

# システム情報第一研究室 (嵯峨山/守谷) の紹介 2004

東京大学大学院情報理工学研究科システム情報学専攻

嵯峨山茂樹 / 守谷健弘 (NTT) / 西本卓也 / 酒向慎司 (21cCOE)

## ■ 研究要素と分野

### 三大要素

- 信号処理
- 確率モデル
- ヒューマンインタフェース

×

### 三大分野

- 音声
- 文字
- 音楽

## ■ 環境

- Linux, C, sh, perl, gnuplot, tgif, T<sub>E</sub>X; Windows
- PC × 2 + Notebook PC, network, 音響・音楽機器
- 勉強会 (信号処理、パターン認識、確率モデル、音楽理論、etc.)

## ■ 目標

- 創造的な研究、学会発表、特許

# 分野1: 音声・音響信号処理の研究

音声、音響信号、さらに一般に多チャンネル信号の処理の研究をしています。主なテーマは以下の通りです。

- **マイクロフォンアレイ**: 雑音中から音声を分離し、音声認識の性能を向上させる。  
当研究室で開発した **CSCC 法 (複素スペクトル円心法)** は、フーリエ変換の性質を巧みに用い、複数マイクロフォン入力から雑音を除去する方式で、その斬新なアイデアは学会から注目されている。
- **多チャンネル信号圧縮**: 多チャンネルの音楽信号や医用信号などの情報を圧縮する。
- **雑音環境下の音声認識**: 音素モデルの雑音適応の新方式を考案し、従来法 (**PMC, VTS** など) を大きく上回る性能を得た。
- **残響環境下の音声認識**: 残響環境に適応する音声認識の方式を開発し、残響を含む音声の認識精度が向上した。
- **歌声合成**: 楽譜と歌詞の情報を与えて、歌声を合成する方式を開発した。 **HMM** に基づく方式。

# 複素スペクトル円心法 (CSCC法) (2004)

## ■ 原理

- 目的信号  $s(t)$ 、雑音  $n(t)$  が別角度から到来
- 第  $i$  マイクロフォン信号  $m_i(t)$ :

$$m_i(t) = s(t) + n(t + \tau_i)$$

- 信号のフーリエ変換:

$$M_i(\omega) = S(\omega) + N(\omega)e^{-j\tau_i\omega}$$

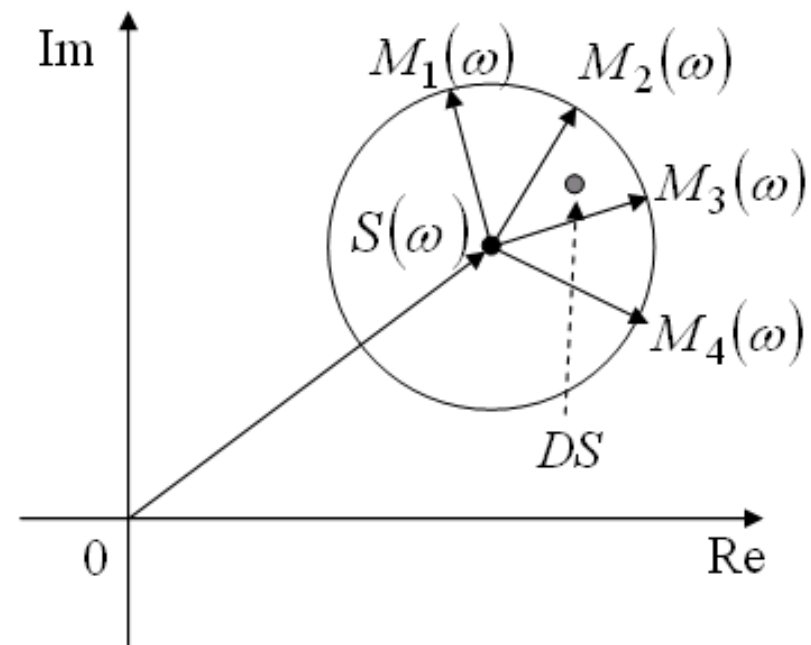
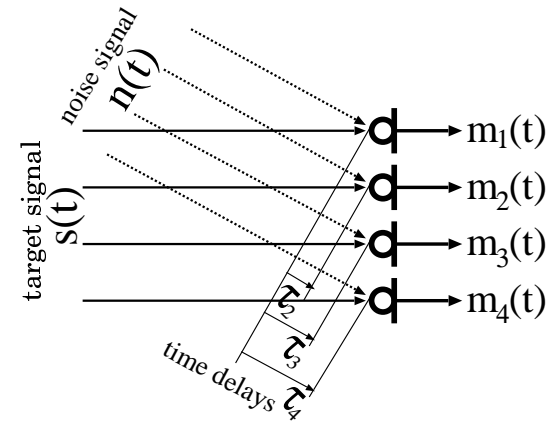
$S(\omega)$  を中心、 $|N(\omega)|$  を半径とする円周に分布 (右図)

- 円心を求めれば、雑音除去

## ■ 特徴

- 周波数ごとに別の到来角度も OK
- 学習不要: 動く音源も OK

## ■ 音声認識入力で大きな効果

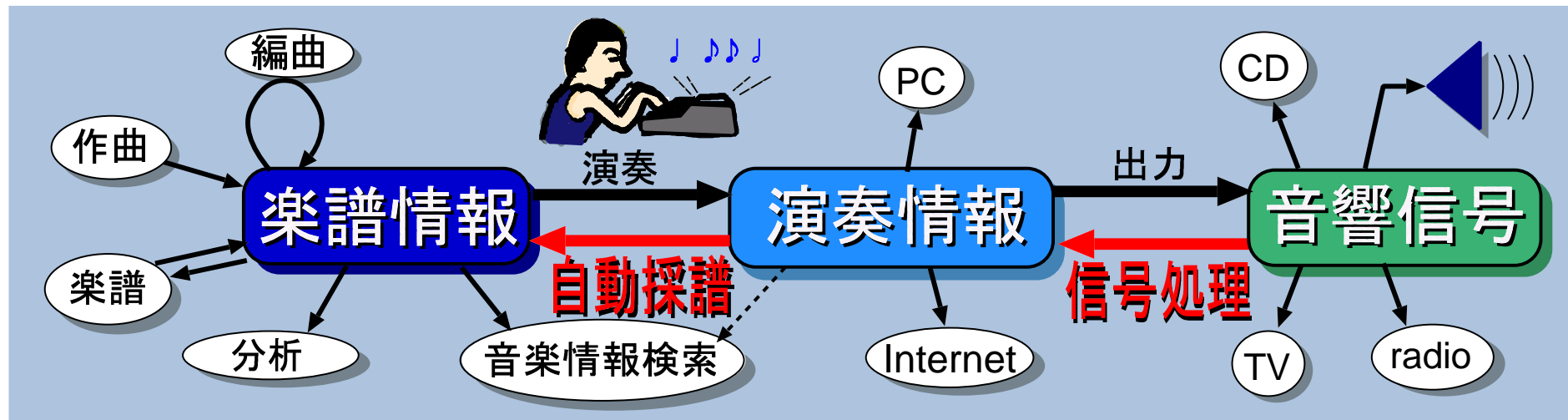


## 分野2: 音楽信号解析と自動楽譜化

PCに音楽を聴かせたら、そのMIDIファイルを作ってくれたり、楽譜を印刷してくれるならば、大変ありがたいでしょう。音楽は、音声と並んで、音に多大な情報を載せる媒体である点で双璧をなしている。しかし、その研究は音声に比べてまだ多くない。我々は、音楽信号処理と情報処理の分野で活発な研究を行っている。

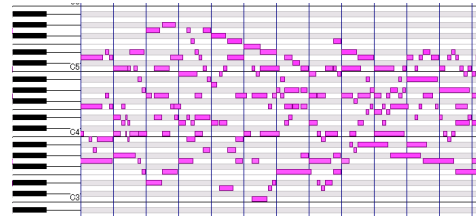
- 音楽の視覚化：Specmurt 法によりMIDIのピアノロール表示に相当する情報が得られ、これからMIDI信号に変換するなどが可能である。
- 多重音ピッチの推定と分離：Harmonic Clustering 法により、高精度の多重音ピッチの推定と、それに基づく音の分離を可能にした。
- 楽譜推定：隠れマルコフモデル(HMM)によるリズム認識により、MIDIファイル(あるいはMIDI化された音響信号)を、楽譜に変換する。
- これらの発展として、音楽信号からMIDIへの変換、楽譜化、演奏と楽譜のマッチング、主要旋律の抽出などを行う。
- 他のテーマとしては、与えられた旋律に対する自動的和声(コード)付け、MIDIファイルから最適指使い決定、コード抽出、演奏の解析など、興味と音楽知識の多寡により多数のテーマの中から詳細を決定する。

# 音楽信号/情報処理の概念



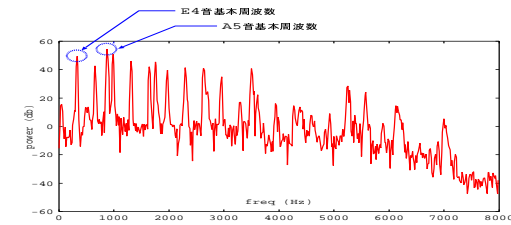
楽譜

- 編曲 (和声、対位法、伴奏)
- 楽曲解析、自動演奏
- 印刷
- 楽譜読み取り (認識)



演奏  
(MIDI ピアノロール表示)

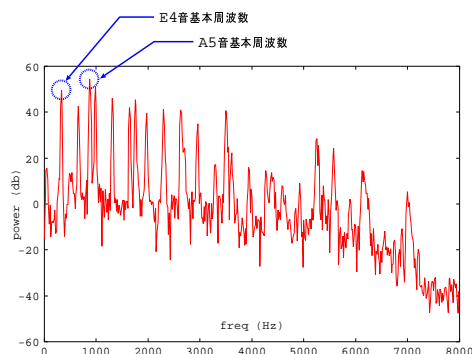
- 音色変換
- 編集、加工
- 歌声合成
- 楽譜化 (Video)



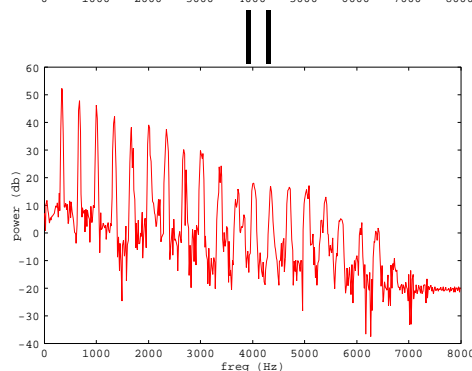
音 (例 15")  
(パワースペクトル)

- 多重音解析、分離
- MIDI化 (Video)
- カラオケ、着メロ作成

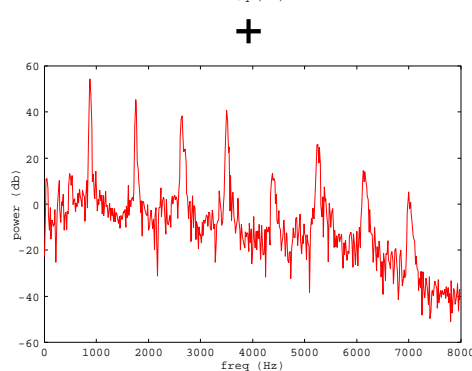
# 多重音の specmurt 解析の原理 (2003)



バイオリン E4 と A5 の混合音のスペクトル



E4音のみのスペクトル



A5音のみのスペクトル

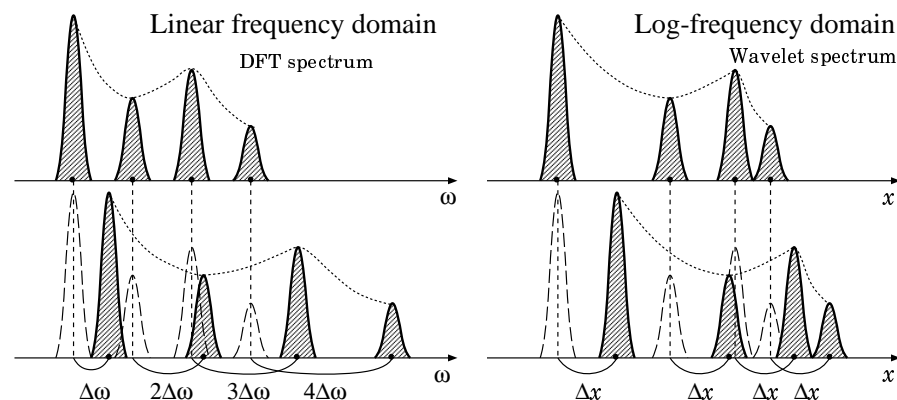


図1. 周波数軸を対数化する (wavelet 変換)

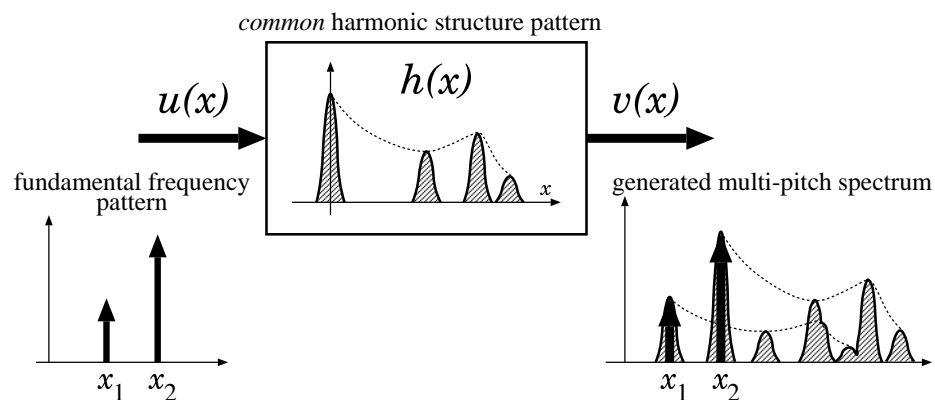
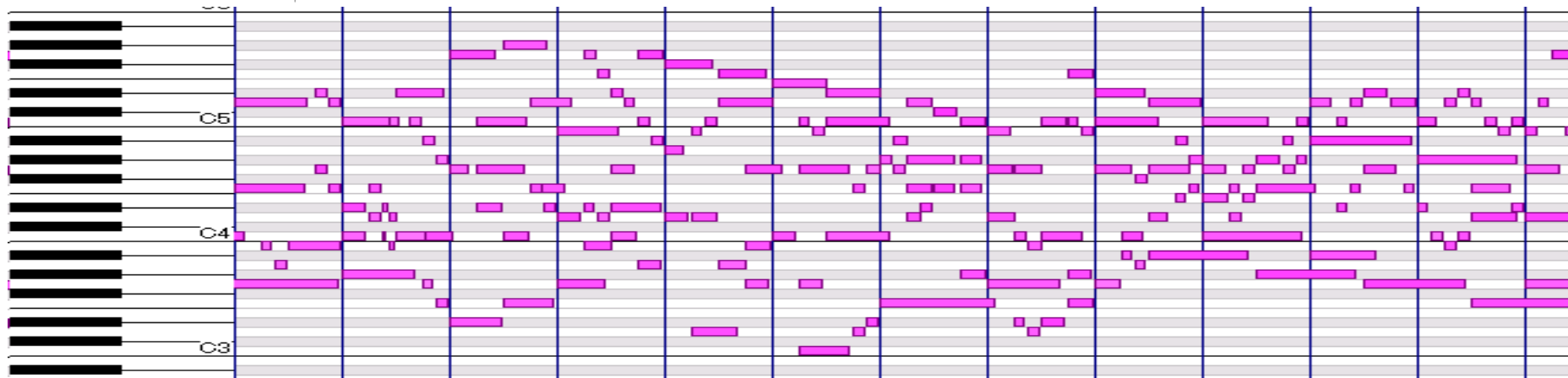


図2. 基本周波数分布と共通調波構造の畳み込み(対数周波数上)と見なせる

# 音楽例

## ■ J. S. Bach: Ricercare à 6 from Musikalisches Opfer, BWV 1079, (バッハ「音楽の捧げ物」から「六声のリチェルカーレ」) (管弦楽演奏)

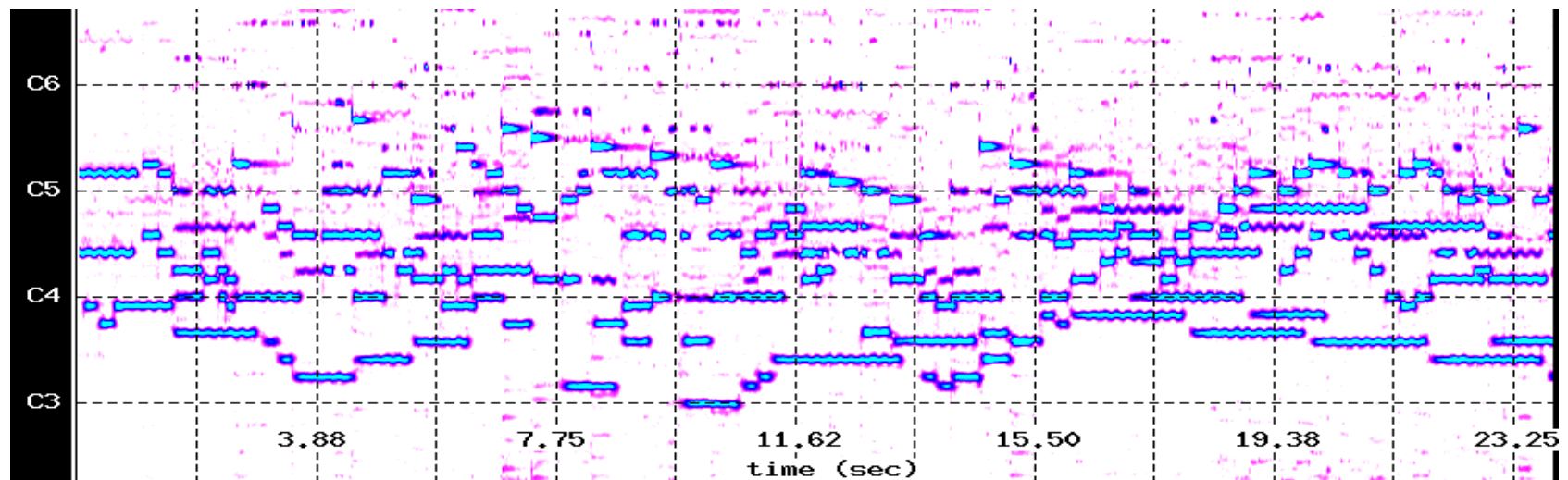
### 楽譜情報



### 演奏情報 (手作業で入力したMIDIファイルのピアノロール表示)

# Specmurt 解析例

- **管弦楽演奏** (J. S. バッハ「音楽の捧げ物」から「六声のリチェルカーレ」, J. S. Bach: Ricercare à 6 from Musikalisches Opfer, BWV 1079) の信号の **Specmurt** 解析結果 : 下のピアノロール表示とほぼ一致する。 **自動MIDI作成** (現在はさらに改良)

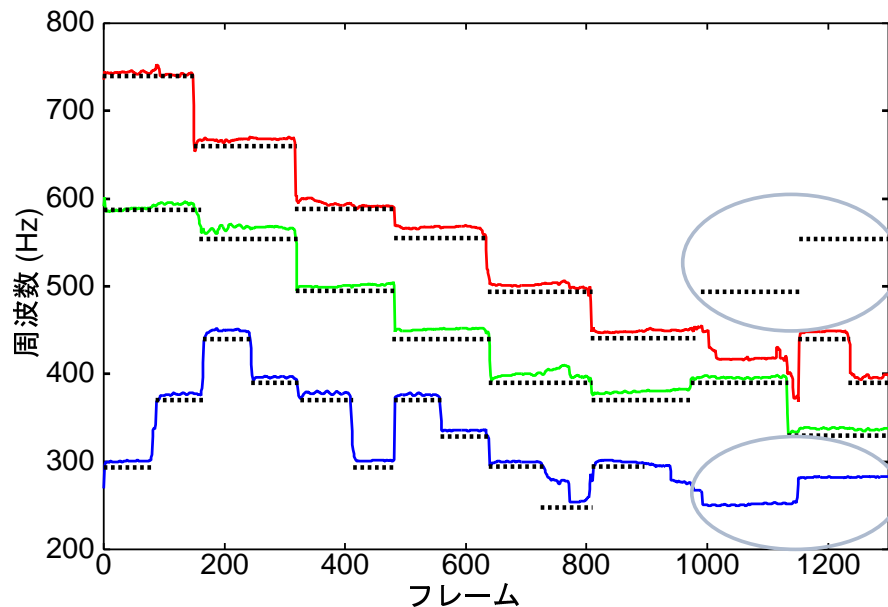




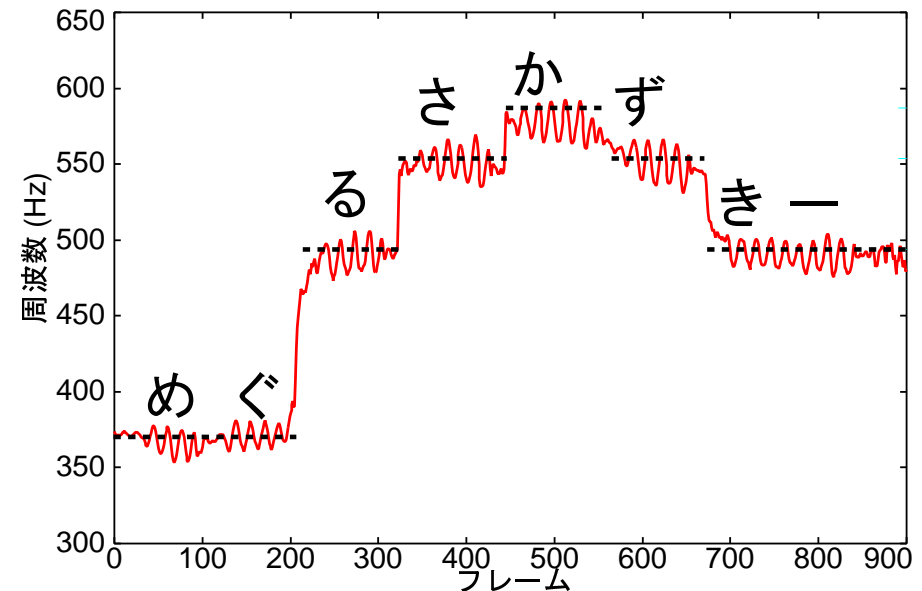
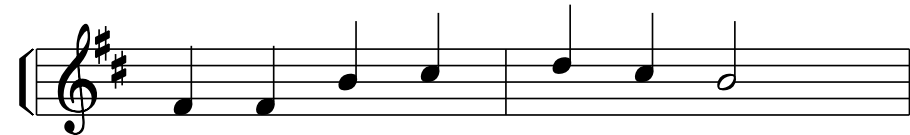
# Harmonic Clustering 解析 (2002)

- スペクトル成分を調波構造へクラスタリング (反復・収束)
- 音源数推定 (情報量規準)
- 高精度の基本周波数 (ピッチ) 抽出
- 音源分離

KANON  
Johann Pachelbel



荒城の月 滝廉太郎



## 分野3: 手書き文字・数式の認識

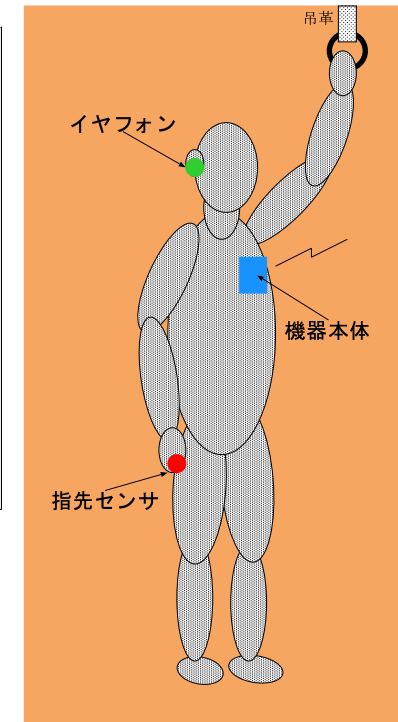
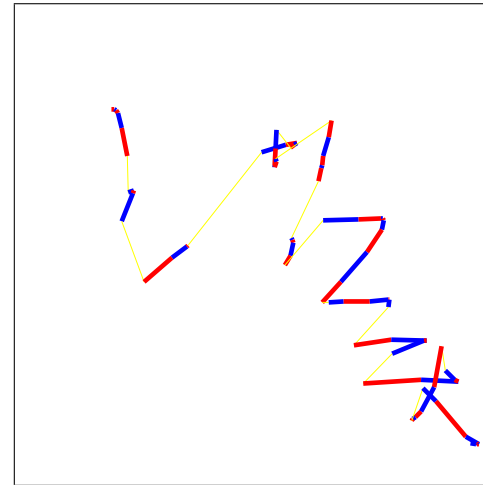
音声認識で開発された隠れマルコフモデル(HMM)と、漢字構造あるいは数式文法記述を用いて、手書き漢字や数式を認識する研究を行っています。

- オンライン手書き漢字認識：HMMと文字構造文法に基づく高精度な手書き漢字認識方式を開発した。筆者適応、文字ヒューマンインタフェース、環境依存型ストロークモデルなど、様々な研究を行い、高精度、高速性、省メモリを達成している。(Video, Online デモ)
- オンライン手書き数式認識：HMMと数式構造文法に基づく高精度な手書き数式認識技術を開発中である。その結果は、清書、計算、表示などに用いられる。

# 隠れマルコフモデルによる手書き漢字認識

## ■ 手書き文字認識アルゴリズム

- 連続音声認識手法を用いたオンライン手書き文字認識
- モデル単位の自動獲得、筆者適応、筆順学習、局所情報と大局情報の融合
- (予備評価実験: Zaurus < 我々の手法 < Windows IME)
- 昨年: 「文字スペクトル」、LSPパラメータ



## ■ 非目視手書き文字認識

指先の動きから意志を認識 視覚障害者文字認識 (デモ: [Video](#), [Online](#))

## ■ 視覚障害者のための文字コミュニケーションプロジェクト

オンライン非目視手書き文字認識、音声合成、対話制御、指先トランスデューサ (北陸先端大、金沢大、静岡県立大などと共同)

# HMM と確率文法による手書き数式認識

## ■ 手書き数式の曖昧さ

$$(a + 1)(2b - 1)$$

解釈 { “ca + 1x2b - 1)” …… 数式文法に反する  
 { “(a + 1)(2b - 1)” …… 正解らしい

- ストロークの解釈     HMM(隠れマルコフモデル)尤度
- ストロークの組合せ     CFG(確率文脈自由文法)尤度

## ■ 位置関係の文法

$$A = \int_0^1 \frac{dx}{\sqrt{1+x^2}}$$

# 分野4: 音声対話擬人化エージェントシステム

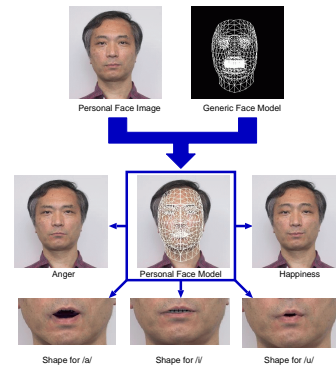
人間と同じように対話するコンピュータを目指し、**擬人化音声対話エージェント**の研究を進めています。

- 音声対話擬人化エージェントの共通基盤を作るプロジェクト(代表者：嵯峨山)により、オープンソース、無償ソフトを作成。
- 音声認識、音声合成、顔画像合成、全体制御の4つの部分からなる。
- 当研究室では、対話の進行に合わせたエージェントの**視線方向制御**、**表情・ジェスチャ合成**、使用者の反応(応答、割り込み、あいずちなど)に応じて各モジュールを即時に制御する**統合システム技術**の研究を行っている。

# 擬人化音声対話エージェント

## ■ 構成要素

- 音声認識
- 音声合成
- 顔画像合成
- 対話制御



## ■ 研究課題例

- 表情・視線制御
- 合成音声・歌唱の音色(表情)改善

## ■ デモ(落語(庭の蟹)2分)

- **家来:** ある殿様が、巷で評判の洒落の名人をお城に呼び寄せた。
- **殿様:** おぬしは洒落の名人だそうだな。
- **客人:** はい、そのように呼ばれております。
- **殿様:** では腕前を見せてみよ。庭の前の池に蟹が戯れているが、あれで何か洒落を言って見るが良い。
- **客人:** 庭の前の池の蟹でございますか?仰せではございますが、そう**にわか**には洒落られません。
- **殿様:** なんだと!? わしの申しつけなのに洒落られないとは何事か!
- **客人:** お殿様...
- **家来:** お殿様! いま彼が「**にわか**には洒落られない」と申したのは、庭と蟹をあしらって、「**にわ・かに**」と申しました。あれが言葉の洒落でございます。
- **殿様:** うぬ!なるほど!これはわしが悪かった。わしの誤りだ。許してくれ。
- **客人:** とんでもないことです。誠に恐れ入ります。
- **殿様:** いや、はははは。これは見事だ。
- **家来:** もう一度聴きたいですか?

## ■ HMM 歌唱音声合成 [Vol]

- **げんこつやまのたぬきさん**
- **幸せなら手を叩こう**

# 嵯峨山研究室の基礎知識 (勉強会項目)

- 統計的信号処理・学習・確率的逆問題・認識
  - デジタル信号解析 (自己相関、スペクトル、ケプストラム、LPC、PAR-COR、LSP)
  - 隠れマルコフモデル (Viterbi アルゴリズム、学習、音声・文字認識)
  - パターン認識理論 (パターン空間、距離尺度、識別関数、学習)
  - ニューラルネット (MLP、BP 学習)
  - 多変量解析 (分散分析、主成分分析、弁別)
  - 統計学習 (分布推定、Bayes 推定、EM アルゴリズム)
  - クラスタリング ( $k$ -means クラスタリング)
  - 動的計画法 (DP マッチング)
  - 文法理論、構文解析アルゴリズム、統計文法、探索手法

# 嵯峨山研究室：こんな人に向いている

- システム的問題思考、数理的定式化が得意な人
- 自分で面白い問題を考えるアイデア豊富な人
- 信号処理、確率モデルに関心がある人
- C-programming, Linux 環境が得意な人
- 音楽(クラシック)の理論に詳しい人
- 自分のアイデアを特許にしたい人