

News Release

2022.4.21

NEDO(国立研究開発法人新エネルギー・産業技術総合開発機構)

株式会社ヒューマンテクノシステム

国立大学法人東北大学

障害のある話者の声で明瞭な発音を生成する技術を開発 —発話障害者の個人音声合成器の活用を後押しし、生活の質(QOL)を向上—

NEDOが福祉用具の実用化開発を目的に推進する「課題解決型福祉用具実用化開発支援事業」で、このたび(株)ヒューマンテクノシステムと東北大学は、発話障害者の声で明瞭な音声を合成する技術の開発に成功しました。この技術は、音を作る器官の問題で発音が不明瞭になってしまう障害のある人の声と健常者の声を学習し、前者の声色で後者のように明瞭な音声を生成するものです。

これにより、構音障害のために音声合成器の製作・活用をためらっていた障害のある話者に対し、個人音声合成器による明瞭な音声の生成が可能になり、自分の声で話すことによる生活の質(QOL)向上を実現します。

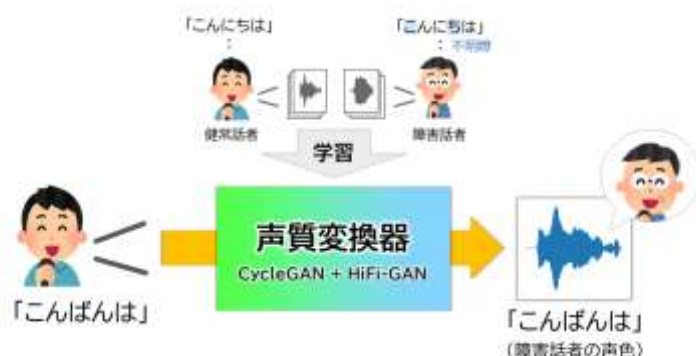


図1 音声変換のイメージ

1. 概要

喉頭がんや筋委縮性側索硬化症(ALS)などのため声を失った人の生活の質(Quality of Life: QOL)を向上する手段として、あらかじめ録音した自分の音声データを使って任意のテキストを読み上げる「個人音声合成器」が注目されています。しかし、現在の個人音声合成器は録音した音声の特徴を忠実に再現するため、構音障害^{※1}などで発音が不明瞭な場合、合成器の音声も不明瞭になってしまう課題がありました。

この課題の解決を目指し、株式会社ヒューマンテクノシステムと国立大学法人東北大学はNEDO(国立研究開発法人新エネルギー・産業技術総合開発機構)の「課題解決型福祉用具実用化開発支援事業^{※2}」の下、明瞭な声で自分らしく話すことができる発話障害者向け個人音声合成器の開発に関する共同研究を行ってきました。このたび東北大学が開発した声質変換技術^{※3}を基に、その機能を拡張した上で健常者の音声を変換することによって、その人らしい声を保持しながら明瞭な音声を生成する声質変換技術の開発に成功しました。

なお(株)ヒューマンテクノシステムは、この成果の一部を2022年3月8日に開催された第116回電子情報通信学会・福祉情報工学研究会(WIT)で発表しました。

2. 今回の成果

(1)本人の声色かつ明瞭な音声をつくる技術の開発に成功

従来の敵対的生成ネットワーク(GAN)^{※4}の一種であるCycleGAN^{※5}を利用した声質変換技術は、ノンパラレルデータによる学習^{※6}が可能です。(株)ヒューマンテクノシステム、東北大学はこの技術の音韻や韻律を保ったまま声色を変換できる性質に着目し、健常者の音声を構音障害者の声色へと変換することにより本人の声色による明瞭な音声の生成を実現しました。

(2)学習データの前処理を導入し、さらに高い品質の音声変換を実現

学習する構音障害者の音声は健常者の音声と比べて話速が非常に遅い・抑揚に乏しいといった音響的な特徴があり、健常者同士の声質変換を実現する従来の方法では学習がうまく進まず品質の劣化を招くという問題がありました。そこで学習データに話速や声の高さを調整する前処理を導入することで、より高い品質の音声を生成することが可能になりました。

(3)限られた音声でも高品質な音声を生成

従来大量データによる学習が必要だった音声生成部をマルチ話者ボコーダ^{※7}「HiFi-GAN」の学習済みモデルで置き換えることにより、限られた量しか入手できない構音障害者の音声でも高品質な音声を生成することが可能となりました。

これにより、これまで構音障害などによる不明瞭な発音のために個人音声合成器での音声の製作・活用をためらっていたケースへも、その人らしさを保ったまま明瞭な声を生成する個人音声合成器を提供することが期待できます。

なお、(株)ヒューマンテクノシステムが昨年11月に開催された「第48回 国際福祉機器展H.C.R.2021」のNEDOブースでの発表の中で自分の声ソフト「ボイスター」(自社開発)での健常話者の実際の収録音源と合成音の比較聴取ができます。詳細については下記をご参照ください。

<https://www.youtube.com/watch?v=FdhIGN7CVdc>

[本研究の実施体制]

(株)ヒューマンテクノシステム 担当分野:声質変換技術の障害音声への適用
代表研究者:自分の声ソフトウェア担当 渡辺 聡

東北大学 担当分野:声質変換技術の開発
代表研究者:大学院工学研究科 通信工学専攻 伊藤・能勢研究室 准教授 能勢 隆

3. 今後の予定

本事業において複数拠点での実証実験を実施し、性能の向上とALSや脳性麻痺を含むさまざまな症状の障害音声や収録環境への対応を進めます。これを個人音声合成器と組み合わせ、発話障害者が明瞭な声で自分らしく話せるシステムを実現し、2023年度の実用化を目指します。



図 2 個人音声合成器への適用例

【注釈】

※1 構音障害

口唇・舌・口蓋や脳機能などの障害により、話しことばを正確・明瞭に発音できない状態を指します。(大辞林「構音障害」より)

※2 課題解決型福祉用具実用化開発支援事業

研究開発項目: 明瞭な声で自分らしく話せる発話障害者向け個人音声合成器の開発

事業期間: 最長2～3年

事業予算: 63百万円(2021年度総事業額)

事業・プロジェクト概要: https://www.nedo.go.jp/activities/EP_00375.html

※3 声質変換技術

ある人の声を別の人のように変換する技術です。双方の音声データを機械学習することで変換モデルを作成し、これを用いて変換を行います。

※4 敵対的生成ネットワーク(GAN)

敵対的生成ネットワーク(GAN: Generative Adversarial Networks)は、画像の分野で発展したニューラルネットワークで、たとえば写真から絵画を生成するようなタスクで成果を上げています。生成器と識別器の二つのニューラルネットワークが互いにだまし合うようにして学習を進めることで、本物と見間違ふような画像を生成できるようになります。音声の分野においても音響特徴量を画像と見立てることでGANの活用が進んでいます。

※5 CycleGAN

画像分野において、2種類のデータセットを学習することでスタイル変換を実現するニューラルネットワークです。写真を仏画家・モネの絵画風に変換、男性の顔を女性の顔に変換といった用例が報告されています。

※6 ノンパラレルデータによる学習

学習データにおいて、変換元のデータとそれに対応した変換後の正解データを与える形をとる学習をパラレルデータによる学習と呼びます。声質変換においては変換後の音声を入力することが困難であるため、こういった対応関係を必要としないノンパラレルデータによる学習が可能なモデルが使用されます。

※7 マルチ話者ボコーダ

ボコーダとはスペクトログラムなど、音声の特徴量から音声波形を生成する手段のことです。深層学習を使ったボコーダは、通常、特定話者の音声波形だけを生成できるよう構成しますが、これを任意の話者の音声波形を生成できるように構成したものをマルチ話者ボコーダと呼びます。

4. 問い合わせ先

(本ニュースリリースの内容についての問い合わせ先)

NEDO イノベーション推進部 担当:只野、白井 TEL:044-520-5175

(株)ヒューマンテクノシステム プロダクツ事業部 担当:渡辺 TEL:03-5847-1206

東北大学大学院工学研究科 情報広報室 担当:沼澤

TEL:022-795-5898 E-mail:eng-pr@grp.tohoku.ac.jp

(その他NEDO事業についての一般的な問い合わせ先)

NEDO 広報部 担当:坂本、橋本、鈴木、根本

TEL:044-520-5151 E-mail:nedo_press@ml.nedo.go.jp

※新聞、TV などでは弊機構の名称をご紹介いただく際は、“NEDO(国立研究開発法人新エネルギー・産業技術総合開発機構)”または“NEDO”のご使用をお願いいたします。