

音響学入門ペディア

Q. 様々な音響特徴量
それぞれの使い方や意味を教えてください

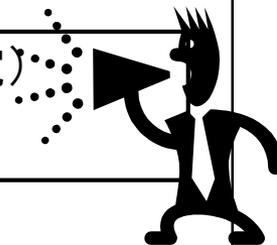
千葉 祐弥

東北大学 大学院 工学研究科
博士後期課程2年

特徴量って何に使うもの？

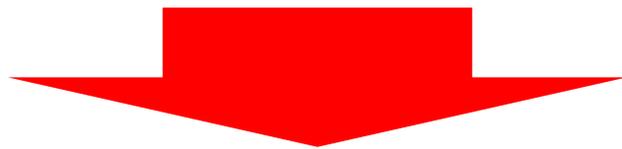
• 統計的分析

- 人間が音を聞く仕組みを解明する(方向、高さ、大きさ、音色...の知覚)
- データの符号化、圧縮への応用 など



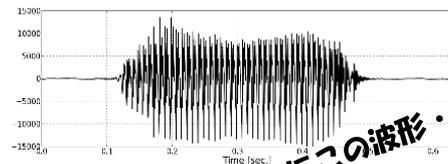
• 機械学習、パターン認識

- 音声認識、音声インターフェースの作成
- 楽曲のジャンル推定、楽曲検索、推薦等への応用 など



生のデータそのものを利用するのは無理がある

- 収録条件、個人差、データ差などの違いが大きくて本質的な部分がわかりにくい
- データ量や計算量が多くて取り扱いにくい



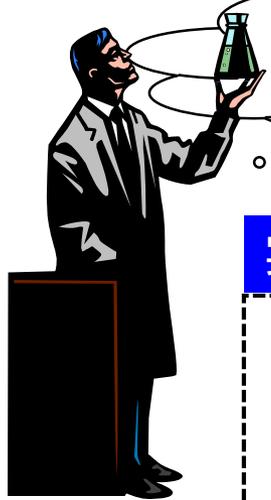
より無駄が少なく、データの本質をあらわした表現が欲しい！
⇒ データからなるべく良い特徴量を抽出する必要がある

特徴量の選び方

よい特徴量とは？ ⇒ 問題によって異なる

どのような特徴量を使うかは分析から得られた示唆やひらめきなどで決まる

例)



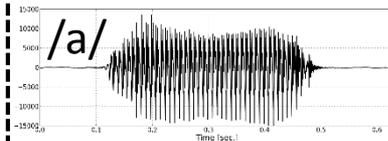
音素の判別に有効なのは、
スペクトルの概形に違いない！

積み上げられた知見

突然の思いつき

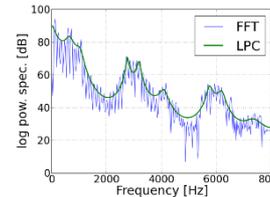
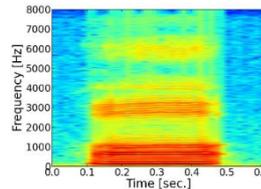
実際の作業

データ
(音声波形)



特徴抽出

LPC係数やMFCC
(スペクトル概形の表現)



分析or実験



統計的分析や認識実験
により理論や効果を実証

自分が対象の説明に有効であると思えば何でも特徴量になり得るので、
無限に種類が考えられる！ よし悪しは別ですが…

とは言っても、 やはり実績のある音響特徴量はある

このスライドでは特に、音声認識関連の話を中心に代表的なものを紹介します

以降のスライドの構成

- 音に関する人間の知覚について
- 音の基本属性にのっとった代表的な特徴量の紹介
- 音声、音楽の認識分野での典型的な事例の紹介
- まとめ

特徴量の紹介、その前に...

知っててお得な人間の聴覚特性

- 何かの仕組みを考えると、生物を参考にするのは良い方法
- 特に音声認識に関連する分野では、人間がどのように音を聞いているのかを特徴量に反映させることが多い
- このような知見も、統計的な分析によって得られる

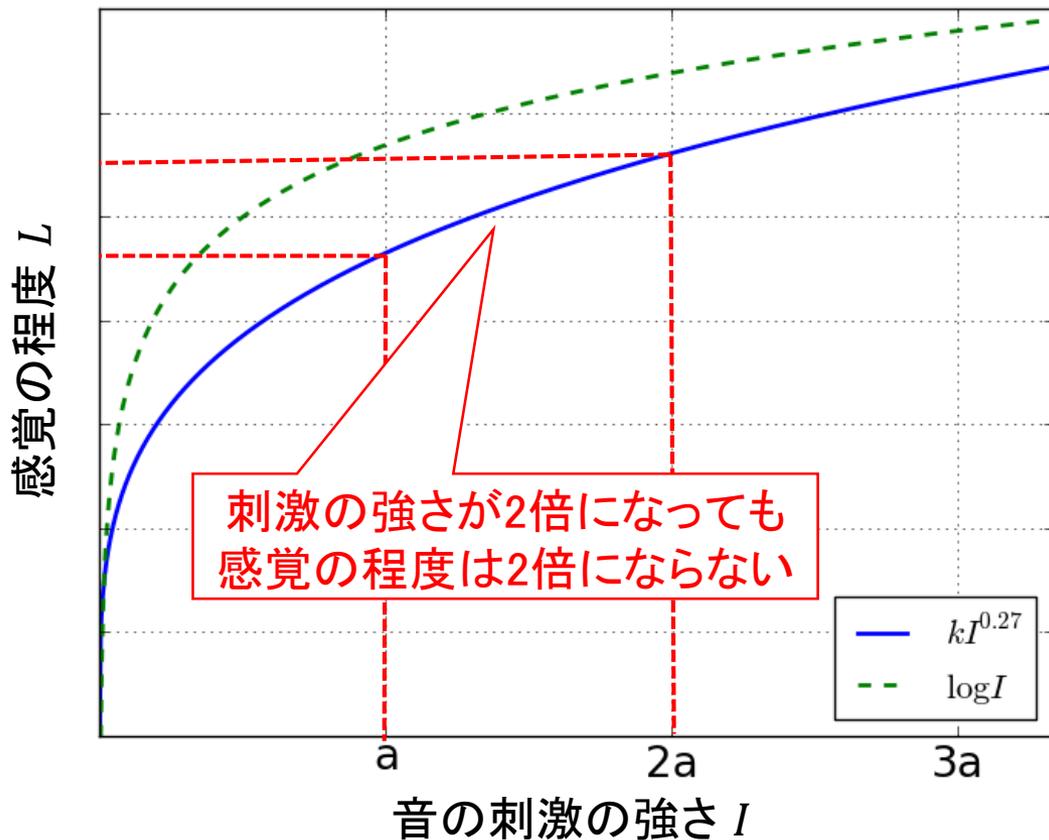


アノマロカリスに耳はない

音の大きさの知覚

音の大きさの心理尺度 ラウドネス

⇒音の刺激の強さと大きさの感覚の程度のおおむね下図の青線の様になる(だいたい音のレベルが10dB上がると二倍の大きさに聞こえる)

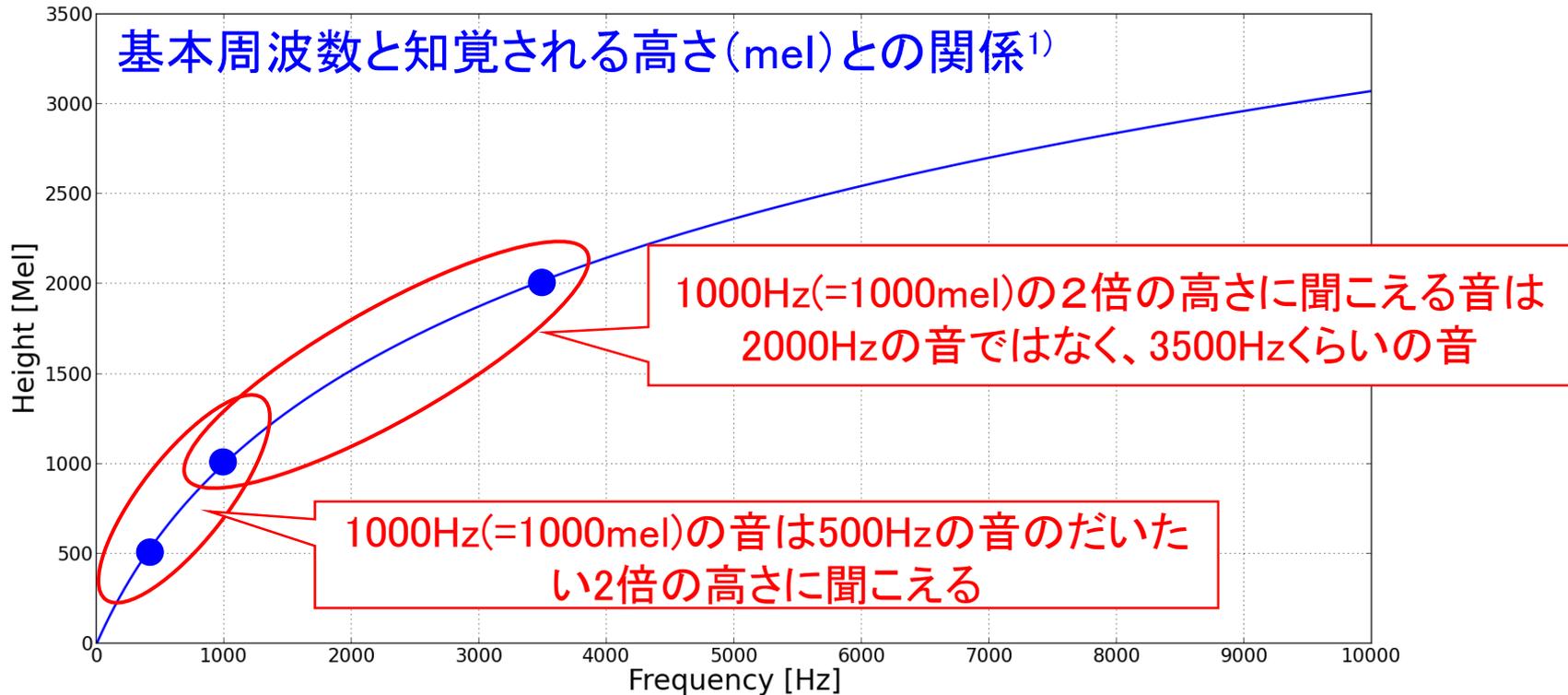


また、周波数ごとに音の大きさの感じ方も異なる(参考:等ラウドネス曲線)

音の高さの知覚

高さの心理尺度 mel

⇒ 例えば1000 melの2倍の高さに感じる音を2000 melとする尺度



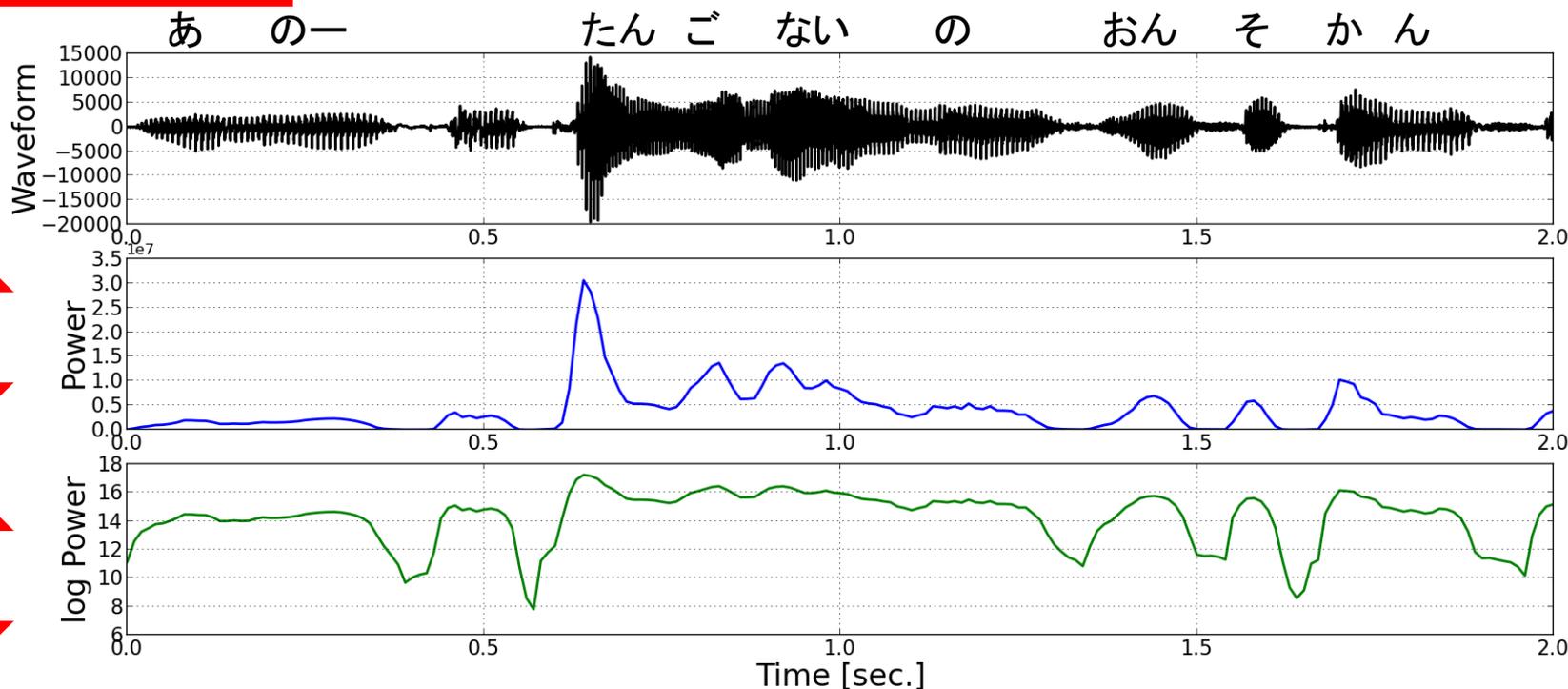
¹⁾ 鹿野ら, 「音声認識システム」, オーム社より

音の高さも、単に刺激が k 倍になれば k 倍に知覚されるわけではない

1. 音の大きさに関する特徴量

- 音の大きさは**パワー**で表現
 - パワー = 波形振幅の二乗の時間平均値
 - 波形全体、もしくは一部分で計算される
 - 下の図では25ミリ秒区間のパワーを10ミリ秒毎に計算し、時間方向に並べている
- 人間の感覚を考慮して対数尺度に変換したものを特徴量とする場合が多い

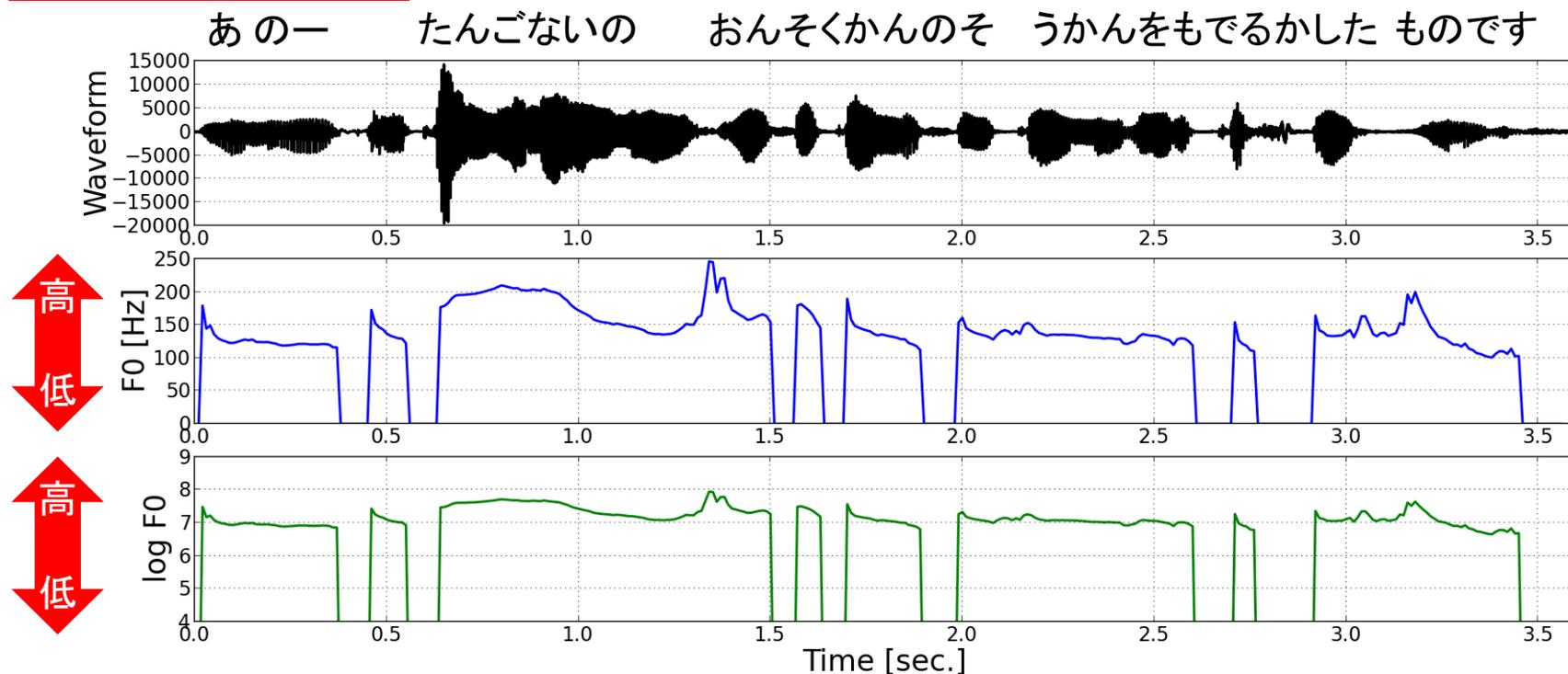
パワー抽出の例



2. 音の高さに関する特徴量

- 音の高さの基本は**基本周波数(F0)**
- ベーシックなF0抽出方法としては自己相関法、ケプストラム法などがある
- こちらもF0を対数尺度に変換したものを特徴量とする場合が多い

音声のF0抽出の例

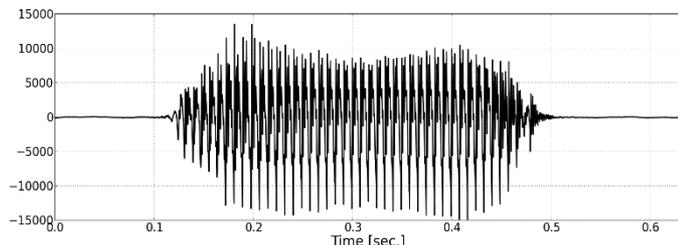


基本周波数については、「音響特徴量ってなんですか？」(http://abcpedia.acoustics.jp/bs13_a_q4.pdf)

ケプストラムについては、「ケプストラムって要するに何ですか？」(http://abcpedia.acoustics.jp/bs13_q4.pdf)

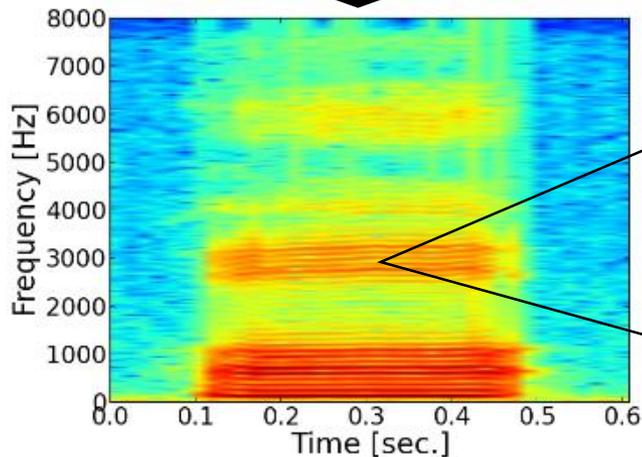
3. 音色に関する特徴量

- 音色は基本的には含まれる音に含まれる周波数成分によって決まる (⇒ **周波数スペクトル**)



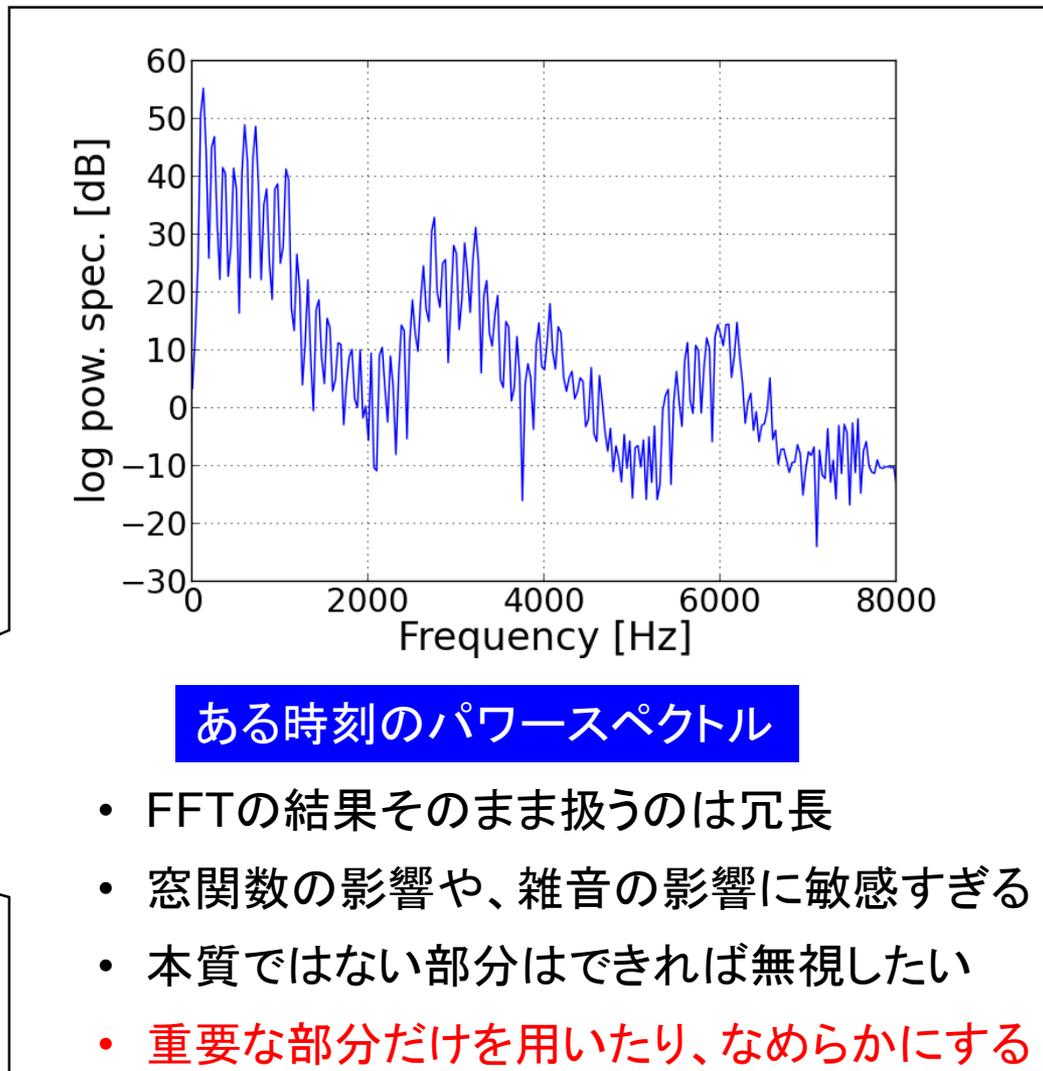
/a/の音声波形

STFT



スペクトログラム

(短時間パワースペクトルを時間方向に並べたもの)



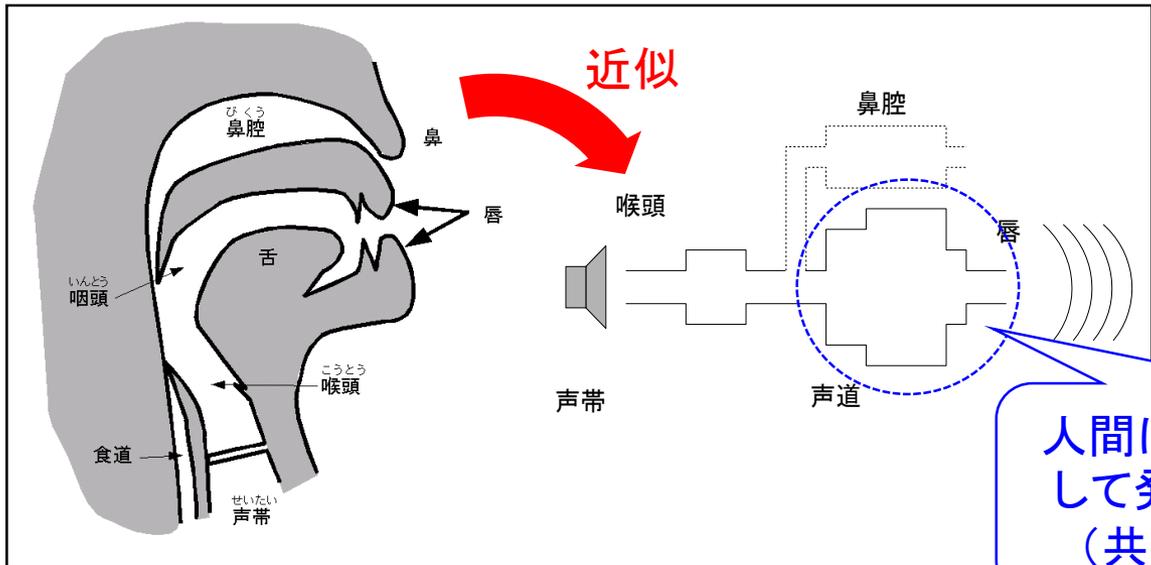
ある時刻のパワースペクトル

- FFTの結果そのまま扱うのは冗長
- 窓関数の影響や、雑音の影響に敏感すぎる
- 本質ではない部分はできれば無視したい
- **重要な部分だけを用いたり、なめらかにする**

3. 音色に関する特徴量

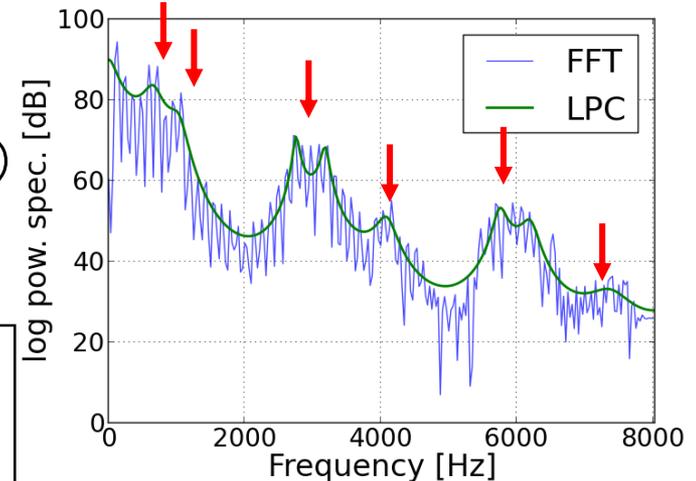
フォルマント周波数

- 音声の特徴量
- 声道の共振周波数のこと(右図赤い矢印の辺り)
- 発声器官は管に近似できる



人間はこの辺りの形を制御して発音を切り替えている(共振周波数が変わる)

音素/a/のスペクトル



- 音素によって特徴的な共振周波数があるため、簡単な音素認識が可能
- 有声音(特に母音)には顕著に現れるが、一部の子音の分別はできないので、現実の音声認識にはあまり使われない

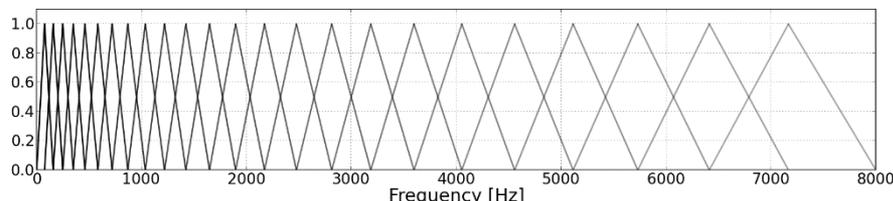
3. 音色に関する特徴量

- 重要なのはスペクトルの全体的な形(概形)
- スペクトルの概形を表現するものとして、現在の音声認識はMFCCがよく使われる
- **MFCC (Mel-Frequency Cepstrum Coefficient, メル周波数ケプストラム係数)**
人の聴覚特性を考慮しながらスペクトルの概形を表現する特徴量。以下で抽出方法(意味)を説明。

STEP 1: メルフィルタバンクをスペクトルに掛ける

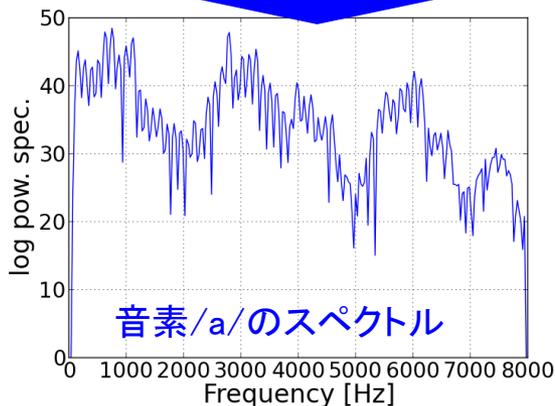
メル周波数ケプストラム係数

メルフィルタバンク



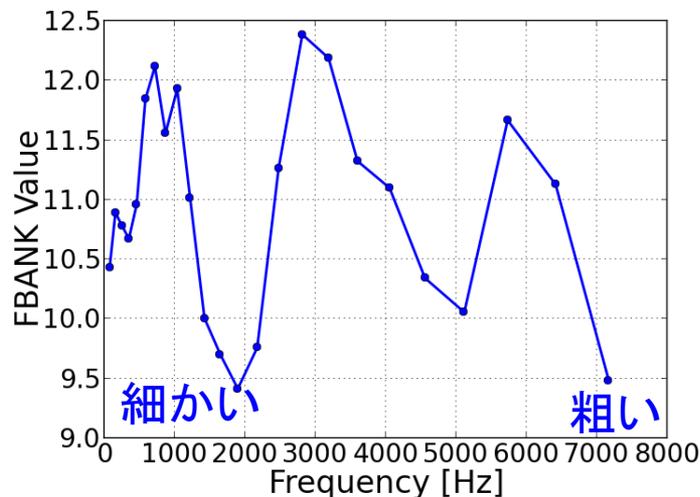
- 7ページで紹介したmel周波数に変換すると等間隔に並ぶフィルタバンク
- 人間の耳の特性を表現(低いところはよくわかる。高いところは雑音)

掛け算



音素/a/のスペクトル

結果



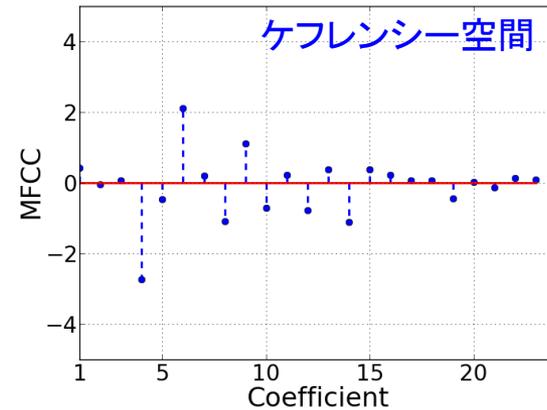
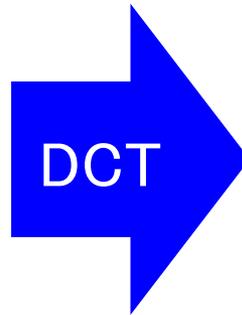
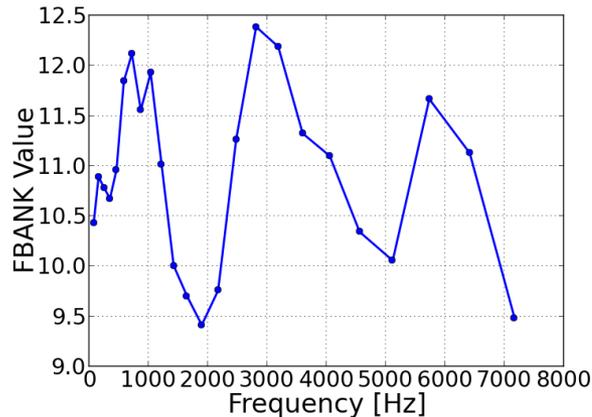
得られたスペクトルの概形

注)ここでは高い方の周波数成分を強調する処理を施している
また、この図は対数スペクトルだが、本当は真値に掛けて対数にする

3. 音色に関する特徴量

STEP 2: スペクトル概形を離散コサイン変換(DCT)する

メル周波数ケプストラム係数



離散コサイン変換する理由

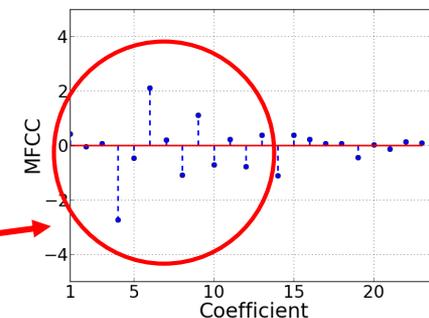
- 係数間の相関が減る(≒特徴量としての性能が向上)
- 概形成分は係数の下の方に集まるので、高い係数を減らせる可能性がある

STEP 3: 下から12個くらいの係数を取り出す

メル周波数ケプストラム係数

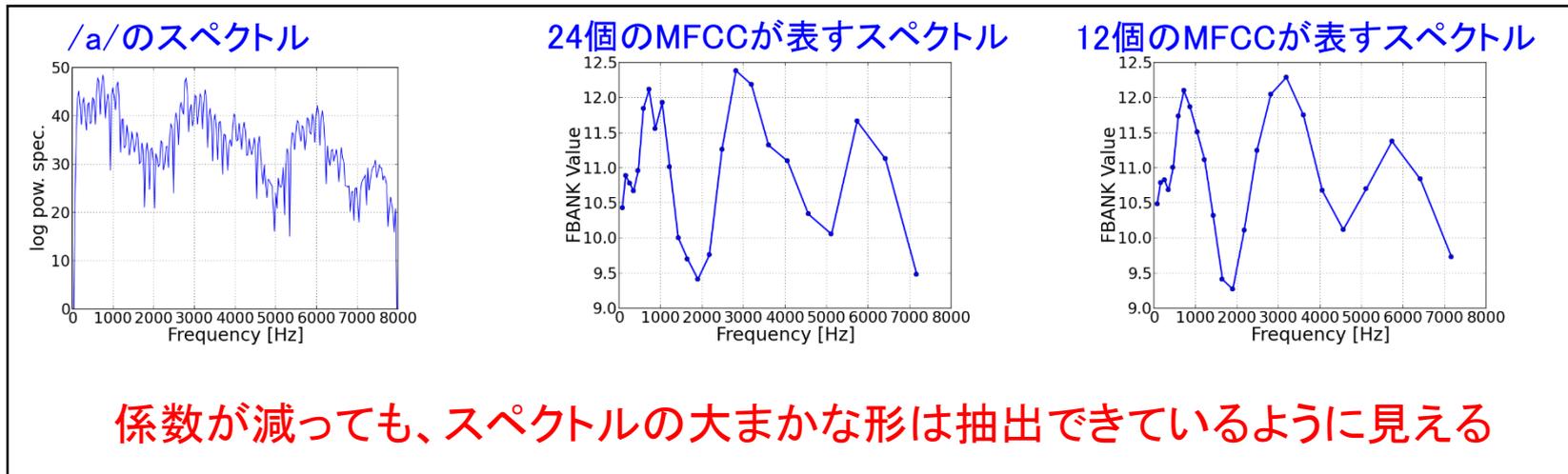
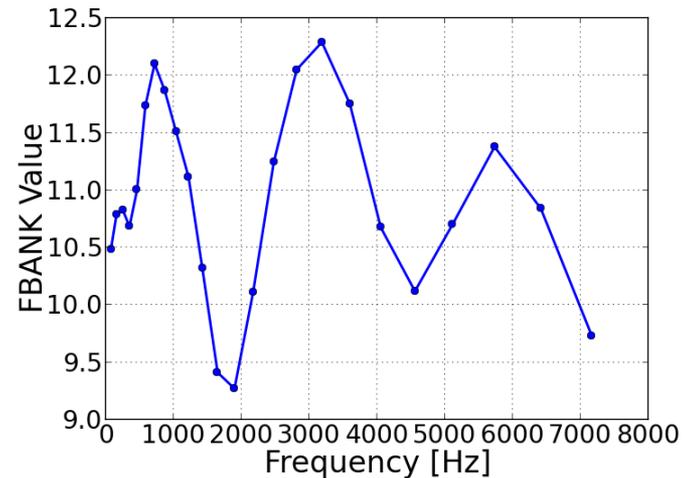
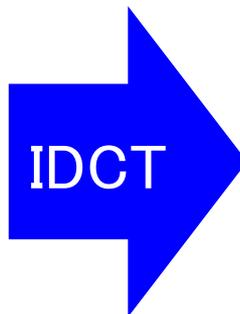
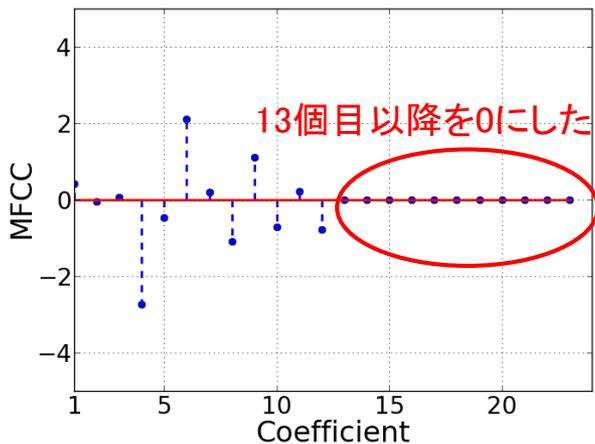
0番目の係数は直流成分で、波形を表していないので除く場合が多い
(0番目の係数はこのグラフにはそもそも書いていない)。

このへんの値を音声認識などに使う



3. 音色に関する特徴量

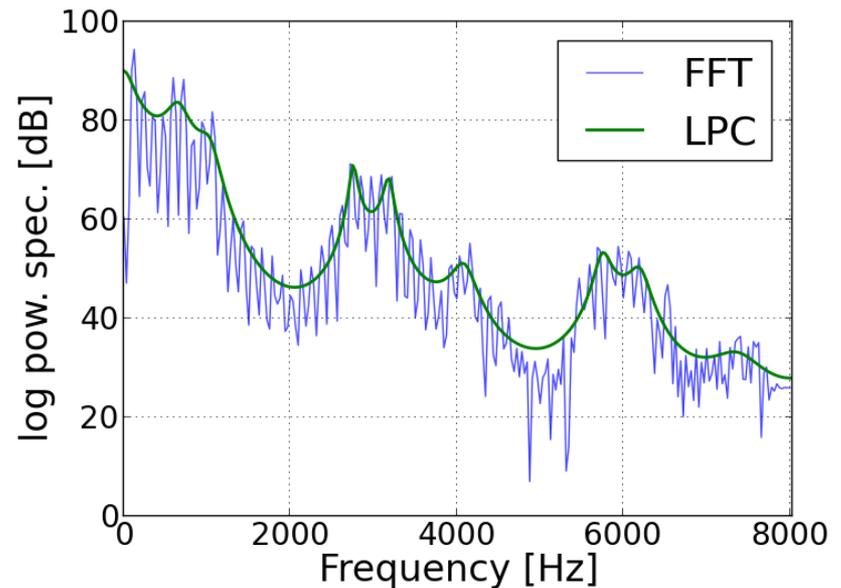
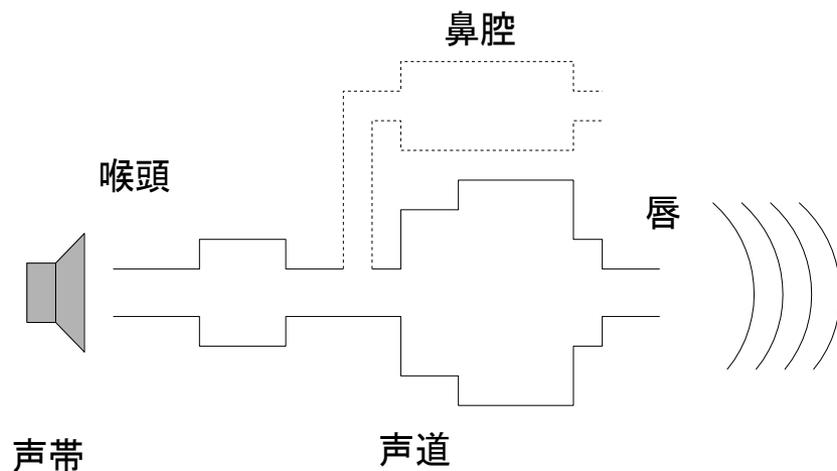
おまけ: 下から12個のMFCCを逆離散コサイン変換してスペクトル概形に戻すと...



3. 音色に関する特徴量

LPC (Linear Predictive Coding, 線形予測符号)

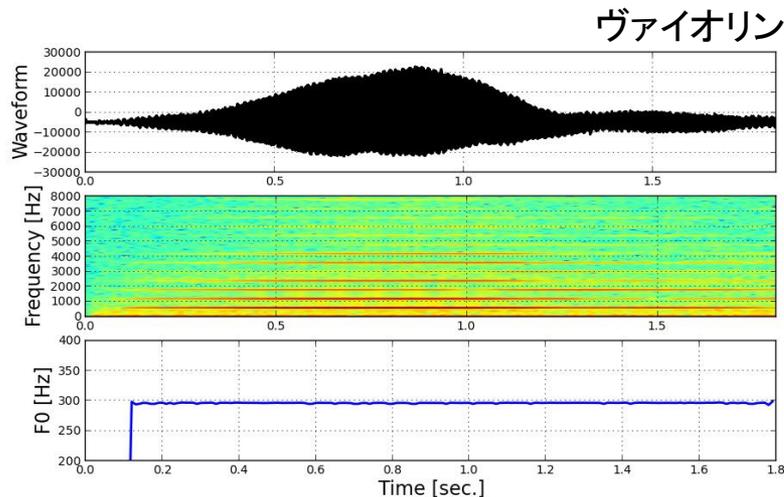
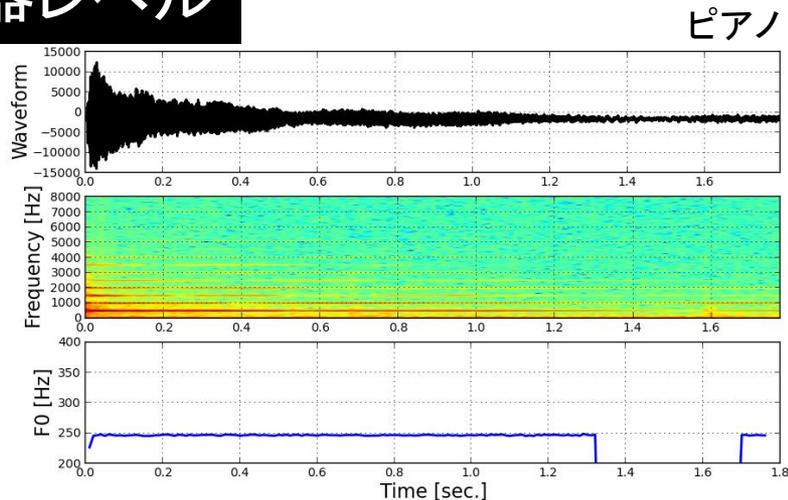
- スペクトル包絡を全極フィルタで近似したもの
- 現在の時刻の音声サンプルを過去のサンプルの重み付き足しあわせ (線形結合) から予測 (重みがフィルタの係数)
- 下図の緑線が予測されたフィルタで、スペクトル包絡が表現されている
- 全極形のパワースペクトルにケプストラムの概念を導入したLPCケプストラムというものもある



声道の共振特性は全極形フィルタで表現すると都合が良い

4. 音楽信号の解析

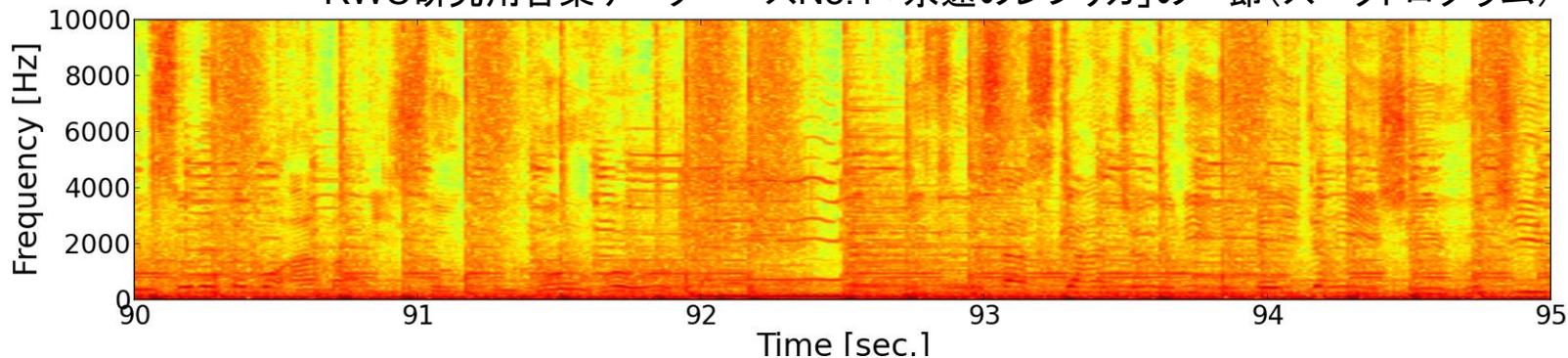
楽器レベル



- 楽器ごとに音響的な性質(調波構造など)は異なる

楽曲レベル

RWC研究用音楽データベースNo.1「永遠のレプリカ」の一節(スペクトログラム)



- 楽曲はもっと複雑。**テンポ**や**コード進行**なども特徴量になる

事例の紹介

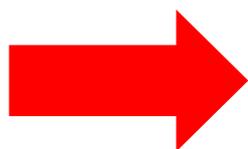
- 最後に、音声・音楽認識関連の典型的な事例について抜粋し、問題解決に有効とされている音の性質をまとめました

	事例	特徴量に組み込む音の性質	今回挙げたもので使えそうなもの
音声	音声認識 話者識別 感情認識 (喜び、怒り、悲しみ・・・)	音色、大きさ 音色 高さ、大きさの変化	MFCC, パワー MFCC 基本周波数, パワー
音楽	音楽情報検索 (ジャンル、アーティスト、楽器・・・) 楽曲間類似度の推定 ハミング検索	音色、音色の時間変化、 高さ、テンポ、コード進行 音色、リズム、高さ 高さ	MFCC, 基本周波数 MFCC, 基本周波数 基本周波数

- あくまで例なので、この問題に対してこのような特徴量を取れば間違いない、ということを示したものではありません
- それぞれの問題に関してもっと効果のある特徴量が考案されている場合もあります
- 表にはテンポやコード進行など、必ずしも同一のレイヤーで議論するべきではない特徴量も含まれています

Q. 様々な音響特徴量

それぞれの使い方や意味を教えてください



特に音声認識周辺で使われている特徴量の代表的なもの、その使い方を紹介

どのような特徴量が良いかは問題設定によって決まるため、
使い方やその種類は無限にあることも事実

データから対象とする問題を上手く説明できそうな特徴量を考えることが重要

例) 人間はどのようにして音が鳴った方向を知覚するのかを考えたいなら、
人間のように音声をステレオで録音
⇒2つの波形の時間のズレを特徴量として分析する、など