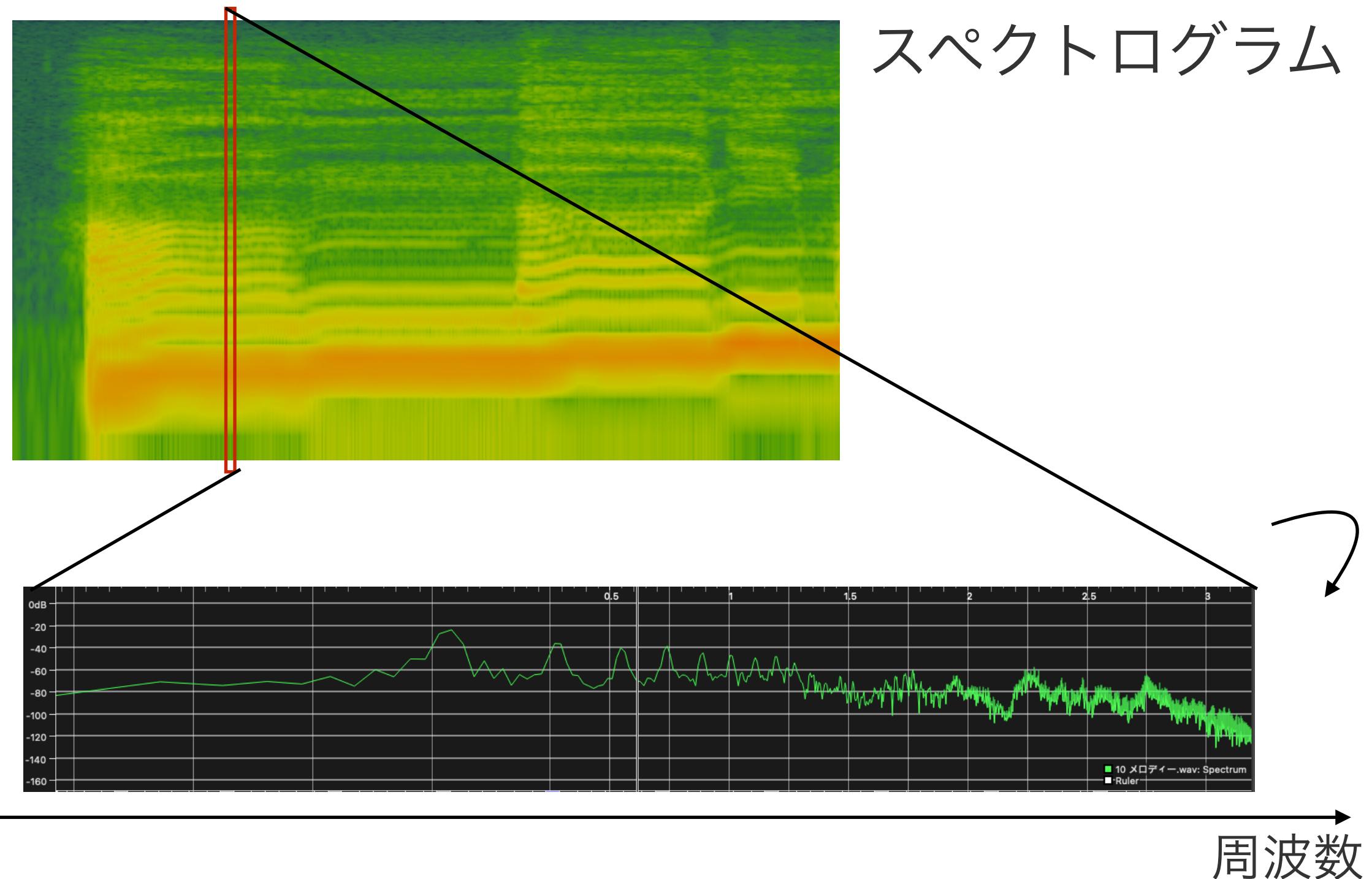
- 信号の振幅のエンベロープ = 音量に関する特徴量
- 計算方法；下記
 - 上：信号波形から
 - 下：STFT計算後から

$$\text{RMS}(l) = \sqrt{\frac{1}{N} \sum_{m=0}^{N-1} (x(m - l \cdot R) w[n])^2}$$
$$\text{RMS}(l) = \sqrt{\frac{1}{N} \sum_{k=0}^{N-1} |X(k, l)|^2}$$

$w[n]$: window
 N : window size
 R : hop size



- 振幅スペクトラム (STFTの1フレーム分の結果, 緑線) から得られる統計量

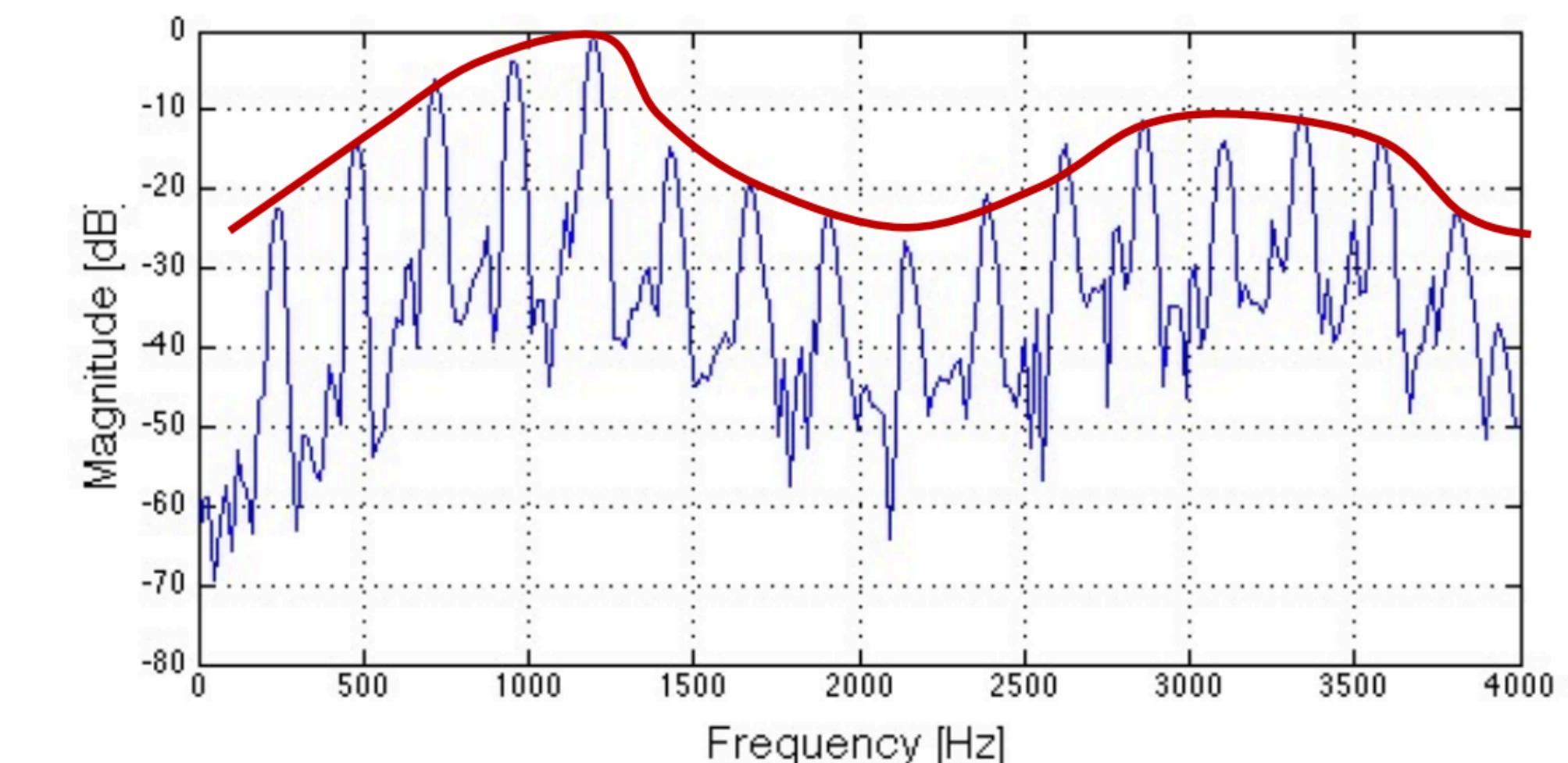
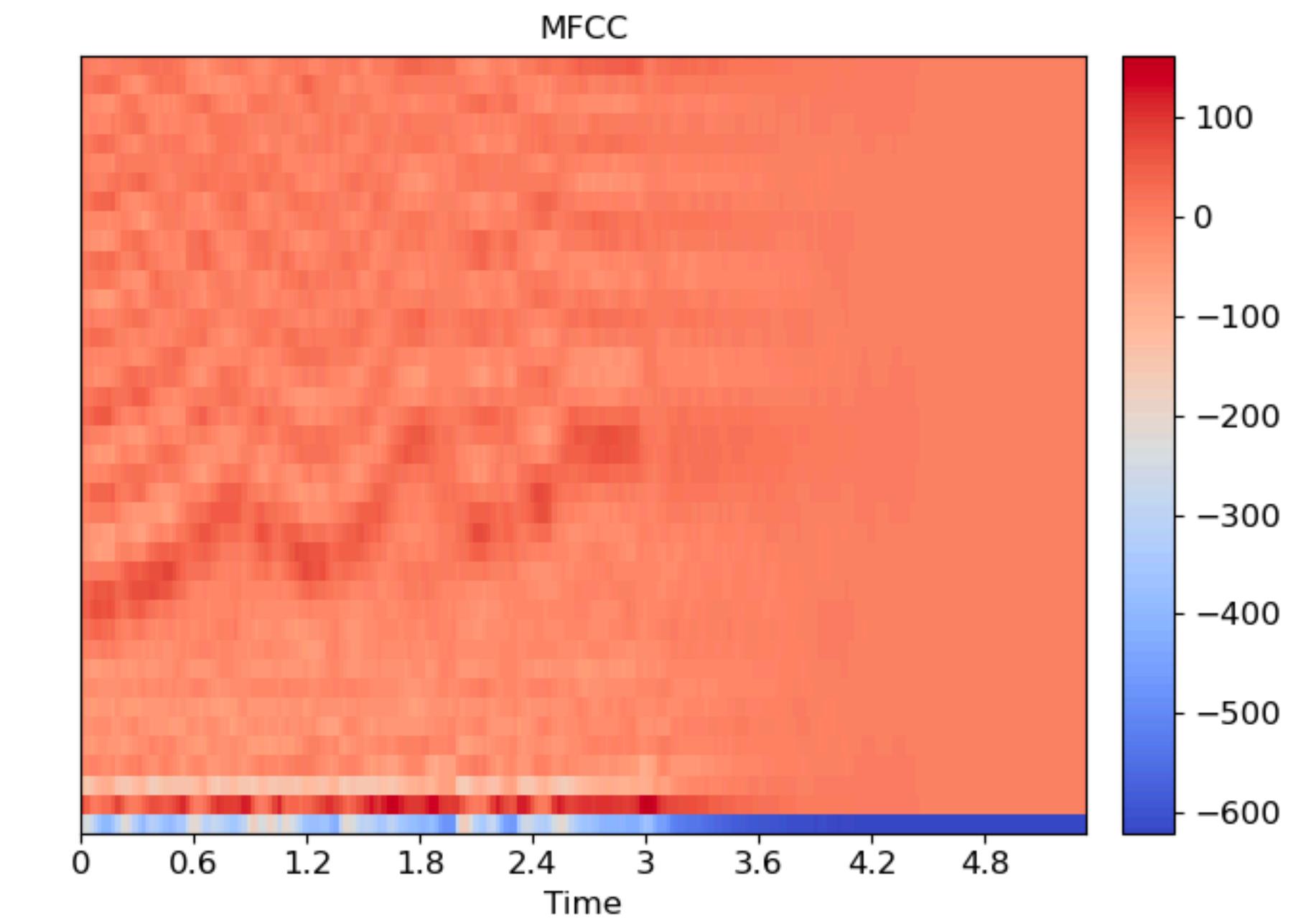


- スペクトルセントロイド
 - 重心. 音の明るさと関係があるといわれる
 - スペクトルフラットネス
 - スペクトラムの平坦さ. ノイズらしさと関係があるといわれる
- などなど...

MFCC (Mel-Frequency Cepstral Coefficients)

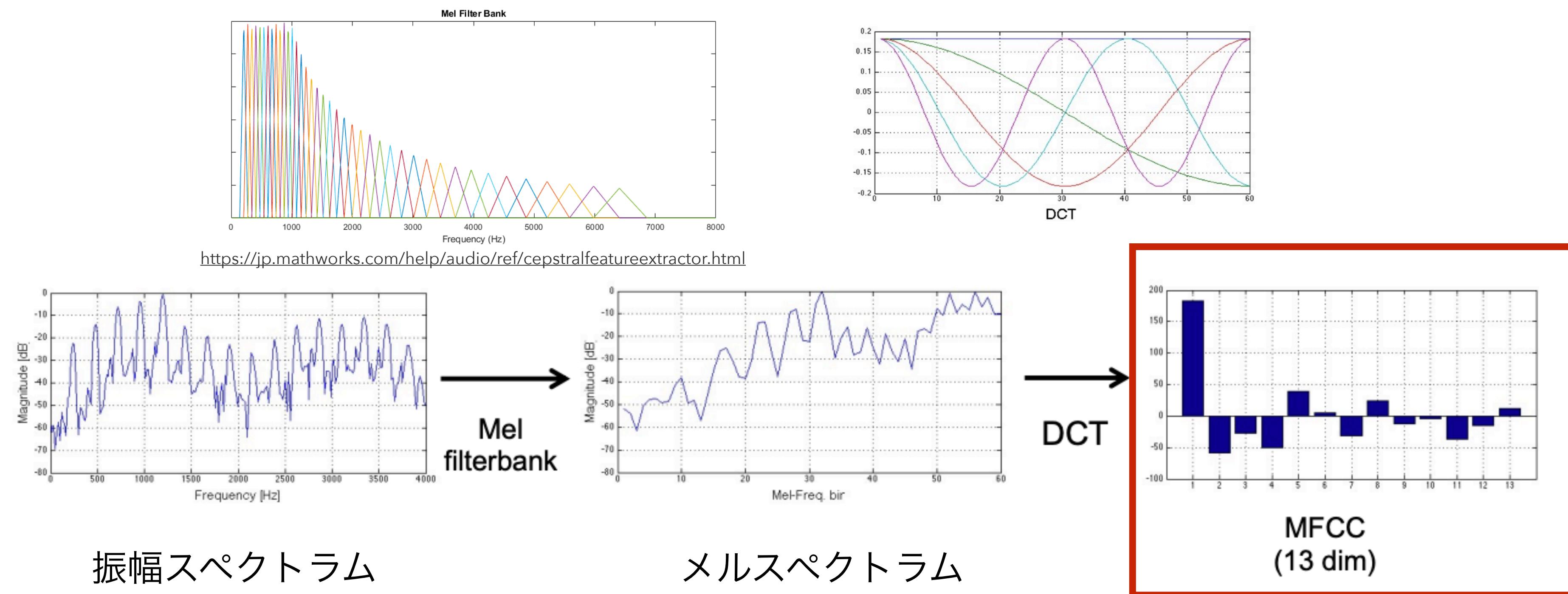
15

- 音色に関する特徴量
- 人間の聴覚に基づいた**メル尺度**の周波数を利用して算出
- 音では最もメジャーな特徴量
- 低次特徴には音色, 高次特徴にはピッチの特徴があらわれる
- スペクトル構造の概形を表現できる



MFCCの計算手順

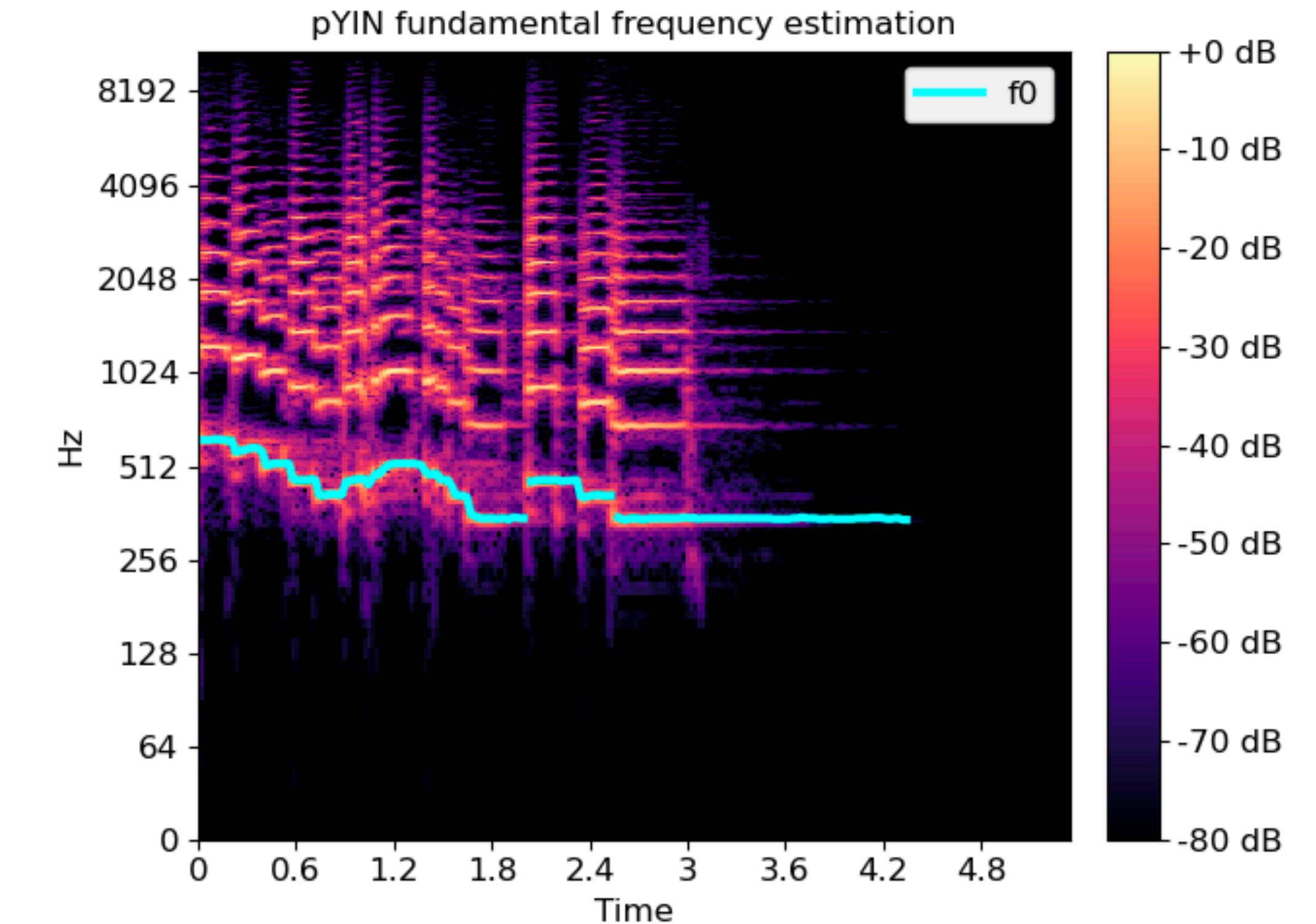
- 各フレームの振幅スペクトラムに対し以下の2つを適用
 - 1. メルフィルタバンクの重畠
 - 2. 離散コサイン変換 (DCT) の適用



基本周波数 (Fundamental frequency; F0)

17

- 音高に関する特徴量
- 音の波形周期の逆数であり
- 倍音の中から基本となっている波の周波数
 - 音高と一致する場合が多い
 - (F0 ≠ ピッチであることに注意！！
(物理量と心理量))
- 抽出法もさまざま



- 基本的な方法

- ゼロ交差法, 自己相関法, ケプストラム法

- より実践的な方法

- Yin [1], pYin [2], HARVEST [3], CREPE [4], ほか

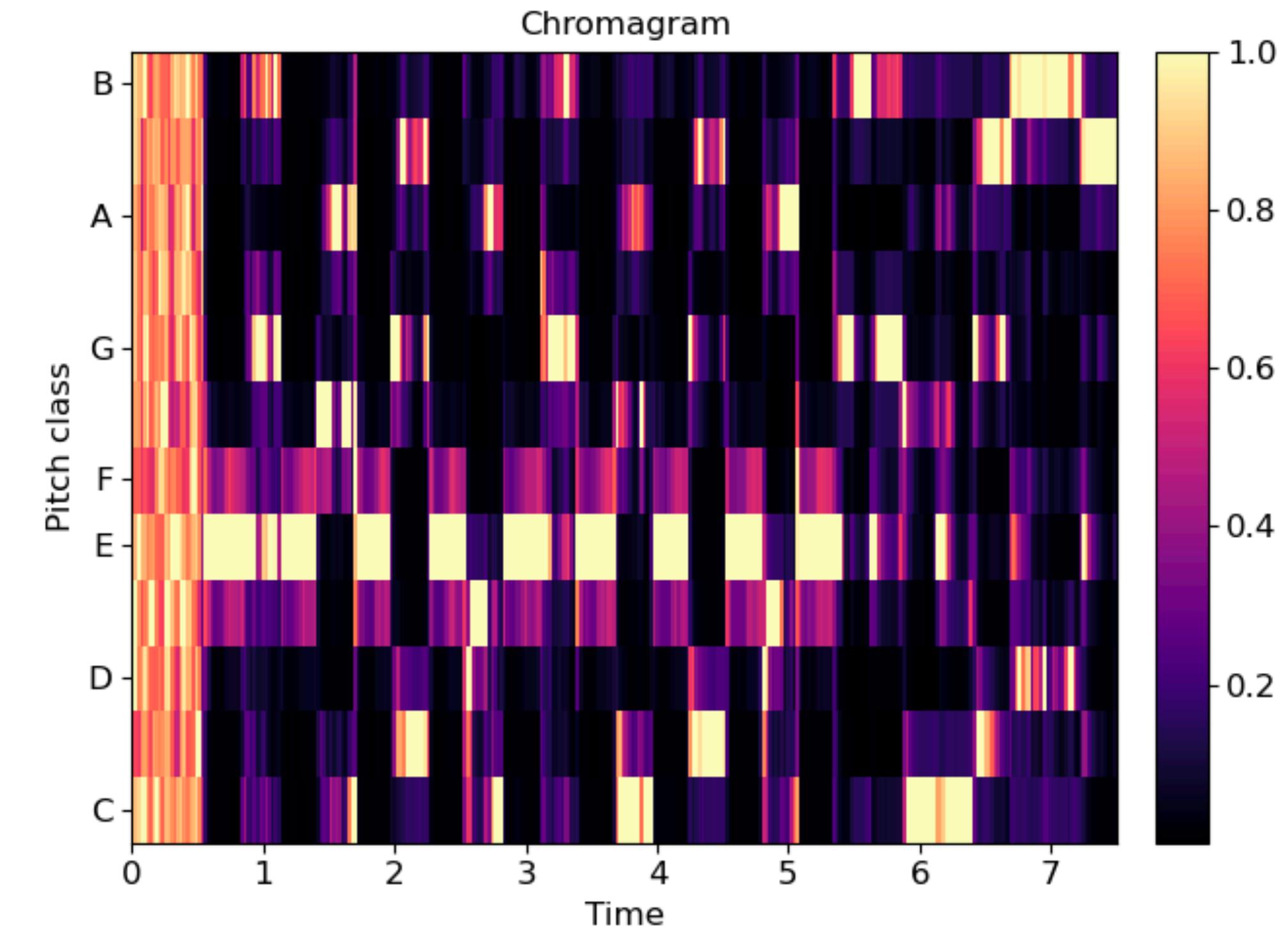
[1] De Cheveigné, Alain, and Hideki Kawahara. "YIN, a fundamental frequency estimator for speech and music." *The Journal of the Acoustical Society of America* 111.4 (2002): 1917-1930.

[2] Mauch, Matthias, and Simon Dixon. "pYIN: A fundamental frequency estimator using probabilistic threshold distributions." *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014.

[3] Morise, Masanori. "Harvest: A High-Performance Fundamental Frequency Estimator from Speech Signals." *INTERSPEECH*. 2017.

[4] J. W. Kim, J. Salamon, P. Li and J. P. Bello, "Crepe: A Convolutional Representation for Pitch Estimation," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 161-165,

- 調性に関する特徴量
- 各音の音の強さを12次元のベクトルで表現
- 音楽で、同じ音名の音（オクターブ違い）は同じように扱いたい時に使う（コード進行推定など）

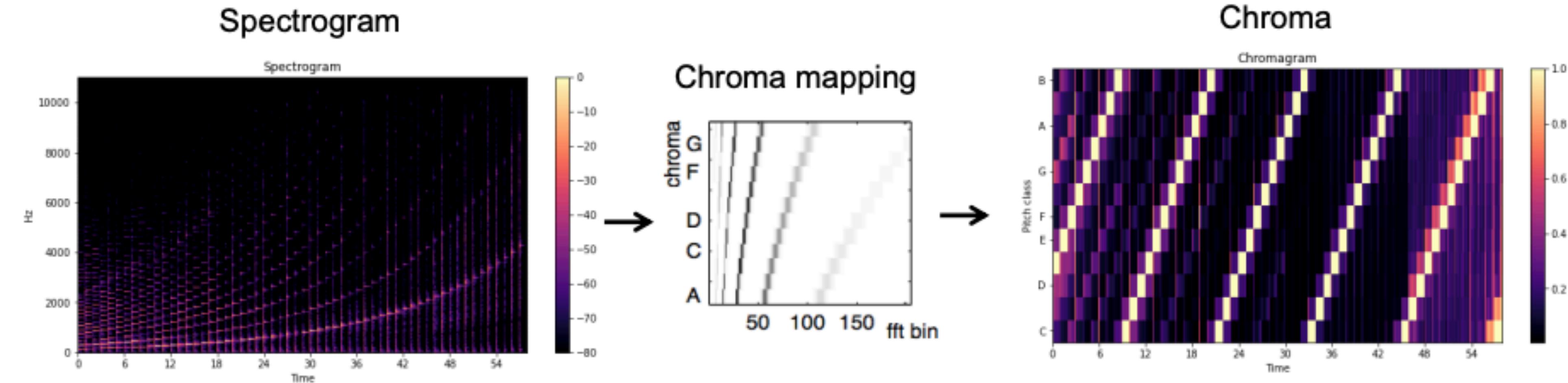


https://librosa.org/doc/main/generated/librosa.feature.chroma_stft.html

クロマベクトルの算出処理概要

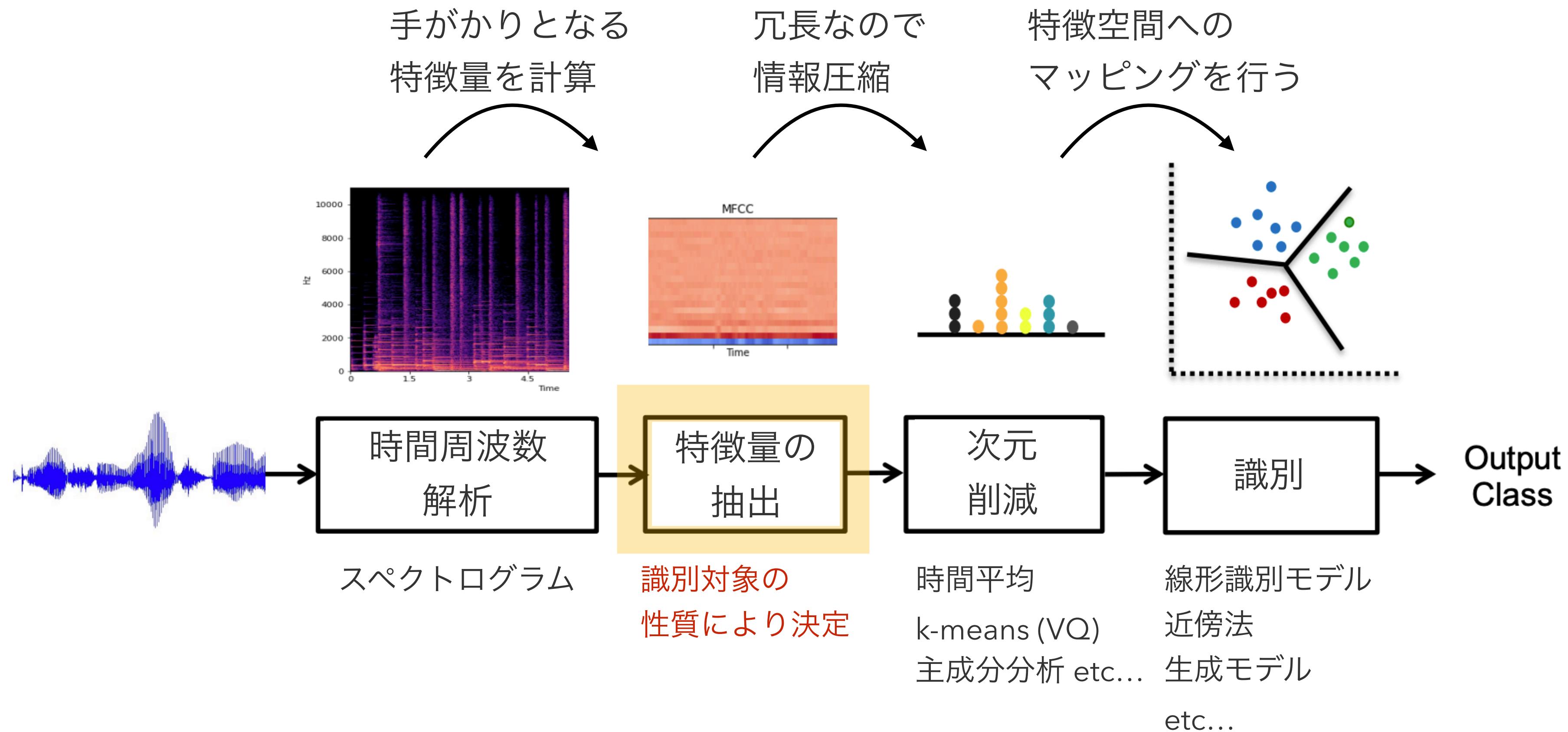
20

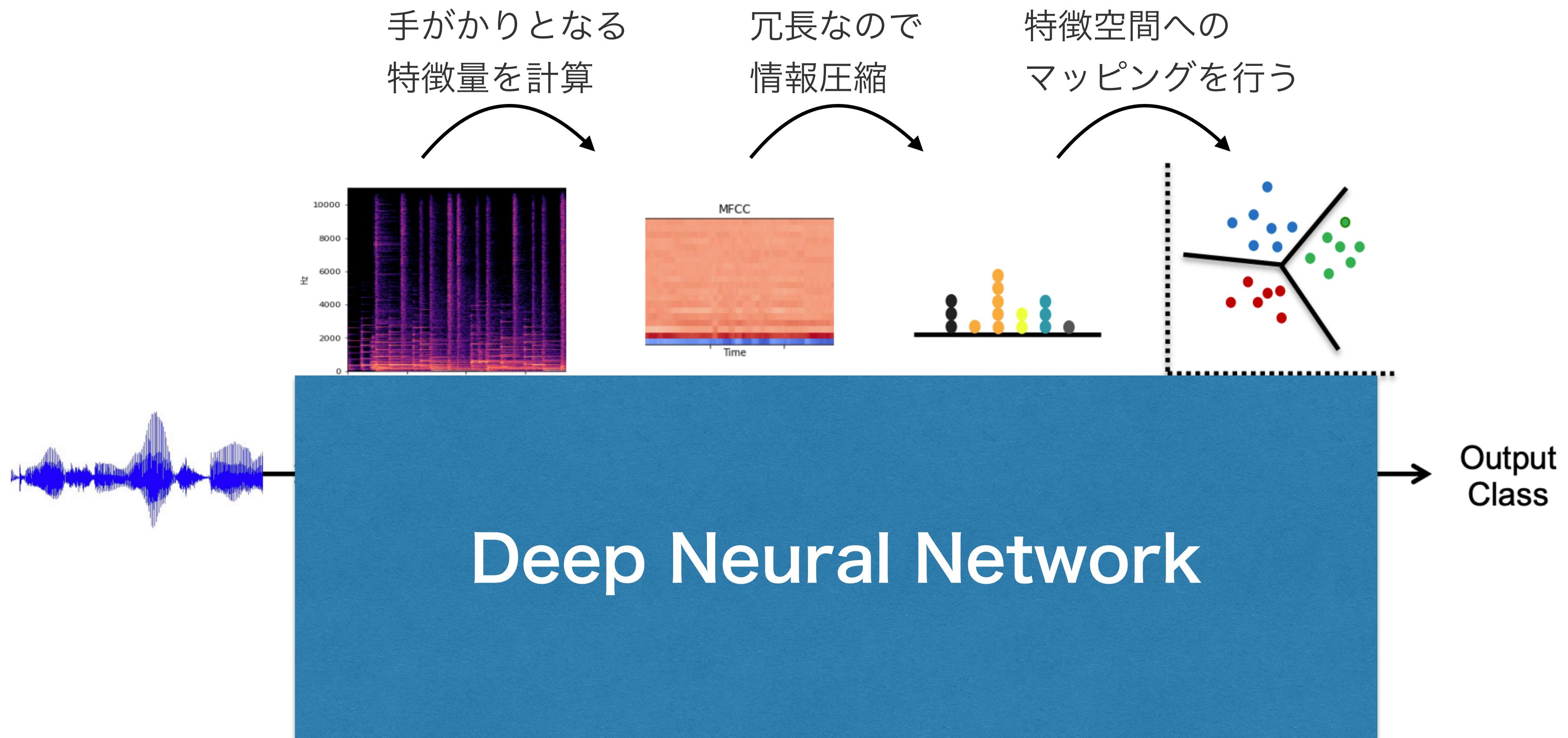
- スペクトログラムを入力に、**クロママッピング**というスケールへの変換処理を行う
- 周波数をピッチクラスに割り当て



- Librosa
 - もう定番中の定番
- Essentia
 - <https://essentia.upf.edu/>
 - 音楽に特化した特徴量たち。
 - 山本はビブラート特徴量を計算した
- Opensmile
 - <https://www.audeering.com/research/opensmile/>

最後にちょっとだけ
深層学習に片足をつっこみます





ここまで任意のステップをDNNに任せることができる

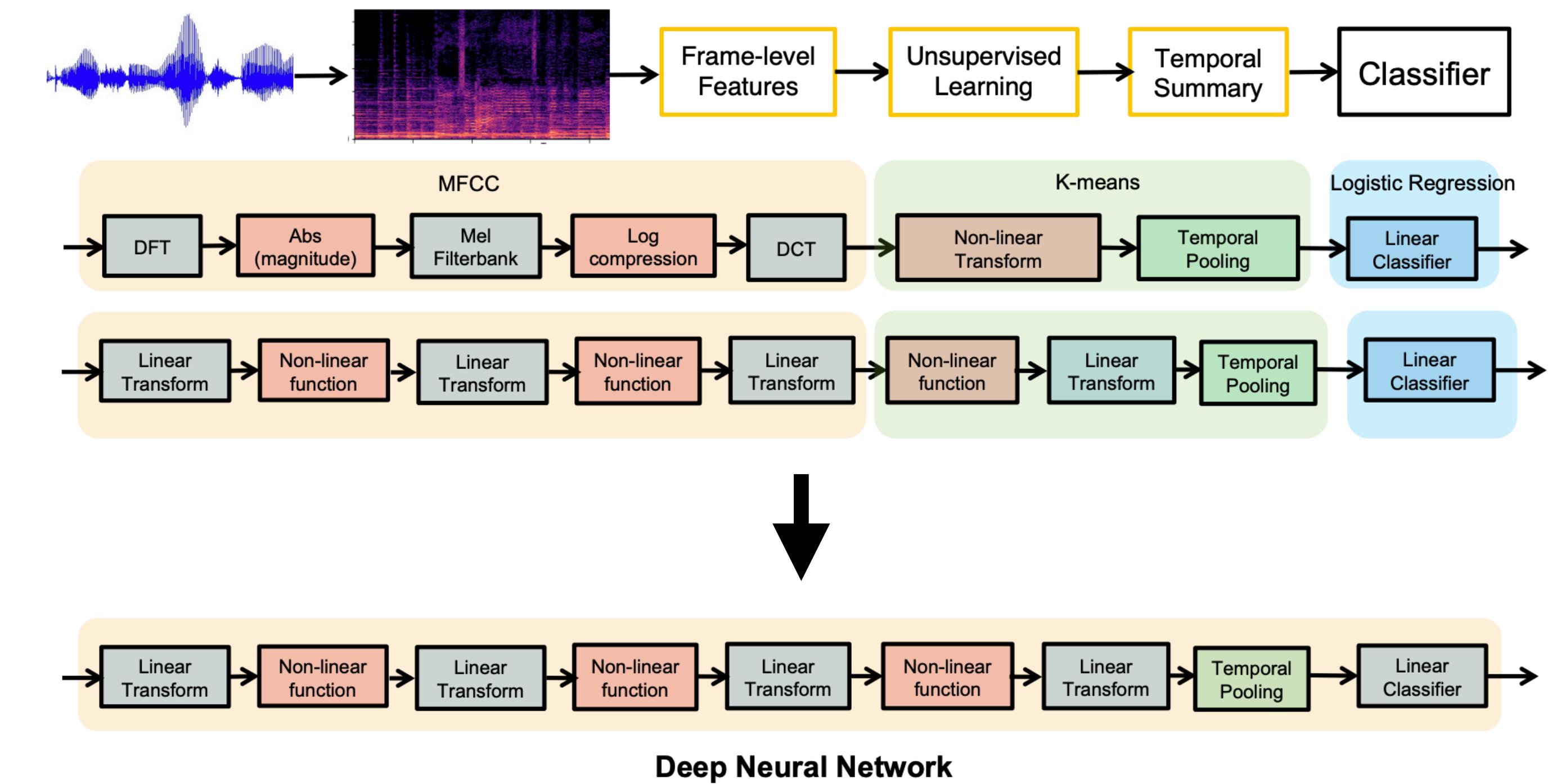
=自分でルールを設計するのではなく、データからルールを獲得する

従来手法と、線形処理 + 非線形処理の繰り返しという共通点がある

- Ex; MFCC + k-means 次元圧縮 &

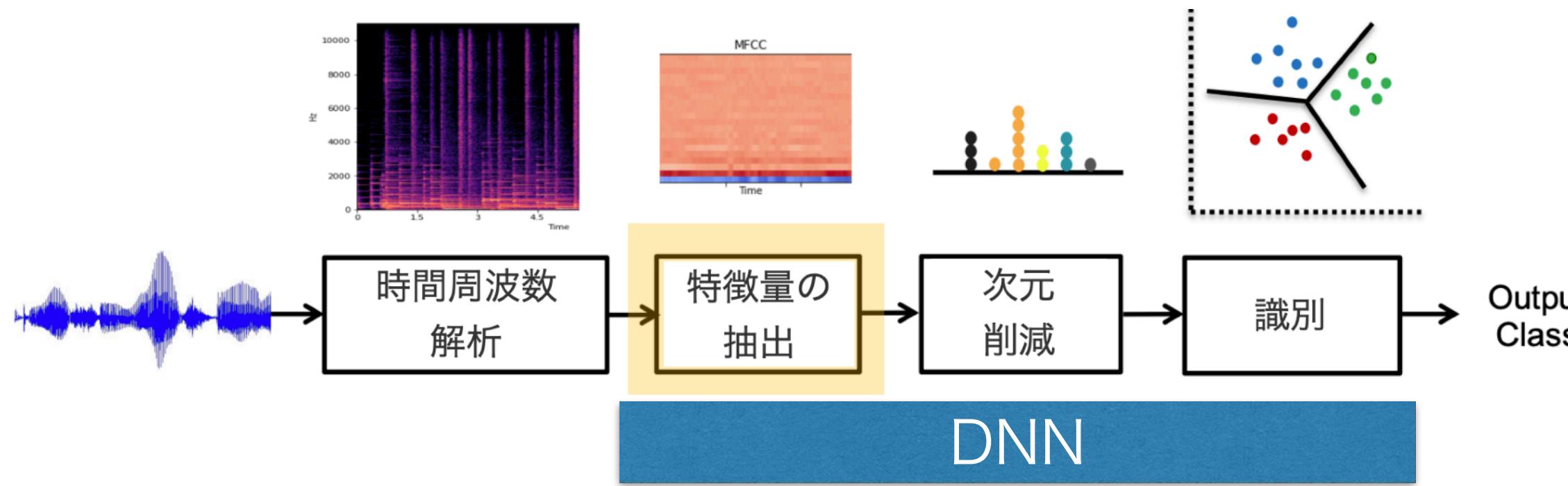
線形識別モデル

- MFCC, k-meansはともに線形処理
+非線形処理を繰り返す処理

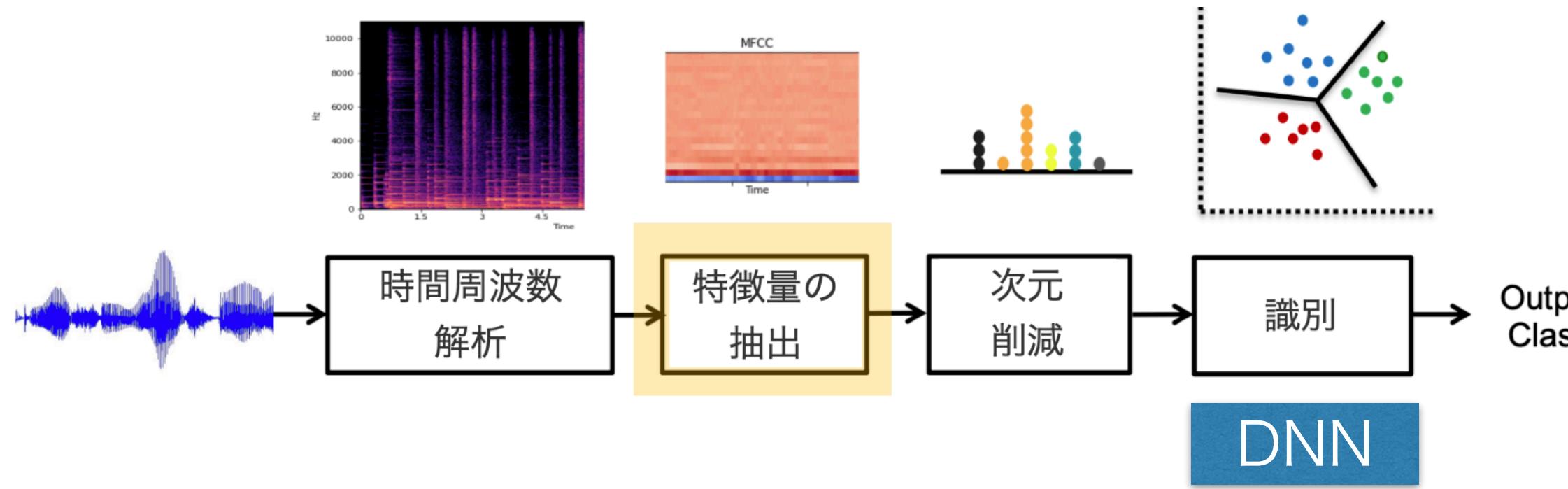


- DNNも同じく線形処理（重み付き和） +
非線形処理（活性化関数）の繰り返し

一部だけDNN, というモデル化も可能

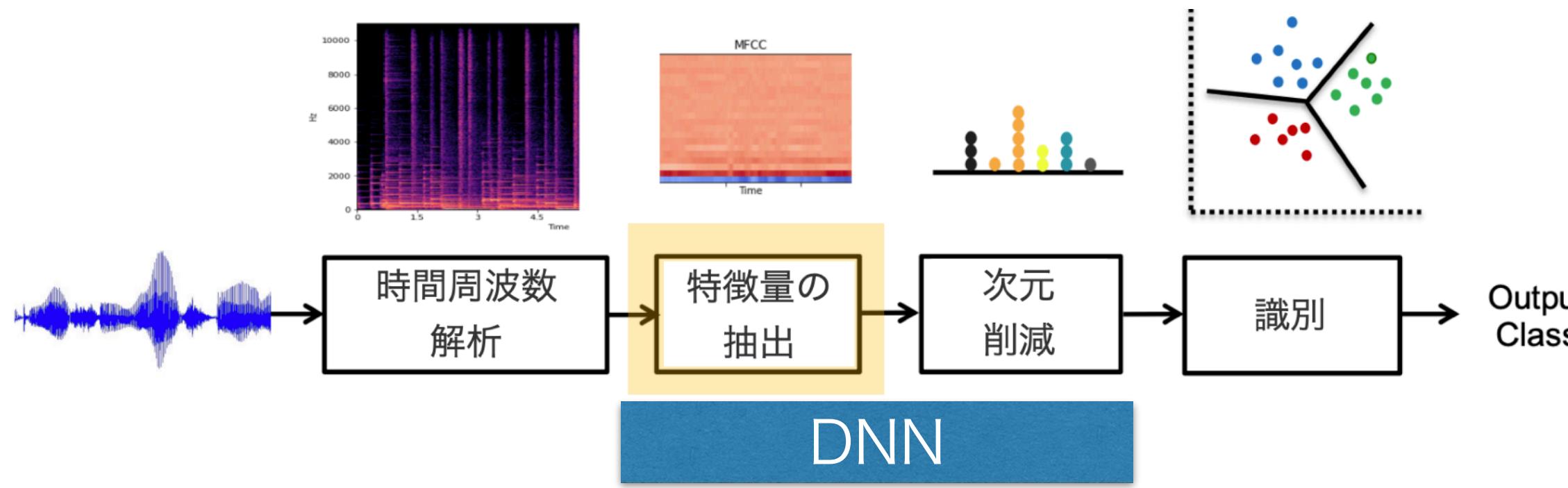


- ・ パターン1：スペクトログラム入力型
 - ・ スペクトログラムを画像のように扱い, CNN等を適用
 - ・ 現在一番メジャーな使い方
 - ・ メルスペクトログラム, CQTスペクトログラムが多い

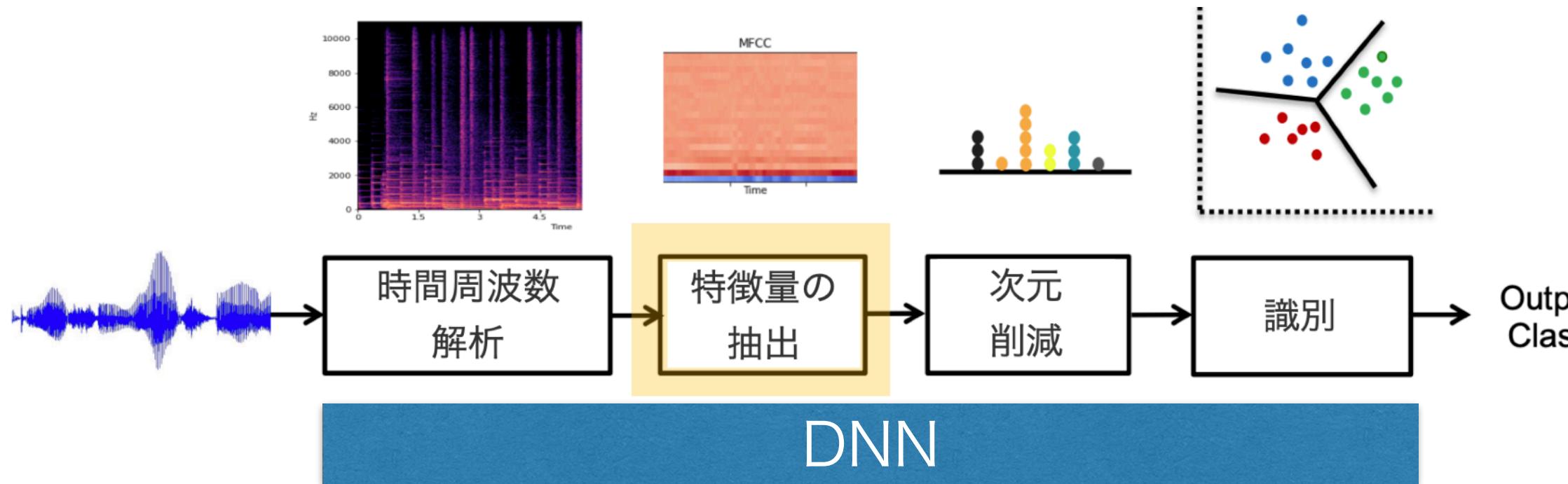


- ・ パターン2：識別器のみDNN型
 - ・ MFCCやF0等を入力するなど
 - ・ 時系列推定問題（音声認識等）に多い
 - ・ ↑の特徴量 + LSTM等. 系列依存関係だけDNNにモデルリングする場合等

一部だけDNN, というモデル化も可能



- **パターン3：特徴量学習 (feature learning)**
 - 特徴量設計をDNNに肩代わりしてもらう手法
 - 識別器にはSVMやRandom Forest等だったり、他のNNだったり
 - マイナーだが、少量データセットには効果的
 - (山本APSIPA論文はこれを採用)



- **パターン4：End-to-End**
 - 信号波形を入力して、直接出力を得る
 - STFTのような計算もDNNに肩代わりさせる
 - データ量がたくさんある場合に効果的

- 特徴量とは、対象の特徴を数値化したもの
- 自分の対象に合わせて用いる特徴を変える
 - 識別の性能に大きく影響する
- 音響特徴量の紹介
- 深層学習は特徴量抽出 - 識別までの処理の肩代わりを可能にする

EOF