雑音下でも頑健に動作する音声UI技術とその応用

辻川 剛範・岡部 浩司・花沢 健

要　旨

本稿では、周囲の雑音が大きい場所でもユーザーの音声を正しく認識し、瞬時に応答するための技術とその応用について紹介します。音声UIは、手が離せないとき、目が離せないときに有用ですが、周囲の雑音が原因で誤った動作をするという問題があります。2つのマイクを用いて的確にユーザーの音声だけを検出する技術、音声のパターンを用いて音声認識に適した形で雑音を除去する技術によって、これまでは難しかった用途への適用が可能になります。また、より自然に音声UIが使えるようになります。

キーワード

●音声UI ●音声認識 ●音声検出 ●雑音除去 ●応答速度 ●自動通訳 ●キャラクター ●会話

1. まえがき

スマートフォンやタブレット端末で、音声UIを利用したアプリケーションが普及してきています。例えば、ユーザーが音声で機器に質問をすると、回答を画面に表示する、または音声で返答するといったアプリケーション[1) 2)]です。ユーザーは、手で文字を入力する手間が省けるなど、効率的に欲しい情報を得ることができます。

しかし、雑音が大きい場所で利用すると、ユーザーの音声に反応しないことや、誤認識により誤動作をすることがあります。また、雑音に反応して誤動作をしないために、音声を発する前にボタン操作を求められることがあります。これらは、音声UIの利用範囲を狭める要因やアプリケーションを使ううえでの効率を落とす要因になります。

本稿では、NECが開発してきた雑音下で頑健に音声を認識するための技術（耐雑音音声認識技術）[3) 4) 5)]について説明します。また、その技術を活用して試作したアプリケーション[6) 7)]を紹介します。

2. 耐雑音音声認識技術

音声UIを利用した基本的なシステムを**図1**に示します。マイクを通して得たユーザーの音声を、音響モデル（音の情報）、言語モデル（言葉の情報）と照合することにより、認識します。そして、認識結果に応じて応答の制御を行い、画面出力やスピーカーから音声出力を行うことにより、ユーザーに対して応答します。

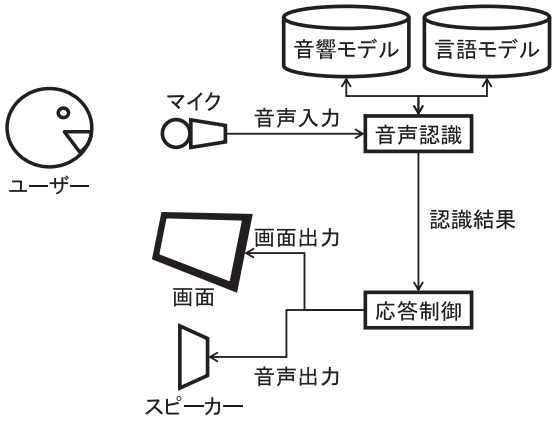

図1　音声UIを利用した基本的なシステム

雑音が大きい場所でシステムを利用する場合、ユーザーの音声とともに雑音をマイクで拾ってしまいます。雑音によるシステムの誤動作を防ぐために、ユーザーが発話している時間を検出する技術（音声検出）、混入した雑音を除去する技術（雑音除去）が用いられます。

これらの技術による効果の例を**図2**に示します。図2（a）はユーザーの音声のスペクトログラムです。（a）に雑音が混入すると（b）になります。音声検出により、システムへの入力音声（b）は（c）になります。更に雑音除去により、（c）が（d）になります。適用する音声検出、雑音除去によって、雑音の影響を軽減できる度合いが異なります。

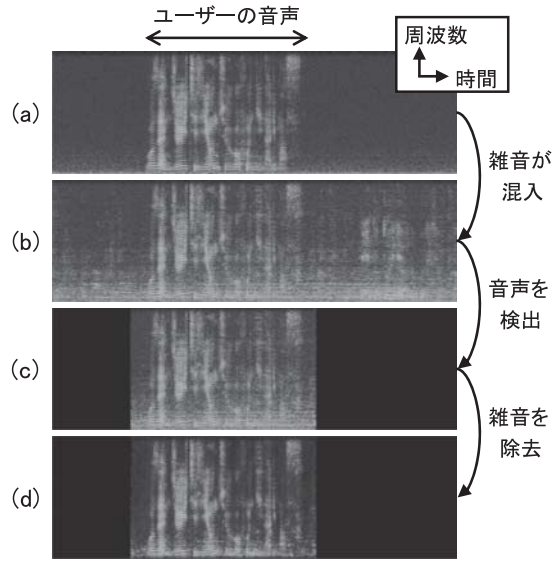

図2　音声検出と雑音除去による効果の例

2.1 2マイク音声検出

2マイク音声検出では、2つのマイクを使ってユーザーの音声と雑音を空間的に区別して、音声を検出します。ユーザー以外の人の音声が雑音に含まれる場合に特に有効です。利用シーンに応じて、以下で説明する2種類の音声検出方式を使い分けます。

(1) 位相差を利用する音声検出方式

図3（a）は、2つのマイクに到来する音の位相差を利用して、ユーザーの音声と雑音を区別する場合のマイクの配置例です。マイク1とマイク2に同時に入力される音はユーザーの音声、同時に入力されない音は雑音、と区別することができます。2つのマイクを小さいスペースに配置できる利点があります。

弊社が開発した方式[3]では、ユーザーの方向（図3（a）の例では正面方向）から到来する音を強調するフィルタの出力と、その方向から到来する音を除去するフィルタの出力との比を用います。出力の比が閾値より大きい場合には、ユーザーの音声が到来していると判定します。特長は、複素スペクトル領域と振幅スペクトル領域の2段階で、ユーザーの方向から到来する音を除去することです。これにより、ユーザーの方向が想定からずれた場合にも、頑健に音声を検出できます。

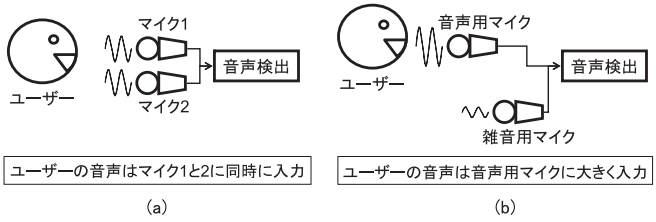

図3　2マイク音声検出におけるマイクの配置例

(2) 振幅差を利用する音声検出方式

図3（b）は2つのマイクに到来する音の振幅差を利用して、ユーザーの音声と雑音を区別する場合のマイクの配置例です。音声用マイクに大きく入力される音はユーザーの音声、雑音用マイクに大きくまたは同程度で入力される音は雑音、と区別することができます。2つのマイク配置の自由度が比較的高いという利点があります。

弊社が開発した方式[4]では、音声用マイクと雑音用マイクそれぞれの入力の比を用います。比が閾値より大きい場合には、ユーザーの音声が入力されていると判定します。特長は、周波数サブバンド単位で比を計算し、最大の比を用いて音声を検出することです。これにより、雑音の大きさがユーザーの音声と同程度の場合にも、頑健に音声を検出できます。

2.2 雑音除去

振幅スペクトル領域で雑音を除去する技術が、音声認識の前処理として効果が高く、よく用いられます。入力される雑音混じりの音声スペクトルから、雑音のスペクトルを推定・除去することにより、音声のスペクトルを得ます。

弊社が開発した方式[5]では、雑音混じりの音声スペクトルから、推定した音声と雑音のスペクトルの比を用います。比の値が小さい、すなわち雑音が大きいほど、雑音除去フィルタの値はゼロに近い値をとります。その雑音除去フィルタを雑音混じりの音声スペクトルに乗算することにより、雑音を除去します。特長は、仮推定した音声のスペクトルを事前に準備した音声のパターンを用いて補正し、雑音除去フィルタを計算することです。これにより、音声認識に適した形で雑音を除去できます。

3. 耐雑音音声認識技術を活用したアプリケーションの試作

3.1 スムーズな会話を実現する自動通訳システム [6)]

耐雑音音声認識技術、特に位相差を利用する音声検出方式（2.1の（1））を活用して、スムーズな会話を実現する自動通訳システムを試作しました。音声の到来方向に応じた対象言語の切り替えと音声検出の自動化により、発話ごとに要求されるボタン操作を省略できます。その結果、会話のスムーズさの向上が期待できます。

(1) 試作した自動通訳システムの概要

ホテルの受付ロビー、商業施設のカウンターなどで、母語の異なる2人の話者が向かい合い、自動通訳アプリケーションを搭載したタブレット端末を介して会話することを想定します。**図4** はその使用例です。お互いに音声認識結果とその翻訳結果を閲覧しながら会話します。

試作したシステムでは、タブレット端末に搭載された2つのマイクから見て、ユーザーAの方向を方向1、ユーザーBの方向を方向2として、各方向から到来する音声のみを検出します。それぞれの方向以外から到来する音を棄却することにより、雑音による誤検出を減らすことができます。更に、方向1と方向2がお互いに十分離れていれば、方向1の音声検出器は方向2からの音声を棄却できます。この特性を利用し、方向1からの音声は日本語、方向2からの音声は英語、といったように言語を識別することができます。

(2) 音声認識評価

2つのマイクを用いた音声検出と、音声の到来方向による言語識別の効果を確認するために、音声認識評価を行いました。

評価用音声は、7インチタブレット端末の形状、大きさを備えるモックを用いて収録しました。2つのマイクは3cm間隔で配置しました。マイクの正面方向を0度とし、日本語音声を再生するスピーカーを-45度方向、英語音声を再生するスピーカーを45度方向に設置しました。スピーカーからマイクまでの距離は20cm及び40cmとしました。日本語音声、英語音声ともに4名×5発声=20発声の旅行会話読み上げ音声をスピーカーから再生し、収録しました。更に、実利用場面の雑音を模擬するために、同じマイクで収録した雑音データを音声データに重畳して、評価用音声を作成しました。

図5 に音声認識評価の結果を示します。1つのマイクを用いて音声検出及び言語識別を行う従来手法に比べて、今回の手法は高い音声認識精度が得られています。従来手法の音声認識精度が低い主な原因は、雑音を誤検出しているためです。また、ユーザーが発話している時間と言語が既知の場合と比べて、今回の手法は同等の音声認識精度が得られています。この結果により、2つのマイクを用いた音声検出と音声の到来方向による言語識別の効果を確認できます。ボタン操作を省略することにより会話のスムーズさを向上しつつ、80%という実用レベルの音声認識精度を確保できています。

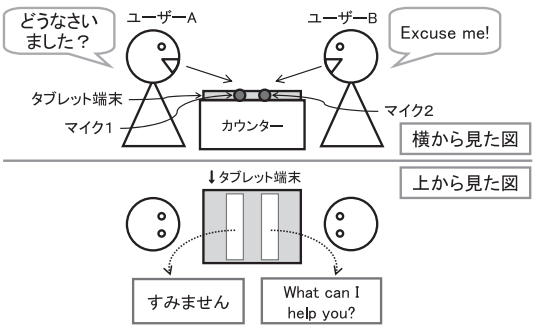

図4　自動通訳アプリケーションを搭載したタブレット端末の使用例

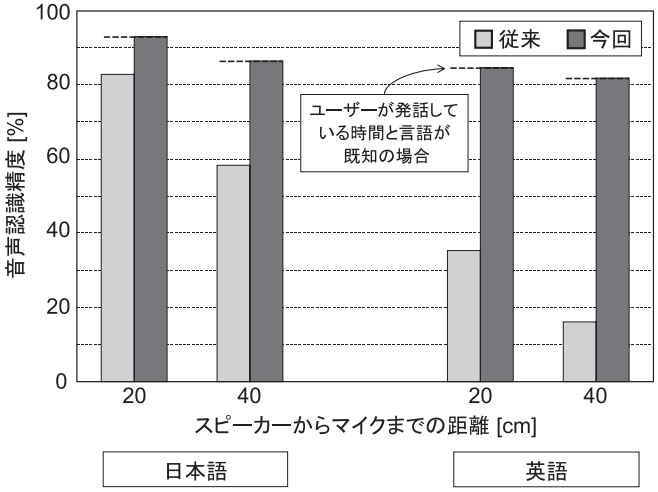

図5　音声認識評価の結果

3.2 エンターテインメント領域での音声応答の試験サービス[7)]

耐雑音音声認識技術をエンターテインメント領域へ適用することを目的として、キャラクターとの会話体験を提供する音声応答の試験サービスを実施しました。試験サービスは屋内型テーマパークであるサンリオピューロランドにて行いました。

(1) 試験サービスの概要

今回の試験サービスは、テーマパーク（遊園地）内に設置されたキャラクター「シナモロール（愛称：シナモン）」と簡単な会話体験を行うものです。シナモンの人形に話しかけると、認識した音声の内容に合わせて人形が返事をすることで、お客様に会話体験を提供します。会話体験は、あらかじめ用意された数十種類のあいさつ・質疑応答です。

今回は室内型パークであるため風雨の影響は受けませんが、常に背景音楽が存在し、また混雑した状況では周囲のお客様の話し声や足音といった雑音が大きくなるため、音声認識にとっては過酷な状況です。

設置する音声応答システムは、振幅差を利用する音声検出方式（2.1の（2））を活用します。音声用のマイクの入力と雑音用マイクの入力とをPCに2チャンネルで入力し、スピーカーから応答音声を出力します。音声の入力から認識、応答まで全ての処理は1台のPCで行います。このように実現した音声応答システムの外観を**図6**に示します。人形はアクリルの筒の中に納められており、音声用マイクを正面に、雑音用マイクを筒の上部に、スピーカーを筒の下部にそれぞれ設置します。音声用マイクの位置はお子様の顔の高さに設定してあり、雑音用マイクとは距離を置くことで耐雑音効果を出しています。なお、PCは別途バックヤードに設置します。

図6　音声応答システムの外観

(2) 試験サービスの結果

音声応答システムは、2012年4月～9月までの全営業日において稼働させました。稼働日数はおよそ140日間、システムが応答を行った回数としての応対件数は約13万6,000件です。テーマパークという性質上、平日と比較して休日の応対件数は数倍規模でした。休日の多いときで1日あたり2,000件ほどの応対件数が記録されています。休日の営業時間が10時間程度であることを考えると、平均1分間あたり3回以上応答を行っており、人気の高さがうかがえます。

システムは営業開始時に起動させ、営業終了時に終了しました。その間、スタッフの作業は不要です。スタッフは起動及び終了手順を実施すればよく、トラブルやクレームなどの発生も見られませんでした。

収集したログからサンプリングにより、応答速度と応答精度の評価を行いました。応答速度については、お客様の発声終了直後に応答音声が出力できていることから、リアルタイム処理が実現されており、お客様に対してストレスの無い会話体験を提供しました。応答精度については、約1,000発声を検聴したところ、7割程度正しく応答を返していました。応答精度は決して高くありませんが、応答が無かった場合や誤った場合でも、即座に言い直してもらうことで正しい応答が得られることが少なくなく、結果としてお客様には会話体験を提供できていることが多かったとみています。

一方、当初懸念されていた、雑音や背景音声を誤検出して誤動作する問題は、ほとんど発生しませんでした。2つのマイクを用いた音声検出方式による耐雑音効果が表れています。

最後に、お客様とスタッフのコメントを紹介します。お客様からいただいたコメントとしては、「可愛いので何回も話しかけてしまう」「本当にシナモンと会話しているみたいでびっくりした」といった好意的なものが多くみられました。スタッフからも「認識精度が高く、お客様が喜んでくれるのでとても嬉しい」という評価をいただいています。

4. むすび

本稿では、弊社が開発してきた耐雑音音声認識技術について説明しました。また、その技術を活用して試作したアプリケーションとして、スムーズな会話を実現する自動通訳システムとエンターテインメント領域での音声応答試験サービスを紹介しました。今後、耐雑音音声認識技術の更なる進展と応用範囲の拡大が期待されます。

参考文献

1) Apple「Siri」
http://www.apple.com/jp/ios/siri/
2) NTTドコモ「しゃべってコンシェル」
http://www.nttdocomo.co.jp/service/information/shabette_concier/
3) 辻川剛範：ハンズフリー音声認識のための2マイクロホンによる頑健な音声区間検出法, 2005年春季日本音響学会講演論文集, 2005.3
4) 江森正ほか：法廷音声認識システムの開発　ー複数マイクロフォンを用いた音声検出ー, 2010年春季日本音響学会講演論文集, 2010.3
5) 辻川剛範ほか：Model-Based Wiener FilterとMulti-Condition学習の併用による車内音声認識, 2008年春季日本音響学会講演論文集, 2008.3
6) 岡部浩司ほか：音声の到来方向を用いてスムーズな会話を実現する自動通訳システム, 2012年秋季日本音響学会講演論文集, 2012.9
7) 花沢健ほか：キャラクターとの会話体験を提供する音声応答の試験サービス, 第8回音声言語情報処理技術デベロッパーズフォーラム, 2012.10

執筆者プロフィール

辻川 剛範
中央研究所
情報・メディアプロセッシング研究所
主任
IEEE、電子情報通信学会、音響学会各会員

岡部 浩司
中央研究所
情報・メディアプロセッシング研究所
音響学会会員

花沢 健
中央研究所
情報・メディアプロセッシング研究所
主任研究員
情報処理学会、音響学会各会員

NEC 技報のご案内

NEC技報の論文をご覧いただきありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報(日本語)　　NEC Technical Journal(英語)

Vol.65 No.3　スマートデバイス活用ソリューション特集

スマートデバイス活用ソリューション特集によせて
スマートデバイス活用に向けたNECグループの取り組み

◇ 特集論文

サービス基盤
OSやキャリア不問のスマートデバイスの管理・セキュリティソリューション
スマートデバイスの活用を支えるソリューションと導入事例
スマートデバイスに最適な認証ソリューション
スマートデバイスの利活用に貢献する「Smart Mobile Cloud」
高品質なサービスの構築を支える「BIGLOBEクラウドホスティング」
スマートデバイス向けコンテンツ配信サービス「Contents Director」
BYODに最適なスマートデバイス活用基盤「UNIVERGEモバイルポータルサービス」
スマートデバイスの利用を促進するリモートデスクトップ・ソフトウェア
スマートデバイス対応アプリケーション開発を効率化する業務システム構築基盤「SystemDirector Enterprise」
BIGLOBEホスティングを活用したスマートフォン向けコンテンツ配信基盤サービス

スマートデバイス
Android搭載タブレット「LifeTouch」シリーズの概要
Windows 8搭載 大画面タブレットPC「VersaPro タイプVZ」
Android搭載タブレット型パネルコンピュータの開発

ソリューション
スマートデバイス対応のペーパーレス会議システム「ConforMeeting」
スマートフォンを活用したBusinessView保守業務ソリューション
UNIVERGE遠隔相談ソリューションの見守りサービスへの適用
画像認識サービス「GAZIRU」の紹介
インストア・コンシェルジュ～究極の接客ソリューション～
スマートデバイスを活用した業務システム向けテンプレートの開発
マルチデバイス対応のビデオコミュニケーションクラウドの紹介

先端技術研究
ユーザーフレンドリーなセキュリティ強化BYODソリューションに向けて
OpenFlowを活用した業務用スマートデバイスのセキュアな通信の実現
映像投影とジェスチャー入力によるインタラクション技術
雑音下でも頑健に動作する音声UI技術とその応用

◇ 普通論文

大規模災害における移動通信サービスの輻輳解決に向けた取り組み

◇ NEC Information

C&Cユーザーフォーラム&iEXPO2012
人と地球にやさしい情報社会へ ～あらゆる情報を社会の力に～
NEC講演
展示会報告

NEWS
2012年度C&C賞表彰式典開催

Vol.65 No.3
(2013年2月)

特集TOP