

発話訓練動画を用いた言語障害の 遠隔リハビリテーション支援システム

平山 大貴[†] 長尾 確[‡]

名古屋大学工学部電気電子・情報工学科[†] 名古屋大学大学院情報学研究科[‡]

1. はじめに

言語障害のリハビリテーションにおいて反復して発話訓練を行うことは重要である。一般に、発話訓練は言語聴覚士と対面しながら行われる。近年は人口の高齢化に伴い患者数が増加傾向にあり、発話訓練を指導する言語聴覚士の負担が増大している。また言語聴覚士の数は不足するとされており、患者が十分な発話訓練指導を受けられなくなる可能性が高い。このような問題を解決する手段として、テレビ電話やビデオチャット等を利用した遠隔リハビリテーションが注目されている[1]。

そこで本研究では、患者の1発話ごとの録画が可能な遠隔発話訓練アプリケーションと、録画した動画をもとに言語聴覚士が診断結果を入力するWebアプリケーションを開発した。これらのアプリケーションにより発話データとその発話に対する診断データの収集を行い、機械学習を用いてそれらの関係を学習し、発話の動画から自動的に評価する機械学習モデルを生成する。

また、自動評価の結果を用いて言語聴覚士を支援するシステムの実現方法を検討した。

2. 遠隔リハビリテーション支援システム

本システムはビデオチャットによる発話訓練アプリケーションと動画の管理・分析を行うサーバから構成される(図1)。

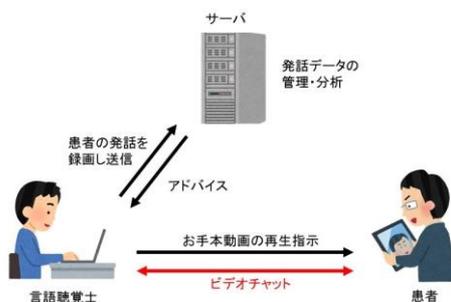


図1. システム全体像

2.1. ビデオチャットによる発話訓練システム

遠隔で発話訓練を行うために、患者と言語聴覚士がビデオチャットを介して発話訓練を行うためのWebアプリケーションを開発した(図2)。本アプリケーションの開発にはWebRTC[2]を用いた。

本アプリケーションには患者側と言語聴覚士側共通の機能として、カメラのON/OFFのボタンと通話の開始/終了のボタンがある。お互いがカメラとマイクをONにし、通話開始ボタンを押すことで通話が開始され、通話している相手の顔と自分の顔が表示される。言語聴覚士側のアプリケーションには、患者側の端末に発話サンプル動画の再生を指示するボタンと、患者の発話を撮影するためのボタンがある。また、患者側のアプリケーションには再生指示を受けて発話サンプル動画を再生する機能がある。

発話サンプルの動画は「おはよう」「ありがとう」「みずをのむ」等の課題語30語に対して、1名の言語聴覚士が正しく発話しているシーンを正面からビデオカメラで撮影したものである(図3)。発話の訓練は、患者による発話サンプルを真似しての発話とそれに対する言語聴覚士の診断・指導の繰り返しにより行われる。

まず、発話訓練指導を行う言語聴覚士は発話サンプルを自分の端末で選択し、再生指示ボタンを押す。患者側の端末では、指示を受け発話サンプルとなる動画が再生される。言語聴覚士は撮影開始ボタンを押した後患者に合図し、患者は発話サンプルの真似をして発話する。発話の終了を確認した後、言語聴覚士が撮影終了ボタンを押す。撮影された動画は発話した患者、担当した言語聴覚士、どの課題語を発話したか、発話訓練の日時といったデータと関連付けられ、自動的にサーバにアップロードされる。その後言語聴覚士が行われた発話に対しての診断および指導を行う。

Remote Rehabilitation Support System for Language Disorder using Speech Training Video

[†]HIRAYAMA, Daiki, Department of Information Engineering, Nagoya University

[‡]NAGAO, Katashi, Graduate School of Informatics, Nagoya University



図 2. 発話訓練アプリケーションの画面例



図 3. 課題語「おはよう」に対する発話サンプルの動画例

2.2. 動画管理・分析サーバ

サーバは発話の動画がアップロードされると、ストレージに動画を保存し、データベースに動画に関連したデータを登録する。そして動画の口の動き、音声から特徴を抽出する。

口の動きは動画をフレームごとに分割した画像に対して機械学習ライブラリ dlib[3]の face landmark detector を用いて、口の輪郭の座標 20 点の時系列データとして抽出した。

音声からは sptk[4]を用いて、MFCC（メル周波数ケプストラム係数）を抽出した。

3. 診断データの収集

言語聴覚士の診断を支援するためには言語聴覚士の診断データが必要であるが、診断および指導はビデオチャットを通じた口頭で行われるため、データとして記録することが難しい。そこで本研究では、1 課題語ごとの発話データに対する診断データを収集するために、言語聴覚士が前述の発話訓練用アプリケーションを用いて収集された発話動画をもとに診断を行い、その結果を記録するためのアプリケーションを開発した（図 4）。

言語聴覚士は撮影された動画を閲覧し、「お」「は」「よう」のような音節ごとの発話の明瞭度、「お-は」「は-よう」のような音と音のつながり（渡り）の発話の明瞭度を評価する。発話の明瞭度は 1.0 点（問題なし）から 5.0 点（問題あり）までの 0.5 点刻み 9 段階で評価する。

点数の入力後、診断終了ボタンを押すことで点数が診断の結果として記録される。

発話が途中で途切れている、課題語の発話以外の音が入っている等の動画は、1 課題語の発話の動画として扱うことができないため、評価から除外した。



図 4. 診断入力アプリケーションの画面例

4. 言語聴覚士に対する支援

発話動画から抽出されたデータとそれに対する診断データを組み合わせて機械学習を行い、音声を含む発話動画から自動的に発話の評価する機械学習モデルを作成する。

サーバ上で動画の発話の自動評価を行い、その評価に基づいて言語聴覚士に情報を提示することにより、言語聴覚士の活動を支援する予定である。具体的な情報提示としては、過去の発話の評価との比較に基づいて、上達が見られる部分や重点的に指導すべき項目を、一覧として言語聴覚士側のアプリケーションに表示することを検討している。

5. おわりに

本研究では、ビデオチャットを介した遠隔での発話訓練システムおよび言語聴覚士による診断データを収集するシステムを開発した。

今後は、言語聴覚士の協力のもと本システムを用いて、患者の発話データおよび言語聴覚士による診断データの収集を行う予定である。その後、機械学習モデルの評価を行い、患者および言語聴覚士を支援する仕組みを実現する。

参考文献

- [1] 鈴木昂太郎, 村田嘉利, 鈴木彰真, 佐藤永欣: Kinect を用いた遠隔リハビリテーションシステムの研究, マルチメディア分散協調とモバイルシンポジウム 2016 論文集, pp. 314-321 (2016-07-06)
- [2] WebRTC, <https://webrtc.org/> (2018/1/10 閲覧)
- [3] dlib, <http://dlib.net/> (2018/1/10 閲覧)
- [4] sptk, <http://sp-tk.sourceforge.net/> (2018/1/10 閲覧)