【研究区分：先端的研究】

研究テーマ：構音障がい者の発声を訂正するための声質変換研究	
研究代表者：地域創生学部 地域創生学科 地域産業コース 准教授 陳金輝	連絡先：kinki@pu-hiroshima.ac.jp
共同研究者：神戸大学都市安全研究センター 大学院システム情報学研究科 教授 滝口哲也	

【研究概要】
構音障がい者の人々は一人一人その障がい特性が多様であり，コミュニケーションが困難な場合がある。そのため構音障がい者のコミュニケーションへの支援が可能，汎用性高い AI 技術が有意義である。しかし，この技術の開発については十分検討してきたわけではなく，このことは他の関連文献においても同様である。本研究では，発話コミュニケーション障がい者独自の課題に対しディープラーニングの１種である敵対生成ニューラルネットワークを用いた障がい者の発声の声質改善技術を提案し，障がい者の身体能力の壁を超越する事が出来るコミュニケーション支援技術の実現を目指す。研究期間内，主な成果として海外専門誌 Signal Image and Video Processing に研究論文を１編発表した。

【研究内容・成果】

1. 研究背景・研究内容

近年，家庭生活，学校生活，社会生活において様々な機器の情報化が進み，情報機器が身の回りの生活環境にて浸透しつつある。しかし，そのような機器は操作が複雑であり，障がい者が自立して使いこなすには困難である場合が多い。平成 25 年 6 月には，障害を理由とする差別の解消の推進に関する法律が公布され，全ての国民が，障がいの有無によって分け隔てられることなく，共生社会の実現に資することを目指すものとなってきた。このような状況を踏まえ，本課題では，障がい者の自立生活を AI 技術で支援する新しいユニバーサルコミュニケーション技術を提案し，実証実験まで行った。

2. 提案手法の概要

本研究では構音障がい者の発話内容（話者特徴，テキスト）に対する分解，再構成，音声合成の統合による新しい声質変換手法を提案する。提案手法が機械学習を基盤技術にし，学習フレームワークは敵対生成ニューラルネットワーク(GAN: Generative Adversarial Network, NIPS 2014)に基づいて開発したものである。

GAN はゲーム理論を元に触発されたものであり，生成器と識別器は互いにナッシュ均衡を達成しようとする。図 1 に示すように，生成器 G は健常者声データ A の潜在的分布に合わせて入力から合成音声データを生成するように学習される。一方識別器 D は健常者声データと合成音声データを正しく区別するように学習される。生成器 G の入力はランダムノイズベクトル x（通常，一様分布か正規分布）である。ノイズは生成器 G を介して新しいデータ空間にマッピングされ多次元データ（ベクトル）である音声サンプル $G(A)$ を得る。そして識別器 D はバイナリ分類器であり，データセットから実際の健常者声サンプルと生成器 G によって生成された合成音声のサンプルを入力として受け取り，生成のサンプルが健常者の声データである確信度・確率を出力する。識別器 D が生成データが健常者の声か合成音声かを判断できない場合最適な状態となる。その時，健常者声データ分布を学習した生成器 G を得る。言い換えると，G は健常者声の声質と遜色ないデータを生成できる能力を獲得する。

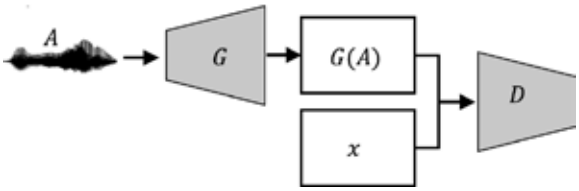

図 1. GAN の構成

図２に示すよう，「声」が話者特徴 h_S（話者性，発話速度，抑揚等特徴）とテキスト h_P 情報（発話内容）から構成されるとされている。構音障がい者の場合は構音器官が必要な筋肉

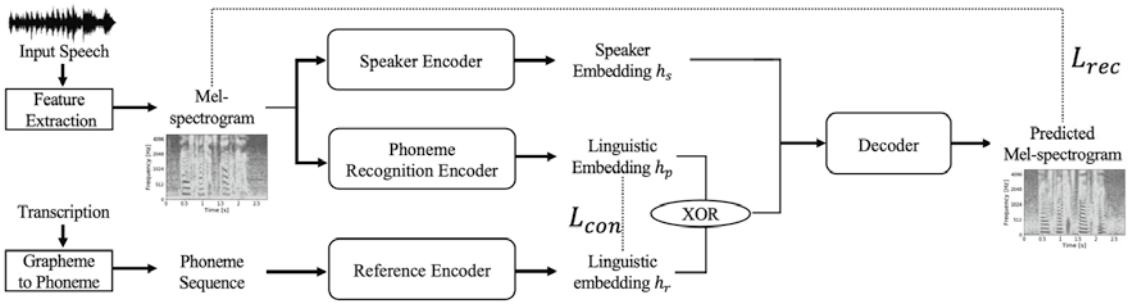

図 2. 提案学習フレームワーク

が麻痺して上手く動かせない等のため発話が不明瞭となり，話者特徴に深くかかわると考えられる。そこで，self-attention 機構の Encoder を用いて，スピーチ A から話者特徴とテキストを分解し，Decoder で分解した話者特徴とテキスト情報を再構成する。入力音声から抽出した特徴空間を，変換した出力音声からの特徴空間に近くなるように，損失関数 L_{rec} で制約することにより，音声分解方法をモデルに学習させる。なお，健常者と構音障がい者のテキストを同じ分布を共有するように音素レベルまでテキスト情報を識別する。Encoder に分解を正しく学習されるため，レファレンステキスト情報（標準テキスト）を介してコンテンツ損失関数 L_{con} の制約により学習スピーチデータ A のテキストを標準テキスト空間に近づくように訂正しながら，学習を行う。これで，テキストの正確性を確保できる。音声変換段階では Decoder で健常話者特徴と構音障がいのテキストを再構成することにより構音障がい者の発話を改善する。（詳細は，成果論文に参照）

3. 評価・考察

実証実験では公開データベースの TORGA を用いて検証を行った。提案の深層学習モデルでは実験には主観評価実験を用いており明瞭性（聞き取りやすさ），話者性に関する評価を行った（本学及び神大学生の候補被験者からランダム 10 名分)。明瞭性は構音障害者の生音声と比較しどちらが聞き取りやすいかを AB 評価，話者性は DMOS (Degradation Mean Opinion Score) により構音障害者の生音声と比較を行い，「5:非常に良い（非常に近い），4:良い（近い），3:普通, 2:悪い（遠い），1:非常に悪い（非常に遠い)」の 5 段階評価を行った。MelCD 法と MSD 法を用いた主観評価実験では，再学習及び声質変換に用いていない音素バランス文をランダムに抽出しテストデータとし，平均を取った。

実験結果（詳細は，成果論文に参照）により各話者において高い明瞭性を示した。また，本手法では変換したデータの基本周波数が健常者の概形を持つため，明瞭性が向上したと考える（成果論文，Fig.2，A2-C2)。また声質変換時，健常者の話者特徴を用いることで安定した発話を合成できたと考える（成果論文, Table.1)。なお，得られた合成音と構音障害者の生音声のスペクトログラムの比較（成果論文，Fig.2，A1-C1）によりも，音障害者発話の低周波数帯域においては生音声の概形を維持しつつ，高周波数帯域においてはパワーが失われておらず，明瞭性の改善が期待できることを確認した。

4. まとめと今後の予定

本研究では，構音障がい者を対象として，発話分析を行い，その分析から判明した不明瞭性の原因を基に声質変換の枠組みを提案することで，聞き取りやすい合成音声を生成した。また，今回は構音障がい者の話者性を維持しながら，聞き取りやすい音声の合成に至らないため，今後はその検討が必要である。

研究成果論文（査読あり・下線：本研究プロジェクトの構成員，*付：協力学生）

1. X. Chen*, A. Oshiro*, J. Chen, R. Takashima, and T. Takiguchi: “Phoneme-Guided Dysarthric Speech Conversion with Non-parallel Data by Joint Training”, Signal, Image and Video Processing, 2022. (forthcoming)