

201122057A

別添1

厚生労働科学研究費補助金
障害者対策総合研究事業

発話障害者のための
ハンズフリー支援機器の開発とその臨床評価

平成23年度 総括研究報告書

研究代表者 伊福部 達
平成24（2012）年 5月

目 次

I. 総括研究報告	
発話障害者のためのハンズフリー支援機器の開発とその臨床評価	----- 3
伊福部 達	
(資料1) 試作マイクの評価実験手順・結果	----- 11
(資料2) 音声生成器の抑揚制御についての評価実験の手順と結果	----- 13
II. 研究成果の刊行に関する一覧表	----- 15
III. 研究成果の刊行物・別刷	----- 17

研究要旨

代表者らは、過去10年以上にわたり喉頭摘出者のための「電気式人工喉頭」の実用化とその普及に努めてきた。その間に多くのユーザから、仕事をしながら使えるように手を束縛しない「ハンズフリー型」にして欲しいという要望が続いていた。一方、発話のための筋・神経系に疾患のある構音障害者からは、残された身体機能の動きにより実時間で感情豊かな音声を生成することができるような「発話障害支援のための音声生成器」を開発してほしいという要望が出てきている。本課題の目的は、これらの要望に応じて、①ハンズフリー電気喉頭の改良研究を進めるとともに、②構音障害支援器の開発研究を行うものであり、喉頭摘出者から脳・神経系疾患による発話障害者の全般をカバーできるようにする。

ハンズフリー電気喉頭について、平成22年度には、発声音の音量拡大のためにマイクロホン、アンプ、スピーカが一体化した小型拡声器2種類の試作と使い勝手の評価を行った。それを踏まえ、平成23年度は、さらに、拡声器使用時の周囲雑音や本体からの雑音を低減し発声音を選択的に取得でき、使用時にハウリングを防止できるようにマイクロホンを改良した。そして、人工喉頭によって発声された音声について4モーラの単語リストを用いて明瞭さの定量的な評価を行った。評価の実験から、改良型マイクを用いることで増幅音量を従来よりも10dB以上向上でき、聞き取り易さも向上することが分かった。このことから、特に、周囲に雑音のある環境下での音量増幅の効果が期待される。

一方、発話支援音声生成器については、平成22年度は/サ/などの「摩擦音」や/ナ/などの「鼻子音」などの子音を実時間で指先の動きだけで生成できるようにし、ほぼすべての音素を不明瞭ながらも表現可能にしたことを踏まえ、平成23年度には、抑揚の制御できるようにする点と、小型化に向けたタッチパネル型PCに対応する点との2点に焦点を当てた。

抑揚については、指に取り付けたフィルム型の圧力センサの押圧によって音程を制御する方法と、使用者の声を頸部に取り付けた皮膚伝導マイクロホンによって取得しホルマント合成の原音に用いる方法との2種類による実験を行い、双方とも、使用者が意図する音程の変化を付加できることが分かった。また、聴取実験から、抑揚のある音声のほうがわずかに単語を了解しやすくなることも示された。押圧によって抑揚を付加する手法については、前述の人工喉頭による抑揚制御にも応用可能と考えている。ただし、操作するセンサが増えるため、今後、慣れや使いやすさの個人差については、今後検討していく必要がある。また、タッチパネルへの対応については、本体単体だけでも操作可能となったことにより、携帯性や手軽さが向上した。ただし、表面の指の滑りの良さや、操作盤面の大きさ等を使用者に合わせて調整する必要があるため、動作する機種を選定については、今後さらに検討していく。押圧センサによる抑揚制御方式については、電気喉頭にも応用でき、試作器を作成中である。また、音声生成機能の一部をiPhone上でも動作できるようにし、実用化への手掛かりとしている。

以上について、プロトタイプ機器を他の治療法やリハビリ法を補完するという立場で臨床的な観点から有用性と問題点を評価し、実用化への道筋について検討した。

分担者

伊福 部達
東京大学 高齢社会総合研究機構 特任教授

田中 敏明
東京大学 先端科学技術研究センター 特任教授

上田 一貴
東京大学 工学系研究科 特任講師

A. 研究目的

代表者らは今まで、喉頭摘出者のための「電気式人工喉頭」の実用化とその普及に努めてきたが、その間に多くのユーザからこの電気喉頭を手を束縛しない「ハンズフリー型」にして欲しいという強い要望が続いている。一方、脳・神経系疾患による構音障害患者からは、実時間で感情表現もできる「構音障害支援器」を開発してほしいという要望が出ていた。

本課題はこれらの目的に応えるために、①ハンズフリー電気喉頭の改良研究を進めるとともに、②構音障害支援器の開発研究を行うものである。この2つの支援技術により喉頭摘出者と脳・神経系

疾患による発話障害者の両方をカバーする。試作済みの発話支援機器を、他の治療法やリハビリ法を補完するという立場で臨床的な観点から有用性と問題点を評価し、製品化を目指して研究を進める。

とくに、①については、ハンズフリー化するために首かけ式の装具によって人工喉頭の加振部を頸部に当てる方法を検討しているが、それによる音量の低下を防ぐために、マイクロホン、アンプ、スピーカが一体化した小型拡声器の開発・改良に取り組んでいる。

平成23年度は、拡声機のマイクロホンについて、拡声器使用時の周囲雑音や本体からの雑音を低減し発声音を選択的に取得でき、使用時にハウリングを防止できるように改良し、人工喉頭と拡声機によって生成された声の定量的な評価を行なう事を目的とした。

また、②については、操作盤面をペンや指でなぞって操作する音声生成器の開発を行い、平成22年度までに、ぼすべての音素を不明瞭ながらも表現可能にしたことを踏まえ、平成23年度には、抑揚を制御できる手の開発と実装を行うことと、小型化に向けてタッチパネル型PCに対応させ携帯性の向上を実現することとの2点を目的とした。

以上のことから、臨床的な立場から本方式によって、両手を使う作業時においても十分に使用できるかを確認した。

B. 研究方法

B-1 ハンズフリー電気喉頭に関する検討

ハンズフリー電気喉頭の実用化と評価の研究においては、(1)首バンド装着用の薄型軽量振動子の開発、(2)発声時にのみ働く高感度の抑揚制御呼吸センサの開発、(3)音量が大きく音漏れがしない小型拡声器の開発の3点が、過去の実績・経験から、解決すべき課題として挙がっている。

この3課題のうち特に22年度以前にはほぼ未着手であった、マイクロホン、アンプ、及びスピーカからなる小型拡声器の開発とその使い勝手の評価に焦点を絞り込み、長年拡声器開発に取り組んでいた協力研究員が本課題を担当した。

ハンズフリー首バンド装着型の電気喉頭の開発の際には、既に実用に至っている手持ち型の電気喉頭に比べて、装置を頸部に当てる強さを強くできず、頸部に当たった装置から直接外へ漏れだす音が雑音となったり、大きな声が出せないといった課題が強くなることが懸念されている。

拡声器は、これらの課題を解決するためのものである。この場合、使用者が拡声器をポケットへ装着したり、首から下げたりして携帯して使用することになる(例として、図1に平成23年度に開発した手持ち型の電気喉頭のための拡声器の使用

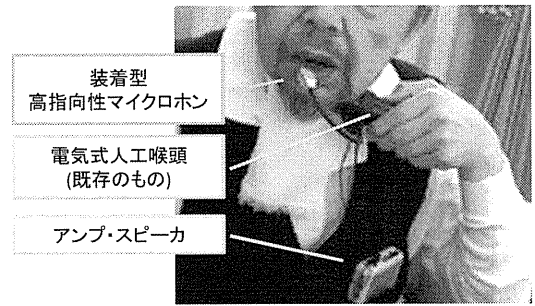


図1 開発中の拡声機の使用例

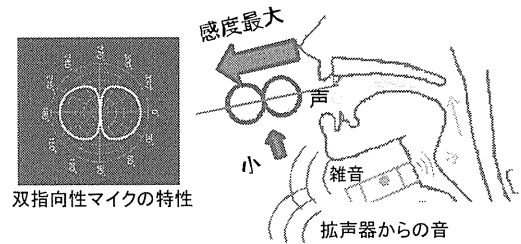


図2 マイクの指向性と雑音源の関係

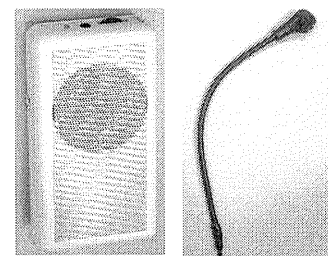


図3 電気人工喉頭のための拡声器
(協力: ㈱サザン音響)

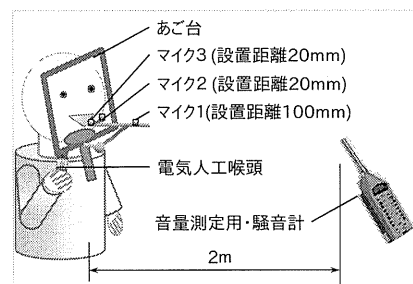


図4 録音用マイクの設置位置

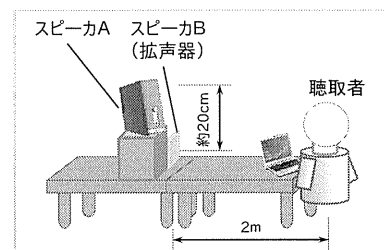


図5 聴取実験装置

例を示す)。このとき、マイクと出力スピーカとの距離が近いために、市販される汎用的なマイクによる拡声器では、音量を上げるとハウリングを生じてしまって十分な音量を得ることが困難であった。

そこで代表者らは、平成22年度から23年度にかけて、使用者の至近距離にスピーカがある場合でも、ハウリングを抑えたまま十分な拡声効果を得られる拡声器の開発を試み、拡声器の試作を行った。平成23年度は特に、マイクの特性に焦点を絞り込み、必要な音のみを増幅させるように、改良を加えて人工喉頭による声の簡易的な評価を行った。

試作マイクの具体的な改良方法は、図2に示すように口に向けた方向と平行な方向だけの音を取得できるような、双指向の特性を持つマイクを用いて、雑音源が位置する方向への感度がほぼゼロとなるように配置して、ハウリングマージンを上げ、同時に人工喉頭の直接音を抑える工夫を施した。

評価方法としては、人工喉頭を使用して単語を読み上げた場合の、①拡声器を用いない場合、②無指向マイク(汎用マイクを想定)の拡声器Ⅰを用いた場合、③双指向マイク(改良マイクを想定)の拡声器Ⅱを用いた場合との3種類について、比較し、拡声器の効果を調べた。また同時に、これまで定量的に調べられていなかった、電気人工喉頭で発話される声の明瞭さについて、文脈の無い単語音声の了解度で、調べた。

評価実験は、人工喉頭で単語を読み上げた声について、拡声器を用いない場合(条件Ⅰ)、未改良の例である無指向性マイクの拡声器Ⅰを用いる場合(条件2)、改良型である双指向性マイクの拡声器Ⅱを用いる場合(条件3)を、発話・聴取実験で比較した。その際、話者が話した声の質を3条件で一致させる必要がある。そのため、図4に示すような装置で、あらかじめ条件1～3のマイクからの音声を同時に収録した上で、聴取実験で各条件をシミュレートした音量で再生・提示する方法をとった。

実験から、各条件の了解度をもって、人工喉頭の音声の評価とし、実用上の展望と臨床に向けた指針を得た。

なお、実験の詳細な手順と結果については、補足資料Aに記述する。

倫理面への配慮事項として、話者に対して、体力的な負担が無いように十分な休憩をとりながら、録音実験を行った。また、次年度以降に予定している患者による試用実験を想定し、東京大学の倫理指針に基づき、臨床評価に向けて倫理審査の申請を行い委員会からの認可を受けた。

B-2 音声生成器に関する検討について

音声生成器については、前年度までの研究に引

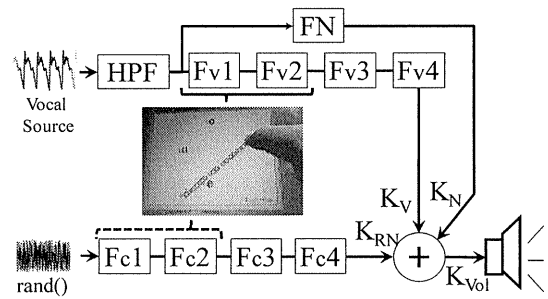


図6 音声生成器のブロック図

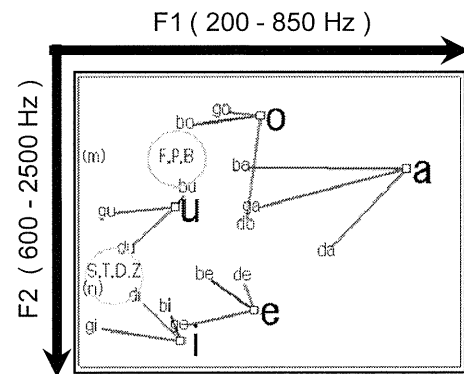


図7 音声生成器の操作盤面

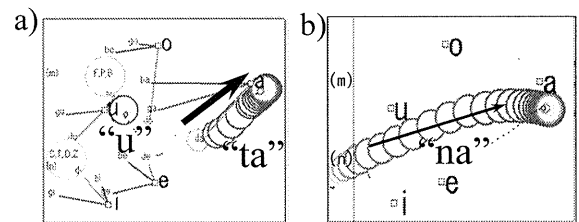


図8 子音生成時の入力軌跡

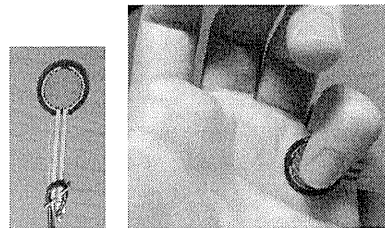


図9 押圧センサ FSR と取り付け位置

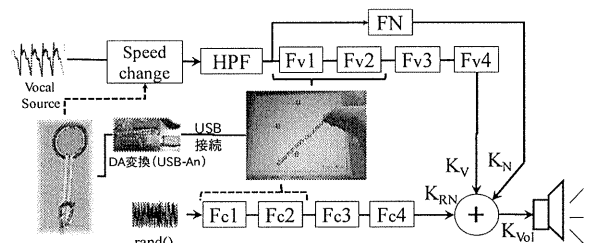


図10 押圧センサによる抑揚制御

き続き、文字や単語に頼らずに、指やペンで面をなぞる動きで、楽器のように使用者自身がリアルタイムに音声をコントロールしながら音声を生成できる音声生成器の改良を進めている。

この音声生成器は、ヒトの母音知覚に必要な第1ホルマントと第2ホルマントの周波数(以下、F1,F2)を、音声生成器の操作面の縦軸と横軸とに割り当て、使用者が図7のような操作盤面を指やペンを操作面に置いた瞬間に、その位置に対応したホルマント周波数を持った音を発するものである。

平成22年度までに、母音音声の出力をベースとした本方式を用いて、疑似子音音声を生成する手法を提案し改良を進め、図8のように操作盤面上をタップしたりなぞったりすることで、ほぼすべての子音を表現可能としてきた。そこで、平成23年度は、会話中の感情表現等に影響の大きい、抑揚に関する点と、使いやすさに影響する、静電容量方式のタッチパネル型PCへの対応に焦点を絞って、改良を進めた。

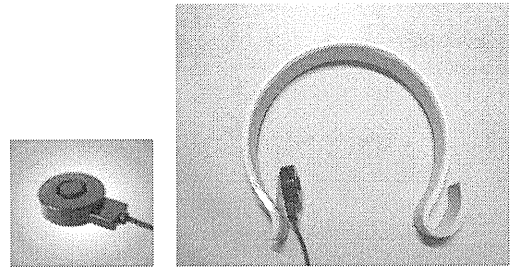


図 11 皮膚伝導マイクと試作の頸部への固定具

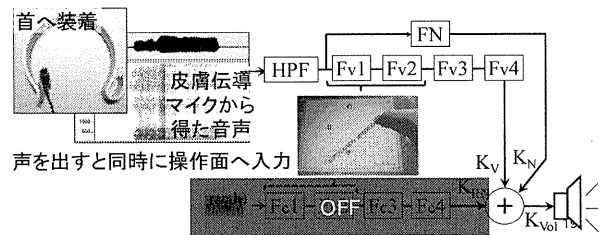


図 12 声を用いた抑揚制御

B-2-a 抑揚の制御について

抑揚の制御方式としては、次の2つの方式を提案・検討した。すなわち、タッチパネルを人差し指で操作しながら、(1)人差し指と独立に動かし易い小指の握力抑揚をコントロールする方式と、(2)タッチパネルの操作と同時に使用者が声を出し、その声の抑揚をそのまま用いる方式との、2種類である。これらの抑揚制御方式をこれまでの音声生成ソフトウェアに組み込み、簡易的な音声生成実験とその聴取実験を行った。

なお、後者の方式は構音機能のみに主な障害をもち、声帯音の発声には障害を持たない場合を想定したものである。そのような話者にとっては、もっとも自然な抑揚付加方法であるといえる。

○小指握力による抑揚制御

音声生成方式では、操作面をタッチする場所を素早い動きで移動したり、100~200 msの単位で一瞬操作面から指を離したりして、音韻を制御する。これらの操作により、操作面を押す力や、タッチする指にかかる力が自然に変化してしまうため、この音韻の制御の動きと抑揚制御の動きが無関係に制御できる必要がある。そこで、図9のように小型の押圧センサを用いて、操作面をタッチする人差し指からもっとも離れた、小指の握力によって抑揚を制御する方式を試みた。

小指の握力によってこの押圧センサを押すことによって、抑揚が制御される。押圧センサから得られた電圧信号は、図10に示すようにUSBインタフェースによってPCに取り込まれ、抑揚が変化する。使用者ははじめに最大の握力かけ、握力無付加時と最大握力時との間で、1オクターブの声を制御可能にした。

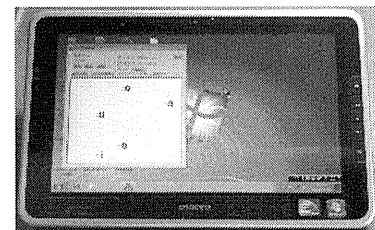


図 13 タブレット PC 上の動作画面

表 1 タブレット PC の仕様

本体仕様	TW217A5
OS	Windows® 7 Home Premium 32 ビット
内蔵ディスプレイ	タッチパネル付 10.1 型 ワイド 最大 1,024×600 ドット
キーボード	なし
ポインティングデバイス	静電容量式タッチパネル (マルチタッチ対応)
寸法	274(×173×18.5 mm
質量	約 850g
バッテリー	動作時間 約 6.4 時間

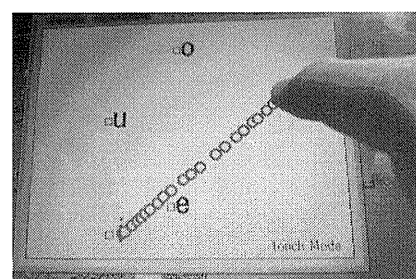


図 14 "ya"出力時のタッチパネル表示

○使用者の声による抑揚制御

本研究では、ホルマント合成方式を採用しているため、声帯音を模擬した原音と、音韻を与える共鳴の制御を分離できる。実際のヒトの発話においても、抑揚や声の音量は声帯音によって出され、音韻については構音器官で特徴づけられるため、構音器官のみに障害を持つ場合もある。そのような患者では、自分の正常な声帯から出される抑揚や強弱によって発話できることが有用である。

そこで、図11に示すような皮膚伝導マイクロホンによって、頸部から取得した音声波形をホルマント合成の原音として用い、その音に対してタッチパッドからの入力によって制御されるホルマントの特徴づけを行う手法を導入した(図12)。

マイクロホンの固定具としては、過去にハンズフリー電気喉頭の固定具として作成されたものと同様のものを試作し、電気喉頭加振器の代わりにマイクロホンを取り付けた。マイクロホンの信号はUSB音声入出力デバイスを介して音声生成器が動作するタブレットPCへ入力し、そのままホルマント合成の原音として用いた。

以上の2方式について、発話障害を持ちこの研究に関わってきた研究員による試用と、生成された音声の聴取実験によって評価をおこなった。

B-2-b タッチパネル型PCへの対応について

平成22年度までの研究では、主に小型のペンタブレットを用いた単語音声の生成実験を行い、擬似子音や鼻音付加の効果を調べてきた。ペンタブレットが採用された最大の理由としては、専用のデバイスを用意することなく、絶対座標が入力可能であることであったが、一方でパーソナルコンピュータ(PC)と一体でないために、携帯しづらいという問題や、表示画面と操作面が分かれてしまうという問題があった。また、抵抗膜式のタッチパネルを備えたPCや、タッチパッドを用いた実験では、操作に、ある程度の圧力が必要であるため、指が滑らせづらいという難点があった。

一方で最近の数年の間に、静電容量式のタッチパネルを備えた、携帯性に優れた汎用ノートPC(タブレットPC)製品が急速に増えている状況にある。タッチパネルが内蔵された小型のコンピュータや、携帯電話のようなポータブルデバイスで本研究の音声生成器を動作させることができれば、接続の煩雑さの軽減や、表示画面と操作面との一体化が可能になる。

具体的には、PCとして、ONKYO TW217A5を使用した。静電容量式のタッチパネルを持つ10.1インチワイドサイズの液晶ディスプレイが内蔵されている。Windows 7搭載であるため、Windows

タッチの機能を使用でき、マルチタッチ操作が可能となっている。

一般にタブレットPCは、キーボードを持たずほぼ全ての動作をタッチパネルからの入力によって行う。タッチパネルからの入力は多くの場合、マウス入力として処理され、タッチパネル専用の機能を持たないソフトウェアでも動作可能になっている。

タッチパネル対応のPCでは、画面タッチの情報をマウスポインタの機能として扱う処理方法が一般的である。しかし、その方式ではタッチパネルの操作からマウス処理へ変換される過程に、原理的な遅延がある。マウスでは、タッチの情報からクリック、ダブルクリック、右クリックへ振り分ける。長押し等によってそれらを判別するため、画面にタッチしてすぐに左クリックとして認識させることが不可能となっている。

以上のことから、タッチパネルからの情報を直接処理をするために、“Windowsタッチ”機能を用いるためのソフトウェアの改良を加えた。この方式を用いることで、発話のための遅延を以前に比べて抑えることができ、スムーズな音声出力を可能にした。

C. 研究結果

C-1 ハンズフリー電気喉頭に関する検討

前年度に引き続いて、マイク・拡声器の開発に長年かかわる研究員の協力により、改良型のマイクを使った拡声器を制作できた。

その結果、マイクの指向性を、口の方向に平行な方向からの音だけを取得できるような、双指向の特性を持つマイクに改良することにより、実際に、雑音の低減と、ハウリングマージンを上げることができた。具体的には、実験から、スピーカをマイクに向けて、マイクの下20cmの距離に置いた状態でも、ハウリングを起こさずに、約60dBAの声を約70dBAにまで増幅でき、従来よりも音量を上げることが確認できた。

また、電気喉頭で発話された声の聴取実験での了解度の結果では、単語全体が正答された数が、拡声器を使わない条件よりも改良型の拡声器を使ったほうが僅かに上昇した。特に、1単語を音節ごとに分割した場合に、1つの単語中の1音節も理解されない単語が約半数に減少し、拡声器とマイクの効果によって、特に聴きづらかった音が、聞きやすくなったことが伺えた。

C-2 音声生成器に関する検討について

C-2-a 抑揚の制御について

実験により、小指握力による抑揚制御と皮膚伝導マイクからの声を用いた抑揚制御の2つによる抑揚制御を音声生成ソフトウェアへ実装させ、動作させることができた。

音声生成実験では、音声生成器で声を生成する生成者により、26語を対象に所望の音声が生じた。生成された音声の例を図15-16に示す。図15は、小指に当てた押圧センサにかける握力によって抑揚を制御しながら、音韻を制御して生成した音声。図16は、生成者が声を出しながら、ペンによって音韻を制御し、皮膚伝導マイクによってピックアップされた声に対して音声生成器で言葉になるように音韻が付加されて出された音声である。

生成された音声のうち、いずれの単語も、通常の発話と変わらない自然なタイミングで抑揚が付けられた。ただし、抑揚を操るタイミングと合わせて音韻のコントロールを加えるため、ややゆっくりとした口調の音声が生じた。

小指圧力による制御では抑揚は強め、使用者の声による方式ではやや弱めの抑揚であった。握力を試用した方式の場合には微調整が難しかったためと考えられる。

生成された単語の聴取実験による理解度の比較では、大きな差が見られなかった。この結果から、音韻の明瞭さについては、全体としては抑揚の有無の影響は少なかったといえる。ただし、本研究で用いた単語リストは、アクセントの位置が0型もしくは4型(LHHH)に限られていたため、抑揚の有無の影響を受けにくかった可能性もある。

本項目で述べた押圧センサによる抑揚のコントロールは、音声生成器だけでなく電気喉頭にも応用することができ、実際に、押圧センサによって抑揚をコントロールできる電気喉頭の試作器も開発中となっている(図17)。

C-2-b タッチパネル型PCへの対応について

ソフトウェアの改良により、“Windowsタッチ”の機能を導入することで静電容量式のタッチパネルを持つPC単体で、容易に操作盤面を制御可能とすることができた。

音声の生成実験で試用されたPCでは、操作盤面の画面上の大きさは、約 87 mm × 68 mm とし、筆者による簡単な音声生成実験を行った。擬似子音を加えない母音フィルタのみの音声出力では、ペンタブレットを使用した場合と同様に、母音、半母音に加えて「おはよう」等の簡単な単語を生成可能であった。

一方で、擬似子音を加えた場合には、最初の母音が短くなってしまいう傾向があった。例えば「おはよう」という語の場合には、初めの「お」の音が短くなり、その次の「は」の子音も短めになってしまい、「おっはよう」という声が出された。ゆっくりと操作した場合には、「おはよう」という出力を出すことが可能であった。この点については、試用したPCのハードウェアの性能のためである可能

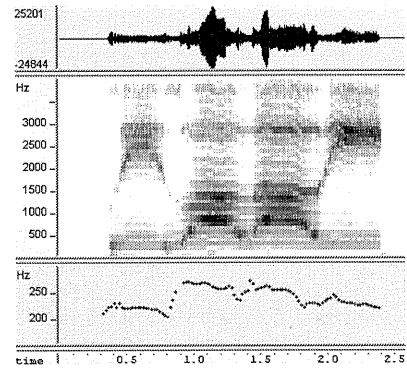


図 15 小指の握力による抑揚制御の音声 (上から,波形,スペクトログラム,ピッチ,時間)

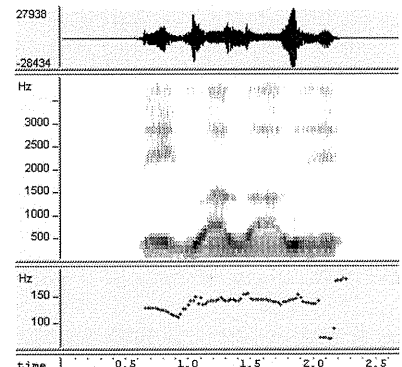


図 16 皮膚伝導マイクを使った音韻付加音声 (上から,波形,スペクトログラム,ピッチ,時間)

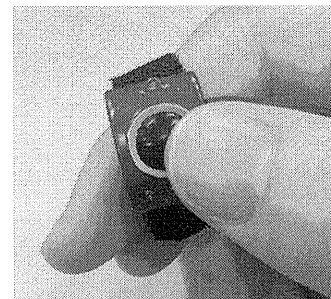


図 17 ハンズフリー電気喉頭の抑揚制御センサ試作器

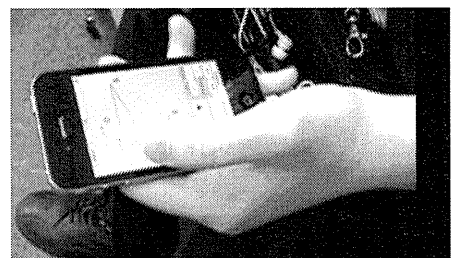


図 18 iPhone 上で動作する音声生成器

性があり、調査中である。

指で操作する場合にはペンの場合よりも手を動かす範囲が広くなり、腕を動かす必要が出てくるため、速い動きでの位置決めが難しくなることも

分かった。また、操作盤面上の指の滑り易さの検討や、操作盤面から指が外に出ないようにするためのガードの必要性が示唆された。

また、上記の実験とは異なるが、iPhone 上でも一部の機能を実行できるようなソフトウェアを開発できており、実用化に向けて準備を進めることができている(図18)。

D. 考察

ハンズフリー電気喉頭については、マイクに特定の指向性を行い、ハウリングの除去と電気喉頭からの直接音による雑音を低減できるように、拡声器を改良した。電気喉頭による音声の聴取実験では、再現音量を、改良型マイクでスピーカ位置を20cmに設定した状態でハウリングを起こさずに最大増幅できる音量に設定した。そのため、静かな部屋ではやや煩く感じる程度の音量であった。この結果から、改良型マイクによって音量を上げられることを示された。

しかし一方で、単語の了解度については、静かな部屋で実験をしたため、提示音が環境音に対して大きすぎた可能性がある。そのため、単語の了解度が、拡声機の有無では大きな差が出なかったものと考えられる。音量の増幅は、本来は環境雑音がある時に効果を発揮するものと考えられ、今後の評価項目として、環境雑音の条件も入れた実験を行う必要があると考えている。

また、拡声器については、マイクやスピーカだけでなく、それらの接続ケーブルや電池の持ち時間の管理、装着・脱着の手軽さといったことが、実用上かなり重要となることが、発話障害を持った研究者自身の一定期間の拡声器の試用から、明らかとなった。

そのため、次年度は拡声器全体の形状や装着感等も考慮し改良を加えて、臨床的な評価へ結び付けていきたい。

音声生成器については、音声生成の実験から、2つの方式によって、単語における抑揚コントロールが可能であることが示された。ただし、特に小指圧力による方法は、操作に慣れを要することが明らかとなった。その難易度には、個人差があると考えられるため、今後、複数の生成器による検証の必要がある。

また、今回行なった実験では、抑揚の有無による了解度には差がほとんどない結果となったが、文脈がある場合では、抑揚も聞き取りの手掛かりになる可能性があり、今後の検証が必要である。

音質については、聴取者から、小指の握力による制御の音声では明るい声、使用者の声を聞いた方式では低く暗い声であるという感想があった。

この理由としては、小指の握力制御では抑揚の制御幅が広がったために明るく感じられ、使用者の声ではゆっくりと低い声で音声生成され暗く感じられたことが挙げられる。聴取者からの意見として、声質がもっと高く明るい声が出ると良いのではないかという意見もあったため、発話の速度を補ったり、声質をある程度変更できるような機能が、今後必要となってくると考えている。

タッチパネル型PCへの対応については、改良によって本体の単体でもスムーズに音を出力できるようになったが、依然として、タッチパネルに触れてから音が鳴るまでの遅れ時間があつた。その結果、ゆっくりした音声の出力では、ペンやマウスを用いた場合と同等の音声を出力できたが、速い動きになると上手く生成されない傾向があつた。

この遅れ時間は、実験に使用したPC固有のものである可能性もあり、別な機種による調査を計画している。特に、指が画面に触れてから音が鳴りだすまでの時間と、指が画面から離れてから音が停止するまでの時間とにずれがあると、リズムが崩れてしまうと考えられ、避けなければならない。

また、タッチパネルを指で操作する場合、指の付け根から指の先端までの距離が、ペンによる操作の場合に比べて短くなるため、腕の動きが必要になることが分かった。このことから、腕の動きを少なくするために操作盤面を小さくする必要があることが示された。

タッチパネルによる操作では、操作盤面に自らがたどった軌跡を表示できるため、タッチした位置をすぐに確認でき分かりやすい。また、前述したが、接続の煩雑さが軽減されることの、携帯性への影響は大きいと考えている。

また、初めての使用者に対して、単語を出すための軌跡を画面に表示して、その上をたどらせるような、補助を可能にすることができ、操作方法の学習の助けとすることが期待できる。

本研究の音声生成器は、操作の方法が他に類を見ない者であり、母音を出すだけであればだれでも一瞬で操作できる一方で、上手く使いこなすには音素に対する知識や操作方法の習得が必要であるという特徴がある。そのため、今後は、初心者でも機器の操作方法を習得できるような操作マニュアルを作成していく必要がある。

E. 結論

電気人工喉頭については、その発声音について課題となっている、声の音量の問題と、人工喉頭からの直接音の雑音の問題について、適切な拡声器を開発することで問題を低減することを提案した。

具体的には、双指向性マイクを接話マイクに似た方法で用いることで、人工喉頭からの直接音によ

るノイズを低減すると同時に、マイクとスピーカが至近距離にある場合でもハウリングが起りにくいような拡声器を開発した。その結果、スピーカをマイクに向けて、マイクの下20cmの距離に置いた状態でも、ハウリングを起こさずに、約60dBAの声を約70dBAにまで増幅できた。

今後の課題として、環境雑音の大きい場合の各席の効果の検証と、拡声器自体の形状などの改良を行い、拡声機の量産・実用化へ結びつけていきたい。

また、音声生成器については、小指握力による押圧センサの操作によって抑揚を制御する方法と、使用者自身の声を皮膚伝導マイクロホンを通じて取得しホルマント合成の原音とする方法との2種類を導入し、簡易的な音声の生成実験と聴取実験とを行った。

実験の結果から、健常の発話の抑揚と同様の抑揚を各単語に付けられたことから、この2つの方式において音韻と独立して抑揚を制御が可能であることが示された。

ただし、音声生成の際の慣れや使いやすさには個人差があると思われるため、今後の検証が必要である。また、多様なアクセント型による音声の伝わりやすさや、文脈を持った場合の影響についても調べる必要があると考えている。

さらに、タブレットPCを用いたタッチパネルによる操作を可能にし、その動作状況について報告した。実験の結果から、タッチパネル特有と思われる反応遅れの問題やサイズ調整の必要性の課題があることが分かった。

音声生成器の次年度への課題として、抑揚を含めた実用に耐えうる実機での音声出力動作を実現できるように改良を行なうと同時に、初心者にもむけた使用マニュアルを作成し、評価を行っていくつもりである。

電気喉頭で用いられる技術と、音声生成器で用いられる技術や評価方法には、多くの面で共通の点がある。例えば、電気喉頭の拡声器のスピーカ部分は音声生成器の音声出力部に応用でき、音声成績の押圧センサをもちいた抑揚制御の方式は、電気喉頭の抑揚制御にも応用できると考えられ、現在、開発を進めている最中である。このような技術を双方で活かしながら、臨床的な立場から従来方式との比較や、従来方式との併用によってどこまで本方式が有用かの評価を進めていきたい。

F. 健康危険情報

該当するものは無い

G. 研究発表

1. 論文発表

1) Tohru IFUKUBE, "Sound-based Assistive Technology Supporting "Seeing", "Hearing" and "Speaking" for the Disabled and the Elderly", Proceedings of Key note speech ,pp. 11-19, InterSpeach 2010 Tokyo (2010.9)

2) 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—”電子情報通信学会、信学技報. SP, 音声 110(220), 41-46, 2010

3) 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—”電子情報通信学会、信学技報, SP, 110(452), 67-72, 2011

4) 藪 謙一郎、伊福部 達：“構音機能障害者のための音声生成器の抑揚制御方式に関する基礎的検討”，電子情報通信学会技術研究報告，電子情報通信学会，音声，111(225)，43-48，2011-10-06,2011

5) 藪 謙一郎，上田 一貴，稲永 潔文，伊福部 達 電気人工喉頭のための拡声器に関する一考察，電子情報通信学会、信学技報，SP，2011(161)，35-38，2012

2. 学会発表

1) 藪謙一郎、伊福部達”「なぞり」の入力による発話障害者支援のためのリアルタイム音声生成器 —阻害音と鼻音付加の操作方法の検討—”日本音響学会・3-4-14・pp.1623-1626・2011

2) 藪 謙一郎、伊福部 達：“タブレットPCのなぞり動作で操作する発話障害者支援リアルタイム音声生成器の試作”，ヒューマンインタフェースシンポジウム2011 論文集，ヒューマンインタフェース学会，CD-ROM，677-680,2011

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得

該当なし

2. 実用新案登録

該当なし

(補足資料A)

人工喉頭のための改良型マイクの 評価実験手順および結果

実験手順

本編で述べたように、評価実験は、人工喉頭で単語を読み上げた声について、拡声器を用いない場合(条件I)、未改良の例である無指向性マイクの拡声器Iを用いる場合(条件2)、改良型である双指向性マイクの拡声器IIを用いる場合(条件3)を、発話・聴取実験で比較した。その際、話者が話した声の質を3条件で一致させる必要がある。そのため、あらかじめ条件1～3のマイクからの音声を同時に収録した上で、聴取実験で各条件をシミュレートした音量で再生・提示する方法をとった。

・録音実験装置

音声の録音装置を図A-1に示す。実験用のあご台に3種類のマイクを固定した。話者はマイクに向かって、人工喉頭で単語を読み上げ、その声を録音した。その際、話者の口からマイクの距離を一定にするために、額をあご台上部へ当てて発話させた。

まず、マイク1として口元から約100mmの位置に無指向性マイクを設置した。このマイクは、拡声器を使わない場合の音を得るためのもので、人工喉頭からの直接音を含んだ音声を得られる。次に、マイク2として、マイク1と同じ無指向性マイクを口元から約20mmの位置に設置した。これは、拡声器Iのマイク音声を意図している。最後に、マイク3として、双指向性マイクを口元から約20mmの距離に設置した。これは、改良型の拡声器IIのマイクを想定したものである。

これらのマイクは、プリアンプ(AT-MA2: audiotechnica)とUSBオーディオインタフェース(UA-101: Roland)を介して、汎用パソコンにより44.1 kHz 16bitで多チャンネル録音した

なお、実験に用いた電気人工喉頭は、(株)電制製の「ユアトーンII ゆらぎ」を、ノーマルモード、音量“5”、ピッチ“5”の設定で使用した。

・録音手順

話者は人工喉頭の使い方を熟知する30歳代の健康男性1名とし、録音は防音室内で行なわれた。初めに、話者が人工喉頭を使って「あー」という声を出し、それを録音すると同時にその音量を2m離れた位置で測定した。その結果、約60 dBAであった。

読み上げる単語として、坂本らによる難聴者の

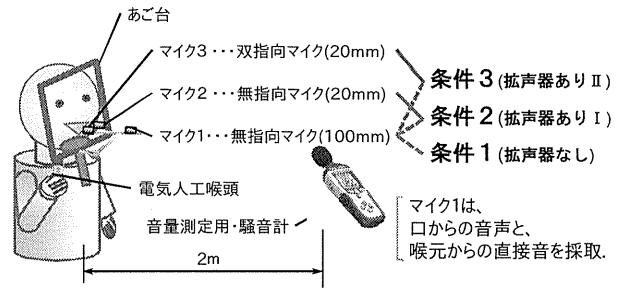


図 A-1 録音用マイクの設置位置

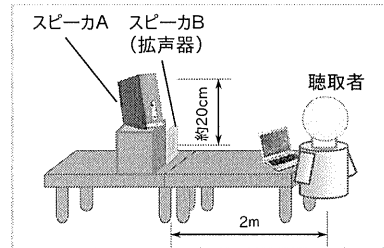


図 A-2 聴取実験装置

ための単語理解度試験用単語リストの中から、親密度7.0～5.5の単語50語を用いた。これらは、全てアクセント型が0型又は4型(LHHH)の、4モーラ音声からなる単語音声である。

話者は、録音の前に1時間程度、単語リストを見ながら聞き取り易い発音で話せるように練習をしたのち、前節で述べたようにマイクに向かって、単語リストに記載されている単語を順に読み上げ、その音声を録音した。読み上げは10語ごとに、数秒から十数秒の休憩をはさみながら行なった。

・聴取実験装置

聴取実験は、録音された音声を聴取者にランダムに提示して行なった。提示するスピーカの種類と位置を、図A-1に示す。スピーカA(BOSE Companion 2 series II)は、マイク1の録音音声を再生するためのもので、拡声器無しの場合の音声を再現するものである。また、スピーカ2は、マイク2またはマイク3の録音音声を再生するためのもので、拡声器から出される音を再現する。

すなわち、聴取実験の条件として、マイク1の音をスピーカAから再生する条件1、それに追加してマイク2の音声をスピーカBからも再生する条件2、マイク3の音声をスピーカBから再生する条件3の、3条件とした。それぞれの音は全て同期して同時に再生される。

スピーカAの音量設定は、録音時の音量とほぼ同値となるように、前節で録音された「あー」の音声再生時に2m離れた位置で58~60dBAとなる音量を設定した。

また、スピーカBの音量は、条件2と条件3とで異なる値を設定した。前節の「あー」を再生時にス

表 A-1 想定する拡声器と実験条件

	拡声器	マイク 1 ⇒スピーカ A (口の声+電気喉頭の 直接音を再現する)	マイク 2 ⇒スピーカ B (未改良マイク拡声器 を再現する)	マイク 3 ⇒スピーカ B (改良型マイク拡声器 を再現する)	混合音量 [dBA]
条件1	無	使用する	使用しない	使用しない	58-60
条件2	有(I)	使用する	使用する	使用しない	63-65
条件3	有(II)	使用する	使用しない	使用する	68-70

スピーカAとスピーカBとの混合音声が、距離2mの位置でそれぞれ、条件2で63~65dBA, 条件3で68~70dBAとなるように設定した。このスピーカBの音量設定は、各マイクの拡声器でリアルタイムの拡声動作時に、スピーカがマイクの下20cmの位置でマイクの方に向けて置かれた条件で、ハウリングを起こさない範囲の最大のゲインを再現した音量である。

・聴取実験手順

聴取者は30歳代の健常男性1名とした。実験は静かな部屋で行なわれた。聴取者は、スピーカから2mの位置に耳が来るように椅子に座り、一語ずつ提示音声を聴取した。

提示音声は、録音された50語について、条件1、条件2、条件3のもとで再生される150音声を、ランダムに提示するものとした。すなわち、聴取者はランダム化された150刺激の中で、同じ語を3回、異なる条件で聴くことになる。ただし、ランダム化の際には、同じ語が連続しないようにした。

聴取者には、提示される音が全て4モーラから成る意味のある単語であることを伝え、聞こえた単語が何であるかを聴取者の前に置かれたノートパソコンへキーボードからひらがなでテキスト入力をさせた。なるべく意味のある単語を記入することとし、どうしてもわからない場合は、わかる範囲で聞こえた音を記入するように教示した。音声は5秒おきに最大3回再生されるように設定し、少ない回数で入力できた場合は、聴取者が入力をした時点で次の後へ移るようにした。初めの4語は「あいうえお」「かきくけこ」「さしすせそ」「たちつてと」の音声を提示し、聴取者は聴取や文字入力の練習を行った。

以上の実験から、各条件の了解度をもって、人工喉頭の音声の評価とした。

実験結果

図A-3に結果を示す。50語中、完全に正答された単語数は、条件1では16語(32%)、条件2では20語(40%)、条件3では18語(36%)であった。150語全ての回答に要した時間は、約24分30秒で、1語あたり9.8秒であった。

また、単語に含まれる4モーラのうち、部分的に

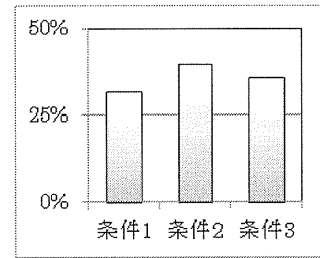


図 A-3 正答した単語数

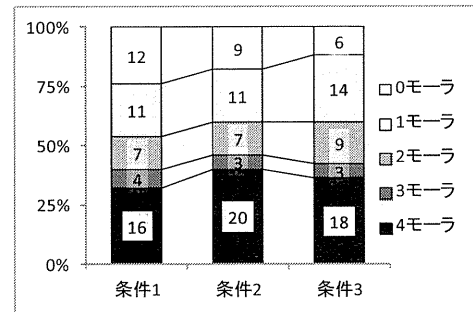


図 A-4 正答したモーラ数ごとの語数

正答していた場合の単語数を、正答したモーラごとにカウントすると、図A-4のようになった。

拡声器無しの場合である条件1と、拡声器を用いた場合である条件2、条件3を比較すると、拡声器を用いる条件のほうが拡声器無しの条件よりも正答数が多かった。被験者数が1名であるので断定はできないが、これら結果から、ある程度の拡声器の効果が見える。また、1モーラも正答しなかった単語数を比較すると、条件1に比べて条件3では約半数になっていることから、拡声器とマイクの効果によって、特に聴きづらかった音が、聞きやすくなったものと考えられる。

この実験から、条件3の音量を、ハウリングを起こしにくいマイク3で、スピーカ位置を20cmに設定した状態でハウリングを起こさずに最大増幅できる音量に設定した。そのため、静かな部屋ではやや煩く感じる程度の音量であった。単語の正当数が条件2よりも条件3のほうが低い結果となったのは、このことが影響したと思われる。音量の増幅は、本来は環境雑音がある時に効果を発揮するものと考えられ、今後、環境雑音の影響も考慮していきたい。

(補足資料B)

音声生成器の抑揚制御についての 評価実験の手順と結果

実験手順

本編で述べたように、音声成績の抑揚の付加方法について、(1)小指の握力を押圧センサから取得し抑揚制御パラメータとする方法と、(2)使用者の声を頸部に取り付けた皮膚伝導マイクロホンによって取得しホルマント合成の原音に用いる方法とを導入し、試作器による簡易的な実験を行った。

抑揚付加によって、抑揚のついた音声を十分に制御可能であるのかどうかと、抑揚付加によって音声の明瞭度が変化するかどうかを調べるため生成実験と聴取実験とに分けて実験を行った。

・音声の生成実験

生成は、本インタフェースの開発に関わり使用方法を熟知している1名(以下、生成者と呼ぶ)が行った。生成者は、左耳50dBHL、右耳30dBHL程度の聴力レベルであり、器質性および運動性の構音障害を持つ。手や指の運動機能は正常である。

対象とする単語は、先の研究で用いた100語の単語のうち、特に正答率が低かった単語26語を使用した。

生成者は、表記された対象の単語を見ながら、(1)抑揚無しの場合、(2)小指握力による抑揚制御の場合、(3)自身の声を用いる場合との3通りで、音声生成を行った。それぞれの単語の生成回数は自由とし、生成者自身が納得できる音声を生成できた時に、次の単語へ移るようにした。なお、疑似子音の付加機能については、(1)、(2)では有効、(3)の時には無効とした。生成実験(3)では、生成者は口を半開きの状態にして動かさずに、声を出すと同時にタッチパネルを操作した。また、子音部において、相応する/h/の摩擦音を発声した。なお、生成作業中の音声は、イヤホンで生成者へ提示すると同時に、PC上で動作する音声波形エディタWavesurferによって録音された。

・聴取実験

聴取実験は、録音された生成音1語ずつを聴取者へランダムに提示し、聴取者自身がPCのキー入力によって、聞こえた単語が何であるかを記入する方法で行った。聴取者は全員、聴力に関する大きな病気をしたことがなく正常な聴力を持つ20代から30代の男性4名とした。

音声の提示は、汎用PCに接続されたアームレスヘッドホンVictor HP-AL202を通して行われた。聴取者には、実験の前に提示音声日本語の4音節の単語であることを教示し、単語が推測可能な

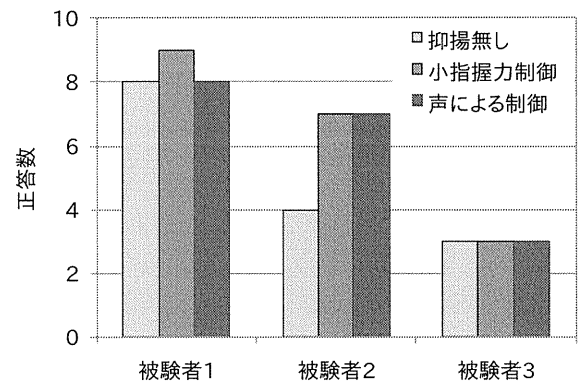


図 B-1 聴取実験結果

場合には、聞こえたままの文字を記入するのではなく、その単語を答えるように指示した。

実験結果

音声生成実験では、生成者により3つのパターンで所望の音声が生成了。いずれの単語でも、単語内での抑揚の位置は、通常の発話される単語と同様の出力がされた。小指圧力による制御では抑揚は強め、使用者の声による方式ではやや弱めの抑揚であった。音声生成の実験から、2つの方式によって、単語における抑揚コントロールが可能であることが示された。ただし、特に小指圧力による方法の難易度には、個人差があると考えられるため、今後、複数の生成者による検証の必要がある。

図 B-1 に聴取実験の結果を示す。抑揚の有無による正答数の違いをみると、3人の被験者のうち、1名では、抑揚無しに比べて抑揚を付加した場合のほうが高い正答率となった。しかし、他の2人では大きな差が見られなかった。この結果から、音韻の明瞭さについては、全体としては抑揚の有無の影響は少なかった。ただし、本研究で用いた単語リストは、アクセントの位置が0型もしくは4型(LHHH)に限られていたため、抑揚の有無の影響を受けにくかった可能性もある。また、文脈がある場合では、抑揚も聞き取りの手掛かりになる可能性がある。

また、聴取者から、小指の握力による制御の音声では明るい声、使用者の声を用いた方式では低く暗い声であるという感想があった。この理由としては、小指の握力制御では抑揚の制御幅が広がったために明るく感じられ、使用者の声ではゆっくりと低い声で音声が生成了暗く感じられたことが挙げられる。

研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
伊福部達 他	8.2システムの評価と設計	日本バーチャルリアリティ学会	バーチャルリアリティ学	工業調査会	東京	2010	323-336
伊福部達 他	VI 高齢者支援機器・技術、第7章2 将来技術 IT技術－「見る」、「聞く」、「話す」機能を支援するIT、	大内 尉義 秋山 弘子 折茂 肇	新老年学 第3版	東京大学出版	東京	2010	2099-2104
伊福部達	伊福部達教授が語る、長寿社会の「テクノロジー」－長寿社会を明るくする福祉工学－	東京大学高齢社会総合研究機構	2030年超高齢未来	東洋経済新報社	東京	2010	100-107
伊福部達 (監修) 筒井 信介 (著)	第6章 福祉工学が秘める可能性	有限会社イデア	ゴジラ音楽と緊急地震速報、あの警報チャイムに込められた福祉工学のメッセージ	ヤマハメディアミュージック	東京	2012	157-162

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Tohru IFUKUBE	Sound-based Assistive Technology Supporting “Seeing”, “Hearing” and “Speaking” for the Disabled and the Elderly	InterSpeech	2010	11-19	2010
萩謙一郎 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－音声改善のための基礎的検討－	電子情報通信学会技術研究報告, SP, 音声	110(220)	41-46	2010

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 -	電子情報通信 学会、信学技 報, SP	110(452)	67-72	2011
藪謙一郎 伊福部達	構音機能障害者のため の音声生成器の抑揚制 御方式に関する基礎的 検討	電子情報通信 学会技術研究 報告	111(225)	43-48	2011
藪謙一郎 上田一貴 稲永潔文 伊福部達	電気人工喉頭のため の拡声器に関する一考察	電子情報通信 学会技術研究 報告, SP	2011(161),	35-38	2012
伊福部達	「見る」「聴く」「話 す」を助ける放送技術 ～感覚のナゾ解きから 生まれたモノ～	NHK技研R& D	No.123	36-47	2010
伊福部達	ふぉーかす 指で見る「夢」	映像メディア 情報学会誌	Vol.64 No.11	巻頭言	2010
伊福部達	「視聴覚・発声障害 のためのバリアフリー 技術」に際して(巻頭言)	ライフサポー ト学会誌	Vol.22 No.4	2	2010
伊福部達	福祉工学への夢 人の耳をだまらず物まね 鳥たち	ミネルヴァ通 信	2012 No.10	29-32	2012
伊福部達	超腹話術のナゾ解きか ら生まれたモノ	ミネルヴァ通 信	2012 No.11	29-32	2012

研究成果の刊行物・別刷



Sound-based Assistive Technology Supporting “Seeing”, “Hearing” and “Speaking” for the Disabled and the Elderly

Tohru Ifukube¹

¹Research Institute for Advanced Science and Technology, University of Tokyo, Japan

ifukube@rcast.u-tokyo.ac.jp

Abstract

With a rapid increase of a population rate of the elderly, disabled people also have been increasing in Japan. Over a period of 40 years, author has developed a basic research approach of assistive technology, especially for people with seeing, hearing, and speaking disorders. Although some of the required tools have been practically used for the disabled in Japan, the author has experienced how insufficient a function of the tools is for supporting them. Moreover, the author has been impressed by amazingly potential ability of the human brain that compensates for the disorders.

In this report, the author shows some compensation abilities formed by “brain plasticity”, and also shows extraordinary ability such as voice imitation of mynah bird and obstacle sense of the blind. Furthermore, the author introduces six assistive tools borne by solving mystery of the compensation function and the extraordinary animal. Finally, the author emphasizes that these assistive tools will contribute to design a new human interface that may support the elderly as well as the disabled.

Index Terms: assistive technology, the blind, the deaf, speech disorders

1. Introduction

As shown in figure 1, nowadays, around 1/4 of population are elderly in Japan, so actually near 30 millions people are more than 65 years old. With the rapid increase of the elderly, disabled people who need care are also increasing. Actually, in 2005, the number of the disabled was beyond 6 millions. Most of them have disability regarding hearing, speaking, reading, thinking and moving functions.

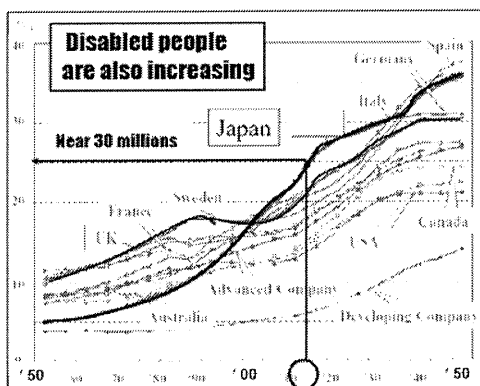


Figure 1: Population rate of the elderly from 1950 to 2050.

In this situation, assistive technologies should be urgently designed for supporting these people in addition to medical side technologies. However, the disability is so diverse and

complex that it has been difficult to construct a research approach of the assistive technology.

Our research approach has three steps as shown in figure 2. The first step is to analyze the human function such as sensory, brain and motor functions based on neuroscience and cognitive science. The second step is to design various assistive tools based on the findings obtained by the basic research. If the assistive tools are insufficient for supporting the disabled, the research should go back to the first step. The third step is to apply the basic findings and the assistive tools to human interface systems such as robotics and virtual reality systems. The third step is important to open a big market and to make a price of the tools cheap [1], [2].

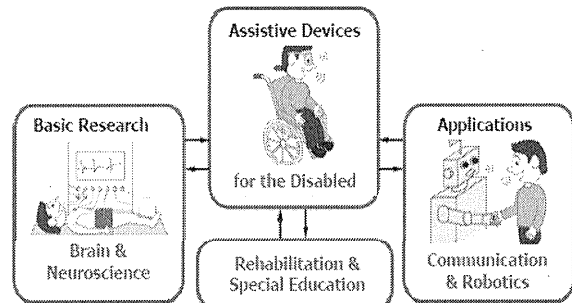


Figure 2: Three steps of our assistive technology research.

The assistive technology for elderly disabled people is mainly called “Geron-technology” that is somewhat different from “Barrier-free design” that includes supporting young disabled people. As shown in figure 3, the barrier-free design for the young should positively use “plasticity” of the human body, especially the brain because residual functions work to compensate the disordered function by the help of brain plasticity.

On the other hand, in general, the plasticity function decreases in the elderly, whereas, they will acquire abilities using their “experience”. The acquired experiences should positively be used in the geron-technology. The following examples of six assistive tools, which we have designed, were mainly designed for the young disabled.

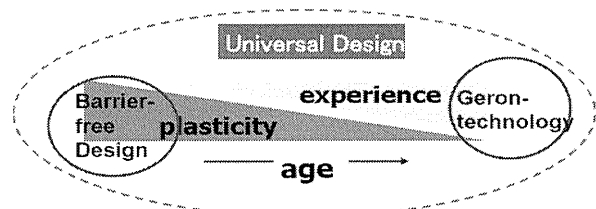


Figure 3: Barrier-free design and Geron-technology.

2. Assistive tools supporting “Hearing”

To assist “hearing” for auditory disorders, there are three approaches as shown in figure 4. One is to convert sound signals into tactile stimulations, the second is to electrically stimulate the surviving auditory nerves, and the third is to convert speech signals into letters. Two examples of assistive tools for hearing impairments are shown in this section.

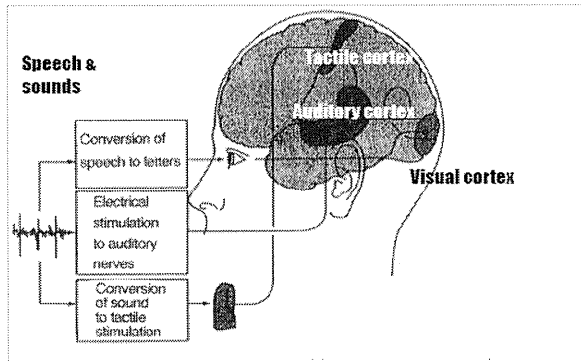


Figure 4: Three research approaches of assistive tool design for the hearing impaired.

2.1. Tactile communication aids for the deaf and/or the blind

The first topic is an assistive tool called “tactile aid” that is the author’s first research started from 1972. Tactile information is combineant for suppoting hearing as well as seeing because it can be used together with the visual and/or hearing sense. Through fundamental research regarding similarities between auditory and tactile information processing, we developed a prototype tactile aid named “tactile vocoder”, as shown in figure 5, in 1975.

The tactile vocoder is a device which produces sound spectral patterns of which frequency analysis method is modeled after cochlear mechanism. In the tactile vocoder, spectral patterns of speech sounds were divided into 16 frequency components and each component was transmitted to lateral inhibition circuit in order to sharpen the spectral patterns. Finally, the 16 components were converted into 16 vibratory stimuli of which level of the is corresponding to each intensity of the 16 components.

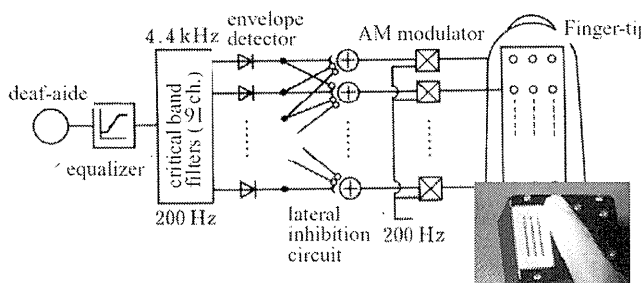


Figure 5: Block-diagram of tactile vocoder and a vibrator array (3x16).

When an index fingertip touches a vibrator array that was composed of 16 rows and 3 columns, the device made it possible to discriminate the first and the second formants of vowels as well as some consonants such as fricatives, semi-vowels and nasals [3].

The device was used for a telephone communication aid at a deaf school after deaf boys learnt tactile patterns corresponding to several words such as “hello”, “yes”, “no”, and etc. Although the device was manufactured in Japan about 30 years ago, it has not been widely used because nobody knows whether or not the tactile information is associated with speech understanding area in the brain.

After 25 years ago of the tactile aid research, the author moved to the University of Tokyo to attend at barrier-free project. The project has been conducted by a professor Satoru Fukushima who became blind at the age of 9 years old and deaf at 18 years old. From a viewpoint of the deaf-blind, he has suggested how we should perform the barrier-free research. He ordinarily communicates with us using both the tactile sense of his 6 fingers by a help of interpreters as shown in figure 6(lower left). This tactile communication method is called “finger Braille” that he first investigated in the world.

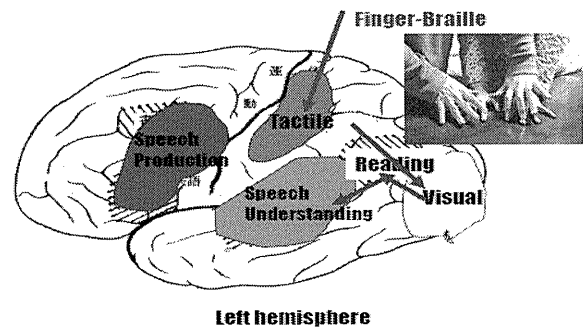


Figure 6: Finger-Braille information routs acquired by “plasticity” of the human brain.

From recent neuroscience researches, it is anticipated that the finger Braille patterns would be reached to the visual cortex through the tactile cortex and then the patterns would be associated with language understanding cortex as shown in figure 6. By the “plasticity” of neural network in the human brain, the lost functions might be compensated by the other sensory cortex. Actually, in 1998, by using a functional MRI, Swedish brain researchers showed that tactile stimulation activates the auditory cortex of the acquired deaf [4]. Furthermore, it has also been found that the visual cortex is activated by Braille stimulation on a fingertip of acquired blind subjects [5].

Encouraged by a professor Fukushima and the neuro-science results, we re-designed a new model of the tactile vocoder in 2006.

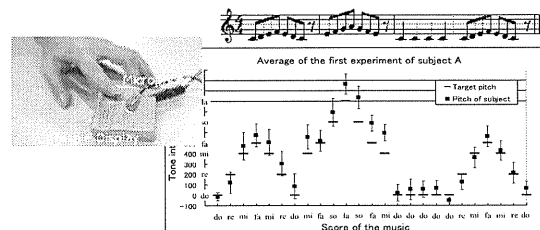


Figure 7: Left: New tactile vocoder made of a microprocessor and a vibrator array. Right: Pitch pattern of a deaf-blind produced by a “Frog’s song”

The new device consisted of a DSP and a piezo-electric vibrator array so that the signal processing can be changed only by software. For the deaf-blind use, the new device was set to transform a voice pitch frequency into musical scale that corresponds to vibrating point of a fingertip.

A lady, who has been deaf-blind since she was 40 years old and is now 67 years old, attended at our experiment. As she was a teacher of Japanese musical instrument and folk songs until she lost her visual and auditory senses, we expected that she can easily handle the tactile pitch display and can sing songs by using a feedback of the songs through her fingertip. Actually, after she learnt musical scale using the tactile device for short time, she could sing some Japanese songs as shown figure 6. This example shows that the tactile information might help to get a feedback of melodies of the songs [6]. However, congenital deaf-blind people were not able to sing any song.

These findings give us many suggestions to analyze whether or not the tactile information could be transmitted to auditory cortex and speech understanding area in the brain. It is ascertained how brain and cognitive researches are important to design assistive tools.

2.2 Captioning system for the hearing impaired

The second example is a voice recognition system to assist listening of the acquired deaf who have lost their hearing when they became adult. We were asked to design the voice typewriter by the acquired deaf group in 1975.

Fortunately, all Japanese voices are represented as a series of 68 monosyllables (/a/, /ka/, /sa/, /ta/, /na/,.....) each of which consists of 5 vowels and 14 consonants and represented by one of Japanese Kana. This simplicity makes easy to design a Japanese voice typewriter that converts each monosyllable into a corresponding Kana.

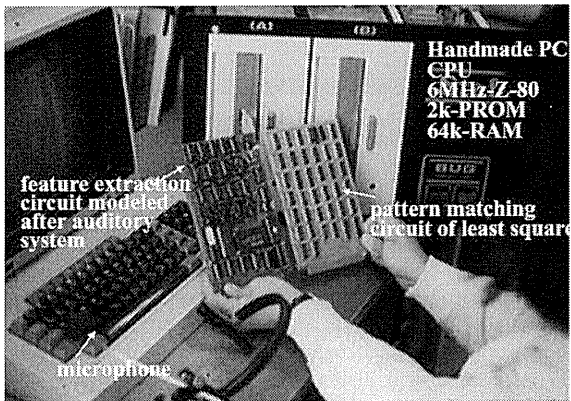


Figure 8: Monosyllabic voice typewriter using a Z80 microprocessor and 32kbyte memories.

However, Japanese sentences are represented by combining Japanese Kana and Chinese characters that have different pronunciations. This complexity makes difficult to design a Japanese voice typewriter.

In 1976, we designed a monosyllabic voice typewriter system using a Z80 microprocessor and 32kbytes IC memories, as shown in figure 8, which was used together with a Japanese word processor so that Japanese monosyllabic voices are automatically converted into both Japanese and Chinese letters.

Our voice typewriter could recognize separately pronounced 68 Japanese monosyllabic voices into Kana at a correct recognition rate of 96% in the case of well trained speakers. However, as the voice typewriter system was too expensive for deaf users to buy, it was only used for an input device of printing machines in 1977 [7], not for an assistive tool for the deaf.

In 2001, 25 years after of the voice typewriter research, we were asked to design a captioning system by DPI (Disabled Peoples' International) conference held in 2002. As about 3000 disabled people from more than 100 countries attend at the conference, we had to design the captioning system that can automatically convert various languages into both Japanese and English captions. As the speech recognition technology was not perfect, we expected an ability that the hearing impaired can often guess the correct meaning of spoken sentences by observing a speaker's face and lip-movement.

With this in mind, we designed the captioning system in such a way that both the series of letters and the speaker's face simultaneously appear on a large screen. Where, we used a commercially available speech recognition software "Via Voice". Furthermore, we adopted "re-speak" method that the speaker's speeches are sent to a well trained re-speaker who repeats the speeches. In our system, the recognized outputs were checked by humans and then the corrected letters were displayed on a screen. The system was first used at a pre-conference in 2001 and then at the DPI in 2002. Both Japanese and International sign languages, both Japanese and English captions as well as speaker's face were displayed on a same screen as shown in figure 9.

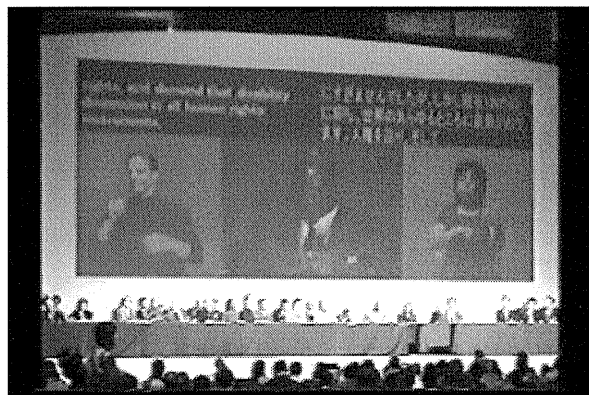


Figure 9: Captions (English and Japanese) presented on a screen together with sign languages and a speaker's face at DPI conference in 2002.

From the analysis of captioning results, the accuracy of the caption was about 98% and captioning speeds were 4 sec from English to English, 11 sec from Japanese to Japanese, 12 sec from Japanese to English and 17 sec from English to Japanese. After the use at DPI, we investigated whether or not such non-verbal information as a speaker's lip-movements and facial expressions as shown in figure 10 could improve the comprehension of spoken sentences that contain incorrect words.

From the results of the speech comprehension, the improvement was observed only when the facial expression was displayed roughly one second after the incomplete presentation of a speaker's words. This fact was indirectly

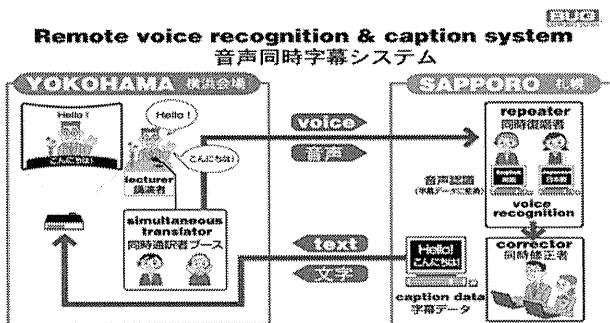
proven by an experimental result that eye balls of the deaf subject mainly stayed on the captions, not on the speaker's face [8].



Figure 10: left: Caption only, center: Caption + face, right: Caption + lip-movement.

No improvement was obtained for people with normal hearing. It is ascertained that combining both incomplete verbal with non-verbal information is indeed significant in facilitating comprehension for the hearing impaired .

After the DPI, the caption system was used in the various conferences in Japan through internet (ISDN). For example, at International Conference of Universal Design held at Yokohama in 2002, both English or Japanese speeches picked up at Yokohama were simultaneously interpreted, and then the interpreted speeches were sent to Sapporo where the speeches were converted into both Japanese and English text information which were displayed on screens at Yokohama as shown in figure 11.



Powered by E.U.G., Inc. / Hokkaido Univ. / Tokyo Univ.
共同開発：(株) ビー・ユー・ジー、北海道大学、東京大学
2002. 11. 30 : 国際ユニバーサルデザイン会議2002 (ロマンフィロ横浜)

Figure 11: Captioning system and mobile phone (lower right) using internet

Furthermore, we tried to apply the captioning system for communication aid for mobile phone in 2004 in order to be used by general people. These experiences suggest that the assistive tools should be widely used as possible as we can not only for deaf users but also for general users in order to make the system cheap and better model.

3. Assistive tools for speech disorders

Speech disorders are roughly divided into three causes shown in figure 12. The first is a caused mainly by laryngeal cancer. The laryngeal cancer patients have sometime lost their larynx by surgical operation, losing their vocal folds. Therefore, they become “laryngectomee” who lost a sound source for speech. The second is articulation disorders who are difficult to control their speech organ such as tongue, jaw and lips. The third is an aphasia that is caused by a disorder of brain nervous system in Broca area in the cortex.

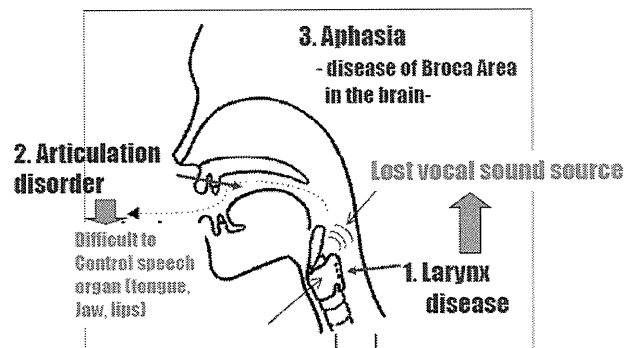


Figure 12: Three causes of speech disorders

3.1. Artificial electro-larynx for laryngectomee

An electronic artificial larynx is one of substitute vocalization methods. This method applies a vibrator to the lower jaw, which sends a vibrating sound into a voice tract through soft subcutaneous tissue. We focused on the electro-larynx because this method can be easily learned, although the voice quality is very poor like a buzzer sound.

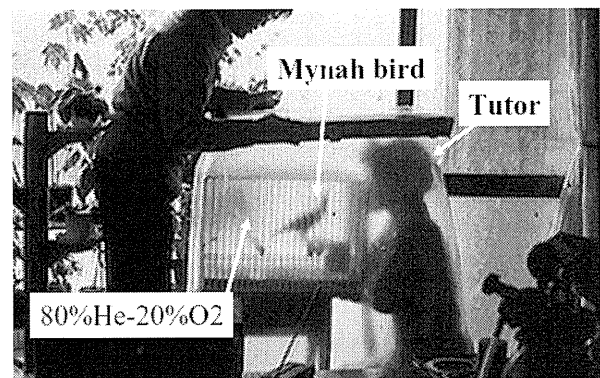


Figure 13: Mynah bird and her tutor in He-O2 gas

The ability to generate voice sounds is not limited to human beings. Among the animals which can generate voice sounds are certain birds, including mynahs, parrots, and parakeets. For researchers of speech assistive technology, it is a mystery as to why these birds, whose mouths and ears are shaped in a totally different way from humans, are capable of distinguishing and vocalizing certain human words. Especially the imitations of a mynah's voice sounds quite smooth and natural. Therefore, a research of the vocalization mechanism of mynahs might lead to new ideas for improvements of the electro-larynx.

From an analytical study of the mynah bird's voice imitations using He-O2 gas, about 30 years ago as shown in figure 13, we found that the mynah's imitations in the He-O2 gas are heard as the same as in normal air. Since it was reported that mynah bird has two sound sources (syrinx) corresponds to a human vocal fold, it was ascertained that he produces speech signals by combining the two sounds sources. Therefore, a formant pattern of the mynah was quite different from the human's formant pattern. We also found the reason why the mimicking of the mynah can be heard as natural voices is because she is able to faithfully imitate the fluctuation and intonation of the human voice [9].