

## Commencement Lecture

# 音環境理解・ロボット聴覚の 水平展開

京都大学  
大学院 情報学研究科知能情報学専攻  
工学部情報学科計算機科学コース（兼担）  
奥乃 博  
<http://winnie.kuis.kyoto-u.ac.jp/>  
okuno@nue.org  
2014年3月17日 電気総合館大講義室



## 自己略歴（続）

- 1986年 Stanford Univ. CSD, Knowledge Systems Lab 2年半海外出張  
- Ed. Feigenbaum先生主宰のParallel AIプロジェクト参加。
- John McCarthy先生主宰のParallel Lisp Qlisp にもユーザーとして参加。
- Anoop Gupta先生（現Microsoft）とAIシステムの並列化に従事。
- Johan de Kleer, Ken Forbus 氏らがATMSのデータを提供。
- TAO/ELIS を ARPANET に接続。村上健一郎氏作成のTCPプログラム。
- NTT・スタンフォードの直接リンク保守。IMAP2 クライアントをTAOで作成。
- 「研究プロジェクトの進め方」を学ぶように Feigenbaum先生から示唆
- 1988年 NTTソフトウエア研究所に戻り、マネージメント、Internetの活用を訴えるためにNTTデータ関係の部署を全国行脚。
- 1992年 NTT基礎研究所で研究に戻る  
- 日本で初めて、Webページ作成（日本国憲法、NTTの事務処理、バス時刻表）
- TMS (Truth Maintenance System) の高速化
- BDDを使って、パズルを解く。
- 研究のストーリ付けを行い、ジャーナル論文に。 (Don Knuth著作で引用)
- TMSにBDDを応用。石塚満先生と共同研究。
- 博士論文の一部に。
- 多重文脈型真偽維持システムの高速化（1996年博士（工学）授与）
- この頃から多作となる（書きながら考える）

## 自己略歴

- 1950年 生まれ（神戸市）
- 1962年 大阪市立阪南小学校卒業
- 1965年 大阪市立阪南中学校卒業
- 1968年 大阪府立天王寺高等学校卒業
- 1968年 東京大学教養学部理科1類入学
- 1970年 同大教養学部基礎科学科進学
- 1972年 同大学卒業 教養学士
- 1972年 日本電信電話公社入所
- 1972年 同社武蔵野研究所配属 基礎研究所池野信一研究室配属  
- 甫宮真人氏（九大名誉教授）、竹内郁雄氏（東大名誉教授）の下でプログラミング言語の研究。LIPQでパズルを解く。（Martin Gardner著書2001で引用）
- 日比野靖氏（北陸先端大副学長）、竹内郁雄氏、大里延康氏（大阪工業大学）らで Lisp workstation ELIS とマルチパラダイム言語 TAO を開発
- 竹内郁雄氏から「研究を多方面から展開する」方法論、発表法、文章作法を学ぶ
- NTT基礎研究所の成果が、NTT-IT社を設立し、OKIとともにTAO/ELISを商品化へ。
- 村井純氏（当時東大）、浅見徹氏（KDD）と知り合い、JUNETに参加。NTTは国内のUUCPハブを運用。Sendmailのメンテ、故島久雄氏、後藤滋樹氏（早稲田）らと。
- TAO/ELIS の実用化で横須賀研究所に兼務になった時に、横須賀研究所にモデルを持ち込みネットワークを構築。
- DEC Tops-20 binary hack（第五世代プロジェクト中ずっと使用される）

## 自己略歴（続々）

- 1993年頃 音環境理解研究（CASA）の構想を固め、中谷智広氏と研究を開始。  
- NTT研究所は音声研究、音声認識の牙城 → 「NTTの恥」と酷評。
- 1994年 マルチエージェントによる音源定位・音源分離をAAAI-94に発表  
- AAAI-94, IJCAI-95, AAAI-96, IJCAI-97, AAAI-98  
- IJCAI-95, AAAI-96で CASA workshop を運営
- 後藤真孝氏をインターに迎え、共同研究。
- 1995年 東京大学工学部電子工学科知能工学（富士通）寄付講座客員助教授で半年間出向  
- ホストの田中英彦研究室で音楽情報処理を進めていた柏野邦夫氏、中臺一博氏と知遇
- 田中先生から「学生に好きなことをさせ、伸ばす」という研究室運営を学ぶ
- 1998年 NTT早期退職、科学技術振興機構ERATO北野共生システムプロジェクトに技術参事・グループリーダーとして参加。  
- AAAI-99, AAAI-00, IJCAI-01, AAAI-02  
- 中臺一博氏とRobot audition, active audition, 聖徳太子ロボットなどを提案
- 北野宏明氏から「勝てる研究ストーリづくり」を学ぶ
- 1999年 東京理科大学理工学部情報科学科教授に採用、北野プロジェクト兼務  
- 田中研方式で研究室運営：セキュリティの研究も始め、短期間で多数の成果
- 2001年 京都大学大学院情報学研究科知能情報学専攻知能メディア講座音声メディア分野 教授に採用

## 研究方針

1. 素人のように考え、玄人のように進める（金出武雄先生）
2. 研究のidentityを確立  
- 聞き分ける
3. 研究のストーリ  
- 学生諸君のやりたいこと（素人的発想）をプロ風にストーリづけ、専門家として取り組む。
4. 研究テーマは創発的戦略と意図的戦略のバランス
5. 未開拓に分野、未消費の市場に進出
6. 既確立の研究分野に参入する場合には未消費の分野を狙う。

## 研究を進める上での注意（1）

1. 3ヶ月研究すればその成果を論文として発表する。  
➢ 目的、手法、実験内容、実験結果までひとまず書いてみる。  
➢ ストーリーが通っているかどうかを判断し、考察を洗練化するか再実験
2. 要素技術では、関連研究との比較が重要  
➢ 関連技術を実装し、同じベンチマークで比較するのがベスト
3. システム技術では、要素技術選択の根拠・ストーリーが重要  
➢ 適用可能な要素技術をリストアップして根拠を示す  
➢ 同じ研究内容を複数の切り口から考察した論文を書く
4. 日本語で報告・論文を書くときには、「起承転結」  
➢ Paragraph writing: 結論 (topic sentence) をまず書く  
➢ 英語に訳しやすいかどうかが一つの判断基準  
➢ negativeな記述は極力避ける(単語に否定の意味を含める)
5. 締め切りぎりぎりに書いた論文は通りにくい  
➢ typo, grammatical errors が多い。1週間前には英文添削  
➢ Self-plagiarism を避ける
6. 特許は論文投稿3ヶ月前には申請を

## 教育方針

### 学生が大学に期待することに極力答える

1. 成功したという達成感を得る  
- 論文執筆（卒論→国際会議論文→ジャーナル論文）  
- 考え方をOJTで学ぶ  
- 学生自身の問題意識が重要（動機付け要因）  
× プロジェクトで雇用は達成感を？？？（衛生要因）
2. 人のつながり、友人を作る  
- 授業では友達を作るよう出席票を工夫  
- 研究室では飲み会、BBQ、旅行などで運営経験・親睦
3. 知識・技術の習得（高等教育として）  
- メタ知識、メタ学力を伸ばす  
- 研究、文章化、発表の3つの側面でトレーニング



## 研究を進める上での注意（2）

7. 国際会議論文は必ずジャーナル論文に投稿  
➢ 全国大会・研究会で発表  
➢ 國際会議論文に投稿  
➢ ジャーナル論文を投稿
8. 研究会では、必ず質問すること  
➢ 質問するため熱心に聴く  
➢ つまらない発表は、どう改善すればよいかを考えてあげる  
➢ 分野違いでも「耳学問」が後で役に立つことがある
9. 発表ではわかりやすく説明する。  
➢ 相手は素人と思うこと  
➢ 時間が限られているので、コンパクトに要点を話す  
➢ 質問は復唱してから、回答する  
➢ 回答には必ず理由を述べる
10. 教員が獲得した資金に関するプロジェクトに協力を

## 音環境理解とは

### Computational Auditory Scene Analysis (CASA)

#### 1. 音による環境認識

- 音源定位, 音源分離, 分離音認識
- 多様な音の認識  
音声認識,  
音楽採譜,  
擬音語認識



#### 2. 「聞き分け」がキーとなる技術

- 大部分の音の応用では聞き分けた後の処理
- マルチモーダル情報統合の基礎 (感情, 印象)

#### 3. 混合音を聞き分ける

- 同時発話
- 多重奏音楽音響信号, 伴奏付歌唱, 合唱

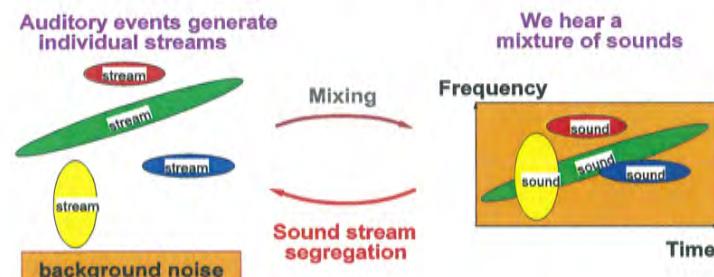
## なぜ音環境理解が必要か (2)

- ・聴覚障害者, 高齢者, 中途失聴者には聞き分け機能が必要.
- ・1対1対話は何とかこなせるが, 混合音はダメ
  - 複数人
  - 広い場所
  - 遠方
- ・カクテルパーティ・コンピュータは不十分
- ・聖徳太子のように聞き分ける機能が不可欠



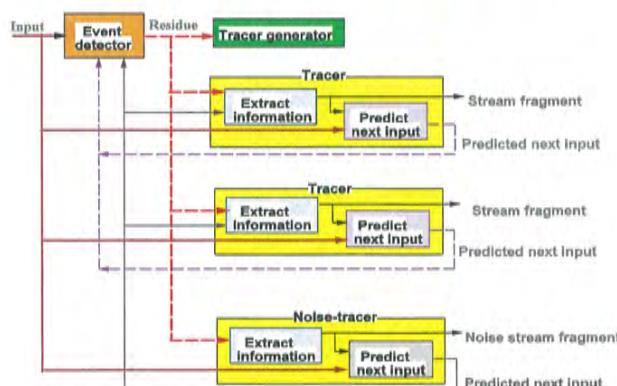
## マルチエージェント研究として

- ・音響信号処理は100年以上の歴史.
- ・新規参入は多大な障壁.
- ・自分の強いAIから参入. 【中谷さんと】



## 残差駆動型アーキテクチャ

- ・HBSS (Harmonics-Based Stream Separation)



## なぜ音環境理解が必要か (1)

- ・音声認識技術を実際に使えるようにするためにには, 混合音の処理が必要.



- ・クリーンな音を入力として想定し, 音声認識の応用を考えるのは, これから消費が発生する「未消費の市場」の放棄.

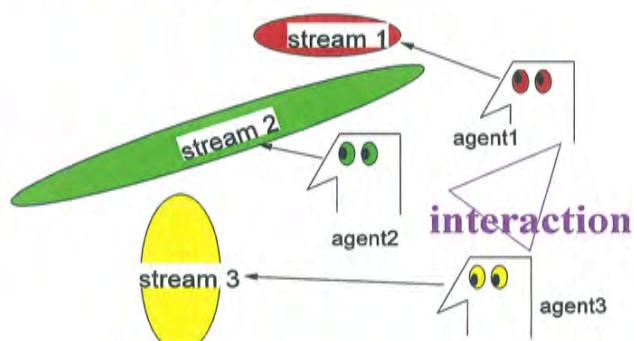
聞き分ける技術としての音環境理解研究

## 聖徳太子が聞き分けたのは何人?

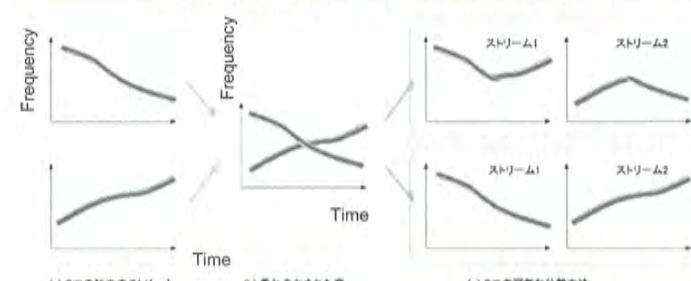


## マルチエージェントによる分離

- ・各エージェントは自分が追跡する音響ストリームに集中し、相互作用を通じて調整.



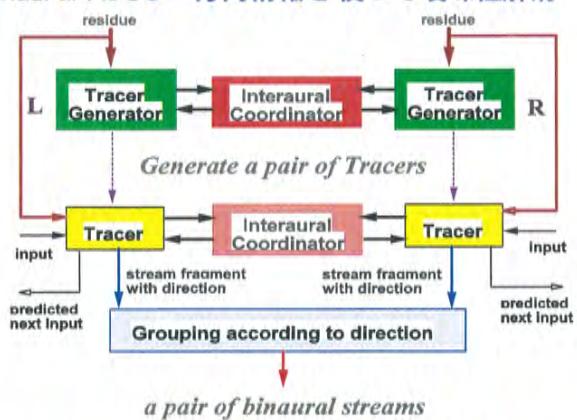
## 音源分離するまでの曖昧性



- ・音の高さ情報 (周波数) だけを使って分離すると, 2つのストリームの関係の曖昧性が残る
- ・さまざまな特徴を使って曖昧性を解消

## 残差駆動型アーキテクチャ

- Binaural HBSS: 方向情報を使って曖昧性解消



## Bi-HBSSででききたこと

1. 調波構造の音は分離できる

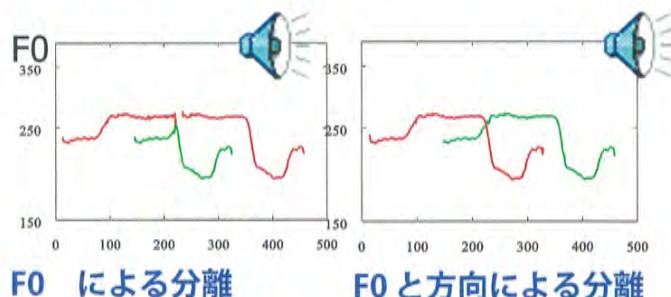
- 母音
- 有声子音

2. 無声子音が難しい

3. 入力音から調波構造を引いた残差はある

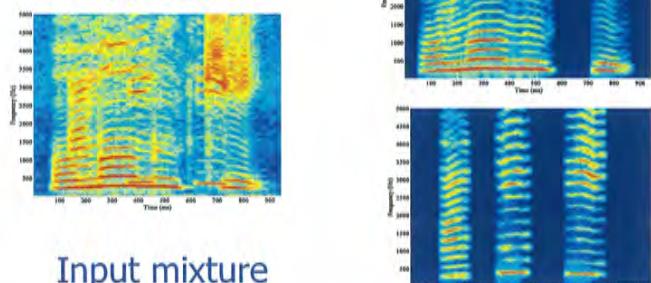
## 音源分離での曖昧性

- 2名の女性の声（同じ声をずらす）：入力、分離音（女性1, 女性2）

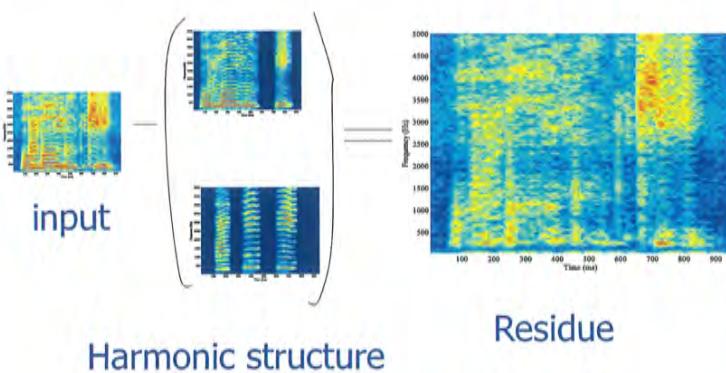


## 音源分離の第1ステップ

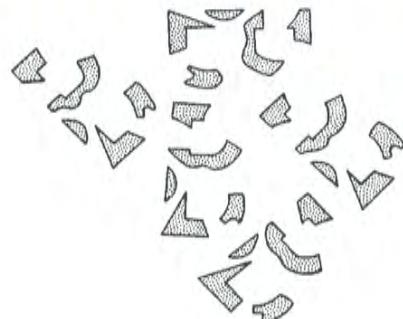
Separated by BiHBSS



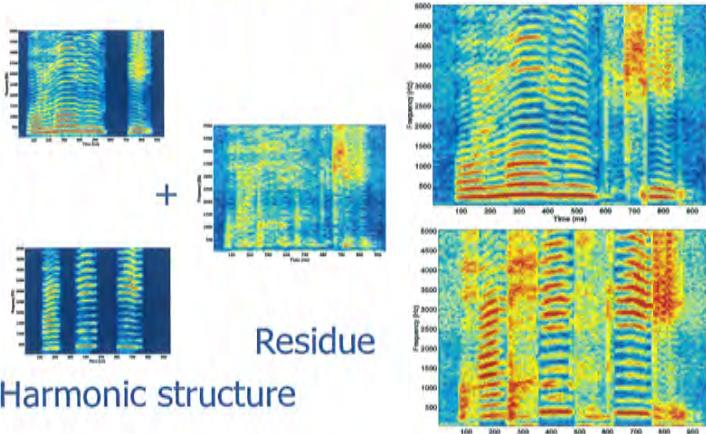
## 音源分離の第2ステップ



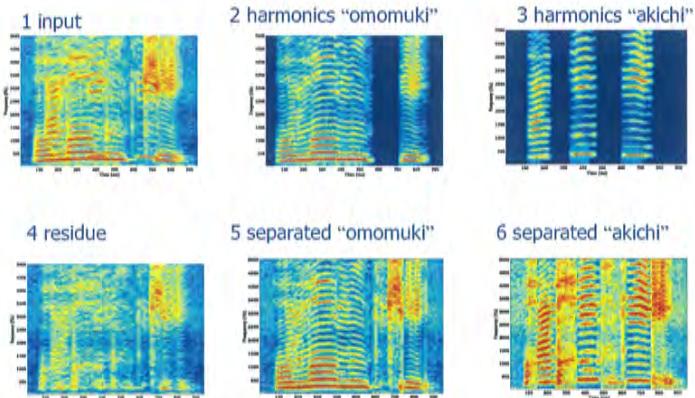
## 同じ大文字のアルファベットが複数



## 音源分離の第3ステップ



## 音源分離: demo



# 京都大学 with 中臺さん (ERATO & HRI-JP)

## 1. 音環境理解の基礎を詰める

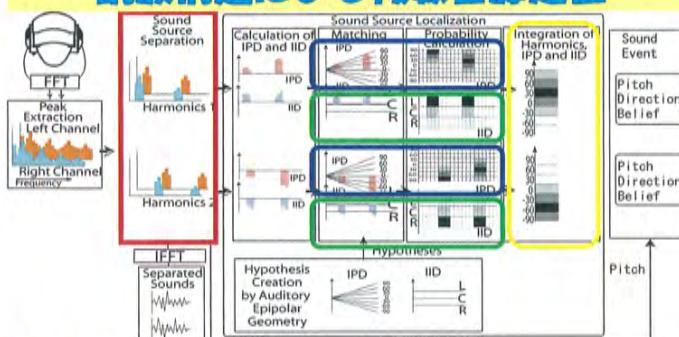
- 信号処理系の充実
- 音楽情報処理・擬音語認識

## 2. ロボット聴覚への展開

- 20世紀のロボットには「耳がなかった」
- 接話型マイクによるロボットとのインタラクション
- 未消費の市場への参入
- 「聖徳太子ロボット」という分かりやすいスローガン
- SIG, SIG-2
- Robovie [尾形先生]
- 両耳聴からマイクロフォンアレイへ



## 調波構造による複数音源定位



- 調波構造に注目した音源分離
- IPD(両耳間位相差)と聴覚エピポーラ幾何による仮説生成・照合(1500Hz以下)
- IID(両耳間強度差)を利用した右左の判断(1500Hz以上)
- IPDとIIDをDempster-Shaferで統合し確信度をつき方向情報を算出  
→ロバスト性の確保

## 2002: 60° 間隔の3話者



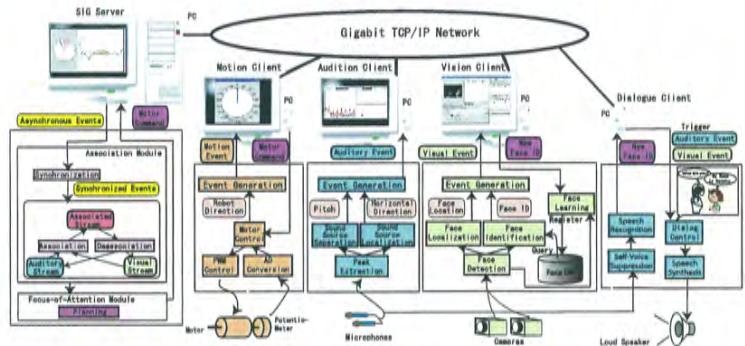
聞こえるのはSIGのモーター雑音

## ロボット聴覚のアプローチと主要機能



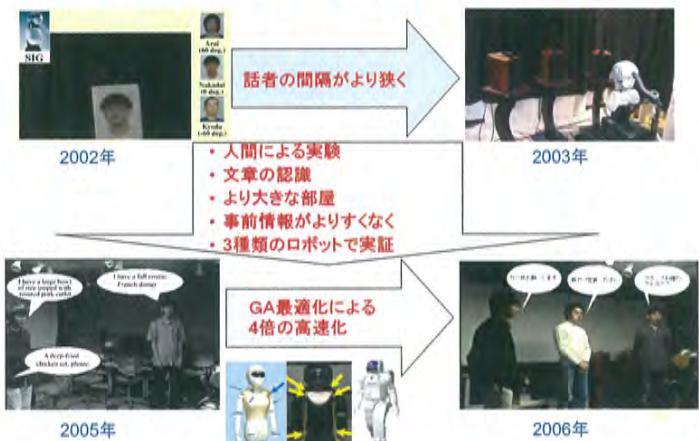
- マイクロフォンアレイ処理の導入
  - 音がどこから来るのか推定する (複数音源定位)
  - 特定方向の音に耳を傾ける (音源分離)
  - 耳を傾けた音を音声認識する (同時発話音声認識)
- 前処理と音声認識の統合、音声認識の頑健化
  - ミッシングフィーチャー理論、スペクトル特徴量、音響モデル適応、バージイン対応、発話区間検出

## 両耳聴による音環境理解システム



5台のPCをGigabit Etherで接続し、記号レベルで情報統合

## 三話者同時発話認識技術の進展



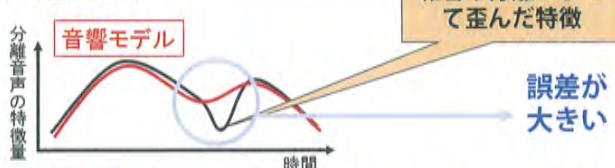
## 2003: 3話者 at 30° 間隔



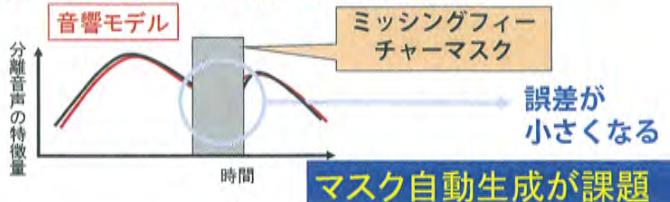
作り込みのシステムであり、汎用システムにできない。

## ミッシングフィーチャー理論 (MFT)

通常の音声認識

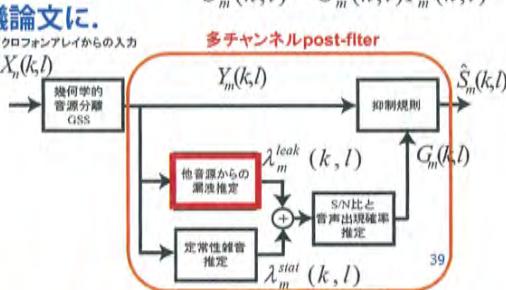


ミッシングフィーチャー理論に基づく音声認識



## MFM (マスク) の作り方

- マイクロフォンアレイ処理での情報を使用し、特徴量の信頼度を求める。
- 両耳聴でのMFMの構築法を研究する山本氏とマイクロフォンアレイ処理を研究するValin氏（短期留学）とが1週間で構築。
- 3か月で国際会議論文に。
- 二人は博士に。



## ロボットが人間と共生の場に進出

- 近年のロボット研究の目覚しい発展



- 種の淘汰: ソーシャルインタラクションが選択原理
- ロボットの知覚処理、特に聴覚機能は、ソーシャルインタラクションに不可欠

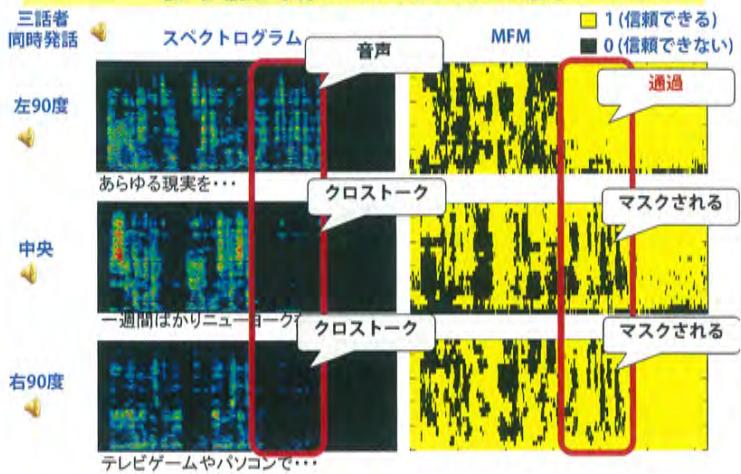
## HARKとは？

- モジュラーアーキテクチャに則ったロボット聴覚用オープンソースソフトウェア
- HRI-JP Audition for Robots with Kyoto Univ.
- 研究用途: Free (商用: Licensing)
- 事前知識を最少にしたポータビリティの高いシステムを目指して、設計・開発
  - ロボットの形状、マイクロフォンの数・配置
  - ミドルウェア HarkDesigner (FlowDesigner) 上でのモジュラー性
  - 多様な機能 (音源定位・音源分離・分離音認識) の実時間提供
  - 多様なプラットフォーム Ubuntu, Windows

## HARK 1.2.0 の機能

- 最新の実環境向け音響信号処理・音声処理技術
  - 音源定位:
    - GEVD/GSVD による3D定位 [Nakamura'11]
    - 3D超解像音源定位 [中村 11]
  - 音源分離:
    - GHDSS [Nakajima '09], HRLE [Nakajima '10]
    - HARK-SSS (10種類の音源分離アルゴリズム)
    - 自己発話抑圧 (Semi-blind ICA [Takeda'08])
  - 音声認識:
    - ミッシングフィーチャ理論
    - マイクロホンアレイにあった音響特徴量
    - 日本語・英語音響モデル
- 豊富な日英マニュアル・ドキュメント (300p超)
- 過去10回の国内外での講習会

## 自動生成したMFMの例



## ロボット聴覚ソフトウェアの設計

### 世界のロボットの聖地太子ロボット化

#### 1. ハードウェアに依存しないソフトウェア

- ロボット聴覚のOpenCV化
- Visual programming
- マイク配置・マイク応答のデータベース化
- 多様な音響処理モジュールの提供 (適材適所)



#### 2. コミュニティを作る

- ロボット聴覚の学会でのorganized session
- JSPS日仏研究交流「ヒューマノイドロボットのための両耳聴」

#### 3. 普及・支援活動

- 無料講習会実施
- 出張支援

## Ariel's Song, The Tempest

Come unto these yellow sands,  
And then tale hands;  
Curt'sied when you have, and kiss'd,  
(The wild waves whist;)  
Foot it feately here and there;  
And sweet sprites the burden bear.  
[Burden dispersedly.]



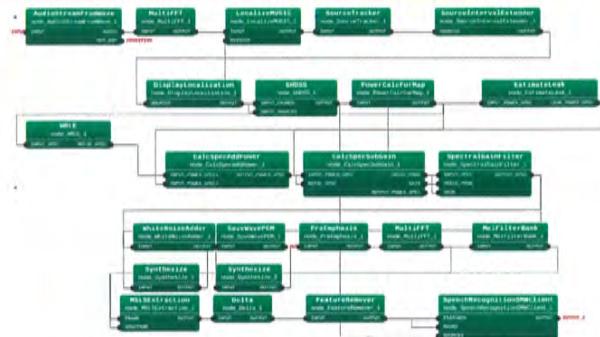
Hark, hark! the watch-dogs bark,  
Bowgh-wowgh.  
Ariel. Hark, hark! I hear  
The strain of strutting chanticleer  
Cry cock-a-doodle-doo.

## HARKのソフトウェアレイヤ

HARK Designer			CLI		
Original Modules For FlowDesigner	FD-dep. HARK packages	HARK-ROS package	ROS-dep. HARK packages (e.g. HARK-ROS stacks)	Indep. HARK packages (e.g. ASR)	HARK Tools (e.g. WIOS)
	BatchFlow		ROS		
OS : Linux/Windows 7/8/(Mac OS)					

- BatchFlow (FlowDesigner (FD))
  - データフロー指向ミドルウェア
  - Pull アーキテクチャ、オープンソース、低レイテンシ
- Robot Operating System (ROS)
  - ロボットミドルウェアのデファクトスタンダード
  - ネットワークベース通信、実装上の自由度が高い

## HARKネットワーク例 (同時発話認識)



- 音源定位 (MUSIC)
- 音源分離 (GHSS, HRLE, BF, Postfilter)
- 音声認識 (MFT-ASR)

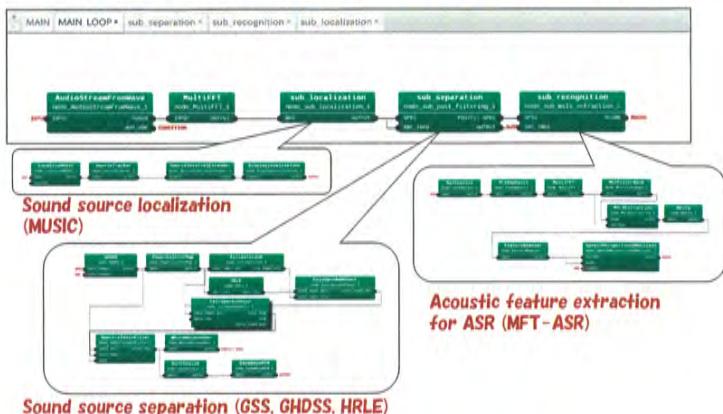
## HARKの応用



3人の注文  
終了後、  
ロボットの  
応答時間に  
注意！

Before:  
7.9秒  
After:  
1.9秒

## HARKの階層的なネットワーク構成



## 同時発話認識 2012



Copyright (C) 2012 Honda Research Institute Japan Co., Ltd. All Rights Reserved.

- 16ch マイクロホンアレイ使用
- 音源定位情報は所与

## 音環境理解・ロボット聴覚の展開

- Willow Garage社の Telepresence robot に HARK の移植+音環境可視化 (招待)
- 自己発話を抑制する
- 劣決定問題に挑戦 (音源数 $\geq$ マイク本数)
  - ノンパラメトリックベイズ
  - Sparse Factor Analysis
  - 補助関数法
  - Non-negative Matrix Factorization
  - 調波非調和混合モデル
  - 二次錐計画



## Lifelog音響版での音の可視化 [吉田]

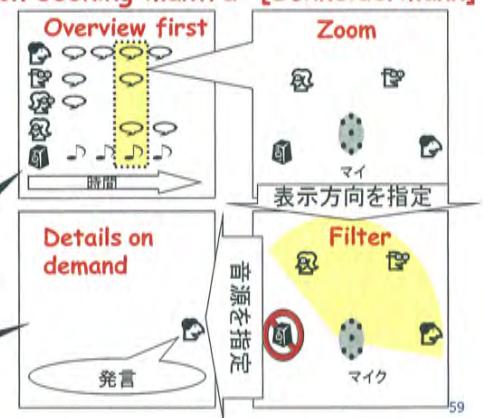
"visual-information seeking mantra" [Schneidermann]

ロボット聴覚  
システムと3D  
viewer を組合  
わせる。

O-ZF-Dレベル

時間的一覧  
性の支援

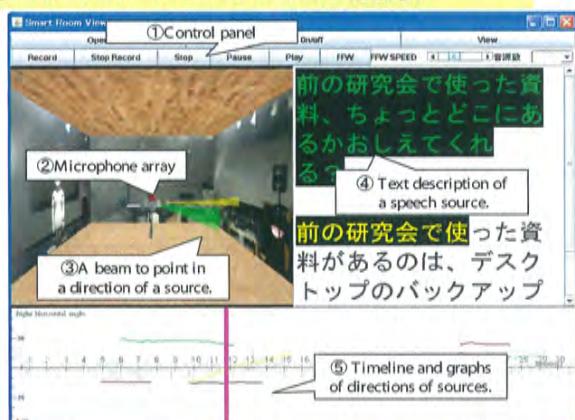
音の弁別  
の支援



59

## 音のライフログの2D表示

音声認識結  
果のカラオケ  
風表示



## HARK on Telepresence robot

Texai, Willow Garages's telepresence robots



## Demo: Texai with 3 people & Texai



<http://www.willowgarage.com/>

### 自分の音を抑制する

1. *Semi-blind separation* により自分の発話を抑制 [武田]
2. 音声対話システムでのバージイン（割り込み発話）で自分の声を抑制し、相手の声だけを聞く [松山, 駒谷]
3. 音楽共演ロボットが自分の演奏を抑制し、相手の演奏だけを聞く [水本, 糸原, Lim]
4. ペコッパ：相手の音だけに反応 [柳楽 with 渡辺先生 (岡山県立大) ]

#### Semi-blind ICA 既知信号抑圧を用いて 相手の声だけに反応してうなずくシステム：

- ・ 入力：目的信号と既知信号の混合信号
- ・ 出力：分離した目的信号
- ・ 難しさ：部屋の残響・目的信号（外乱）の影響  
→単純な減算×、従来手法：雑音区間検出が必要
- ・ 独立成分分析（ICA）を短時間周波数領域で適用
  - 目的信号と既知信号の独立性を利用
  - 残響と外乱に頑健な遂次処理を実現!



### 方言認識の勝てるストーリー

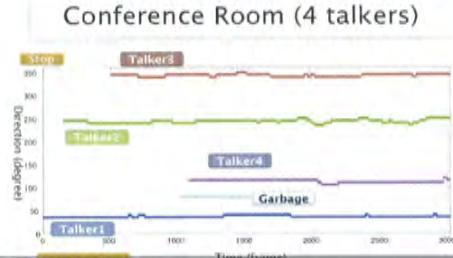
- 【批判】すでに終わっている研究 [K教授]  
【事実】人は複数の方言を混合して使用 [平山]  
【方法】方言混合比を推定し、方言認識を向上



- 【効果】複数の用語グループを混在して使用する比率がわかれば、個人の特徴がつかめる。  
【応用】用語違いの個人性が分かる  
⇒SNSに応用可。

## Demo: Texai with 3 people & Texai

Texai, 20 Talkers and 10000 words



Sound Separation

<http://www.willowgarage.com/>

### 自己発話抑制の勝てるストーリー

- 【問題】システム発話中の割込発話（バージイン）  
【課題】自己発話の元信号は既知、反射残響を推定して自己発話を抑制する必要がある [武田]  
【方法】semi-blind separation を独立成分分析（ICA）で構成、実時間処理のための工夫。  
【効果】マイクロフォンを常時ONにしても影響がなく、音情報が使える。  
【応用】・音声対話システムで割込発話から得られる情報を活用し、対話をスムーズに [松山, 駒谷]  
・音楽合奏ロボットで、自分の演奏音を抑制し、相手の演奏だけを聞きわける

### 音声対話に新しい視点

1. マルチドメインへの展開 [駒谷, 福林, 池田, 勝丸]
2. 顕現性に基づく談話文脈のモデル化 [白松]
3. バージインの処理 [松山]
4. 複数方言の混合モデル [平山]
5. 音響特徴量による言い換え [浜辺]

- ・新しい視点で未消費の分野に

### 感情・印象をとらえる

1. 認識・生成の双方向感情モデル [Lim]
2. 音響と歌詞の印象軌跡の統合 [西川]
3. 音楽と映像の調和度計算モデル [西山]
4. ロボット応答の個性 [奥乃]
5. 歌声の個人性の抽出 [池宮]

- ・複数モダリティから取り組む。
- ・モデル化・表現法に取り組み、一般化。
- ・より難しい問題に挑戦。

## 感情知性の勝てるストーリー

【問題】感情の認識・識別が別々の処理、モダリティごとに別々の処理

【課題】複数のモダリティを超えた、感情の認識・識別を可能とするモデル化 [Lim]

【方法】Multimedia Emotional Intelligence by SIRE (Speed, Intensity, irRegularity, and Extent) model.

【効果】音声、画像、歩容、身振りで実証。

【応用】・感情情報が統一的に扱えるので、モダリティを超えたインタラクションが可能に。  
・感情情報のデータマイニング等にも活用が期待。

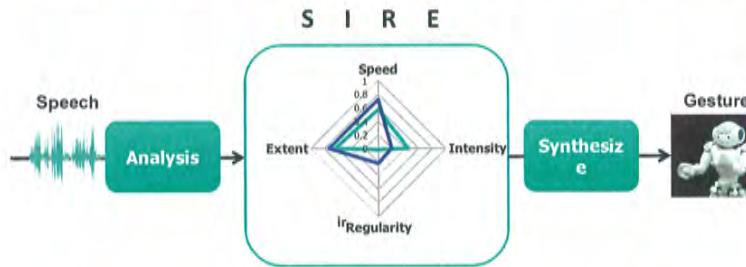
## 音声をSIREでパラメータ化

Speed	Intensity	irRegularity	Extent
<i>Syllables per second</i>	<i>Voice onset rapidity</i> $p(k) = \sum_{i=0}^{n-1} x(k+i)^2$ $\max_{k=1, \dots, N/n} p(k) - p(k-1)$	<i>(Inverse) jitter</i> $\frac{1}{N-1} \sum_{i=1}^N  x(t) - x(t-1) $	<i>Pitch range</i>



## The SIRE Emotion Transfer System

SIREモデルを通じて、感情を異なるモダリティに変換



low-dimensional • analysis/synthesis • for each modality

## 音楽情報処理への展開

特集

### CGMの現在と未来

初音ミク、ニコニコ動画、ピアプロの切り拓いた世界

#### 編集にあたって

後藤真作 (慶應法政総合研究所)  
島乃 博 (京都大学)

日本の技術、社会、文化の進歩の推進力  
一般的な人々が活発な創作活動を展開するための  
Consumer Generated Media (消費者生成メディア)  
N次創作  
能動的音楽鑑賞  
映像・音響の融合

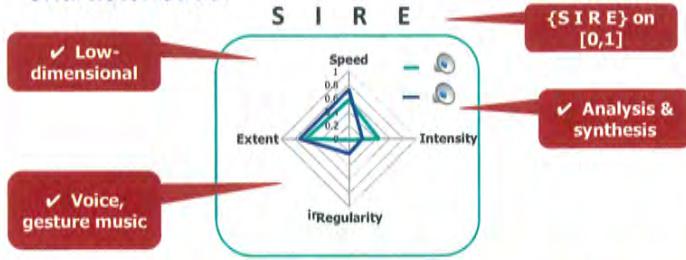
- CGM: Consumer Generated Media (消費者生成メディア)
- N次創作
- 能動的音楽鑑賞
- 映像・音響の融合



音楽情報処理の基礎と応用

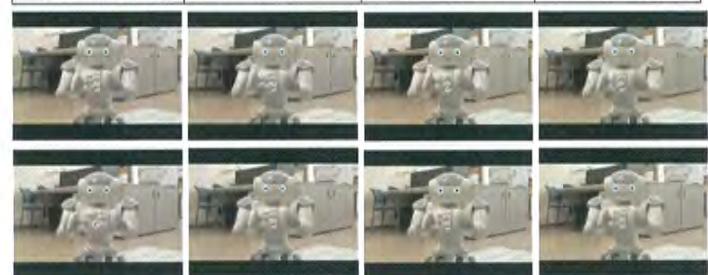
## Approach: The SIRE Model

**SIRE:** Description of Emotion based on Speed, Intensity, irRegularity and Extent. We represent emotion based on its dynamic characteristics.



## ジェスチャをSIREでパラメータ化

Speed	Intensity	irRegularity	Extent
<i>Joint speed</i> $t_0 = t_0$ $t_1 = \max(S, t_1, m)$ $t_2 = \max(S, t_2, m)$	<i>Joint acceleration</i> $t_0 = t_0$ $t_1 = \max(I, t_1, m)$ $t_2 = t_2$	<i>Phase offset</i> $\delta_t = (1 - R) \cdot t$ $\delta_t = \delta_t + t_f$	<i>Gesture size</i> $p_1 = p_0 + E \cdot (p_1 - p_0)$



## 楽器音やボーカルを聞き分ける

- ビートトラッキング [吉井]
  - 調波・非調波混合モデル [糸山, 安倍]
  - NMFによる楽器音分離 [糸山, 坂上]
  - 補助関数法による分離 [安良岡]
  - ボーカルを聞き分ける [藤原]
  - 歌い方を聞き分ける [池宮]
  - T-Fマスクで聞き分ける [池宮]
  - ギター演奏を聞き分ける [矢澤]
- 実演奏、楽器数で、より難しい問題に挑戦。

## 音楽パートの置換 [安倍]

- 混合音から置換する楽器パートを抽出するための音源分離

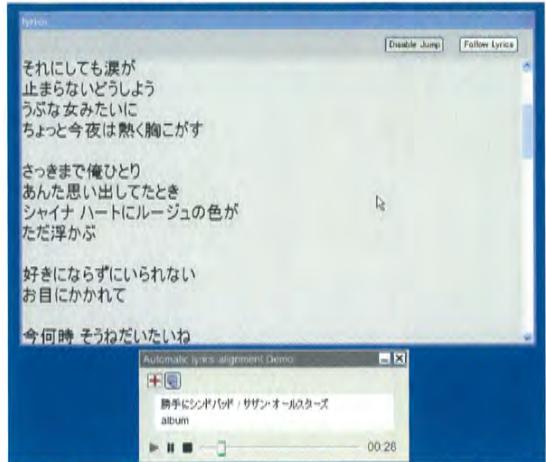


- 別の楽器音から楽器パートの構成に必要な音の音合成



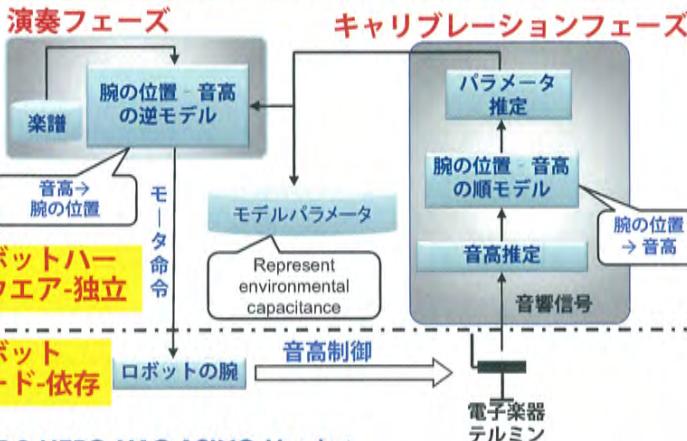
音の高さ、長さを変えてはもちろん、  
発音時の印象を変えるetc.  
様々な加工を施した音で音色置換可能  
※音量差がなくなるよう調節済み

## CD音楽に歌詞を時間的対応付け [藤原]

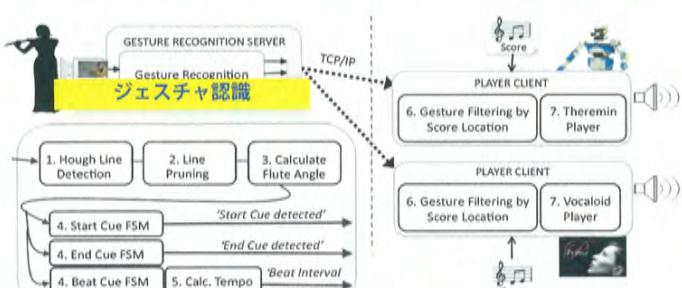


1. 文字ごとに音楽と対応付け
  2. 歌詞をクリックすればそこから演奏
- 原理: ボーカルパートを抽出  
応用: 電子辞書の対話データ(シャープ・パピルスで商用)  
松下幸之助人間力のページ(Dコモ, AU)

## テルミン演奏ロボットの構成[水本]



## ロボットアンサンブルシステム[Lim]



ロボットがテルミンを演奏し、歌うと同時に  
1) 開始と終了を同期  
2) 人のテンポに追従

93

## トリオ: NAO が Theremin を演奏



Nao knows the tempo and can groove with the humans.

## 音楽演奏ロボットの勝てるストーリー

【問題】 音楽ロボットはハードウェアとともに提供

【課題】 ポータビリティの高いテルミン演奏システムの開発[水本]

【方法】 ハードウェア非依存部とハードウェア依存部との分割.

【効果】 いろいろなロボットがテルミンを演奏できる.

【応用】 音楽共演ロボットの開発

⇒ 孤立化しがちな人に楽しみを

⇒ 人のサークルへ.

## HRP-2 が電子楽器テルミンを演奏



92

## Quartet: 2 Robots and 2 humans



Copyright 2011 Honda Research Institute Japan, Co. Ltd

Audio-Visual Integration for Beat Tracking [IEEE Humanoids 2012]

## 「聞き分ける」能力を室内から屋外へ

基盤研究(S) H19~H23



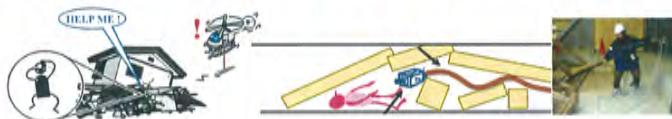
## 東日本大震災を目のあたりにして世の中の役立つことをしたい

1. 災害用ロボットでなぜ耳は使われていないのか
  - ホース型ロボット [坂東 with 田所]
  - 無人飛行ロボット [古川, 中臺, 公文]
2. サイバーフィジカルによる安全安心へのアプローチ
  - 音環境の可視化 [井山, 杉山]
  - 無人飛行ロボットによる音マップ [古川]
3. Human-in-the-loopによる人口ロボットの協調
  - ホース型ロボット用音源探索 [坂東, 井山, 杉山]
  - 無人飛行ロボットによる音探索 [京大, 東工大, 熊大]

## 災害用ロボットに耳をつける

【問題】画像だけでは不十分

【課題】実環境に対応できる音響信号処理

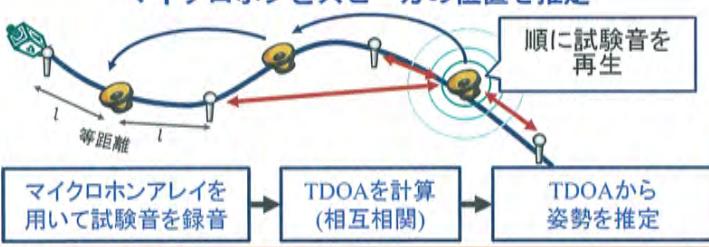


【方法】・無人飛行ロボット (UAV) で空中から音を取得  
・がれきの下はホース型ロボットで音を取得

【効果】これまで未活用であった音の情報を利用し、探索活動を効率化できる可能性がある。

## 姿勢推定 → マイクロホン位置推定

マイクロホンアレイと複数のスピーカーをホース上に搭載  
音の到達時間差(TDOA)の計算結果からマイクロホンとスピーカーの位置を推定



累積誤差の修正が可能  
マイクロホンアレイは声による被災者位置推定にも応用可

## 無人飛行ロボット・ヘビ型に耳を

東日本大震災を目のあたりにして世の中の役立つことをしたい



小型ロボ、空飛び回る

災害救助や警備など期待

## ホース型ロボットの勝てるストーリー

【問題】ホース型ロボットの姿勢推定

【課題】Navigation Usability と Mission Usability.



姿勢推定



音源定位  
音源分離  
⇒マイク位置

【方法】マイクロフォンとスピーカーを交互に配置し、TSP信号を使って、到達時間差から各マイク位置推定。

【結果】障害物が存在する狭路でも0.05mの高精度な姿勢推定を実現。突然の姿勢変化にもOK.

【今後】音源定位・音源分離を開発中。

## Experimental room for UAV

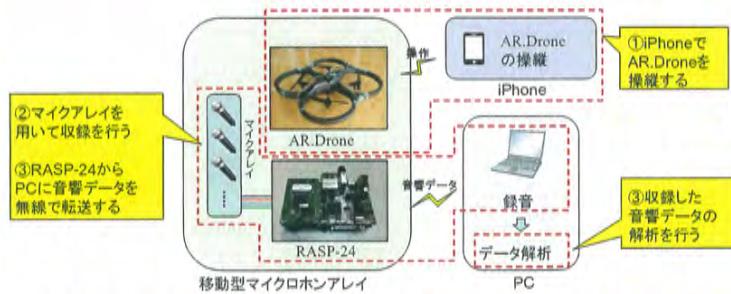


Indoor field for UAV



## 飛行体による音源定位システム[中臺]

製作した移動型マイクロホンアレイを用いて屋外環境音解析システムを作る



- 収録した音響データ解析にはロボット聴覚ソフトウェアHARKを使用  
- 解析ではGEVD-MUSIC法による音源定位を行う

## クアッドロコプターに聴覚機能を [公文]

Autonomous UAV with microphone array

- Sensor readings, motor command and audio signal can be measured synchronously.
- Accurate localization (reference data) : RTK-GPS
- 1.4kg+ payload, 10min flight time

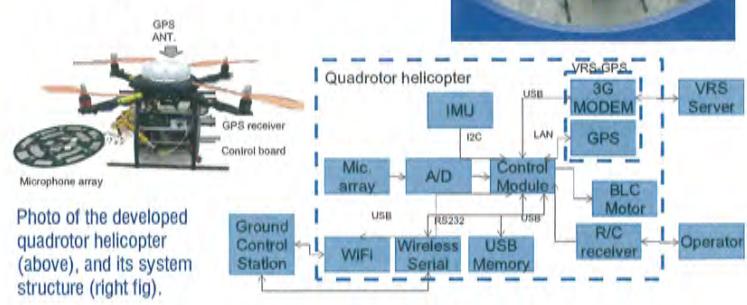
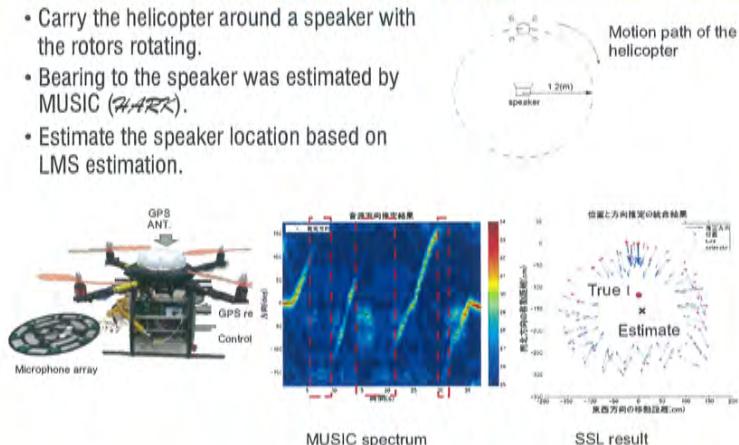


Photo of the developed quadrotor helicopter (above), and its system structure (right fig.).

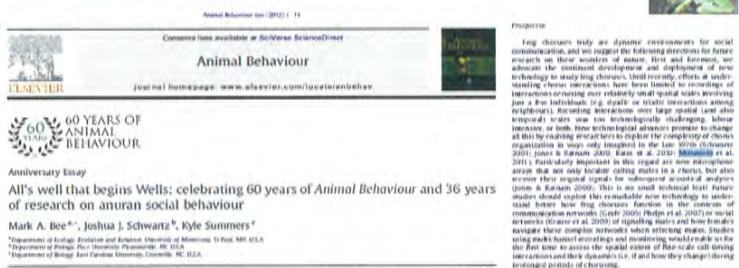
## 音源定位の性能

- Carry the helicopter around a speaker with the rotors rotating.
- Bearing to the speaker was estimated by MUSIC (HARK).
- Estimate the speaker location based on LMS estimation.

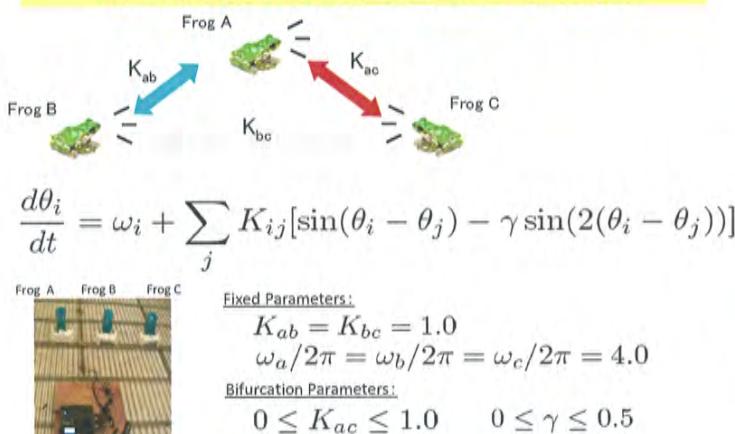


## 動物行動学への展開

- 音響信号処理は、小型夜行性動物の観測には不可欠の機能（画像処理、データロッガー不可）
- 動物行動学は、音環境理解・ロボット聴覚にとっては、未消費の市場、未開拓の市場。



## 3匹のカエルの合唱のモデル化



## Scientific Report 掲載での報道

科学技術報道（大阪）2014年1月28日 朝刊23~24面  
 科学技術報道（大阪）2014年1月28日 朝刊30~34面  
 引用許諾範囲範囲外禁止

### カエルの歌 輪唱してた



## カエルの合唱機構の観測での課題

### 1. 見つけるのが難しい

夜行性の小型動物：コウモリ、バッタ、カエル

### 2. 種同定

音源定位、音源分離が必要

### 3. 歌同定

分離音の歪に対応する必要。

### 4. ミクロ視観測とマクロ視観測の統合

ミクロ視的活動：音源分離による種同定・歌同定

マクロ視的活動：カエルホタルによる観測。



## カエルの合唱機構の測定による解明

### ・アマガエルの発声行動の数理モデリング[合原08]

- 2匹・3匹のアマガエルが同期して鳴く現象

- より多いカエル(多体系)への拡張

→ 水田で鳴き声を収録 (H20年6~7月)



アマガエル

### 1. 無数の雑音源

- 別種カエル・コウモリ・自動車

→ 雜音の位置が仮定できない

### 2. 広い空間

- 大量のマイクアレイが必要

→ 設置・回収が困難

### ・音光変換デバイスの開発



115

## 水田でのカエルホタルによる観測



音光変換デバイス  
「カエルホタル」の開発 [水本]



We can see

where the sound is.  
that two frogs are calling alternately.

Okuno Lab, Kyoto University



118

## 個別カエルのミクロ視観測

### ・カエルホタルによるマクロ観測

### ・HARKによるミクロ視観測



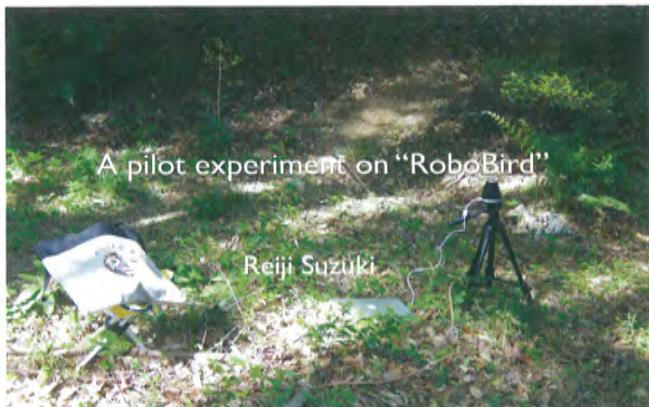
オーストラリア Gold Coast にて

## 野鳥の合唱機構の観測での課題

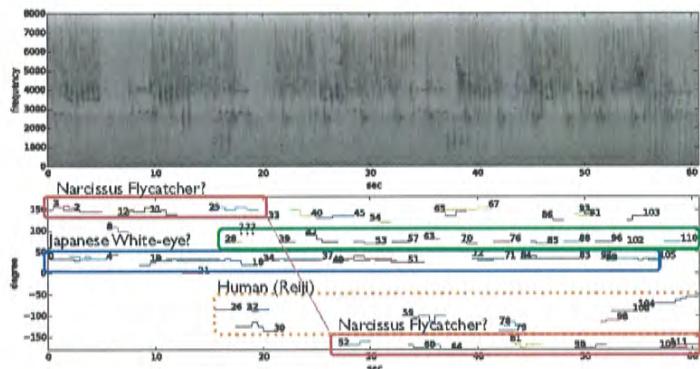
1. 見つけるのが難しい  
隠れていることが多い
  2. 種同定  
音源定位、音源分離が必要
  3. 歌同定  
分離音の歪に対応する必要
  4. 歌声発話区間認識 (Which species sing where, when, why and How) 歌重複回避行動の解明  
分散マイクロフォンアレイシステムによる継続的観測
- 【応用】野鳥の鳴き声の文法と意味の理解

## HARK を使って野鳥の観測

名古屋大学鈴木麗豊先生



## 野鳥の発話区間認識



# 日本国憲法

昭和21年11月3日

施行 昭和22年5月3日 (補則)

## 日本国憲法 前文

### 第1章 天皇

[第1条 天皇の地位](#)

[第2条 皇位の継承](#)

[第3条 天皇の国事行為に対する責任](#)

[第4条 天皇の機能](#)

[第5条 摂政](#)

[第6条 天皇の任命権](#)

[第7条 国事行為](#)

[第8条 皇室の財産授受](#)

[第2章 戦争放棄](#)

[第9条 戦争放棄、軍備及び交戦権否認](#)

[第3章 国民の権利及び義務](#)

[第10条 国民の要件](#)

[第11条 基本人権の不可侵](#)

[第12条 自由・権利の保持の責任とその濫用の禁止](#)

[第13条 個人の尊重](#)

[第14条 法の下の平等](#)

[第15条 公務員の選定及び罷免の権利、普通選挙と秘密選挙の保障](#)

[第16条 誓願権](#)

[第17条 国及び公共団体の賠償責任](#)

[第18条 奴隸的拘束及び苦役からの自由](#)

[第19条 思想及び良心の自由](#)

[第20条 信教の自由](#)

[第21条 集会・結社・表現の自由と通信の秘密](#)

[第22条 居住・移転及び職業選択の自由、外国移住及び国籍離脱の自由](#)

[第23条 学問の自由](#)

[第24条 家族生活における個人の尊厳と両性の平等](#)

[第25条 生存権、国の社会的使命](#)

[第26条 教育に関する権利と義務](#)

[第27条 勤労の権利・義務、労働条件、児童酷使の禁止](#)

[第28条 勤労者の団結権](#)

[第29条 財産権](#)

[第30条 納税の義務](#)

[第31条 法定の手続の保障](#)

[第32条 裁判を受ける権利](#)

[第33条 不当な逮捕をされない権利](#)

[第34条 抑留・拘束の禁止](#)

[第35条 住居不可侵](#)

[第36条 捜問と残虐刑の禁止](#)

[第37条 刑事被告人の権利](#)

[第38条 自白の証拠能力](#)

[第39条 遷及処罰・二重処罰の禁止](#)

[第40条 刑事補償](#)

[第4章 国会](#)

[第41条 国会の地位](#)

[第42条 両院制](#)

[第43条 両議院の組織](#)

[第44条 平等選挙](#)

[第45条 衆議院議員の任期](#)

[第46条 参議院議員の任期](#)

[第47条 選挙に関する事項](#)

[第48条 両院議員兼職の禁止](#)

修了年度	氏名	修士論文題目
2001	安達 史博	VoiceXMLの動的生成に基づく自然言語による電話音声対話システム
	山肩 洋子	ロボットとの音声対話のための信念ネットワークを用いた適応的言語理解
	Antoine Raux	Intelligibility Assessment and Adaptive Drill Generation for a Computer-Assisted Pronunciation Learning System
2002	伊藤 亮介	音声対話における感情の自動判別と会話ロボットへの実装
	住吉 貴志	情報検索HTMLフォームへの音声入力機能を有するWebブラウジングシステム
	長谷川 将宏	討論・講演音声に対するMPEG-7タグの設計と自動付与
	吉田 尚史	調波構造と音源方向に基づく音源分離システムにおける残響耐性の改善
	Ian R. Lane	Language Model Switching Based on Topic Detection for Multi-Domain Dialog Speech Recognition
2003	秋田 祐哉	討論音声を対象とした教師なし話者インデキシングとそれに基づいた自動書き起こし
	北原 鉄朗	音高による音色変化と未知楽器を考慮した楽器音の音源同定の研究
	櫻庭 洋平	音色・定位・音楽知識を用いた多重奏の自動探譜
2004	戸田 充彦	距離を利用したマルチモーダル情報統合によるヒューマノイドロボットの挙動選択
	石原 一志	環境音からの擬音語への自動変換と環境音検索システムへの応用
	小鷹 研理	一次反射音と身体的ダイナミクスを利用した仮想空間上の移動間隔提示
	山本 俊一	音源分離との統合によるミッシングフィーチャーマスク自動生成に基づく同時発話音声認識
	山口 健	唇画像を用いた母音認識における方向と距離の多用性に対する顔健化
2005	吉井 和佳	Drum Sound Recognition for Polyphonic Audio Signals by Adaptation of Spectral Templates and Suppression of Harmonic Structure
	神田 直之	マルチドメイン音声対話システムにおける対話履歴を利用したドメイン選択法
	田崎 豪	人間親密度の空間マッピングによるロボットと複数人とのインタラクション
	服部 佑哉	モダリティ間マッピングに基づく環境音を表現するロボット動作生成
	松本 祥平	人間ロボット協調のためのRNNPBを用いた疑似シンボルの獲得
2006	吉岡 拓也	室内伝達系と音声生成系の逆フィルタの同時推定によるブラインド残響除去
	西山 正紘	マルチメディアコンテンツにおける音楽と映像の調和度計算モデル
	藤原 弘将	調波構造の分離と歌声信頼度の推定に基づく歌声の認識に関する研究
2007	村瀬 昌満	RNNPBによる言語列と動作列の意味的結合と人間ロボットインタラクションへの応用
	糸山 克寿	調波・非調波統合モデルの制約付きパラメータ推定による多重奏音楽音響信号の音源分離
	武田 龍	ロボット聴覚におけるICAとMFT-ASRIに基づくバージイン発話認識
	徳田 浩一	聴覚障害者支援のためのHMDを用いた音環境可視化システム
	福林 雄一朗	WFSTに基づくラピッドプロトタイピング向け音声言語理解と想定外発話からのユーザ知識推定への応用
2008	横矢 龍之介	ロボットによる自己モデル投影に基づく他者の発見と模倣
	安部 武宏	インハーモニシティを組み込んだ調波・非調波統合モデルに基づく楽器音の音高・音長・音色操作
	池田 智志	想定外発話に頑健なマルチドメイン音声対話システムのためのドメイン選択とヘルプ提示戦略
	神田 尚	人工神経回路モデルと声道物理モデルを用いた母音模倣モデルに基づく音素獲得過程シミュレーション
	久保田 祐史	聴覚的アウエアネスの改善：顔動作による3次元音環境可視化システムの設計と実装
2009	王 凱平	Query-By-Example music retrieval approach based on music genre change by adjusting instrument volumes
	勝丸 真樹	複数の言語モデルと言語理解モデルによる音声理解の高精度化とそのラピッドプロトタイピングへの適用
	水本 武志	ロボットによるテルミン演奏のための音高・音量特性のモデル化とフィードフォワード制御
2010	大塚 琢馬	Real-time Audio-to-Score Alignment using Particle Filter for Co-player Robots
	日下 航	神経力学モデルを用いたロボットによる言語と感覚運動系の統合的認知
	前澤 彰	Score-Aided Inference of Classical Music Interpretation (楽譜を援用したクラシック音楽解釈の推論)
	松山 匠子	発語行為レベルの情報をユーザの意図解釈に用いた音声対話システム
	安良岡 直希	調波パラメトリックNMF及びダイバージェンス規準残響推定に基づく音響信号モデリングとフレーズ置換への応用
	張 陽	Prediction and Classification of Environmental Sounds using Recurrent Neural Network
2011	阿曾 慎平	歌声話声自動識別及び歌声の話声自動変換への応用
	西川 直毅	音楽情報検索のための歌詞と音響特徴量を用いた楽曲印象軌跡推定
	平澤 恭治	Under-Determined Blind Speech Separation Using a GMM-Based Sound Spectral Model
	Angelica Lim	Design and Implementation of Emotions for Humanoid Robots based on the Modality-independent DESIRE Model
	高野 秀樹	信頼度による調選択を用いた転調に適応可能な音楽音響信号に対する自動和音進行認識システム
	山川 暢英	突発性音事象の Matching Pursuit と Formant-wave function による認識とその擬音語変換のための特徴量設計と音源適応
2012	糸原 達彦	Beat-tracking method for music robot ensembles with a human guitarist
	阪上 大地	Optimization-Free Multipitch Analyzer based on Psychoacoustical and Musical Criteria
	柳楽 浩平	Nonparametric Bayesian Blind Source Separation in Frequency Domain with Permutation Resolution
	信田 春満	神経力学モデルを用いた予測可能性による身体識別と視覚運動系環境学習における身体の優先的獲得
	安周 慶	Real-Time Lead Sheet Following for Human Accompaniment by Particle Filter
2013	黄 楊暘	追体験メディアにおける注意誘導の音響提示手法
	佐野 正太郎	Designing Security Guidelines for COntinuous Visual and Audio CAPTCHAs with HMM-based Breaker
	平山 直樹	複数方言言語モデル混合による複数方言音声認識の高機能化
	望月 敏太	神経力学モデルを用いた描画運動における人間の認知モデルに基づくロボットの発達的模倣
	山口 雄紀	Deep Neural Network とマルチストリームHMMを用いた視聴覚音声認識
	矢澤 一樹	ギター演奏音響信号からの運指時の身体的制約と演奏難易度に基づくタブ譜自動生成

修了年度	氏名	博士論文題目
2002	天野 真家	Studies on Natural Language Processing for Kana-to-Kanji Conversion and Machine Translation (仮名漢字変換及び機械翻訳のための自然言語処理の研究)
	駒谷 和範	Spoken Dialogue Systems for Information Retrieval with Domain-Independent Dialogue Strategies (ドメイン非依存な対話戦略を備えた情報検索音声対話システム)
2003	渡邊 太郎	Example-Based Statistical Machine Translation (用例に基づく統計的機械翻訳の研究)
2004	秋葉 泰弘	Automatic Evaluation Methods for Machine Translation Systems (機械翻訳システムの自動評価法)
2006	北原 鉄朗	Computational Musical Instrument Recognition and Its Application to Content-based Music Information Retrieval (計算機による楽器音認識および内容に基づく音楽情報検索への応用)
2007	白松 俊	Salience-based Modeling of Discourse Context (顕現性に基づく談話文脈のモデル化)
	山本 俊一	Real-Time Robot Audition Software Based on Missing Feature Theory for Multiple Simultaneous Talkers in Real Environments (実環境における同時発話認識のためのミッシングフィーチャ理論に基づくリアルタイムロボット聴覚ソフトウェア)
	吉井 和佳	Studies on Hybrid Music Recommendation Using Timbral and Rhythmic Features (音色とリズムの特徴量を用いたハイブリッド型音楽推薦に関する研究)
2008	西出 俊	Self-Organization of Invariants for Motion Generation based on Reliable Predictability (予測信頼性に基づく動作生成のための不変項の自己組織化)
	金 鉉	Binaural Active Audition for Humanoid Robots (ヒューマノイドロボットのためのバイノーラルアクティブオーディション)
2009	藤原 弘将	Statistical Modeling for Recognizing Singing Voices in Polyphonic Music (統計的モデルによる多重奏音楽中の歌声の認識)
	吉岡 拓也	Speech Enhancement in Reverberant Environments (残響環境下での音声強調)
2010	糸山 克寿	Source Separation of Musical Instrument Sounds in Polyphonic Musical Audio Signal and Its Application (多重奏音楽音響信号に対する楽器音の分離とその応用)
	武田 龍	A Unified Framework of Blind Separation, Blind Dereverberation and Self-Voice Cancellation for Real-Time Robot Audition (実時間ロボット聴覚のためのブラインド分離、ブラインド残響除去、自己発話抑制の統一的枠組み)
2012	水本 武志	Synchronization among Interacting Individuals: Case Studies of Human-Robot Ensembles and Frog Choruses (人とロボットの合奏及びカエルの合唱における相互作用する個体間の同期)
	中村 圭佑	Robust Sound Source Localization Based on Microphone Array Processing for Robot Audition (ロボット聴覚のためのマイクロホンアレイ処理による頑健な音源定位)
2013	田崎 豪	People Detection based on Points Tracked by an Omnidirectional Camera and Interaction Distance for Service Robots System (サービスロボットシステムのための全方位カメラによるトラッキング可能特徴点とインタラクション距離情報を用いた人物検出)
	金 宜鉉	Improvement of Sound Source Localization for a Binaural Robot of Spherical Head with Pinnae (耳介付球状頭部を持つ両耳聴ロボットのための音源定位の高性能化)
	大塚 琢馬	Bayesian Microphone Array Processing (ベイズ法によるマイクロフォンアレイ処理)
	Angelica Lim	MEI: Multimedia Emotional Intelligence (MEI: マルチメディア・エモーショナル・インテリジェンス)
	神田 直之	Open-ended Spoken Language Technology: Studies on Spoken Dialogue Systems and Spoken Document Retrieval Systems (拡張可能な音声言語技術: 音声対話システムと音声文書検索システムにおける研究)

	学部卒業	修士課程修了	博士課程修了	教員
2001	上野 晋一 櫻庭 洋平 下岡 和也 横尾 昌弘	安達 史博 山肩 洋子 Antoine Raux		西井美季秘書 河原達也助教授 奥乃博教授
2002	石原 一志 北出 祐 翠 輝久 山本 俊一 吉井 和佳	伊藤 亮介 住吉 貴志 長谷川 将宏 吉田 尚史 Ian R. Lane 秋田 祐哉	天野 真家† 駒谷 和範	駒谷和範助手 助教
2003	神田 直之 田崎 豪 服部 佑哉 松本 祥平 吉岡 拓也	北原 鉄朗 櫻庭 洋平 戸田 充彦	渡邊 太郎	尾形哲也講師 助教授
2004	大庭 隼人 大谷 拓 海尻 聰 浜辺 良二 藤原 弘将 村瀬 昌満	石原 一志 小鷹 研理 山本 俊一 山口 健 吉井 和佳	秋葉 泰弘*	准教授
2005	糸山 克寿 田口 明裕 武田 龍 福林 雄一朗 横矢 龍之介	神田 直之 田崎 豪 服部 佑哉 松本 祥平 吉岡 拓也		
2006	池田 智志 神田 尚 清水 敬太 吉田 雅敏	西山 正紘 藤原 弘将 村瀬 昌満	北原 鉄朗(DC2) 丹羽治彦‡	高橋徹GCOE助教
2007	須見 康平 勝丸 真樹 斎藤 博己 張陽 水本 武志	糸山 克寿 武田 龍 徳田 浩一 福林 雄一朗 横矢 龍之介	白松 俊(DC2) 山本 俊一(DC1) 吉井 和佳(DC1) 内藤昭三*‡	西岡裕美秘書 西出白眉助
2008	中川 達裕 大塚 琢馬 日下 航 松山 匠子 安良岡 直希	安部 武宏 池田 智志 神田 尚 久保田 祐史 王 凱平	西出 俊(DC2) 金 鉉墩	糸山克寿助教
2009	穂山 空道 阿曾 慎平 栗野 皓光 平澤 恭治	水本 武志 勝丸 真樹	藤原 弘将* 吉岡 拓也*	吉井和佳講師
2010	糸原 達彦 佐野 正太郎 柳樂 浩平 信田 春満	大塚 琢馬 日下 航 前澤 彰 松山 匠子 安良岡 直希 張陽	糸山 克寿(DC1) 武田 龍(DC2)	
2011	壺内 将之 黄 楊暘 平山 直樹 望月 敬太 山口 雄紀 山村 祐介	阿曾 慎平 西川 直毅 平澤 恭治 Angelica Lim 高野 秀樹 山川 暢英		
2012	池宮 由楽 古川 孝太郎 田島 照久 水野 俊一郎	糸原 達彦 柳樂 浩平 阪上 大地 信田 春満 安 周煥	水本 武志(DC2) 中村 圭佑*	
2013	井山 貴裕 絵本 詩織 山崎 健史 中村 秀樹	佐野 正太郎 黄 楊暘 平山 直樹 望月 敬太 山口 雄紀 矢澤 一樹	田崎 豪* 金 宜鉉 大塚 琢馬(DC1) Angelica Lim 神田 直之* Louis-Kenzo Cahier‡	
在籍中		池宮 由楽 西牟田 勇哉 古川 孝太郎 坂東 宜昭 井山 貴裕	前澤 陽 *	

†論文博士 \*社会人博士 ‡研究指導認定退学