

山梨中央銀行は、大学等の研究機関が保有する技術シーズと企業ニーズを結びつけ、新技術の開発や新規事業の創出を支援するリエゾン（橋渡し）活動に取り組んでいます。

本リポートが、中小企業の皆さまが抱える経営課題の解決や新産業創出の“ヒント”となり、ビジネスチャンスに繋がればと考えております。

<第90回>



研究室の西崎先生

深層学習（ディープラーニング） を用いたマルチメディア処理

～音声認識・音声対話・スマート
農業の応用研究～

西崎 博光 先生

（工学域 機械工学系 准教授）

■ 研究の概要について教えてください。

人工知能技術の急速な発展を担う深層学習（ディープラーニング）を活用して、人間が持つ五感のうちの視覚・聴覚において、人間が持つ認知的な作業をする能力などを代替する技術への応用や、基礎的な研究を行っています。

この深層学習を使ったマルチメディア情報処理が研究テーマで、主に音声認識、音声対話などの音声インターフェースやスマート農業への応用などの研究に取り組んでいます。

■ 研究テーマである「深層学習（ディープラーニング）を用いたマルチメディア処理」とは具体的にはどのような技術ですか。

最近 AI（人工知能）という言葉をよく聞くようになりました。コンピュータが将棋や囲碁でプロの人間に勝ったり、テレビゲームや東大入試へ挑戦したり、また、車の自動運転を可能にしたのがAIですが、その人工知能技術の一つが深層学習（ディープラーニング）です。

深層学習は、ニューラルネットワーク（人間の脳の仕組み）をコンピュータ上に実現した機械学習の一つです。これまでの機械学習より優れた技術で、一部の作業では、すでに人間を超えたレベルまでに達成していると言われています。

また、マルチメディア処理とは、コンピュータなどの情報機器を用いて、デジタル化された文字や画像、動画、音声など、様々な種類・形式の情報を組み合わせて複合的に扱うことができる処理方法です。具体的にはリアルタイムで動作する音声認識技術や音声認識を使った音声対話技術などで、接客訓練音声対話システム等の開発やスマート農業などにも応用されています。

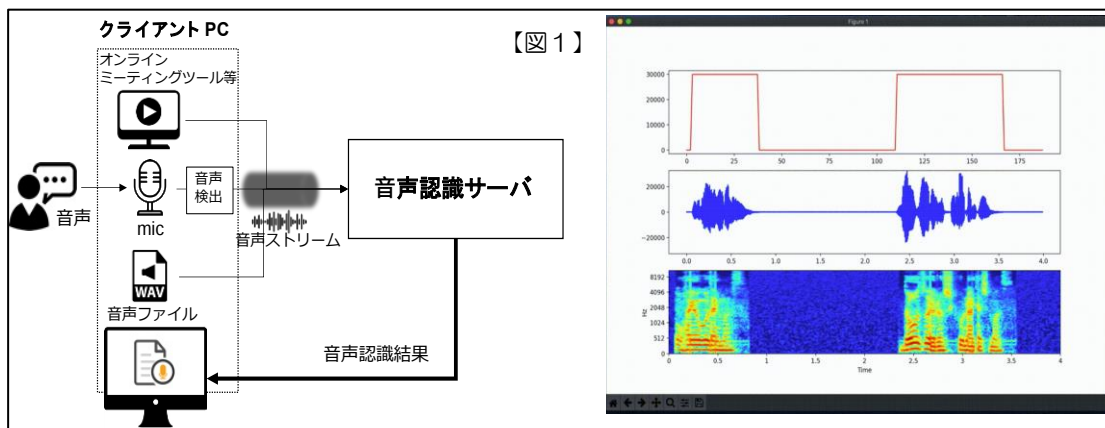
■リアルタイム音声認識とはどのような技術でしょうか。

人の音声(会話)をコンピュータに認識させることで、「話し言葉をテキストデータに変換」あるいは「音声の特徴を捉えて声を出している人を識別」などを行う機能で、それを瞬時(リアルタイム)で動作することができる音声認識の技術です。(図1)

人の声を聴いて文字を起こすことや字幕、自動翻訳などリアルタイム音声認識の技術としてすでにスマホなどで実用化され、Google などから提供されています。論文ベースですが、正確さはすでに90%程度に達していると言われています。

しかしながら、自由に機能をカスタマイズできないことやインターネット接続が必要などの制約があり、利用しにくい場合も多く、また特定の企業では、専門用語が飛び交う会議で使う場合や方言など、語彙を登録しなければならないなどの課題もあります。

そのため、利用者がカスタマイズできる低遅延のリアルタイム音声認識システムを研究開発し、その手法を世界中に公開しています。



※リアルタイム音声認識の仕組み

■接客訓練音声対話システムとはどのようなシステムでしょうか。

音声認識の技術を使った「接客訓練音声対話システム」は、仮想空間を活用してひとりで接客の訓練が可能なシステムで、筑波大学との共同研究により開発いたしました。

VR(バーチャルリアリティ)技術を使って仮想空間にエージェント(※1)を配置し、そのアバター(※2)との対話を通じて接客した内容をAIが自動評価・判定することで、接客訓練の環境を実現しました。

このシステムは、お客様がアバターになり、お客様がクレームを言ったら訓練者がどのように対応したかを自動的に数値で判定します。その仕組みは、訓練者の声、感情、敬語等を総合的にAIが評価し、接客対応結果を点数化して一目でわかるように表示します。(図2) 訓練にはHMD(ヘッドマウントディスプレイ)も使うことができます。

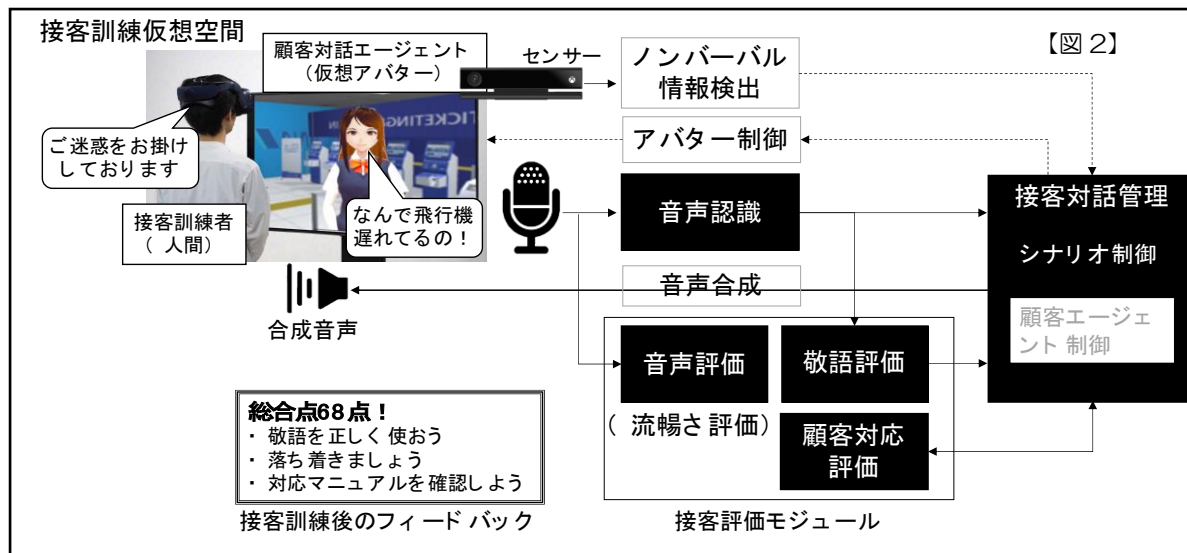
接客は日本においては、特に重要なテーマです。訓練の方法としては、マニュアル、OJT、指導員による研修など様々ですが、どれも相当なコストと時間がかかります。

一方、このシステムは訓練者がひとりで、まるで現場にいるかのような臨場感で実践しながらの訓練ができることから、人材の早期育成を短時間・低コストで行うことが可能です。そのため、新人研修やコンビニのアルバイト向け研修等のニーズが高いです。

訓練結果をより正確に評価するために、さらに音声認識率を上げる必要があります。

※1 エージェント：コンピュータのユーザーが連続した操作をしなくても、自律的に情報収集や状況判断を行い、適切な処理動作を実行する機能またはそのソフトウェア

※2 アバター：ゲームやネットの中でユーザーの「分身」を表すキャラクターの名称



※接客訓練音声対話システム

■ スマート農業への取組とはどのようなことですか

ブドウの色づきを推定し、収穫に適しているかどうかを判定する技術を開発しました。

ブドウの色は最適な収穫時期を判断するのに特に重要で、良い色で収穫すると良い品質のブドウが出荷できるようです。色は個人によって見え方が違うので、それを統一的に判断するために人工知能技術、深層学習を使って色の判断を可能にしています。

フィールドテストを実施した結果、人間が判断した色と深層学習が判断した色で、だいたい相関が0.7程度の推定精度が出ています。(※相関が1の場合、人間と深層学習の判断が同じ)。



この技術は、箱詰めするときにも利用が可能で、品質が高いブドウとして出荷できるようになりました。

また、この技術は、長年の経験が不要なブドウの摘粒（実の間引き）技術とともにスマートグラスとして開発され、実用化に向けた実証実験が行われています。



※本技術は農研機構「スマート農業実証プロジェクト(ローカル5G)」プロジェクトにおける匠の技による高品質シャインマスカット生産実装コンソーシアムで開発した技術です

※ブドウの色見本

■ 他の技術についてはどうですか。

ベテランの技術者が持つ技能・ノウハウ・暗黙知を後世に伝える「技術伝承」が問題となっており、技能・ノウハウ・暗黙知をマニュアルとして残しておきたいというニーズが企業にあります。このため、ある企業さまと共同開発したのが「技術伝承システム」です。

このシステムの特徴は、技術者の目線と解説音声を収録してマニュアルの作成支援を行うことができます。具体的には、カメラやゴーグルをつけたまま作業しているところが動画で記録され、技術者が話した作業のポイントは、わざわざメモしなくても自動的に文書化されるなど、動画や文字で技術を正確に伝えることができる画期的なシステムです。

通常作業後などにマニュアルの作成時における、作業説明の抜け落ちを防止することにも役立ちます。

このシステムは、製造業に限らず農業や医療など多くの産業・業種において技術保存や人材育成にも活用できるようです。

■研究に対する課題はなんですか

音声認識については、様々な課題があります。具体的には、うるさい環境の中で、どうやって声だけを認識したらいいのか（雑音処理）、高齢者の場合、ろれつが回らないことや発声器官の衰退などにより認識率が低下することなどへの対策、また、ドアのノック音や水道が流れる音など生活の中で発生する環境音認識などのシーン検出、周囲の音を認識することで全体の状況を把握できる技術の確立など、その課題解決には音声認識率をどのようにして高めていくかが大きなポイントとなっています。

また、スマート農業においては、気象状況や光の影響への対応などが課題となっています。

■地域（企業）との連携（共同研究等）の可能性についてお聞かせください。

音声認識率の向上が課題の接客訓練音声システムは、研究室内での評価にとどまっており実証が十分でないことから、実証実験等に協力してくれる企業などを探しています。

本テーマである音声認識技術やスマート農業など「深層学習（ディープラーニング）」を用いたマルチメディア処理」に興味のある方は、ご連絡をお待ちしております。

【トピック】

西崎博光メカトロニクス学科准教授の研究チームがIEEE GCCE2020においてExcellent Paper Award, Outstanding Prizeを受賞しました。

詳しくは、<https://www.yamanashi.ac.jp/28883> を御覧ください。

（取材～地域連携コーディネータ 内藤）

山梨大学との共同研究、技術的な相談や指導のご要望は

山梨中央銀行コンサルティング営業部 地方創生推進室

TEL: 055-224-1091 まで、お気軽にご連絡・ご相談ください。