

発話困難者のためのリアルタイム音声生成アプリケーションの開発

橋場 参生, 高瀬 雅由*, 須貝 保徳*, 藪 謙一郎**, 伊福部 達***

Development of Real-time Speech Synthesis Application Software for Speech Disorders

Mitsuo HASHIBA, Masayoshi TAKASE*, Yasunori SUGAI*
Kenichirou YABU**, Tohru IFUKUBE***

キーワード：音声，発話障害，構音，ホルマント，音声合成，福祉機器

1. はじめに

我々が普段何気なく使用している音声は、主に、肺からの「呼気の送付」、喉頭での「原音生成」、そして、口唇や舌を動かして原音に変化を加える「構音」という一連の過程を経て作り出されている。従って、舌・唇・顎などに欠損や変形が生じてしまった場合や、筋肉や神経系の疾患などで、これらの動きに制約が生じてしまった場合には、音声の生成に必要な「構音」機能に障害が生じ、発話が困難になる。

筆者らはこれまで、喉頭癌などで生じた「原音生成」の障害を支援するための福祉機器『電気式人工喉頭』の製品化に取り組んできたが^{1,2)}、この度、「構音」の障害により発話が困難な方々も支援するために、新たな音声生成技術の研究開発を実施した。本技術は、二次元平面上で指示した座標の軌跡を基にしてリアルタイムに音声を生成する技術で、キーボード等による文字入力が必要とせず、発話内容にも制限が生じないという特長がある。

以下、本報では、まず、発話の仕組みと構音の役割について説明し、続いて、開発した音声生成術の原理と概要について述べる。さらに、本研究の成果を基にして、一般ユーザも利用可能な形で商品化したスマートフォン向けアプリケーションソフトウェアについて報告する。

2. 発話の仕組み

2.1 発話器官

音声の生成に関わる発話器官を図1に示す。音声は主として、肺から喉頭への「呼気の送付」、喉頭内の声帯が呼気で振動して起こる「原音生成」、舌・唇・顎などの動きによって原音に変化が加えられて音声言語となる「構音」という三つの過程を経て生成されており、何れの過程に障害が生じていても、音声によるコミュニケーションが困難になる。発話障害を引き起こす原因としては、舌・唇・顎などの欠損や変形の他、脳血管障害、筋ジストロフィ、筋萎縮性側索硬化症(ALS)などで生じる神経系や筋肉系の疾患、また、喉頭癌による喉頭の摘出などがある。

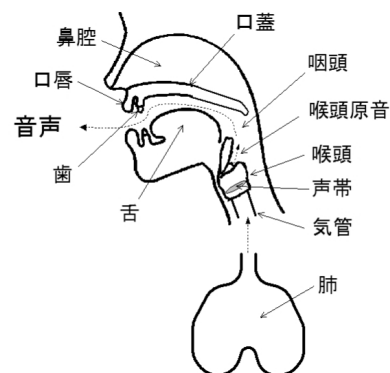


図1 発話器官

2.2 構音とホルマント

音声は、母音と子音によって構成されており、このうち母音は構音運動によって生じる複数の共鳴周波数によって特徴づけられている。この共鳴周波数は、周波数の低い方から第1ホルマント(F1)、第2ホルマント(F2)と呼ばれており、

* 株式会社 電制

** 東京大学 先端科学技術研究センター

*** 東京大学 高齢社会総合研究機構

事業名：重点研究

課題名：人間親和型ICTシステム開発のためのインタラクティブ
インタフェース技術に関する研究

F1とF2を両軸に取って、日本語5母音との対応をグラフ化すると図2に示すように5角形の配置が得られる³⁾。そして、図3に示すように、この5角形の配置は、顎の開閉や舌の前後の動きによる母音の構音位置とも密接な関係を持っている。また、子音も同様に、口唇や舌による狭めや閉鎖といった構音運動によって作り出されている。

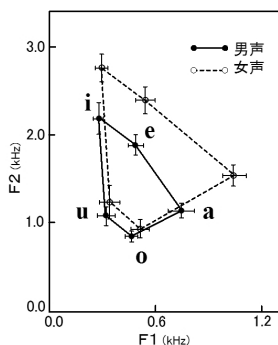


図2 F1-F2平面³⁾

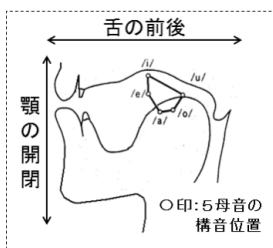


図3 日本語母音の構音位置

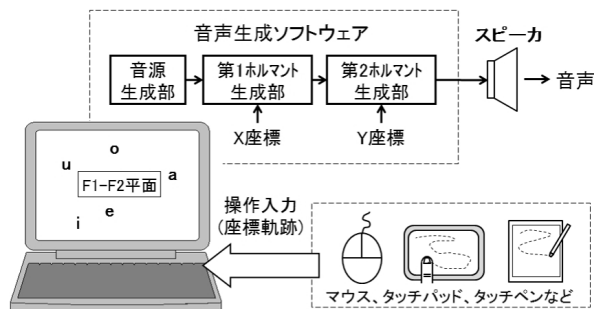


図4 開発した音声生成技術の概要

3. リアルタイム音声生成技術の開発

3.1 発話困難者を支援する既存技術

発話が困難になった場合のコミュニケーション手段としては、筆談の他、文字盤やメッセージカードを利用する方法、手話などがあるが、これらは何れも相手の視覚を通じたコミュニケーションとなり、音声は伴わない。音声によるコミュニケーションを可能にする電子的な支援機器はVOCA (Voice Output Communication Aids) と呼ばれており、大別して、事前に登録されたメッセージのみを再生できる方式と、キーボードから入力した文章を音声合成して読み上げる方式がある。しかし、前者の方式には、発話内容が限定されるという課題があり、後者の方式には、会話のリアルタイム性に制約が生じるという課題がある。

3.2 開発した音声生成技術の概要

開発した音声生成技術の概要を図4に示す。F1-F2平面を操作座標平面に見立て、マウスやタッチパッド等で指示されたX座標から第1ホルマント生成部の共鳴周波数を計算し、Y座標から第2ホルマント生成部の共鳴周波数を計算する。そして、音源生成部で生成された原音に、第1・第2ホルマント生成部で共鳴の特性を加えてスピーカから出力する。図3に示したように、F1-F2平面は、顎の開閉と舌の前後の動きに対応していることから、図4の音声生成技術を用いると、マウス等を動かす方向や速度によって、あたかも顎や舌が動いて生じるような様々な音声の変化を作り出せることになる。なお、図4では省略しているが、子音については、疑似ホワイトノイズを原音として、破裂音や摩擦音などに相当する音を生成する処理を開発している。

3.3 音声の生成

図4の音声生成技術をパーソナルコンピュータ上で動作するソフトウェアとして実装し、タッチパッドを接続して音声の生成試験を行った⁴⁾。まず、画面に表示された/i/, /e/, /a/, /o/, /u/の位置を順に指示し、5母音を離散的に発生させた場合の時間波形と周波数分析結果を図5(a)に示す。また、同図(b)に、成人男性による同様の発声結果を示す。両図の周波数分析結果の比較から、本技術で生成された5母音は、健常者と類似した第1・第2ホルマントを有する音声になっていることがわかる。

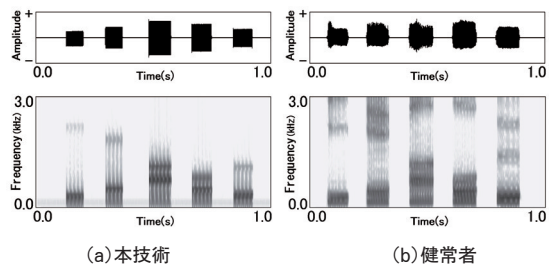


図5 /i/, /e/, /a/, /o/, /u/の離散生成結果

次に、画面に表示された/i/, /e/, /a/, /o/, /u/の位置を、指を離さずに順になぞって行くことによって、5母音を連続的に発生させた場合の時間波形と周波数分析結果を図6(a)に示す。同図(b)に示す健常者の発声結果と比較することにより、母音の特徴である第1・第2ホルマントはもちろんのこと、母音から他の母音へと移行する際のホルマントの連続的な推移(調音結合)も、良好に生成できていることがわかる。

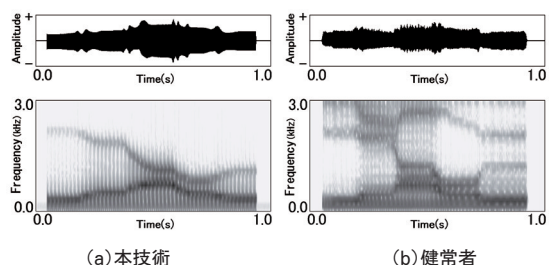


図6 /i/, /e/, /a/, /o/, /u/の連続生成結果

4. スマートフォン向けアプリケーションの開発

研究成果の早期実用化を図るため、図4の音声生成技術を基にして、近年普及が著しいスマートフォンやタブレット端末上で動作する音声生成アプリケーションを開発した。

4.1 概要

スマートフォンやタブレット端末向けに開発したアプリケーションの操作画面を図7に、操作方法を図8に示す。操作画面の中央部には、F1-F2平面に相当する操作エリアがあり、日本語5母音や子音の一部を表すローマ字が配置されている。また、その左側には「カ (ka)」の/k/、「サ (sa)」の/s/などの破裂音や摩擦音を生成するための補助ボタンが配置されている。この操作エリアを、図8のように指先やタッチペンでなぞることにより、様々な音声をリアルタイムに生成することができる。さらに、本アプリケーションは端末の傾斜角度を基にして音声の基本周波数を変化させる機能を備えており、この機能を併用することで「雨」と「鈴」などのイントネーションの違いを表出することもできる。この他、画面上部には後述する設定画面などに移動するためのボタンや、音声の生成にあわせて口や舌が動く横顔のアニメーション画像が表示されており、画面下部には生成中の音声の基本周波数などの情報が表示されるようになっている。

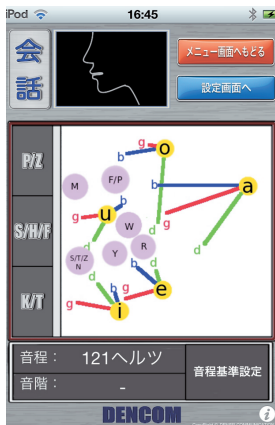


図7 操作画面



図8 操作方法

4.2 発声操作例

図9から図12に基本的な操作例を示す。図9は「アイウエオ」と音声生成する場合の操作軌跡で、画面上の/a/, /i/, /u/, /e/, /o/の記号を連続して順になぞれば良い。図10はマ行を生成するための軌跡で、/M/の記号から/a/, /i/, /u/, /e/, /o/に向けて素早く指を動かすことで「マ・ミ・ム・メ・モ」を生成できる。図11はヤ行を生成するための軌跡で、ヤ行は/i/の口形から/a/, /u/, /o/の口形に変化させて発話することから、画面上でも/i/を起点にして/a/, /u/, /o/の方向へ素早く指を動かすことで「ヤ・ユ・ヨ」

の生成が可能となる。同様に「ワ」は/u/の口形から/a/の口形に変化させて発話することから、/u/の記号から/a/の記号へと素早く指を動かせば良い。図12は「バ・ガ・ダ」に相当する音声を生成するための操作で、これらは、/a/に向かって引かれている3本の導線に沿って指を動かすことで実現できる。

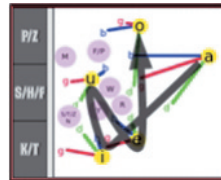


図9 「アイウエオ」

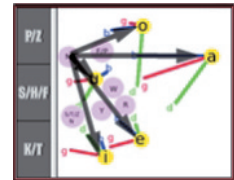


図10 「マ・ミ・ム・メ・モ」

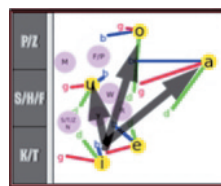


図11 「ヤ・ユ・ヨ」

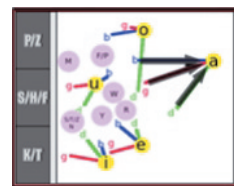


図12 「バ・ガ・ダ」

次に、具体的な言葉を生成するための操作例を図13から図18に示す。図13は「青い海」と出力する場合の操作軌跡で、/a/, /o/, /i/, /u/, /M/, /i/の順に記号をなぞれば良い。図14と図15は、それぞれ「暑い」と「寒い」の操作軌跡で、操作エリア左側の補助ボタンも利用することで「ツ (tu)」の子音/t/に相当する破裂音や「サ (sa)」の子音/s/に相当する摩擦音を含んだ音声を生成できる。図16は「ありがとう」の操作軌跡である。

なお、本技術は、子音の生成に関して必ずしも厳密に操作する必要は無く、例えば、図17のように円を描く操作で「おはよう」と聞こえる音声を、図18のように左右に指を動かす操作で「こんばんは」と聞こえる音声などを生成することができる。このような特徴から、発話内容がある程度推定できるような場面では、F1-F2平面上をなぞる操作だけでも、意思疎通の支援に十分役立つ音声を生成できると考えている。

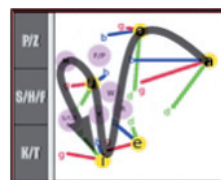


図13 「青い海」

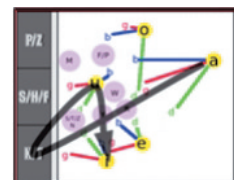


図14 「暑い」

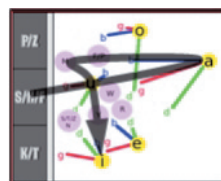


図15 「寒い」

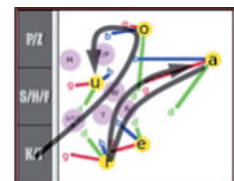


図16 「ありがとう」

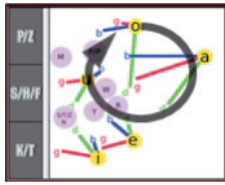


図17 「おはよう」

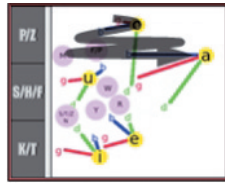


図18 「こんばんは」

4.3 その他の機能

図19はアプリケーションの各種設定を行う画面で、操作エリアの導線の有無や、端末の傾斜で基本周波数を変化させる機能のオン・オフ、男性・女性・合成音などの音源の選択などを行うことができる。また、音源に関しては、図20の録音・分析機能を開発し、録音された音声からユーザ固有の音源を作成することで、ユーザの特徴に近い声質での音声生成もできるようにした。この他、日常会話などの操作の練習サンプル(図21)や、歌詞に沿って画面をなぞると、自動的に音程が付加されて歌を歌える練習サンプル(図22)なども開発した。

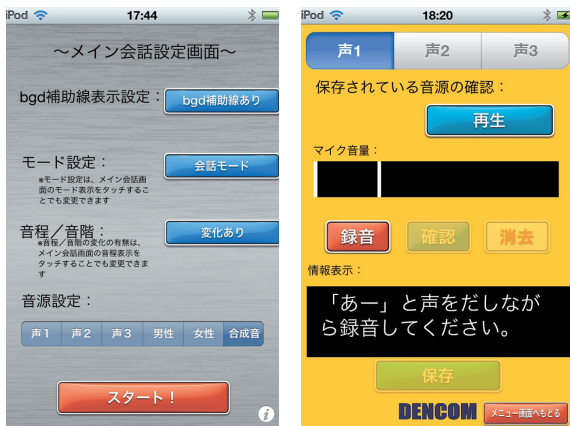


図19 設定画面

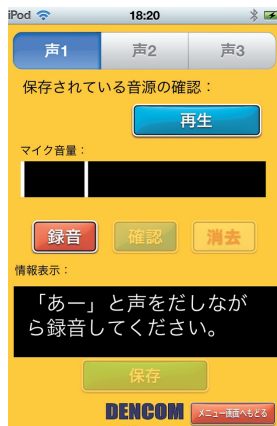


図20 録音分析画面

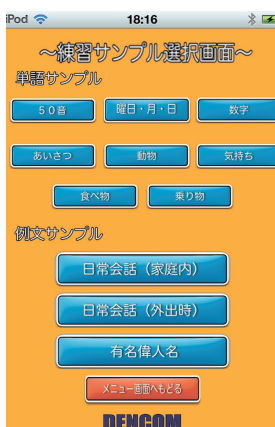


図21 練習サンプル選択画面

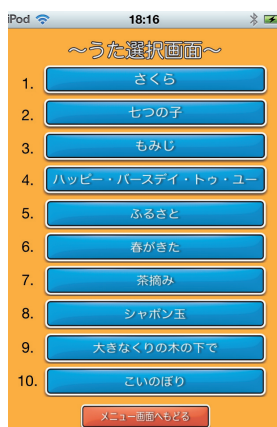


図22 うた選択画面

5. 商品化

開発した音声生成アプリは、iPhoneやiPad、iPodTouch等のiOS端末上で動作するアプリケーションソフトウェア

『ゆびで話そう』として2013年4月15日に商品化された(図23)。現在、アップル社が運営するApp Store(アップストア)より、有料版と無料体験版がダウンロード可能になっている。



図23 音声生成アプリ「ゆびで話そう」(株電制HPより)

6. おわりに

発話困難者の支援を目的に、画面をなぞる指やペンの軌跡からリアルタイムに音声生成する技術の研究開発に取り組み、一般ユーザも利用可能なスマートフォン向けアプリケーションとして商品化した。今後は、子音の生成方法等に関する先行研究^{5,6)}の成果を反映させて一層の音質向上を図ると共に、手指の動きに制約がある発話困難者の支援を可能にするための各種操作インターフェースの開発を進める計画である。

引用文献

- 1) 橋場参生・上見憲弘・及川雅稔・山口悦範・須貝保徳・伊福部達：抑揚制御機能を備えた電気式人工喉頭の製品化と評価，電子情報通信学会論文誌D-II，Vol. J84-D2，No. 6，pp. 1240-1247，(2001)
- 2) 橋場参生：喉頭摘出者の音声コミュニケーションを支援する電気式人工喉頭の製品化と展開，第25回バイオエンジニアリング講演会講演論文集，pp. 31-32，(2013)
- 3) 佐藤大和：男女声の声質情報を決める要素，通研実報，vol.24，No5，pp.977-993，(1975)
- 4) 橋場参生・藪謙一郎・井野秀一・泉隆・伊福部達：F1-F2平面で操作する会話補助装置の提案と試作，ヒューマンインタフェースシンポジウム2005論文集，pp. 585-588，(2005)
- 5) 藪謙一郎・伊福部 達：発話障害者支援のための連続タッチ平面で操作する音声生成器—子音改善のための基礎的検討—，信学技報 WIT2010-40，pp. 41-46，(2010)
- 6) 藪謙一郎・伊福部達：発話障害者支援のためのペン入力座標によるリアルタイム音声生成方式—鼻子音出力の操作方法と音声生成方法の検討—，信学技報，SP2010-126，pp. 67-72，(2011)