総合報告

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の総合開発

E-Society Software Development Project for Speech Recognition and Synthesis

鹿野清宏　武田一哉　河原達也　河原英紀　猿渡　洋
徳田恵一　李　晃伸　川波弘道　西村竜一　Randy GOMEZ
戸田智基　西浦敬信　高橋　徹　坂野秀樹　全　炳河

abstract

ITの急速な普及に伴い，だれもが容易に情報にアクセスできる手段として，人と機械との自然な音声対話の実現が望まれている．2003年度から5年間，文部科学省のリーディングプロジェクトとして「e-Society基盤ソフトウェアの総合開発」が実施され，その一環として音声認識・合成技術を本格的に普及利用させるための頑健かつ高精度な音声認識・合成基盤ソフトウェアを開発する本プロジェクト研究が実施された．本稿では，大語彙連続音声認識プログラム，話者環境適応プログラム，ハンズフリー音声認識・電話プログラム，更にハンズフリー音声収録DSP，大語彙連続音声認識プログラムのマイコンへの実装，多様な音声を合成可能な音声合成プログラムの開発など，プロジェクトで得られた成果を概説する．

キーワード：音声認識，話者適応，対話システム，マイクロホンアレー，非可聴つぶやき

1. まえがき

ネットワーク技術・社会の急速な進展に伴い，様々な情報やサービスに多様な手段でアクセスする機会が増えている．このような社会背景の中，「世界最高水準の高度情報通信システム形成のための鍵となるソフトウェア開発を実現させ，いつでもどこでもだれでも安心して参加できるIT社会を構築する」ことが我が国の産官学にとって極めて重要な目標である．文部科学省リーディングプロジェクト「e-Society基盤ソフトウェアの総合開発」（表1）は，この目標を達成するために「高い生産性を持つ高信頼ソフトウェア作成技術の開発」と「情報の高信頼蓄積・検索技術等の開発」を進めるべく，2003年度より5年間，18の大学研究グループを中心に実施された．本プロジェクトは「情報の高信頼蓄積・検索技術等の開発」の三課題の一つとして，情報インタフェースの観点から進められたものである．

いつでもどこでもだれでも安心して参加できるIT社会実現のためには，より多くの利用者が携帯電話や携帯端末など，多様な情報端末を利用できるようなインタフェースの実現が重要である．人と機械との自然なコミュニケーション手段の一つとして音声対話，すなわち音声の認識や合成技術が本格的に利用されるには，頑健かつ高精度な音声認識・合成基盤ソフトウェアを開発し，かつ廉価に利用できるようにすることが重要である．本プロジェクトでは，音声認識が広く利用され，IT社会の基盤となるには，音声認識に関して少なくとも以下の三つの技術が必要と考えた．

① 利用環境によらずユーザに負担をかけない話者・環境適応技術

だれでも・どんなときでも最高に近いシステム性能が享受できるよう，様々な環境変動や話者の違いを瞬時に学習しこれを吸収する技術．

② マイクを意識させない自然なハンズフリー音声認識技術

鹿野清宏　正員：フェロー　奈良先端科学技術大学院大学情報科学研究科情報処理学専攻
E-mail　shikano@is.naist.jp
武田一哉　正員　名古屋大学大学院情報科学研究科メディア科学専攻
河原達也　正員　京都大学学術情報メディアセンター
河原英紀　正員　和歌山大学システム工学部デザイン情報学科
猿渡　洋　正員　奈良先端科学技術大学院大学情報科学研究科情報処理学専攻
徳田恵一　正員　名古屋工業大学大学院工学研究科創生シミュレーション工学専攻
李　晃伸　正員　名古屋工業大学大学院工学研究科創生シミュレーション工学専攻
川波弘道　正員　奈良先端科学技術大学院大学情報科学研究科情報処理学専攻
西村竜一　正員　和歌山大学システム工学部デザイン情報学科
Randy GOMEZ　京都大学学術情報メディアセンター
戸田智基　正員　奈良先端科学技術大学院大学情報科学研究科情報処理学専攻
西浦敬信　正員　立命館大学情報理工学部メディア情報学科
高橋　徹　正員　京都大学大学院情報学研究科知能情報学専攻
坂野秀樹　正員　名城大学理工学部情報工学科
全　炳河　東芝ケンブリッジ研究所
Kiyohiro SHIKANO, Fellow (Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0192, Japan).
電子情報通信学会誌　Vol.92 No.6 pp.475-491 2009年6月

表1　e-Society基盤ソフトウェアの総合開発

高い生産性を持つ高信頼ソフトウェア作成技術の開発
プロジェクトリーダー：片山　卓也（北陸先端大）
　高信頼組込みソフトウェア構築技術
　　片山　卓也（北陸先端大）
　　中島　達夫（早大）
　　湯淺　太一（京大）
　次世代高性能コンピュータシステム上の高信頼ソフトウェアシステムの開発支援技術
　　石川　裕（東大）
　プログラム自動解析に基づく高信頼ソフトウェアシステムの構築技術
　　大堀　淳（東北大）
　安全なシステム記述言語および高信頼OS
　　米澤　明憲（東大）
　データ収集に基づくソフトウェア開発支援システム
　　鳥居　宏次（奈良先端大）
　　井上　克郎（阪大）
　高信頼構造化文書変換技術
　　武市　正人（東大）
情報の高信頼蓄積・検索技術等の開発
プロジェクトリーダー：村岡　洋一（早大）
　インターネット上の知識集約を可能にするプラットフォーム構築技術
　　村岡　洋一（早大）
　先進的なストレージ技術およびWeb解析技術
　　喜連川　優（東大）
　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
　　鹿野　清宏（奈良先端大）
　　河原　達也（京大）
　　猿渡　洋（奈良先端大）
　　武田　一哉（名大）
　　河原　英紀（和歌山大）

利用者がマイクを持ったりマイクに近づいたりせずに利用できるよう，離れた位置から高い品質の音声を取得する技術．

③　高精度連続音声認識プログラムと研究開発ワークベンチ

多様な応用に利用可能なはん用性の高い音声認識アルゴリズムを開発し，これをマイコンDSP等多様なプラットホーム上で動作させる技術．

更に，これらに加えて研究開発の有効性を評価するためには④人と機械の音声対話の実証実験を行い，実際の利用環境における誤りに対応するための様々なシステムの改良が必要である．

一方音声合成では，⑤多様な声質を合成可能な音声合成・声質変換プログラムの開発が重要と判断した（図1）．そこで本プロジェクトではこれらの課題を網羅的に進めるために，研究開発項目ごとに図2に示す中間目標，最終目標を設定した．（網の中は研究の進展に伴い追加した研究目標である．）

本稿では2.で音声認識エンジンJuliusとその対話音声への拡張について，3.で教師なし話者適応と無音声処理技術について，4.でブラインド音源分離とマイクロホンアレー技術を用いる遠隔音声認識技術について，5.では高品質音声分析合成系STRAIGHTに基づく声質の変換技術について，6.では隠れマルコフモデルに基づく音声合成技術について，それぞれ開発経緯と成果を説明する．7.では得られた成果をまとめるとともに，今後の展望について述べる．　（鹿野清宏）

2.　大語彙連続音声認識プログラムJuliusの音声対話への展開

2.1　ねらい

ユーザに負担のない自然な音声対話を実現するには，

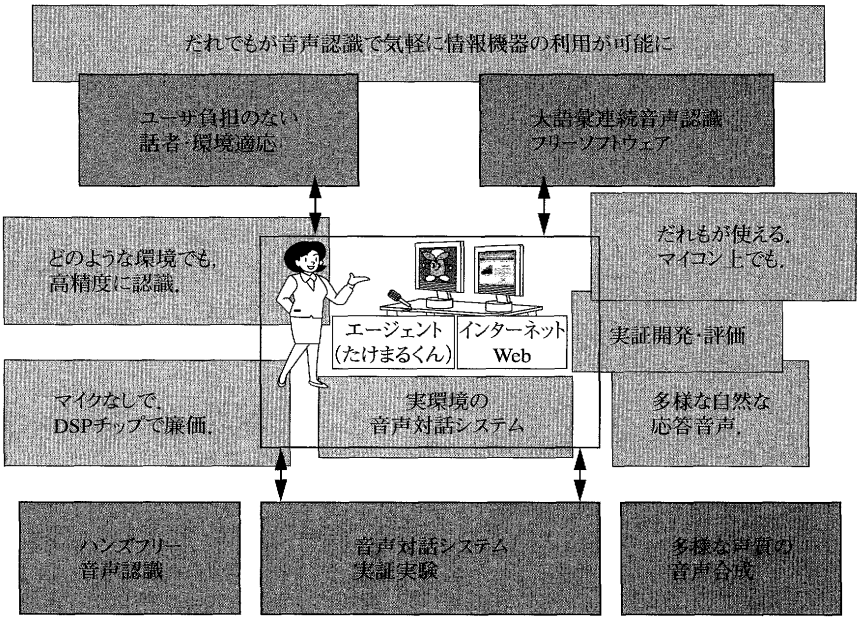


図1　社会基盤としての音声認識・合成プログラム

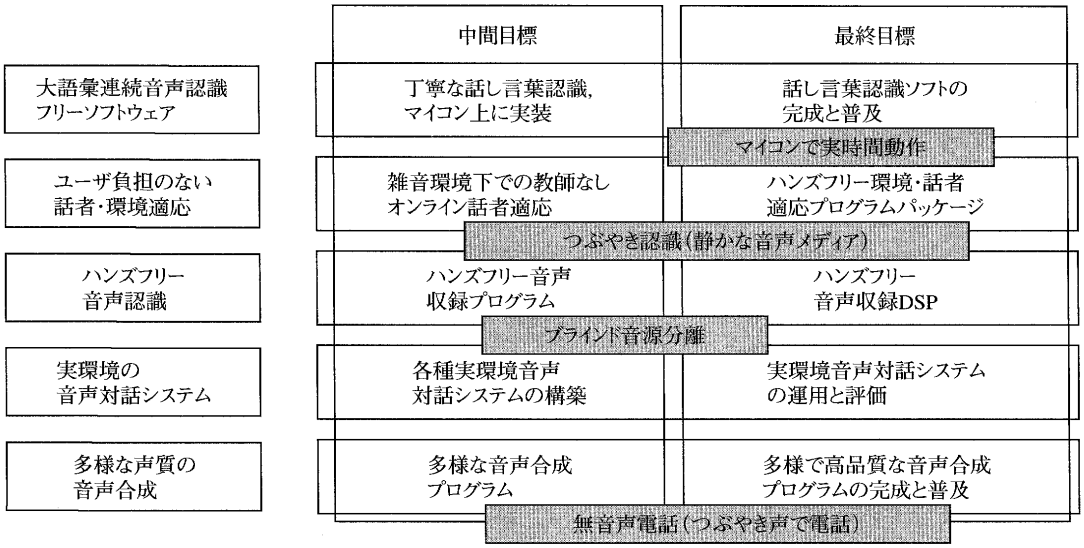


図2　プロジェクト目標

音声認識システムがアプリケーションや話者・環境に適応するとともに，話し言葉に対応できる必要がある[1]．このような大語彙連続音声認識を行うプログラムの開発を行った．

本プロジェクトのメンバーは，これまでにIPAのプロジェクト（1997-1999年度；http://www.ar.media.kyoto-u.ac.jp/dictation/）で「日本語ディクテーションツールキット」[2]を開発し，その後，情報処理学会連続音声認識コンソーシアム（2000-2003年度；http://www.lang.astem.or.jp/CSRC）でその維持と普及を行ってきた．それまでのシステムは主にディクテーションを指向して読み上げ音声を対象としていたので，本プロジェクトでは音声対話を指向して発展させた．また，組込み機器にも利用できるようにマイコンへの実装も行った．

2.2　アプローチ

図3に本研究開発の概要を示す．左下がベースラインとなる音声認識システムで，音声認識エンジンJuliusと音韻モデル及び言語モデルから構成される．音韻モデル[用語]・言語モデル[用語]を，音声対話を想定して，話し言葉の発話・表現に対応できるようにした．音声認識エンジンJuliusについても，対話システム向けの機能強化を行った．次に，多様なアプリケーションへの適用を容易にするために，音声認識システムを容易にカスタマイズできるような枠組みを研究開発した．特に言語モデルは，アプリケーションごとに語彙や言い回しが変わるため，個別に用意する必要がある．本プロジェクトではWeb上のテキストを用いることで，多様な言語モデルを効率的に構築する方法を実現した．音韻モデルについても，正規化や適応を指向したものを開発した．

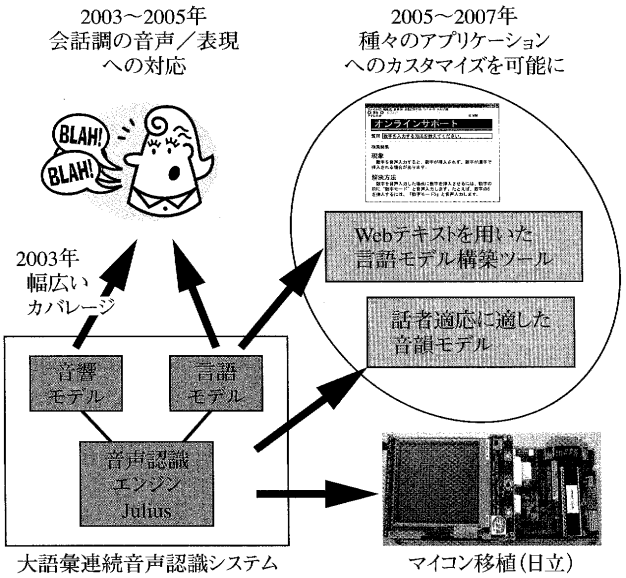


図3　大語彙連続音声認識システムの開発経過

用語解説

音韻モデル　音声信号の特徴を，5ms程度ごとにスペクトル分析を行うことで得られる短時間特徴の時系列により表現し，この時系列の生起確率を，音韻ごとに，複数状態からなる確率モデルである隠れマルコフモデル（HMM）を用いてモデル化したもの．音韻モデルを連結することで，任意の単語のHMMを構成することができる．

言語モデル　単語連鎖の出現頻度（確率）$P(w_i|w_{i-2}, w_{i-1})$のように，先行2単語が与えられたもとで，ある単語が出現する確率（トライグラム確率）などから，任意の単語系列が出現する確率を以下のように近似的に連鎖的に計算することができる．

$$P(w_0, w_1, \cdots, w_i) = P(w_i|w_{i-2}, w_{i-1})P(w_{i-1}|w_{i-3}, w_{i-2})\cdots P(w_2|w_0, w_1)P(w_1|w_0)P(w_0)$$

表2　音響モデルの改良結果　『日本語話し言葉コーパス』に適用した場合の単語認識精度の改善.

ベースライン（ケプストラム平均正規化）	76.0%
＋話者正規化（CVN＋VTLN）	77.0%
＋識別学習（MPE）	79.4%
＋話者適応（MLLR）	81.3%
＋話者適応学習（SAT）	81.8%

2.3　話し言葉向け音韻モデル

多様な話者の話し言葉音声を高精度に認識するためには，音響特徴量を話者ごとに正規化したり，音韻モデルを話者ごとに適応させる必要がある．本研究開発では，話者正規化手法としてCVN（ケプストラム分散正規化）とVTLN（声道長正規化），話者適応手法としてMLLR（最ゆう線形回帰），更にMLLRにより話者正規化を行うSAT（話者適応学習）の評価を行った．更に高精度な音韻モデルを実現するために，有力な識別学習法の一つであるMPE（音韻誤り最小化学習）も実装した．標準的な音韻モデルの構成である3,000状態16混合からなる状態共有トライフォンHMM（隠れマルコフモデル）(注1)にのっとり，『日本語話し言葉コーパス（CSJ）』の講演音声で学習・評価を行った結果（単語認識精度）を表2に示す．ベースラインに比べて5%以上の改善を実現した．国会審議の音声に対して適用した場合にも，同様の結果が得られることを確認している．

2.4　Webテキストを活用した話し言葉向け言語モデルの構築

音声対話を利用したインタフェースを構築するには，レストラン検索や観光案内などの当該アプリケーションに特化した言語モデルが不可欠であるが，人手で記述した有限状態文法は受理できる発話パターンが限定され，負担が大きい．一方，頑健性・柔軟性の点で優れている統計的言語モデル（単語n-gram）を構築するには，当該ドメインの学習用テキストを大量に用意する必要がある．この言語モデル構築を効率的にするために，Web上に存在するテキストの利用を考えた．ただし，音声認識のための言語モデル学習には，単に検索エンジン等で関連テキストを収集するだけでなく，話し言葉スタイルの文を選択的に用いる必要がある．そこで，既存の（当該ドメインとは異なる）音声対話コーパスと検索対象の文書テキストを混合してできる言語モデルによるゆう度を文選択の指標として用いる．この枠組みを図4に示す(3)．これら一連の処理を行うツールキットを作成した．

(注1) 3音素連鎖ごとに中心音素をモデル化する場合，40（音素の概数）の三乗のHMMが必要となる．各HMMが3状態から構成される場合，総計1万5,000以上のHMM状態を学習する必要が生じる．そこで，この状態のバリエーションをクラスタリングすることで3,000状態に集約し，それぞれの状態ごとに16個のガウス分布の線形結合でスペクトルの出力確率を表現する．

表3　対話システムに対して発せられた発声の認識性能比較（単語認識精度）

対象ドメイン	ソフトサポート	観光情報案内
対象文書＋対話コーパス	71.5%	71.6%
対象文書＋対話コーパス＋Web	77.2%	77.4%

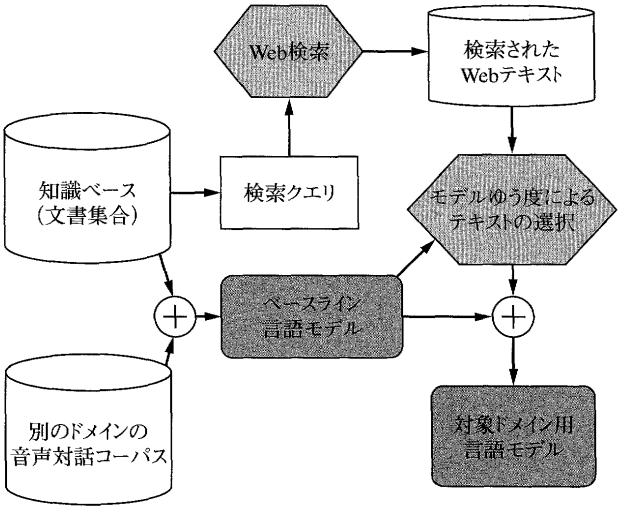


図4　Webを活用した言語モデルの構築

ソフトウェアサポートと観光情報案内の二つのドメインに対して，それぞれのアプリケーションに関する発話を一切収集することなく，高い性能を有する言語モデルを構築することができた．評価結果（単語認識精度）を表3に示す．適切な選択基準に基づいて学習用の文を選択することが効果的であった．

2.5　音声認識エンジンJuliusの改善

本プロジェクトのメンバーは，これまで音声認識に関する研究開発プラットホームとして，オープンソースの大語彙連続音声認識エンジンJulius（http://julius.sourceforge.jp/）(4)の開発を行ってきた．これは，国内の研究機関では事実上のベースラインとして広く使われており，国外にも広がりを見せている．本プロジェクトでは，音声対話システムを指向して，このソフトウェアの機能強化及び性能改善を継続的に行った．主な機能強化は以下のとおりである．

- 非音声区間の棄却
- 音声区間の頑健な検出
- 複数の音響・言語モデルの併用・切換
- 単語グラフの出力
- 認識結果の信頼度の付与

更に，大幅な高速化や省メモリ化も実現し，標準的なノートPCで大語彙連続音声認識（ディクテーション）

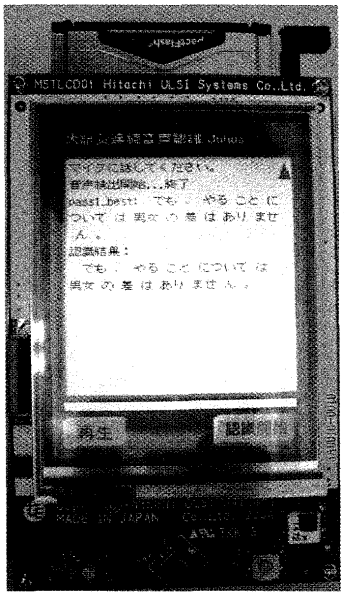


図5　SH-4A マイコン上の音声認識プログラムの実装

の実時間動作を可能にし，この集大成として，2007 年 12 月に約 8 年ぶりのメジャーバージョンアップとなる Julius-4.0 をリリースした．(現在の最新バージョンは Julius-4.1)

2.6　マイコンへの実装

カーナビや家電製品などの様々な機器への組込みを容易にするために，日立製作所の協力を得て，音声認識エンジン Julius の同社のマイコンへの移植を行った．SH-4 に続いて，新しいマイコンプラットホーム SH-4A (CPU 400MHz，メモリ 128MByte) への移植を進めた (図 5)．その結果，表 4 に示すような性能が得られ，2 万語彙の連続音声認識がマイコン上で実時間動作可能であることを示した．また，サンプルアプリケーションも作成した．

2.7　音声認識システムのフィールドテスト

実用的な音声対話システムを構築し，幅広い一般のユーザを対象として，フィールドテストを行った．まず，電話により京都市バスの運行情報を案内するシステムを約 4 年間にわたり公開・運用した．その結果，約 1 万対話 (コール)，約 6 万発話を収集することができた．この結果，長期間にわたるユーザの振舞いを調べることもできた．次に，エージェント／ロボットにより京都の観光情報を案内するシステム『京都版ダイアログナビ』を構築し，京都大学博物館の企画展示において 3 か月にわたり運用した．このシステムの対話例を図 6 に示す．延べ約 2,500 人のユーザの利用があり，約 2 万 5,000 発話を収集できた．このデータを用いて対話戦略を最適化する学習や，発話のタイミングに着目して対話の齟齬を検出する研究が可能になった．更に，インターネットを介して実施した PC 上での音楽検索システムの利用実験では，

表 4　マイコンでの音声認識性能 (新聞記事読み上げ文による認識精度と実時間比)

対象語彙サイズ	5,000 語	2 万語
SH-4 (240MHz，64MByte)	89.7% 1.48×RT	NA
SH-4A (400MHz，128MByte)	89.7% 0.73×RT	90.9% 1.15×RT

利 用 者：銀閣寺について教えて下さい(検索)
システム：銀閣寺は京都市左京区にある東山文化を代表する臨済宗相国寺派の寺院です．通称は銀閣寺，山号は東山です．銀閣寺の歴史や見所について何か知りたいことはありますか?(要約提示)
利 用 者：いつ建てられましたか?(質問)
システム：銀閣寺の建立は1482年です．(応答)
システム：では，銀閣寺はだれが建てたか知っていますか?(システムからの情報推薦)
利 用 者：知りません．だれでしょうか?
システム：銀閣寺の創立者は室町時代８代将軍の足利義政，開山は夢窓疎石とされています．

図6　『京都版ダイアログナビ』の対話例

マイクロホンや環境雑音などユーザ利用環境の多様性や，利用満足度などの個人の性向に関する研究が可能となった．

(河原達也，李　晃伸，武田一哉，鹿野清宏，川波弘道，西村竜一)

3.　教師なし話者適応と無音声認識／無音声電話

3.1　ねらい

IT 技術の普及に伴い，だれもが容易に情報機器を利用できるようになることが望まれている．そのための入力手段として音声認識技術が有効である．音声認識技術をより広範な応用で利用するためには，話者や利用環境に瞬時に (利用者に負担をかけることなく) 適応することで，いつでも・どこでも頑健に動作させる技術が望まれる．この適応を実現するために，教師なし話者・環境適応技術に関して研究開発を行った．話者への適応技術により，広範な話者・環境に対して高精度な音声認識を提供する基盤技術が確立された．

3.2　アプローチ

教師なし話者適応は，発話者に負担をかけないで，音声認識の性能を向上させ得る手法である．特に，認識性能の低い話者に対して効果がある．そこで任意の 1 文発話に基づくユーザに負担をかけないオンライン教師なし話者適応アルゴリズムの雑音環境下での研究開発を進めた．

具体的には，雑音に頑健な音声認識アルゴリズムと HMM (隠れマルコフモデル) 十分統計量に基づく教師なし話者適応アルゴリズムを開発した．この雑音環境下

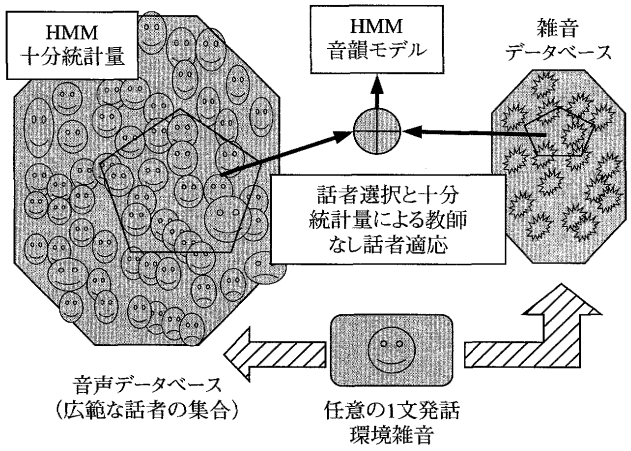


図7　雑音環境下での教師なし話者適応

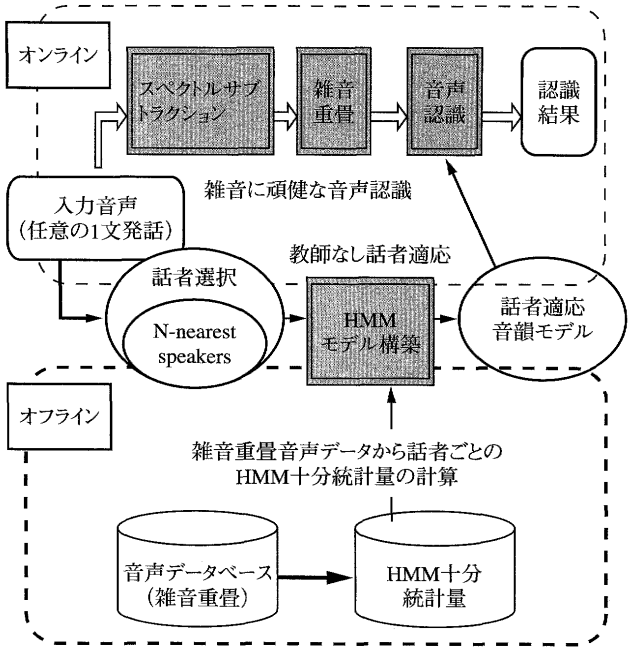


図8　雑音環境下での HMM 十分統計量に基づく話者適応

での教師なし話者適応アルゴリズムの概念を図7に，フローチャートを図8に示す．このアルゴリズムは，雑音重畳音声データベースからあらかじめ HMM 十分統計量を計算するオフラインパートと，任意の1発話に基づいて類似話者の HMM 十分統計量を合成するオンラインパートからなる．最終年度までには，従来の教師あり適応アルゴリズムの 10 文発話の性能を上回り，かつ，オンラインで動作する話者・環境適応プログラムを完成させる．

この雑音環境下での教師なし話者適応技術をオンライン化して，パッケージとして提供する．3年度後半から，この適応技術を実環境の音声対話システムに適用して，性能評価及び改良を行い，オンラインで動作する完成度の高い環境・話者適応プログラムをパッケージとして提供するほか，新しい音声メディアとして，外部の人に聞こえないつぶやき声の音声認識（無音声認識）と電話による伝送（無音声電話）の研究を行う．

3.3　研究成果の概要

雑音環境下での HMM 十分統計量に基づく教師なし話者適応を約5秒で実行できるシステムを PC 上に実装した．5年間にわたり音声情報案内システム「たけまるくん」を公共施設において運用し，多くの子供や幼児を含む広範な話者からの音声データを収集した．収集した音声データのうち，2年間分の 30 万発話の書き起こしを

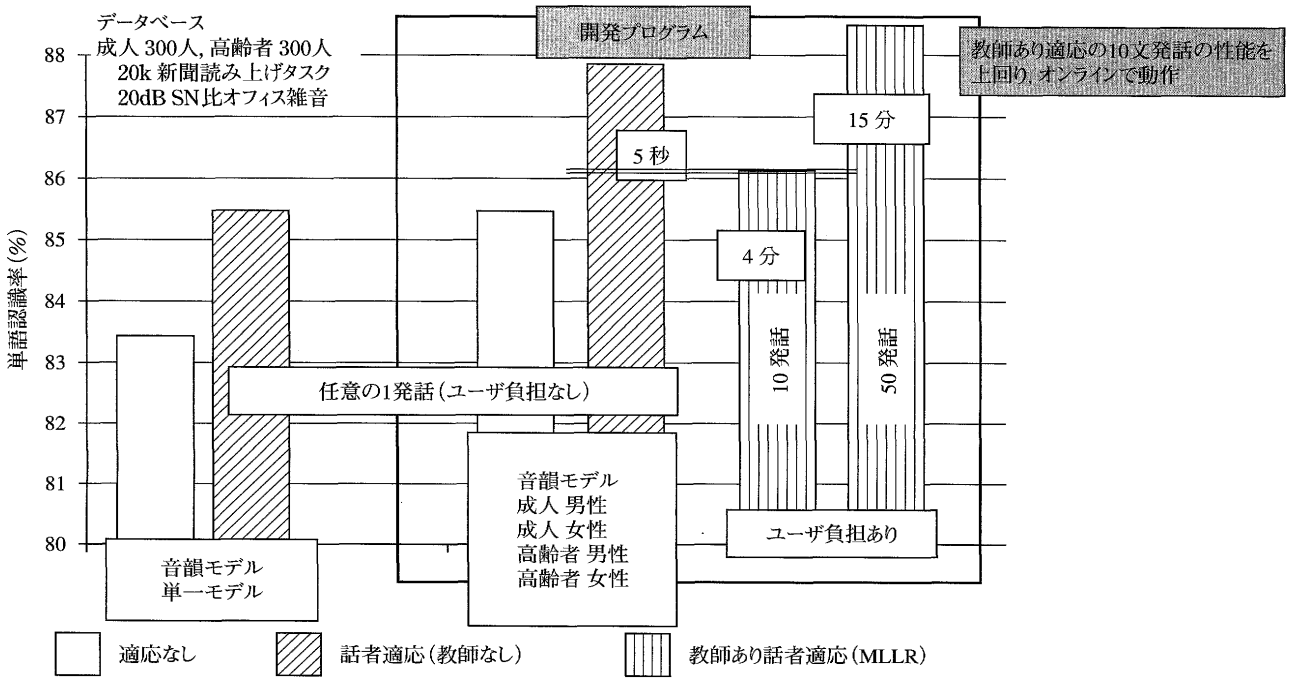


図9　教師なしオンライン話者適応システムの性能

達成し，高性能な音韻モデルの構築に成功した．また，駅構内にロボット型の対話システム「キタロボ」とエージェント型対話システムの「キタちゃん」を設置して，乗換え案内などの情報案内を行うサービスを2年間運用した．更に，既存の音声データベースからタスクに適した音韻モデルを構築するアルゴリズムを考案した．

外部に聞こえない非可聴つぶやき（NAM；Non-Audible Murmur）による音声認識（無音声認識）と電話（無音声電話）の研究も進めた．

以下に，研究成果をまとめる．

3.4 教師なし高速話者適応アルゴリズム[5]

話者適応アルゴリズムの精度の向上，計算量の削減を達成し，5秒で教師なし話者適応が実行できる話者適応システムを構築できた．新聞記事読み上げタスクの評価で，単語認識率が83.5%から87.9%に向上し，50文発話の教師あり学習に匹敵する性能を達成した（図9）．更に，オンライン教師なし話者適応プログラムを，不特定話者モデルのデコーダと話者適応モデルのデコーダを並列に実装することによって，オンライン動作を実現した．

3.5 音声情報案内システムの運用[6]

音声情報案内システム「たけまるくん」（図10）の運用を継続するとともに，2年間分の30万発話の書き起こしと応答文の付与を終了し，各種音韻モデルの精度の大幅な向上を達成した．「たけまるくん」音声データベースの配布を開始した．「たけまるくん」は，2004年4月の改善により，利用者が2倍以上に増え，1日平均で約600入力となっている．利用状況を図11に，月ごとの入力数で表しておく．コミュニティセンターに設置している音声情報案内システム「たけまるくん」に加えて，より厳しい騒音レベルの高い実環境での実証実験を行うために，近鉄「学研北生駒駅」に音声情報案内システム「キタロボ」とロボット案内システム「キタちゃん」（図12）を設置して，2年間の運用を行った．この「キタロボ」と「キタちゃん」で，「たけまるくん」からの音声情報案内システムのポータビリティの研究を行った．「たけまるくん」システムパッケージも無料公開して，10機関以上に移植されている．

3.6 音韻モデルの構築のコスト削減手法

音韻モデルの構築のコストを減らすことを目指して，

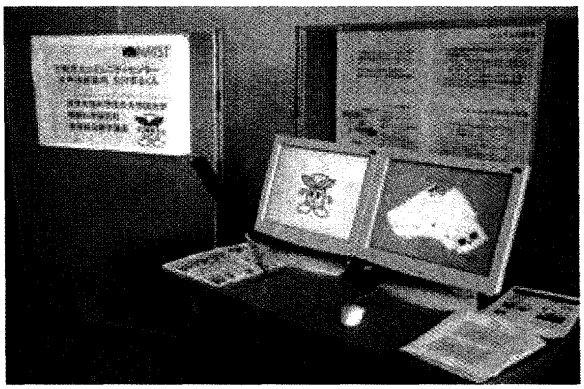


図10 生駒市コミュニティセンターに常設されている音声情報案内システム「たけまるくん」

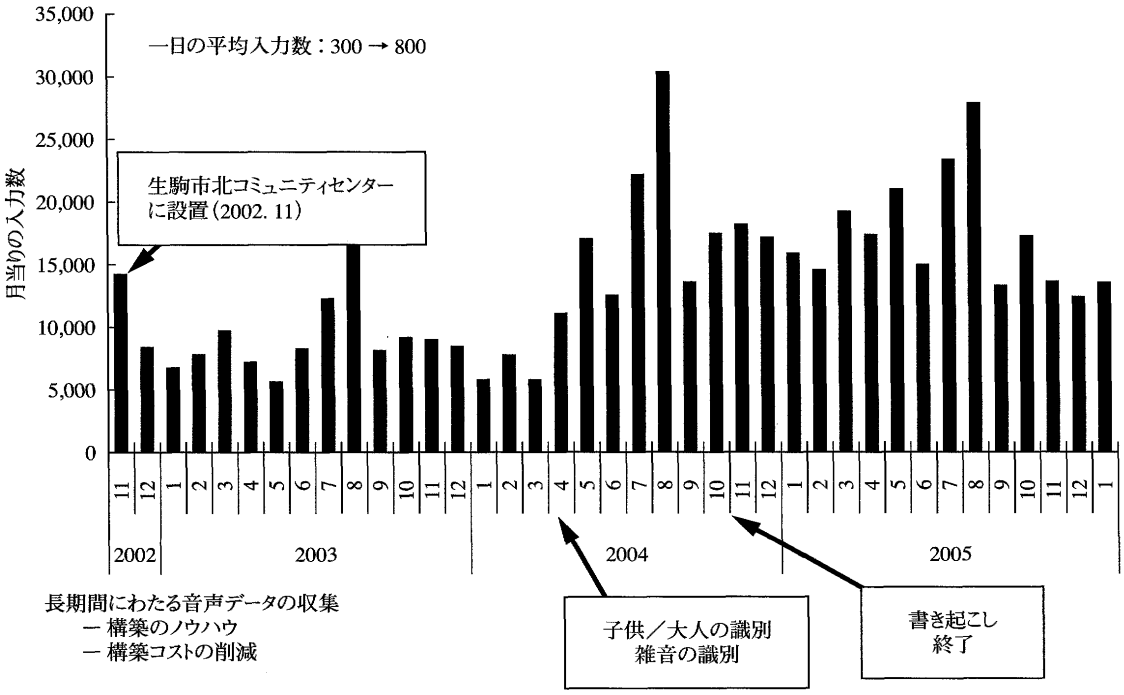
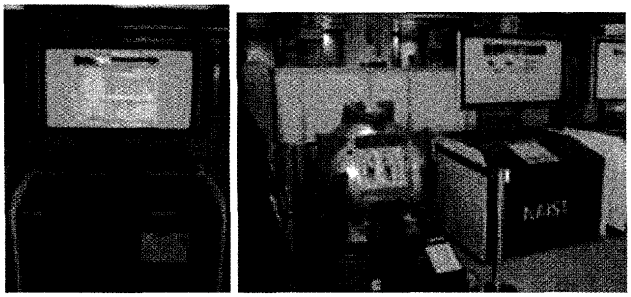


図11 「たけまるくん」の月別利用数遷移

膨大な既存の音声データベースからタスクに適した音韻モデルを自動構築するアルゴリズム（図13）を考案した．



(a)「キタちゃん」　(b)「キタロボ」

図12　駅に設置した音声情報案内システム「キタちゃん」と「キタロボ」

高齢者及び幼児の音韻モデルの構築を行い，有効性を実証した．このアルゴリズムは，既存の膨大な音声データベースから，設定タスクの音声のゆう度の向上を評価関数として，学習音声データを選択できる手法である[7]．「たけまるくん」での実環境での幼児（図14）や高齢者の音声認識に適用して，有効性を確認した．

3.7　非可聴つぶやき（NAM）[8]

非可聴つぶやき（NAM：Non-Audible Murmur）は，話し手の近くでも聞こえない非常に小さな声である．本プロジェクトのメンバーは，これまでに，身体とのインピーダンス整合を図ることができるシリコンなどの材料を介して音響トランスデューサを首周辺の適切な位置に接触させることで，NAM信号を検出可能であることを

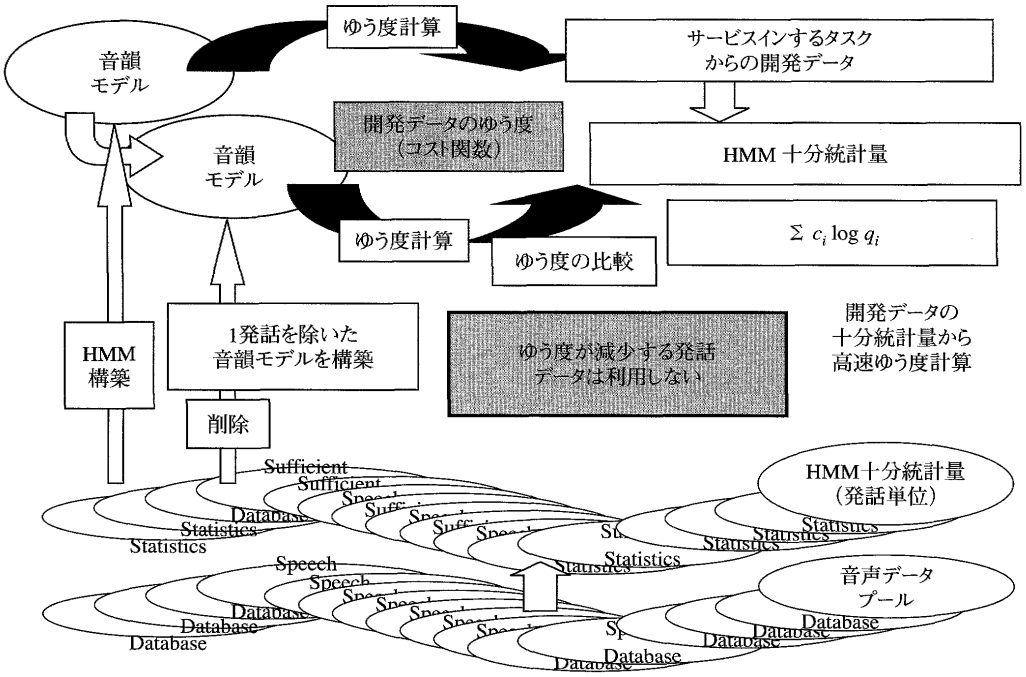


図13　音声データプールからの選択学習

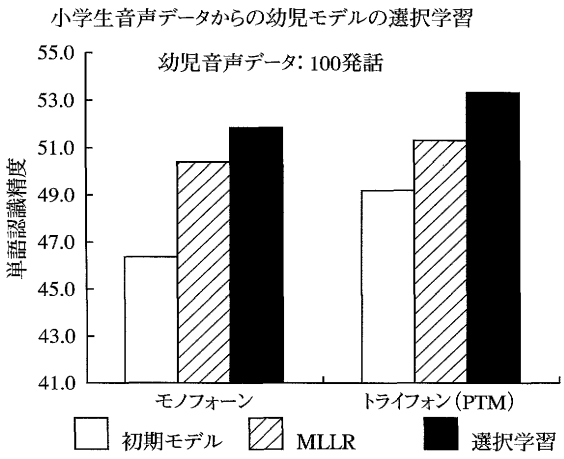


図14　選択学習の効果

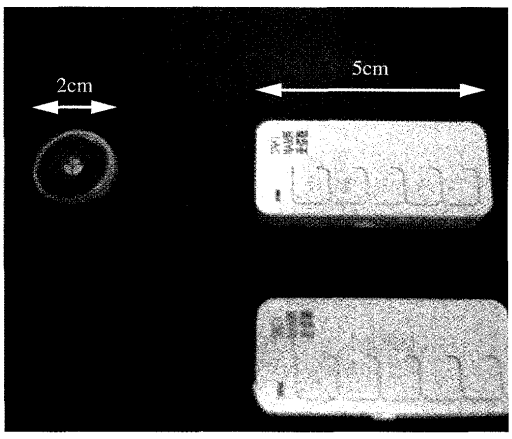


図15　Bluetooth 型 NAM マイクロホン

明らかにした．このNAM検出デバイス（NAMマイク）を利用することで，様々なNAM信号を計測記録し，NAMによる音声認識（無音声認識）[8]や，声を出さない電話（無音声電話）[9]の研究を進めた．NAMマイクにより計測されるNAMは，原理上環境雑『音』の影響をほとんど受けないため，環境雑音に耐性の高い音声入力装置に応用することも可能と考えられるほか，咽頭の機能障害の補償など様々な応用が考えられる．

本プロジェクトでは，NAM音声による音声認識（無音声認識）に応用するために，Bluetoothインタフェースを持つNAMマイク（図15）を開発した．更に，統計的音声モーフィングの研究とNAMマイクロホンとを活用して，発話障害者音声からささやき声への変換（図16）を試み，良好な結果が得られた[10]．様々な発話に対応できるNAMの認識プログラムの開発を行った．そのほか，NAMが他人に聞こえないとの特質を利用したNAMのキーワード発話での個人認証の研究を行い，良好な結果が得られた[11]．

（鹿野清宏，Randy GOMEZ，戸田智基）

4. ハンズフリー音声認識

4.1 ねらい

人と情報機器間においてシームレスなコミュニケーションを確立するためには，「自然な（普段人間同士が行うような）」マン・マシンインタフェースを構築する必要がある．本項目では，上記を達成するべく，ユーザに負担をかけない自然な音声入力系として，遠隔発話を高精度に収音・認識するハンズフリー音声認識システムを開発した．特に，複数マイクロホンを並べて多点受音を行う「マイクロホンアレー」による音声収録技術に着目し，音声認識性能の向上を目指した．また，コンパクトかつ廉価なマイクロホンアレーアルゴリズムの開発を行った．

4.2 アプローチ

一般に，従来のマイクロホンアレーを音声認識の前処理に用いる場合，以下の問題点があった．

① 高い雑音抑圧性能を得るためには，非常に多くのマイクロホン素子を広く配置する必要があった．
② 適応フィルタを導入した適応マイクロホンアレーの場合，1チャネル当り数百個のフィルタ係数を逐次的に学習・計算する必要がある．よって，実時間処理には適していないという問題があった．
③ 従来マイクロホンアレーの出力は音響信号波形であり，音声認識処理系にとっては冗長かつ非最適であった．例えば，音声認識処理系では，通常，波形そのものではなくそのスペクトル包絡情報のみを入力し，位相に関する情報は無視される．

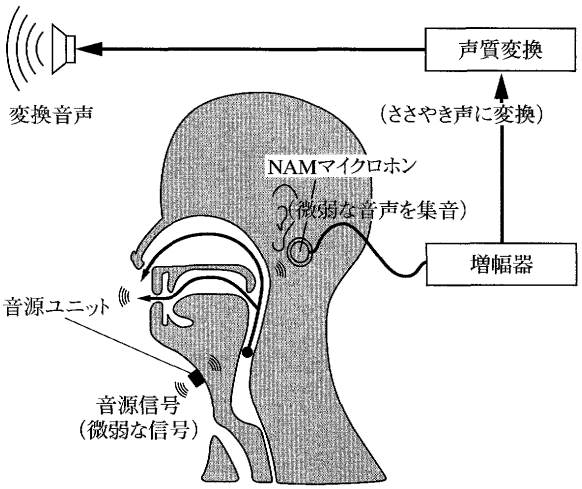


図16 NAMマイクとささやき声変換による発話障害者の発話補助

上記問題点を解決するため，本プロジェクトでは，音声認識に適したマイクロホンアレー信号処理の研究開発を行った．具体的には，音声認識入力段にて必要とされるスペクトル包絡情報を直接出力するアレー信号処理を確立し，これにより演算量を削減し実時間処理を目指した．本研究では，上記アレー信号処理を「空間スペクトル減算アレー（SSA：Spatial Subtraction Array）[12]」と呼ぶ（図17）．本プロジェクトでは，SSAの開発とともに，実環境における様々な雑音のデータベースの収録・整備や，音声認識器側におけるSSAに適した音響モデルの適応に関しても検討を行った．また，SSAよりも負荷は多いが更に自由度の大きなアレー信号処理である「ブラインド音源分離（BSS：Blind Source Separation）」に関しても検討を行い，実環境評価や実時間動作可能なアルゴリズムの開発を進めた．同時に，SSAとBSSを融合したアルゴリズムに関しても提案を行った．

以上述べた処理を検討し，ユーザからの距離1m以下で高性能に動作するハンズフリー音声認識システムを，8チャネル以下のマイクロホンアレーを用いて開発した．認識性能は，1m離れた音声入力で，従来の接話マイクとほぼ同等の認識性能を目指す．更に，マイクロホンアレーのコストを20分の1以下にするため，ハンズフリー音声収録用DSPを開発した．

4.3 空間スペクトル演算アレーSSA[12]

効率的かつ高精度な雑音抑圧アルゴリズムとして，空間スペクトル演算アレーSSA（図17）を新たに提案し，その実環境評価及び改良を行った．SSAは，認識対象とする音声に2種類（主パス，副パス）の指向特性を形成し，空間的に異なる二つの信号を受音する．主パスでは，広い指向性を持つ遅延和方式により，音声及び雑音

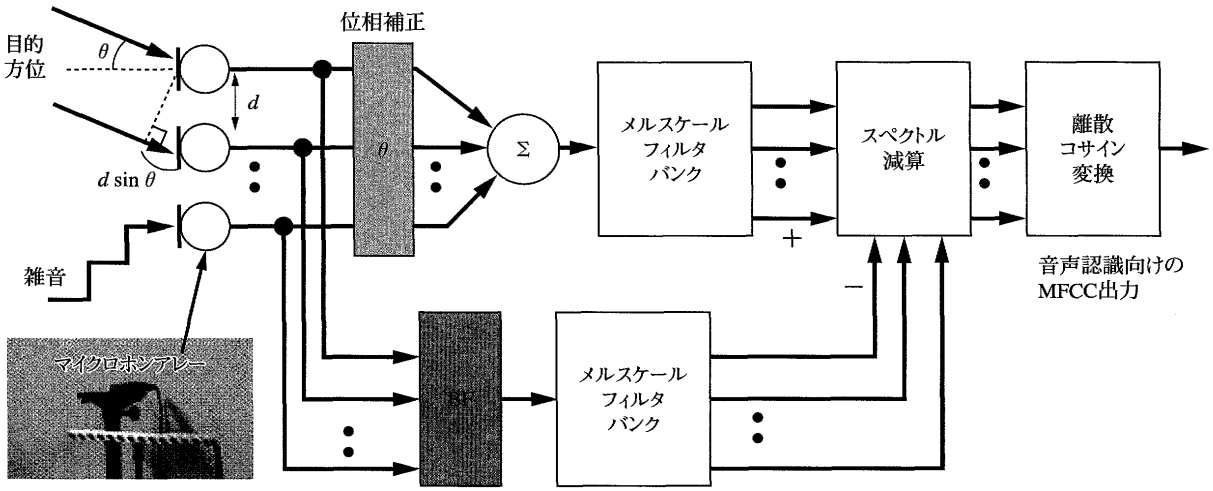


図17　空間スペクトル減算アレーの全体構成

を取得する．副パスでは，急しゅんな死角指向特性を用いて音声のみをキャンセルし，推定雑音信号を得る．それぞれの指向特性を介して受音された信号はフィルタバンク分析によりパワースペクトルに変換され，パワースペクトル領域で推定音声信号から推定雑音信号の減算を行うことで雑音抑圧が実現される．本手法の特徴は，音声信号の音韻的特徴が振幅スペクトルの包絡に表れることに着目し，従来高品質な音声信号の回復には必須とされた信号の位相成分の回復処理を考慮しないことで，高速かつ環境変動に頑健な処理を実現していることである．

SSA を用いた雑音抑圧の性能は，SSA に適した音響モデルを学習し音声認識システムで利用することで更に向上する．例えば，模擬室内音響特性や環境雑音を付加した音声に SSA 処理を適用し隠れマルコフモデルを学習することで，SSA に適した音響モデルが学習できることが示された．また，後述する BSS 技術を SSA の雑音推定処理部に導入することにより，未知の素子誤差や室内残響により頑健な SSA アルゴリズム（Blind SSA：BSSA）[13]も開発した．

4.4　発話検出アルゴリズム

音声認識の対象となる区間を断続的に入力される音声信号から正しく検出すること（発話検出）は，音声認識において重要な問題である．本プロジェクトでは，発話者の位置（空間的発話検出）と発話区間の同定（時間的発話検出）双方に着目して開発を行った．まず，発話者位置検出法では，あらかじめ発話者位置ごとに学習された音声の長時間平均スペクトルの分布を用いて，最ゆうの発話者位置系列を高精度に推定する．更に発話区間検出技術を導入し，時間・空間情報を併用する新しい発話検出法を SSA における指向形成と一体化させることに成功した（図18）[14]．時空間情報を積極的に融合することで，周囲雑音により耐性の高い発話検出が実現された．

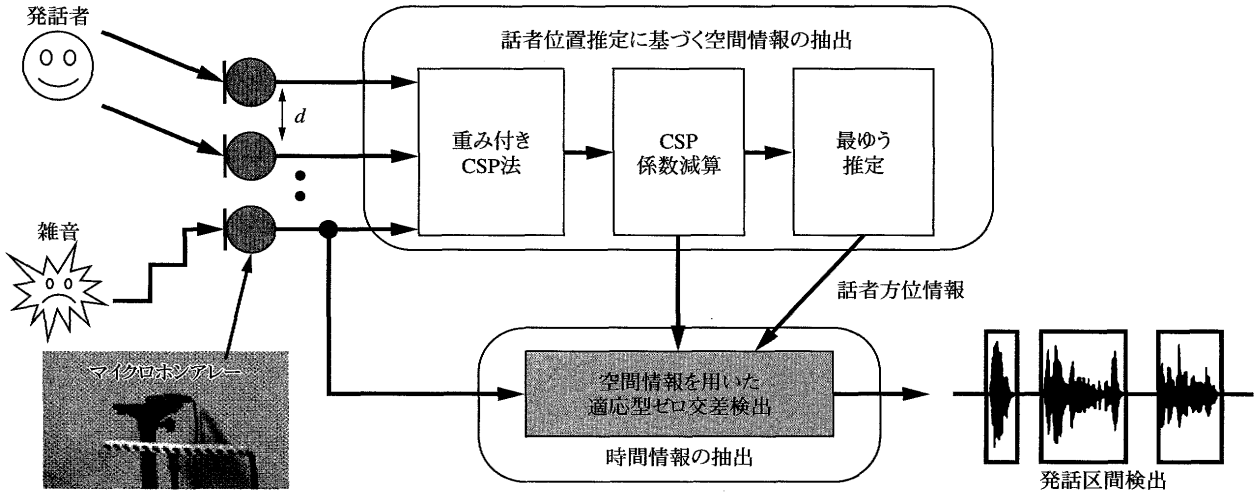


図18　発話区間検出アルゴリズムの構成図　　話者の空間的な位置の推定結果を利用して，特定方向について音声区間を検出する．

4.5 実時間 SSA 処理系の DSP 実装

開発した SSA は実時間処理に適しており，これを浮動小数点 DSP モジュール上に実装した(注2)．音声案内システム「たけまるくん」と結合することにより，ハンズフリー音声対話デモシステムを構築した（図 19）．本 DSP モジュール内部には，(a)多チャネル A-D/D-A 変換器，(b)話者方位推定に基づく実時間発話検出器，(c)実時間動作可能な SSA アルゴリズム，(d)音声認識特徴量パラメータを TCP/IP にて伝送するネットワークモジュールなどが実装されている．本モジュールを直接ネットワーク経由で音声認識装置へ接続することにより，ハンズフリー音声認識システムを容易に構成することが可能となっている．ここでは，想定される規模（4～8 素子アレー）において実時間処理が可能であることを確認できた．

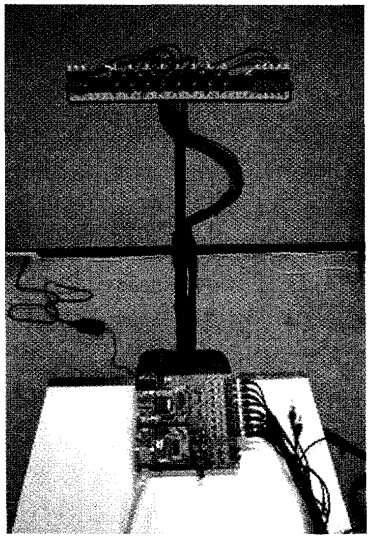


図 19　SSA を実装した実時間 DSP モジュール

4.6 ブラインド音源分離 BSS

本プロジェクトでは，音源間の独立性のみに基づいて混合信号の分離を行う BSS の研究開発を行い，これまで実験室環境にとどまっていた本技術の適用範囲を実環境下で動作するレベルへ高度化することに成功した．特に，分離尺度に独立性だけでなく空間特性保持を導入した SIMO-ICA の提案や，バイナリマスキングとの併用による高速化などが特筆すべき成果として挙げられる．これらは，図 20 に示す BSS 実時間モジュールとして市販されるに至っている．

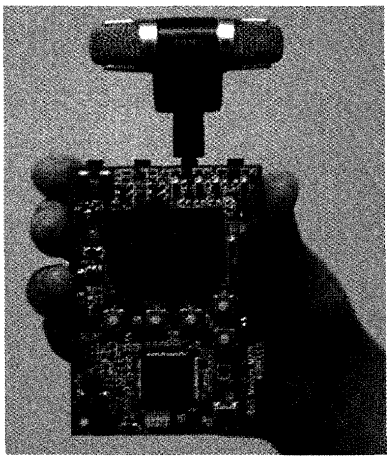


図 20　BSS 実時間 DSP モジュール

4.7 ハンズフリー音声対話システム

上述した SSA に基づく雑音抑圧手法や音声検出法を用いることにより，日常的な雑音環境下における遠隔発話（マイクロホンから 1m 程度以上離れた場所での発話）であっても，高い精度で音声を認識することが可能となった．更にハードウェアモジュール化されたこれらの技術を利用して，通常室内や駅などにロボット音声対話システムを設置し，実環境における雑音データの収録及び評価システムを構築した．実証実験を通じて，ハンズフリー環境におけるリアルタイム音声認識システムが，高精度に動作することが確認された．

（猿渡　洋，西浦敬信）

5. STRAIGHT 分析合成系に基づく多様な音声の音声合成

5.1 ねらい

音声によるコミュニケーションを文字によるコミュニケーション以上のものとしているのは，音声に含まれる個人性のような非言語情報と感情のようなパラ言語情報である．本プロジェクトでは，合成音声にこれらの情報を適切に付与するソフトウェアを開発することにより，ユーザの負担が少ない，より自然な音声対話処理技術の確立を図った．

研究にあたっては，音声処理に関する最新の工学的，科学的成果を実際のシステムに組み込むことをねらい，二つの目標となる応用形態を設定した．例えば，カーナビ等の応答音声に多様な声質や話し方の音声を作り込むような応用に必要な，非リアルタイムではあるが高い品質を実現する技術がその一つである．もう一つは，リアルタイムでの声質や話し方の変換である．ここでは，処理の軽さが重要であり，非リアルタイム応用のように高い品質は要求されない．

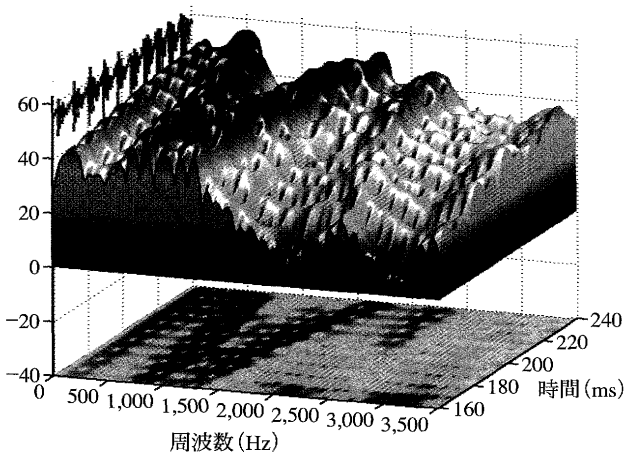
5.2 アプローチ

音声に含まれる非言語情報やパラ言語情報は，言語情報よりも品質の劣化の影響を受けやすい．これらのデリ

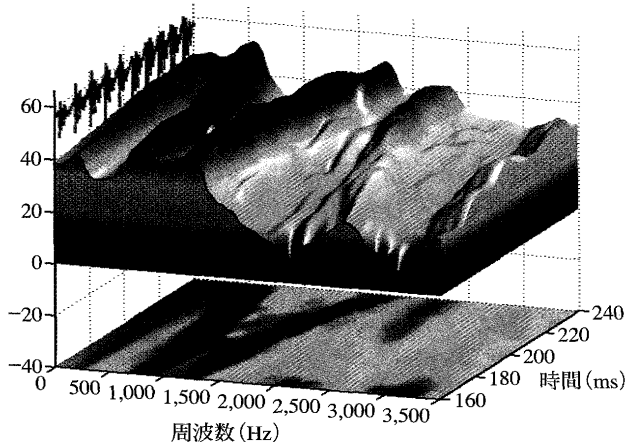
(注 2)　SSA は本来実時間処理に向いており，軽量な演算処理のみで実現できるアルゴリズムであるが，試験開発の効率性を重視しここでは浮動小数点 DSP 上に実装しその性能を確認した．

ケートな情報の加工のためのソフトウェアの開発には，研究基盤として品質劣化の少ない音声処理技術が必要である．本プロジェクトでは，我々が開発してきた高品質音声分析変換合成システム STRAIGHT (Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrum) を研究の基盤とすることで，この条件をクリアした．

STRAIGHT[16] は，有声音における周期的な励振を，時間－周波数表現からの情報の標本化機構であるととらえることで導出された方法である．短時間フーリエ変換により求められるスペクトログラムは，図 21 の上の図に示すように，音声の周期性の影響を受けて，時間方向と周波数方向の双方に周期的な変動を含む．STRAIGHT は，分析位置に依存しないパワースペクトル推定法と，周波数領域での適応的平滑化処理により，下の図に示すように，この周期的な変動だけを選択的に除去した音声のスペクトル包絡を与える．この方法を用いることにより，音声波形は，時間とともに滑らかに変化するスペクトル包絡と，基本周波数の時系列，音源の周期性の程度を表す指標の 3 個の成分に分解される．この 3 個の成分から再合成された音声は，元の音声と同等の自然性を有しており，処理による品質の劣化は実用上無視できるレベルにある．しかも，これらの成分は，いずれも正の実数として表現されており，相互の干渉に煩わされることなく独立に加工し再合成のために用いることができる．これらの特長が，本プロジェクトにおける重要な研究手法である音声の高品質モーフィングを可能にしている[15]．



(a) 通常のスペクトル分析



(b) STRAIGHTによる分析

図 21　通常のスペクトル分析と STRAIGHT による分析
STRAIGHT 分析では，より滑らかなスペクトルが推定されている．

本プロジェクトでは，感情そのものの研究は対象外とした．声優が目的に応じて声音を使い分けるのと同様に，用途に応じて多様な声質を合成音声に付与することのできるソフトウェアの開発を目的とした．具体的には，現実に合成音声が使われる場面を想定して目的に応じた声音を声優に演じ分けてもらい，その場合の声質の変化と同様な変化を合成音声に付与することのできる音声変換ソフトウェアの開発を行った．このようなアプローチを採ることにより，「感情の推定や記述」という人間にとってさえ困難な未解決の問題を直接扱わずに，状況に応じて多様な声質を付与することのできる音声合成ソフトウェアの開発を可能とした．

5.3　研究経過

研究の序盤では，多様な声質と話し方の音声試料の収集と，モーフィングを中心とした音声変換方式の研究を進めた．具体的には，まず，組込み機器の応答音声への応用を想定し，家電製品のインタフェースとして用いるメッセージを，多様な話者及び多様な話し方により収録した．この音声データベースの拡充は，研究の序盤と中盤を通じて行われた．音声変換方式については，リアルタイム版のプログラムに適した簡易な変換法の検討を進めるとともに，話し方の多様な変換を可能にする HMM に基づく韻律制御の研究を進めた．

研究の中盤では，リアルタイムの音声変換プログラムの実装の検討を先行させて進めるとともに，高品質の非リアルタイム版のプログラムの C 言語化のための機能分割・プログラムインタフェースの整理を進めた．

研究の終盤では，リアルタイム版及び非リアルタイム版プログラムの実装と，主観及び客観評価に基づく高品質化，設計資料及び声質や話し方のプリセットデータの蓄積を進め，開発されたプログラムや資料をオンラインで利用できるサイトとして公開した．

また，これらの当初計画に従った進ちょくだけではなく，音声変換技術を非可聴つぶやき声の変換に応用することで全く新しい形態のコミュニケーションが可能になるという，大きな発見が加わったことが特記される[18]．

5.4 リアルタイム音声変換プログラム

研究用途での柔軟性を重視して，科学技術計算用の特殊なシステムである Matlab により作成されていた STRAIGHT のアルゴリズムを，リアルタイム処理用に一部を簡略化するとともに処理の流れを新たに設計した．合成時に必要なタイミングでのみ分析を行うアーキテクチャの採用などの工夫と併せて，処理量と必要とする記憶容量の大幅な削減を実現した．こうして再構成されたアルゴリズムを，広く用いられている C 言語で記述することにより，容易に様々な応用システムに組み込むことのできるプログラムとして実装した．また，リアルタイム処理に適した簡易な音声変換アルゴリズムを開発した[17]．図 22 に，開発したリアルタイム版のインタフェースと，品質評価結果を示す．

5.5 非リアルタイム音声変換と応用

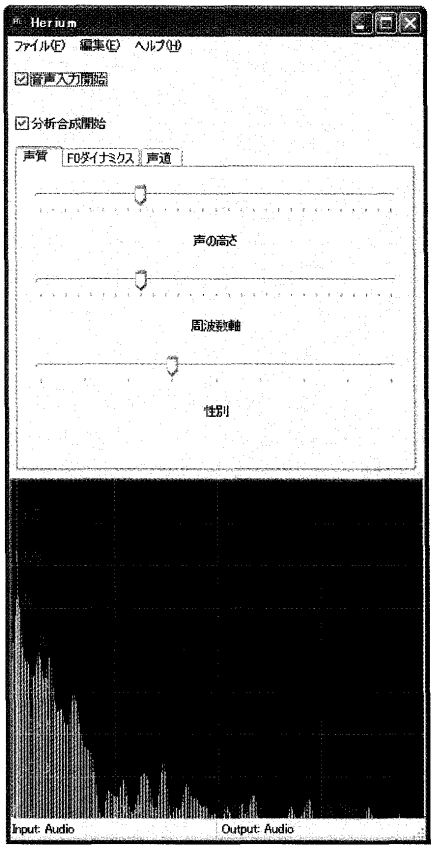
高品質な非リアルタイム版では，応用プログラム開発が容易となるように，Matlab 版の処理を機能モジュールに分割・整理し直し，拡張性のあるプログラムインタフェースを定義した．こうして再設計されたアルゴリズムを，リアルタイム版と同様に C 言語により実装した．また，合成音声の品質を大きく左右する音源情報抽出における品質改善用オプションとして，高精度の基本周波数抽出プログラムを開発し実装した．

こうして開発された非リアルタイム版のアルゴリズムは，後述の隠れマルコフモデルに基づく音声合成技術や統計的音声変換[19]，非可聴つぶやき声による音声コミュニケーションをはじめとして，様々な形で応用されている．

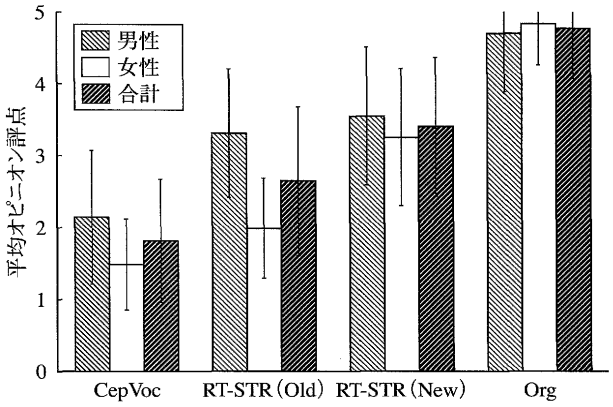
5.6 その他の成果と成果展開

STRAIGHT を応用した統計的音声変換アルゴリズムは，柔軟で高品質な音声変換を可能にした．この変換と，非可聴ささやき声を検出できる特殊なマイクを組み合わせることにより，ほとんど聴こえない声を使ったコミュニケーションという，当初計画には含まれていなかった全く新しいメディアが発明された．また，喉頭摘出等で声を失った方の人工喉頭による機械的な声をささやき声に変換することで，自然なコミュニケーションを回復するなど，新しい応用が次々と生まれている．これらの技術の応用範囲は広大であり，生活の質を向上させるなど社会にも大きなインパクトを与えるものである．

本プロジェクトの成果である STRAIGHT の水準は，対抗する様々な技術を大きく引き離しており，学術の分野では事実上の世界標準となっている．STRAIGHT の基本文献の引用数は，この分野では際立った数であり，現在も増加し続けている．日本科学技術未来館での展示等，メディアへの露出の機会も増加しており，既に数社が本プロジェクトの成果に基づいた商用システムへの応用を計画している．



(a) 操作インタフェース



(b) 品質評価結果

図 22 リアルタイム STRAIGHT の操作インタフェースと品質評価結果　従来法(CepVoc)，リアルタイム版 (RT-STR (Old))，改良 (音源情報抽出部) 後のリアルタイム版 (RT-STR (New)) と，元音声(Org)の主観評価値を比較している．

(河原英紀，高橋　徹，戸田智基，坂野秀樹)

6. HMM に基づく音声合成システム (HMM 音声合成)

6.1 ねらい

近年，劇的に進化する情報通信社会において，人と機械の間の滑らかな意思疎通の実現は重要な課題であり，

ディジタルディバイド等の問題を解決するためにも，人間の最も基本的なコミュニケーション手段である音声を用いたヒューマンインタフェースに大きな期待が寄せられている．中でも音声合成技術は，機械から人への情報提示手段として極めて重要な役割を担うものである．このような背景から，多様な音声表現を自在に実現可能な音声合成方式の確立を目指し，隠れマルコフモデルに基づいた音声合成方式について研究開発を行った．

6.2 アプローチ

従来の代表的な音声合成技術は，単位選択方式と呼ばれ，事前に収録された音声データから適切な波形素片を選択し接続することで音声が生成されるものである．肉声感に優れた自然な音声が得られる反面，①多様な話者性，感情表現等を持った音声を生成することが容易でない，②携帯電話等の計算資源が限られたモバイルデバイスでの実現が困難である，などの問題があった．これらの問題を解決するため，新しい音声合成方式「HMM に基づく音声合成（以下，HMM 音声合成）」を提案し，開発を進めてきた（図 23）．本方式は，これまでに提案した「動的特徴量を用いた HMM からの音声パラメータ生成法」，「多空間確率分布 HMM による基本周波数のモデル化手法」を基幹技術とするものである．本方式には，データに基づいたシステムの自動構築が可能，言語依存性が低く，多くの言語への適用が容易，統計量に基づいて音声を生成するため，限られた計算資源の中で動作可能，などの利点のほか，従来方式にはない利点として，HMM の統計量を表すモデルパラメータを適切に変換することにより，様々な話者の声質，感情表現を伴った音声等，多様な音声を容易に生成できることが挙げられる．

6.3 モデル化精度の向上と理論の深化

隠れマルコフモデル（HMM）を用いた韻律制御モデルについて，パラメータ間の相関を考慮することにより，モデル化精度，声質・話法・感情等の補間性能の向上を図った．その結果，従来システムに対して，プレファレンススコア 35：65 の音声品質の改善を達成した（図 24）．更に，合成音声のスペクトル概形の表現法として，本プロジェクトメンバーの開発による，一般化対数に基づく，メル一般化ケプストラムの線スペクトル対表現（MGC-LSP）を適用することにより，信号の補間性能が向上し，多様な音声の生成において性能向上が確認された（図 25）．

継続的に理論を深化させ，関連アルゴリズムを整備し

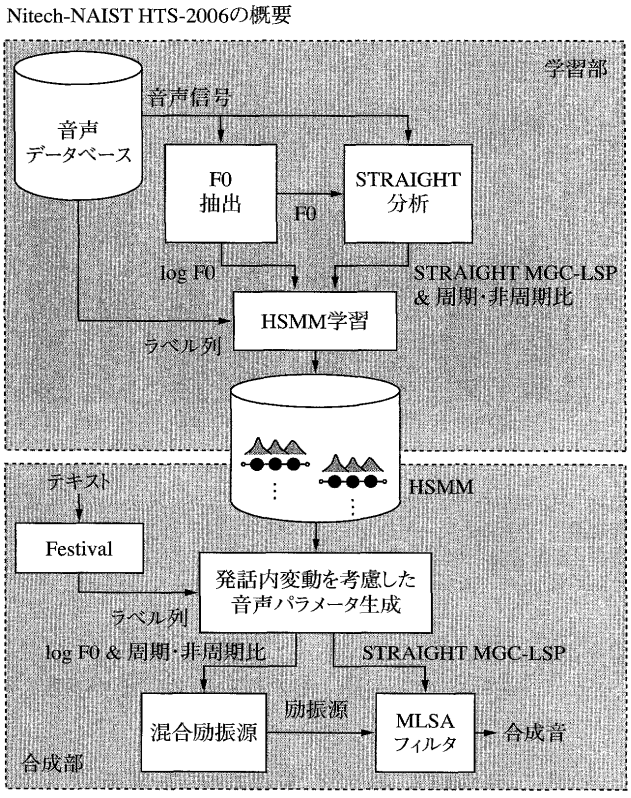


図 23　HTS（隠れマルコフモデルに基づく音声合成システム）の概要　学習用音声データから統計的な音韻モデル（HSMM）が学習される．入力テキストに従って音韻モデルを連結することで，音声の情報表現であるメル一般化メルケプストラムの線スペクトル対（MGC-LSP），基本周波数（F0）と音源の周期性が，時系列として生成される．

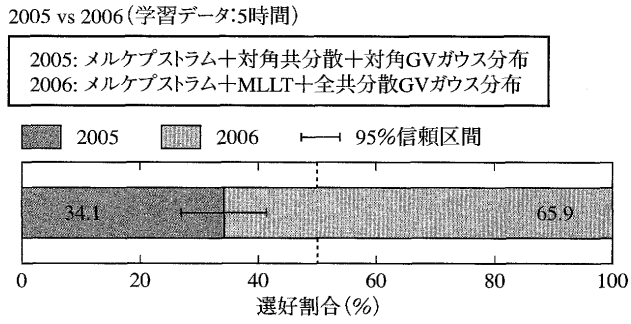


図 24　パラメータ間相関の利用による合成品質の向上

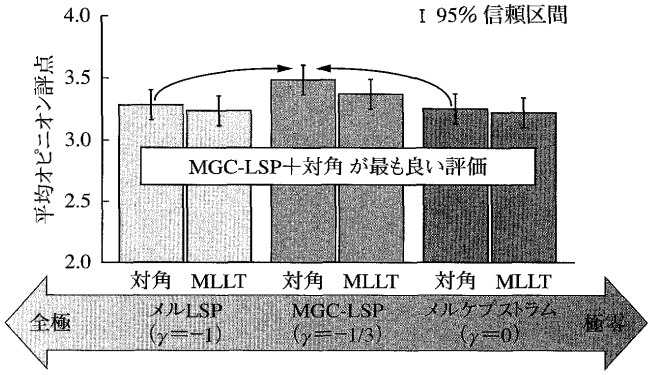


図 25　音声スペクトル概形の表現形式が音声合成品質に与える影響の比較　提案法である，メル一般化対数ケプストラム分析の線スペクトル対表現（MGC-LSP）により，最も高い品質が得られている．

たことから，HMM 音声合成は，新音声合成方式として一つのスタイルを確立しつつある．特にこの過程で「動的特徴量を用いた HMM」が「トラジェクトリ HMM」と呼ぶ全く新しい音声の統計モデルに帰着されることを示した点は，音声合成研究分野のみならず音声処理研究分野全体に反響を呼ぶ結果となった．

6.4 HMM 音声合成のためのソフトウェアツールキットの開発と公開

本プロジェクトで，多様な声質・話法・感情を生成可能な韻律生成プログラムとして開発した「隠れマルコフモデルを用いた韻律制御モデル」に関するソフトウェアは世界的な標準研究基盤ソフトウェアとして定着した．本ソフトウェアを更に改良するとともに，継続的に改良バージョンを公開し，世界的な普及を更に促進した．

多様な声質・話法・感情を持った音声の合成が可能な音声合成システムの新バージョンを HMM-based Speech Synthesis System (HTS)バージョン 2.0.1，2.1α，2.1β，2.1 として公開した．特に，バージョン 2.1 では，話者適応の機能を完全に実装しており，様々な話者の声質や話法を容易に実現することが可能となっている(図 26)．当該ソフトウェアは国内外の有力研究機関より多数のダウンロードがあり，多くの反響を呼んだ．

HMM に基づく韻律制御モデルを中心とする音声合成システム HTS の集大成として，HTS-2 を公開した．このシステムは，STRAIGHT のスペクトル推定を組み込むことにより，世界最先端の音声合成システムとなるように構成されている．公開されたシステムでは，これまでに整備してきた音声データベースを用いて，高品質化のためのモデル化精度及び補間特性の向上が行われている．

6.5 音声合成技術に関する国際評価会の開催

共通のデータベースを用いることにより，様々な音声合成技術を比較・評価するための国際評価会 Blizzard Challenge を，カーネギーメロン大学のアランブラック准教授らと例年行事として 2005 年から 2008 年まで毎年主催した．

また，結果発表及び議論のため，毎年，ワークショップあるいは国際会議のスペシャルセッションを開催した．いずれも多くの参加者を得，大変な活況を呈した．評価会とともに世界的な恒例イベントとして定着した感がある．

HMM 音声合成による音声品質は単位選択型によるものに遠く及ばないであろうとの観測もあったが，STRAIGHT を組み込んだ HMM 音声合成システムは，期せずして Blizzard Challenge 2005 において圧倒的高評価を達成したことから，この通説は覆り，関係分野に衝撃を与えた．大学等の研究機関のみならず，大企業・ベンチャー企業における利用も急速に広がっている．

(徳田恵一，全　炳河)

7. あ と が き

本プロジェクトでは，音声対話による情報インタフェースを広く IT 社会の基盤として普及させるために，音声認識・合成技術について，国内 7 大学が協力して五つの視点から研究開発を進めた．

① 利用環境によらずユーザに負担をかけない話者・環境適応技術
② マイクを意識させない自然なハンズフリー音声認識技術
③ 高精度連続音声認識プログラムと研究開発ワークベンチ
④ 人と機械の音声対話の実証実験
⑤ 多様な声質を合成可能な音声合成・声質変換プログラム

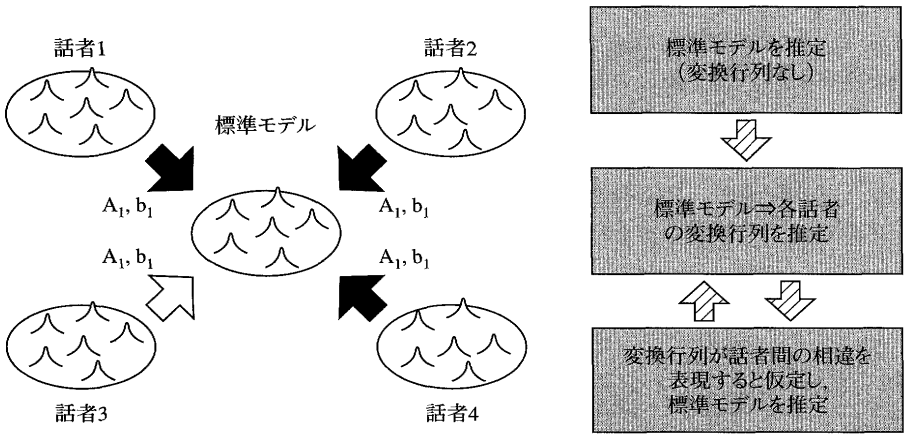


図 26　話者変換の原理　話者の性質を（標準話者パターンから当該話者パターンへの）スペクトル表現パラメータ間の変換行列として学習することで，当該話者の音声を合成する．

表5　フリーソフトウェアサイト（2007年度）

大語彙連続音声認識ソフトウェア　Julius 4.0
http://julius.sourceforge.jp/　ダウンロード数　3万件以上
HMM音声合成ソフトウェア　HTS（HMM-Based Speech Synthesis System）
http://hts.ics.nitech.ac.jp/　ダウンロード数　2万5,000件以上
音声分析合成　STRAIGHT　約350拠点で利用
http://straight-suite.sys.wakayama-u.ac.jp/

それぞれの開発項目において，当初計画以上に研究が進展したことに加えて，新しい音声メディア（NAM：Non-Audible Murmur）やブラインド音源分離（BSS：Blind Source Separation）の新しい原理SIMO-ICAを発見できた．また，大語彙連続音声認識プログラムJuliusや音声合成プログラムSTRAIGHTやHTSの研究開発も大きく進展し，多くの企業や研究機関において使用されるに至っており，国際的にも高く評価されている．

本プロジェクトの研究成果は，学術論文や国際会議を通して積極的に発表してきた．非可聴つぶやき，ブラインド音源分離，音声対話システムなどで多くの賞を受賞した．

フリーソフトウェアの普及活動として，音声認識では，Juliusのサイトを，音声規則合成では，HTS（HMM-Based Speech Synthesis System）のサイトを立ち上げ，オープンソース，フリーソフトウェアの活動を継続している．本プロジェクトで開発された主なフリーソフトウェアについて，2007年度の配布実績を表5にまとめる．

本プロジェクトの成功により，音声認識・対話技術を多様な場面で情報インタフェースに活用する基盤が確立された．しかし，これらの技術を最大限に生かすためのシステム設計指針やヒューマンファクタなど，今後の研究進展が期待される領域も依然として残っている．今後とも音声認識，音声合成プログムの普及に努め，だれでもが容易に音声認識・合成技術が利用できるソフトウェア環境を確立する努力を継続している．

謝辞　本稿で述べた研究開発は，文部科学省のリーディングプロジェクト「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」で実施されたものである．本プロジェクトに参加・協力頂いた奈良先端科学技術大学院大学，京都大学，名古屋大学，和歌山大学，名古屋工業大学，立命館大学，日立製作所，旭化成，パナソニック，オムロン，パナソニック電工の皆様に感謝します．

文　献

（1）　河原達也，荒木雅弘，音声対話システム，オーム社，2006.
（2）　鹿野清宏，伊藤克亘，河原達也，武田一哉，山本幹雄，音声認識システム，オーム社，2001.
（3）　翠　輝久，河原達也，“ドメインとスタイルを考慮したWebテキストの選択による音声対話システム用言語モデルの構築，”信学論(D)，vol.J90-D, no.11, pp.3024-3032, Nov. 2007.
（4）　河原達也，李　晃伸，“連続音声認識ソフトウェアJulius，”人工知能誌，vol.20, no.1, pp.41-49, 2005.
（5）　R. Gomez, T. Toda, H. Saruwatari, and K. Shikano, “Reducing computation time of the rapid unsupervised speaker adaptation based on HMM-sufficient statistics,” IEICE Trans. Inf. & Syst., vol.E90-D, no.2, pp.554-561, Feb. 2007.
（6）　T. Cincarek, H. Kawanami, R. Nishimura, A. Lee, H. Saruwatari, and K. Shikano, “Development, long-term operation and portability of a real-environment speech-oriented guidance system,” IEICE Trans. Inf. & Syst., pp.576-587, March 2008.
（7）　T. Cincarek, T. Toda, H. Saruwatari, and K. Shikano, “Utterance-based selective training for the automatic creation of task-dependent acoustic models,” IEICE Trans. Inf. & Syst., vol.E89-D, no.3, pp.962-969, March 2006.
（8）　中島淑貴，柏岡秀紀，ニックキャンベル，鹿野清宏，“非可聴つぶやき認識，”信学論(D-Ⅱ)，vol.J87-D-Ⅱ，no.9, pp.1757-1764, Sept. 2004.
（9）　M. Nakagiri, T. Toda, H. Kashioka, and K. Shikano, “Improving body transmitted unvoiced speech with statistical voice conversion,” The 9th International Conference on Spoken Language Processing, pp.2270-2273, Sept. 2006.
（10）　中村圭吾，戸田智基，猿渡　洋，鹿野清宏，“肉伝導人工音声の変換に基づく喉頭全摘出者のための音声コミュニケーション支援システム，”信学論(D)，vol.J90-D, no.3, pp.780-787, March 2007.
（11）　小島摩里子，川波弘道，猿渡　洋，松井知子，鹿野清宏，“非可聴つぶやき声を用いた個人認証，”ユビキタスネットワーク社会におけるバイオメトリクスセキュリティ研究会，no.2D1-4, Jan. 2006.
（12）　Y. Ohashi, T. Nishikawa, H. Saruwatari, A. Lee, and K. Shikano, “Noise-robust hands-free speech recognition based on spatial subtraction array and known noise superimposition,” Proc. International Conference on Intelligent Robots and Systems (IROS2005), pp.533-537, 2005.
（13）　Y. Takahashi, T. Takatani, H. Saruwatari, and K. Shikano, “Blind spatial subtraction array with independent component analysys for hands-free speech recognition,” Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC), 2006.
（14）　傳田遊亀，田中貴雅，溝口　遊，中山雅人，西浦敬信，山下洋一，“話者方位推定を利用した動的時間領域処理に基づく遠隔発話区間検出，”信学論(D)，vol.J92-D, no.1, pp.112-122, Jan. 2009.
（15）　H. Kawahara and H. Matsui, “Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation,” ICASSP2003, pp.6-10, 2003.
（16）　河原英紀，“Vocoderのもう一つの可能性を探る―音声分析変換合成システムSTRAIGHTの背景と展開―，”音響誌，vol.63, no.8, pp.442-449, 2007.
（17）　H. Banno, H. Hata, M. Morise, T. Takahashi, T. Irino, and H. Kawahara, “Implementatioin of realtime STRAIGHT speech manipulation system; Report on its first implementation,” Acoustical Science and Technology, vol.28, no.3, pp.140-146, 2007.
（18）　中村圭吾，戸田智基，猿渡　洋，鹿野清宏，“肉伝導人工音声の変換に基づく喉頭全摘出者のための音声コミュニケーション支援システム，”信学論(D)，vol.J90-D, no.3, pp.780-787, March 2007.
（19）　T. Toda, A.W. Black, and K. Tokuda, “Voice conversion based on maximum likelihood estimation of spectral parameter trajectory,” IEEE Trans. Audio Speech and Language Processing, vol.15, no.8, pp.2222-2235, 2007.

（平成20年11月10日受付　平成20年12月10日最終受付）

鹿野 清宏（正員：フェロー）
昭45名大・工・電気卒．昭47同大学院修士課程了．同年日本電信電話公社(現NTT)武蔵野電気通信研究所入所．昭61～平2 ATR自動翻訳電話研究所音声情報処理研究室長．平6より奈良先端科学技術大学院大学情報科学研究科教授．本会平16，17年度論文賞，平16年度猪瀬賞．情報処理学会，IEEEフェロー

武田 一哉（正員）
昭58名大・工・電気卒．昭60同大学院修士課程了．同年KDD入社．昭61～平2 ATR自動翻訳電話研究所．平6名大・工・助教授を経て，現在名大大学院情報科学研究科教授．音響・音声・行動の信号処理に関する研究に従事．平12年度本会論文賞．日本音響学会，情報処理学会，IEEE，自動車技術会各会員．

河原 達也（正員）
昭62京大・工・情報卒．平2同大学院博士後期課程退学．同年京大・工・助手．同助教授を経て，現在京大学術情報メディアセンター教授．音声言語処理，特に音声認識及び対話システムに関する研究に従事．京大博士(工学)．著書「音声対話システム」など．

河原 英紀（正員）
昭47北大・工・電子卒．昭52同大学院博士課程了．同年日本電信電話公社(現NTT)武蔵野電気通信研究所．電話通話品質，音声知覚等の研究に従事．平4 ATR，平9和歌山大・システム工・教授．工博．平7 EURASIP最優秀論文賞．

猿渡 洋（正員）
平3名大・工・電気卒．平5同大学院修士課程了．平12同大学院博士課程了．工博．平5セコム(株)入社．アレー信号処理の研究に従事．平12奈良先端大助教授．現在同准教授．音声信号処理の研究に従事．本会平12，17年度論文賞受賞．

徳田 恵一（正員）
昭59名工大・工・電子卒．平元東工大大学院博士課程了．同年東工大助手．現在，名工大教授．工博．音声言語情報処理，統計的学習理論の研究に従事．本会平12年度論文賞，猪瀬賞，平19年度本会情報・システムソサイエティ論文賞，平成13,20電気通信普及財団賞各受賞．

李 晃伸（正員）
平8京大・工・情報卒．平10同大学院修士課程了．平12同大学院情報学研究科博士課程了．同年より奈良先端大情報科学研究科助手．平17より名工大大学院工学研究科助教授．主として音声認識・音声言語理解の研究に従事．博士(情報学)．平14日本音響学会粟屋潔学術奨励賞受賞．平19情報処理学会山下記念研究賞受賞．情報処理学会，日本音響学会，IEEE，ISCA各会員．

川波 弘道（正員）
平6東大・工・電気卒．平12同大学院工学系研究科博士課程了．博士(工学)．同年電子技術総合研究所入所．平13より奈良先端科学技術大学院大学情報科学研究科助手，平19同助教．現在，音声分析，音声対話の研究に従事．日本音響学会会員．

西村 竜一（正員）
平11名大・工・電気系卒．平16奈良先端大博士課程了．同年和歌山大・システム工・助手．現在，同助教．博士(工学)．情報処理学会第65回全国大会奨励賞，FIT2007ヤングリサーチャー賞各受賞．情報処理学会，日本音響学会，ISCA各会員．

Randy GOMEZ
平18奈良先端大博士後期課程了．同大学研究員を経て，現在京大研究員．

戸田 智基（正員）
平11名大・工・電気卒．平15奈良先端大情報科学研究科博士課程了．同年JSPS特別研究員．平17奈良先端大・情報・助手．平19同助教．工博．音声情報処理の研究に従事．平14，19年度電気通信普及財団賞，平19年度本会情報・システムソサイエティ論文賞，平20年度エリクソン・アワード各受賞．

西浦 敬信（正員）
平13奈良先端大情報科学研究科博士後期課程了．同年和歌山大・システム工・助手．平16立命館大・情報理工・助教授．現在，同准教授．音響信号処理，主として音環境の理解・生成に関する研究に従事．博士(工学)．

高橋 徹（正員）
平8名工大・工・知能情報システム卒．平16同大学院博士課程電気情報工学専攻了．同年和歌山大・システム工・産学官連携研究員．音声合成の研究に従事．平20京大大学院情報学研究科GCOE助教．ロボット聴覚の研究に従事．博士(工学)．

坂野 秀樹（正員）
平8名大・工・電気系卒．平13同大学院工学研究科博士(後期)課程了．名大，和歌山大助手を経て，現在名城大准教授．音声信号処理とその応用に関する研究に従事．博士(工学)．

全 炳河
平18名工大大学院工学研究科博士課程了．現在東芝ケンブリッジ研究所研究員．音声認識・合成の研究に従事．平18日本音響学会粟屋潔学術奨励賞，平20日本音響学会独創研究奨励賞板倉記念，電気通信普及財団テレコムシステム技術賞，本会情報・システムソサイエティ論文賞各受賞．