

人にやさしい音声インタフェース構築の 基盤ソフトウェアの開発

奈良先端科学技術大学院大学
鹿野清宏

H17.9.12

第11回 NAIST産学連携フォーラム

ユーザ負担のない話者・環境適応性 を実現する自然な音声対話処理技術

代表:

奈良先端科学技術大学院大学
鹿野清宏

参加機関:

奈良先端大、京大、和歌山大、名大、名工大、立命館大
日立製作所、旭化成、松下電器、松下電工、オムロン、ASTEM

期間: 2003.4 ~ 2008.3

音情報処理学講座

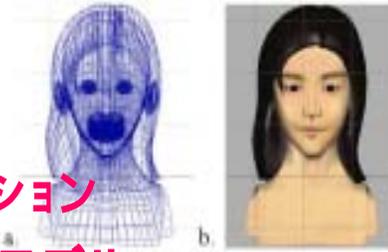
Speech and Acoustic processing Lab.

マルチモーダル情報処理

ブラインド音源分離
マイクロホンアレーによる
ハンズフリー音声認識

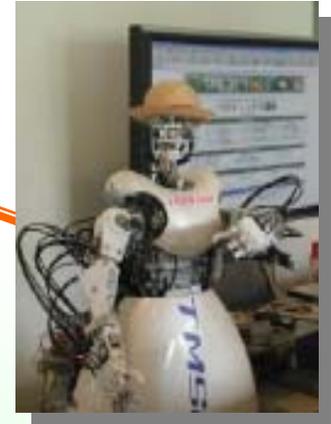


日本語ディクテーション
統計的音韻・言語モデル



ビジュアルエージェント

音声対話ロボット



音声認識
ウェブブラウザ

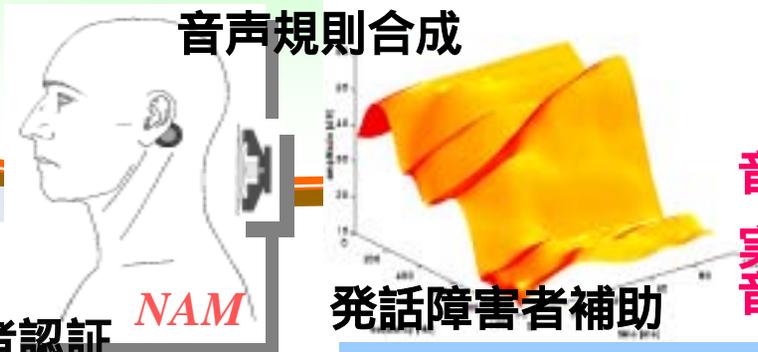


音声情報案内システム
実環境における頑健な
音声認識

自由な音声割り込みがで
きる音声対話システム

仮想音源の合成

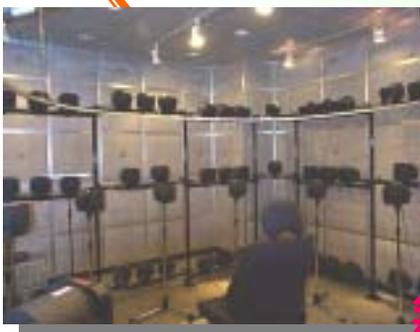
声質変換
音声モーフィング
音声規則合成



発話障害者補助

無音声認識
無音声電話
NAM発話者認証

音場再現



音声認識・合成により、誰もが気軽に情報機器の利用できる社会へ！

大語彙連続音声認識
フリーソフトウェア

ユーザ負担のない
話者・環境適応

どのような環境でも、
高精度に認識。
マイコン上でも。

誰もが使える
子どもからお年寄りまで



実証開発、
評価

マイクなしで、
DSPチップで廉価。

多様な自然な
応答音声。

ハンズフリー
音声認識

ヒューマンインタフェース、
構築ガイドライン

多様な声質の
音声合成

大語彙連続音声認識プログラム

サブリーダー 河原 京大教授

奈良先端大、名古屋大、和歌山大、日立、ASTEM

H16-17

カスタマイズ
化を容易に

H15-16

幅広い話者層



H16

非音声のGMM
による棄却

多様な端末・機器からの入力



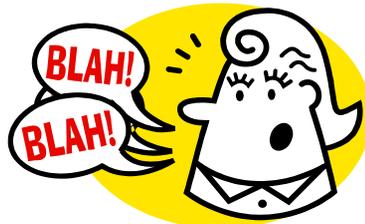
H15

H16

SAPI/SALT対応
マイコン移植

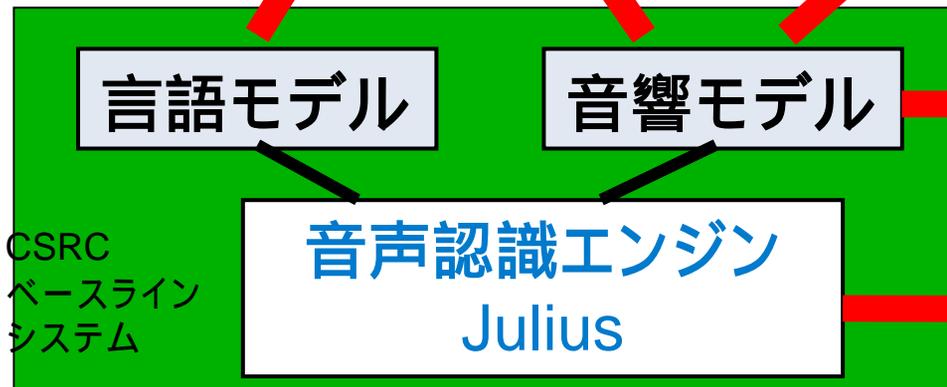
H16

会話調の音声 / 表現への対応



H15

幅広いカバーレージ



- 京都大学で開発
- SAPI・SALTの利用
- コンピュータの使い方に関する質問を音声で受付・回答
 - 「両面印刷の方法」
 - 「Wordで数式を入力したい」

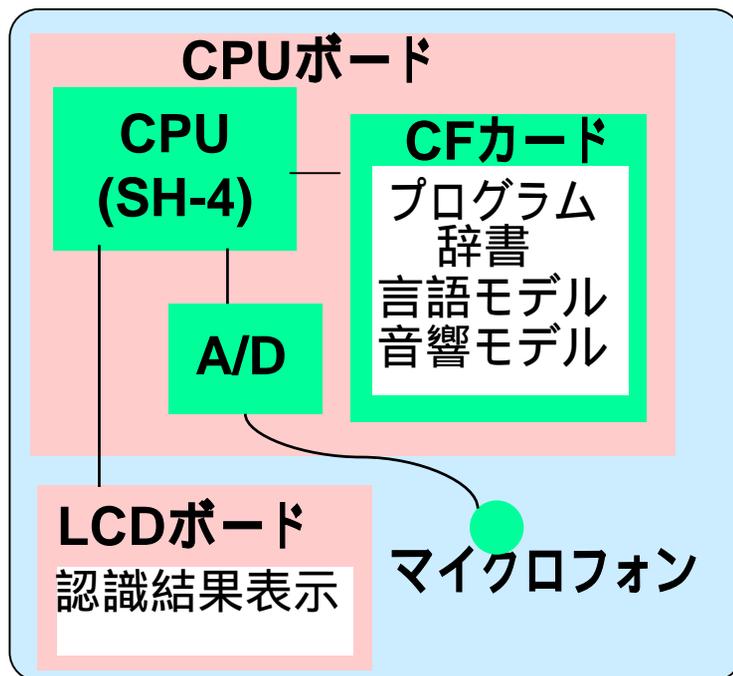
京都大学メディアセンター
オープンスペース端末室に設置



大語彙連続音声認識ソフトJuliusの マイコンへの移植

日立との共同研究

- 開発プラットフォーム 日立「T-engine」 SH - 4
 - TRONアーキテクチャに対応した共通開発ボード



SH-4 → SH-4A、アルゴリズムの高速化→実時間動作 (H19.3)

マイコンボードの仕様

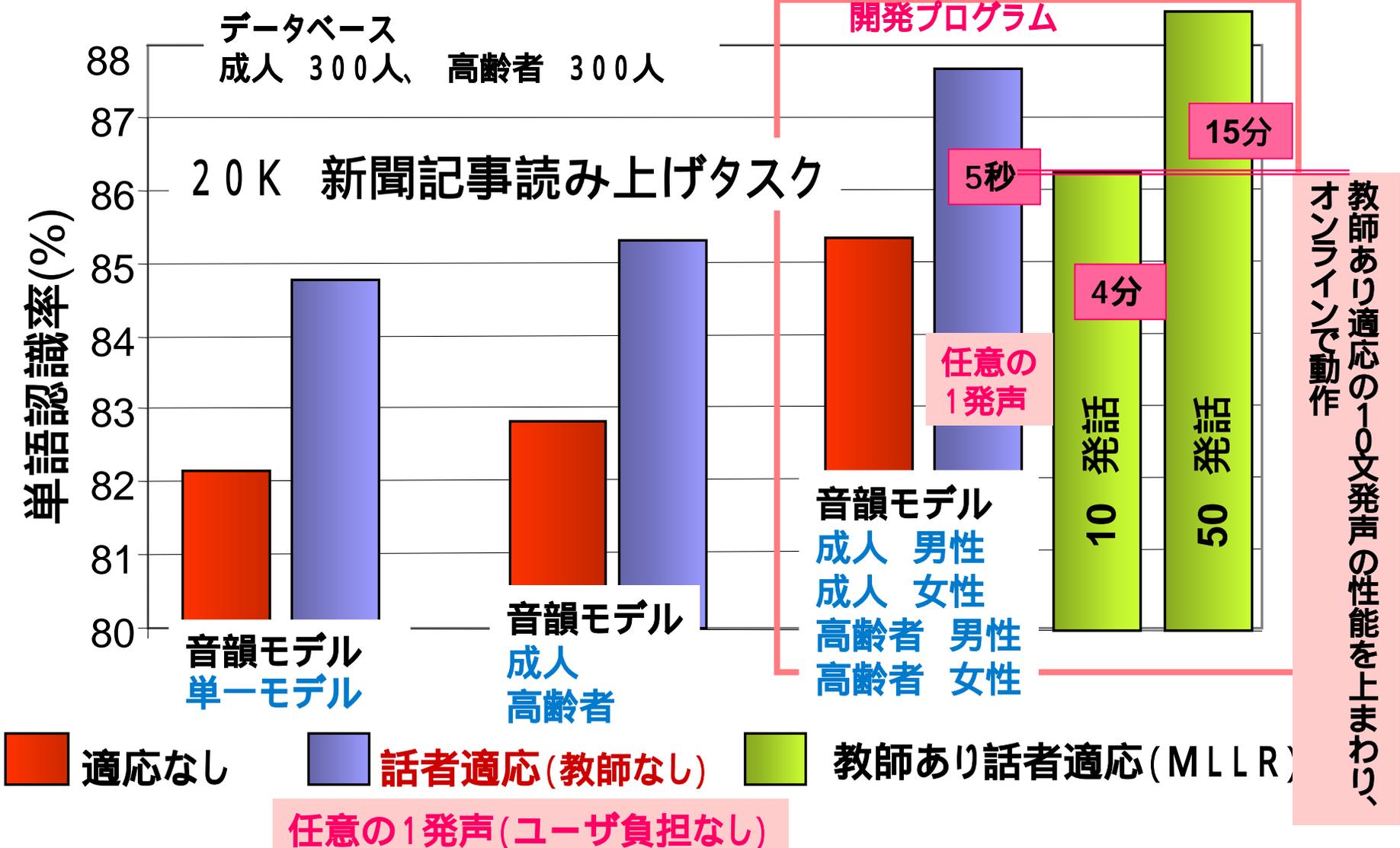
性能 (2004年11月現在)

CPU	SH-4 (240MHz) SH-4A 400MHz
フラッシュメモリ	8Mbyte
ワークメモリ	64Mbyte (SDRAM)
OS	T-Kernel
入出力I/F	USB, PCMCIAカード, シリアル, ヘッドフォン, マイク, LCD, 拡張バス, 他
LCDボード	TFTカラー 240x320
外形	120mm x 75mm

	簡易版	高精度版
語彙サイズ	5,000	5,000 SH-4A 20,000
音響モデル	モノフォン	トライフォン (PTM)
言語モデル	trigram	trigram
ビーム幅	400	400
単語認識精度	86.1%	90.7%
処理速度 (T-engine)	1.5 × RT	3.6 × RT SH-4A 実時間

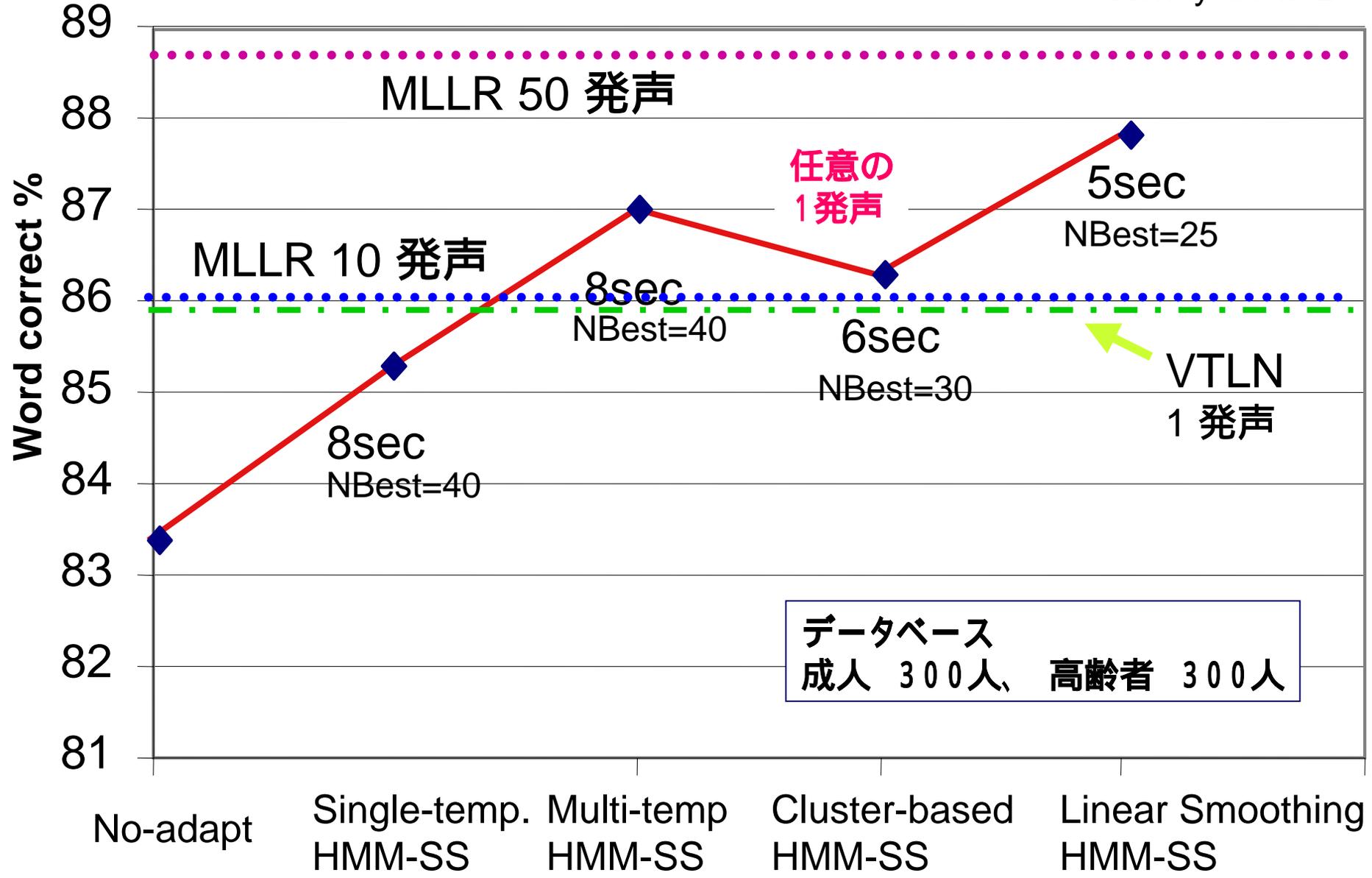
ユーザ負担のない話者・環境適応

サブリーダー 鹿野 奈良先端大教授
松下電器、松下電工



ユーザ負担のない話者・環境適応

Randy Gomez



無音声認識・無音声電話 Non-Audible Murmur (NAM)

電子情報通信学会 論文賞 猪瀬賞

つぶやき声(NAM)は、話し手の近くでも聞こえない声。
この声を認識(無音声認識)したり、電話で送る(無音声電話)ことができる。

NAM マイクロフォン



無音声電話

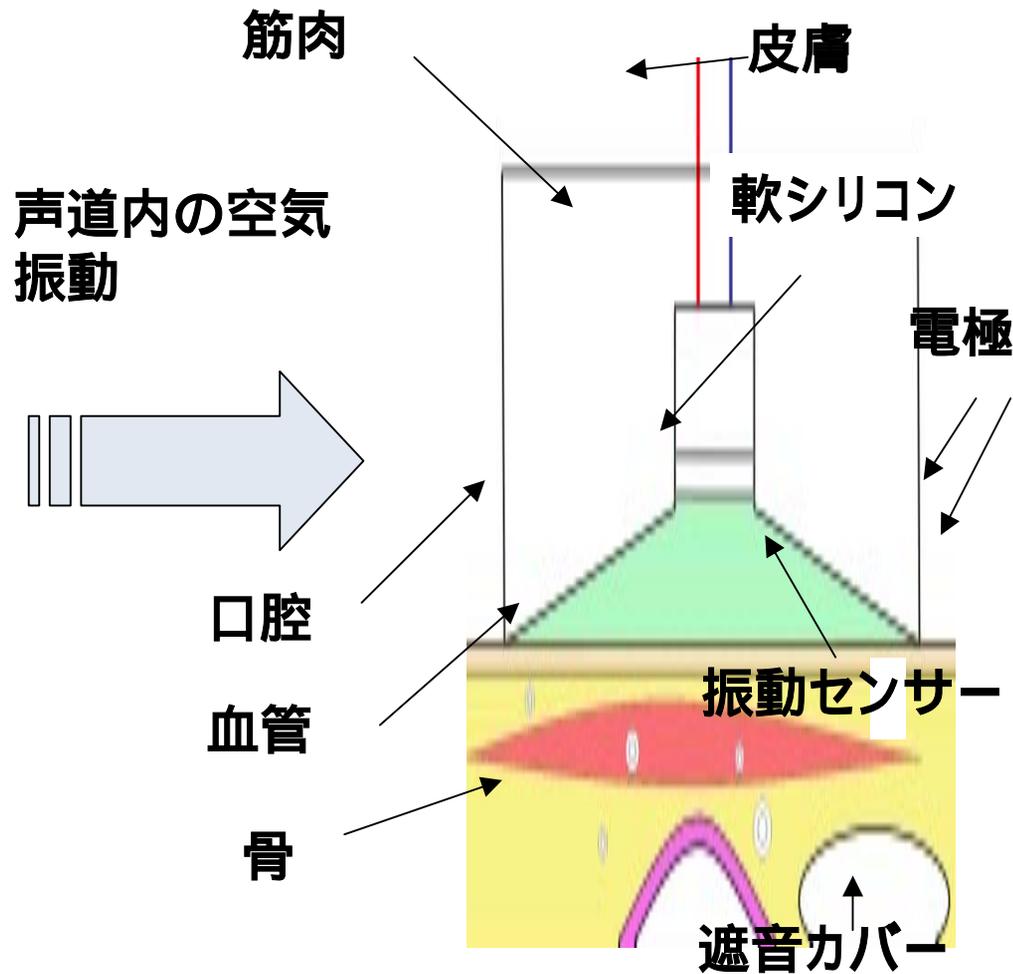


私は名義を貸しただけで株購入をしていない。たとえ逮捕されても裁判で徹底的に闘う



NAMマイクロフォン

Y.Nakajima



NAM (Non-Audible Murmur: 非可聴つぶやき音) や微弱振動音源による微弱な声道共振音の収録が可能

NAM音声

NAMマイク

外部マイク

NAM
無音声
つぶやき声



Murmur
ささやき声

Usual
Speech
通常音声

NAMは音声の応用範囲を 拡大する新しい静かな音声メディア

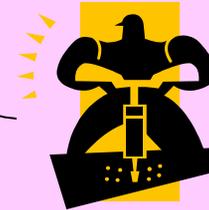
静かな音声メディア



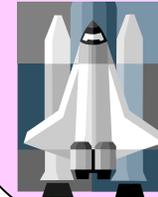
プライバシー
オフィスでの音声認識
会議中でも電話が可能



耐雑音性あり



口の中の音波を
直接に集音



NAM マイクロフォン

発話障害者補助

障害者の音声
コミュニケーション



これも重要

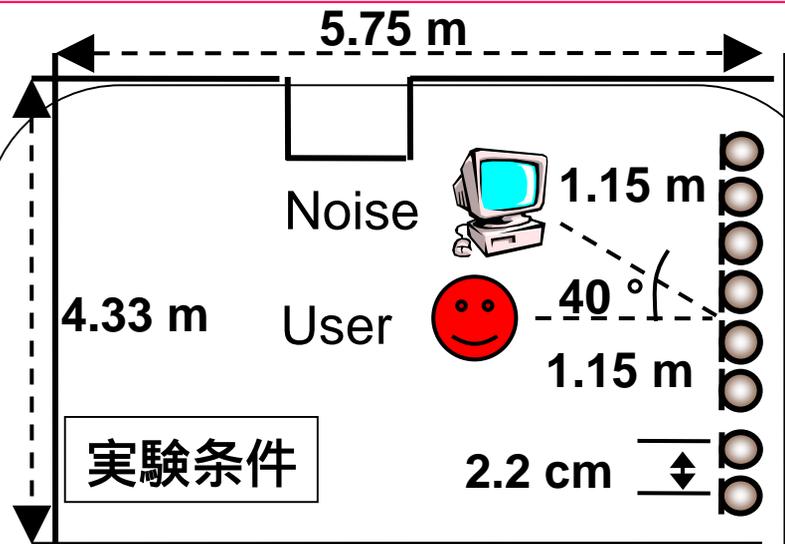
個人認証
列車の中でも電話



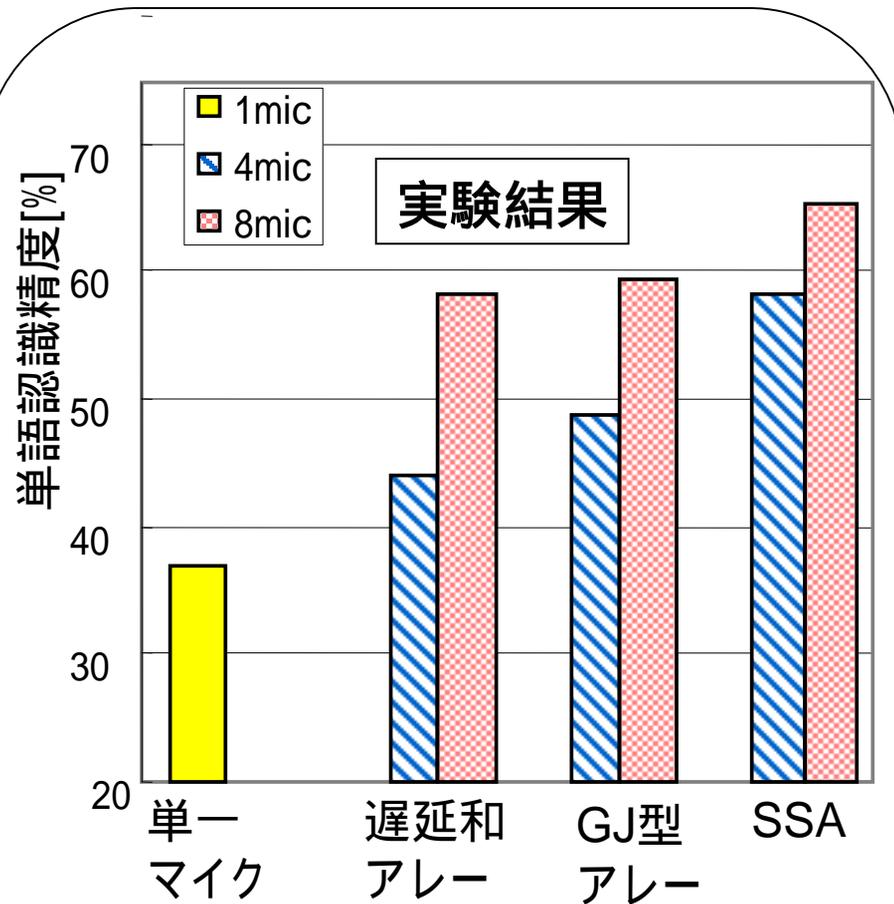
What's
NEW?



空間スペクトル減算アレー (SSA)

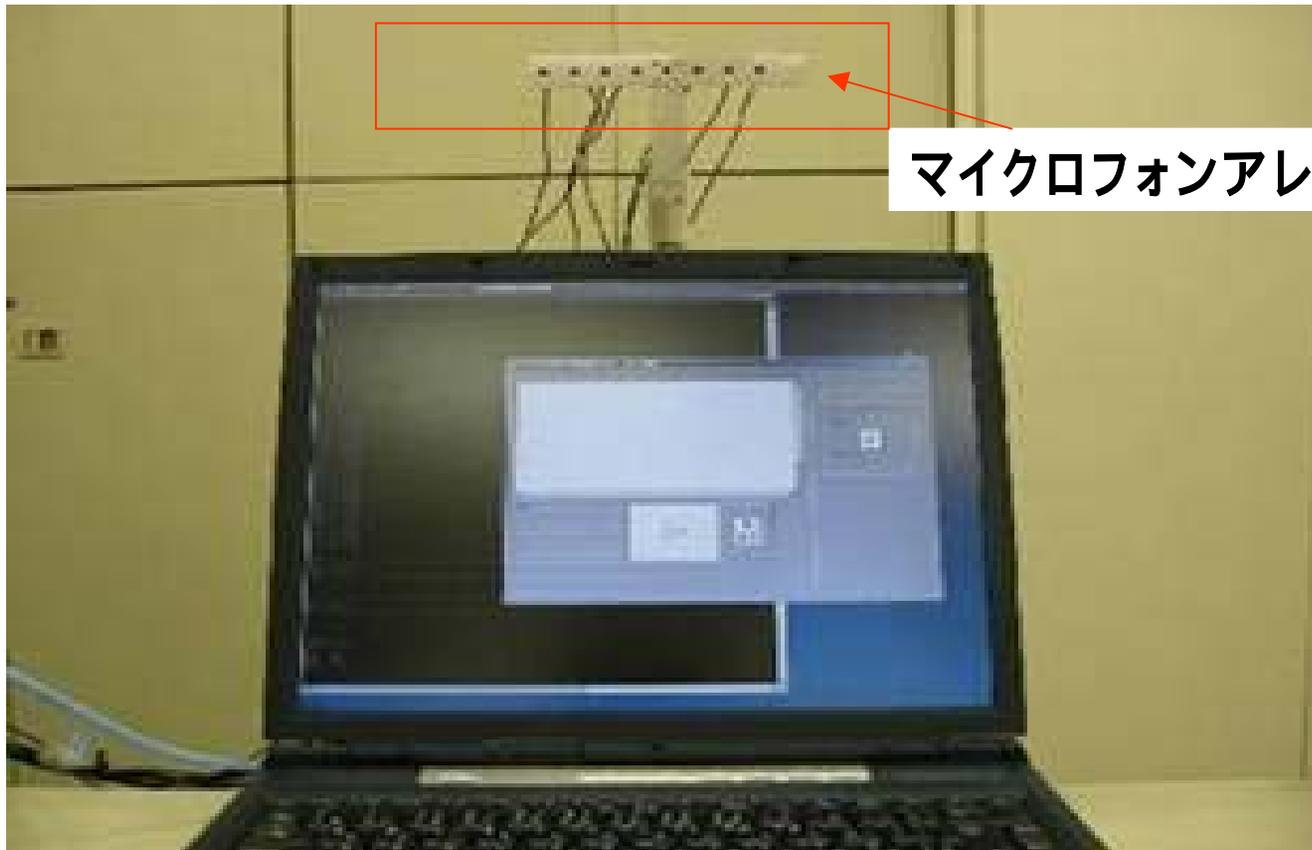


- タスク: 新聞読み上げ (語彙サイズ20K)
- 音響モデル: PTM (JNAS 不特定話者)
- 評価データ: JNAS テストデータ200文章
- 残響時間: 200 ms
- SN比: 平均 5 dB



ブラインド音源分離 (SIMO-ICA) も並行して研究

少ないパラメータで従来法よりも優れた性能が得られている



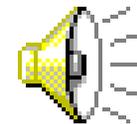
旭化成のハンズフリー音声認識実験システム

DSPさらにFPGAによる実装 (H19.3)

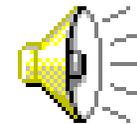
BSS音源分離プログラムオンラインバージョン (H18.3)

◆ 残響時間 (Reverberation time) 300 ms

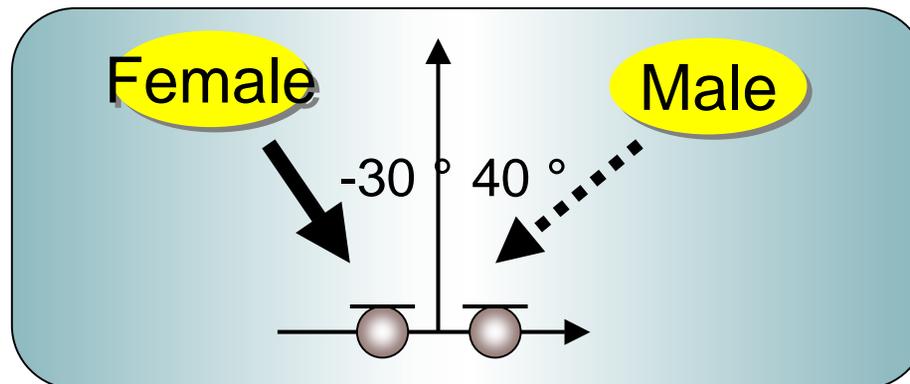
● 混合音声 (Female , Male)



● 分離音声 (Female)



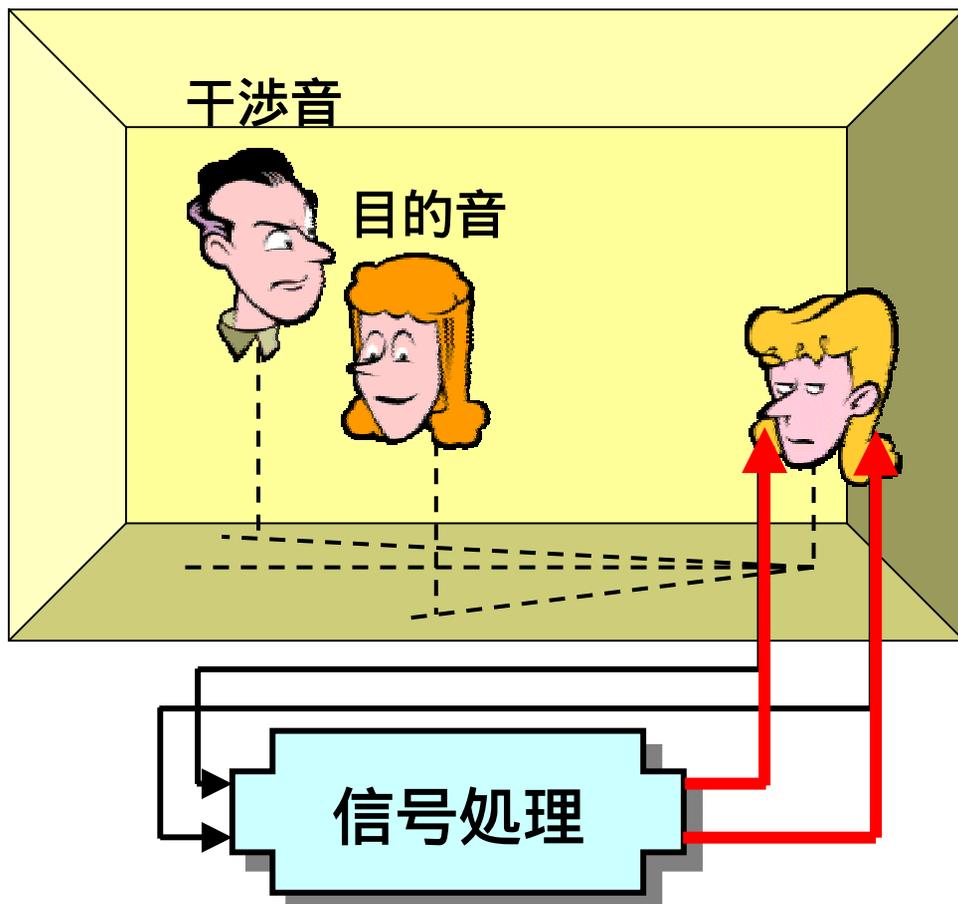
● 分離音声 (Male)



2 elements , 4 cm distance

Single-Input Multiple-Outputモデルに基づく独立成分分析 を用いたバイノーラル混合信号のブラインド音源分離

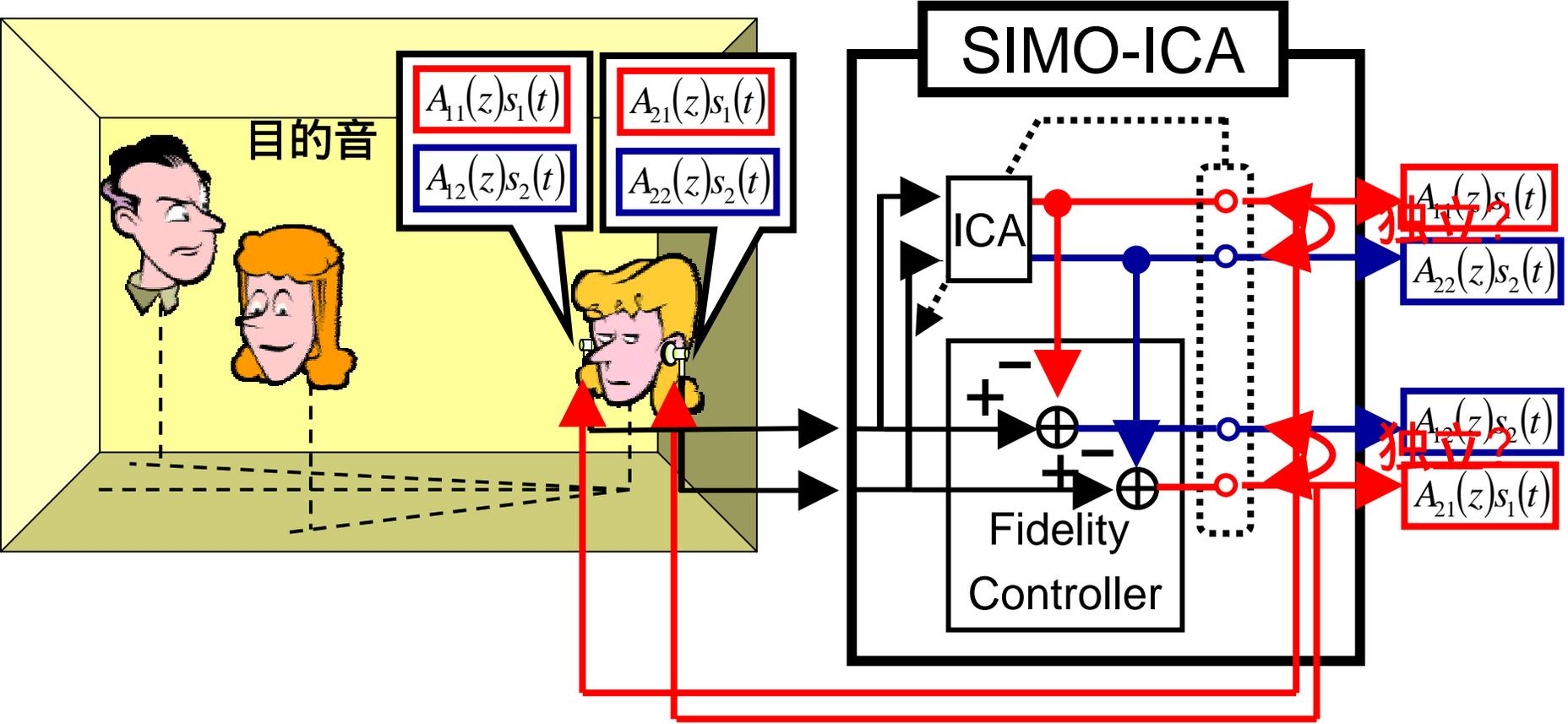
Tomoya Takatani



両耳介入り口付近の音を收音

目的音のバイノーラル信号のみを再現する
音響拡張現実感システムの実現

SIMOモデルに基づくICA (SIMO-ICA)



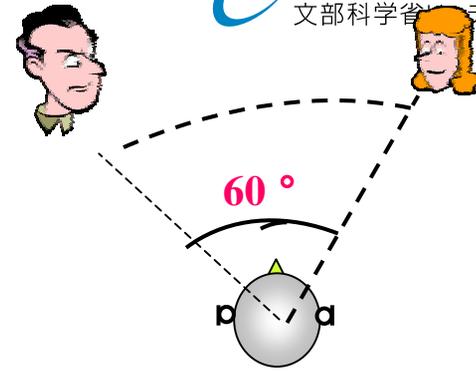
2組の出力信号が互いに独立
出力信号はSIMOモデル信号(数学的に証明)

提案法の有効性を検証するために、バイノーラル混合信号の分解実験を行った。

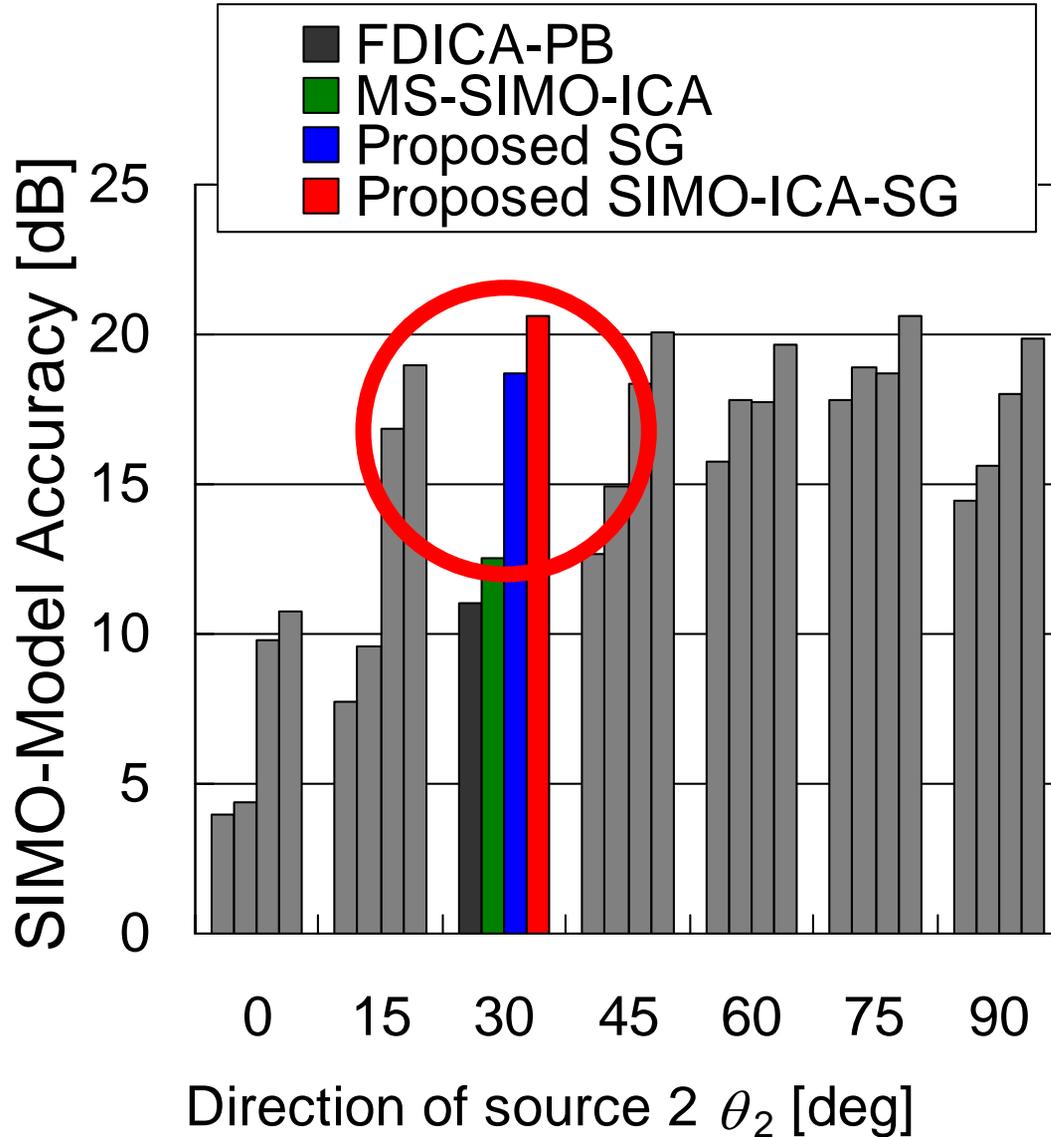
- 従来法1: FDICA-PB
- 従来法2: MS-SIMO-ICA
- 提案法: SIMO-ICA with Self Generator for Initial Filter

* 提案法の総学習回数が、FDICA-PBの学習回数と同数になるように設定

実験結果

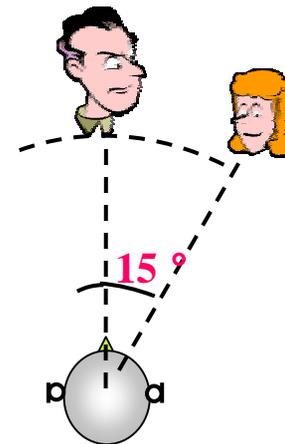
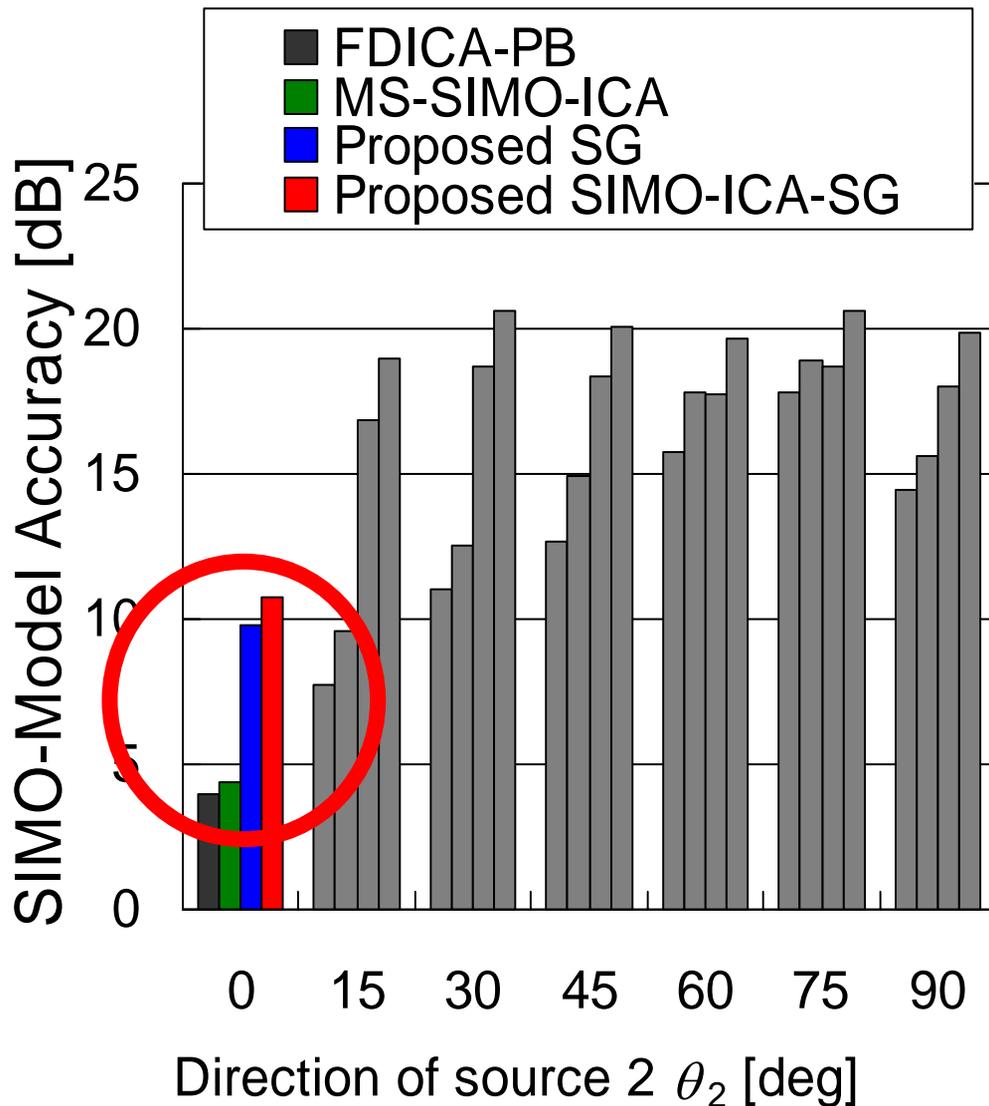


音源1の方位: -15度
音源2の方位: 45度



	Input	
従来法 1 FDICA-PB	Output 1	Output 2
従来法 2 MS-SIMO-ICA	Output 1	Output 2
提案法	Output 1	Output 2

実験結果



	Input	
従来法 1 FDICA-PB	Output 1	Output 2
従来法 2 MS-SIMO-ICA	Output 1	Output 2
提案法	Output 1	Output 2

実時間BSSのDSPへの実装



DSP: TI
TMS320C6713
200MHz

神戸製鋼との共同研究

サブリーダー 武田 名大教授
奈良先端大、京大、オムロン



公共の場の情報案内システム
(奈良先端大)

バス運行問い合わせシステム
(京都大)

車内情報案内システム
(名古屋大)

情報家電の音声制御システム
(オムロン)

2年以上運用、
200,000 発注

2年以上運用、
20,000 発注

900人、3モードで
冬モード100,000 発注

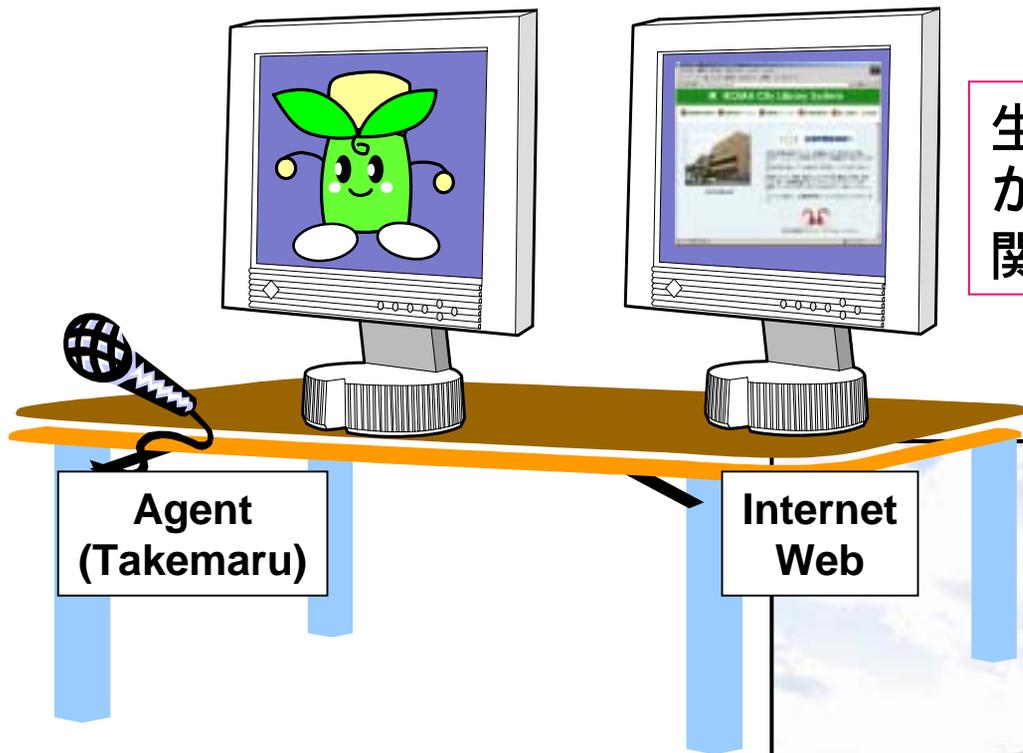
実環境音声対話システムで実証、評価のフィードバック

Dialog Navi.
展示場案内、
万博ロボット

プログラム開発

多量のデータに基づく開発、
実システムでの評価

音声情報案内システム「たけまるくん」



生駒市のキャラクター「たけまるくん」が規則合成音で答えて、Webなどの関連情報を表示する。

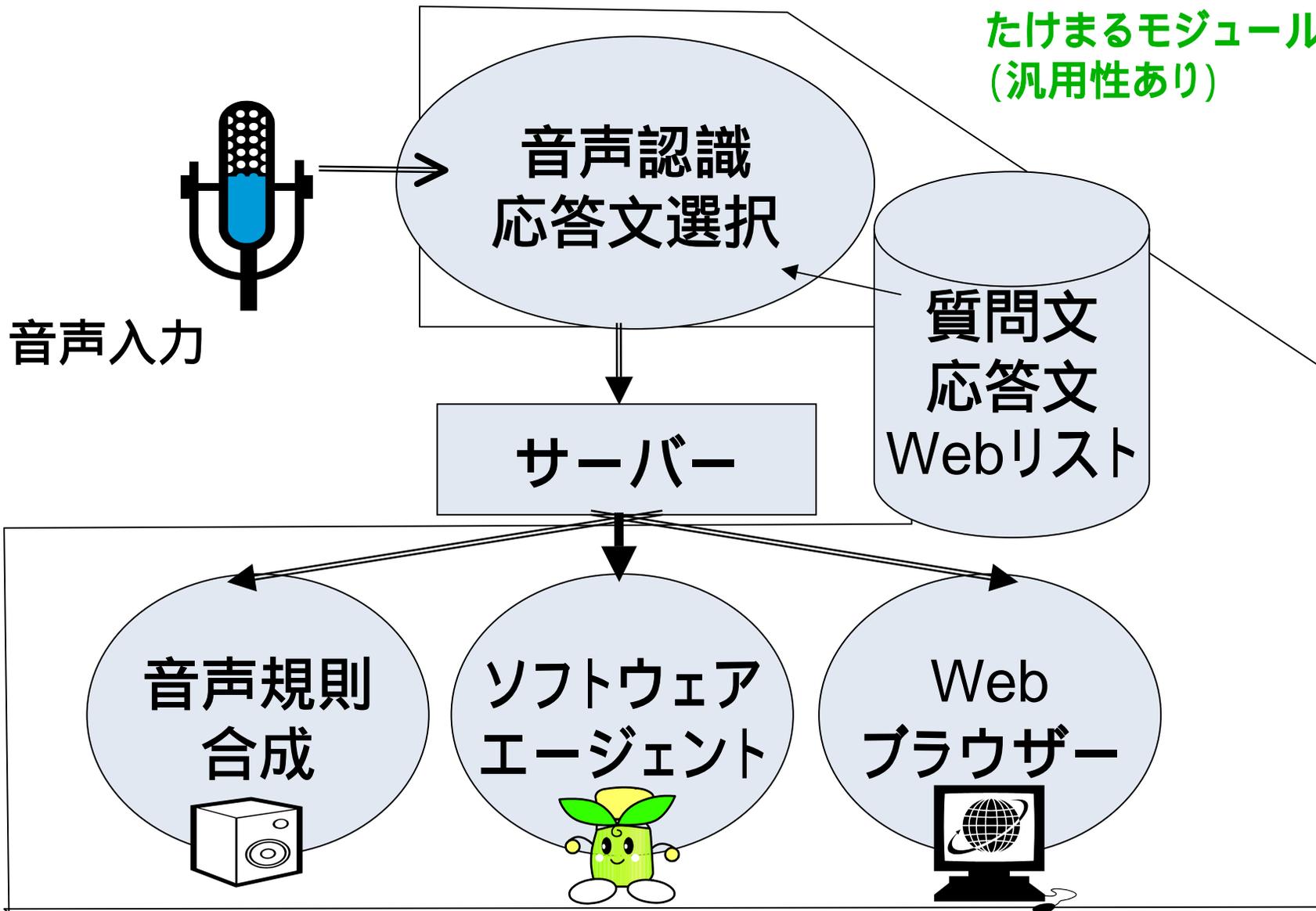
生駒市北コミュニティセンター
に設置 (2002.11)

2年8ヶ月 連続稼動



プログラムの構成

たけまるモジュール
(汎用性あり)

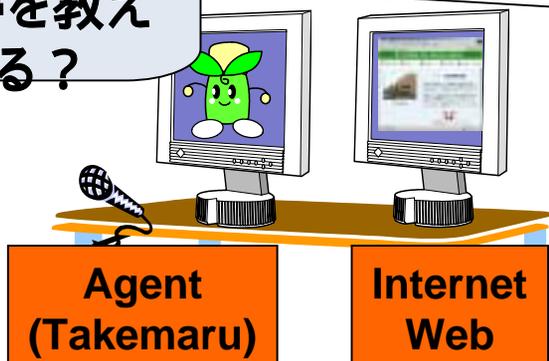
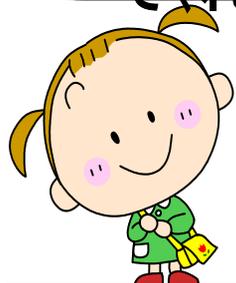


情報案内システム「たけまるくん」

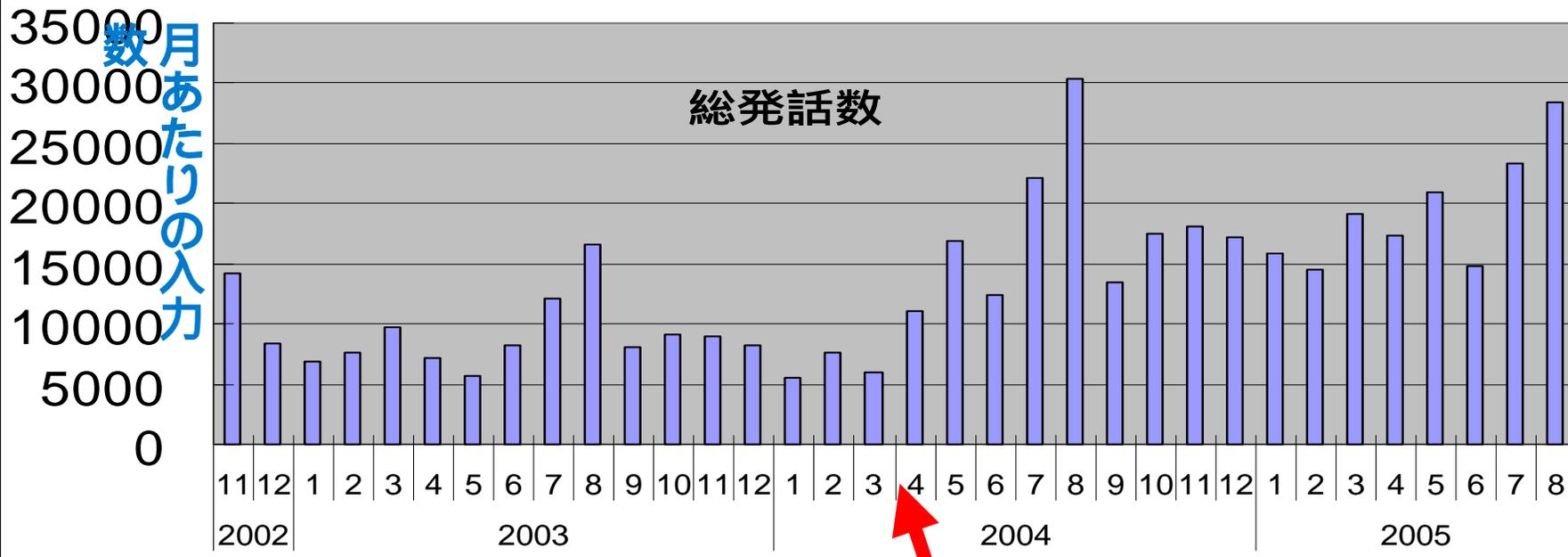
こんにちは、近くのバス停を教えてくださいませんか？

バス停は川を横切ったところにあります....

生駒市北コミュニティセンターに設置 (2002.11)

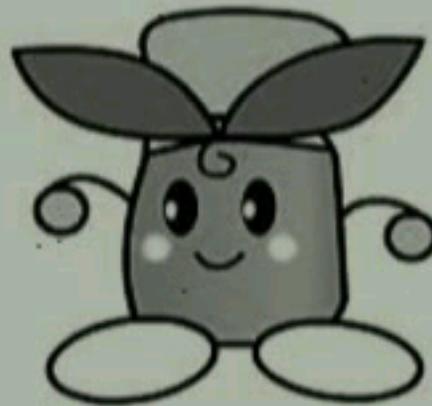


- 2005年8月
- 総入力数: 330,000
- 一日あたりの平均入力数: 800



- 長期間にわたる音声データの収集
- 構築のノウハウ
- 構築コストの削減

子供/大人の識別
雑音の識別



応答文の選択

- **一問一答タイプ**: 入力発声に近い質問文を、質問データから選択し、対応する応答文で応答し、Webを表示。
- 応答文の種類: 500, (スロットタイプ: 3)

001 こんにちは。

208 今は、<<hour>>時<<min>>分です。

212 バスの時刻表を表示します。

301 トイレは、左の奥か、はばたきホール、入り口の近くにあります。

305 図書館の入り口は、市民サービスコーナーの横です。

- 質問データベースの総数: 約10,000(子ども、大人)

トイレ+トイレ+2 は+ワ+65、+、+79 どこ+ドコ+14 です+デス
+74/56/1 か+カ+70? +? +77#301

食堂+シヨクドー+2 は+ワ+65、+、+79 あり+アリ+47/17/5 ます
+マス+74/58/1 か+カ+70? +? +77#332

- 年齢層別の発話パターンを応答生成に反映
 - 収集した書き起こしテキストから用例テキストの作成
- 年齢層の性質を考慮した応答
 - 年齢層別応答候補の作成

たばこを吸いたい

大人「喫煙コーナーは、この壁の裏側にございます」

子供「たばこなんてすったら駄目」

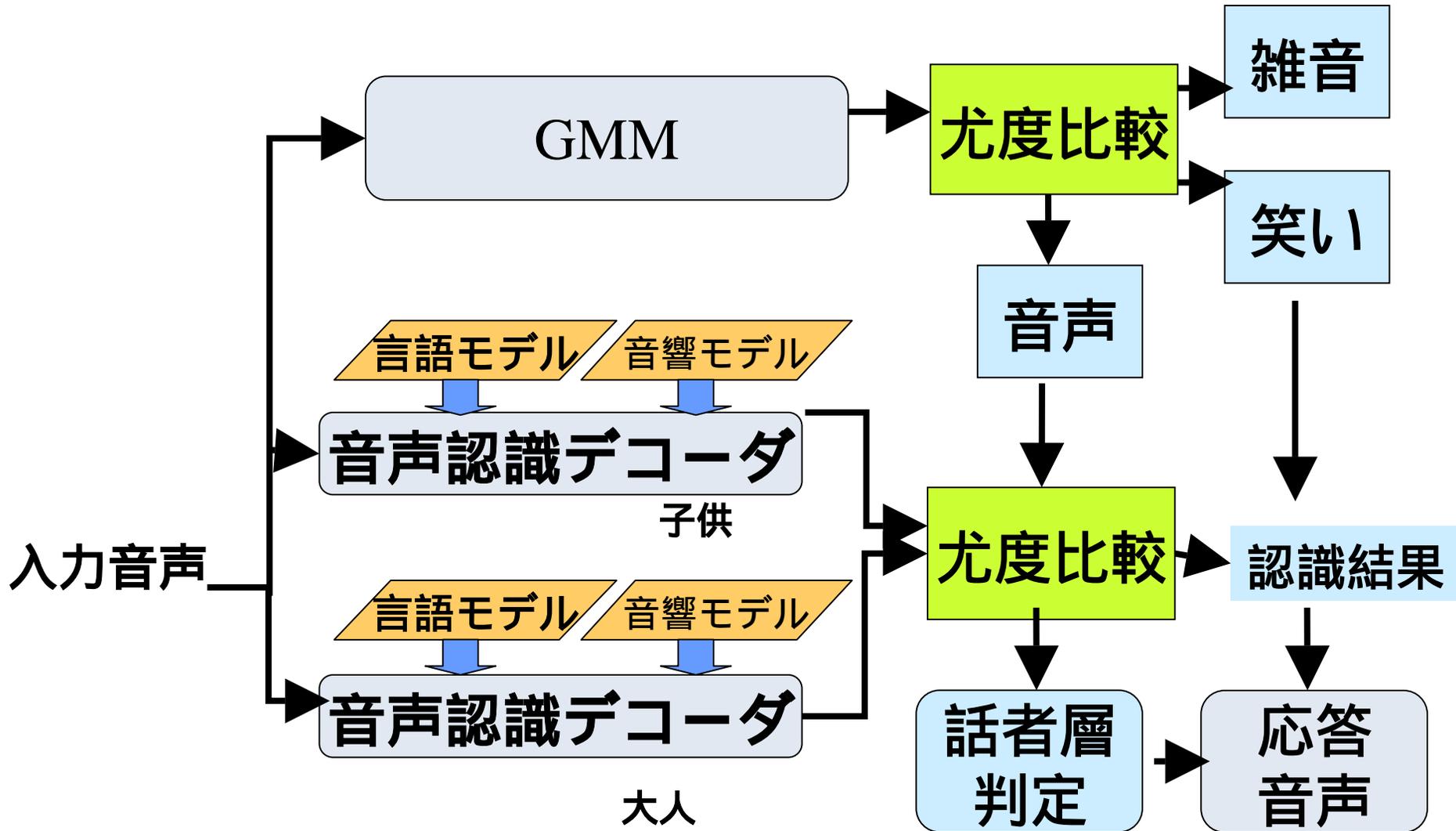
くろんど池はどこですか？

大人「くろんど池は、玄関前の道をしばらく北上したところにあります」

子供「くろんど池に行くには、バスに乗ってね」

- 何歳？   
- こんにちは  
- おはよう  
- 各種案内  
- その他 (Others)    

並列デコードデコーディングによる実装

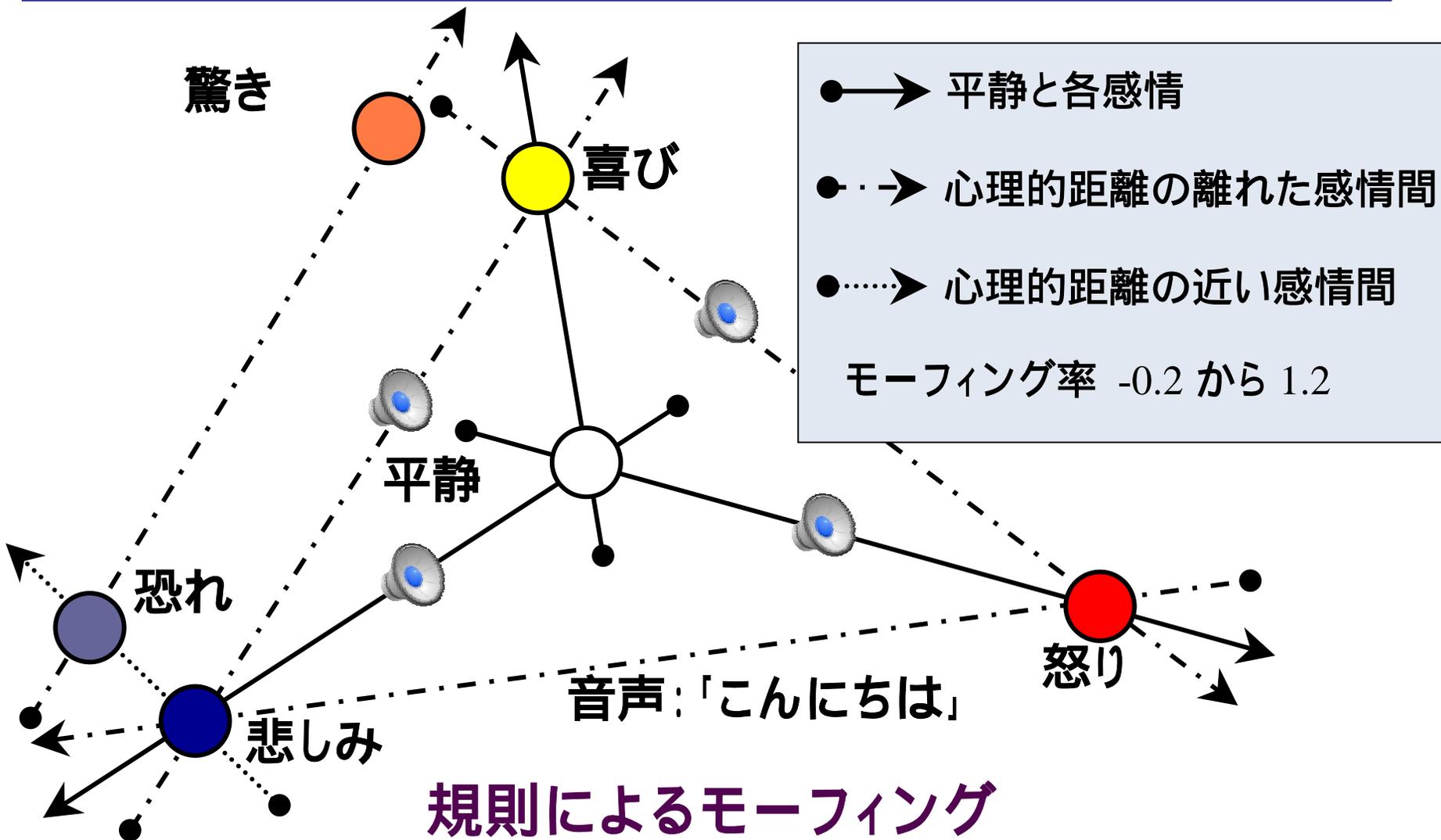


サブリーダー 河原 和歌山大教授
名古屋工大、奈良先端大、松下電器

- 「多様な**話者性**を持った音声合成や**感情音声**を表現できる音声合成、ユーザが好みの声に変換できる音声合成プログラムの作成」
- 最終目標：
 - オフライン版のプログラムでは、原音声とほぼ変わらない品質 (**主観評価値 MOS値で4以上**) を実現
 - 実時間のオンライン版のプログラムでは、**主観評価値 MOS値で3以上** を実現

話者性変換プログラムのオンライン版を開発 (H18.3)

目標感情間のモーフィング



フリーソフトウェアサイト

ダウンロード数

Julius	http://julius.sourceforge.jp	大語彙連続音声認識プログラム	7000
HTS	http://hts.ics.nitech.ac.jp	HMMベース音声規則合成	700
Festival	http://www.cstr.ed.ac.uk/projects/festival/	世界標準の音声合成フレームワーク HTS,音声モーフィングが採用	-

製品化によるコストダウン

Julius	SH-4A	Juliusのマイコンへの実装
SSA	DSP, FPGA	ハンズフリー音声収録(SSA)の実装

標準インタフェース: SAPI, SALTの実装

H16

報道発表	展示会	共同/受託研究
11	13	22

講習会	H15	H16
参加者、日数	70名、4日間	88名、5日間

博士の取得者: 11名(社会人4名)、学生をRAとして雇用、インターシップなど

学術活動(研究論文、受賞)

	学術論文	国際会議	報道発表	受賞
H15	18	66	5	7
H16	30	81	11	6
合計	48	147	16	13

主な受賞リスト

H15.6	京大、人工知能学会、大会優秀賞、研究会優秀賞、対話システム
H16.3	奈良先端大、情処、大会奨励賞、「たけまるくん」
H16.3	奈良先端大、電気通信普及財団、技術賞、技術学生賞、BSS
H16.3,H16.9	鹿野、情処、電子情報通信学会、各フェロー
H16.10	京大、情処、山下記念研究賞、バス運行音声案内システム
H17.2,H17.3	奈良先端大、C&C若手優秀論文賞、船井情報科学奨励賞、BSS
H17.5	奈良先端大、電子情報通信学会 論文賞、猪瀬賞、NAM

進捗状況

- (1) 順調に進捗、当初計画以上の成果が達成できる。マイコンによる実時間大語彙連続音声認識など
- (2) 新しい手法の発見、研究：「つぶやき声」による静かな音声メディア(NAM)、ハンズフリーにおける音源分離(BSS)。
- (3) 多くの論文、国際会議発表。学会の表彰(論文賞、猪瀬賞など)
- (4) 予算：企業への発注、開発、人件費。研究には他の予算も活用。企業の実力によるマイコン、DSPへの実装。
- (5) フリーソフトウェアサイトの運用、講習会、展示会などで普及活動。共同研究などによる産業界への技術移転活動も活発化。