コミュニケーションの知識源化を実現する音声認識技術

近年，コンタクトセンタの通話分析や議会録の作成支援など，音声認識技術を活用し，これまで人が行ってきた作業を支援・代替するシーンが増えてきました．私たちは，人にもっとも馴染みやすいコミュニケーション手段である音声が，今後さらに，人，特に企業における活動の支援に大きく貢献するものと考え，音声認識技術の研究開発を進めています．本稿では，私たちNTT研究所が培ってきた音声認識技術のこれまでの発展から，これからの人の活動，企業活動における音声認識技術の貢献，役割を述べるとともに，近年注目を浴びる，感情や性別，年齢といった音声から読み取る非言語情報の活用についても紹介します．

中澤　裕一（なかざわ　ゆういち）
森　岳至（もり　たけし）
山口　義和（やまぐち　よしかず）
宮崎　昇（みやざき　のぼる）
篠原　雄介（しのはら　ゆうすけ）
NTTメディアインテリジェンス研究所

音声認識技術の発展

「Hey Siri」，「Ok Google」．これらは音声アシスタントに最初に話しかける言葉ですが，皆さんも利用したことがあるのではないでしょうか．

スマートフォンや，AI（人工知能）スピーカーに話しかけて機器の操作や，欲しい情報を教えてくれる音声アシスタントの登場により，音声認識技術が世の中に急激に普及しました．このような人とコンピュータとの対話を実現する音声認識技術は，古くは1980年代の自動音声応答装置（IVR: Interactive Voice Response），1990年代のカーナビゲーションへの導入など実用化がなされてきましたが，近年の深層学習技術の導入により音声認識精度が大幅に向上したことで，音声アシスタントや，グローバル化の流れから機械翻訳と組み合わせた音声翻訳など，さまざまなシーンで活用が始まっています．

一方，音声は人どうしのコミュニケーションにおける重要な情報伝達手段の1つです．

前述の音声アシスタントなどでは比較的短い音声を扱いますが，長い音声（長文，会話）を対象とした音声認識の実用化も検討されてきました．2000年以降，当初はニュース番組の字幕化，議会における議会録作成の支援など，いずれも手元に原稿が存在するシーンが多く，比較的明瞭な発話が対象でしたが，近年では，コールセンタでのオペレータと顧客の会話内容分析や，リアルタイムの会話支援など，人と人との自然なコミュニケーションで現れる音声が対象になってきています．

このように音声認識技術は，音声認識精度の向上とともに，対象とする音声をさらに多様なものに拡大することで発展を遂げてきました（**図1**）．

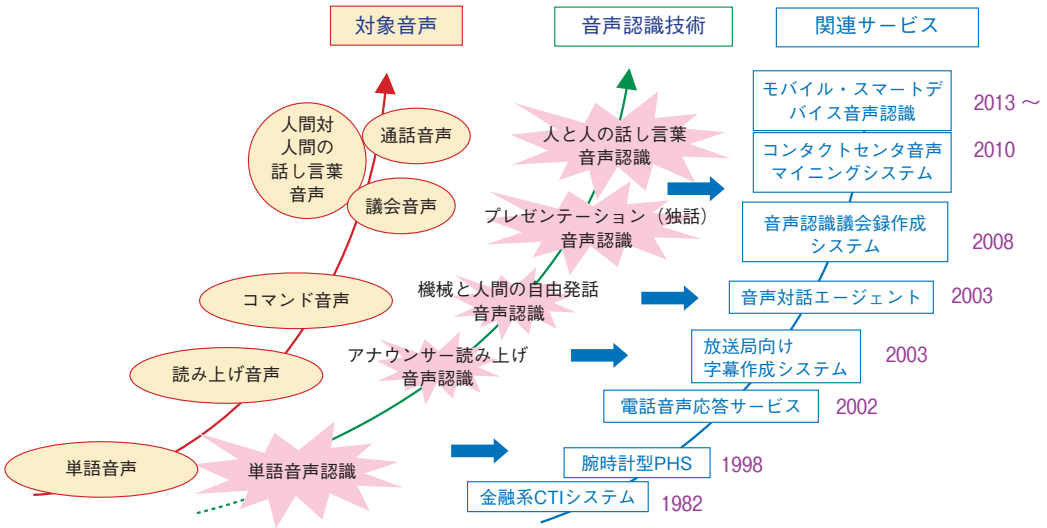

図1　NTTにおける音声認識技術の取り組み

ビジネスのDXを支える音声認識技術の役割

私たちは，音声認識技術が対象とする音声をさらに拡大することで，企業活動の変革を推進する役割を担うことができると考えています．

近年，AI（人工知能）技術を含むITを活用した業務プロセスの変革がデジタルトランスフォーメーション（DX）と呼ばれており，業種・業態を問わずその取り組みへの重要性が注目されています．DXの推進にあたっては，ITを用いた業務プロセスの合理化や自動化と，業務とITとのシームレスな連携による新たな価値創出といった取り組みが必要とされますが，業務プロセスの合理化や自動化を進める際に音声認識技術が力を発揮します．

私たちは，他者とコミュニケーションをとる際に音声を多用します．SNSやメール，チャットといったテキストによるコミュニケーションツールが充実してきたとはいえ，複雑な内容を伝えたり確認したりする場合，また複数のメンバーの合意を必要とするような意思決定においては，対面や電話によるリアルタイム音声コミュニケーションを選択する方が多いのではないでしょうか．

テキストによるコミュニケーションに比べ，音声コミュニケーションは，そのリアルタイム性や，音声のニュアンスを通じて言外に表現される情報の伝達といったメリットがあります．一方で，録音やメモを残さない限り，発声されたそばからすぐにその情報が消えてしまうという揮発性を併せ持っています．企業活動に伴って膨大な量の音声コミュニケーションが日々発生していますが，現在はコンタクトセンタにおけるお客さまとの通話のように，限られた音声コミュニケーションのみがデータ分析の対象として活かされるにとどまっており，ほとんどの音声コミュニケーショ

ンに含まれるデータは活用されずにいます．

一方で，対面接客や営業担当から顧客への電話連絡など，お客さまとの接点で生じる音声コミュニケーションには，マーケティングやコンプライアンス管理などさまざまな観点から有用な情報が含まれています．また会議やちょっとした相談のような社員間のコミュニケーションにも，新たなビジネスアイデアの種や業務改善のヒント，メンタルヘルスの傾向など，企業活動の改善に有用な情報が含まれています．

これらの情報を揮発させることなく音声認識技術によってテキスト化し，業務改善につなげるさまざまな処理の知識源とすることが，業務プロセスの合理化や自動化の推進に貢献すると考えられます．

次の章では，業務プロセスの合理化につながる音声認識技術の利用例をいくつか紹介します．

音声認識技術のユースケース

音声認識技術の研究開発は，今，これまでよりも技術的に難易度が一段高い，砕けた発話を対象としており，今後，より多くの場面でビジネスのDXを進めることが可能となっていきます．ここでは，そのような砕けた発話の音声認識精度を高めることで広がるユースケース例を紹介します．

■会議音声認識

ビジネス会議では議事録を残していることが多いと思いますが，議事録を作成した方の多くが感じているとおり，議事「メモ」ではなく議事「録」を残そうとすると予想以上に稼働がかかります．会議中に丁寧な議事メモをつくろうとすると，会議への参加や議論が手薄になってしまいますし，簡単なメモを残してそこから議事録を作成する場合は会議終了から記憶が鮮明な間に済ませてしまわないと議事を網羅的に残せているか不安になります．かといって，会議をすべて録音して後で聞きながら議事録をつくるなどということをすると，会議時間以上に時間がかかりますし，それなら議事録作成要員を１人追加して会議に参加してもらったほうがよいでしょう．早く議事録が自動でつくられるようになればいいのに．そう思ったことのある人は少なくないはずです．

これまでの音声認識では精度が不十分であったため，重要な単語が認識されていることを期待して，時間情報で対応させた音声の検索を行うくらいの用途にとどまっていました．しかし砕けた発話への音声認識を実現することで，シンプルな議事録作成の支援に加えて，要約技術と連携した議事録の自動作成，宿題事項の自動抽出による課題管理システム連携，議事進行や議論の論点整理など，人間（ファシリテータ）が担っていた役割をAIがこなしていくことが期待されます．

■遠隔作業支援

遠隔での業務や応対が今後広がっていくと考えられる医療や教育などにおいては，対面でないがゆえに発生してしまう不便さを解消していくことが求められます．遠隔機器の操作はもちろんボタンやレバーで行うことは技術的に可能ですが，医療現場において，画面越しで患者から得られる情報量が通常より少

ない医師に診療に集中してもらうためには，会話の記録はもちろん，それ以外の部分でAIによるさり気ないサポートが必要となります．例えば，「お熱を測りましょうね」に反応して体温計が患者に渡される，「口を開けてください」に反応して患者の口腔内への照明の点灯と自動消灯など．また，方言の特徴が強い地方への遠隔医療では，方言変換技術と連携させることにより，スムーズなコミュニケーションの実現が期待できます．

教育の現場の基本である1対多数の授業においては，生徒全体への呼びかけと生徒たちの反応による理解度の把握を行いますが，そのときに発生するクロストークはオンライン音声コミュニケーションでは成立しづらいことは明らかです．リアルタイム音声認識テキストによる生徒の発話内容の把握はもちろん，生徒の「はーい」に対応した挙手コマンドの実行など，生徒の集中力を遮るような機器操作を強制しないさり気ないAIは，医療現場における医師と同様に必要不可欠なものとなっていくでしょう．

■コンタクトセンタ

従来活用されてきた分野であるコンタクトセンタにおいても，これまで積極的に活用されてきたオペレータ音声の認識結果に加えて，お客さま音声の音声認識結果が充分な精度で得られるようになれば，業務支援のさらなる効率化，オペレータ業務の削減，応対通話数の増加，お客さま満足度の向上等，今後さまざまなサービスのオンライン化により高まるコンタクトセンタの需要を満たすために，音声認識技術が従来以上の貢献をすることが期待されます．

非言語情報の活用

音声コミュニケーションを通じて伝達される情報には，言語情報（テキスト情報）だけではなく非言語情報（性別，年齢など）やパラ言語情報（感情，意図，態度など）も含まれており，実業務における音声サービスの高度化に向け，非言語・パラ言語情報の積極的な活用も求められています．

私たちは，音声からテキスト情報を高精度に認識する取り組みとともに，非言語・パラ言語情報の認識・活用技術についても検討を進め，音声の非言語・パラ言語情報を抽出できるソフトウェアエンジンRexSense®を開発しました．本ソフトウェアエンジンにより，①話者属性（成人男性・成人女性・子供），②感情（喜・怒・哀・平静），③疑問・非疑問，④緊急度，を音声データから高精度に認識・推定することが可能です．また，コンタクトセンタ高度化などの活用に向け，本エンジンと音声認識と統合したWeb API（Application Programming Interface）サービスを実現できるRexSense®システムを開発しました．

RexSense®を活用することにより，例えば人の感情に応じてロボットが適切な反応やレコメンドを返すといった高度な応対サービスの提供や，音声から判別した話者属性等の非言語情報に基づき，より適切なコンテンツ（案内，広告等）を提示する高度なデジタルサイネージなどの実現が可能となります．

また，コンタクトセンタにおける高度な

VoC（Voice of Customer）分析の実現やIVRにおける自動応答サービスの高度化，将来的には非言語・パラ言語情報を活用したより高度な音声会議ソリューションの実現も期待できます（**図 2**）．

その他，コンタクトセンタにおけるオペレータとお客さまとの通話音声から，お客さまの声の特徴やさまざまな会話の特徴を分析し，お客さまの満足感情（満足・不満）を抽出する顧客満足度推定技術を開発し，コンタクトセンタAIソリューション「ForeSight Voice Mining®」に導入，2019年4月よりサービス提供を開始しました．また，これに加え，オペレータの応対の好感度を評価する応対好感度推定技術を開発，サービス化に向け検討を進めています．

これらの技術を活用することで，例えば通話分析（オペレータ応対の優良事例の検索や顧客満足度の分析等）やオペレータ支援，オペレータやコンタクトセンタの評価，オペレータ教育などへの応用が期待されます．

今後の展望

これまで紹介してきた音声認識技術は，適用領域をビジネスシーンからさらに拡大し，あらゆる音声コミュニケーションを対象とすることで，NTTグループが進めるIOWN（Innovative Optical and Wireless Network）構想の1つであるDTC(Digital Twin Computing)[(1)]において，ヒトDTCを実現するための必須技術となります．

DTCのアーキテクチャ（**図 3**）におけるサイバー・フィジカルインタラクション層では，実空間のモノやヒトのセンシングにより，

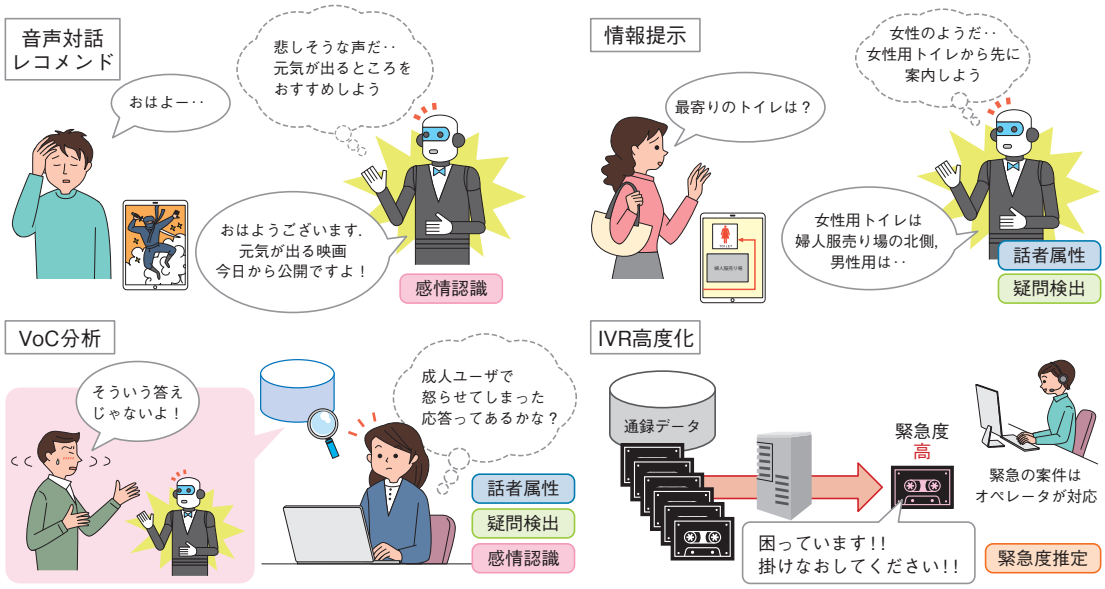

図 2　Rexsense® 応用例

図3　デジタルツインコンピューティングアーキテクチャ

デジタルツインの生成に必要なデータを収集する必要があり，ヒトの思考をセンシングするうえで音声認識技術は重要な役割を担うこととなります．

企業活動のDXやヒトDTCの実現が進むことで，社会はより便利に，豊かに，安全にと変容していきます．私たちは，ヒトとヒトとのコミュニケーションを対象とする音声認識技術の研究開発によって，このような社会の実現に貢献していきます．

■参考文献

⑴　戸嶋・小橋川・能登・倉橋・廣田・小澤：“ヒトDTCの挑戦と今後の展望，”NTT技術ジャーナル，Vol. 32, No. 7, pp. 12-17, 2020.

（左上から）宮崎　昇／中澤　裕一／森　岳至
（左下から）山口　義和／篠原　雄介

今後あらゆる分野で進むＤＸを支えるため，新たな価値の提供をめざし研究開発に取り組んでいます．一方でNTTの音声認識技術は活用が進み，手軽に利用できるAPI環境も整備されています．ぜひ，「NTT　音声認識」と検索していただき，新しいアイデアの検討にご活用ください．

◆問い合わせ先
NTTメディアインテリジェンス研究所
心理情報処理プロジェクト
E-mail　noboru.miyazaki.mt@hco.ntt.co.jp