

発声障害者補助のための 統計的声質変換技術の研究開発 (082107009)

研究代表者

○戸田 智基 奈良先端科学技術大学院大学

研究分担者

阪口 剛史 奈良県立医科大学（平成20年度）

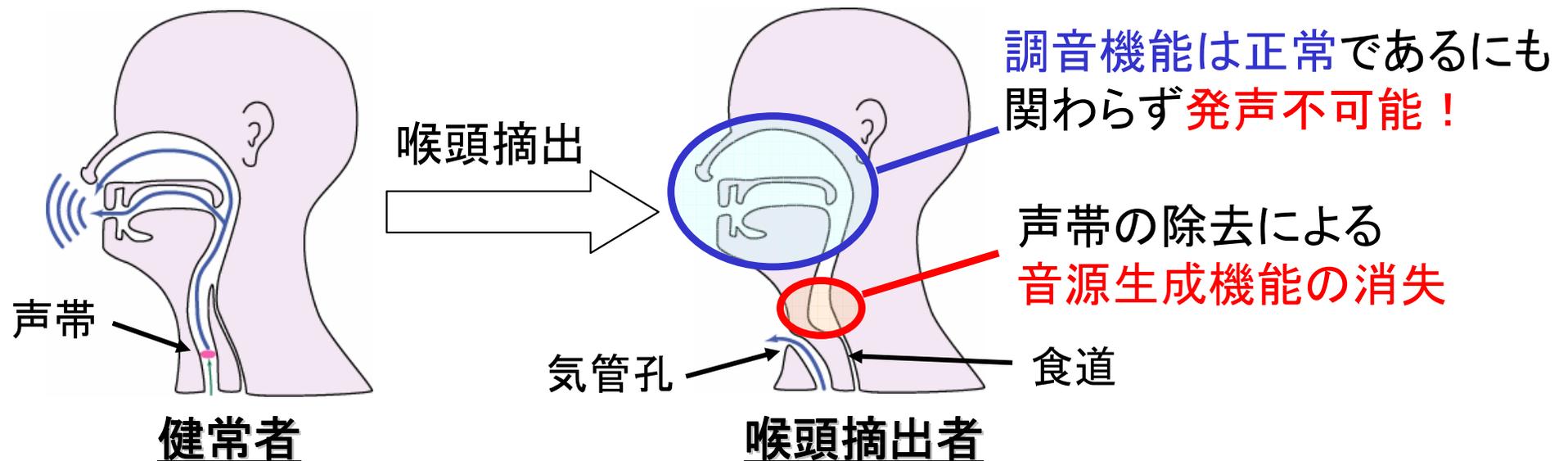
研究期間 平成20～22年度

2011年10月4日

喉頭摘出による発声障害

● 喉頭摘出者

- 喉頭癌等の理由により喉頭を摘出した人
- 全国に約2万人(高齢になるにつれ増加)
- 気管と食道を完全に分離



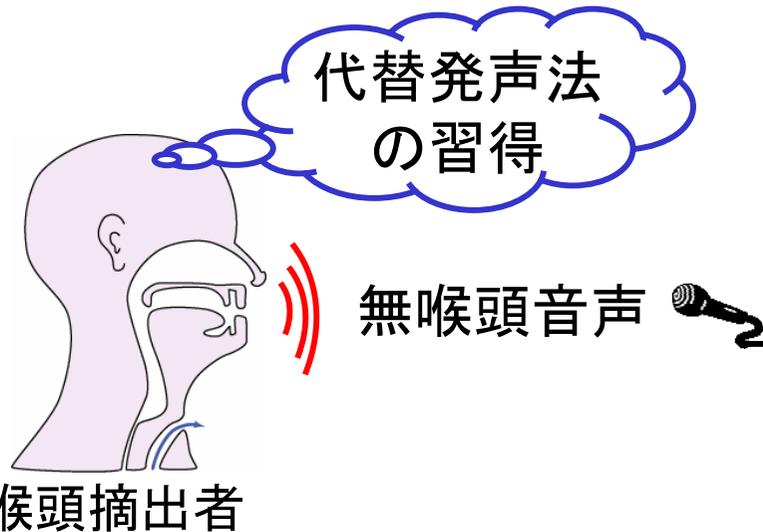
発声能力が失われることによるQoL (Quality of Life)の低下は極めて深刻な問題！

研究目的

喉頭摘出者の音声生成機能を拡張する 情報通信技術の研究開発

既存技術

代替発声法による無喉頭音声生成



健常者の通常音声と比較して

- ・ 自然性及び明瞭性が大きく劣化
- ・ 話者性が大幅に欠落

研究開発技術

無喉頭音声をより自然な音声へと
変換できる強調処理

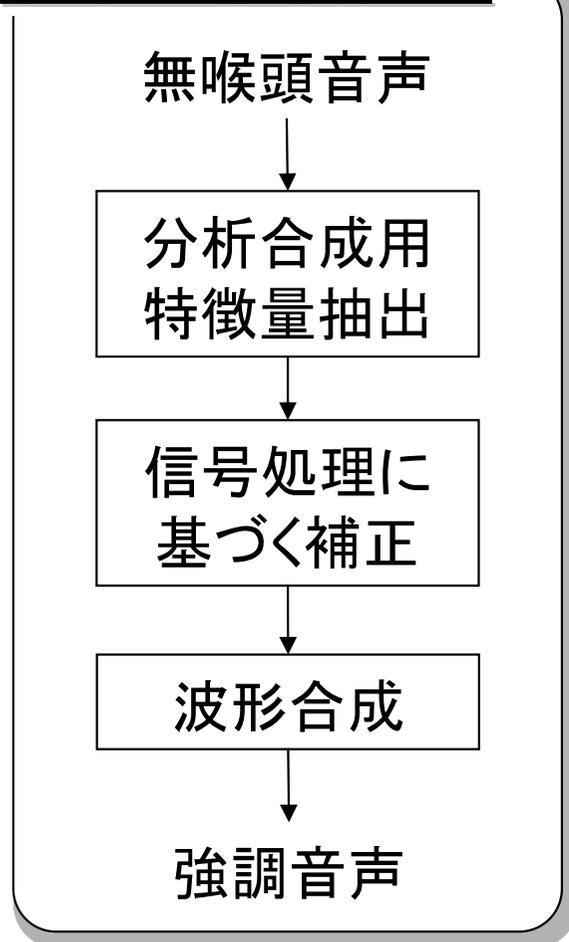


- ・ 自然性, 明瞭性を大幅に改善
- ・ 話者性も柔軟に制御

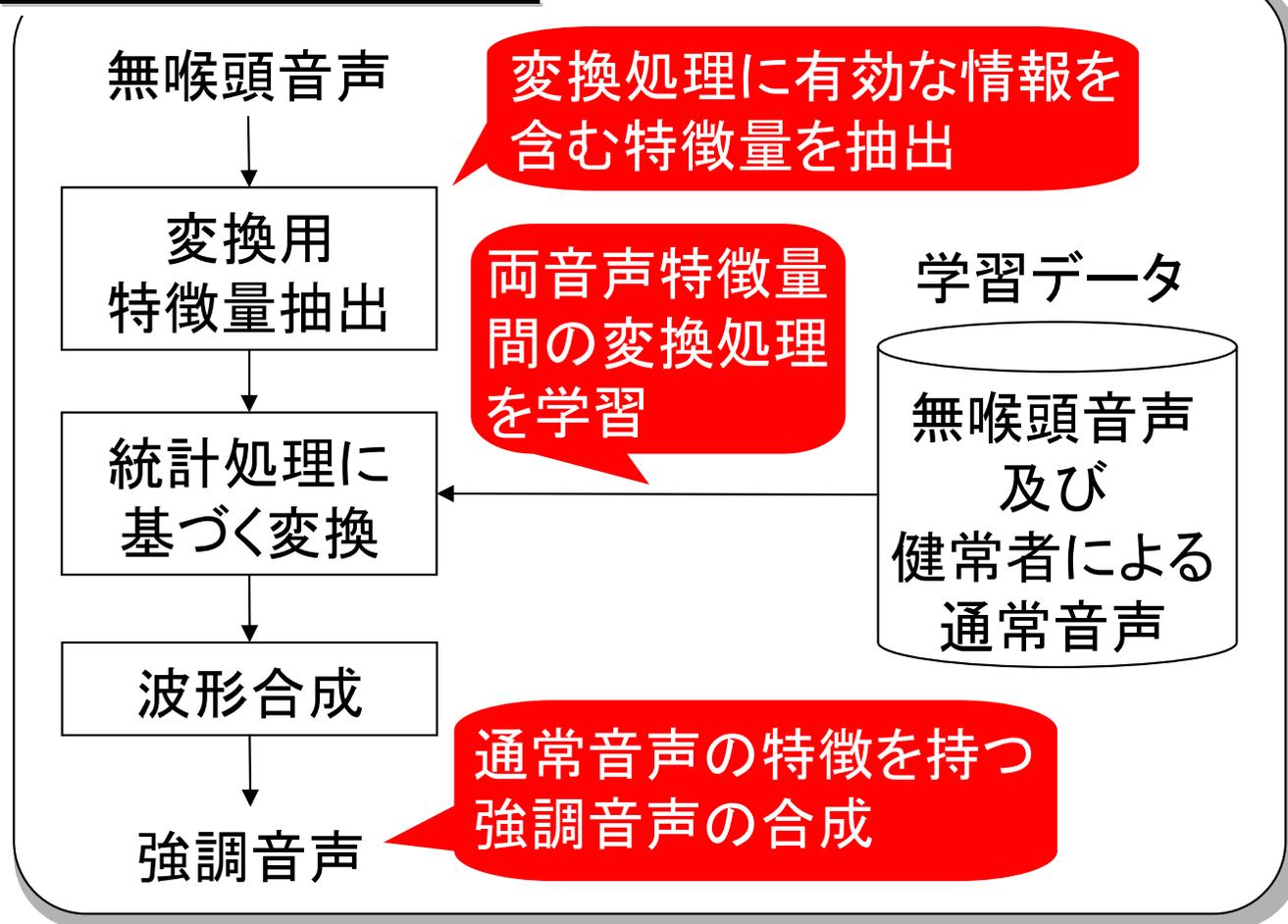
アプローチ

統計的声質変換技術を応用して、無喉頭音声の品質を大幅に改善できる強調処理を構築

従来の強調処理



開発する強調処理



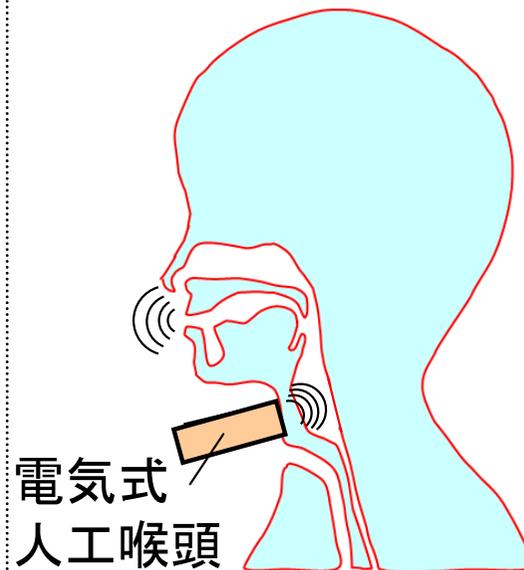
代替発声法

食道発声



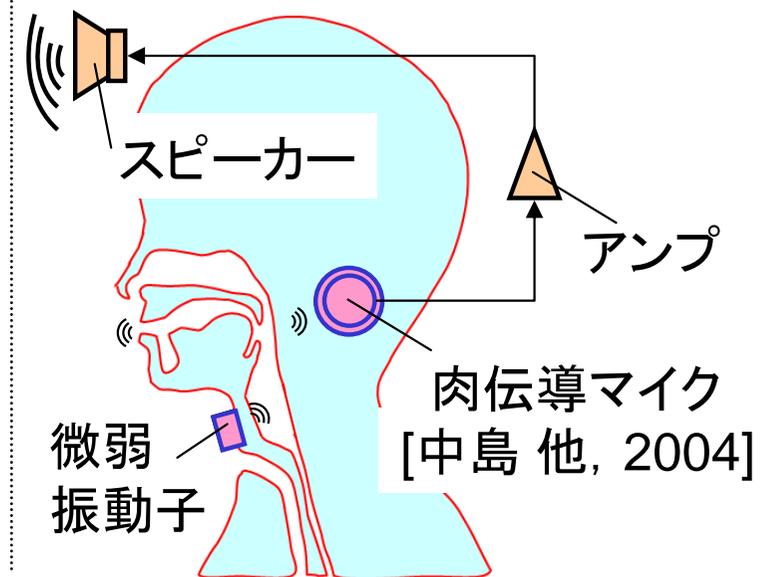
電気式人工喉頭を用いた発声

[橋場 他, 2001]



微弱振動音源を用いた発声

[中村 他, 2007]

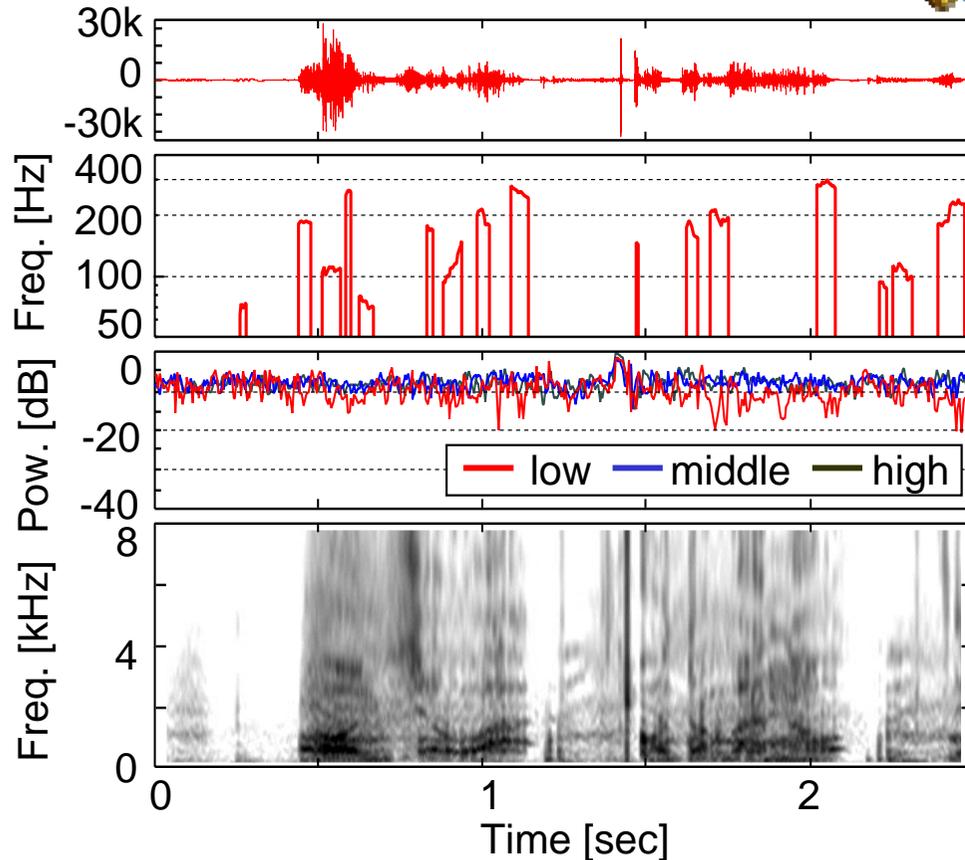


補助器具	不要	必要	必要
習得	困難	容易	比較的容易
伝達される音信号	無喉頭音声	無喉頭音声及び人工喉頭の音源信号	スピーカーから提示される無喉頭音声
無喉頭音声	食道音声	電気音声	肉伝導微弱電気音声

無喉頭音声の音響的特徴の一例

- 食道音声の場合

- 比較的高い自然性及び明瞭性, 乏しい話者性
- 肉声感のある音質
- ピッチの知覚が可能



波形

- 音韻に応じてパワーは変化するが, 不安定

基本周波数 (F_0)

- ピッチに対応する F_0 の抽出は困難

非周期成分

- 極めて強く, 音韻によらずほぼ一定

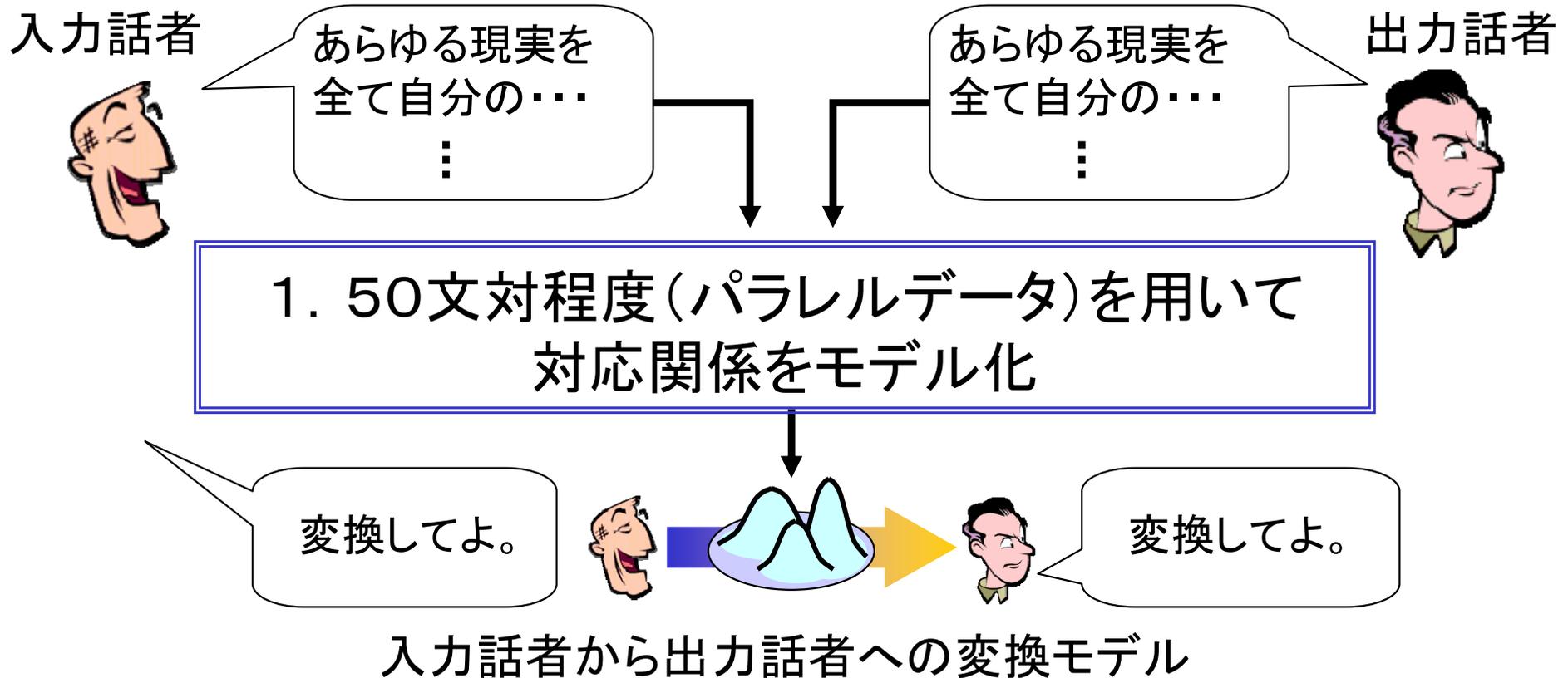
スペクトル包絡 **変換に有用?**

- 音韻に応じて変化するが不安定
- 生成過程で生じる特有の雑音が混入
- ピッチ情報を内包

基礎技術：統計的声質変換 (VC)

[Abe *et al.*, Stylianou *et al.*]

- 学習データから得られる統計的変換規則に基づき、ある話者の声を他の話者の声へと変換する技術

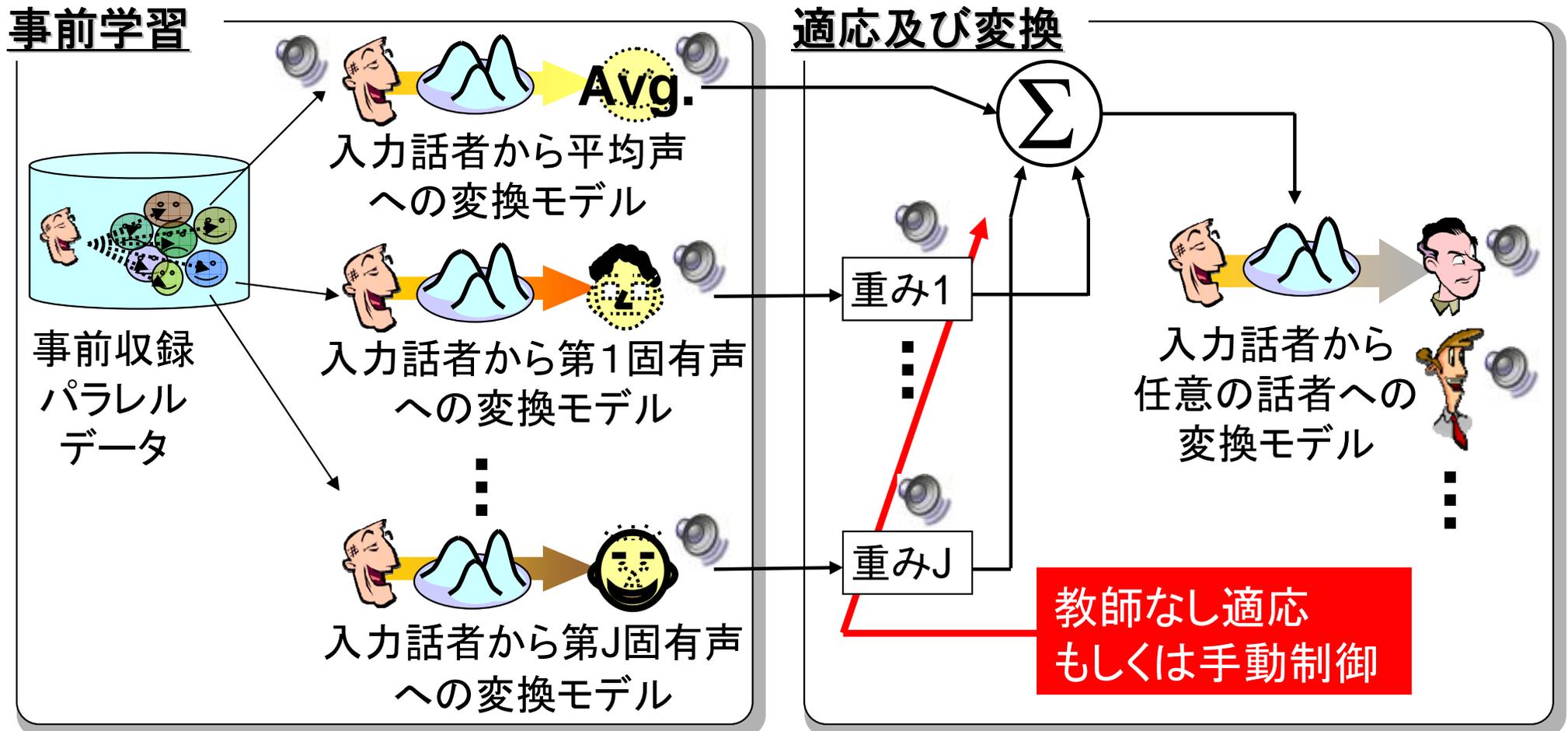


2. いかなる発声も変換可能

基礎技術：一対多固有声変換 (EVC)

[Toda et al.]

- ある話者の声を任意の話者の声へと変換する技術
 - 多数の事前収録出力話者の情報を事前知識として活用
 - 任意の出力話者に対して容易に適用可能な変換モデルを構築

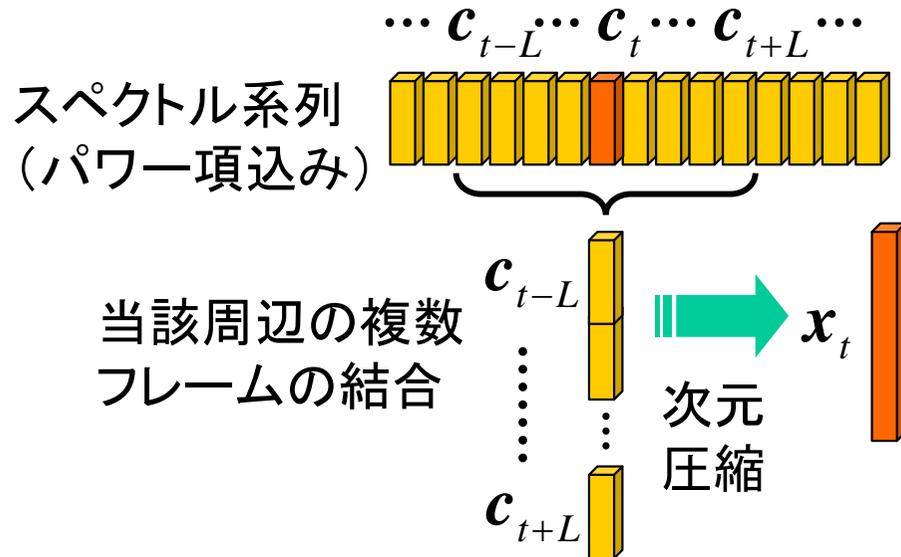


無喉頭音声強調のための声質変換技術

無喉頭音声の音響特徴量

- スペクトル包絡(波形パワー込み)
- ・不安定な変動の除去
- ・音韻情報の欠落

セグメント特徴量の使用!



ささやき声の音響特徴量

- スペクトル包絡(波形パワー込み)

変換モデルの構築!

無喉頭音声の
スペクトル
セグメント



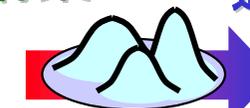
ささやき声の
スペクトル

通常音声の音響特徴量

- スペクトル包絡(波形パワー込み)
- 非周期成分
- 基本周波数(有声/無声情報込み)

変換モデルの構築!

無喉頭音声の
スペクトル
セグメント



通常音声の
スペクトル



非周期成分

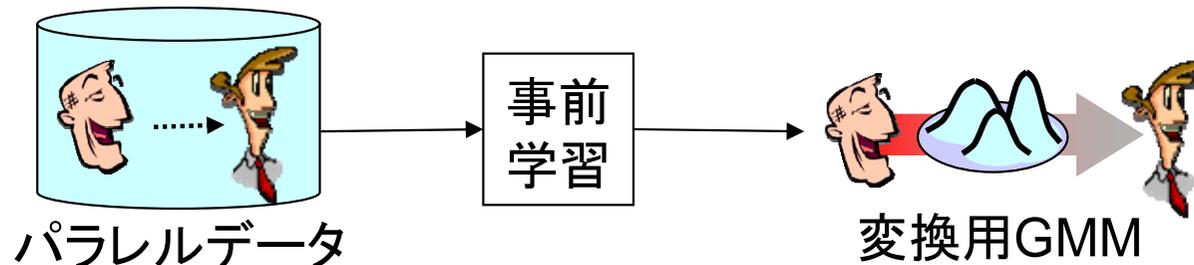


基本周波数

声質制御技術の導入

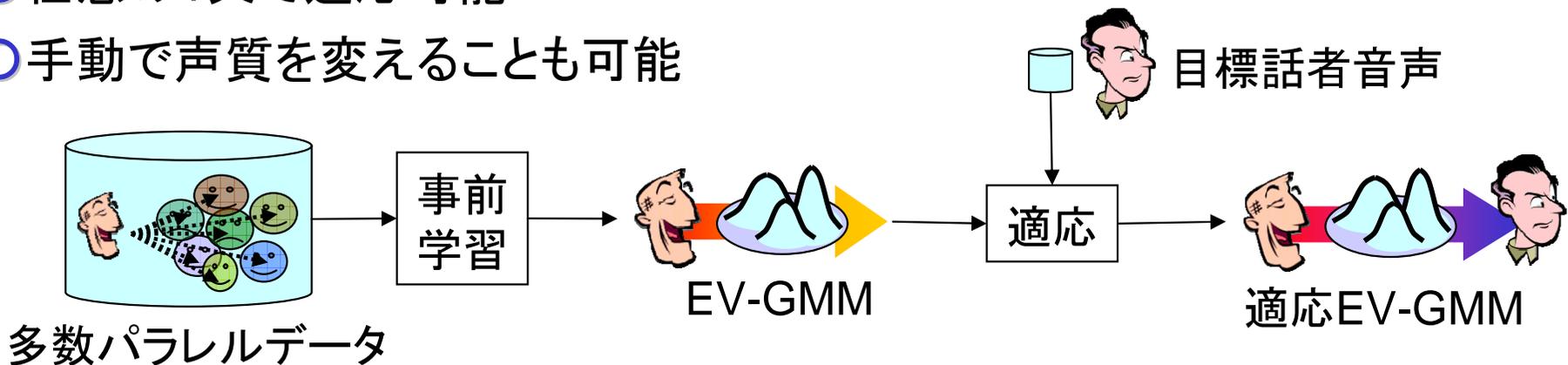
• 通常のVCに基づく無喉頭音声強調

- × 変換音声の話者性は学習に用いた出力話者に依存
- × 学習には50文対程度からなる平行データが必須



• 一対多EVCに基づく無喉頭音声強調

- 所望の声質を持つ話者の音声データに対してモデル適応
- 任意の1文で適応可能
- 手動で声質を変えることも可能



強調結果の一例

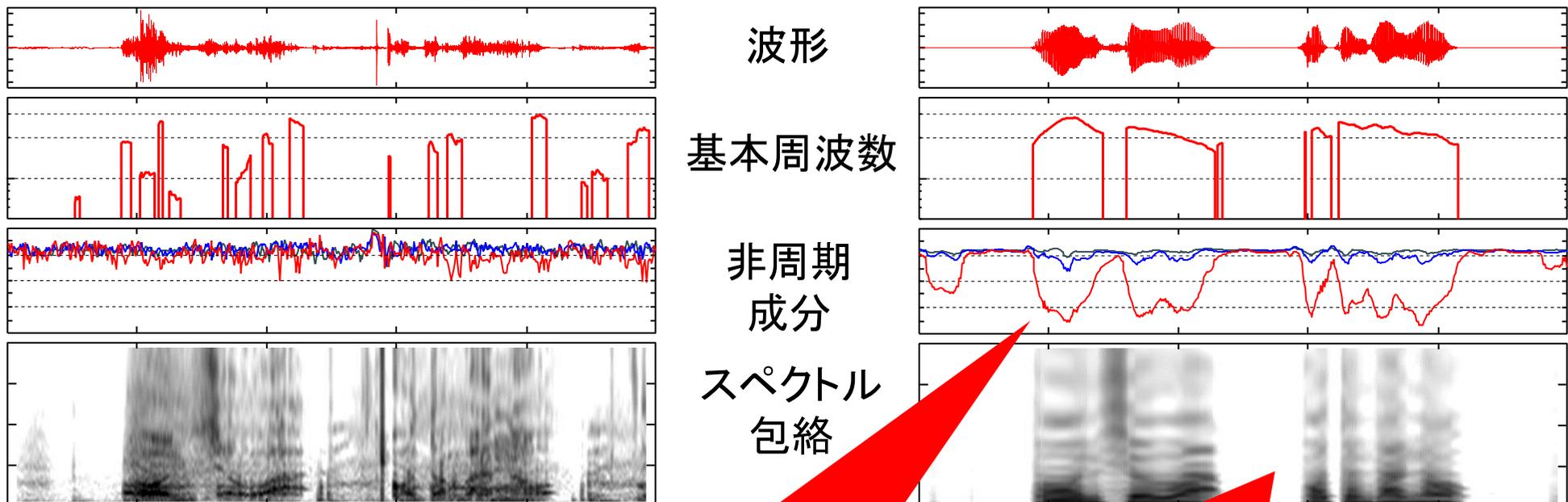
- 食道音声に対する一対多EVCによる強調処理

- 食道音声のピッチに対応した F_0 を推定するために、健常者によるピッチ模擬音声  を出力としてモデル学習
- 目標話者の任意の発声  を用いてモデル適応

目標話者の声質に類似した音声へと変換可能

食道音声 

変換音声 



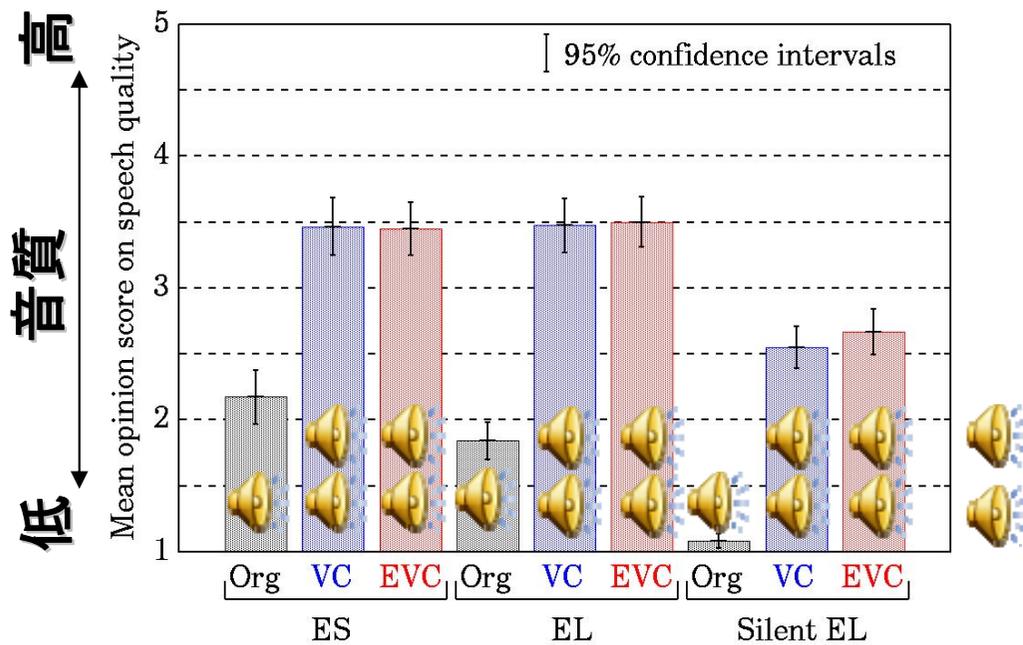
音源特徴量(基本周波数や非周期成分)も推定可能

不安定なスペクトル包絡変動や雑音成分を除去

テストベッドシステムの構築および評価

- 音声データベースの構築
 - 喉頭摘出者Aによる食道音声
 - 喉頭摘出者Bによる電気音声および肉伝導微弱電気音声
 - 健常者75人(男性48人, 女性27人)による通常音声とささやき声
 - 発声内容は音素バランス50文(文中ポーズ位置を上記食道音声と一致させることで, 精度の高いパラレルデータを構築)
- 肉伝導微弱電気音声用の補助装置の試作
- 変換システムの構築
 - 健常者10人を目標話者(実現したい話者性)として使用
 - 40文対/各目標健常者を用いてVCに基づく変換モデルを学習
 - 40文対×健常者30人を用いてEVCに基づく変換モデルを学習し, 任意の1文/各目標健常者を用いて変換モデルを適応
- 事前評価の結果および喉頭摘出者の方のコメントを考慮して, ささやき声への変換ではなく, 通常音声への変換を採用

評価結果



無喉頭
音声

ES: 食道音声

EL: 電気音声

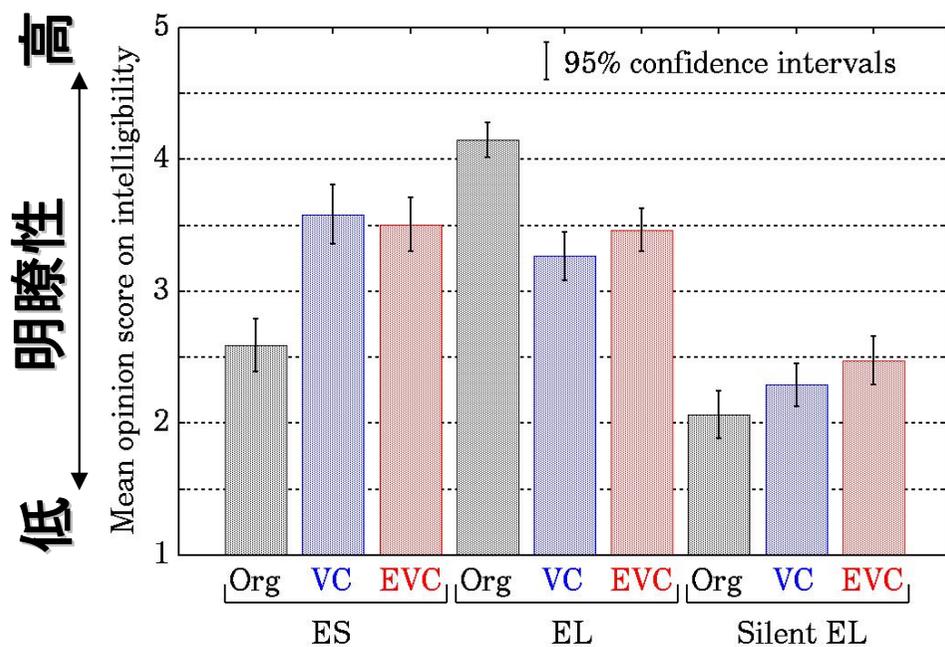
Silent EL: 肉伝導微弱電気音声

各手法

Org: 強調処理なし(従来法)

VC: VCによる強調音声

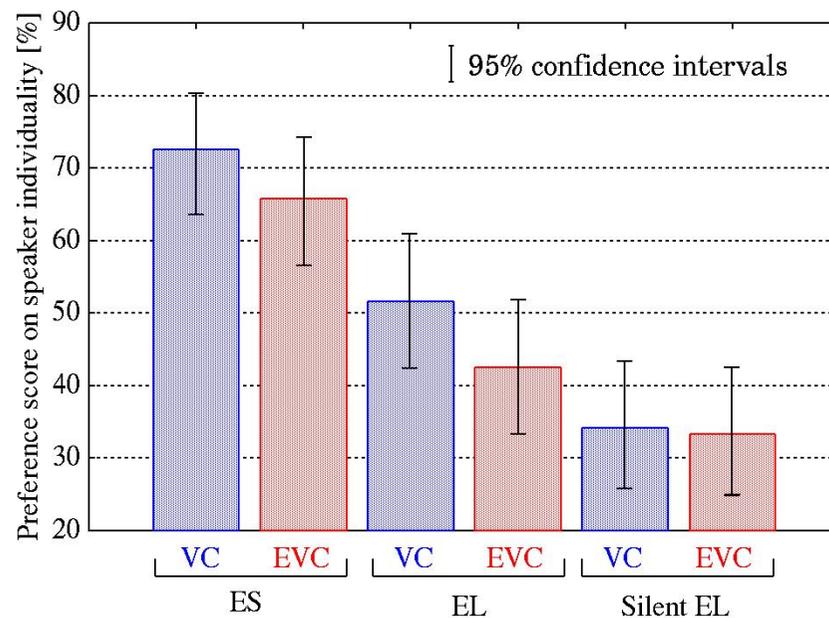
EVC: EVCによる強調音声



話者性

高

低



喉頭摘出者からのコメント:良くなっている!さらに明瞭性を改善して欲しい。

本研究開発による成果

- 本委託研究に基づく発表成果

査読付き学術論文誌		5
海外論文誌	1(2012年1月発行)	
国内英文論文誌	4	
査読付き国際会議論文		17
特別セッション基調講演論文	1	
招待形式特別セッション論文	2	
その他会議		32
海外招待講演	1	
国内招待講演	1	

- 受賞 4件

- 国内 3件
- 国際会議学生論文賞 1件

まとめ

統計的声質変換に基づく無喉頭音声強調技術の研究開発

代替発声法 + 提案技術

1. 食道発声 + 強調

◎外部機器が不要

×修得が困難

- より自然で明瞭な音声
- 任意の1文による声質設定

電話等で効果大



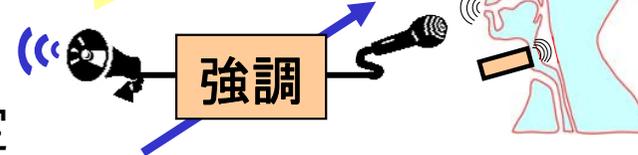
2. 電気式人工喉頭を用いた発声 + 強調

×低い静寂性(人工喉頭の音漏れ)

◎容易に修得可能

- より自然な音声
- 任意の1文による声質設定

電話等で効果大



3. 微弱振動音源を用いた発声 + 強調

◎高い静寂性

◎比較的容易に修得可能

- △比較的自然而明瞭な音声
- 任意の1文による声質設定

対面会話で効果大



今後の課題

- 明瞭性の改善
 - 声質変換アルゴリズムの改良
- リアルタイム変換装置の開発
 - 計算量の削減およびDSP上への実装
- オンライン適応処理の導入
 - 使用時における環境変化への対応
- 実環境への適用