

201027021A

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22年度 総括研究報告書

研究代表者 伊福部 達  
平成23(2011)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----- 伊福部 達	1
II. 研究成果の刊行に関する一覧表 -----	10
III. 研究成果の刊行物・別刷 -----	12

## 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

研究代表者 伊福部 達

### 研究要旨

代表者らは、過去10年間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が、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。

本年度は、ハンズフリー電気喉頭については試作済みの装置の音量を拡大するために、マイクロホン、アンプ、スピーカが一体化した小型拡声器（縦8cm×横5cm×厚さ2cm）を2種類開発し、胸ポケット装着型のプロトタイプ装置の試作により、その使い勝手を評価した。その結果、ポケットや首の動きがあっても口元のマイクロホンやアンプの位置ずれないような固定方法が重要であることが分かった。また、首バンドに拡声器そのものを取り付ける「一体型」の有用性が再確認された。

一方、発話支援音声生成器については、「母音」や「半母音」だけでなく/サ/などの「摩擦音」や/ナ/などの「鼻子音」を含む全て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た。摩擦音の音源として乱流音を生成する処理を加え、パッドを指やペンでタッピングすると生成されるように設定し、乱流音声後に指やペンで母音の位置へ移動させることで多くの摩擦音を生成でき、かつそれを聴取した時の認識率も高くなることが分かった。

以上の2点について、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

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  
東京大学先端科学技術研究センター 特任助教

### 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。

## B. 研究方法

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、過去の実績・経験から以下の3点に課題を解決しなければならないことが分かっている。それらは、（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気センサの開発、（3）音量が大きく音漏れがしない小型拡声器の開発の3点である。

ただし、22年度は、この3課題の中でも、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。なお、モニターとしては一昨年からお願している三笠宮寛仁殿下に引き続き担当して頂き、東大・先端研および赤坂御用邸で評価試験を進めた。なお、その際、倫理規定を十分に考慮し、試験を実施した。

一方、構音支援器については、過去に評価研究から、以下の3点が検討課題であることが分かっている。すなわち、（イ）連続発声音の語頭が摩擦音や破裂音の場合にはその子音の認識が困難であること、（ロ）/ナ/や/マ/のような鼻子音を考慮していなかったことから、鼻子音生成機能を導入する必要性について検討す

ること、（ハ）音声の強度と高さ（ピッチ）を指先の動きや指圧で付ける使い勝手の良い方法を見るけることの3点である。

本年度は、指なぞりパッド上をタッピング（軽く叩くような動作）すると乱流音が生成されるように工夫し、それによる音声明瞭度の改善度と使い勝手の観点から、乱流音付加の有用性を調べた。同様に、鼻子音についても鼻音生成の領域をパッド上に設けて鼻音を再現できるようにし、その有用性を評価した。

さらに、これらの発話支援方法が他の治療法やリハビリ方法とどのような補完的な役割を果たすかを評価した。その評価については臨床的・認知科学的な立場から分担研究者の田中と上田が行った。以上の要素技術の成果に基づいて、代表者の伊福部は支援技術を一般ユーザが利用できるように拡張する道を探り、その場合の市場性などを総合的に評価し、その実用性を考察した。

なお、倫理面の配慮については、代表者が4年前からNIHK技術総合研究所の倫理審査委員を務めているが、その時に作成した「ヘルシンキ宣言に基づいた倫理規定」を発話支援機器評価試験に一部変更し、それを基に評価試験を実施した。

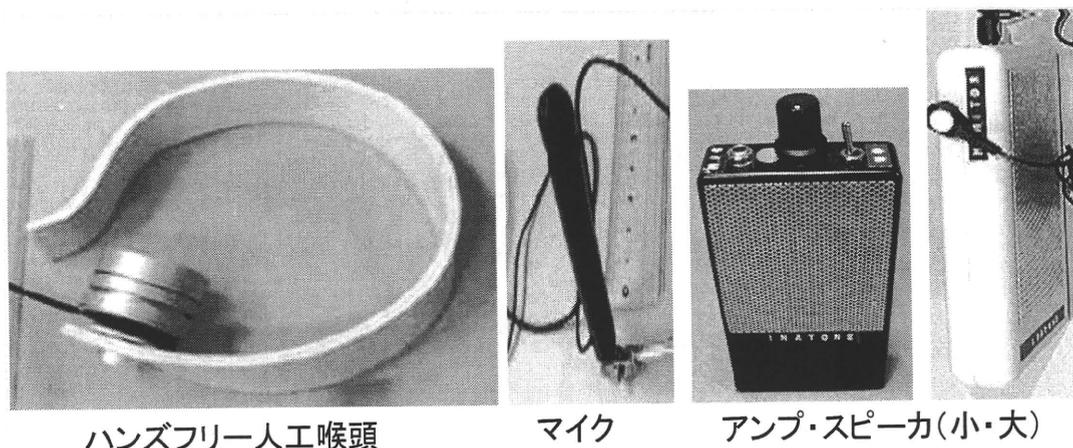


図1 ハンズフリー人工喉頭(左図)と高指向性マイクロホン(中央)およびアンプ本体(右図)

### C. 研究結果

#### <ハンズフリー電気喉頭のための拡声器>

ハンズフリー電気喉頭については試作済みの装置の音量を拡大するために、マイクロホン、アンプ、スピーカが一体化した小型拡声器（縦8cm×横5cm×厚さ2cm）を2種類開発し、胸ポケット装着型のプロトタイプ装置の試作により、その使い勝手を評価した。なお、電気喉頭音声の発生する部位は自然発生音と異なり、頸部からも多くの音が出ていることが浮き彫りにされたことから、マイクロホンについては、口元からの発生音のみを検出するものを開発した。そのため、音の指向性が極めて高い（10度前後）

マイクロホンを開発し図1はハンズフリー人工喉頭（左図）と高指向性マイクロホンおよびアンプ（中央）を示している。図1の右図は開発したアンプ本体（2台）である。

開発者が評価した結果、音質・音量ともに従来のものよりはるかに改善されることを確かめた。その上で、当事者による使い勝手の評価を試みた。また、図2は被験者（殿下）が既存の人工喉頭と一緒にマイクロホンを持ち、音声増幅用アンプを胸ポケットに入れて使用している様子である。これを2-3日付けて使用して頂いた時の使い勝手の評価は以下のようであった。

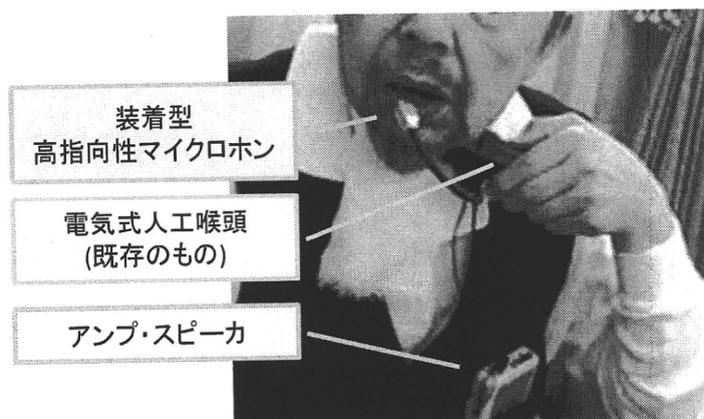


図2 マイクとアンプ・スピーカを使用した発話の様子

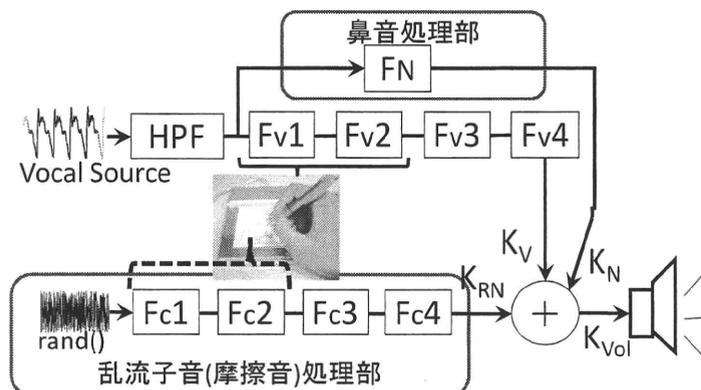


図3 乱流子音(摩擦音)と鼻音の処理部を付加した音声生成器のブロック図

- ①日常、公邸ではポケットの無い服装を着ているので、人と話をしているときに常にスピーカを持たなければならないので疲れる。
- ②ポケットのあるベストや上着を着ているときはスピーカは問題にならないがバイブレータとマイクがピタッとこないことが多いことから、度々当てる位置を変えねばならず、単独でバイブレータを使った方が便利。
- ③客をもてなす大応接ではポケットのある服を着る機会が多いが、上記の理由のため客は聞き取りにくいことが多いらしく、その都度、バイブレータの首に当たる部分と、マイクロホンの口に当たる部分を微調整しなければならないので、結構めんどくさい。

- ④残念ながら、現在の新機器は、市販のヘッドフォン式スピーカと殆ど変わらない、つまり、市販のものは、今でもパーティー等で、飲み食いをあきらめてスピーカを人に向けてしゃべる時使うが、それと、今の機材はそれほど使用上の違いが無い。
- ⑤やはり完全フリーハンド式、つまり首に輪を掛け、バイブレータとスピーカの双方をそのバンドに装着するタイプの物に直接行った方が良いと思う。

以上から、衣服や首の動きがあっても口元のマイクロホンやアンプの位置ずれしないような固定方法が重要であることが分かった。さらに、首バンドに拡声器そのものを取り付ける「一体型」の有用性が明らかとなった。

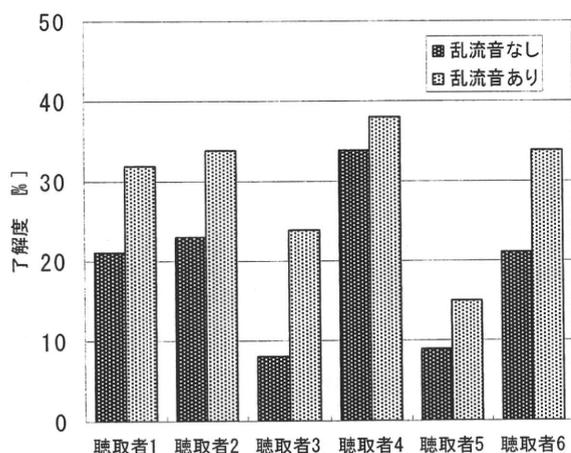


図4 乱流音処理の有無と生成音了解度

#### <発話支援用のための音声生成器>

一方、発話支援音声生成器については、「母音」や「半母音」だけでなく/サ/などの「摩擦音」や/ナ/などの「鼻子音」を含む全て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た（図3）[文献 2-1) 2-2)、3-1)]。摩擦音の音源として乱流音を生成する処理を加え、パッドを指やペンでタッピングすると生成されるように設定し、乱流音声後に指やペンで母音の位置へ移動させることで多くの摩擦音を生成でき、かつそれを聴取した時の認識率も高くなることが分かった（図4）。ただし、/ta/や/pa/の「破裂音」ように乱流音が短時間で終わる子音の場合には、タッピングのタイミングをずらす方法である程度実現できることが分かった。連続音声の認識実験から、摩擦音や破裂音が語頭にあるときにはその明瞭度が大きく改善されることが分かった。ただし、操作の煩雑性を避けるために有声子音のための処理

を除いたことから、有声破裂音等には不明瞭さが残った。

一方、鼻から抜ける「鼻音」についても、パッド上にそれに相当する鼻音音源を設け（図3、図5）、そこから指やペンをスタートさせると鼻子音が発生するように改良した。ただし、聴取実験を行った結果から、音声処理だけに着目すれば、連続音声を生成した場合にはその明瞭度は鼻音処理の有無に大きくは依存しないことが分かった（図6）。一方で、パッド上の鼻音の領域を示した場合と示さなかった場合とを比較すると、鼻子音の直前や直後の音声の音響的特徴の再現性が向上していたことから、操作の助けとして有効であると判断した。

また、上記で述べた評価実験は、分担研究者のPTである田中と認知科学専門の分担者上田が行った。その際、他の治療方法やリハビリ方法との補完的な役割等を考慮しながら臨床的・認知科学的な立場から評価を勧めた。なお、「母音」と「半母音」のみを生成する開発済みの音声生成器はスマートホンの一つであるiPhone上で操作できるようにし、そのソフトをインターネットを介してダウンロードできるようにした。

さらに、他の治療方法やリハビリ方法との補完的な役割等の評価については臨床的・認知科学的な立場から分担研究者のPTである田中と認知科学専門の分担者上田が行った。以上の要素技術の成果に基づいて、代表者の伊福部は利便性、市場性など総合的な立場から、その実用性を評価し、拡声器等の実用性・利便性を確認した。

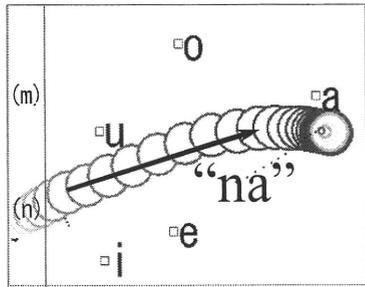


図5 パッド操作面上へ鼻音の配置

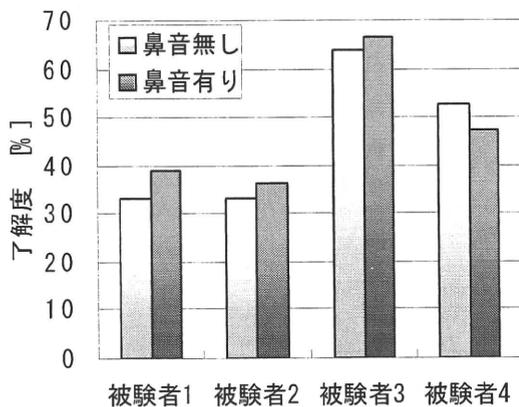


図6 鼻子音処理の有無と生成音了解度

#### D. 考察

本年度は、ハンズフリー電気喉頭の課題である①小型拡声器、②薄型軽量振動子、③抑揚制御用呼気センサの内、①に絞り込んで研究開発とその評価を行ったが、アンプ本体やマイクロホンの性能そのものは大きく改善されたが、装着方法や使い勝手に多くの問題点が残された。現在の電池の技術等を考えると、十分な音量の出せる小型拡声器を開発することは限界があることから、音量が小さくても首バンドに装着できるものを優先すべきであるという結論に

至った。今年は当事者として三笠宮寛仁親王殿下に絞り込んで徹底的に評価して頂いた。しかし、当事者の職業、嗜好、立場などによって、求められる拡声器の大きさや性能が異なる可能性があることから、他の当事者に使い勝手の評価を依頼し、さらに理学療法士（PT）である分担者田中敏明、認知科学者の上田分担者および聴覚言語士（ST）による客観的な臨床評価を進める必要がある。

また、②と③については、人工喉頭「コアトーン」を販売している企業とその共同研究者が既に試作済みのあるものが、それを基礎にしてハンズフリー用に改良し、さらに複数の当事者と上記の医療系専門家による評価を行い、①、②、③が一体となったもので総合的に評価することが重要であると考えている。以上のことを踏まえて、来年度は一体化と総合評価に力点をおいて研究を進める。

一方、発話支援音声生成器では、全ての子音を出せるようになったが、その分、操作が複雑になり訓練に要する時間もかかってしまう恐れが出てきた。連続音声として認識率が下がらない範囲で操作の負担を軽減する研究の必要性がある。同時に、訓練マニュアルを作るとか、インターネット上で訓練の仕方を指導するなどのシステムも必要になる。また、感情表現を豊かにするための抑揚制御についても使い勝手の観点から改良の余地が十分にある。さらに、それ以上に、出力音声聴き手に十分に伝わるように音量を拡大する技術の必要性が表面化している。とくに、拡声器は何処でも何時でも使えるようなウェアラブル機器にするためには避けて通れない課題である。

一方、現在、全ての子音が生成できるようにした改良型をアンドロイド上でもそのソフト

をダウンロードし、希望者は誰でも利用できるようにしている。これにより、自分の持っているスマートホンの表面をなぞることで、実時間で音声を生成できるようになるが、そのユーザによる使い勝手の評価をもらい、その後の改良の手がかりとする。臨床評価についてもPTやSTの協力により客観的に行い、実用化のための課題を明確にし、ネットワークを介して利用者の意見を吸い上げ、それを基に利用者自身が改良化できるような方法も考えるべきである。

さらに音量を拡大する方法、抑揚を付ける方法など両機器に共通する重要な課題が浮き彫りにされたので、来年度からは、特に、この共通課題に焦点を絞りたい。また、マーケティング等を通じて市場性を把握したうえで、発話障害を支援する技術の実用化を検討すべきであると考えている。

## E. 結論

発話障害者のための支援技術である「ハンズフリー電気喉頭」と「ウェアラブル音声生成器」を試作し、その評価を行った。前者では、電気喉頭の音量拡大のための高指向性マイクロホンと胸ポケットに入る小型のアンプを開発し、その評価を行った。その結果、従来に比べて音質・音量ともに大きく改善されたが、マイクロホンやアンプ本体には自然音声の拡声器とは異なった機能と使い勝手が要求されることが分かった。この新しい課題を定量的かつ臨床的な立場から評価した上で、どの程度小型でどの程度の音量が最適かを定める手掛かりが得られた。

また、音声生成器では全ての子音を指やペンの動きで生成できるように改良され、さらに、それらのプログラムはインターネット上で有

料でダウンロードできるようにし、ユーザから生の意見がフィードバックされるようにした。ただし、同時に操作が複雑になり、訓練による時間が増えるので、訓練マニュアルの必要性とともに、その程度の操作性で認識精度が得られるかという課題が残された。どうじに、電気喉頭と共通する拡声器、感情表現のための抑揚制御をウェアラブル機器にどのように組み込むかという課題も残された。

以上、共に、技術的には大きく改良されたものの、多機能・高性能と使い勝手がトレードオフしている状態であり、また拡声器、抑揚制御法など残された課題も多い。来年度からはこれらの残された課題を解決しながら、実用化を想定した価格設定なども行っていきたい。

## F. 健康危険情報

該当するものは無い

## G. 研究発表

### 1、著書

伊福部達：（分担執筆）“バーチャルリアリティ学 日本バーチャルリアリティ学会編” 8.2 システムの評価と設計，工学調査会，pp. 323-336 (2010)

伊福部達：（分担執筆）新老年学 第3版，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、2.1「見る」機能を支援するIT、2.2「聞く」機能を支援するIT、2.3「話す」機能を支援するIT、2.4 IT応用の将来，東京大学出版，pp. 2099-2104，(2010)

伊福部達：（分担執筆）“伊福部達教授が語る、長寿社会の「テクノロジー」ー長寿社会を明るくする福祉工学ー”、「2030年超高齢未来」、東京大学高齢社会総合研究機構，東洋経済新報社，pp. 100-107（2010. 12. 9）

## 2. 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, InterSpeech 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 ー子音改善のための基礎的検討ー” 電子情報通信学会、信学技報. SP, 音声 110(220), 41-46, 2010-10-01

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 ー鼻子音出力の操作方法と音声生成方法の検討ー” 電子情報通信学会、信学技報, SP, 110(452), 67-72, 2011-3

## 3. 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 ー障害音と鼻音付加の操作方法の検討ー” 日本音響学会・3-4-14・pp. 1623-1626・2011-3)

## H. 知的財産権の出願・登録状況 （予定を含む。）

### 1. 特許取得

無し

### 2. 実用新案登録

無し

### 3. その他

#### <雑誌>

1)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, pp. 36-47（2010.9）

2)伊福部達：“指で見る「夢」”，ふぉーかす，映像メディア情報学会誌, Vol. 64, No. 11（2010. 11）

3)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”，ライフサポート学会誌巻頭言, Vol. 22, No. 4,（2010）

#### <特別・招待講演>

1)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放送技術研究所, 東京（2010. 5）特別講演

2)伊福部達「高齢社会を豊かにする産業の創出ー感覚・コミュニケーション・運動の支援技術を例にとってー」特殊衣料株式会社, 札幌（2010. 6）

厚生労働科学研究費補助金（障害者対策総合研究事業）  
（総括）研究報告書

- 3)伊福部達「『見る』『聞く』『動く』を助ける医療・福祉技術 - 福祉工学の挑戦 -」ライフ・プランニング・センター, 新老人の会, 東京 (2010. 4)
- 4)伊福部達「福祉工学の40年 - 当事者研究から学んだこと -」, 科学技術社会論学会2010年年次大会シンポジウム「支援技術開発と当事者研究の出会い」, 東京 (2010. 8)
- 5)伊福部達「音の福祉工学 - 声の障害を支援する最新技術 -」第43回日本甲状腺外科学会学術集会, 倉敷(2010. 10) 特別講演
- 6)伊福部達「福祉工学から健康工学へ - 高齢社会を乗り切る戦略 -」神戸大学 健康工学研究会講演会, 神戸 (2010. 11)
- 7)伊福部達「福祉工学から健康工学へ - 知的な高齢社会を築くために -」NPOウェアラブルコンピュータ研究開発機構東京定例会, NPOウェアラブルコンピュータ研究開発機構, 東京 (2010. 12)
- 8)伊福部達「『見る』『聞く』『話す』を助ける福祉工学」IT機器セミナー&講演会, 日本IBM主催, 東京 (2010. 12)
- 9)伊福部達「(仮) 高齢社会を豊かにするイノベーション - 新しい福祉と産業 -」NPO北海道ジェロントロジー推進協会設立記念講演会, 「生きがい創造する高齢社会に向けて - ジェロントロジー (老年学) の役割 -」, 札幌 (2011. 2)
- 10) Ken' ichiro Yabu and Tohru IFUKUBE, "A Real-time speech generator manipulated by tracing a touch-pad using a finger or a pen.", International Workshop on Performative Speech and Singing Synthesis", (Video Presentation), March 14-15, Vancouver, BC (2011)
- 11) 藪 謙一郎, 「発話障害者のための非言語情報を含む音声生成器」, 東京大学先端科学技術研究センターキャンパス公開ワークショップ(生活支援工学会10周年記念ワークショップ), 「当事者の視点によるバリアフリー研究 - 「見る」「聞く」「話す」を支援する技術 -」, 東京, (2010. 6)

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, 2.1「見る」機能を支援するIT, 2.2「聞く」機能を支援するIT, 2.3「話す」機能を支援するIT,2.4 IT応用の将来	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

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
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－電子音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
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

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 Ifukube<sup>1</sup>

<sup>1</sup> Research 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.

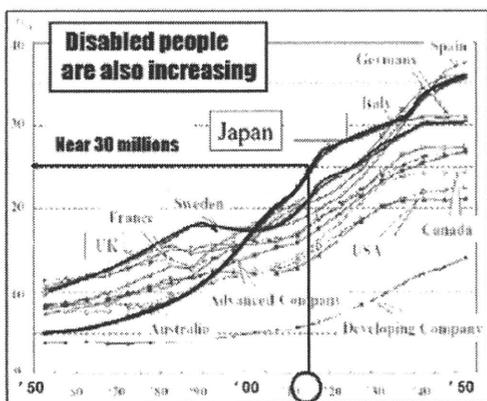


Figure 1: Population rate of the elderly from 1950 to 2050.

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 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].

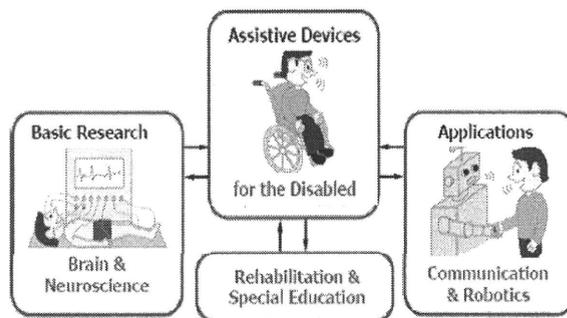


Figure 2: Three steps of our assistive technology research.

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.

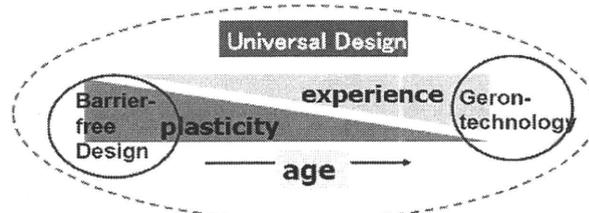


Figure 3: Barrier-free design and Geron-technology.

## 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.

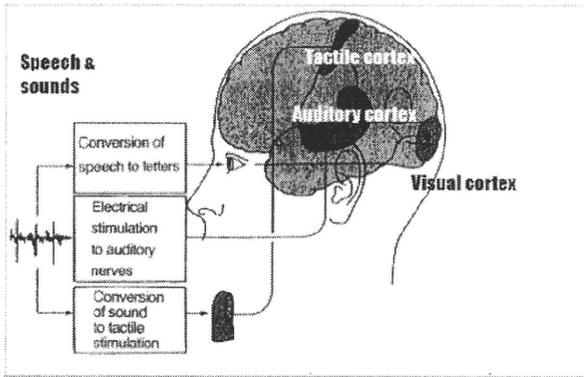


Figure 4: Three research approaches of assistive tool design for the hearing impaired.

### 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 combinate for supporting 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.

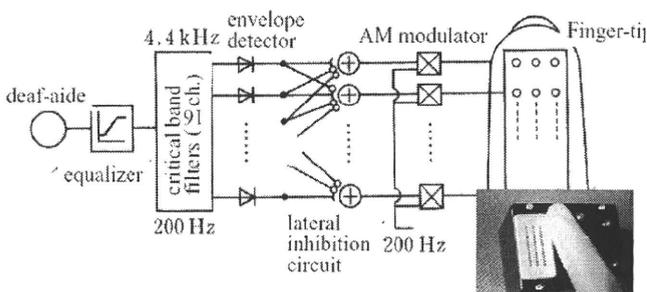


Figure 5: Block-diagram of tactile vocoder and a vibrator array (3x16).

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.

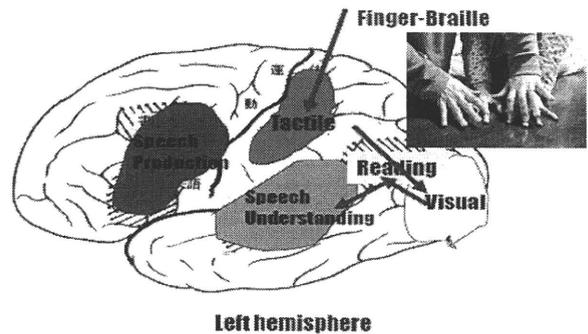


Figure 6: Finger-Braille information routes acquired by “plasticity” of the human brain.

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.

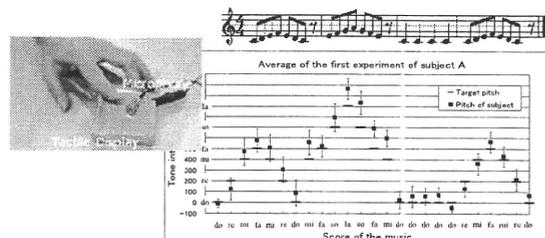


Figure 7: Left: New tactile vocoder made of a microprocessor and a vibrator array. Right: Pitch pattern of a deaf-blind produced by a “Frog’s song”

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 in 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.

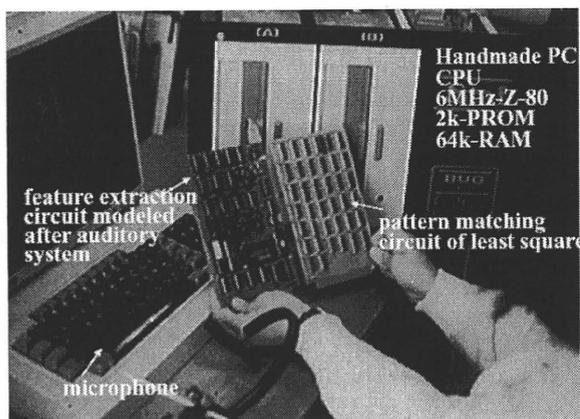


Figure 8: Monosyllabic voice typewriter using a Z80 microprocessor and 32kbyte memories.

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.

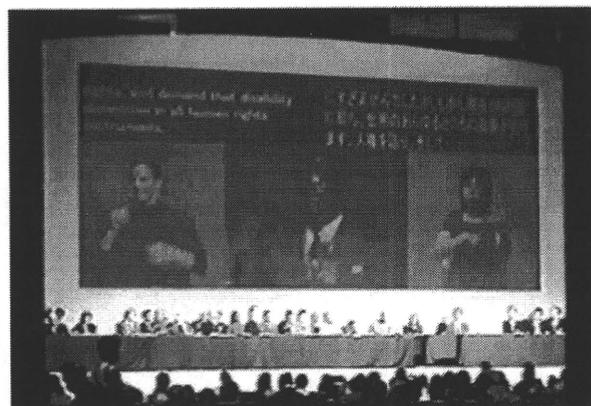


Figure 9: Captions (English and Japanese) presented on a screen together with sign languages and a speaker's face at DPI conference in 2002.

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

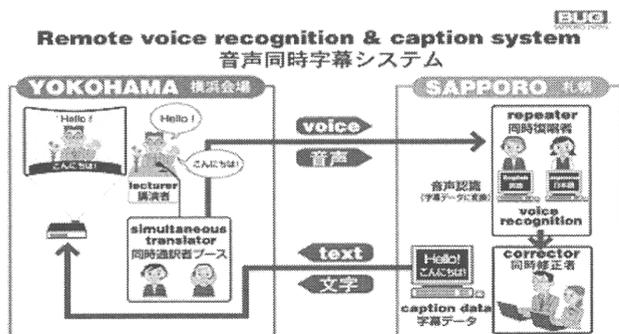
proven by an experimental result that eye balls of the deaf subject mainly stayed on the captions, not on the speaker's face [8].



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 B.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（パシフィコ横浜）

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.

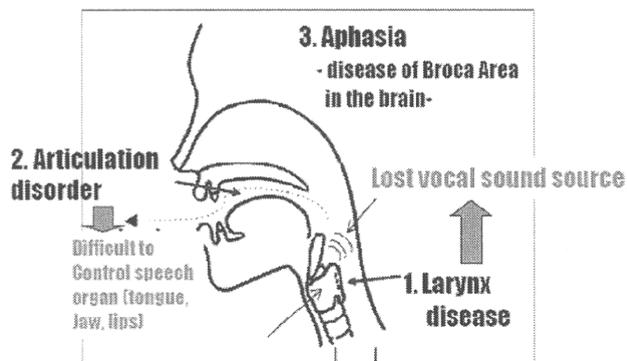


Figure 12: Three causes of speech disorders

#### 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.

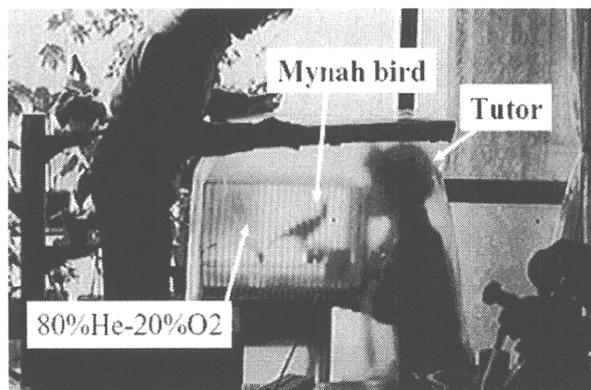


Figure 13: Mynah bird and her tutor in He-O2 gas

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].

After completing this basic research regarding similarity between the human voice and the mynah imitation, we designed a new electro-larynx composed of an air flow sensor, a small computer and a vibrator. An important role in the creation of voice sounds is played by the flow of air from the lungs to the larynx, determining a metrics such as a sound level, an accent and an intonation. In our electro-larynx, the intonation can be controlled by the exhalation flow detected by the air flow sensor as shown in figure 14 [10].

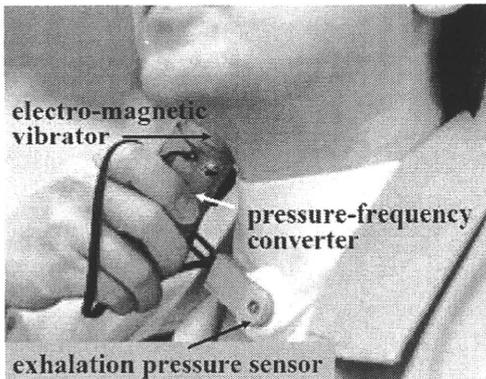


Figure 14: *Pitch controlled electro-larynx*

The electro-larynx voices can be perceived like the human voices with the intonation although the formant structure obtained by the voices is very different from the human's voices. The electro-larynx with pitch-controlled function has been manufactured and more than 4,000 patients have been using the device since 1998 in Japan.

However, since this method requires the use of a hand, it is still somewhat restricting. Many patients have strongly requested us to design a hands-free electro-larynx so that they can use it in daily life and in their office without their hands bound by the control of the electro-larynx while using it.

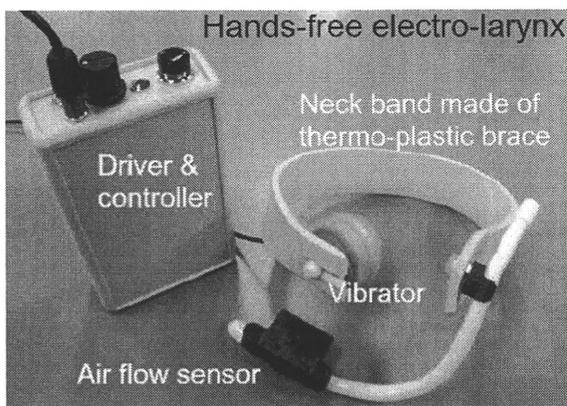


Figure 15: *Hands-free pitch controlled electro-larynx*

Furthermore, the electro-larynx sound is still not human-sounding voice. It was supposed that fluctuations of speech waves are needed to add to the electro-larynx voice.

In order to answer their request, we have designed a hands-free electro-larynx that can be attached to patient's neck. By comparing the various materials for the neck band, a thermo-plastic brace was selected. It becomes soft about at a temperature of 70 degrees centigrade while being hard at a

body temperature so that it is suitable for adjusting individual differences of a shape and hardness of the neck tissue.

We made a smaller vibrator (32mm diameter × 19mm thick and 32g weight) and a much more sensible flow sensor than the conventional one. The vibrator and the sensor can be mounted to the thermo-plastic brace. Figure 15 shows the prototype of the hands-free electro-larynx worn by a patient.

From usability tests, it was confirmed that the electro-larynx voices are heard more naturally than the conventional one and that the activity of the patients significantly improves in daily life. Furthermore, by adding a 1/f fluctuation of the pitch frequency [11] to the electro-larynx, the voice quality apparently exceeded the mynah bird's imitation [12].

### 3.2. Voice synthesizer for the aid of articulation disorders

In order to assist articulation disorders, we designed a speech synthesizer that can be handled by a finger as if a user is playing a musical instrument. One of students, who is difficult to control his speech organ because he has been suffering from muscle dystrophy, designed the synthesizer for himself. He has been disappointed in ordinary speech aids that do not work in real time or can not produce non-verbal information such as intonation and emotional expressions.

Our method was modeled after a speech production mechanism of a parakeet and a ventriloquist that can produce normal speech without moving their mouth. By analyzing their speech production, it was found that they produce speech sounds only by rapidly moving tongue inside the mouth.

This fact means that they can produce bilabial consonants such as /pa/, /ba/, and /ma/ by rapidly moving the tongue without closing their lips. As an analytical result of the ventriloquism vocalization mechanism, it was hypothesized that the ventriloquist quickly moves his/her tongue so that /a/ of /pa/ can be produced after putting the tongue at a front tooth; an articulation point of a consonant /t/, in the case of ventriloquist's /pa/ as shown in figure 16. Human can hear only /pa/ by this tongue movement even if the sound of the consonant /t/ exists. Almost the same mechanism as /pa/ was ascertained in the case of /ba/ and /ma/. This result means that /pa/, /ba/ and /ma/ can be heard only by the tongue movement inside a mouth.

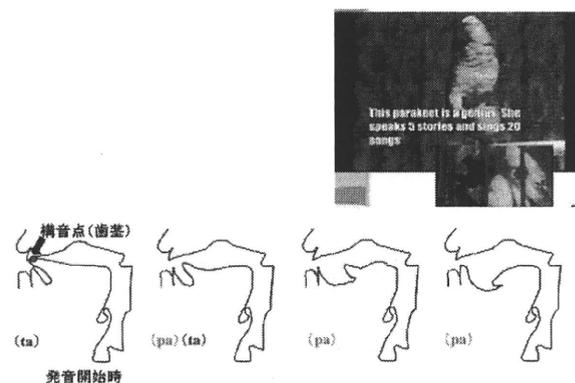


Figure 16: *Parakeet has a flexible and thick tongue (upper), so that she can produce various vowel formants by changing resonant frequencies. Ventriloquist quickly moves a tongue so that /a/ of /pa/ can be produced after putting the tongue at an articulation point of /t/ (lower).*

An idea of our speech synthesizer for the disorders is mainly based on a fact that human recognizes the bilabial consonants produced by the ventriloquist. Parameters of first formant frequency (F1) and the second formant frequency (F2) of a formant synthesis software were controlled by a position and a motion of a user's index finger without any key input.

A touch pad was adopted for the input device that detects the fingertip's position and movement. The F1 and F2 that correspond to the position of the tongue were two dimensionally assigned to the plane of the touch pad as shown in figure 17.

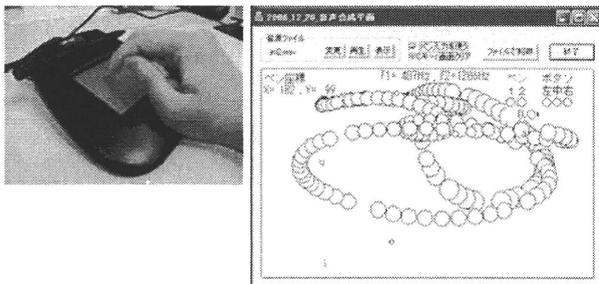


Figure: 17 A touch pad (left) where F1 and F2 frequencies were two dimensionally assigned to the plane (right).

A prototype model was designed on a personal computer using formant synthesis software to get the optimal speech synthesis method for the articulation disorders [13]. In our method, in order for the users to easily find the formant transitions to produce the desired consonants only by changing the formant transitions, "Guiding Lines" were superimposed on the pen-tablet display. Direction and length of each guiding line was assigned at a starting point and a locus of the formant transition on the F1-F2 plane. The starting point and the formant transition locus were decided according to an "Expanded Locus Theory" that we modified based on "Locus Theory" [14].

In the evaluation of sentence recognition, continuous speech sounds were synthesized by tracing the touch pad using the index finger as shown in figures 17 (left). After a few hours training to produce some Japanese sentences such as /ohayo-gozaimasu/ (good morning), /kon'nichiwa/ (good afternoon), /kombanwa/ (good evening) and /arigato-gozaimasu/ (thank you), the consonants /ha/, /go/, /za/, /su/, /ko/, /ni/, /chi/, /wa/, and etc. were apparently heard in the continuous sentences although the synthesized voices had no consonant sounds.

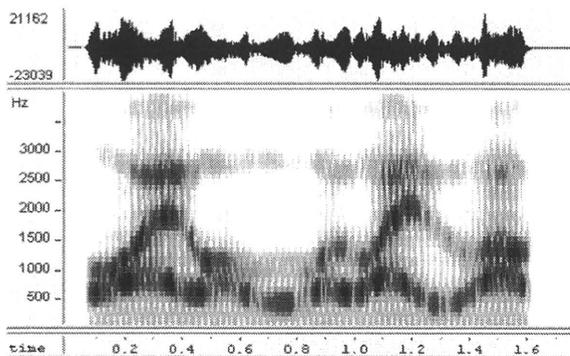


Figure 18: A sound spectrogram of /ohayo-gozaimasu/ obtained by our synthesizer for the speech disorders.

Figure 18 shows a sound spectrogram of /ohayo-gozaimasu/. From the evaluation results, it was ascertained that the users may more easily produce the arbitrary sentences by tracing the guide lines than conventional one [15].

Next we carried out evaluation test of randomly synthesized 100 words composed of 4 monosyllables. Although the identification of vowels and voiced consonants such as semi-vowels, nasals, showed high score around 70 %, the average identification rate showed only 33 %. It was still difficult to perceive some consonants that need random noise component such as /s/, /h/ and /z/. We have been investigating how the noise component should be added and controlled by user's finger.

By using a touch pad with a pressure sensor and by assigning the detected touching pressure to the pitch frequency, the subject could produce some emotional voices such as "laughing", "surprising" and "disappointment" by controlling the pressure of the touch and the speed of finger movement. Furthermore, by connecting this tool with a musical key-board, the subject could sing some Japanese songs after short time training.

We are planning the speech production method to apply a tool for people with aphasia and also to one of new musical instruments for general users.

#### 4. Assistive tools supporting "Seeing"

There are three approaches for visual substitutes as shown in figure 19. One is to convert visual information into tactile stimulations, the second is to electrically stimulate the surviving visual nerves, and the third is to convert text to speech. In this session, two examples to assist the visually disabled are mentioned.

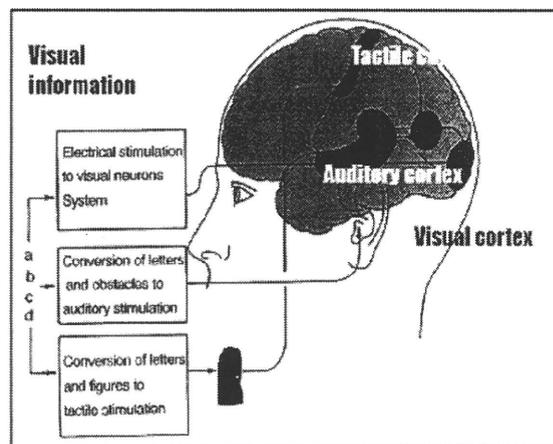


Figure 19: Three approaches supporting "seeing".

##### 4.1. Screen reader for the blind

In 1996, we developed a screen reader, called '95-reader', which can convert text information into synthesized voices for Japanese visually impaired people as shown in figure 20 [16]. However, operating systems based on graphical user interfaces (GUIs) have spread rapidly in personal computer environments. Consequently, it is inherently difficult for visually impaired people to use the screen reader, since they are often unable to detect the necessary visual information.

With this problem in mind, we have proposed a tactile jog-dial (TAJODA) which can control the speech rates of the screen reader while displaying rich text features, such as various