



研究基盤としてのアンドロイド「ERICA (エリカ)」を開発 ～自然な対話が可能な自律対話型アンドロイドの実現に向けて～

ポイント

- 人間と自然に対話するアンドロイド「ERICA (エリカ)」を新たに開発した。
- 見た目は美人顔の特徴を参考にコンピューターで合成され、振る舞いは現時点で最先端の音声認識、音声合成、動作認識、動作生成の技術を統合して作られている。
- 「ERICA」を、「違和感のない自然な対話」を追求する研究プラットフォーム（研究基盤）とすることで、自律対話型アンドロイドの実現が期待される。

JST 戦略的創造研究推進事業において、ERATO石黒共生ヒューマンロボットインタラクションプロジェクトの石黒 浩 研究総括（大阪大学 大学院基礎工学研究科 教授、株式会社国際電気通信基礎技術研究所 石黒浩特別研究所 所長・ATRフェロー）、河原 達也（京都大学 大学院情報学研究科 教授）らは、研究プラットフォーム^{注1)}として人間に酷似したロボットのアンドロイド「ERICA (エリカ)」を開発しました。

従来のアンドロイドは実在の人物をモデルにしてきましたが、「ERICA」の見た目は、美人顔の特徴を参考にコンピューターグラフィックスで合成しています。これまでの研究から、人間は時として人間よりもロボットとの対話を好むことが知られています。姿形を合成することで、ロボットとしての親しみやすさと人間としての存在感を併せ持たせ、見た目において人との自然な対話を実現するための新たな試みを行いました。

また、対話の際の自然な振る舞いを実現するため、「ERICA」には、プロジェクトの参加者がこれまでに開発してきた音声認識技術、音声からの動作生成技術、ロボット制御技術などの要素技術が調和して動くよう統合した上で搭載されています。研究室の来客と対話し自己紹介をするという限られた状況において、相手の動作や音声を認識し、人間らしい振る舞いを伴った対話が可能となりました。音声も現時点で最も完成度の高い技術をもとに作られ、合成音ながらほとんど人間と区別がつかないレベルになっています。

動きを生み出す空気圧アクチュエーター^{注2)}は、頭部を中心として19本が埋め込まれ、2年後を目処に30本まで増やし、腕も含めた上半身が人間らしく動くように改良します。制御回路にはほぼ無音の新しい回路を用いています。

プロジェクトでは、この「ERICA」を研究プラットフォームとして用い、見た目と振る舞いを統合的に進化させることで、日常生活で活躍する自律対話型アンドロイド^{注3)}の実現を目指します。

本成果は、以下の事業・研究領域・研究課題によって得られました。

戦略的創造研究推進事業 総括実施型研究 (ERATO)

研究プロジェクト: 「石黒共生ヒューマンロボットインタラクションプロジェクト」

研究総括: 石黒 浩 (大阪大学 大学院基礎工学研究科 教授、株式会社国際電気通信基礎技術研究所 石黒浩特別研究所 所長・ATRフェロー)

研究期間: 平成26年7月～平成32年3月

上記研究課題では、特定の状況と目的において自律的に対話できる機能、複数の情報伝達手段を用いて社会的状況で複数の人間と対話できる機能等、実社会において人間と親和的に関わり、人間と共生するための自律型ロボットの実現を目指しています。

<研究の背景と経緯>

人型ロボットは日常生活の場で活躍することが期待されており、人間と関わる技術の開発が特に重要です。人型ロボットが人間との対話を行う際には、音声認識や発話などの言語的要素と、視線の動きやしぐさ、顔の表情などの非言語的要素の両面において、違和感がない応答が求められます。しかし、音声認識、発話、反動的動作、見た目など個々の要素技術を追求し、寄せ集めたとしても、そのロボットが違和感なく対話を行えるとは限りません。また、従来の人型ロボットで自然な動作生成技術を研究しても、見た目が人間に酷似したアンドロイドには適用できない可能性があります。特定の技術を改善して自然な対話を追求する研究は進められてきましたが、すべての要素技術を不調和を起こすことなく統合し、自然な対話を実現する技術は研究されてきませんでした。

プロジェクトは、対話に適した見た目を持つアンドロイドに、さまざまな要素技術を組み合わせることで、新たな研究プラットフォームを開発することが必要だと考えました。これにより、話者の動作を認識することによって相手の発話内容のある程度絞り込むことで、音声言語理解の精度を高める、音声言語理解に失敗した場合でも視線や相づちといったしぐさにより対話を継続するといった具合に、複数の要素技術を統合し、それぞれが対話を継続させるように進化する研究が可能になるからです。

<研究の内容>

プロジェクトでは、自律対話型アンドロイドの実現に向けた統合的な技術開発の研究プラットフォームとして、アンドロイド「ERICA（エリカ）」を新たに開発しました（図1）。「ERICA」は、対話に適した人間らしい見た目を持ったアンドロイドです。これに音声認識や動作生成などのさまざまな要素技術を組み合わせることにより、人間と自然な対話を行うための統合的な技術開発を可能にする研究プラットフォームとなっています。

対話に適した見た目を実現するため、「ERICA」では姿形をコンピューターグラフィックス（CG）で合成することを試みました。これはアンドロイドとしては新しい特徴です。従来のアンドロイドは実在の人物をモデルにしてきましたが、人間は時として人間よりもロボットにより親しみを感じるということが知られています。そこで、ロボットの姿形を人工的に作り出すことで、人間としての存在感とロボット特有の親しみやすさを併せ持つ魅力的な容姿を実現できると考えました。鼻と口とあごが一直線上に並ぶというビーナスラインの法則など、多くの美人顔に見られる共通的な特徴を持たせつつ、全体としてバランスの取れた顔を合成しています。顔の左右は対称で、日本人とヨーロッパ人のハーフとして見られるように、彫りの深さや鼻の大きさ、高さを調整しました。人間との対話に適した姿形に加え、CGであるため肖像権を持たない点も、人間と関わる技術の研究プラットフォームに適しています。

機能面については、大阪大学とATRで開発されたアンドロイド制御システム、音声に基づく動作生成システム、マイクロフォンアレイ技術^{注4)}、京都大学で開発された音声認識システム、対話生成システムが実装されています。音声に基づく動作生成システムは、ATR独自の技術で、「ERICA」の発声から唇の動きや頭部の動きを自動的に再現することで、発話と一致した自然なしぐさを生み出すことができます。音声認識技術は京都大学が開発したシステムで、「ERICA」への実装にあたり、「Julius（ジュリアス）」

と呼ばれる日本で最も利用されているオープンソース音声認識エンジン^{注5)}にディープラーニング（深層学習）^{注6)}と呼ばれる最新の技術を導入しています。これにより、多様な発音の音声を認識することが可能になりました。

音声についても現時点で最も完成度の高い技術をもとに作られており、合成音声ながらほとんど人間の声と区別がつかないレベルになっています。音声合成はHOYAサービス株式会社の協力を得て、同社のVoiceTextをベースに作成しました。「ERICA」のイメージに合う声優の音声を20時間以上収録し、録音した声を音素に分解し、再合成することで、人間らしい音声を作り出しています。

さらに、体には空気圧アクチュエーターが頭部を中心として19本埋め込まれており、頭部の滑らかな動きを実現しています。2年後を目処に30本まで増やし、腕も含めた上半身が人間らしく動くように改良します。これらの空気圧アクチュエーターの制御に用いる回路（サーボ回路）もほぼ無音の新しい回路で、大きさも従来のアンドロイドに用いてきた回路の半分以下のサイズです。今後アクチュエーターの数を増やしてもアンドロイドの外見に影響を与えることなく、体内に埋め込むことができる見通しです。

現在搭載されている機能により、「ERICA」は研究室の来客と対話し、自己紹介をするという限定された状況において、人間の動作や音声を認識し、人間と対話することができます。適切な受け答えという発話内容に加え、発話者を見る、相手の発話リズムに合わせて相づちを打つ、自分の発話リズムに合わせて頭や表情を動かすなど、音声と動作が結びついた自然な対話が可能です。

＜今後の展開＞

プロジェクトでは、アンドロイド研究で世界を先導する大阪大学・ATRと、国会（衆議院）の会議録作成に採用されるなど日本語で最高水準の音声認識システムを開発している京都大学との連携により、自律的対話能力を持つアンドロイドの実現をさらに進めます。

音声のみの対話では、人間からの問いに対しシステムが的外れな応答をすれば、人間は対話を諦めてしまうかもしれません。しかし、ロボットが視線やしぐさなどで対話を続けたいという「意図」を示すことにより、人間も対話を続ける気になります。このように音声認識など言語コミュニケーション技術と視線や振る舞いなどの非言語コミュニケーション技術を同時に進化させることで、より自然な対話が可能になります。今後、研究プラットフォーム「ERICA」を使用した研究により、要素技術を統合的に進化させ、より多様な状況で、違和感のない自然な対話を実現することが期待されます。

<参考図>



図1 アンドロイド「ERICA (エリカ)」の外観

<ロボットの仕様>

アンドロイド「ERICA (エリカ)」

身長：166センチメートル（立位時）

自由度：能動関節は19カ所、受動関節は30カ所

アクチュエーター：空気圧アクチュエーター

センサー：CMOSカメラ2個（左右の眼球に1個ずつ）、マイクロフォン2個（左右外耳に1個ずつ）

皮膚素材：シリコン樹脂

頭髪素材：人毛髪

CPU：外部PC（Windows、Linuxなど）による制御

電源：空気圧レギュレーター用に外部電源（AC24V）

周辺機器：エアーコンプレッサー（AC100V）、制御用PC多数

人追跡システム

複数の2次元レーザー距離センサーにより人の位置を計測するシステム（時間分解能33ミリ秒、空間分解能10センチメートル）

発話者追跡システム

複数の16チャンネルマイクロフォンアレイと人追跡システムの連携により発話者の発話区間を検出するシステム（時間分解能100ミリ秒、音源の3次元位置分解能20センチメートル）

音声認識

大語彙連続音声認識エンジン Julius

音声合成

VoiceText

<用語解説>

注1) 研究プラットフォーム

技術を研究・開発するための基盤（土台）。

注2) 空気圧アクチュエーター

圧縮空気を利用した駆動装置。アンドロイドの体の部位を動かすために用いられる。

注3) 自律対話型アンドロイド

操作者なしで人間と対話する機能を備えた人間酷似型ロボット（アンドロイド）。対話する機能には発話内容が適切であるという言語的側面に加え、視線方向、表情や振る舞いといった非言語的側面も含まれる。

注4) マイクロフォンアレイ技術

複数のマイクロフォンを備えた機器により、音声分離や雑音除去などを行う技術。

注5) オープンソース音声認識エンジン

コンピューターで音声を認識する機構（エンジン）のうち、プログラム（ソースコード）が公開されているもの。

注6) ディープラーニング（深層学習）

機械を人間のように学習させる技術の1つで、入力層、出力層と多数の中間層からなる深層構造を備えた人工神経回路網（ニューラルネットワーク）を利用するもの。従来のニューラルネットワークと比べ、より抽象的な概念の学習や特徴抽出が可能。