5 ボタンレス音声インタフェースのための音声コマンド検知技術

大淵 康成
日立製作所中央研究所

ボタンレス音声インタフェースの意義

音声インタフェースのメリットを語るときによく使われるのが，”Hands-free/Eyes-free” というフレーズである．たとえば運転中のカーナビの操作を考えると，手はハンドル操作に，目は前方注視に専念することができれば，安全性の確保という意味で大変望ましい．

それでは，今の世の中を見渡したときに，手と目とを 100% 頼らずに使える音声インタフェースが，どれだけあるだろうか．カーナビ製品に音声認識機能が付いているものは珍しくないが，「発話ボタン」のない機種は，さすがに見当たらない．最近では Web の検索だって音声でできるが，それでもアプリの起動はタッチパネルで行う．これらのアプリケーションは，「起動ボタンを押さざるを得ないとしても，その後が音声だけでできるならば嬉しい」という性質を持っているために，いちはやく世に出ることができたわけだが，逆に言うと，「起動も含めて音声だけでできないと嬉しくない」というようなニーズに応えるアプリケーションは，残念ながらまだ広く普及するに至っていない．

真に Hands-free/Eyes-free なアプリケーションとして期待されるものの典型的な例が，テレビやエアコンなどの家電品の操作であろう．家電メーカが「テレビのリモコンで困ったことは？」という消費者調査を行うと，「必要なときにリモコンが見つからない」という答えが必ず上位にくることからも分かるように，リモコン不要で遠隔操作ができるようなテレビがあれば，多くの人に喜ばれることは間違いない．あるいは，真っ暗な部屋で「電気をつけて」と言うだけで照明が点灯されるようになれば，手探りでスイッチを探して苦労することもない．

それでは，そのような音声インタフェースは，なぜ実現しないのであろうか．答えは簡単である．誤作動が多すぎるのだ．1 日のうちでテレビやエアコンを操作する回数は，少ない人で十数回，多い人でも百〜二百回程度だろうか．一方，常に音を取り込む状態になっているマイクロホンには，多いときで 1 日数千回の信号が入ってくる．したがって，わずか 1% の誤作動率でも，「何も命令していないのに，テレビのチャンネルが勝手に変わってしまう」というようなことが頻繁に起きてしまうわけである．

このような誤作動を起こさず，機器の操作を目的として発せられた声だけを聞き分ける技術を，本稿では音声コマンド検知技術と呼ぶことにする．音声コマンド検知技術は，検知した音声コマンドを聞き分ける音声認識技術と不可分であるが，後者の研究が，耐雑音性の向上も含めて幅広く行われているのに対し，前者についてはこれまで必ずしも十分な研究が行われてこなかった．しかし，真にボタンレスの音声インタフェースが実現すれば，たかだか数単語程度しか認識できないようなものであっても十分に役立つ場面が数多く存在する．つまり，耐雑音音声認識技術とは切り離したかたちで，音声コマンド検知技術を議論することの重要性は明白である．

以下では，音声コマンド検知を困難たらしめるさまざまな環境音についての分析を行ったのち，それらと音声コマンドとを聞き分けるための技術について，特に特徴抽出という視点で詳しく解説する．また，実際のシステムを想定して行った実証実験などの例と合わせて，実用へ向けての現状と将来への課題を示す．

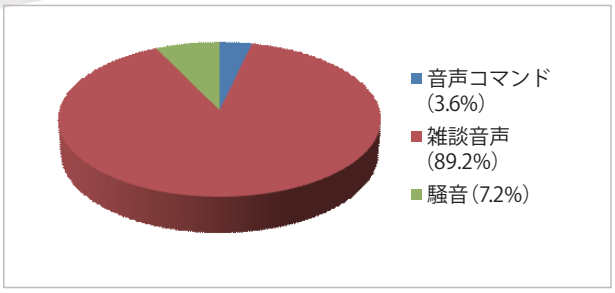

図 -1　生活環境で検出される音の分類例

日常生活の中で検出される音

人間の聴覚は環境に対する適応性が高く，日常の生活において，当たり前のように検出される音の多くは，意識にすら上らないことが多い．しかし，音声コマンド検知の誤作動を防ぐには，およそ検出される可能性のあるすべての音に対し，それらの性質を知っておくことが不可欠である．

図 -1 は，一般家庭のリビングルームを模擬した実験環境で，2 ～ 3 名の被験者が実際に生活した際に，検出された音を分類したものである [1)]．ここでは，実際にテレビ操作用の音声認識装置を設置し，それを用いるための音声コマンドの検出頻度も測定している．実験では，フレーム処理をした入力信号に対するパワーを計算し，その結果に対して単純な閾値処理で音声コマンド候補フレームを抽出した．その後，そのような候補フレームが一定時間以上継続するもの（ただし短時間のポーズを含むことは許容する）を，音声コマンド候補セグメントとして取り出し，それを人間が実際に聴取して，音声コマンド・それ以外の雑談音声・人間の声以外の騒音の 3 種類に分類した．図から分かるように，誤作動のもとになる妨害音の大半が雑談音声であり，掃除機や洗濯機が出すような騒音は，それに比べると頻度が少なかった．

もちろん，このような分布は，実験環境によって大きく変動する．この実験は比較的閑静な住宅で行われたが，それより遙かにうるさい環境の住宅も存在する．また，複数名の被験者がリビングルーム内に滞在し続けるという状況は，雑談音声を生じさせやすいと思われるが，部屋に 1 人しか居ない場合や，何らかの行為に専念していて雑談を行わない場合などには，この分布は大きく変わると思われる．さらに，上述した候補セグメントの抽出アルゴリズムに用いた，パワー閾値や継続長閾値の設定によっても，この分布は変わる（パワー閾値を小さく，あるいは継続長閾値を短くするほど，関係のない雑音・騒音を拾いやすくなる）．また，マイクの設置位置も大きな要因の 1 つであり，被験者が滞留しやすい位置（ソファーやダイニングテーブルの近くなど）にマイクを設置すると，雑談音声が検知される確率が高まる．言うまでもなく，音声インタフェースとしての利便性を高めるためには，多くの時間を過ごす場所の近くにマイクがあることが望ましいが，その場合，雑談音声を拾ってしまうケースもそれだけ多くなるというわけである．

なお，ここでは明示的に示されていないが，テレビの音声がマイクに入ってくることも多い．音声コマンド検知システムがテレビ本体に組み込まれている場合には，エコーキャンセラによってこれを取り除くことは比較的容易だが，独立したシステムとして音声コマンド検知を行う際には，テレビ音声の棄却も重要かつ困難な問題の 1 つとなる．

音声コマンド検知のための特徴量

音声コマンド検知は，本質的には音声コマンドと非音声コマンドの二値の分類問題であり，特徴量抽出と分類器の組合せで考えることができる．以下では，音声コマンド検知に有効と思われるさまざまな特徴量について詳しく述べる．

●音声パワーに基づく特徴量

音声通信などの分野においては，古くから，通話中の音声区間と無音区間とを区別し，音声区間の情報だけを伝送することによって，帯域幅を節約することが行われてきた．このような仕組みは，音声アクティビティ検出（Voice Activity Detection：VAD）と呼ばれ，その後，音声認識の分野でも活用

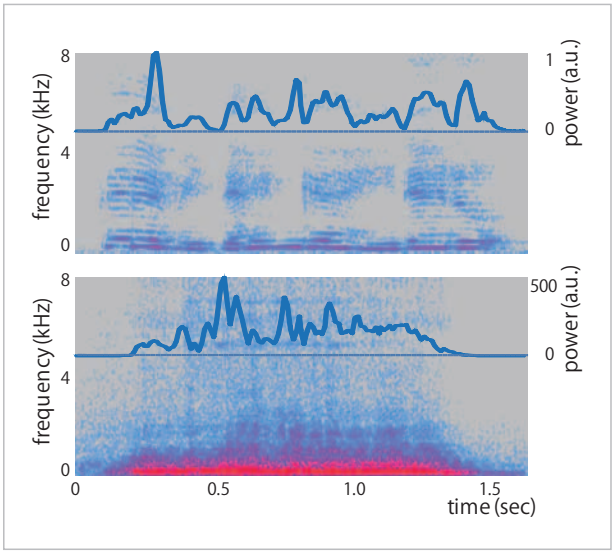

図 -2　音声コマンド（上）と雑音（下）のパワーおよびスペクトルの比較．音声コマンドのパワーは，見やすいように 500 倍に拡大してある．

されている．VAD で用いる最も基本的な特徴量は，短時間の音声パワーである．パワーの計算は処理量が少なく，なおかつ静音環境においてはパワーの有無が音声信号の有無に直結していることから，多くのケースにおいて十分な検出性能が得られる．ただし，雑音環境においては，想定される音声コマンドよりも大きなパワーを持つ雑音が検出されることも頻繁にあり，パワーだけによる音声コマンド検知は容易ではない．図 -2 は，テレビ操作用の音声コマンド発声（上）と，椅子を引きずった際の雑音（下）のパワー（青）とスペクトル（赤）の例である．単純なパワーで見ると，雑音の方が約 500 倍も大きく，パワーだけを使う方式では誤受理が起きてしまう．しかし，スペクトルの形状をよく見ると，両者の間に顕著な違いがあることが見て取れる．このような違いを検知するため，以下に述べるようなさまざまな特徴量を導入する．

なお，本来の音声コマンド検知は，連続的な音声信号の中から音声コマンド部分だけを抽出するというものであるが，本稿では，問題を単純化するため，まずはじめに音声パワーによる粗いセグメント抽出を行った後，二値分類問題として音声コマンド検知を定義した．

●音韻性を表す特徴量

音声認識の分野で使われる特徴量の代表的なものとして，メル周波数ケプストラム係数（Mel Frequency Cepstral Coefficient：MFCC）がある．MFCC は，対数パワースペクトルの逆コサイン変換によって得られる特徴量で，低次の係数にスペクトル包絡の情報が，高次の係数にスペクトルの微細構造が反映される．前者は主に声道の共振特性や口唇の放射特性などに対応し，異なる音素を発声するための口の動きに応じて変化する．一方，後者は声帯から発せられる音源の特徴に対応し，声の高さを調節する動作に応じて変化する．そこで，前者を表す低次の係数だけを取り出すことで，音声に含まれる韻律情報の影響を取り除き，個々の単語を構成する音韻性の違いを識別することができる．一般的には，20 ～ 25ms 程度の窓幅のフレームで切り出した音声から，12 ～ 15 次程度の MFCC を求める．さらに，隣接フレーム間の差分（必要に応じて 2 階差分）を加えることにより，音声認識率が向上することも知られている．

これらの特徴量が「入力音がどの音素に似ているか」という識別に有効であるとすると，それをうまく使うことにより，「入力音が何らかの音素に似ているか」の識別も可能なはずである．実際，MFCC を使った単純な識別器でも実用的な VAD 性能が得られることが知られているし，大量の音声データベースにおける MFCC 特徴量を統計処理して混合ガウス分布モデルで表すことで，さらに精度を向上させることもできる．このように，MFCC 特徴量を使って高精度の VAD を実現することは，近年の音声認識研究の重要なテーマの 1 つであり，特徴量＋分類器という静的なモデルだけでなく，音声や雑音の発生源の時間変化も含めた動的なモデルも数多く提唱されている [2)]．こうした研究の成果により，起動ボタンや対話制御などの助けで音声コマンドの存在がある程度予見できる状況においては，かなりの雑音が存在する環境であっても，音声の始終端の正確な位置を特定することも可能になってきている．

●言語性を表す特徴量

音声コマンドに代表されるような，明瞭に発話された言語音声を，その他の雑音等と識別するための因子として，これまでに述べたような短時間の音韻性だけでなく，それらの時系列変化の様子も挙げられる．言うまでもなく，音声コマンドとして発せられる可能性のあるすべての単語を知っていれば，それらのパターンと入力音とを比較すればよい．実際には，入力音に最も似ているパターン（音声認識装置の出力に等しい）との類似度を見れば十分であり，これは音声認識の分野で発話検証と呼ばれる技術に対応する．発話検証においては，MFCC などで表わされる入力音声の特徴量と，隠れマルコフモデル（Hidden Markov Model : HMM）などで表わされる単語や文章のモデルとの類似度スコアに着目するが，一般に，このスコアが取る値は環境依存性が大きいことから，何らかの正規化処理を行ったのちに閾値処理することになる．

一方，発せられる単語を知らずに音声コマンド検知を行わなければならないケースもある．実際には，ある言語の中に含まれる単語は有限であることから，汎用の音声認識装置との併用で発話検証を行うことも可能だが，多くの場合，処理量の増大に比べて得られる性能はさほど高くならない．

単語の詳細を知らずに入力音声を分類したいというタスクとして，言語識別がある．言語識別の代表的な方式である PRLM（Phone Recognition followed by Language Modeling）[3] は，処理量も比較的軽く，音声コマンド検知に適用することも難しくない．PRLM では，入力された音データに対して制約なし音素認識を行い，得られた音素列に対して，あらかじめ学習した bi-gram（2 音素連鎖）や tri-gram（3 音素連鎖）の出現確率に比例したスコアを得る．音声コマンドにありがちな音素の並びであれば，それだけ音声コマンドである可能性が高いと判定するわけである．

●韻律に基づく特徴量

人が，突然誰かに話しかけられたとき，その内容はまったく聞き取れないにもかかわらず，自分に向かって話しかけられたということだけが分かり，「え，何？」といった反応をすることがある．このような場合，話しかけた言葉の内容より，イントネーションが重要な役割を担っていると思われる．

別の例として，文字で表すとまったく同じ内容であっても，イントネーションによって意味の違いが明白であるケースもある．テレビのリモコンを持っている人に向かって「3 チャンネル」と要求口調でチャンネル変更を求める場合と，単に「3 チャンネル」と独り言を言う場合とは，人間であれば容易に区別できる．

このような例からも分かるように，イントネーション（韻律）を表す特徴量は，音声コマンド検知の重要な要素の 1 つとなり得る．韻律特徴量の代表的なものとして，短時間フレームに対して得られたパワーや基本周波数の最大値・最小値・ダイナミックレンジ・平均・標準偏差・回帰係数などが挙げられる．実際，これらの特徴量を使って，音声に込められた話者の怒り・喜び・悲しみなどの感情を，ある程度識別できることも示されている [4]．発話者の内的心理状態を知るという意味では，音声コマンド検知における発話意図推定も類似の課題であり，韻律特徴量が有効に働くと期待される．

●音響全般の識別のための特徴量

人間の声の識別ではなく，さまざまな環境音などを特定のカテゴリに分類するタスクは，オーディオ分類（Audio Classification）と呼ばれ，これまでにもさまざまな手法が提案されている．音声認識に用いる MFCC などの特徴量を用いるケースもあるが，それ以外の特徴量も数多く用いられる．音声が広帯域の信号であるのに対し，ある種の機械音などは特定の周波数にパワーが集中していることがあり，そのような様子を見るため，帯域を絞ったサブバンドパワーを特徴量として用いることもある．また，スペクトルパターンから見えてくるその他の特徴量として，重心（セントロイド），標準偏差（バンド幅），スペクトラルエントロピーなども有効である．

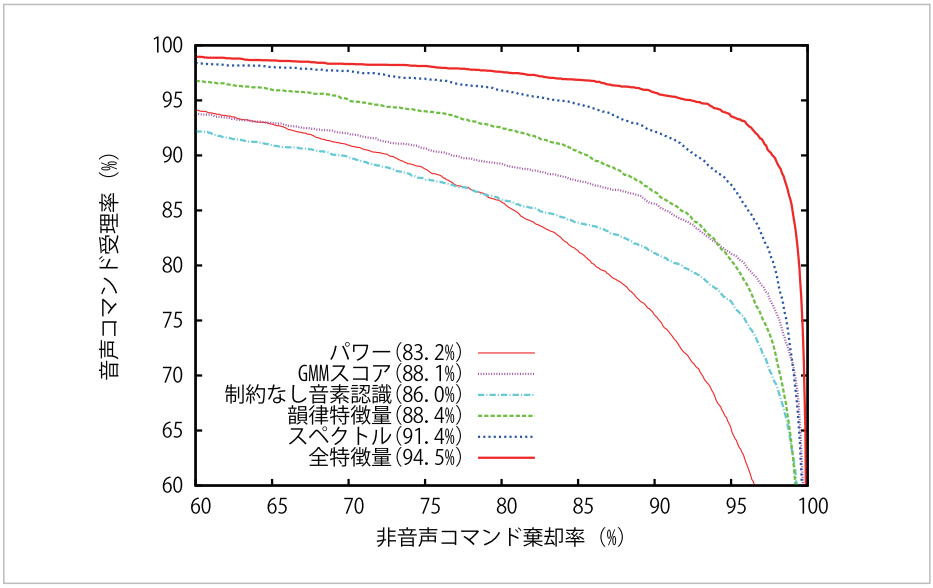

図 -3　音声コマンド検知の実験結果．分類器には LDA を使用．括弧内は閾値を変動させた際の平均判別率の最大値．

これらの特徴量は，短時間フレームに対して求められるものであるが，数十〜数百フレームから成るセグメントに対しては，韻律特徴量の場合と同様に，平均や分散，回帰係数などのかたちで利用可能である．

●特徴量の併用による精度向上

筆者らは，図 -1 に示したデータを対象とし，これまでに挙げた特徴量を使った評価実験を行った[1]．分類器として線形判別分析（Linear Discriminant Analysis : LDA）を用い，閾値をさまざまに変化させることにより，音声コマンド受理率と非音声コマンドの棄却率がどのように変化するかを調べた結果を図 -3 に示す．また，両者の平均値を平均判別率と定義し，閾値の変動に対する平均判別率の最大値を括弧内に示した．ここでは，パワー，MFCC に基づく GMM スコア（音韻性を表す），制約なし音素認識（実際には音節認識を併用）による連鎖のスコア（言語性を表す），韻律特徴量，スペクトル特徴量（音響全般の識別）のそれぞれに対する値に加え，これらすべての特徴量を連結したもの（115 次元）による評価結果を示している．個々の種類の特徴量では，83.2% 〜 91.4% という平均判別率しか得られていないのに対し，さまざまな特徴量を併用することにより，94.5% まで平均判別率を向上させることができた．

システムとしてのアプローチ

ここまで，音声コマンド検知のベースとなるさまざまな特徴量について述べてきた．これらを元に，ボタンレスの音声インタフェースをどのように実現していくべきか，システムとしての観点から整理してみる．

●分類器の選択

個々のセグメントに数十次元程度の特徴量が付与され，それを 2 クラスに分類するという点で，音声コマンド検知は，機械学習のきわめてシンプルな応用問題となる．分類器としては，一般に知られている多くのものが利用可能であるが，当然のことながら，それぞれの分類器には得手不得手がある．

決定木（Decision Tree）は，特徴量の各要素に対する条件比較を重視する場合に適しており，学習の高速性や学習結果の解釈容易性といった長所があるが，特徴量を構成する要素間の相関が大きい場合など，必ずしも効率的な学習ができず，十分な精度が得られないこともある．

前述した線形判別分析は，特徴量空間を 2 つのクラスに分割する超平面を求める方式で，対象となるクラスの分布が綺麗に分かれている場合には，比較的短時間の学習で，高精度の分類が可能である．し

かし，非線形な分布には原理的に適用不可能である．また，特定の特徴量が極端なダイナミックレンジを取る場合などは，分類がその特徴量に過度に依存してしまうこともあるため，対数化などの適切な変換が必要となる．実際，特徴量の中のパワーを対数パワーで置き換えることにより，LDA の分類精度が向上するという様子も見られた．

最後に，サポートベクトルマシン（Support Vector Machine：SVM）に代表される，非線形の分類器を使うことも可能である．SVM は，非線形の分布を持つ 2 クラスの分類も可能であり，安定して高い性能を示す傾向がある．ただし，決定木や LDA などに比べると，特に高次元の特徴量を使う場合には学習の計算量が膨大になり，実装上の工夫が必要となる．

筆者らは，図 -3 に示した全特徴量を用い，上記の 3 つの分類器を比較する実験を行った．その結果，決定木で 91.1%，LDA で 94.6%，SVM で 94.5% という平均判別率を得た [1]．この値からは，音声コマンド検知が，線形判別でも十分な精度を得ることが可能な，線形判別性の比較的高いタスクであることが見てとれる．

●前処理の高度化

ここまで，単一マイクに入力された音データの特徴だけから，音声コマンド検知を行う方式について述べてきた．一方，家電品の操作などを目的とする場合，複数のマイクや，それ以外のセンサからの信号などを活用することによって，利便性を大きく増すことが期待できる．

典型的な例は，マイクロホンアレイによる目的音抽出と音源方向推定である．音声認識の前処理としてマイクロホンアレイを活用することにより，特定音源からの音のみを抽出することが可能になる．音源の特定には，ビームフォーマのように方向で指定するものや，ブラインド音源分離と呼ばれる一連の方法で，方向を明示的に指定せずに行うものなどがある．いずれの場合も，複数のマイクに到達する音響信号の位相差に着目することにより，特定の音だけを抽出することが可能になる．また，音の分離までは行わなくとも，音源の方向を特定することができれば，その場所が「音声コマンドを発する人が滞留しやすい場所かどうか」という知識と照らし合わせることにより，検知精度を高めることができる．家庭環境におけるマイクロホンアレイ処理では，部屋の壁や天井による反響が大きな障害となるが，近年では，反響除去のさまざまなアルゴリズムも提案されており，こうしたアプローチの有効性は高まっていると言えるだろう．

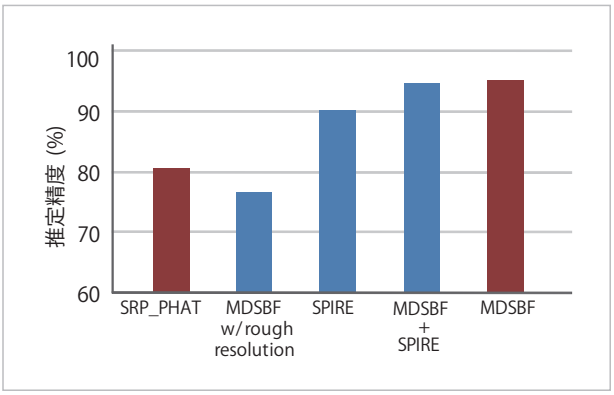

図 -4　音源方向推定の性能評価．赤が高処理量，青が低処理量の方式を示す．低処理量の "MDSBF+SPIRE" でも，95% 近くの推定精度が得られている．

図 -4 は，筆者らがロボットを対象に開発した音源方向推定方式の性能評価実験の結果である [5]．組込み用途を意識した低処理量方式（赤で示された方式に比べておよそ 50 分の 1 の処理量）である MDSBF＋SPIRE 方式でも，95% 近くの精度で音源方向を正しく推定しており，無関係な方向からの非音声コマンド入力を棄却するためのツールとして有効に機能することが期待される．

テレビやラジオなど，電気信号をもとに自ら音を発する機器については，エコーキャンセラによる再生音の抑圧も有効である．特に，機器やマイクの位置関係があまり動かない家庭環境では，いったん学習した伝達特性がさほど変化しないことから，高いエコー抑圧率を得やすい．また，テレビやラジオの音に関しては，特に音声コマンドと間違えやすい人間の声が多数含まれることから，エコーキャンセラへの期待が高い．実際，近年のデジタルテレビのよ

5 ボタンレス音声インタフェースのための音声コマンド検知技術

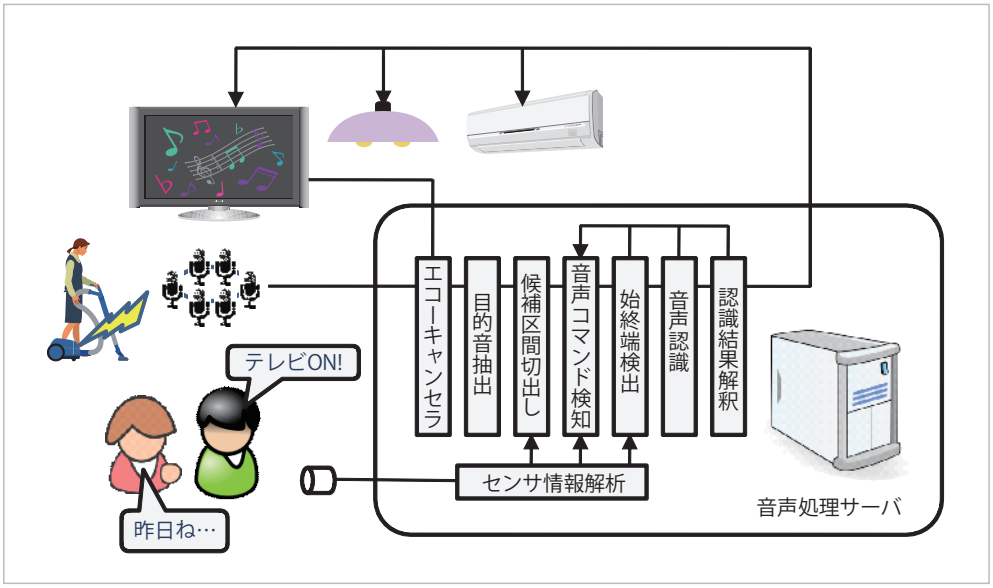

図-5　ボタンレス音声インタフェースを持つ家電システムのイメージ．

うに，ある程度高性能のプロセッサを内蔵し，多様な番組選択のための音声インタフェースが求められるような機器を対象にする場合，マイクロホンアレイ処理やエコーキャンセラを用いた高度な前処理についての研究が活発に行われている[6]．

これらの前処理の有効性については疑問の余地はないが，一方で，これらの処理を加えることにより，システムの複雑度が増し，結果的にコスト高につながってしまうという難点もある．

●状況理解の高度化

図-5に，ボタンレス音声インタフェースを持つ家電システムのイメージを示す．マイク装置には，ユーザによる音声コマンドのほかに，雑談音声，生活騒音，テレビなどからの音などが入ってくる．これらのデータは音声処理サーバに送られ，さまざまな処理が行われていく．そして，これらの処理の結果に基づき，テレビ，照明，エアコンなどの機器に，オン／オフなどの指令が送られる．

前節までは，前処理から音声コマンド検知に至る，信号処理を中心とした技術について述べてきた．一方，音声コマンド検知の最終目的が，検知したコマンドの認識およびそれに基づく機器の操作であることを考えると，それより後段の処理によって，システムの置かれた状況を理解することにより，全体としての使い勝手をさらに向上させることができる．

たとえば，音声以外のセンサにより，状況理解のためのヒントを得ることもできる．現在でも，赤外線などを活用した人感センサにより，人が存在する部屋の電気だけを点灯させるといったことが行われている．同様に，人が存在する部屋の音声取り込みだけを起動しておくことにより，無駄な誤作動と電力消費を減らすことができる．また，カメラによる画像認識を併用して人のいる場所を推定し，音源方向推定と組み合わせて音声コマンド検知の精度を向上させることも可能である．

音声の始終端検出は，そもそも候補区間の切り出しの段階で行われるべきものであるが，処理量の観点からは，簡単な切出しと簡単な音声コマンド検知を行い，そこで検出された音声コマンドだけに対して，処理量の重い高度な始終端検出を行うという方法も有効である．このとき，始終端検出の結果をもとに，再度音声コマンド検知を行ったり，音声コマンド検知の閾値を補正したりといったことも考えられる．同様に，音声認識の結果からのフィードバックとして，たとえば想定されるコマンド以外の発話があまりに多い場合は，閾値を厳しく設定して受理率を低めるといったことも可能であるし，その後の認識結果の解釈（たとえば，同じコマンドに対して普段とは異なる語彙を用いた場合を疑わしいと判定

するなど）に基づいて，パラメータの補正を行うことも考えられるだろう．雑音の環境依存性に応じた音声コマンド検知の感度補正の重要性を考えると，このような後段の情報をいかに適切にフィードバックしていくかによって，システム全体の使い勝手が大きく変わると言っても過言ではない．

今後の展望

本稿では，起動ボタンを使わずとも，いつでも好きなときに声だけで操作することのできるインタフェースの実現に必須となる，音声コマンド検知技術について紹介してきた．雑談音声や機械音などによって誤作動しないインタフェースを実現するためには，さまざまな観点で抽出した特徴を活かした，高度な判別技術が不可欠である．また，マイクロホンアレイ技術に代表される前処理の高度化や，音声認識結果からのフィードバックなどにより，システム全体の性能を向上させることへの期待も大きい．

これまでの音声認識研究は，大規模コーパスの整備に支えられて進展してきたが，不要音も含めた実環境での音声データは，まだまだ十分とはいえない．特に，これらの不要音は，音声よりも遙かに環境依存性が高く，万能の不特定環境モデルを構築することは難しいと思われる．今後，音声コマンド検知の性能を向上させるためには，それぞれの環境でのデータ収集と，それに対する適応処理とが，平易かつ頑健に進められる枠組みを作っていくことが重要であろう．

一方で，システム全体の完成度を上げていくためには，住環境における機器全体のネットワーク化が不可欠である．その中で，機器の配置や個人の嗜好などがデータベース化されることによって，音声インタフェースの高度化もさらに進められていくことであろう．

本稿で紹介した研究成果の一部は，経済産業省「情報家電センサー・ヒューマンインターフェイスデバイス活用技術開発／音声認識基盤技術の開発」(2006-2009)の委託により実施したものです．

参考文献

1) Obuchi, Y. and Sumiyoshi, T. : Intentional Voice Command Detection for Trigger-Free Speech Interface, IEICE Trans. Information and Systems, Vol.E93-D, No.9 (2010).
2) Fujimoto, M. and Ishizuka, K. : Noise Robust Voice Activity Detection Based on Switching Kalman Filter, IEICE Trans. Information and Systems, Vol.E91-D, No.3, pp.467-477 (2008).
3) Zissman, M. A. : Comparison of Four Approaches to Automatic Language Identification of Telephone Speech, IEEE Trans. Speech and Audio Processing, Vol.4, No.1, pp.31-33 (2005).
4) Tato, R., et al. : Emotional Space Improves Emotion Recognition, Proc. INTERSPEECH 2002 - ICSLP, Denver, CO, USA (2002).
5) 戸上真人他：人間共生型ロボット EMIEW2 における音源方向推定機能，日本ロボット学会誌，Vol.28, No.1 (2010).
6) Marquardt, L., et al. : A Natural Acoustic Front-end for Interactive TV in the EU-Project DICIT, 2009 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, BC, Canada (2009).

(平成 22 年 9 月 1 日受付)

大淵 康成（正会員）yasunari.obuchi.jx@hitachi.com
1990 年東京大学大学院理学系研究科修士課程修了．1992 年（株）日立製作所入社．2002 ～ 03 年カーネギーメロン大学客員研究員を経て，現在（株）日立製作所中央研究所主任研究員．博士（情報理工学）．2000 年日本音響学会技術開発賞受賞．IEEE，電子情報通信学会，日本音響学会各会員．