

音・声の認識技術とその応用

名古屋工業大学大学院 准教授

李 晃伸



教育・研究の理念

「ひとづくり」

「ものづくり」

「未来づくり」

私の研究分野：音声認識・音声理解

認識アルゴリズム

- 大語彙探索アルゴリズム
- リアルタイム認識処理
- 並列分散音声認識



音声対話システム

- 統計的QA
- 音声対話ロボット
- 公共案内エージェント

汎用認識ソフトウェア

- 大語彙連続音声認識エンジンJulius



音声インタフェース

- 雑音環境下音声認識
- 音声 UI の感性評価

言語モデル

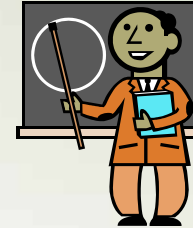
- 話し言葉の言語モデル
- 統計言語モデル



本日の内容

① 技術解説

- 機械による声の分析
- 音声認識のしくみ



② ソフトウェア紹介

- 汎用音声認識エンジン Julius



③ 「環境音認識」の事例紹介



実演を交えてお話します

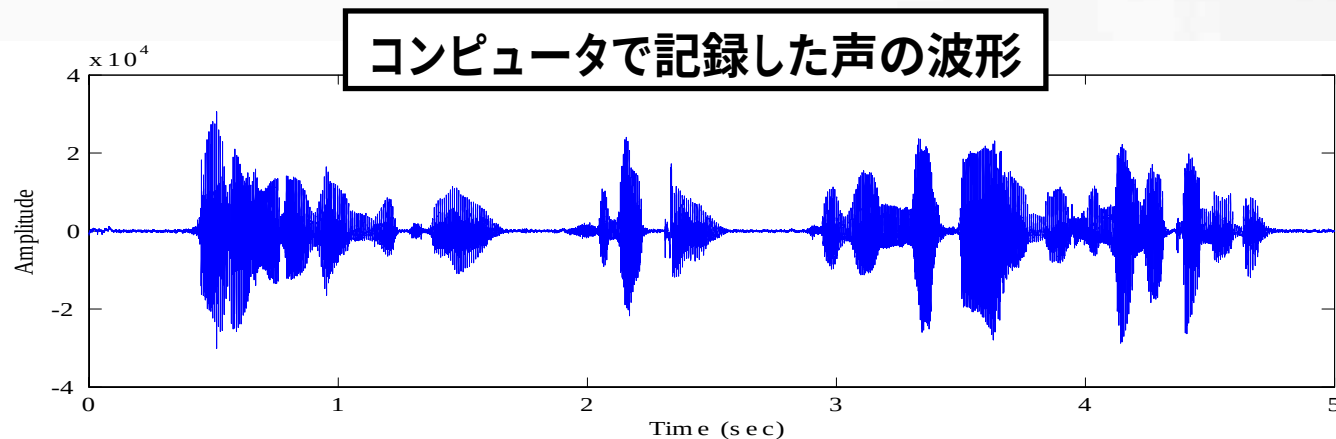
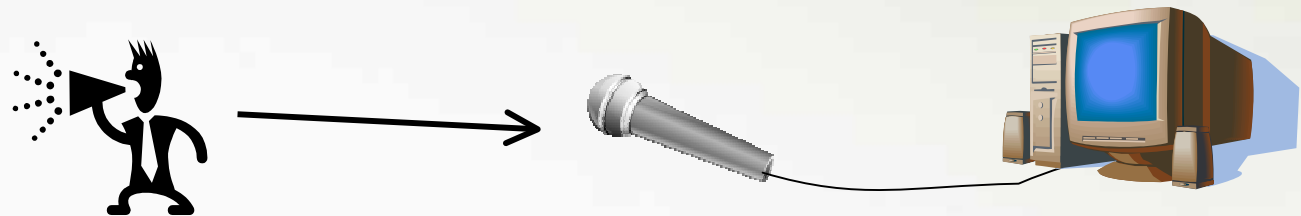
①基礎技術解説

機械による音・声の分析と認識

デジタルサンプリング

マイク: 空気の振動を電気信号に変換

コンピュータ: 電気信号を量子化・サンプリング
→ デジタルデータに変換

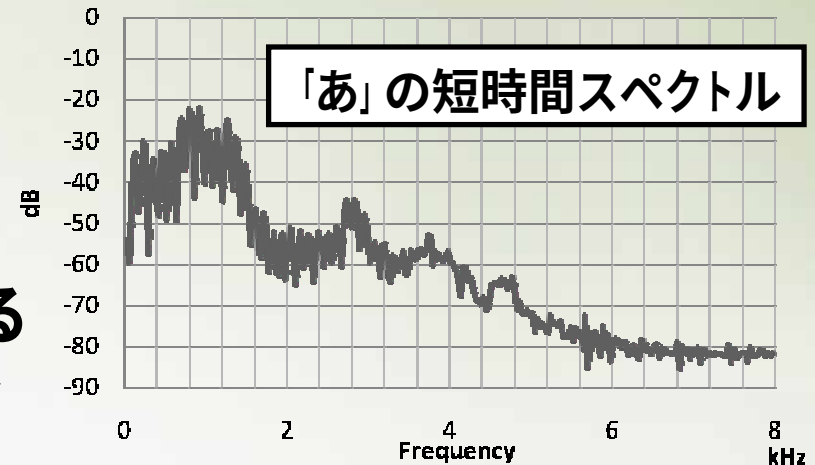


スペクトル分析

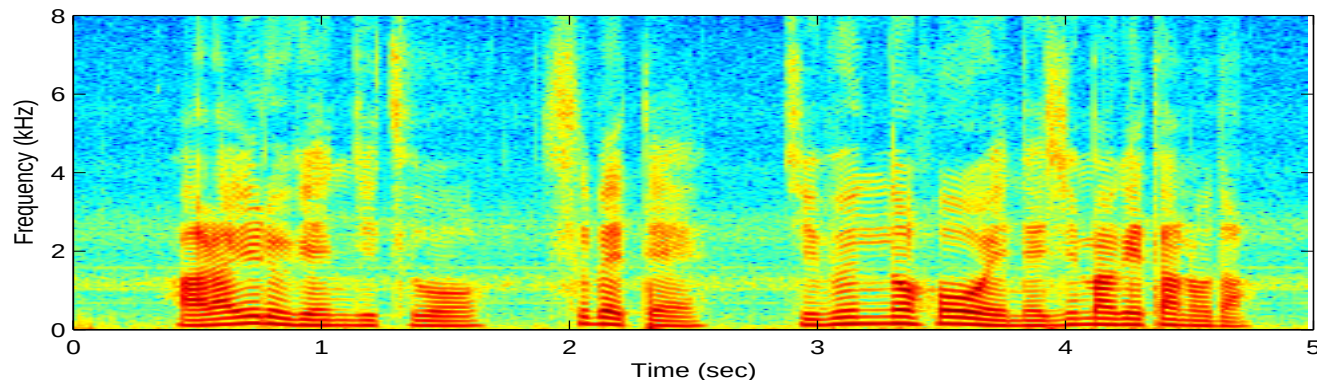
波の信号を周波数領域に変換

- ・ 低い成分＝低い音
- ・ 高い成分＝高い音

スペクトルやスペクトログラムを見ることで、波がどのような周波数成分からなっているのか分析できる。



スペクトログラム：短時間ごとのスペクトルを濃淡で時間順に並べたもの



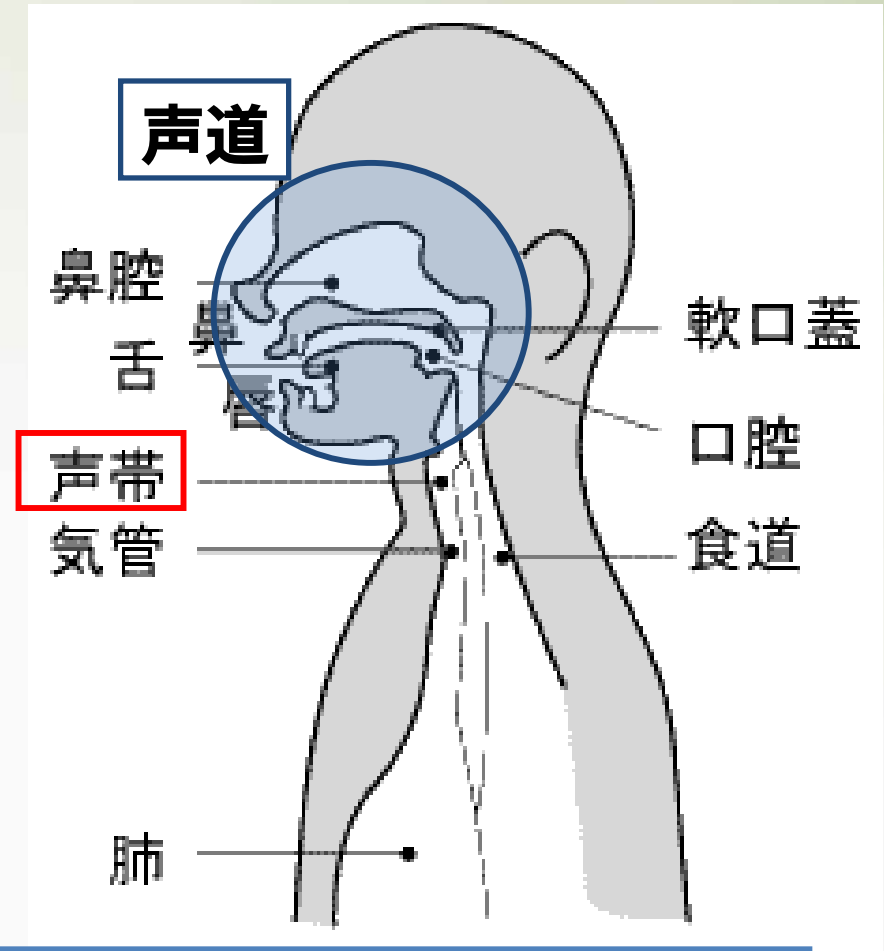
人間の発声のしくみ

声帯: 振動して音を出す

- 声の高さを決める
- 無声音 (「さ」の最初など) では振動しない

声道 (咽喉・口腔)

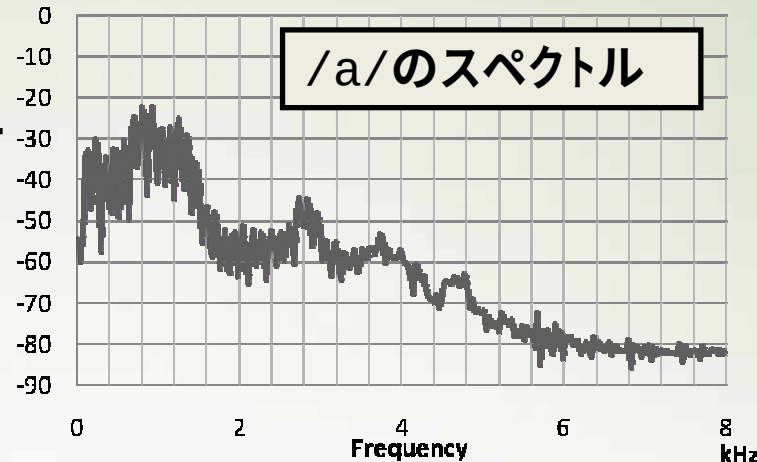
- 出された音を様々に調節する
- 声を特徴づける



この2つのパーツの特徴によって、様々な声が作られる

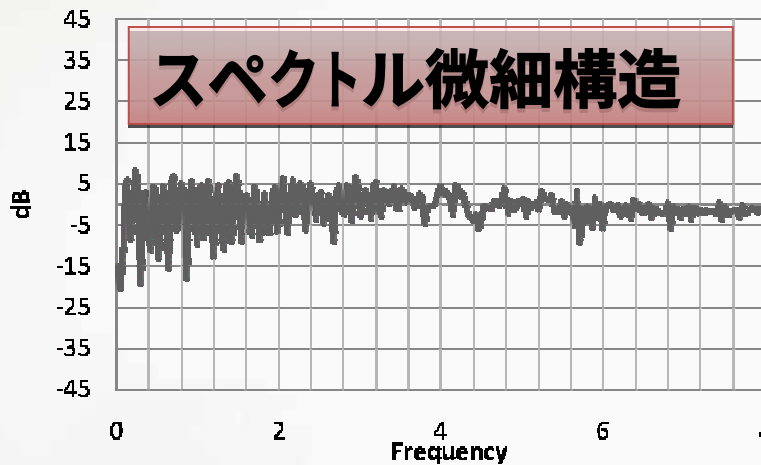
音声スペクトルの分析

コンピュータによって
観測された声の
スペクトル



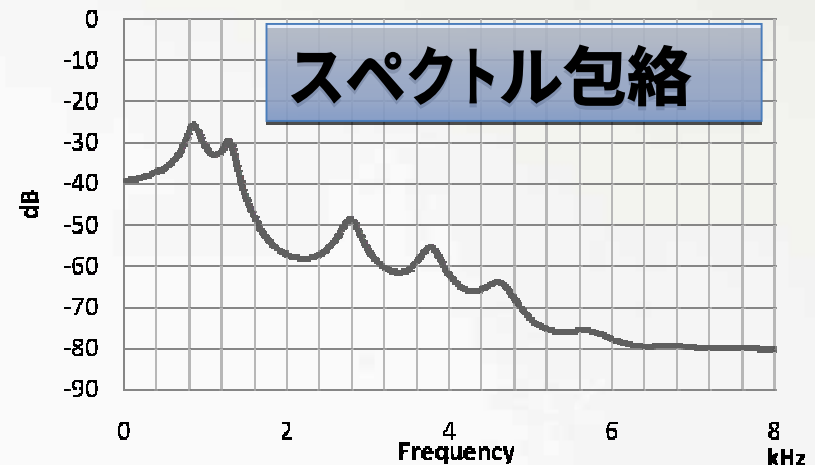
振幅・パワー
＝声の大きさ

スペクトル微細構造



×

スペクトル包絡

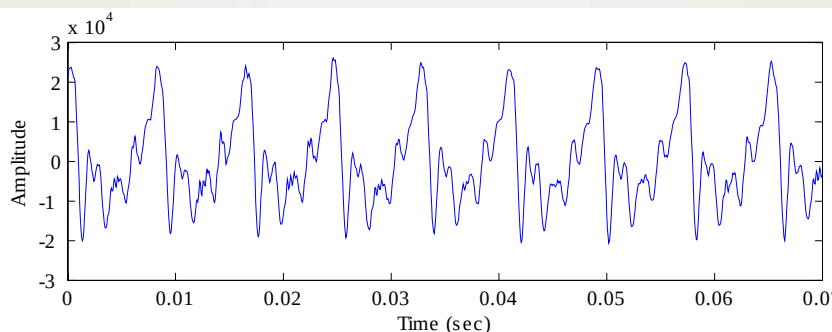


声帯の音源情報 (音高・有声・無声)

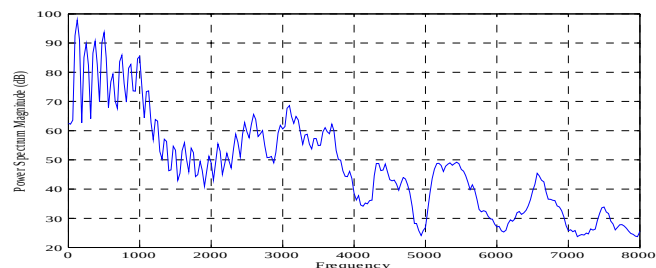
声道の調音特性 (音素の種類等)

このうち、主に**スペクトル包絡**の情報を用いて音声認識を行う

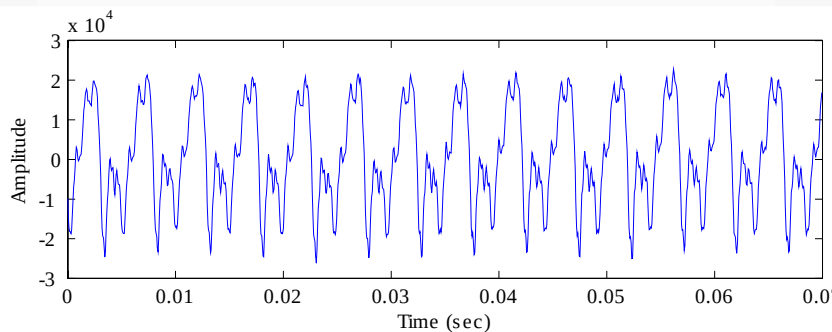
音素の波形とスペクトル: 声の高さ



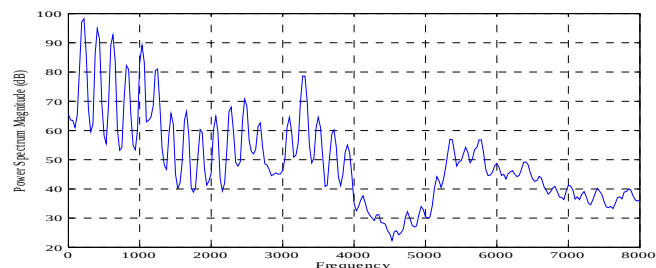
低い声の/a/



(0Hz~8000Hz)



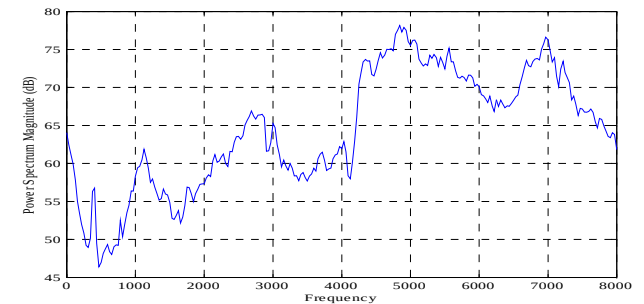
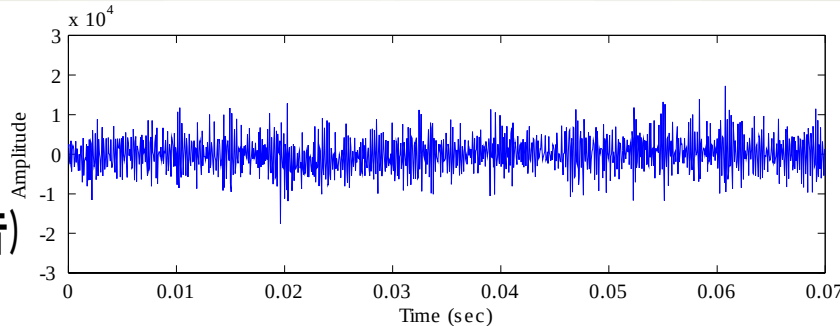
高い声の/a/



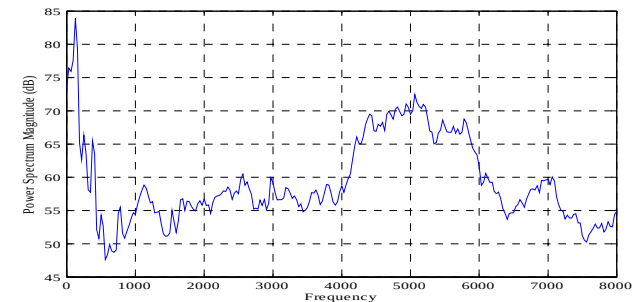
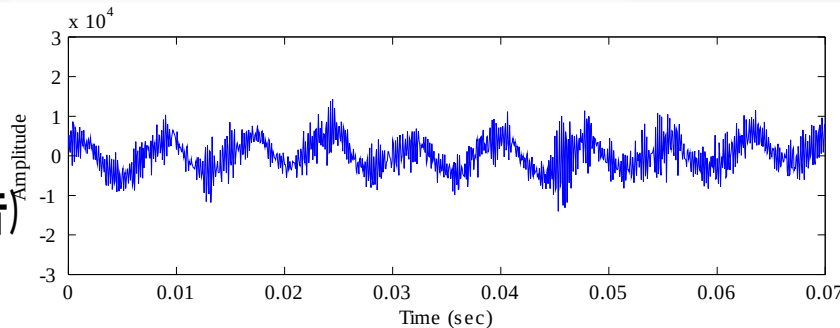
声の高さ (基本周波数/ F_0) は、音声波形では周期波形の長さとして現われ、
周波数領域では、スペクトルの微細構造 (細かいギザギザ) として現われる

音素の波形とスペクトル：有声・無声

/s/
(無声摩擦音)

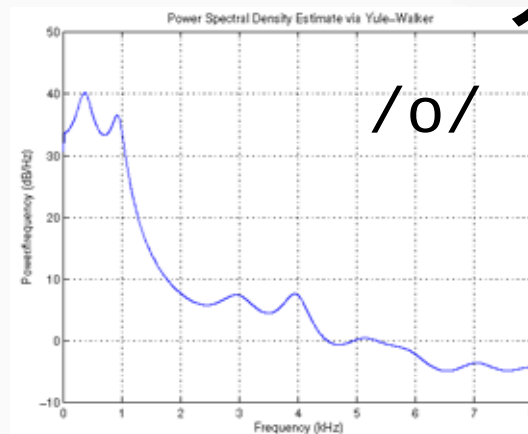
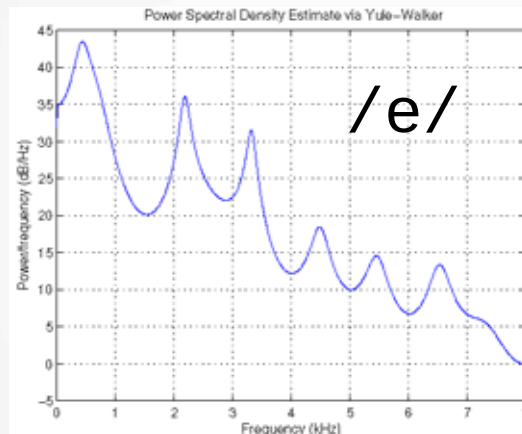
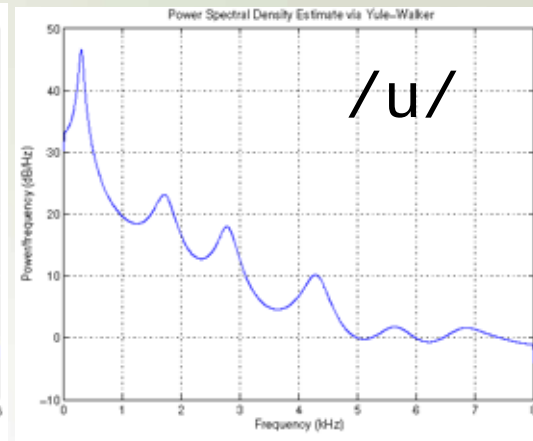
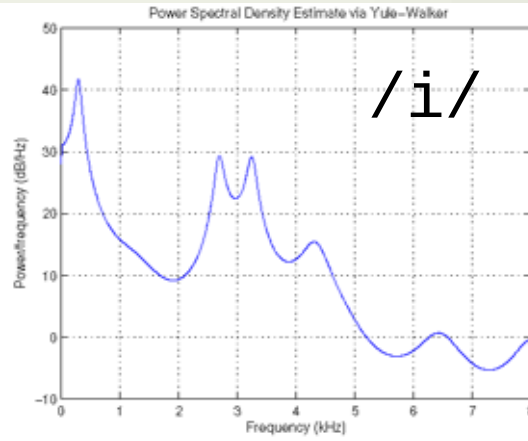
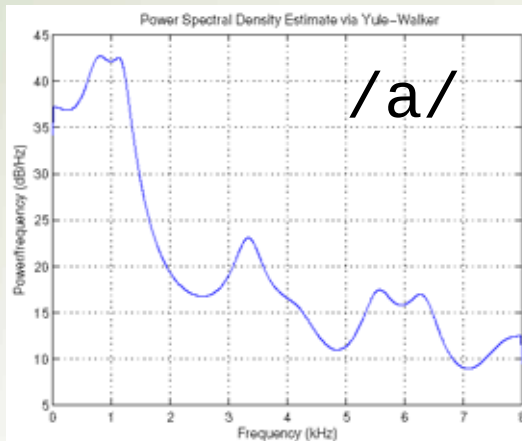


/z/
(有声摩擦音)



有声音：声帯が振動する→音声波形が周期的な成分を含む
無声音：声帯が振動しない (空気の乱流のみ) →非周期的

スペクトル包絡とフォルマント



スペクトル包絡のピークを
「フォルマント」と呼ぶ。
(低いほうから第1, 第2, ...)

なお、子音では
明確なフォルマントは
観察できない

実演

スペクトルがどう変化するか実際に見てみます。

1. 高い音 / 低い音
2. 口笛 / 「あー」
3. 「あー」 / 「いー」 / 「うー」

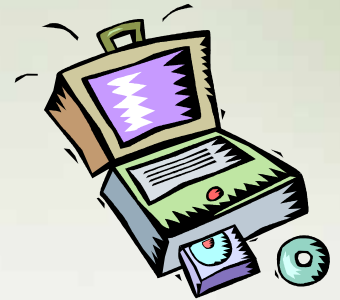
使用ソフトウェア:

高速リアルタイムスペクトルアナライザ WaveSpectra
音声分析ソフトウェア wavesurfer

音声認識のしくみ

• 昔の方法

- 声をあらかじめ登録し、入力された声に対して登録された中から最も「近い」ものを算出する
- ただし、この方法は・・・
 - 事前に認識したい音声を登録する必要
 - 他人では誤動作



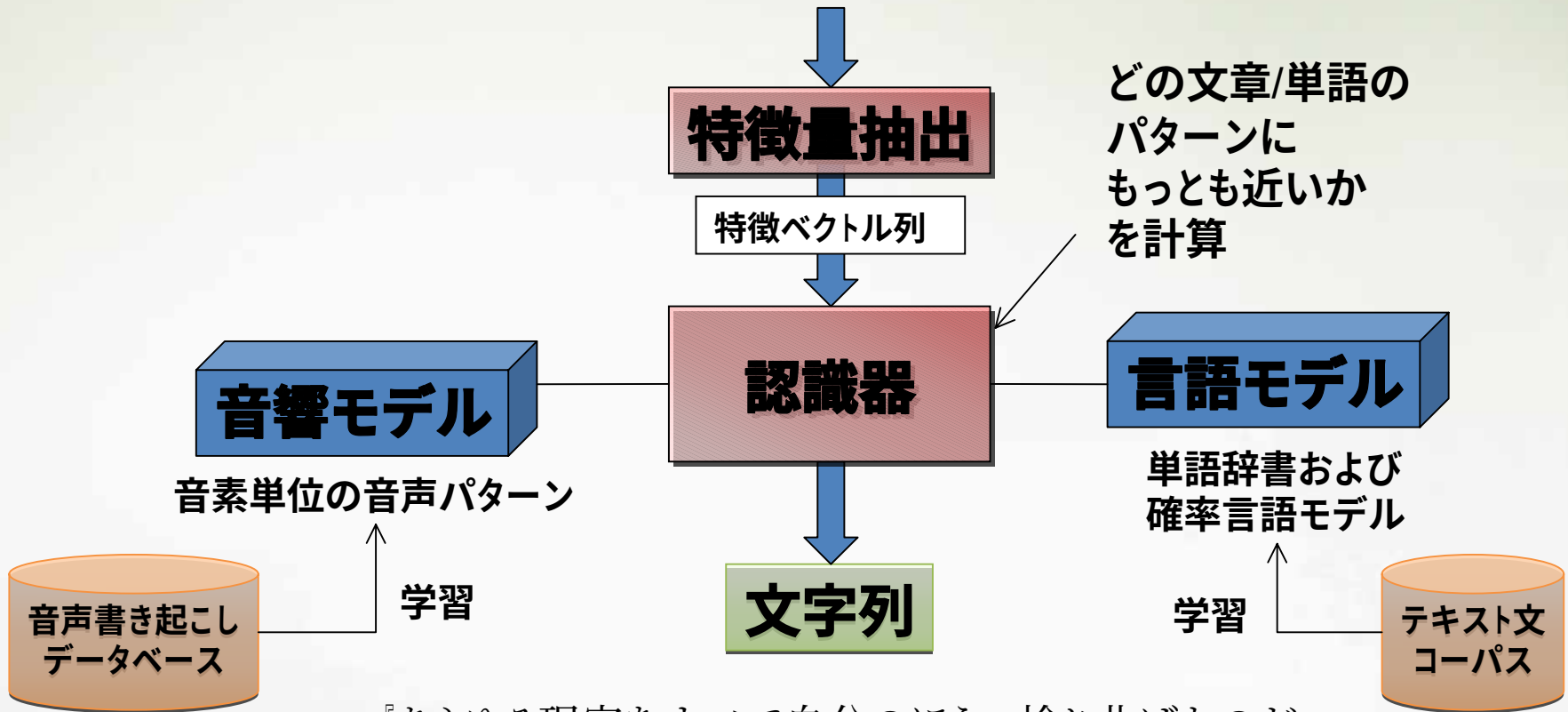
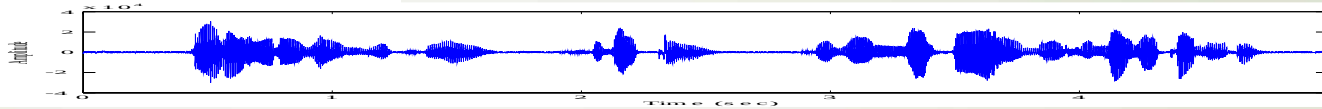
• 現代：大量データによる統計的モデル化

- 数十時間、数百万語の音声データから学習
- 多数話者による不特定話者モデル
⇒ 認識精度の飛躍的向上



音声認識システムの構成

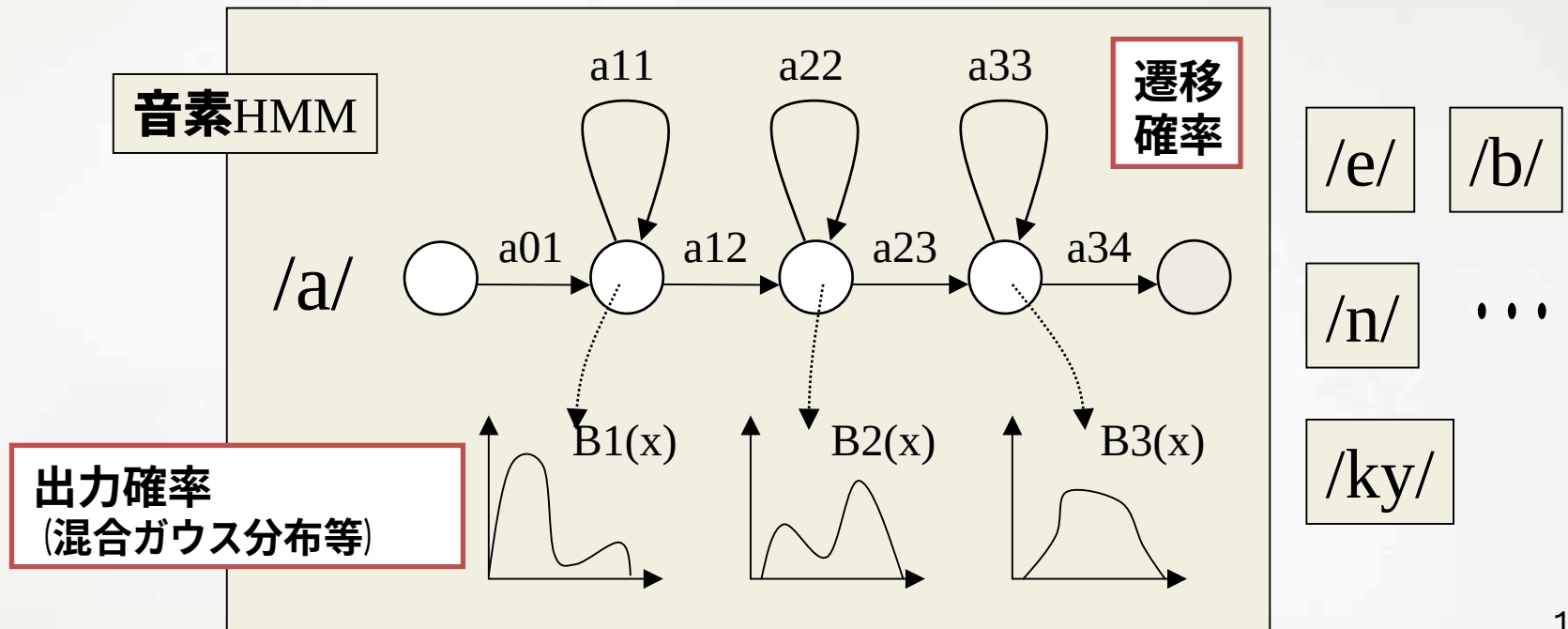
文章あるいは単語発声



『あらゆる現実をすべて自分のほうへ捻じ曲げたのだ』

音響モデル：音素HMM

- HMM: 隠れマルコフモデル (Hidden Markov Model)
- 音素ごとの発声パターンを確率的モデル化
 - 短時間の定常性 ← 状態ごとの出力
 - 長時間の非定常性 ← 状態間の遷移



言語モデル(1) 記述文法

- 現れうる文のパターンを規則で記述
- 認識時には、規則上許される文の中で最も入力にマッチするものを算出

文法ファイル (.grammar)

```
S: NS_B NAME NS_E
S: NS_B NAME FILL_E NS_E
S: NS_B FILL_B NAME FILL_E NS_E
S: NS_B FILL_B NOISE NAME FILL_E NS_E
S: NS_B FILL_B NS_E
```

人手でパターンを記述
小語彙～中語彙 (数百語) 向き

辞書ファイル (.voca)

```
% NAME
小泉      koizumi
森山      moriyama
田中      tanaka

% FILL_B
えーと    e:to
あの一    ano:

% FILL_E
お願い    onegai
さんお願い saNonegai
```

言語モデル(2) 単語N-gram

- 現在最も用いられている統計言語モデル
- ある単語の出現確率が、直前の $N-1$ 単語のみに依存すると仮定

$$P(w_1 \cdots w_n) = \prod_{i=1}^n P(w_i | w_{i-N+1} \cdots w_{i-1})$$

- $N=1$: ユニグラム、 $N=2$: バイグラム、 $N=3$: トライグラム

トライグラムの例:

$$\begin{aligned} P(<S>, \text{学校}, \text{に}, \text{行}, \text{<}, </S>) &= P(\text{学校} | <S>) \\ &\times P(\text{に} | <S>, \text{学校}) \\ &\times P(\text{行} | \text{学校}, \text{に}) \\ &\times P(\text{<} | \text{に}, \text{行}) \\ &\times P(</S> | \text{行}, \text{<}) \end{aligned}$$

大量の学習用テキストデータが必要, 大語彙向き

音声認識の諸相

- 音声を「書き下す」
 - 音声ワープロ

数万語・文章発声が対象
正確さを重視

- 音声で「操作」
 - 音声リモコン
 - 家電・カーナビ

語彙数は少ない
処理速度・反応性
実環境での頑健性が最重要

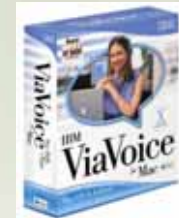
- 音声で「おしゃべり」
 - ロボット
 - 音声対話システム
 - 音声翻訳

より上位の知的処理
いかに「理解」するか？
自然言語処理・人工知能



音声認識ソフトウェア

- 音声認識ソフトウェア
 - IBM ViaVoice
 - Microsoft 音声認識 (Vista には標準搭載)
 - ボイスダイアル (携帯電話)
- 音声翻訳
 - 「しゃべって翻訳」@ドコモ携帯
- 音声操作
 - ほとんどのカーナビゲーションシステム



© ATR-Trek / Docomo

音声認識技術の現状と課題

- よく認識する話者とそうでない話者ができてしまう
 - 本来, 声質・喋り方には**強い個人性**あり
 - 原理的に「平均」から遠い話者ほど不利
 - 話者適応 (エンロール) は手間
- くだけた話し方はとても苦手
 - 意識してしゃべる読み上げは95%以上
⇒ 日常会話文章は70%程度
- 聞こえにくいと間違えやすい
 - 交差点や雑踏, 他者の声など
- 知らない単語は分からない
 - 新しい語彙や地名・人名
- 現状：用途を絞り込んで実用化
 - 例：医療用のカルテ入力



外国語を習い始めて
1年目ぐらいの
大人の言語能力にほぼ近い



②ソフトウェア紹介

汎用音声認識エンジンJULIUS

音声認識エンジンJulius

- ✓ **フリーの音声認識ソフトウェア**
 - 誰でも無償でダウンロード・利用可能
- ✓ **高い認識性能**
 - 数万語の大規模モデルを実時間処理
 - 携帯電話 (iPhone) でも動作
- ✓ **カスタマイズ性**
 - 一般的なモデルが使用可能
 - 他言語でも動作: 英語・フランス語・タイ語・中国語...
- ✓ **学術・応用の両側面で貢献**
 - 国内外の多くの研究・開発機関で利用
 - 日本の音声認識研究の事実上のスタンダード



音声認識の実演

- **単語認識**
 - 全国駅名認識 (約10,000語)
- **自動書き起こし**
 - 新聞記事読み上げ音声の認識
語彙数: 60,000語
- **音識別機能**
 - 雑音棄却
 - 笑い声判別
- **研究紹介: 音声認識過程の「見える化」**

ホームページから
ダウンロードできます

ホームページ

<http://julius.sourceforge.jp/>

The screenshot shows the Julius website in a Mozilla Firefox browser window. The browser's address bar displays <http://julius.sourceforge.jp/>. The website's header features the Julius logo (a microphone icon) and the name "Julius" in large blue letters. To the right of the logo is a search bar labeled "HP内検索 ヘルプ" with a "submit" button. Below the header is a navigation menu with links: HOME, JULIUSについて, 動作環境, BBS, FAQ, ENGLISH HOME, and 開発サイト. The main content area is divided into three columns. The left column contains a "ダウンロード" (Download) section with links to Julius/Julian latest version, dictation kit, grammar recognition kit, custom execution kit, and application kit. Below this is a "派生版JULIUS" (Derived JULIUS) section with links to Julius for SAPI, JuliusLib/JuliusGUI, and Multi-band Julius. The middle column has a "What's New?" section with a "2007/01/10(Wed.) New!" entry for the Julius website renewal, and a "2006/12/29(Fri.)" entry for Julius-3.5.3 release. The right column contains a "QUICK DOWNLOAD" section with links to Source (tarball), Binary for Linux (tarball), and Binary for Windows (zip). Below this is an "ETC" section with a "SOURCEFORGE JP" logo and a "track feed" button. The bottom of the main content area has a note about downloading source code and binaries for Linux/Windows.

大語彙連続音声認識エンジン Julius - Mozilla Firefox

ファイル(F) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

http://julius.sourceforge.jp/

メモ Gmail TR-LAB Groupware ActiveCollab 名工大 徳田・李研 Julius ニュース 検索 講義

HP内検索 ヘルプ

submit

HOME JULIUSについて 動作環境 BBS FAQ ENGLISH HOME 開発サイト

ダウンロード

- Julius/Julian最新版
- ディクテーションキット
- 文法認識キット
- カスタム実行キット
- 応用キット

派生版JULIUS

- Julius for SAPI
- JuliusLib/JuliusGUI
- マルチバンド版 Julius

ドキュメント

- チュートリアル・解説
- インストール

What's New?

2007/01/10(Wed.) **New!**

- ◆ Julius website リニューアル

2006/12/29(Fri.)

- ◆ Julius-3.5.3 リリース

[リリースノート] | [ChangeLog].

- ◆ 高速化:コード最適化により 1.1倍~1.5倍程度高速になった
- ◆ 音響分析条件の HTK Config 読み込み, バイナリHMMへの埋め込み対応
- ◆ 文法最適化ツール: dfa_minimize, dfa_determineize
- ◆ 単語グラフの拡張
- ◆ バグ修正・コードの改善

最新版のページからソースコードおよび Linux/Windows用の実行バイナリがダウンロードできます。

QUICK DOWNLOAD

- ◆ Source (tarball)
- ◆ Binary for Linux (tarball)
- ◆ Binary for Windows (zip)

ETC

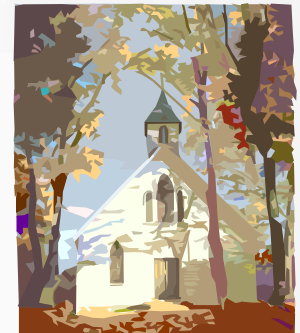
SOURCEFORGE JP

track feed

Julius, 実行キット, チュートリアル, マニュアル, ...

開発の歴史

- 1996年 開発スタート
 - 高性能音声認識アルゴリズムの研究が原点
 - 現在も開発継続中
- 開発体制
 - ほとんど一人
 - アルゴリズムが高度すぎる？ (数万行)
- なぜフリーか
 - アカデミックな研究における共通基盤
 - 実用化・応用の可能性



③

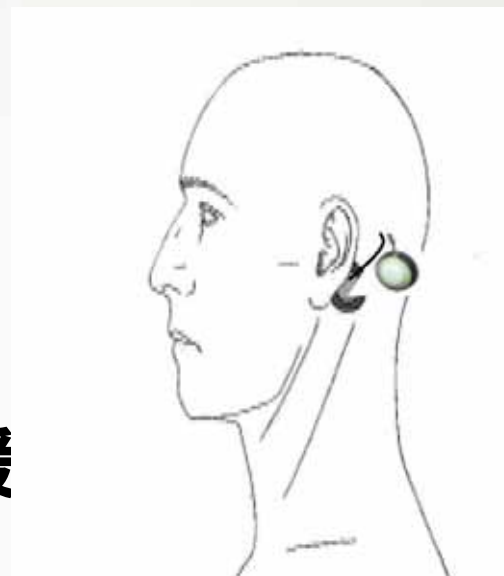
「音の認識」の事例紹介

ご紹介する事例

- **非可聴弦き声 (NAM) 認識**
- **環境音識別**
- **日常環境における警報音識別:
サウンドウォッチャー**

非可聴呟き声 (NAM) 認識

- 特殊マイクを耳の下に圧着
⇒ 声道内の空気振動を直接録音
- 声帯の振動が不要



応用

- 声を出さない「無音声認識」
- 発声障害を持つ方への支援

環境音識別

一般的特徴

- **定常な音はよく識別できる**
 - ファン, 水道のシンク, 掃除機, うなり音
- **突発的な非定常な音は難しい**
 - 雷, クラクション, ドア音
- **音声認識と同じ脆弱性**
 - 雑音が混じると難しい, 未知の音は識別不可など

環境音識別の研究事例

- **HMMを用いた環境音識別**
 - 奈良先端大
- **主成分分析に基づく環境音識別**
 - 名古屋大学
- **パワースペクトルと自己相関関数を用いる方法**
 - 名古屋大学
- **音声認識とほぼ同じしくみ**
 - 学習データが多いほど性能が高くなる傾向
 - 実験室での評価実験のみであるところが多い

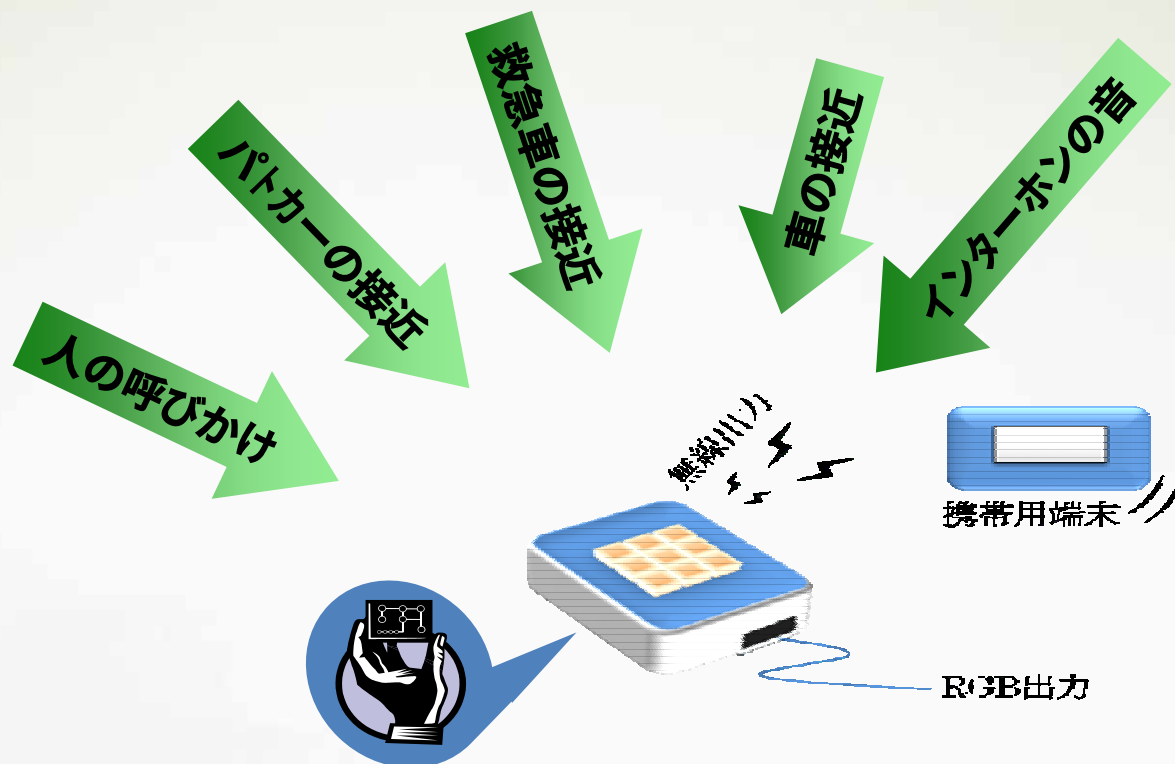
サウンドウォッチャー



名古屋工業大学大学院
創成シミュレーション工学専攻
岩田 彰 黒柳 奨

サウンド・ウォッチャー (音の見張り番)

二つのマイクロホンから収集した音信号から、リアルタイムに音源方向と種別を判定する。

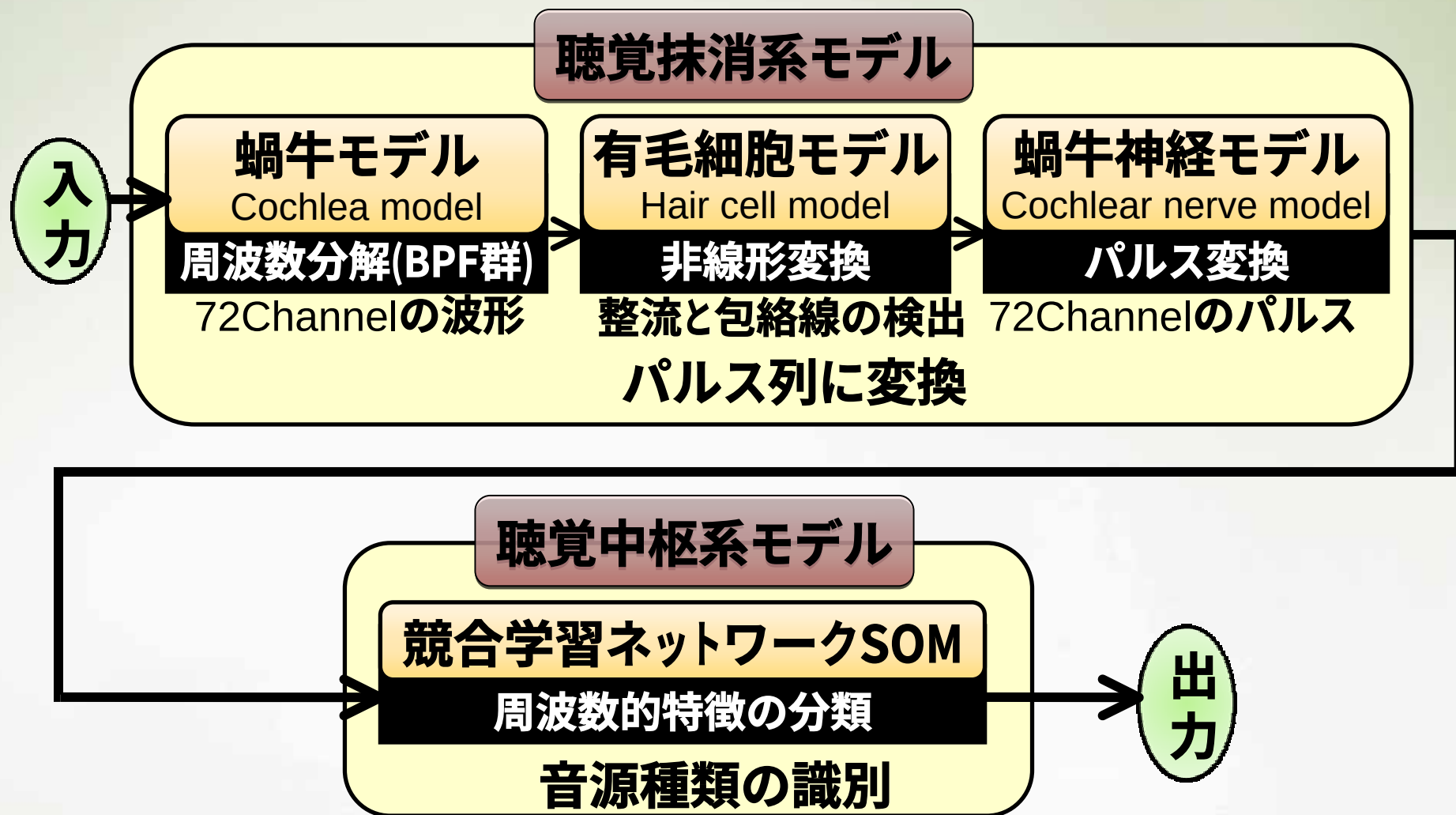


人工脳神経回路網のFPGAによるハードウェア化

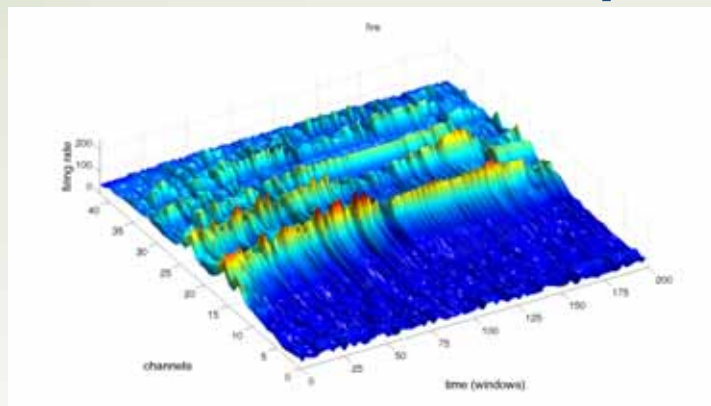
一音の見える化プロジェクト

- ・ 文部科学省知的クラスター創成事業「岐阜・大垣地域ロボティックス先端医療プロジェクト」に平成17年度から採択(平成20年度まで)
- ・ 平成20年度までに事業化を目指す
- ・ 事業化のためのベンチャー企業「(株)エンセファロン」を設立(平成19年10月2日)

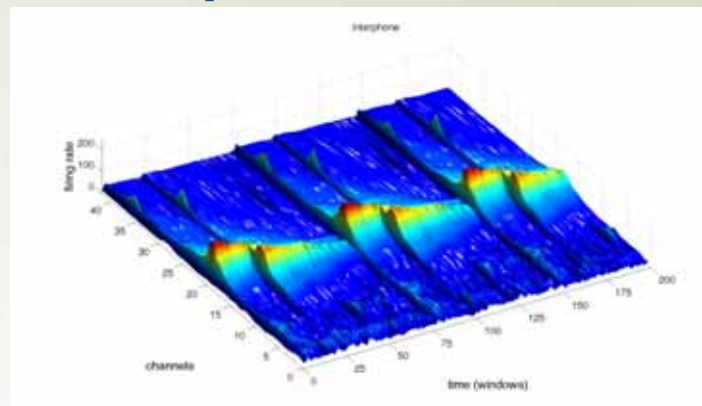
聴覚情報処理システムー音源識別モデルー



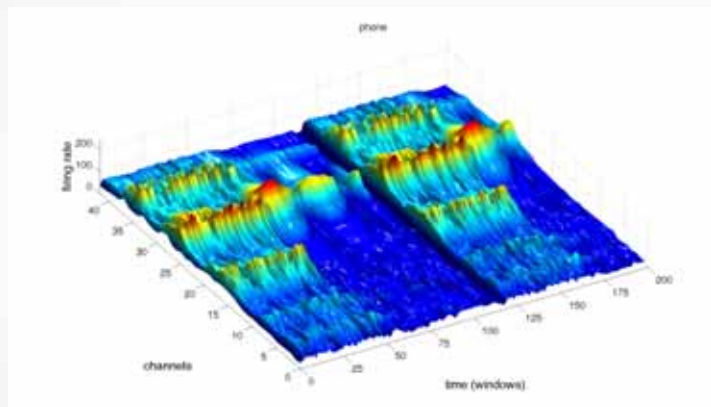
Spike Map



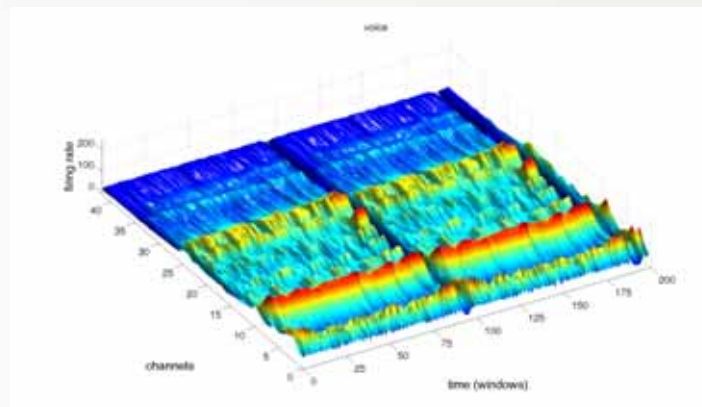
消防車



インターフォン



電話の呼び出し音



人の声

これらのリアルタイム識別を小さなハードウェアで実現

サウンド・ウォッチャー (第2試作基板、機能モデル)

・ 本体仕様

- 音源同定 9種類
- 音源定位 8方向
- マイク2個内蔵、外部マイク端子
- LEDによる結果表示機能
- 携帯端末への無線通信機能
- RGB出力端子
- 内蔵メモリによる収録機能
- リチウムイオン充電電池内蔵



サウンド・ウォッチャー

- **携帯端末仕様**

- 腕時計型またはペンダント型
- LEDによる結果表示機能
- バイブレータ通知機能
- リチウムイオン充電電池内蔵



実用化に向けた課題

- 実用化試作済み
- 聴覚障碍の方へのモニター調査
- 装置の現場使用でのSN特性確認

連絡先

名古屋工業大学大学院
創成シミュレーション工学専攻
岩田・黒柳研究室

<http://www-ailab.elcom.nitech.ac.jp/>

まとめ

- **声の視覚化**
 - **スペクトル分析**
 - **音声認識: 声を文字列に変換**
- **音の視覚化**
 - **声識別: 音声/非音声・性別・笑い声など**
 - **環境音識別**
 - **事例紹介**

参考文献・ドキュメント

- Julius のWebページ
 - <http://julius.sourceforge.jp/>
 - Julius本体・実行キット
 - 各種ドキュメント, ユーザーフォーラム
- 参考書
 - 鹿野清宏 他 編著「IT-Text 音声認識システム」
オーム社, ISBN 4-274-13228-5, 2001.
- サウンドウォッチャー: 岩田 彰／黒柳 奨
 - <http://www.thinkit.co.jp/article/30/4/>
 - <http://www-ailab.elcom.nitech.ac.jp/>