

「音響学入門ペディア」作成に向けて

Q4: ケプストラムって要するに何ですか?

What is cepstrum?

岡本 拓磨

情報通信研究機構

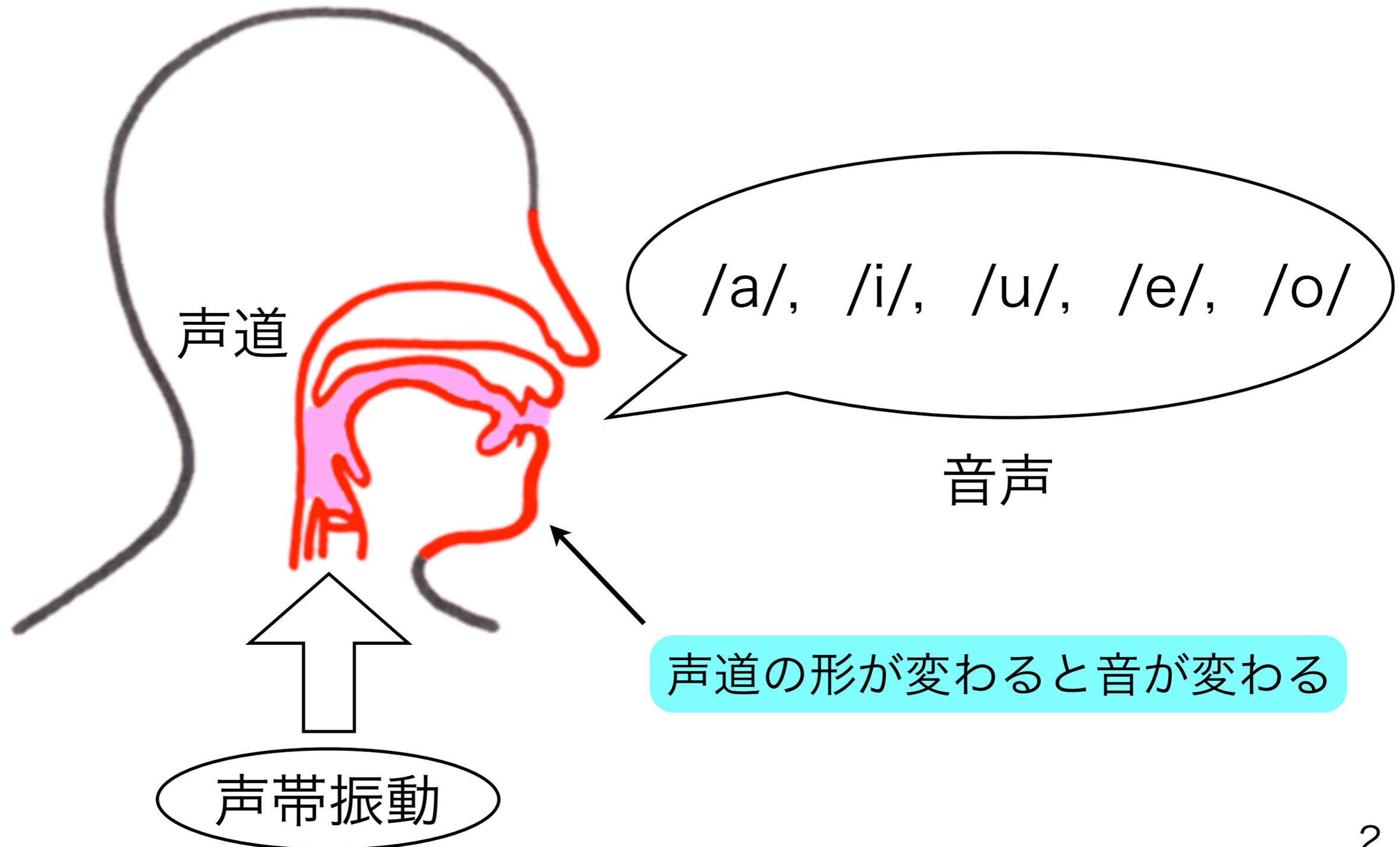
学生・若手フォーラム幹事会

音響学入門ペディア編集委員会

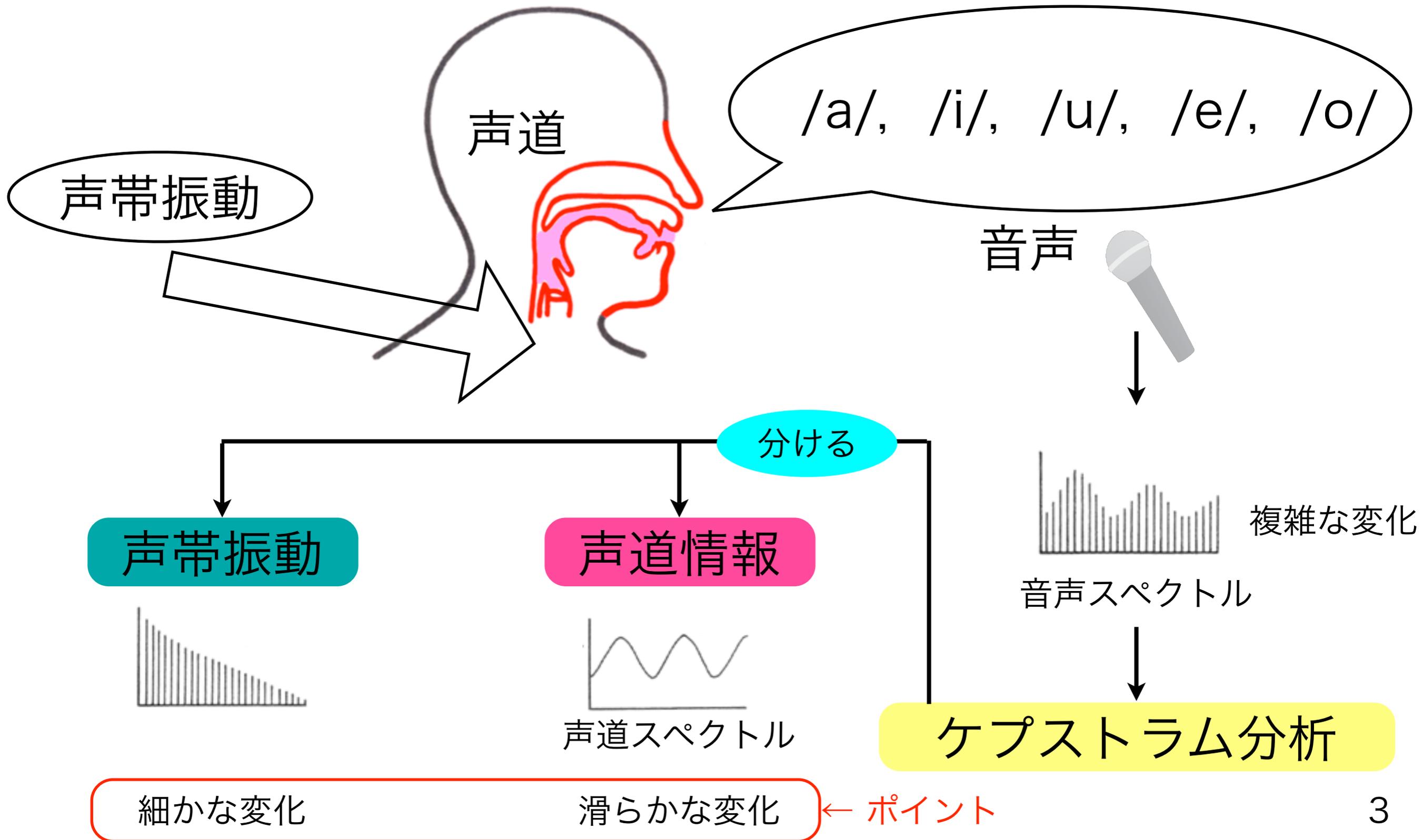
email: [okamoto \(_at_\) nict.go.jp](mailto:okamoto (_at_) nict.go.jp)

HP: <http://www.okamotocamera.com>

声のしくみ



音響学的な「ケプストラム分析」の目的



人の声のしくみ：信号処理として見ると

■ 時間領域

$$a(t) * h(t) = s(t)$$

声帯振動信号 声道フィルタ 音声信号

■ 音源信号と声道フィルタの畳み込み

■ 周波数領域

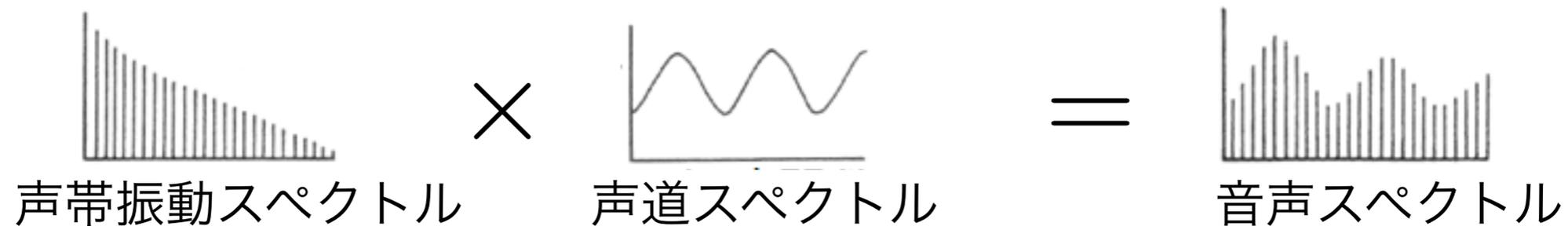
声帯振動スペクトル × 声道スペクトル = 音声スペクトル

■ 音源スペクトルと声道スペクトルのかけ算

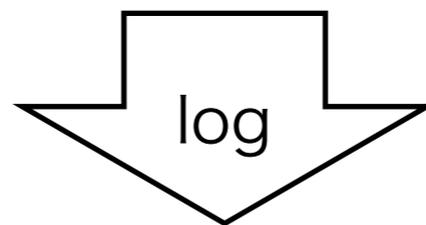
ポイント1：時間領域の畳み込み→周波数領域のかけ算

人の声のしくみ：信号処理として見ると

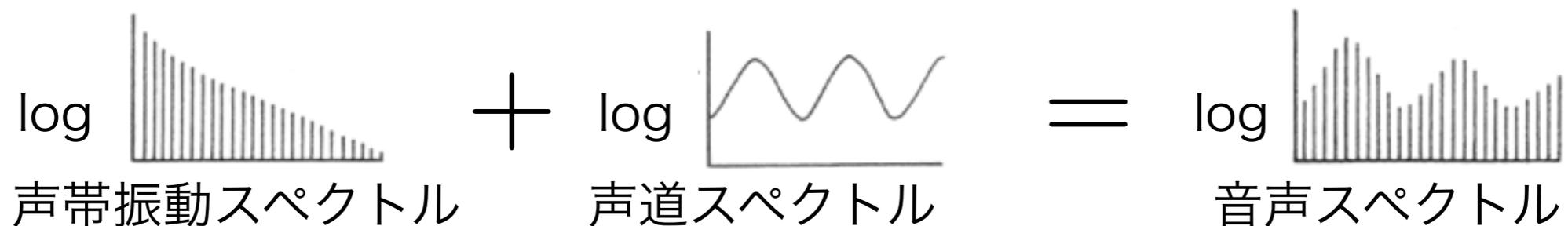
■ 時間領域(縦軸真数軸)



■ 音源スペクトルと声道スペクトルのかけ算



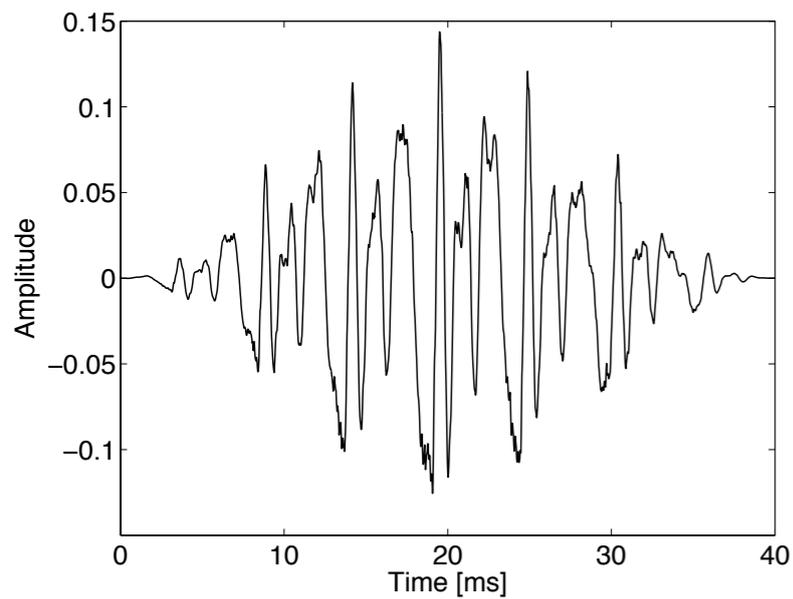
■ 周波数領域(縦軸対数軸)



■ 音源スペクトルと声道スペクトルの足し算

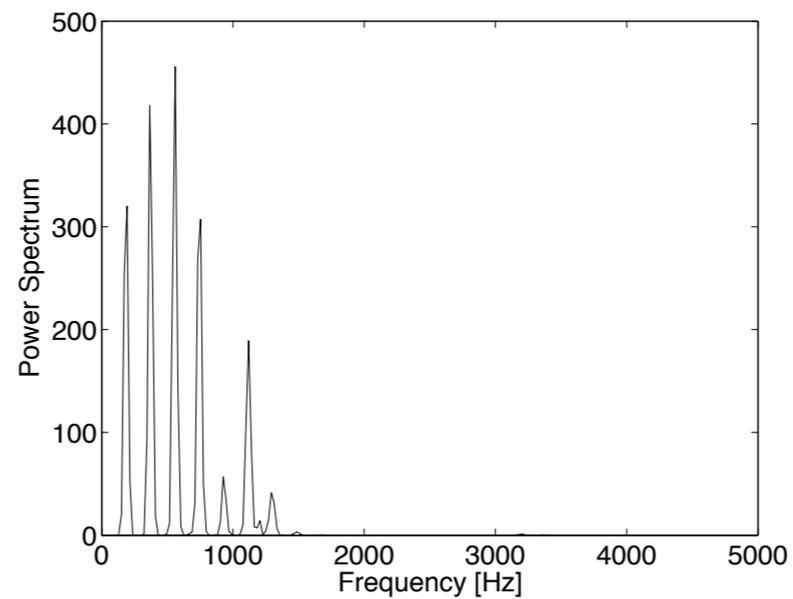
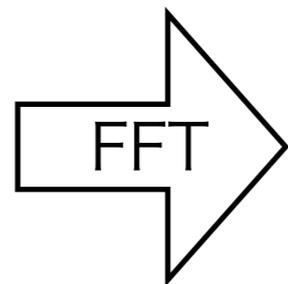
ポイント2：対数領域ではかけ算が足し算になる

前処理



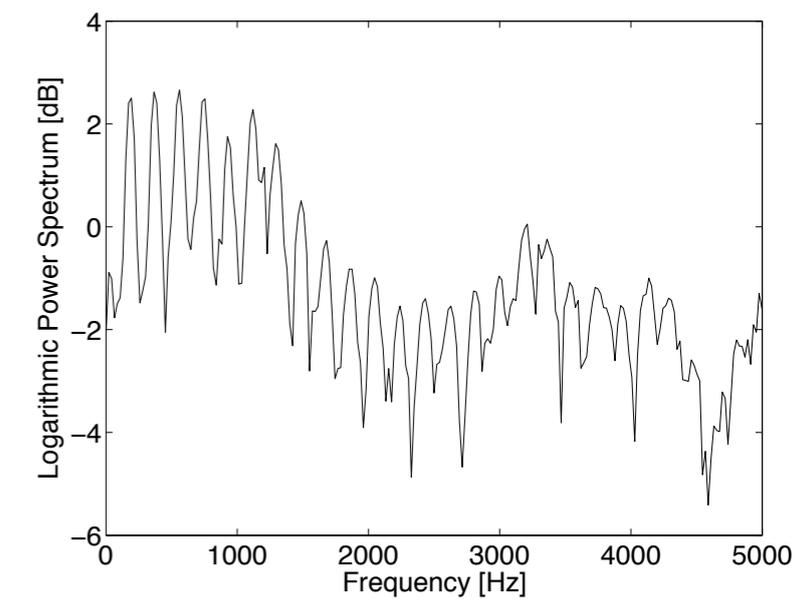
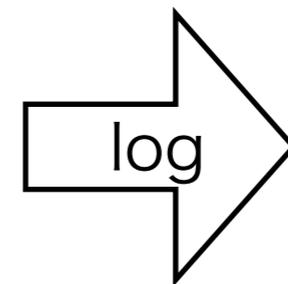
/a/を切り出した時間信号

$$s(t)$$



/a/のパワースペクトル

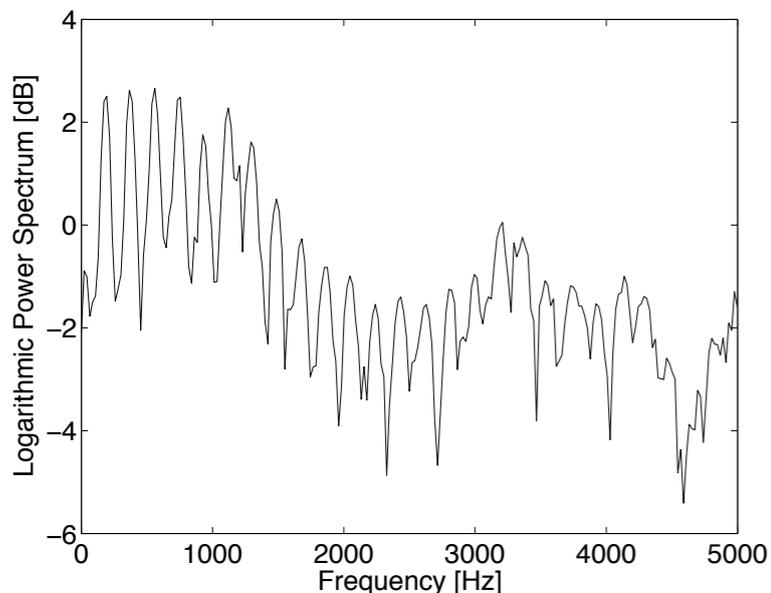
$$|S(f)|^2$$



/a/の対数パワースペクトル

$$\log(|S(f)|^2)$$

2つの成分に分けるには??

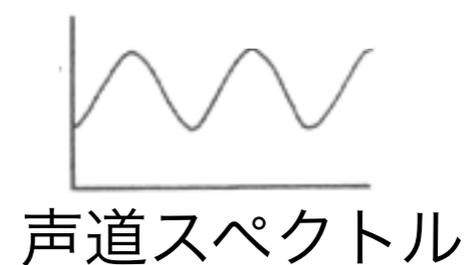
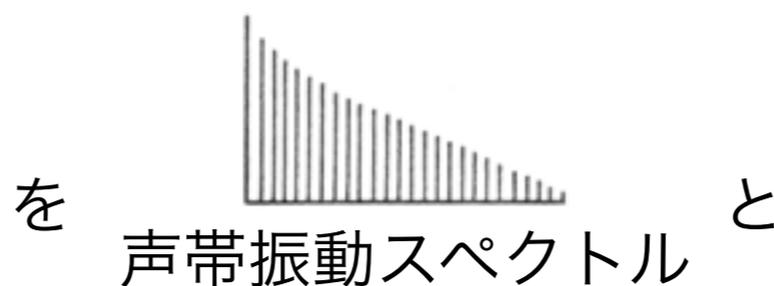


/a/の対数パワースペクトル

$$\log(|S(f)|^2)$$

声帯振動

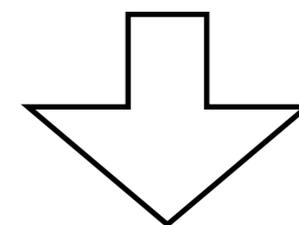
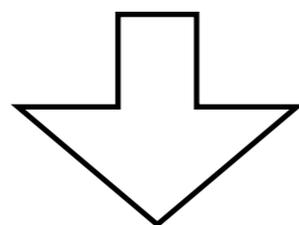
声道情報



に分けたい

細かな変化

滑らかな変化



最大のポイント

もし↑が時間信号だったら...

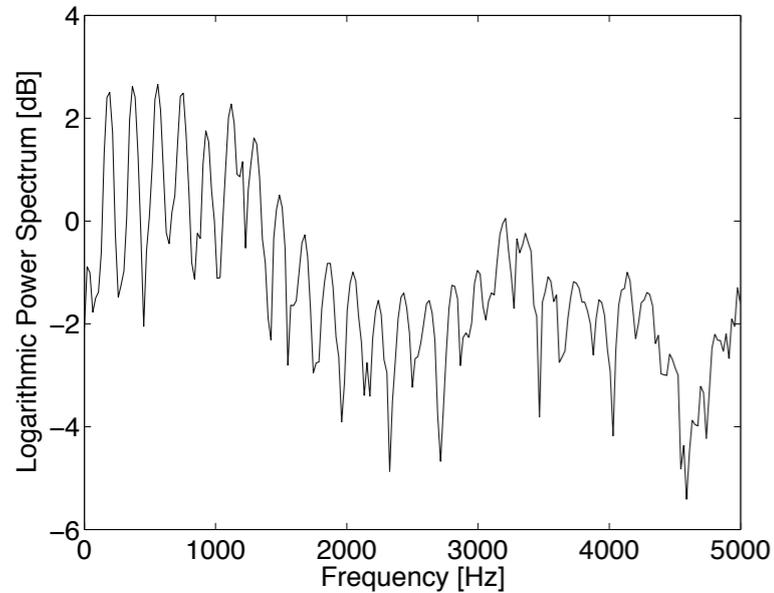
高周波数成分

低周波数成分



$\log(|S(f)|^2)$ をもう一回フーリエ変換すれば分離できる!!

ケプストラム

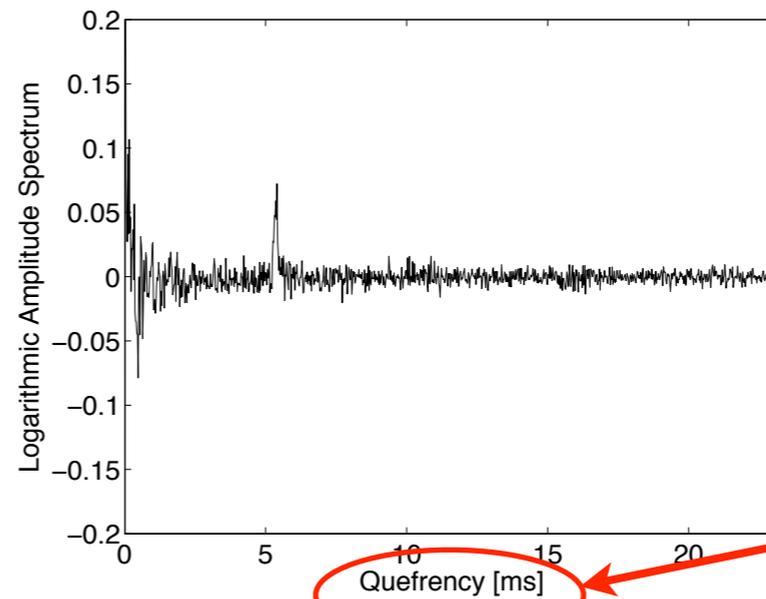
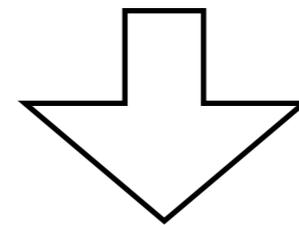


/a/の対数パワースペクトル

$$\log(|S(f)|^2)$$

を時間信号だと思ってFFT!!

※ただ、実際は周波数領域の信号なのでIFFT



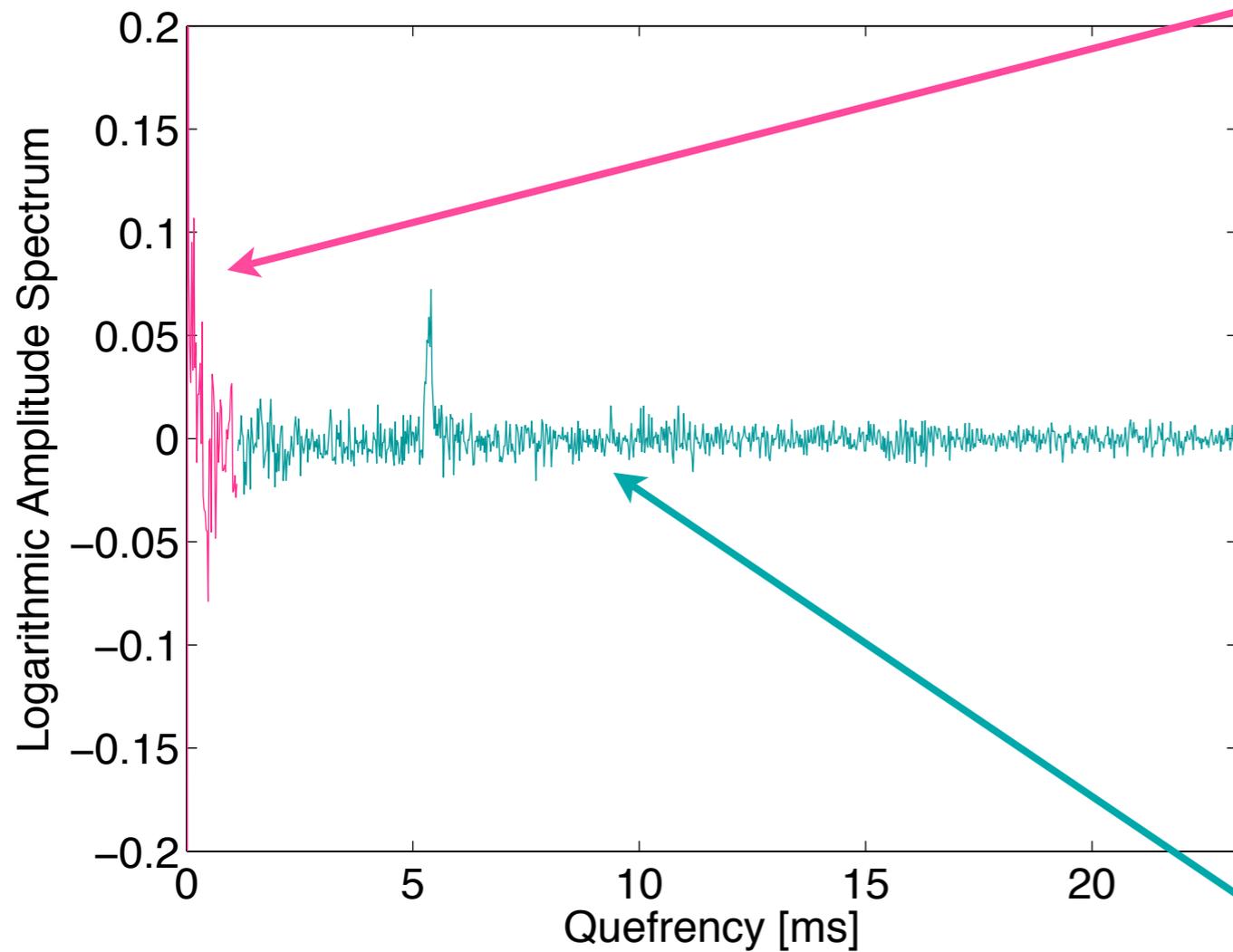
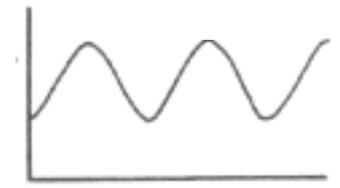
時間領域

/a/のケプストラム

ケプストラム分析

$\log(|S(f)|^2)$ の低周波数成分 →

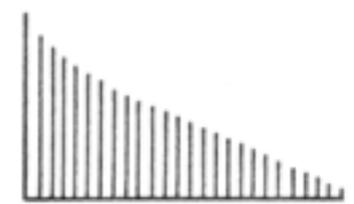
声道情報



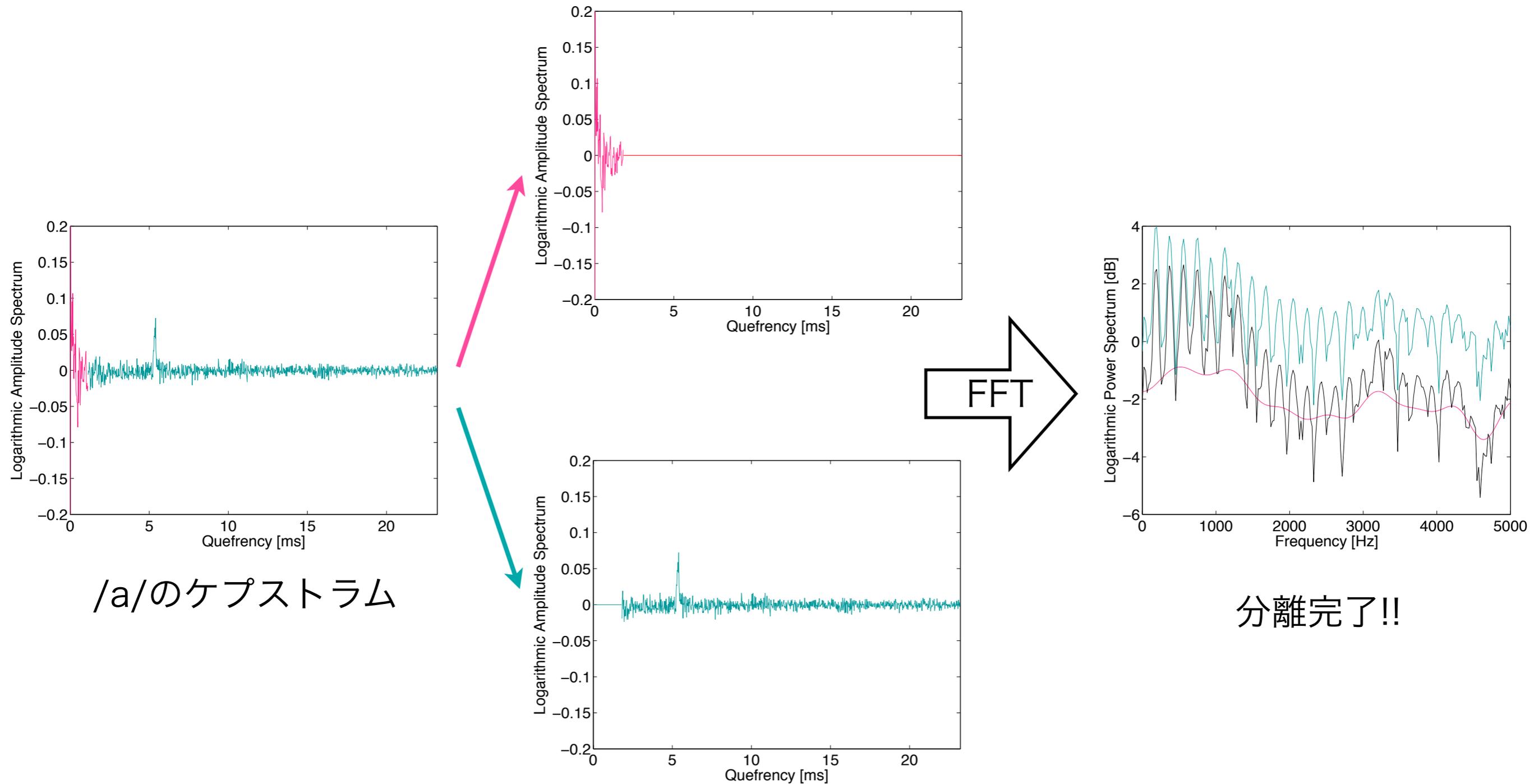
/a/のケプストラム

$\log(|S(f)|^2)$ の高周波数成分 →

声帯振動



ケプストラム分析による情報の分離



/a/のケプストラム

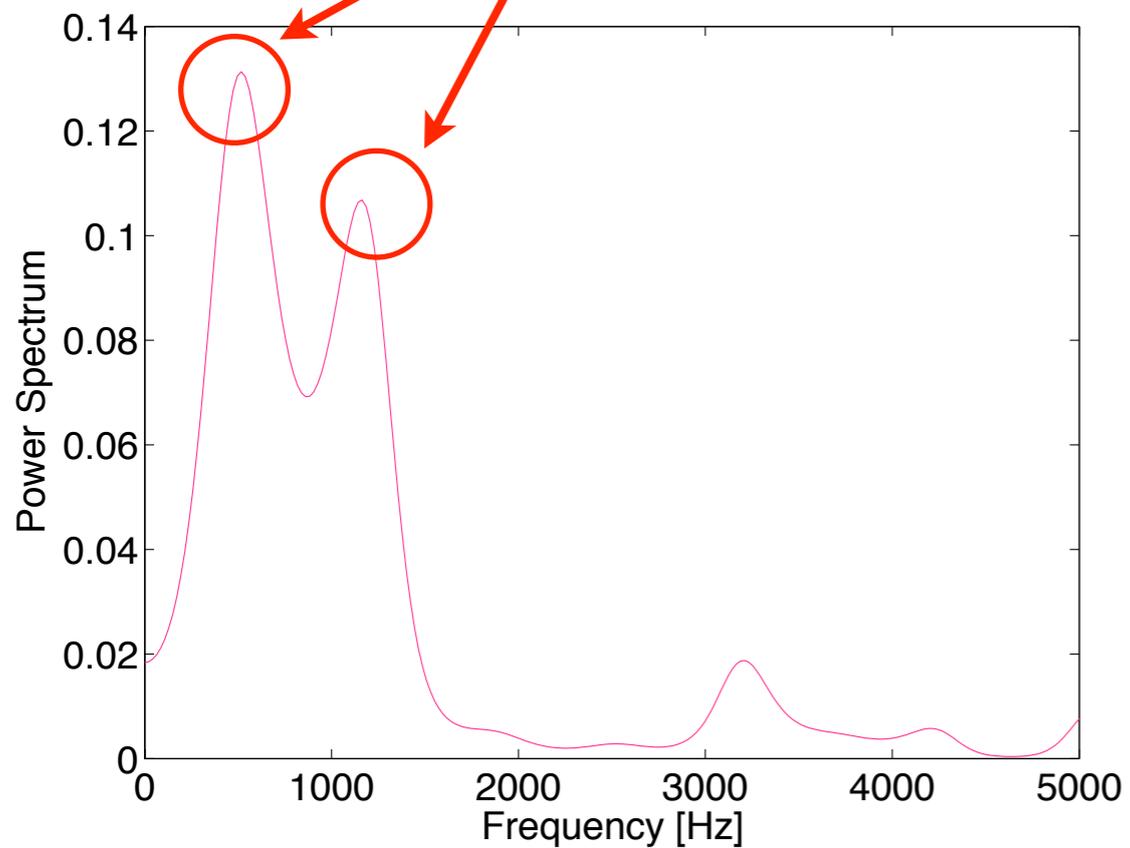
それぞれ0詰め

元に戻す

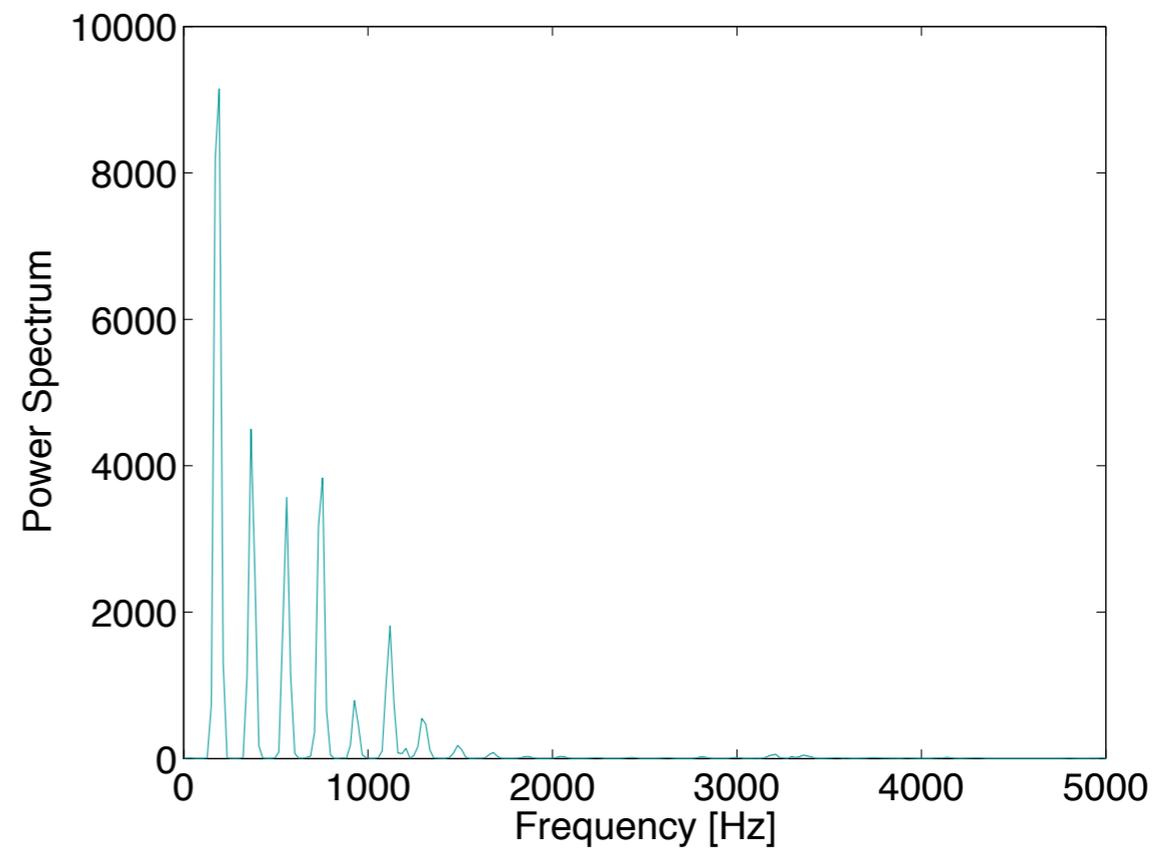
分離完了!!

分離結果

いわゆるフォルマント周波数 → 音声認識などへ



/a/の声道パワースペクトル



/a/の声帯振動パワースペクトル

Q4: ケプストラムって要するに何ですか?

(声帯振動などの)細かな周波数特性の信号が(声道などの)滑らかな周波数特性のフィルタを通過して観測された信号から両者を分離するために



観測信号の対数パワースペクトルをIFFTした信号

3つのポイント

1. 時間領域の畳み込み演算は周波数領域ではかけ算
2. 対数を取るとかけ算が足し算になる
3. 周波数領域の信号であるが時間信号だと思ってFFT

Acknowledgements

■ 正木信夫 (ATR-Promotions / 日本音響学会副会長)

■ デモ機材提供

■ 竹本浩典 (NICT)

■ 声道映像提供