

広島市立大学審査博士学位論文

発声機能障害者支援システムのための音質
明瞭化手法の検討

2016 年 3 月

広島市立大学大学院
情報科学研究科 博士後期課程
情報科学専攻

福井 和敏

目次

第 1 章	緒言	1
1.1	研究背景	1
1.2	発声機能障害の現状	3
1.3	代用発声法の現状と問題点	5
1.4	コミュニケーションエイドの現状と問題点	6
1.5	先行研究の現状と問題点	6
1.6	研究の目的	8
1.7	本論文の構成	11
第 2 章	発声機能障害者支援システム	12
2.1	システムの目的と特徴	12
2.2	システムの課題	12
2.3	提案システムの概要	13
2.4	音源フィルタ理論 ⁽²⁴⁾	14
2.5	体内伝導音	15
2.6	音節区間推定手法 ⁽²³⁾⁽²⁶⁾⁽²⁷⁾	17
第 3 章	システム評価用音声データベースの構築	27
3.1	収録音源と環境	27
3.2	発声障害 ⁽³³⁾	30
3.3	代用発声法	35
3.4	構音障害	36
第 4 章	発声機能障害者支援システムの評価	39
4.1	音声認識評価	39
4.2	音節区間の推定	43
第 5 章	構音障害者への適用とその評価	45

5.1	実験の目的	45
5.2	調音素性	45
5.3	聴取実験による合成音声の評価	51
5.4	SINFA による母音の調音素性伝達率解析	52
5.5	子音の調音素性伝達率解析	53
第 6 章	結言	58
	謝辞	61
	参考文献	62

第 1 章

緒言

1.1 研究背景

音声は人間同士，さらには人間と動物など生物の意思疎通手段として最も重要なものの一つである．病気や事故により発声機能にかかわる障害を罹患すると意思疎通が困難になり日常生活に大きな影響を与えてしまう⁽¹⁾．日本にはこのような発声機能（音声・言語・そしゃく機能）障害者がおよそ 34,000 人存在する⁽²⁾．

我が国における音声・言語・そしゃく機能における障害者認定は 3 級または 4 級であり，それぞれ「音声機能，言語機能又はそしゃく機能の喪失」，「音声機能，言語機能又はそしゃく機能の著しい障害」と定義されている．3 級の音声機能喪失とは，無喉頭，喉頭部外傷，発声筋麻痺による音声機能喪失などをいう．また 4 級の音声機能の著しい障害とは音声，言語のみを用いて意志を疎通することが困難な状況，例えば喉頭の障害または形状異常や構音器官の障害または形態異常をいう．つまり等級に関係なく，音声・言語・そしゃく機能障害と認定されている方は発声によるコミュニケーションが困難な状況であり，何かしらの発声機能障害支援が必要になる．

現在は音声・言語・そしゃく機能障害者のほとんどが食道発声法や電気式人工喉頭などの代用発声法を使ってコミュニケーションをとっている．しかしながら国の調査では聴覚・言語障害者が外出する際，28.5%の方が「人と話をすることが困難」と感じており⁽³⁾，現状の支援で満足しているわけではないことがわかる．

これら発声機能障害の大きな要因である口腔・咽頭・喉頭癌に罹患される方が日本では年々増加傾向にあり，2011 年には 2 万人を超えている（表 1.1）．特に高齢者の割合が多く，全体の 63.1%が 65 歳以上の高齢者となっている．

2012 年における全世界の口腔・咽頭・喉頭癌罹患患者数は 456,000 人にもものぼり全癌患者の 3%を占めている．アメリカでは年間 16,980 人の方が罹患され，そのうち 58.9%が 65 歳以上となっている⁽⁵⁾．イギリスでは新規罹患患者 8,755 人のうち 65 歳以上の割合

は71%にも上る⁽⁶⁾。全罹患者のうち75%はアジア地域の患者であり、特に人口10万人あたりの罹患率（表1.2）では先進7か国の中で日本が最も高い値となっている⁽⁴⁾。

前述のように、発声機能障害になった場合は「食道発声法」や「電気式人工喉頭」、「笛式人工喉頭」などの代用発声法を習得することが多い。代用発声法は、習得が困難、声が小さい、器具が必要（片手がふさがる）、機械的な音になるなど多くの課題がある。また欧米では「T-E シャント法」という代用発声法が多く選ばれるが、この方法も特殊な手術が必要であり、器具の定期交換にコストがかかり、メンテナンスが面倒という課題がある。

他の障害と比較すると、聴力であれば補聴器、視力であれば眼鏡のように道具や機械で機能補完するものがあるが、発声に関しては代用発声法のような擬似的なものしかない。発声障害において自分の意志を伝えるためには手話、筆談といったより確実な手段もあるが、従来のADL（Activities of Daily Living）よりも障害後のQOL（Quality Of Life）向上が重要視されるようになった昨今、やはり自分の声で意志を伝えたいという要望が非常に強い。したがって他の障害に対する支援ツールのように、発声機能障害についても、携帯可能で気軽に自分の声を再現できるような支援ツールを開発することが障害者のQOL向上という大きな社会的意義につながる。

支援ツールの一つとして健常時の肉声を再現する先行研究⁽¹⁵⁾や製品⁽¹²⁾⁽¹³⁾がある。

調査年		1975	1980	1985	1990	1995	2000	2005	2011
患者数	総数（人）	3913	6400	8187	9744	11944	12934	15032	20182
	65歳以上（人）	1777	3053	3873	4738	6370	7561	8983	12744
	65歳以上（%）	45.4	47.7	47.3	48.6	53.3	58.5	59.8	63.1

表 1.1: 日本における口腔・咽頭・喉頭癌罹患患者数の推移

対象国	ASR（人）
アメリカ	5.5
日本	11.1
イギリス	10.0
イタリア	2.1
フランス	6.1
カナダ	4.6
ドイツ	6.9

表 1.2: 先進7か国：10万人当たりの口腔・咽頭・喉頭癌罹患患者数（ASR）

これらは事前に健常時の音声や肉親などの似た音声を元に大量に学習しモデル化しなければならないという課題がある。特に市販製品は非常に高額で音声収録にかなりの時間が必要となるため、導入に金銭的にも体力的にも負担が大きくなってしまう。また、これらは発話する文章をユーザーがパソコン上で入力しなければならないため、会話の応答に時間がかかり、自然な会話の実現には適さない。さらに文章を元に発声するため、その時の体調や感情を反映した音声・口調にすることが難しいという大きな課題がある。

つまり、発声機能障害についてはいまだ有効性の高い支援方法が確立されておらず、この支援方法の確立が学術的にも大きな意味を持つ。

このような背景から、我々は発声機能障害者向けの支援システムを開発中である⁽²²⁾。このシステムは障害者発声時の体内伝導音を抽出し、本人の肉声を再現したうえで聴取者に聞き取りやすい明瞭な音声に変換してリアルタイムで発声するものである。

本システムの大きな特徴として体内伝導音の利用がある。すでに代用発声法使用者の音声を増幅して聞き取りやすくするような支援ツールは市販されているが、これらは実使用環境の騒音が音声より大きな環境下では使えない。そのため騒音の影響を受けにくい骨導をはじめとする体内伝導音を用いている。しかしながら体内伝導音は高周波域が減衰するという特徴があり、そのままでは明瞭な音声としては使えないことがわかっている⁽¹⁷⁾。したがって、本システムの実用化には体内伝導音のリアルタイムな明瞭化手法の開発が必須である。

1.2 発声機能障害の現状

発声機能障害は「発声障害」と「構音障害」の2つに分類される⁽⁷⁾。発声とは肺から吐き出される呼気によって喉頭にある声帯を振動させ音声を作り出す現象であるが、声帯等の音源を生成する器官に関する障害を「発声障害」という。発声障害の症状の多くは声がかかる「嗄声」であり罹患要因としては声帯の器質的病変や運動麻痺が主体である。喉頭で作られた音声を元に、それより上にある声道や鼻腔・口腔などの言語情報を付加（調音）するための構音器官（図 1.1）に関する障害を「構音障害」という。構音障害の要因は形態的障害、運動障害、神経系障害があげられるが、いずれも正確な構音ができなくなる。

このようにそれぞれの障害においてその要因・症状は異なるが、共通して発声した音声が不明瞭となる。

1.2.1 発声障害の現状

発声障害罹患の大きな要因として口腔・咽頭および喉頭癌が挙げられる。図 1.2 はは 1975 年から 2011 年における口腔・咽頭癌および喉頭癌患者数と 65 歳以上の割合を示し

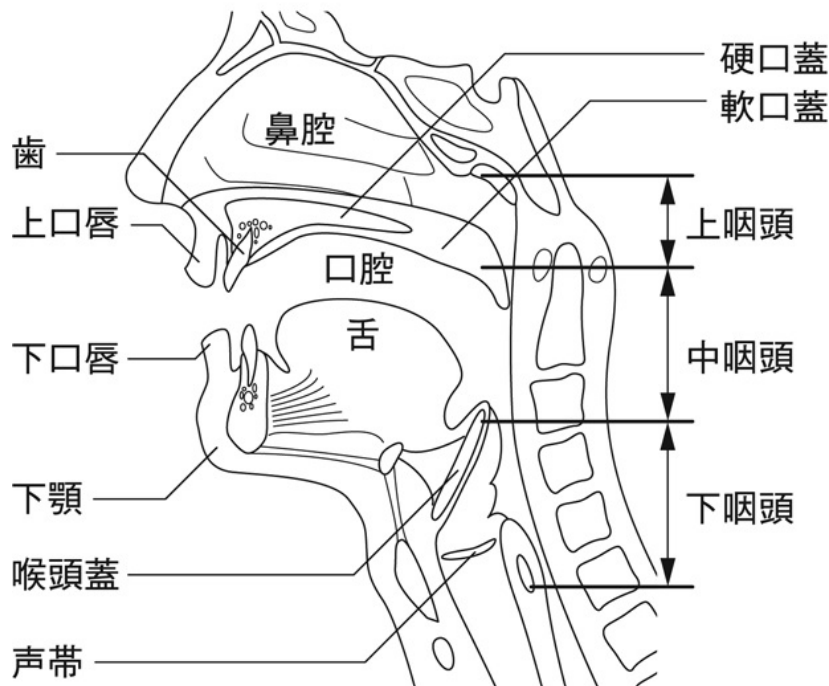


図 1.1: 構音器官 (7)

ている⁽⁸⁾。患者数は2000年の13,026人から2011年では20,182人と10年間でおよそ1.5倍と増加の一途をたどっている。さらに65歳以上の高齢者が全体の63.1%を占め、患者の高齢化が進んでいることがわかる。このような口腔・咽頭および喉頭癌の場合、手術により声帯を摘出する場合がある。下咽頭、喉頭癌を罹患すると声帯摘出を伴う場合が多く、特に下咽頭癌は自覚症状が出にくいため、初診時に60%以上がすでに喉頭に浸潤したり、頸部リンパ節転移を伴うなど進行状態である⁽⁹⁾。その際には声帯を含め喉頭まで切除せざるを得ない場合が多い。

このような発声障害を罹患した場合、声帯機能の低下や肺活量の低下等の要因により音源を生成することが困難になる。

1.2.2 構音障害の現状

構音障害罹患の要因は、調音器官の形態的な要因によるものや発話言語に関わる神経や筋肉の病変によるものなど様々である。形態的要因によるものの例として口蓋裂がある。これは、口蓋の一部が裂け、口腔と鼻腔が直接交通するという先天性の障害である。新生児では哺乳に障害が出たり、成長とともに咀嚼・嚥下・構音に関わる障害が発生する。また、空気が漏れてしまうため発声した音声が開放性鼻声となる。構音障害を引き起こす脳神経変性疾患として筋萎縮性側索硬化症 (Amyotrophic Lateral Sclerosis, ALS) に代表

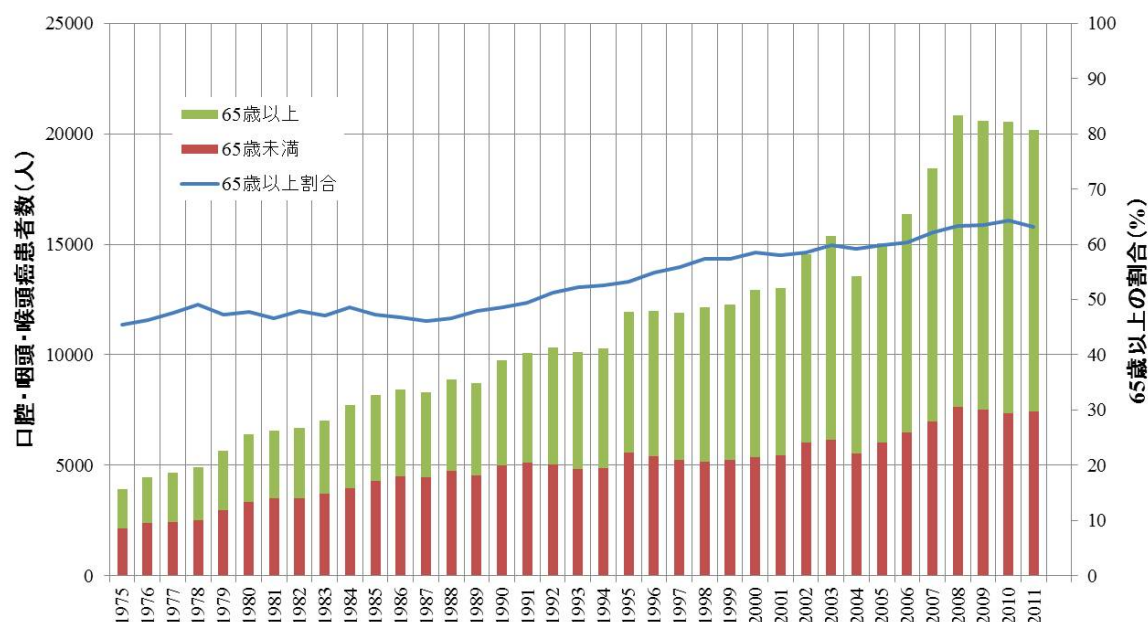


図 1.2: 口腔・咽頭・喉頭癌の患者数と 65 歳以上の割合

される運動ニューロン病 (Motor Neurone Diseases, MND) やパーキンソン病, 多発性硬化症 (Multiple Sclerosis, MS) などがある. これらによって筋肉が痩せたり, 正しく動かすことができないため音声に障害が生じる. そのほか, 事故などによる調音器官の障害・欠損がある. 調音するための器官に損傷を受けたり, 器官そのものが欠損してしまうことで調音機能を失ってしまう. 特に, 本研究の実験にご協力いただいた構音障害者は事故により顎と舌の一部を欠損している.

このように, 構音障害は様々な要因により調音器官に異常が生じ, 正しく言語情報を付加することが困難になる.

1.3 代用発声法の現状と問題点

発声機能障害に罹患し発声が困難になった場合, その代用発声法として食道発声法, 笛式人工喉頭や電気式人工喉頭を用いた代用発声, T-E シャント法などがある. 日本における最も一般的な代用発声法が食道発声法である.

食道発声法は呼吸器官の代わりに食道に空気を吸い込み, 吐き出す時に声門付近の粘膜を振動させ声を出し, 構音器官によって調音を行うことで発声する. 食道発声法は, 声帯の代わりに食道を使って発声するため肉声であり, 器具が不要なのでいつでも発声が可能といった利点がある. しかし音量不足, 雑音混入, 長時間の発声が困難, 苦しそうな発声に見える, 習得が難しいなどの多くの課題がある. また後述のコミュニケーションエイド

を用いた代用発声法として笛式人工喉頭や電気式人工喉頭がある。笛式人工喉頭は声帯摘出時にあけた「永久気管孔」から出た空気で「タピアの笛」と呼ばれる笛を鳴らし、その音をチューブで口腔に導いて発声する。比較的習得が容易で声に抑揚を付けることもできるが、笛の操作が必要で片手がふさがる上にチューブが目立ってしまうという課題がある。電気式人工喉頭は電氣的に振動するバイブレーターを喉に押し当て、その振動音を利用して発声する。習得が簡単ですぐに会話ができるようになるが、機械的な音声になり、機械の保持のため片手が完全にふさがってしまうという課題がある。

その他、気管と食道をチューブでつなぎ、声帯摘出時にあけた「永久気管孔」を指でふさぐことで肺からの空気が発声する T-E シェント法がある。主に欧米で普及している代用発声法であるが、気管と食道をつなぐ手術が必要となり、メンテナンスも面倒なため、日本における普及率は高くない。

1.4 コミュニケーションエイドの現状と問題点

障害者のコミュニケーション支援のための製品をコミュニケーションエイドという。

声帯機能を失った発声障害の方の代用発声に用いる図 1.3 に示すようなコミュニケーションエイドが電気式人工喉頭として商品化されている。この装置は発声する際、声帯に装置を押し当てて口を言葉に合わせて動かすことで発声が可能となるため特別な訓練を必要としない。現在このような装置は出力音声の自然性を高めるために「ゆらぎ」を付加したり声の高さを調節できる機能を追加している⁽¹⁰⁾。したがって、言語として自然性の高い音声を実現できるが、話者自身が発声していた音声との声質の違いにより精神的な苦痛を与える等の問題点がある。

また、言語情報を伝達するために使用する音声出力型コミュニケーションエイド (Voice Output Communication Aid, VOCA) がある。図 1.4 に示すように、現在はモバイル端末に対応した音声合成ソフトウェアも提供されている⁽¹¹⁾。

しかし、電気式人工喉頭では完全に機械的な音声になってしまう等、VOCA には出力される音声が発声者自身の音声と異なってしまうという課題がある。さらに、利用者の年齢や方言、発話の癖等の非言語情報も表現できない。

このように現在のコミュニケーションエイドでは音質や非言語情報といった点で課題があり、障害者個人の自然な音声を再現できる音声合成器の需要が高まっていると考えられる。

1.5 先行研究の現状と問題点

障害者の音声の音質改善を目的とした研究は他研究機関でも行われている。



図 1.3: 電気式人工咽頭「ユアトーン II・UB」(株式会社 電制)



図 1.4: モバイル端末に対応した会話補助ソフトウェア「おしゃべりサポーター」

声質変換手法の一つとして統計的声質変換が知られている．現在でも統計的性質変換について盛んに研究がおこなわれている．

米国オレゴン大学の Kain らは Klatt のホルメント合成器と性質変換技術を組み合わせ，発声機能障害者の音声修復を試みている⁽¹⁴⁾．入力音声に対して有声／無声を判定し，有声部のみ混合正規分布モデル (Gaussian Mixture Model, GMM) を用いて声質変換を行う．また，奈良先端科学技術大学院大学の森口らは GMM に基づく変換により，

発声障害者音声で電気式人工喉頭を用いて発声した電気音声を通常音声に変換するシステムを検討している⁽¹⁵⁾。この研究では電気音声を対象としており我々が対象とする音源と異なるが、入力された電気音声から音声のパラメータ (F0/言語情報/音声の非周期性) を抽出し、それぞれを GMM を用いてパターンマッチングし、対応付けられた通常音声のパラメータを用いて自然な音声を生成する。

統計的声質変換によるシステムでは前述の代用発声法と比較し、より明瞭で自然な音声を再現できるが、障害者の健常状態での音声及び障害者音声を事前に大量学習する必要があるという課題がある。例えば GMM を用いる方法で音声のパターンマッチングを言語情報を用いずに比較的高い精度で変換できるが、自然な音声を作成するには日常会話での音声の非常に多いパターンを網羅し学習させなければならない。したがってシステムを製品化する際、利用者ごとに学習する際の負担が大きい。また、システムの処理も音声を精密に作成するボコーダとなっているため、システム全体としての応答時間に課題があると考えられる。

以上より、体内伝導音明瞭化のための声質変換技術として、事前学習が不要で障害者の自然な肉声をリアルタイムで再現できるものが求められている。

1.6 研究の目的

これまで述べたように、世界的に増え続ける発声機能障害者のコミュニケーションを支援し、QOL の向上を実現するために新たな支援ツールの開発が必須である。我々が考える支援ツールに必要な機能として以下に挙げる項目を満たすことが必要だと考える。

- 事前の大量学習等準備が不要
- 違和感のない自然な音声を実現できる
- 話者の肉声に近い発声で違和感を少なくする
- 話者音声を効率的に識別し高速なシステム応答性を確保する

我々は新たな支援ツールとして騒音に頑健な体内伝導音 (Body Conducted Sound, BCS) を用いた発声機能障害者支援システムの開発を行っている⁽²²⁾。既存の音声や文字列を元に発声するシステムではなく、より実使用環境を考慮し騒音に頑強なシステムとするため、提案システムは、障害者発声時体内伝導音を用いて話者の声質を活かした自然かつ明瞭な音声に変換する機能を持つ。本システムで使用する体内伝導音は外乱の影響を受けづらい反面、高周波数域が減衰してしまいそのままでは音声として使えない。また、障害者の QOL を向上させるため、障害以前の自然な肉声を再現することも求められている。

本研究では提案システムにおける変換音声の音質改善を目的とする。

先行研究⁽¹⁴⁾⁽¹⁵⁾では統計的声質変換で（電気）音声の声質変換による肉声再現をめざしたが、障害以前の肉声による事前の大量学習が必要であり、我々が提唱する新たな支援ツールには適用しづらい。また、先行研究⁽¹⁶⁾⁽¹⁷⁾⁽¹⁸⁾では、代用発声時の気道音と体内伝導音からクロススペクトル法を用いて声道の伝達関数を推定していたが、クロススペクトル法を用いることで単語全体に平均スペクトルが強調された信号となり、その有効性を確認することができなかった。これは音声に複数のサブワードが含まれているため単一のフィルタ特性で表現できず、明瞭度の改善に至らなかったと考えられる。そこで本研究では新たに音声生成の仕組みに着目し、線形予測係数 (Linear Predictive Coefficient, LPC) を用いて推定した伝達関数による音声合成手法⁽¹⁹⁾⁽²⁰⁾を提案し、体内伝導音明瞭化を実現した。

さらに、これまで LPC 分析の単位である音節区間推定は、人間が波形を見ながら音声聴取し決めるヒューリスティックな方法で行っており、システムの実用化に不可欠な音節区間の自動推定が困難であった。本研究ではリアルタイム変換を実現するため、LPC 分析の単位である音節区間について連続サブワード認識 (Continuous Sub-word Recognition, CSR) を用いた推定自動化を提案し、音節区間推定自動化の可能性を確認した。

本システムでは変換前の言語情報は音声認識で判別するが、音声認識の精度を高めるために標準の音響モデルに障害者から抽出した体内伝導音を学習させることで、モデルを一から作成せず特定話者用モデルを作成し使用することを検討している。実現すれば障害者音源のみの学習でモデルを作成できることになる。また、音質変換部では合成フィルタを係数として保持することで演算量を軽減している。したがって、提案システムは効率的に障害者音源を学習し明瞭な障害者音声を実現できる応答性の高い仕組みを持つ。

さらに、LPC 分析や CSR を用いて作成した合成音声がどの程度言語情報を再現できているかを確認するために聴取実験と調音素性分析を行った。これにより、障害者音声と比較して、本システムが具体的に音声のどの部分を再現できるかを特定することができ、また本システムの課題も明確にすることが期待できる。

以上の提案・評価の結果、LPC 分析を用いて合成した音声は、明瞭度の目安とした音声認識の認識率が大幅に向上した。さらに本来、体内伝導音に含まれていない高周波成分の回復がパワースペクトル上で確認できた。また、聴取実験および調音素性分析の結果から本来存在しない調音素性を完全に再現できていることも確認できた。つまり、口腔や鼻、舌、顎といった調音器官を実際に失っても、本システムを使うことでそれら調音素性を回復でき、明瞭な肉声が再現できるということである。

本研究で得た成果を用いることで、開発中の発声機能障害者支援システムの実用化も可能となり、発声機能障害に対する支援ツールとして技術確立できると考える。さらに障害支援だけでなく、育児・介護といった福祉の分野や、将来的に訪れるであろう NUI

(Natural User Interface) による人間と機械のインタフェースの基礎技術の一つとして応用できるはずである.

1.7 本論文の構成

本論文は以下の構成からなっている。

第1章 緒言

本研究の背景・目的を述べる。

第2章 発声機能障害者支援システム

開発を進めている発声機能障害者支援システムの概要と基礎技術を述べる。

第3章 システム評価用音声データベースの構築

発声機能障害者支援システム構築のための評価用音声データベースの作成を行っており、その詳細と収録音声の種類について述べる。

第4章 発声機能障害者支援システムの評価

障害者の体内伝導音を使用した線形予測分析を用いた合成音声について述べる。また、区間推定を自動化するためのサブワード単位の変換処理について述べる。

第5章 構音障害者への適用とその評価

第4章で述べた合成音声生成手法と各変換処理の効果を音声特徴量及び音声特徴量の観点から評価した内容について述べる。本検討では、音声特徴量の観点から音声認識実験、音声特徴量の観点から聴取実験及び調音素性分析を用いて評価した。

第6章 結言

本研究のまとめについて述べる。

第 2 章

発声機能障害者支援システム

本章では開発中の発声機能障害者支援システムの目的・概要とシステム構築に必要な理論・技術について述べる.

2.1 システムの目的と特徴

本システムは下記のような特徴をもった発声機能障害における支援ツールとして実用化を目指す.

障害者の発声時の体内伝導音を使用

骨導などの体内伝導音を利用することで電車内等の騒音環境下でも使用可能.

線形予測係数をデータベース化

処理のたびに線形予測係数を算出するのではなく, あらかじめ線形予測係数をデータベース化し, 対応するものを呼び出すだけで高速に変換音声を生成.

音節単位での音質変換

音節の組み合わせで任意の発音に対応可能.

障害者音声の大量学習が不要

障害前の音声をシステムに大量学習させる必要がないので, 多くの障害者が簡単に利用できる.

2.2 システムの課題

本システムでは耐騒音性向上のため体内伝導音を利用する. 体内伝導音は外乱の影響を受けにくい反面, 高周波成分が減衰するという特徴がある. また, 利用対象が発声機能障害者であるため, 中には事故等により調音器官を失い, 本来調音に必要な調音素性が無い

方もおられる。したがって、体内伝導音から音声を合成する際、これら高周波成分や調音素性を回復させることがシステムの実用化のために必須となる。

先行研究⁽¹⁶⁾⁽¹⁷⁾⁽¹⁸⁾では、代用発声時の気道音と体内伝導音からクロススペクトル法を用いて声道の伝達関数を推定し、駆動音源に伝達関数をフィルタリングすることによる音質変換では音質の向上が困難であることを確認した。そこで、本研究では新たに線形予測係数 (Linear Predictive Coefficient, LPC) を用いて推定した伝達関数によって駆動音源を抽出し、明瞭化を行う手法⁽²¹⁾を提案する。また、これまで体内伝導音の明瞭化は単語単位で行ってきた。しかし、人間の発話内容は多様で不明瞭な発話も多く、単語単位での音質変換ではこれら多様な単語に対応できない。そのため、本研究では単語よりも短いサブワード単位での音質変換システムも提案する。さらに、本システムで生成した合成音声の明瞭度を評価するために聴取実験及び調音素性分析を行い、音声障害者の原音声と比較してどの程度明瞭化されたか確認した。

2.3 提案システムの概要

発声機能障害者支援システムの構成を図 2.1 に示す。

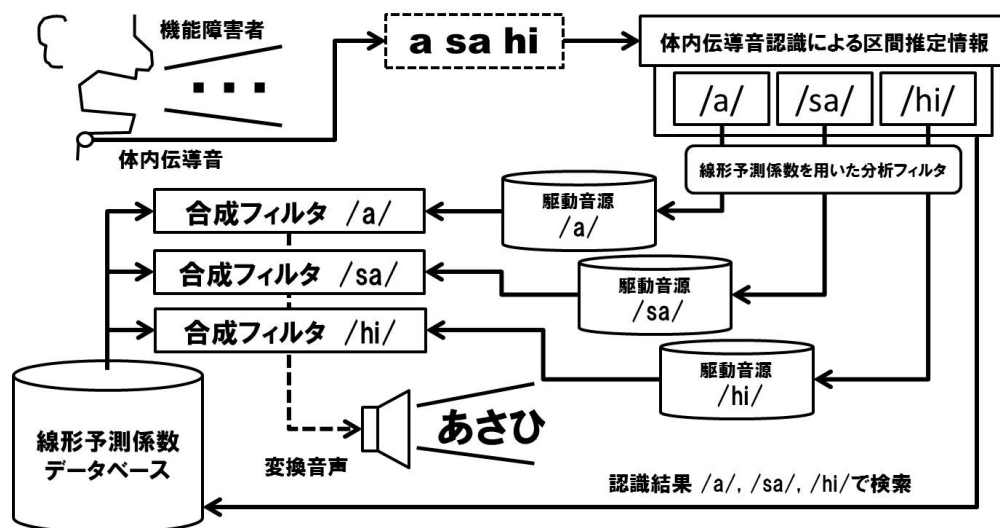


図 2.1: システムの構成

音源フィルタ理論より、音声は声帯の振動などの入力信号を口腔や鼻腔などの声道フィルタに入力した際の出力信号と考えることができる。本システムでは音声が式 2.1 のような全極型の伝達関数を持つ声道フィルタの出力であることを前提として線形予測分析を用いる。

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p}} \quad (2.1)$$

つまり、伝達関数を規定するパラメータである線形予測係数 a_i を決定することで、声道フィルタの共振特性（スペクトル包絡）と音源振動の振幅スペクトル（駆動音源）を推定することができる。このことを利用し、体内伝導音から抽出した駆動音源を、健常な音声から推定した線形予測係数を用いて導いた声道フィルタを通過させることにより音質の明瞭化を検討した。また、線形予測係数は20次とした。

本手法を用いた音声明瞭化の流れを以下に説明する。なお事前準備として、健常者音声の音節単位の線形予測係数を推定し、それらに対応する合成フィルタをあらかじめ作成しデータベース化しておく。

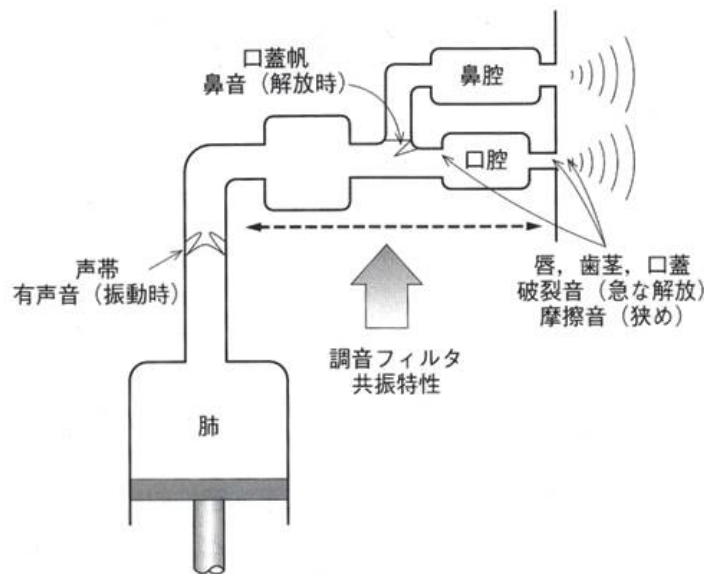
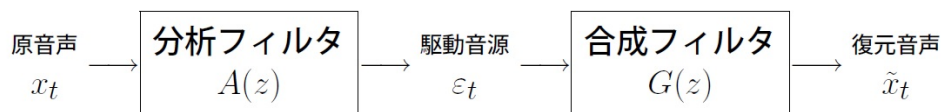
1. 代用音声発声時の体内伝導音を収録し、体内伝導音認識を行い音節の区間情報を推定する
2. 推定した区間情報をもとに収録した体内伝導音から単音節を切り出す
3. 切り出した単音節にLPC分析フィルタを施し、駆動音源を抽出する
4. 体内伝導音認識結果から音節に対応する合成フィルタ（線形予測係数）をデータベースより呼び出し、駆動音源をフィルタリングする
5. 各単音節を結合し、結果を出力する

本システムでは音節の区間推定に音声認識を用いるが、誤認識が発生した場合、誤った合成フィルタを適用してしまう可能性もある。しかし、本システムでは誤ったフィルタを選択したとしても障害者の発声時体内伝導音を元に処理を行うため発声のイントネーションなどから補正されることが期待できる。

2.4 音源フィルタ理論⁽²⁴⁾

音声の生成の仕組みを図2.2に示す。図2.2に示すように人間は声門付近での発声を口腔・鼻腔・口蓋等の調音器官で様々な音色に調音できる。つまり、声帯の振動によって生成された音源が調音器官というフィルタに入力されると共鳴特性によって調音される線形システムと考えることができる。このような線形システムによるモデル化を音源フィルタ理論と呼ぶ。

音源フィルタ理論に基づき音声を音源と調音フィルタの二つに分離したり、逆に二つの情報から音声を合成することを音声分析合成モデルと呼ぶ。図2.3に示した音声分析合成モデルの分析フィルタ $A(z)$ は入力スペクトルに対応する白色化フィルタであり、合成フィルタ $G(z)$ は $A(z)$ の逆フィルタとなるため式2.2のように示すことが出来る。

図 2.2: 音声の生成モデル⁽²³⁾図 2.3: 音声分析合成モデル⁽²⁴⁾

$$G(z) = \frac{1}{A(z)} \quad (2.2)$$

2.5 体内伝導音

発声すると発声者自身は二つの音を聞く．気道音と体内伝導音である．

気道音は空中伝播し聴取者の鼓膜を振動させ聴覚神経に伝わる音である．

体内伝導音は人間の骨などで個体伝播し鼓膜を通さず直接聴覚神経に伝わる音である．気道音よりも周波数特性が低く 2KHz 以上の高周波数域が減衰してしまうため無声音等の弱い音や摩擦音などは伝播しにくいですが，空気中を伝わる外乱の影響を受けにくいという特徴を持つ．図 2.4 に静寂環境下で地名/asahi/を発声した際のスペクトログラムを示す．高周波数域まで伝播している気道音に対して，体内伝導音の低い周波数特性が顕著に表れている．

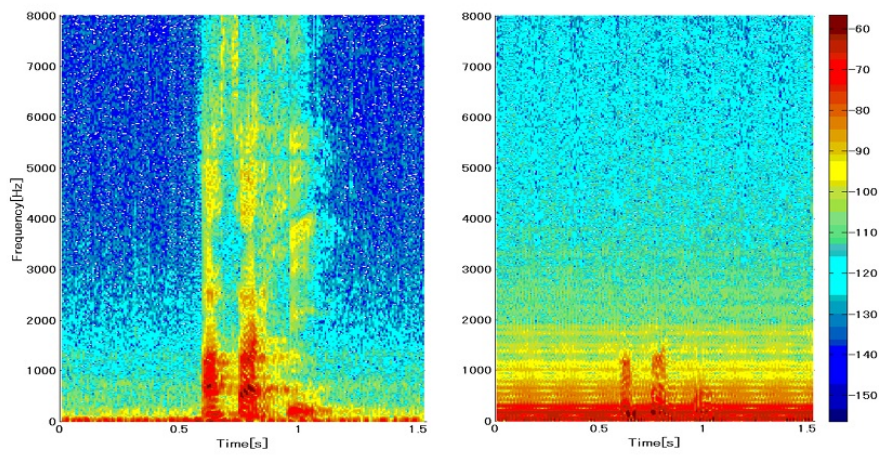


図 2.4: 静寂環境下で発声した/asahi/の気道音声 (左) と体内伝導音 (右)

騒音環境下で地名/asahi/を発声した際のスペクトログラムを図 2.5 に示す．気道音は完全に騒音に埋もれているが体内伝導音は静寂環境下と同等の周波数成分を保持していることが確認できる．同様に先行研究⁽²⁵⁾では体内伝導音を利用した騒音に頑健な音声認

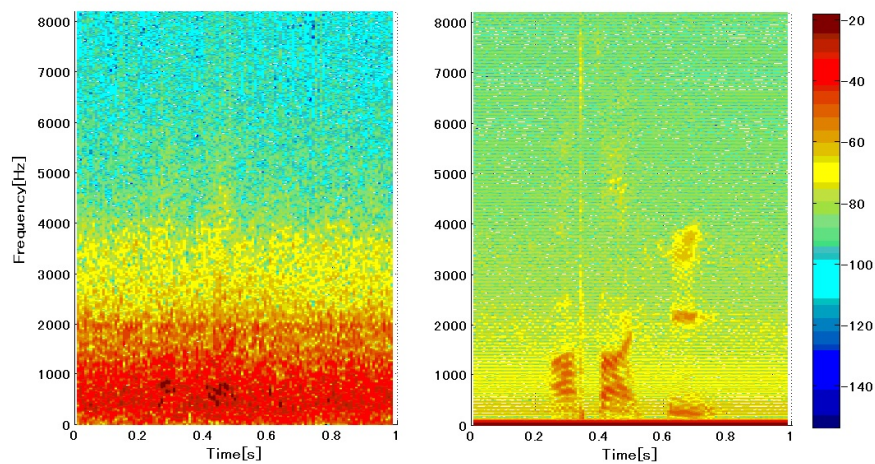


図 2.5: 騒音環境下で発声した/asahi/の気道音声 (左) と体内伝導音 (右)

識システムを構築し，98dB SPL (-20dB SNR) 環境下で 95% の認識率を達成している．

日常生活の様々な環境での使用を前提とした発声機能障害者支援システムを構築するためには体内伝導音の利用が適していると考えられる．

2.6 音節区間推定手法 ⁽²³⁾⁽²⁶⁾⁽²⁷⁾

本システムでは抽出した体内伝導音の音節区間推定に音声認識を用いる．推定した音節区間に基いて分析フィルタおよび合成フィルタを選択するため，例えば誤った音節区間を推定した場合は本来の音声とは若干差異がある合成音声になる可能性がある．つまり本システムでは合成音声の品質に音声認識が与える影響は小さくない．

2.6.1 音声認識の概要

音声認識の基本構成を図 2.6 に示す．

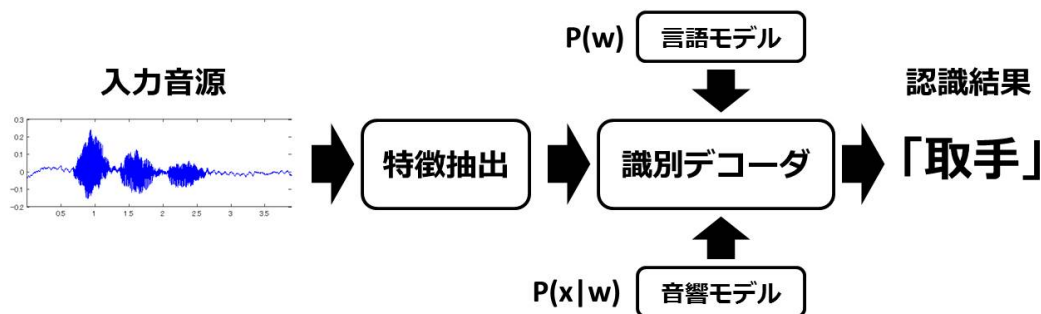


図 2.6: 音声認識システムの構成

音声認識ではまず入力音源から特徴ベクトルを抽出する．抽出された特徴ベクトルを音響モデルおよび言語モデルとのパターン認識で識別子，最も近いと識別されたパターンが結果として出力される．これは入力音声の特徴ベクトル x が観測された時，事後確率 $P(w_i|x)$ が最大になるようなクラス w_i を求めることになる．

$$\arg \max_i P(w_i|x) = \arg \max_i \frac{P(x|w_i)P(w_i)}{P(x)} = \arg \max_i P(x|w_i)P(w_i) \quad (2.3)$$

ここで， x_1, \dots, x_t を長さ t の特徴ベクトルの系列， $w = w_1, \dots, w_n$ を n からなる単語列とする．連続音声認識は，特徴ベクトル系列 x_1, \dots, x_t を観測した時に，事後確率が最大になるような単語列 $w = w_1, \dots, w_n$ を認識結果とするものになる．

$$\arg \max_w P(w_1, \dots, w_n | x_1, \dots, x_t) = \arg \max_w P(x_1, \dots, x_t | w_1, \dots, w_n) P(w_1, \dots, w_n) \quad (2.4)$$

右辺第1項は $P(x_1, \dots, x_t | w_1, \dots, w_n)$ は単語列 $w = w_1, \dots, w_n$ を発声した時に，特徴ベクトル系列 x_1, \dots, x_t がどれくらいの確率で観測されるかということを表した条件付

き確率である．条件付き確率を求めるための確率モデルは，ある単語がどのような音と対応しているかを表したものである．

右辺第2項 $P(w_1, \dots, w_n)$ は，単語列 w_1, \dots, w_n がどれくらいの確率で観察できるかを表しており，言語モデルという．言語モデルでは頻出単語は高確率，低頻度単語は低確率となる．

2.6.2 音声の特徴量抽出

現在の音声認識には，メル周波数ケプストラム係数 (Mel Frequency Cepstrum Coefficient, MFCC) が特徴量として用いられている．メルケプストラムはフーリエ変換によって求めたスペクトルに対して低周波数では細かく，高い周波数では粗い分解能をもつ人間の聴覚特性に合わせたフィルタ群を使って対数変換し，さらに逆フーリエ変換して求める．メルケプストラム空間では，基本周波数の情報 (高周波数領域) と，スペクトル概形 (低周波数領域) に分離できる．MFCC はメルケプストラム空間のスペクトル概形情報を取り出したものである

MFCC 計算におけるスペクトル分析は，図 2.7 に示すように周波数軸上に L 個の三角窓を配置して行う．つまり，窓の幅に対応する周波数帯域の信号のパワーを，単一スペクトルチャンネルの振幅スペクトル $|S'(k)|$ の重み付け和から求める．

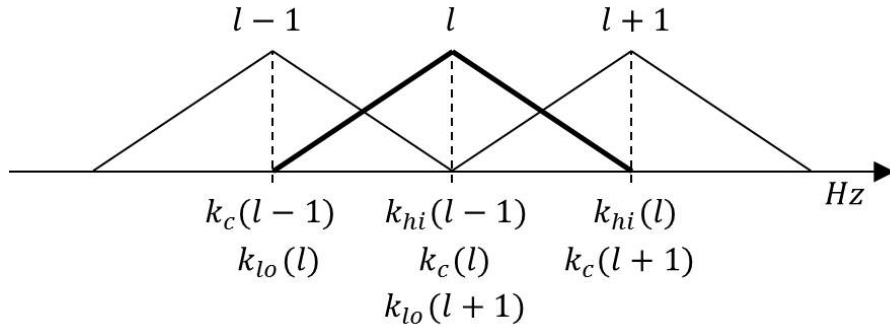


図 2.7: MFCC 分析用のフィルタバンク (23)

$$m(l) = \sum_{k=k_{lo}}^{k_{hi}} W(k; l) |S'(k)| \quad (l = 1, \dots, L) \quad (2.5)$$

$$W(k; l) = \begin{cases} \frac{k - k_{lo}(l)}{k_c(l) - k_{lo}(l)} & k_{lo}(l) \leq k \leq k_c(l) \\ \frac{k_{hi}(l) - k}{k_{hi}(l) - k_c(l)} & k_c(l) \leq k \leq k_{hi}(l) \end{cases} \quad (2.6)$$

ただし, $k_{lo}(l), k_c(l), k_{hi}(l)$ は, それぞれ l 番目のフィルタの下限, 中心, 上限のスペクトルチャンネル番号であり, 隣り合うフィルタ間で式 2.7 の関係がある.

$$k_c(l) = k_{hi}(l-1) = k_{lo}(l+1) \quad (2.7)$$

さらに, $k_c(l)$ は, メル周波数軸上で等間隔に配置される. メル周波数は式 2.8 により算出される.

$$Mel(f) = 2595 \log 10(1 + \frac{f}{700}) \quad (2.8)$$

ただし f の単位は $[Hz]$ であり, $Mel(f)$ は音声の特性を考慮した尺度である. 式 2.9 のようにフィルタバンク分析により得られた L 個の帯域におけるパワーを離散コサイン変換することで MFCC を求める.

$$c_{mfcc}(i) = \sqrt{\frac{2}{N}} \sum_{l=1}^L \log m(l) \cos\{(l - \frac{1}{2})\frac{i\pi}{L}\} \quad (2.9)$$

2.6.3 音響モデル

音響モデルとは人間の発声をモデル化したものである. 人間の発声原理は考慮せず, 単語列 $W = \omega_1, \dots, \omega_n$ を与えると確率的に人間らしい判断で特徴ベクトル系列 $X = x_1, \dots, x_t$ を生成することでモデルを実現する. 図 2.8 に音響モデルの概念図を示す. 一般的な音声認識システムでは, 音響モデルとして確率的非決定性オートマトンである隠れマルコフモデル (Hidden Markov Model, HMM) を利用している⁽²⁶⁾.



図 2.8: 音響モデルの概念図

隠れマルコフモデル (Hidden Markov Model, HMM)

HMM は時系列信号の確率モデルであり, 複数の定常信号源の間を遷移することで非定常な時系列信号をモデル化する.

HMM は発声の時間的な揺らぎ (スペクトル変動) に強く, また, モデル作成過程において多くのデータや処理を必要とする反面, 認識時の処理量が少ないという特徴を持つ.

小語彙の単語認識であれば単語をひとつの HMM に対応付けて認識することは容易に実現できるが, 大語彙の単語認識や連続音声認識において同様の処理を行うと, 単語推定

時の計算量と単語自体の記憶量が膨大になる．そのような大語彙の単語認識等では，学習や語彙の追加が簡単になる音素単位の HMM を結合する方式を用いる．この方式を使えば日本語の音素の数だけ HMM を作成すれば良いことになり，この単位で構成したものをモノフォンという．しかし，発話した個々の音はその前後に発話した音の影響を受けるため，影響を受ける前後の音を考慮した3つ組み音素をトライフォンという．つまり，音節，音素等の単語よりも細かなサブワード単位の音をあらかじめ表現すれば，これらの組み合わせで単語や文章等の長い単位を表現することが可能である．

HMM の原理

HMM による音声認識では入力された音声のスペクトル系列が観測されたとき，この系列が与えられた HMM から出力される最大の確率 (尤度) モデルを認識結果とする．

音声認識に用いられる HMM は，left-to-right 型で一つの初期状態と一つの最終状態がある構造が多い．図 2.9 は最もよく用いられる Bakis モデルと呼ばれる型の例である．

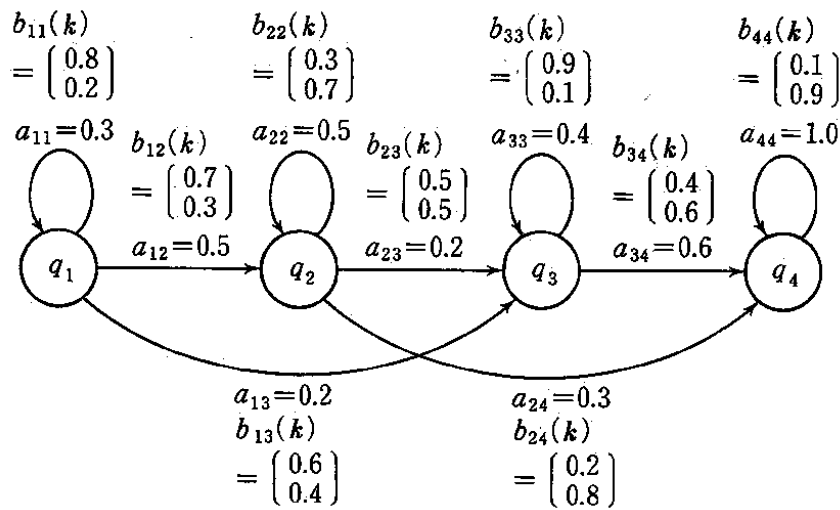


図 2.9: Bakis 型 HMM の例 (q_i : 状態, a_{ij} : 状態遷移確率, b_{ij} : 出現確率)

値 a_{ij} は状態 q_i から状態 q_j への状態遷移確率を表し，状態数を S とすると $S \times S$ の行列ができる．通常，音声パターンには時間的な非可逆性の性質があるので， $i > j$ なら $a_{ij} = 0$ である． $b_{ij}(k)$ は状態 q_i から状態 q_j への遷移で，種々のスペクトルパターンが観測 (出力) される出現確率を表し， $\{b_{ij}(k)\}$ を出現確率行列という．出現するスペクトルパターンには連続値として表す場合 (連続分布モデル) と有限個 (K 個) のシンボルの一つとして表現する場合 (離散分布モデル) がある．また，各状態 q_i の初期確率を π_i で表し，最終状態の集合を F で表す．

実際の音声認識に用いる HMM では対象に応じて適切に状態数やモデル構造 (遷移構

造) を決定するとともに、スペクトルパターンの表現法を決定する必要がある．状態数の数や複雑さ等のモデルの自由度が大きくなればよりきめ細かい変動が表現できるがモデルの推定すべきパラメータが多くなり推定精度も悪くなる．図 2.9 における数値列は，出力を $\{a, b\}$ の二つに限定し， $[]$ 内にそれぞれの出現確率を示している．この場合の遷移確率行列を以下に示す．このとき， $\pi_1 = 1, \pi_i = 0 (i > 1), F = \{q_4\}$ である．

$$A = \begin{bmatrix} 0.3 & 0.5 & 0.2 & 0.0 \\ 0.0 & 0.5 & 0.2 & 0.3 \\ 0.0 & 0.0 & 0.4 & 0.6 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \quad (2.10)$$

認識アルゴリズム

$y = y_1, y_2, \dots, y_r$ を出力シンボルの観測系列 (スペクトルの時系列パターン等) とする．このとき各 HMM モデルによって y が生起する確率 $P(y|M)$ を求め，最大確率を与えるモデルを選んで認識結果とする．この M は HMM により表現される単語や音素に対応している．

$q = q_{i0}, q_{i1}, \dots, q_{iT}$ を状態遷移系列 (ただし $q_{iT} \in F$) とすると，以下のように表現できる．

$$P(y|M) = \sum_{i0, i1, \dots, iT} P(y|q, M) \cdot P(q|M) \quad (2.11)$$

ここで，図 2.9 に基づきシンボル系列 “abb” が出力される場合の確率を求めてみる．この場合の状態遷移系列の経路は以下の 7 通りある．

$$\begin{aligned} q_1 \rightarrow q_1 \rightarrow q_2 \rightarrow q_4 & \quad P_1 = 0.3 \times 0.8 \times 0.5 \times 0.3 \times 0.3 \times 0.8 = 0.008640 \\ q_1 \rightarrow q_1 \rightarrow q_3 \rightarrow q_4 & \quad P_2 = 0.3 \times 0.8 \times 0.2 \times 0.4 \times 0.6 \times 0.6 = 0.006912 \\ q_1 \rightarrow q_2 \rightarrow q_2 \rightarrow q_4 & \quad P_3 = 0.5 \times 0.7 \times 0.5 \times 0.7 \times 0.3 \times 0.8 = 0.029400 \\ q_1 \rightarrow q_2 \rightarrow q_3 \rightarrow q_4 & \quad P_4 = 0.5 \times 0.7 \times 0.2 \times 0.5 \times 0.6 \times 0.6 = 0.012600 \\ q_1 \rightarrow q_2 \rightarrow q_4 \rightarrow q_4 & \quad P_5 = 0.5 \times 0.7 \times 0.3 \times 0.8 \times 1.0 \times 0.9 = 0.075600 \\ q_1 \rightarrow q_3 \rightarrow q_3 \rightarrow q_4 & \quad P_6 = 0.2 \times 0.6 \times 0.4 \times 0.1 \times 0.6 \times 0.6 = 0.001728 \\ q_1 \rightarrow q_3 \rightarrow q_4 \rightarrow q_4 & \quad P_7 = 0.2 \times 0.6 \times 0.6 \times 0.6 \times 1.0 \times 0.9 = 0.038880 \end{aligned}$$

これらの結果から次式を得る．

$$P(abb|M) = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 = 0.17376 \quad (2.12)$$

一般に $P(y|M)$ の値は trellis 法で求められる．forward 変数 $\alpha(i, t)$ を定義し，観測シンボル y_t を出力して状態 q_i がある確率とすれば， $i = 1, 2, \dots, S$ において，

$$\alpha(i, t) = \pi_i \quad (t = 0) = \sum_j \alpha(j, t-1) \cdot a_{ij} \cdot b_{ij}(y_t) \quad (t = 1, 2, \dots, T) \quad (2.13)$$

であり，これを計算することで最終的に

$$P(y|M) = \sum_{i, q_i \in F} \alpha(i, T) \quad (2.14)$$

を求めればよい．

図 2.10 の数値は実際に trellis 法を実行した結果を示したものである．図の縦軸が状態，横軸が時間を表し，図中の太線が最も可能性の高い状態遷移系列を示す．

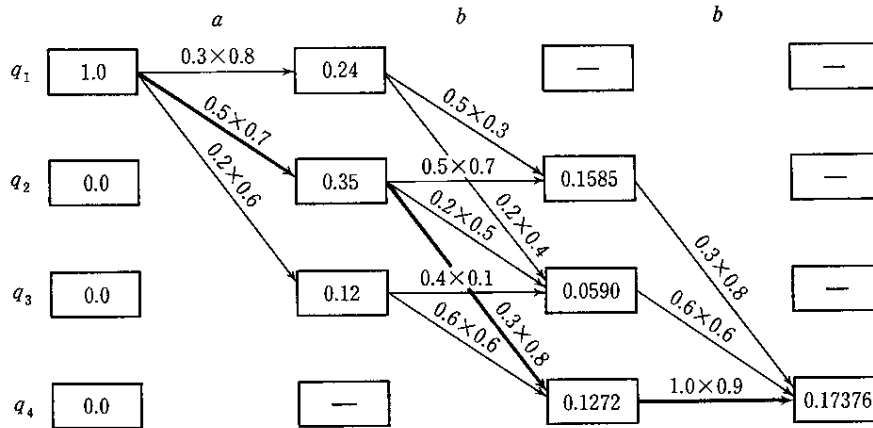


図 2.10: HMM の例でシンボル系列“abb”が出力される確率の trellis 法による計算手順

2.6.4 言語モデル

音声認識における言語モデルとは与えられた単語列 $\omega_1\omega_2\cdots\omega_n$ に対して，その出現確率 $P(\omega_1\omega_2\cdots\omega_n)$ を与えるモデルと考えることができる．現在はコーパスから統計的な手法によって自動的にモデルを作成する統計的言語モデルが主流である．この統計的言語モデルの中で，最も単純でかつ広く利用されているものが N グラムモデルである．

N グラムモデルとは， $P(\omega_1\omega_2\cdots\omega_n)$ の推定をする場合に式 2.15 のような近似を行うモデルである．

$$P(\omega_1\omega_2\cdots\omega_n) = \prod_{i=1}^n P(\omega_i|\omega_{i-N+1}\cdots\omega_{i-1}) \quad (2.15)$$

N グラムモデルでは， i 番目の単語 ω_i の生成確率が直前の $N - 1$ 単語 $\omega_{i-N+1}\cdots\omega_{i-2}\omega_{i-1}$ だけに依存すると考える．

大語彙音声認識デコーダ Julius と音声認識デコーダ Julian

julius は単語 N - gram に基づく高性能汎用大語彙連続音声認識ソフトウェアであり，数万語の語彙を対象とした連続音声認識が行える．最大の特徴は可搬性であり，単語辞書

や言語モデル、音響モデルなどの各モジュールを組み替えることで、様々な規模の音声認識システムを構築することができる。また、オープンソースであるのでソースを自由に入手でき、他プラットフォームへの移植や改造も可能である。*Julian* は有限状態文法 (DFA) に基づき入力音声に対して与えられた文法規則の元で最も尤もらしい単語系列を探索し、文字列として出力する単語認識ソフトウェアである。*Julius* の派生として開発されたもので、認識アルゴリズムはほぼ *Julius* と共通である。*Julian* を用いた認識システムの構成上の特徴を以下に示す。

- 言語モデルとして有限状態文法を、音響モデルとして HMM を使用
- 2 パス探索で高速処理を実現
 - 第1パス：単語バイグラムモデルで高速に粗い照合
 - 第2パス：単語トライグラムモデルで精密認識

2.6.5 線形予測分析 (Linear Predictive Coding, LPC)⁽²³⁾⁽²⁸⁾

線形予測法は音声の生成モデルと密接に関係しており、音声に線形予測法を適用することで、フォルマントに代表される音声の特徴パラメータを抽出することができる。逆に、こうした特徴パラメータから音声を合成することもできる。

音声の生成モデル

音声は時間軸において信号の声質が変化し、音素という基本要素に分解することができる。日本語の音素は図 2.11 に示すようにおよそ 30 個程度あるが、これらを組み合わせることで様々な言語情報を伝えることができる。

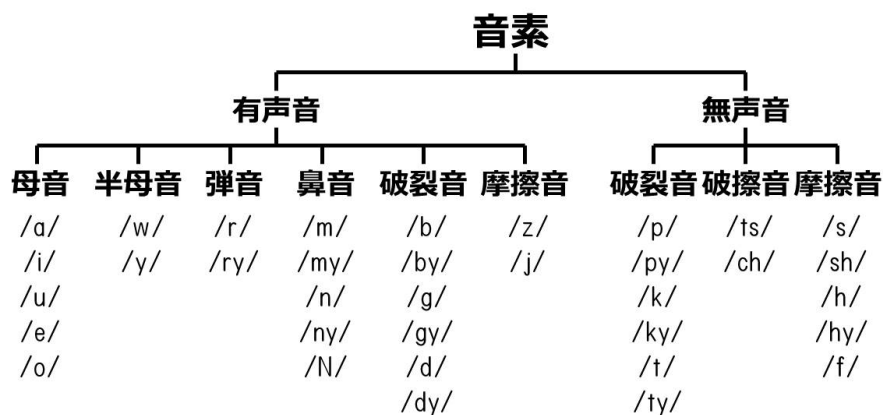


図 2.11: 日本語の音素

図 2.11 のように音素は性質の違いから「有声音」と「無声音」に分類することができる。有声音は声帯の周期振動を音源として生成され、舌の位置など声道の音響特性が反映され母音となる。無声音は声帯振動を伴わず、声道の隙間を呼気が通過するときが発生する気道雑音を音源として生成され子音となる。有声音も無声音も音源の種類は異なるが、声道の音響特性によって音色が特徴づけられるという点では共通である。

図 2.12 に音声生成モデルを示す。

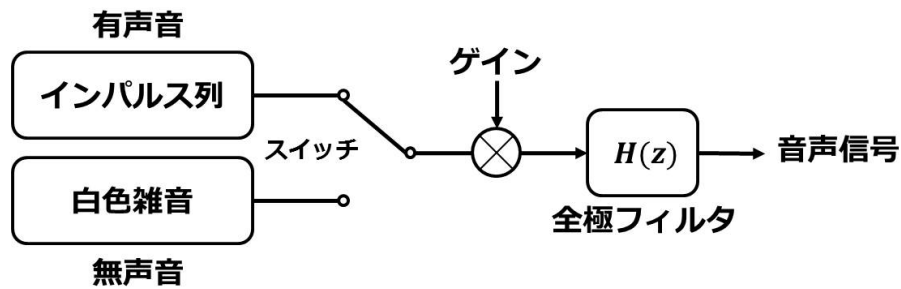


図 2.12: 音声生成モデル

図に示すように音源は有声音の声帯振動をインパルス列、無声音の気道雑音を白色雑音で近似することができる。近似された音源振動は声道の代わりとなる調音フィルタを通過することにより音声となる。LPC は音声に式 2.16 のような全極モデルの伝達関数を持つ調音フィルタの出力であることを前提として、音声スペクトル包絡を求めることができる。LPC では共振特性を決定する線形予測係数 a_i を決定することでスペクトル包絡を求める。

$$H(z) = \frac{1}{1 - a_1 z^{-1} - \dots - a_p z^{-p}} \quad (2.16)$$

具体的には過去の観測信号とモデルから生成される信号の二乗誤差を最小にする線形予測係数を求める。

線形予測分析⁽²³⁾

全極形の伝達関数を仮定することは、調音フィルタを図 2.13 に示すような IIR 形のネットワークで表現することと等価である。

このフィルタの出力信号 $s(n)$ は、過去の p 個のサンプルに重み a_i を掛けてフィードバックした信号の和を式 2.17 と表す。

$$s_w(n) = \sum_{i=1}^p a_i s_w(n-i) + g(n) \quad (2.17)$$

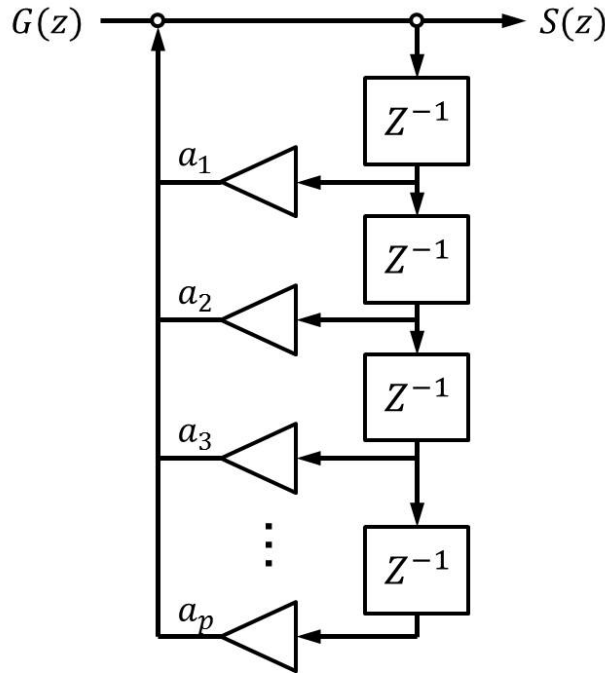


図 2.13: LPC モデル

ここで、フレーム化処理によって得られた長さ N の信号 $s_w(n), s_w(n-1), \dots, s_w(n-N+1)$ が与えられたとする．このとき式 2.17 に基づき、 $s_w(n-1), \dots, s_w(n-p)$ となる p 個のサンプルから計算させた $s_w(n)$ の推定値

$$\hat{s}_w(n) = \sum_{i=1}^p a_i s_w(n-i) \quad (2.18)$$

と観測された $s_w(n)$ との差 $\{s_w(n) - \hat{s}_w(n)\}^2$ ($n = 0, \dots, N-1$) を誤差信号 $e(n)$ と呼ぶ．いま、誤差信号のパワーの期待値 $D = E[\sum_{n=0}^p e^2(n)]$ を最小化する a_i は以下の手順で求めることができる．

誤差信号が、 $\alpha_0 = 1, \alpha_i = -a_i$ とおくことで、式 2.19 と計算でき、その信号のパワーは、式 2.20 で求められる．

$$e(n) = s_w(n) - \hat{s}_w(n) = s_w(n) + \sum_{i=1}^p \alpha_i s_w(n-i) \quad (2.19)$$

$$e^2(n) = \left\{ \sum_{i=0}^p \alpha_i s_w(n-i) \right\} \left\{ \sum_{j=0}^p \alpha_j s_w(n-j) \right\} = \sum_{i=0}^p \sum_{j=0}^p \alpha_i \alpha_j s_w(n-i) s_w(n-j) \quad (2.20)$$

そして、求めた信号のパワーの期待値は式 2.21 のように計算される．

$$D = E\{e^2(n)\} = \sum_{i=0}^p \sum_{j=0}^p \alpha_i \alpha_j s_w(n-i) s_w(n-j) = \sum_{i=0}^p \sum_{j=0}^p \alpha_i \alpha_j R_{ss}(|i-j|) \quad (2.21)$$

ただし $R_{ss}(l)$ は信号 $s_w(n)$ の自己相関関数を表す．そして、 $k = 1, \dots, p$ に対して

$$\frac{\partial D}{\partial \alpha_k} = 2 \sum_{i=0}^p \alpha_i R_{ss}(|i - k|) = 0 \quad (2.22)$$

とおき、

$$\sum_{i=0}^p \alpha_i R_{ss}(|i - k|) = -R_{ss}(|-k|) \quad (k = 1, \dots, p) \quad (2.23)$$

となる p 個の方程式を連立させることで式 2.24 を得る．

$$\begin{bmatrix} R_{ss}(0) & R_{ss}(1) & \cdots & R_{ss}(p-1) \\ R_{ss}(1) & R_{ss}(0) & & \vdots \\ \vdots & & \ddots & R_{ss}(1) \\ R_{ss}(p-1) & \cdots & R_{ss}(1) & R_{ss}(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = - \begin{bmatrix} R_{ss}(1) \\ R_{ss}(2) \\ \vdots \\ R_{ss}(p) \end{bmatrix} \quad (2.24)$$

つまり LPC 信号処理過程では、自己相関関数を介して信号のスペクトルが代数方程式により与えられる．

伝達特性は、求められたパラメータである線形予測係数 a_i を用いて式 2.25 を求めることができる．

$$|H(e^{j\omega})|^2 = \frac{1}{|1 - \sum_{i=1}^p \alpha_i e^{ji\omega}|^2} \quad (2.25)$$

第 3 章

システム評価用音声データベースの構築

発声機能障害者支援システムの評価のため作成した音声データベース⁽²⁹⁾について述べる。提案システムの利用対象者は発声機能障害者であるが、発声機能障害には大別して発声障害および構音障害がある。このように対象となる障害者については障害の程度、また老若男女様々なケースが考えられるため、多くのパターンの評価への使用を目的として、出来る限りの音声および体内伝導音を収録しデータベース化した。

3.1 収録音源と環境

本データベースでは、表 3.1 に示すように健常者 4 名と機能障害者 5 名の気道音声および体内伝導音を収録した。各被験者の気道音声は口元から 30cm の位置で計測用マイクروفोनを用いて収録した。同時に発声時体内伝導音を加速度ピックアップセンサを用いて図 3.1 のように上唇、声帯、乳様突起の三か所から抽出した。気道音声、体内伝導音ともにサンプリング周波数 16KHz とした。また、収録に使用した機器を表 3.2 に示す。

本データベースの機能障害者 1 から機能障害者 4 は声帯に障害のある発声障害であり、機能障害者 5 のみ構音障害である。表 3.1 にあるように機能障害者 1 は声帯ポリープ患者、機能障害者 2 は声帯結核患者、機能障害者 3 は片側声帯麻痺患者、機能障害者 4 は声帯摘出者、機能障害者 5 は構音障害者である。収録時の発話内容を下記に示す。

母音： [あ, い, う, え, お]

撥音： [ん]

単語： 地名 100 単語⁽³⁰⁾

文章： 50 文⁽³¹⁾

被検者	年齢	性別	症状
健常者 1	22	男	異常無し
健常者 2	21	男	異常無し
健常者 3	21	女	異常無し
健常者 4	21	女	異常無し
機能障害者 1	28	男	声帯ポリープ
機能障害者 2	30	女	声帯結末
機能障害者 3	80	男	片側声帯麻痺
機能障害者 4	61	男	声帯全摘出，音声再建法
機能障害者 5	50	男	舌および顎の一部欠損

表 3.1: システム評価用データベース被験者



図 3.1: 収録の様子

また，構音障害者へ対する本システムの有効性を検討するために機能障害者 5 のみ拡大五十音図の発話データも収録した．実験に使用する音声データは，電子協・地名百選から抜粋した地名 10 単語を被験者が静寂下において 3 回繰り返し発話した計 30 語を用いた．発話地名は，“旭”，“恵庭”，“八戸”，“気仙沼”，“北見”，“札幌”，“取手”，“蕨”，“横手”，“行橋”である．

機器	仕様	
計測用マイクロフォン 株式会社 小野測器 MI-1235	周波数範囲	20Hz ~ 20kHz
	最大音圧レベル	143dB
	自己雑音レベル	19dB 以下
プリアンプ内蔵型加速度検出器 株式会社 小野測器 NP-3120	周波数範囲	5Hz ~ 5kHz $\pm 0.5\text{dB}$ 5Hz ~ 12kHz $\pm 3\text{dB}$
	共振周波数	約 50KHz
	最大使用加速度	2200m/s ²
マイクロフォン用プリアンプ 株式会社 小野測器 MI-3310	周波数範囲	10Hz ~ 100kHz
	入力インピーダンス	約 5G Ω
	自己雑音	6.3 μVrms 以下
センサアンプユニット 株式会社 小野測器 PS-602	周波数特性	3Hz - 50kHz $\pm 0.1\text{dB}$ 0.5Hz ~ 100kHz 0~+1 / -3dB
	位相特性	0.5Hz の時 +50° 以内 3Hz の時 +10° 以内 50kHz の時 -20° 以内
	ノイズレベル	50 μVrms (入力換算, DC~100kHz)
レコーディングユニット 株式会社 小野測器 AU-4100A	サンプリング周波数	48KHz / 96kHz
	周波数レンジ	18.75kHz / 37.5kHz
	チャンネル数	5ch

表 3.2: 収録機器仕様

3.2 発声障害⁽³³⁾

発声は肺から吐き出される呼気を用いて喉頭にある声帯を振動させ音源を生成することである。発声障害は音源を生成する器官や機能を妨げられるために、声が出ない、嗄声といった症状が現れる。本音声データベースのうち、機能障害者1から4は発声障害であるが、以下それぞれの症状について述べる。

3.2.1 声帯ポリープ

声帯ポリープは声帯に与える激しい刺激や炎症により粘膜上皮下に出血が起き、声帯の粘膜が一部膨張または突出したもので、声枯れや音域の変化が発生する。図3.2は声帯ポリープ患者の喉の写真である。

3.2.2 声帯結末

指にできるペンだこのように、声帯への慢性的な機械的刺激により炎症性の腫瘤ができ、声枯れや嗄声などの症状が現れる。図3.3に声帯結末患者の喉の写真を示す。

なお、機能障害者1および機能障害者2は声帯ポリープおよび声帯結末と病名こそ違うが、症状が酷似しており実験結果もほぼ同じになったため、以下の実験考察では機能障害者2を対象から除いている。

3.2.3 片側声帯麻痺

声帯を動かす神経が麻痺し、声帯を十分に開閉できなくなる。息漏れするような声枯れや、誤嚥、むせるといった症状が現れる。図3.4に片側声帯麻痺患者の喉の写真を示す。

3.2.4 声帯摘出

口腔・咽頭および喉頭癌に罹患した場合、手術により声帯を摘出する場合がある。特に下咽頭、喉頭癌を罹患すると声帯摘出を伴う場合が多い。声帯を摘出した場合は失声してしまうが、食道発声法等の代用発声法を習得することで声を取り戻すことも可能である。図3.5は声帯を摘出している機能障害者4の喉の写真である。

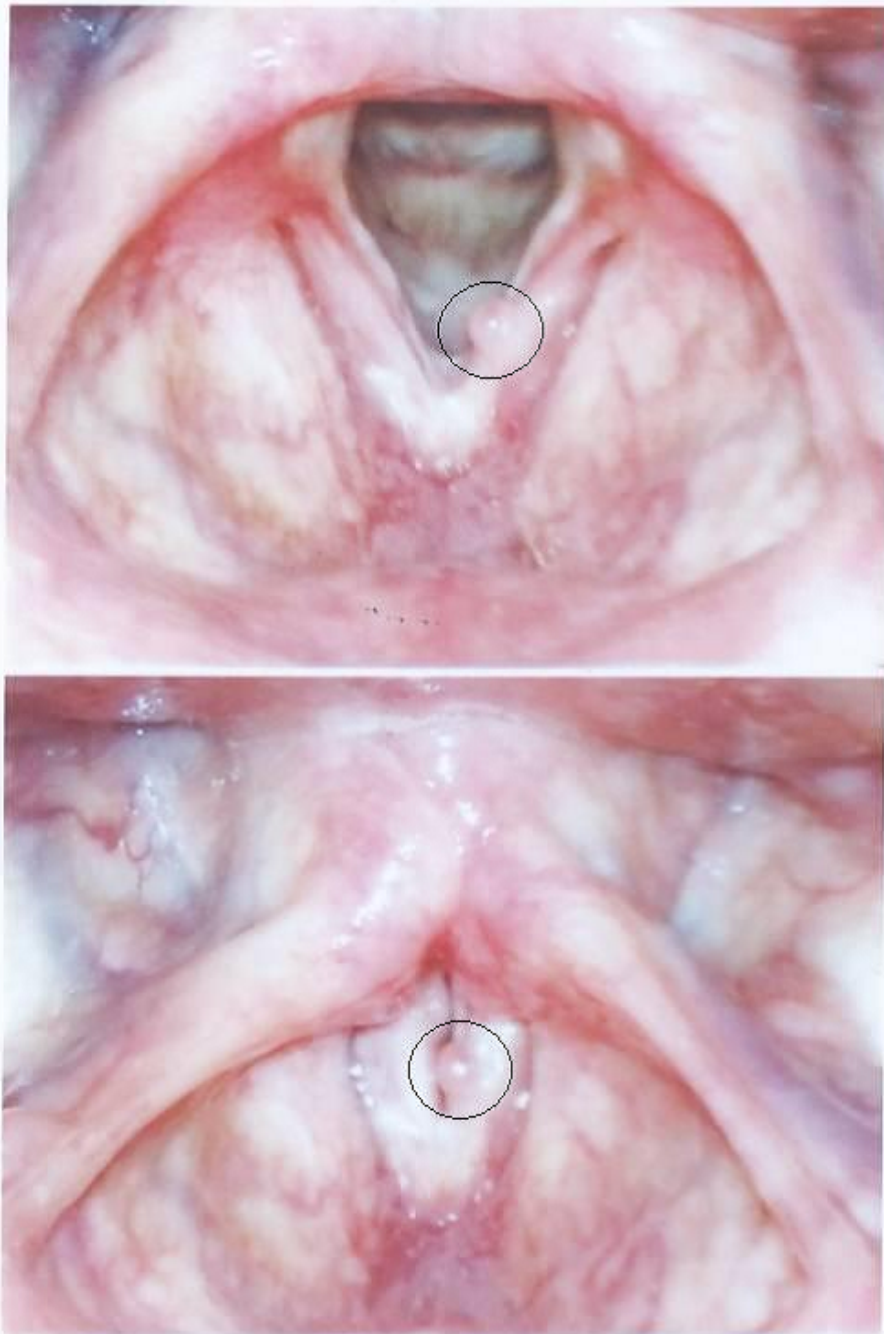


図 3.2: 声帯ポリープ患者の喉の写真（音声データベースの機能障害者1）

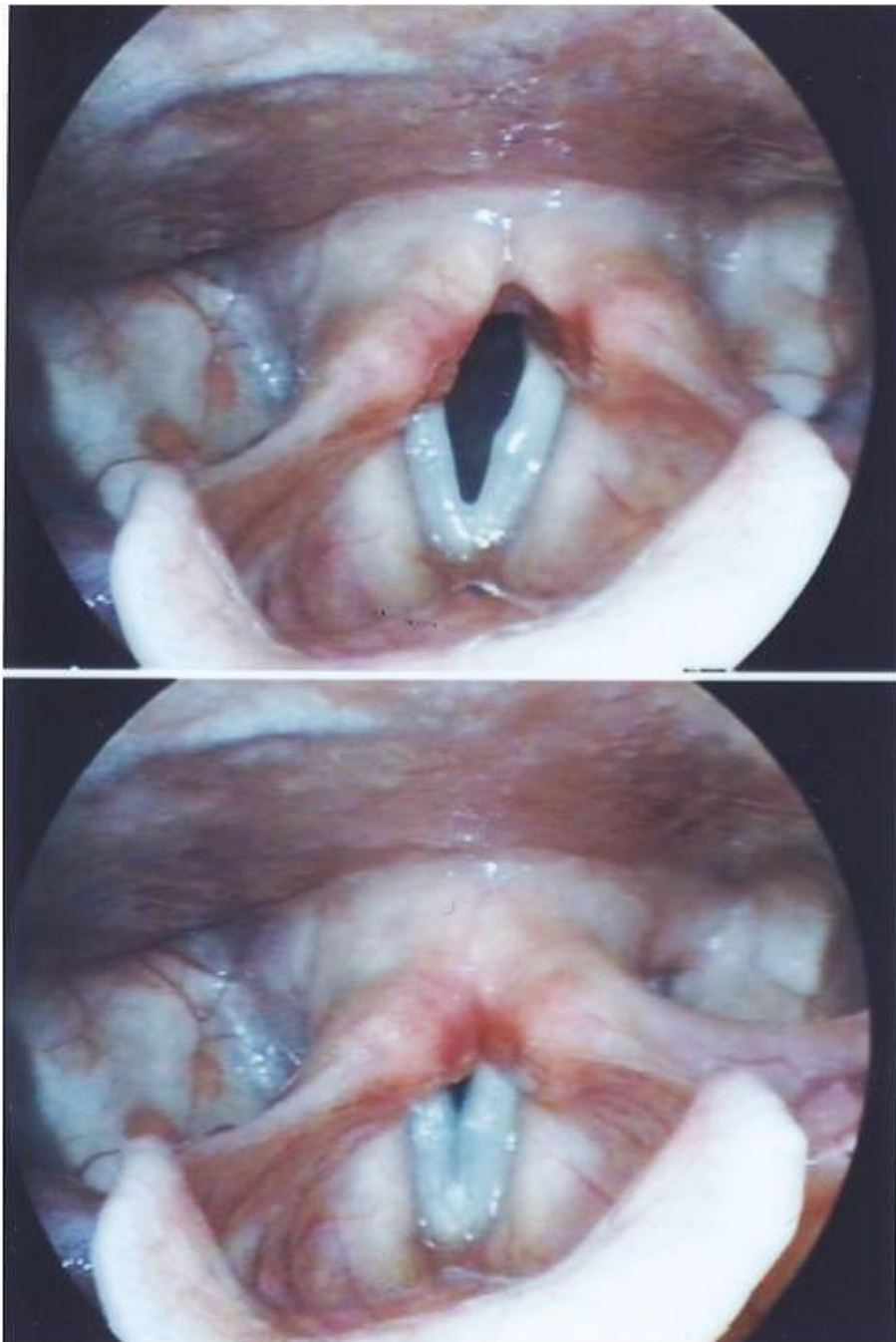


図 3.3: 声帯結束患者の喉の写真（音声データベースの機能障害者 2）

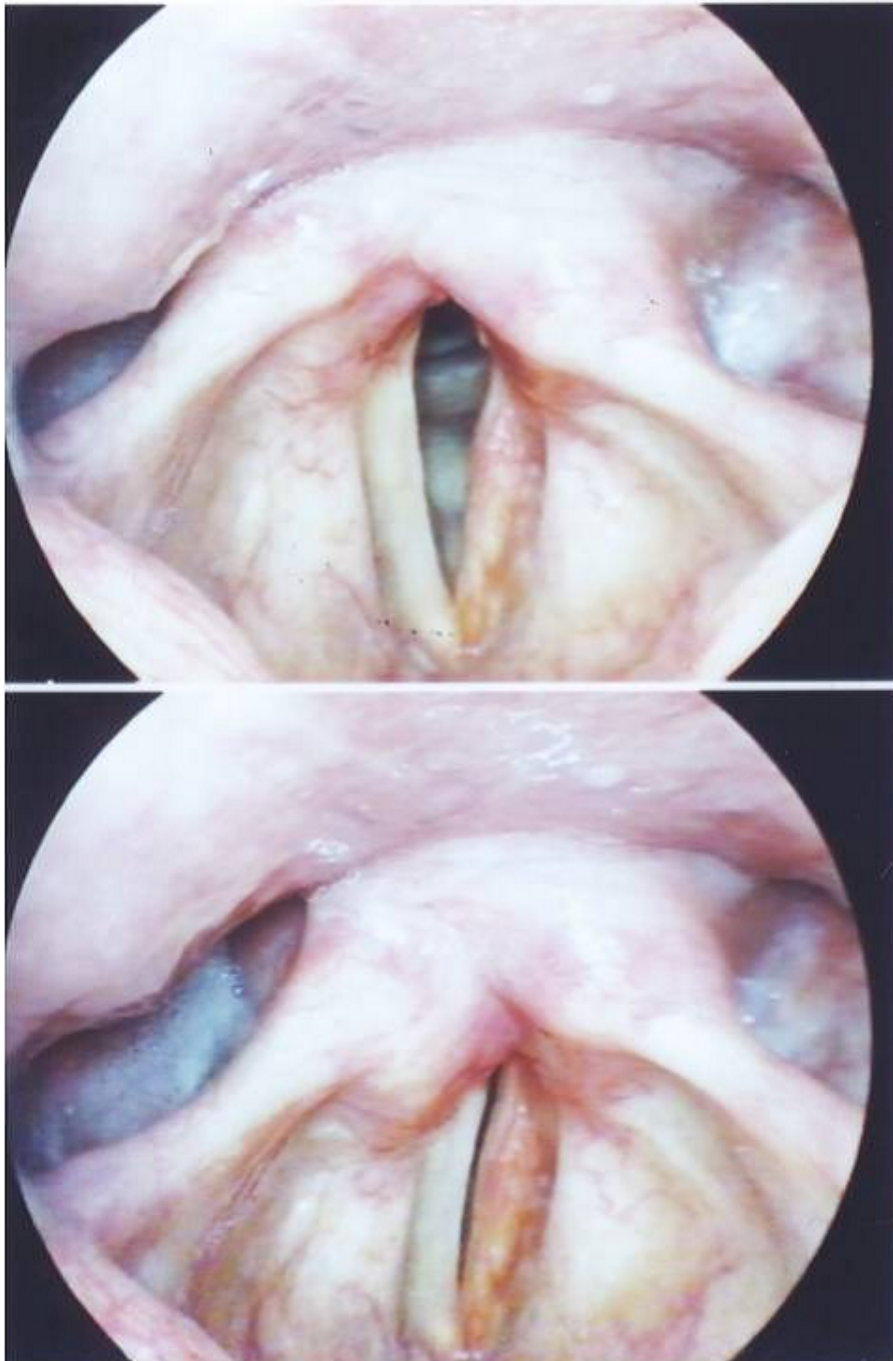


図 3.4: 片側声帯麻痺患者の喉の写真（音声データベースの機能障害者3）

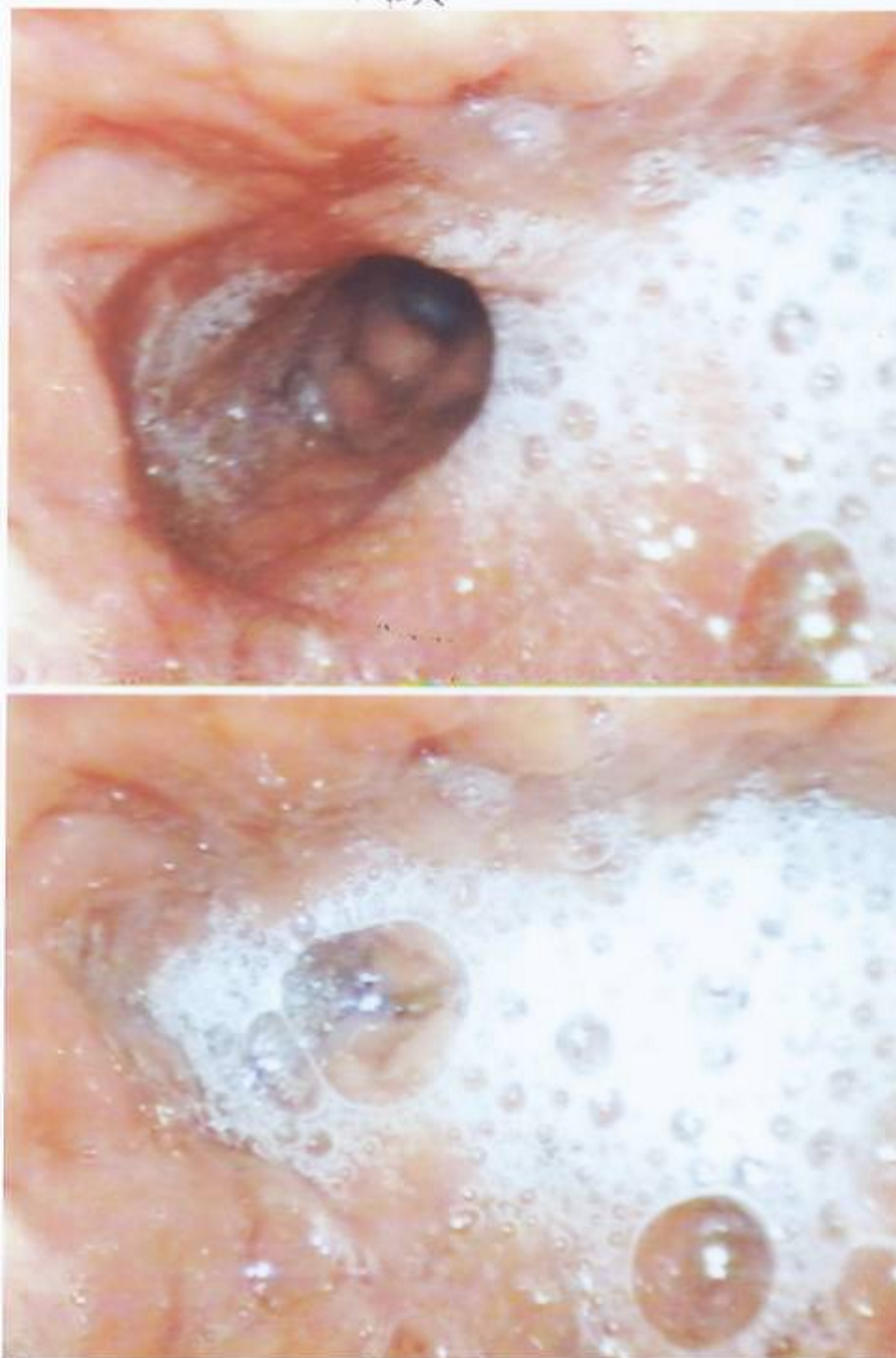


図 3.5: 声帯摘出患者の喉の写真（音声データベースの機能障害者 4）

3.3 代用発声法

声帯摘出等で失声した場合、代用発声法によりコミュニケーションを取ることが可能である。代用発声法はどこの音源を使うか、特別な手術は必要か、器具を使うかなどで分類することが出来る。以下、代表的な代用発声法について述べる。図 3.6 に代用発声法の音源および発声方法を、表 3.3 に各発声法の特徴を示す。

	音源	手術有無	音量	音質	手の使用	使い勝手	体力消耗
食道発声法	食道の空気	不要	×	肉声	不要	○	大
T-E シャント法	肺の呼気	必要	△	肉声	片手	△	中
笛式人工喉頭	肺の呼気	不要	○	良い	片手	×	中
電気式人工喉頭	電気	不要	△	機械的	片手	○	小

表 3.3: 代用発声法の特徴

3.3.1 食道発声法

日本では最も習得者が多い代用発声法である（図 3.6b）。「げっぷ」の原理を応用し、食道に空気をためて吐き出す時の音を利用する。食道による音源で発声する肉声であるため、訓練すれば健常者に近い発声が出来ようになる。特殊な手術や器具が不要で余計な費用もかからない反面、習得に長い時間がかかる。しかし、3000cc 以上ある肺による呼気ではなく、通常 100cc から 150cc 程度と言われる食道へためた空気による発声のため音量が小さくなる。また、一語発声するたびに空気をためる必要があり、体力消耗も激しいため長時間の会話が困難となる。

3.3.2 T-E シャント法

気管 (Trachea) と食道 (Esophagus) を「ヴォイスプロテーゼ」と呼ばれるシリコン製のチューブでつないで連絡路 (シャント) を作る (図 3.6c)。声帯摘出時にあけた「永久気管孔」を指でふさぐことでシャントを通して肺からの空気を食道に引き込み、食道の粘膜を震わせて声を出す。T-E シャント法は肺からの空気を使って発声するため発声音量は健常者の発声と同等である。食道発声法に比べて少しの訓練で自然な会話ができるが、最初の手術とヴォイスプロテーゼの定期的な交換が必要になる。また、発声時に気管孔をふさぐため片手がふさがってしまうという問題もある。欧米では主流の代用発声法だが、シャントを作る手術が必要になる上にメンテナンスも費用がかかり面倒なため、日本国内の普及率は低い。

3.3.3 笛式人工喉頭

気管孔から吐き出される呼気で「タピアの笛」と呼ばれる笛を鳴らし、その音をチューブで口腔内に導いて発声する（図 3.6d）。この方法も肺からの呼気を使って発声するため音量は大きい。特別な手術は不要で簡単に習得できるうえ、熟練するとかなり自然な発声が可能になる。反面、笛やチューブといった器具を使用するため片手がふさがるうえ、見た目が気になる場合もある。さらに口にチューブを入れるため衛生面でも課題がある。

3.3.4 電気式人工喉頭

電氣的に振動する器具を顎下部に押し当てることで声道を振動させ発声する方法である。習得期間は不要ですぐに使えるが、音源が機械振動となるため、改善はされているものの抑揚がつけづらく機械的な音声になるといった問題がある。また、器具を押し付ける必要があるため片手がふさがってしまう。

以上のように、現在主流として使われている代用発声法の中で自然な発声を求めると習得が困難または特殊な手術が必要となる。反面、容易に習得できる手法では機械的な声になってしまう。また、器具（片手）を使うものがほとんどで気軽に会話できる支援ツールとは言い難い。日本国内で最も普及している食道発声法でも音量が小さくなり、長時間話せないなど実用面での課題がある。このように課題のある食道発声法が最も普及しているということから、自分の呼吸器や調音器官を使った肉声の発声が望まれていることがわかる。

3.4 構音障害

喉頭の声帯振動によって生成された音源は、喉頭より上の唇、舌、口蓋といった構音器官によって言語音として意味付けされる。この構音機能を妨げられることを構音障害と言い、正しく言語の意味付けすることができなくなる。構音障害は原因によって以下の3つに分類される。

3.4.1 器質性構音障害

構音器官の形態的な要因によっておこる障害をいう。先天性の形態異常として口蓋裂等がある。口蓋裂は口蓋の一部が裂け、口腔と鼻腔が直接交通するため、発声時に空気が漏れてしまい鼻にかかったような声になる。

3.4.2 運動障害性構音障害

発声に関わる神経や筋肉の病変によって起こる障がいである．運動ニューロン病 (MND) やパーキンソン病，多発性硬化症 (MS) などが原因となる．

3.4.3 機能性構音障害

前述の二つの構音障害に当てはまらない構音障害を機能性構音障害という．言語発達の遅れや構音器官の運動機能の遅れ，言語環境等による構音の稚拙さ等の発達に関わるものが含まれる．また，事故による調音器官欠損なども含まれる．

発声障害では主に声帯の損傷や摘出による駆動音源の発声に関する障害が出るが，構音障害に罹患すると調音機能に障害が発生するため，舌の位置，開口度といった調音素性が欠落し，極めて不明瞭な音声となってしまう．

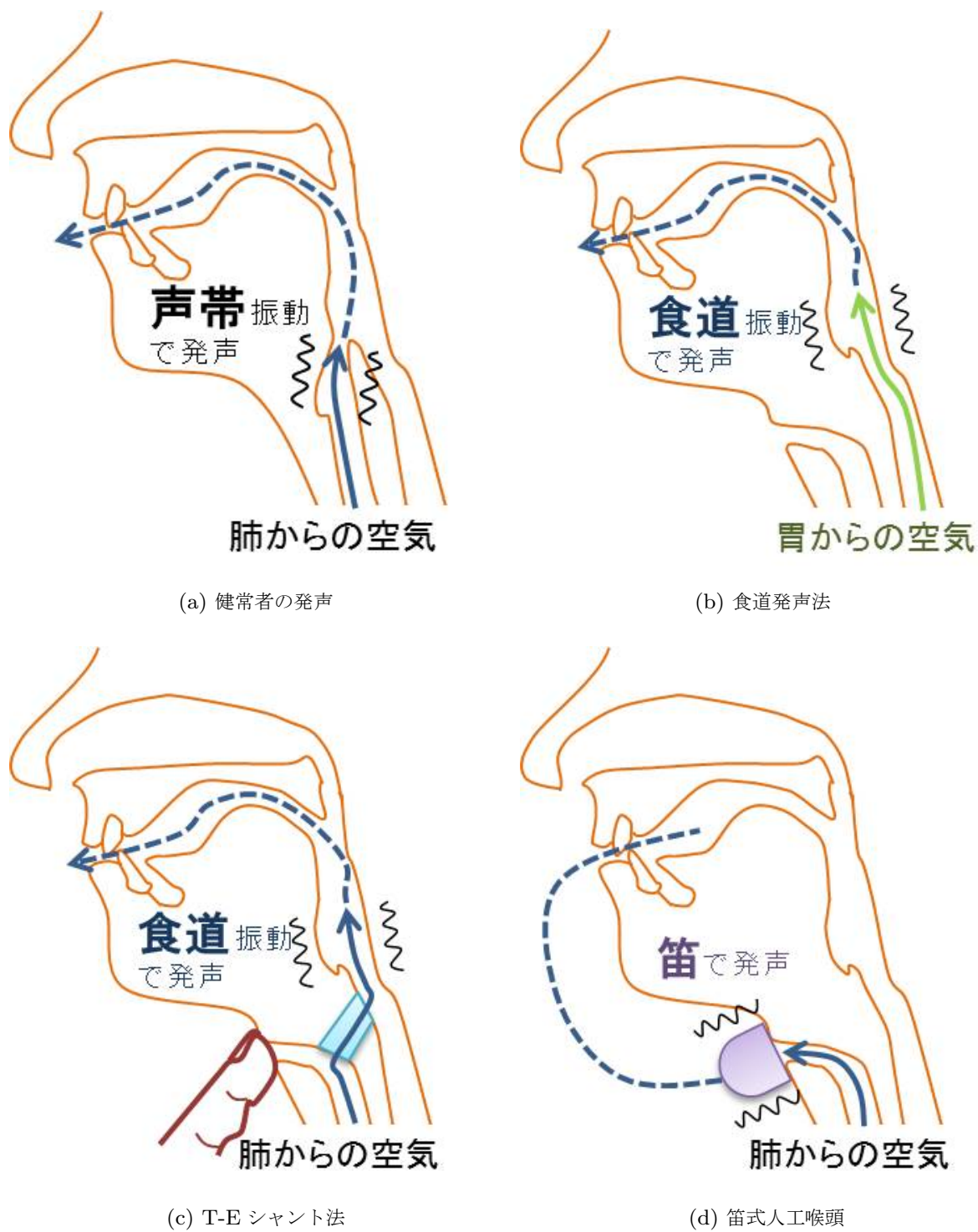


図 3.6: 代用発声法

第4章

発声機能障害者支援システムの評価

4.1 音声認識評価

予備実験として音声データベース中の障害の程度が異なる男性5名（健常者1，機能障害者1，機能障害者3，機能障害者4，機能障害者5）の気道音声および体内伝導音の離散単語認識実験を行った。体内伝導音は上唇，声帯付近，乳様突起の三か所で抽出したものである。各被験者で障害の程度が異なるため，本システムがどの程度の障害にまで適用できるか検討することができる。認識デコーダには julian-v3.1⁽²⁷⁾ を用い，音響モデルは性別非依存型の tri-phone モデル，言語モデルは web テキストから学習した 6 万単語の bi-gram である⁽²³⁾。離散単語認識は発声単語数に対する認識率の割合である単語認識率で評価する。つまり前述の地名 10 単語×3 セット＝計 30 語に対してどれだけ認識できたかが単語認識率である。

各被験者における全信号収録部位の離散単語認識率を図 4.1 に示す。

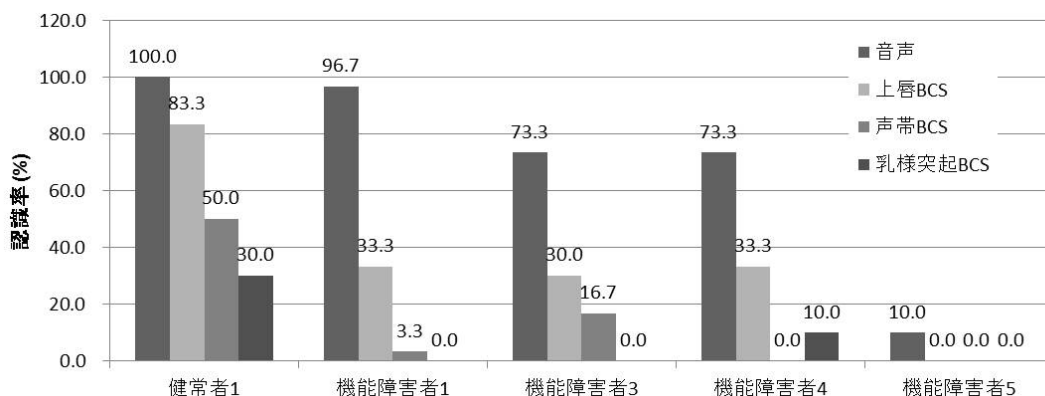
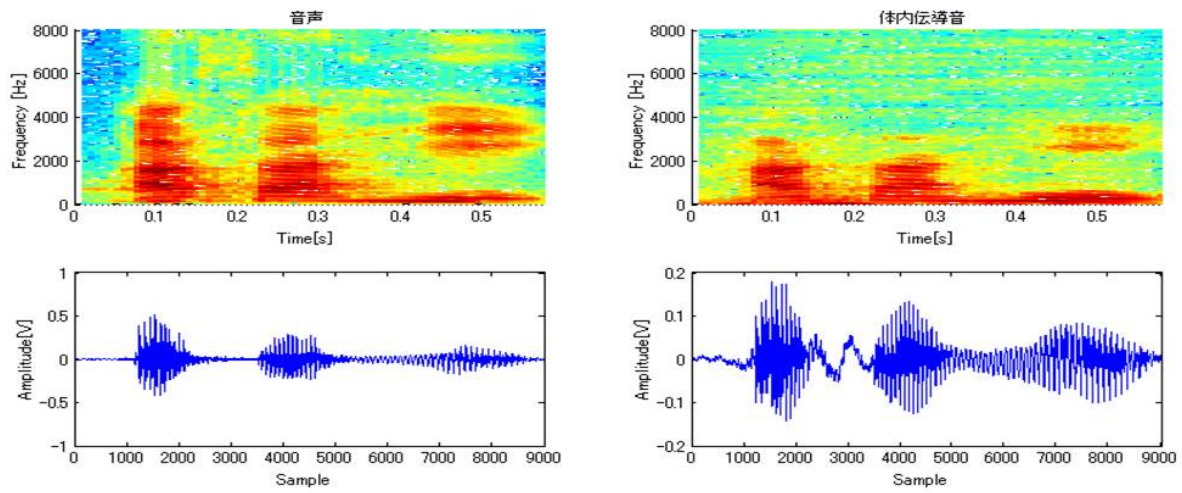


図 4.1: 気道音声と体内伝導音の音声認識率

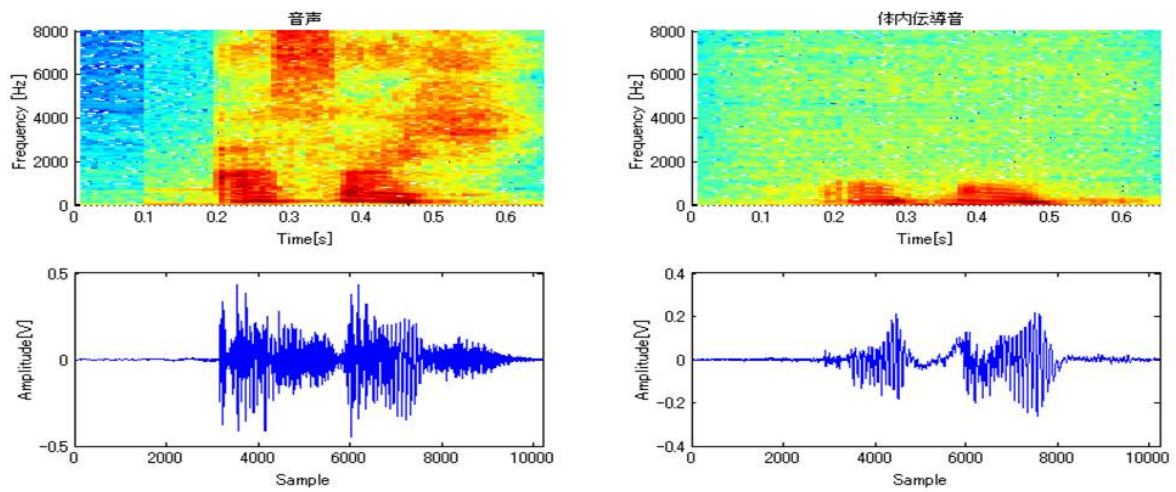
図 4.1 より、健常者においても体内伝導音（BCS）では認識率が低下していることがわかる。最も音声に音響的特徴に近いことが知られている上唇 BCS⁽³⁴⁾ では 83.3% だが、声帯 BCS および乳様突起 BCS ではそれぞれ 50%、30% と非常に低い認識結果となっている。これは気道音声と体内伝導音では音響的特徴が異なるうえ、本実験で使用している音響モデルが気道音声に基づいて作られていることが認識率低下の原因になっている。また障害者において、発声障害者である声帯ポリープ患者（機能障害者 1）、片側声帯麻痺患者（機能障害者 3）、声帯摘出者（機能障害者 4）は気道音声認識率こそ 50% を超えているが、体内伝導音の認識率が非常に悪いことが確認できる。これは健常者と比較して音声不明瞭なため体内伝導音も同様に不明瞭になり、認識率を更に低下させているためだと思われる。

さらに構音障害者（機能障害者 5）の認識率は音声で 10% 程度であり、他の部位では全く認識できていない。これは構音が音声に意味付けを行う器官であり、障害になることで子音等の情報が正確に生成できないためであると考えられる。

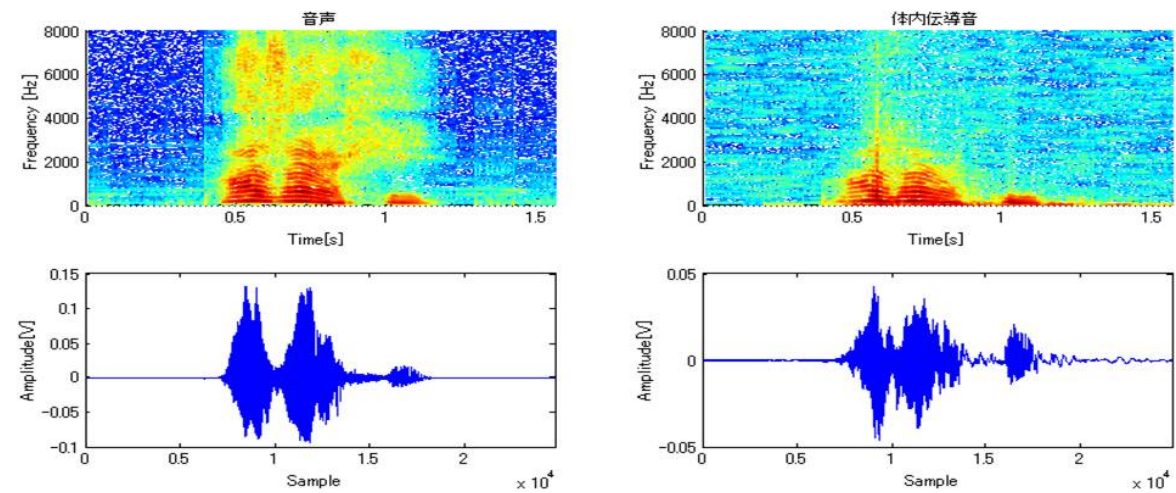
本研究では、これらの結果を踏まえ、明瞭化の対象を上唇で抽出した体内伝導音のみとした。図 4.2a から図 4.2h に各被験者が地名”旭”を発話したときの音声および体内伝導音を示す。それぞれ図の左側が気道音声、右側が体内伝導音で、体内伝導音は上唇で収録したデータを表示している。上はスペクトログラム、下は音声波形である。



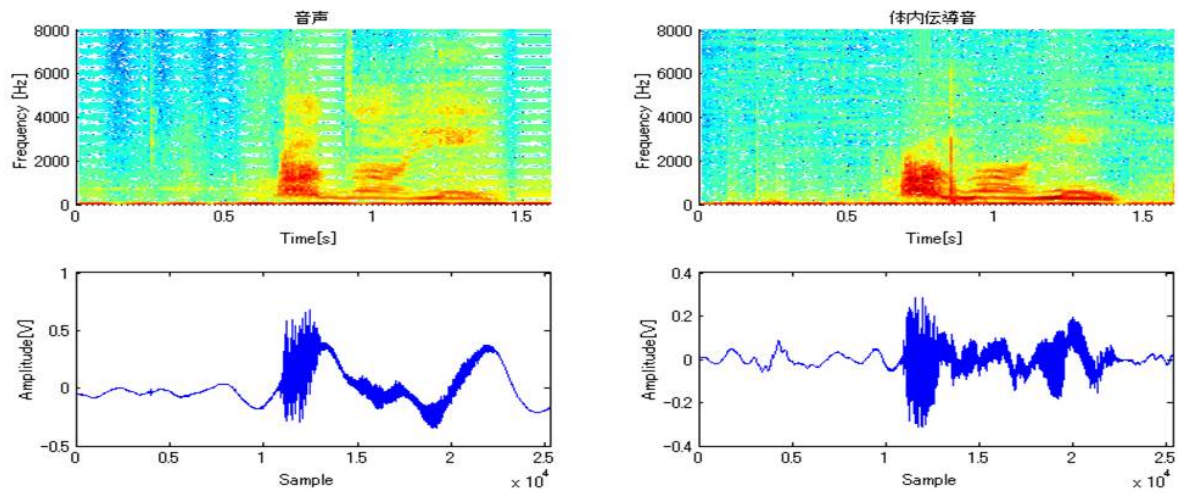
(a) 健常者 1



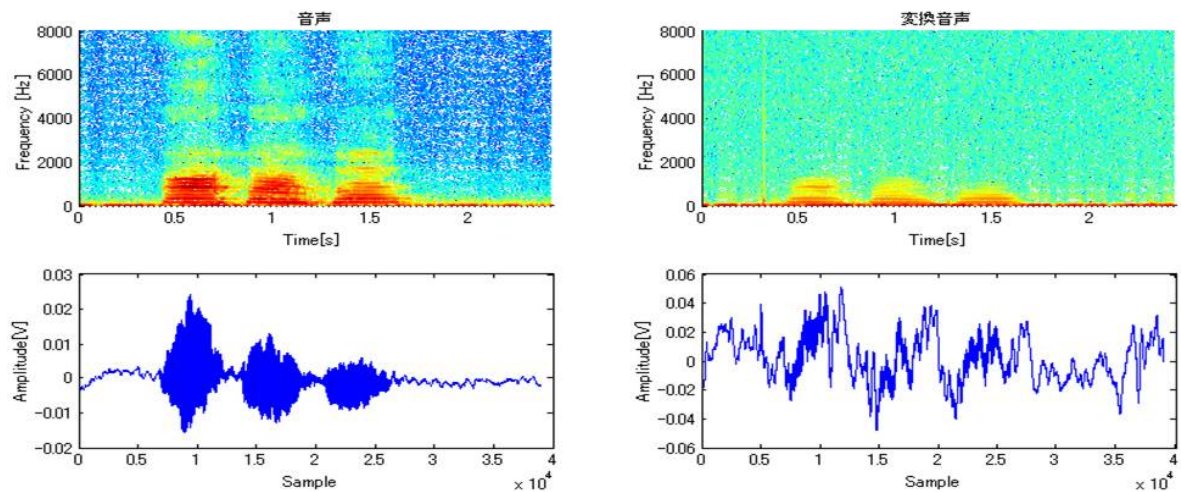
(b) 機能障害者 1 (声帯ポリープ)



(c) 機能障害者 3 (片側声帯麻痺)



(g) 機能障害者 4 (声帯摘出)



(h) 機能障害者 5 (構音障害)

図 4.2: 気道音声と上唇 BCS の音声波形およびスペクトログラム

これらの図より、スペクトログラムでは音声に比べ体内伝導音では高周波成分が弱くなっていることが分かる。特に機能障害者 5 は調音ができないため、気道音声でも高周波成分がかなり減衰している。また、時間波形においては健常者のように一音節ごとにはっきりとした発声ではなく、各音節間がつながるような不明瞭な発声になっていることが分かる。以上のように、障害者は健常者と音声異なるため認識率の低下が生じている。

4.2 音節区間の推定

前述のように本システムは、体内伝導音を音声認識することで音節区間を推定し、音節区間ごとに合成音声を生成する。本研究で提案した LPC による音声合成の適切性を評価するため、人間が体内伝導音を聴取し、その音声波形を確認しながら手作業で各音節区間を決定するヒューリスティックな手法で区間推定して生成した合成音声の認識実験を行った。この手法で区間推定を行うことで、各音節の区間推定がほぼ正確に推定できていると仮定することができる。また、システムによる自動区間推定を実現するため入力音声の連続サブワード認識 (Continuous Sub-word Recognition, CSR) ⁽³⁵⁾⁽³⁶⁾ による区間推定で生成した合成音声の認識実験も行った。CSR は入力音声を 25ms の時間フレームで 10ms ずつシフトしながら連続認識させ、その認識結果をもとに入力音声の単音節との対応付けを行い、各音節区間を推定した。

単語認識結果を図 4.3 に示す。比較のため気道音声および上唇から取得した体内伝導原音の認識結果も記載する。

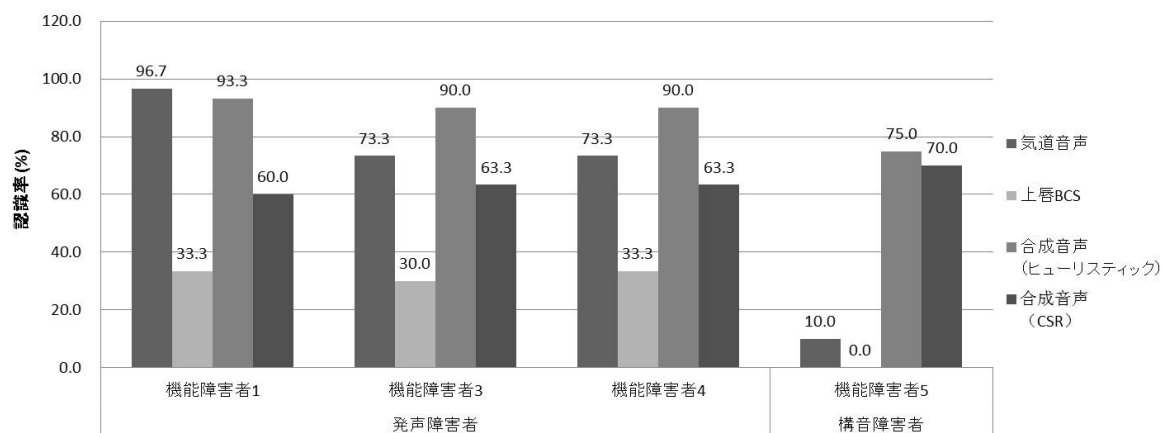
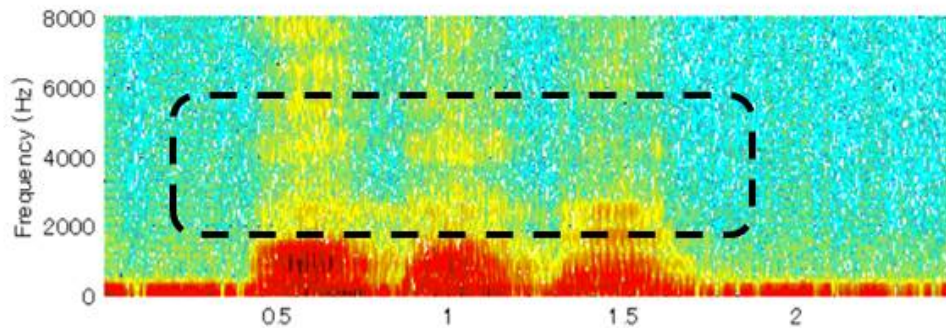


図 4.3: 合成音声の音声認識率

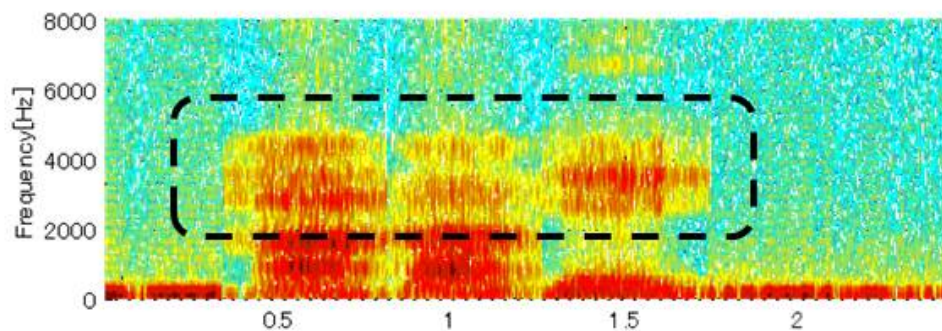
ヒューリスティックな手法および CSR で生成した合成音声は全ての被験者で上唇の体内伝導音と比較して認識率が大幅に向上していることが確認できる。ヒューリスティックな手法で生成した合成音声は気道音声に対して機能障害者 1 (声帯ポリープ患者) では若干低下するが、それ以外は同水準以上の認識率である。特に構音障害者に関しては、元々 10% であった認識率が 75% まで上昇した。これは本来、高域成分がほとんど現れなかった被験者体内伝導音に対し、合成音声では高域成分が回復できたためだと考えられる。

図 4.4 に機能障害者 5 の体内伝導音とヒューリスティックな手法で生成した合成音声の

スペクトログラムを示す.



(a) 体内伝導音



(b) 合成音声 (ヒューリスティック)

図 4.4: 機能障害者 5 の体内伝導音および合成音声のスペクトログラム

体内伝導音では 2KHz 以上の高域部分 (図 4.4a 枠内) 減衰してしまうが, 本システムで合成した音声では図からも 2KHz 以上の高域部分 (図 4.4b 枠内) が回復しているのがわかる.

以上の結果より, 入力された体内伝導音声の音節区間の推定さえ正しく行えれば, LPC による合成音声は障害者気道音声と同水準以上の音声を表現できることが確認できた.

CSR に基づく合成音声は, ヒューリスティックな手法で区間推定した合成音声ほど認識率は向上しなかったが, 一般に実用レベルといわれる認識率 60% 以上を全被験者で達成できたことから, CSR による体内伝導音認識の自動化が可能であることが確認できた.

第 5 章

構音障害者への適用とその評価

5.1 実験の目的

音声データベースの機能障害者 5 は構音障害者であり、音声生成の過程で調音することができず、発声が非常に不明瞭になる。そこで、本支援システムを用いて機能障害者 5 の体内伝導音から生成した合成音声と気道音声に対して聴取実験および調音素性分析を行い、各音声に含まれる調音素性の伝達率を調査した。合成音声の調音素性伝達率が気道音声と比較して向上していれば本システムに調音機能再現性があると確認できる。

5.2 調音素性

人間は口や鼻から呼気を放射するときに、声道の形を変えることで言語音を生成する。この声道の形を変え調音する器官のことを調音器官といい、詳細には図 5.1 に示すようなものがある。

5.2.1 調音の仕組み

人間が発声に用いる器官は声帯と声道、鼻腔の 3 つだけである。我々は無意識のうちにこの 3 つの器官を巧みに変化させて、例えば日本語では 126 種類もの語を発声している。

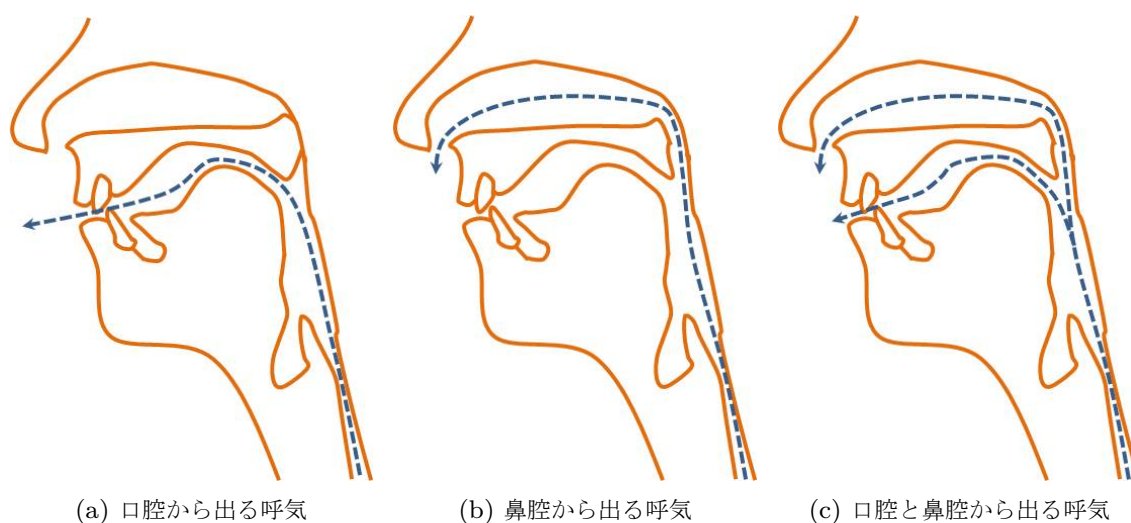
発声時、声帯は有声または無声という二つの状態のどちらかになる。有声は声帯を振動させ声を伴う音のことである。無声は反対に声帯を振動させず肺から吐き出された呼気そのまま通り抜けて出る音のことである。

声道は声帯から上の呼気の通路のことをいい、舌や唇、口蓋を動かしたり、鼻腔を使い声道の形を巧みに変化させることで呼気の調音を行う。調音は呼気が口や鼻から放射される過程で行われるが、呼気を通る経路は図 5.2 に示す 3 種類が存在する。

図 5.2a は口音といい、呼気を口腔のみから出す方法である。図 5.2b は鼻音といい、呼



図 5.1: 調音器官の名称と場所



(a) 口腔から出る呼気

(b) 鼻腔から出る呼気

(c) 口腔と鼻腔から出る呼気

図 5.2: 呼気の通り道

気を鼻腔のみから出す方法である。図 5.2c は鼻母音といい、呼気を鼻と口の両方から出す方法である。

5.2.2 母音の調音素性

調音素性とは「調音位置」「有声／無声」「調音方法」といった音声の基本的な分析単位である⁽³⁷⁾。このうち「調音位置」は音声を発声（調音）する際、舌の一部が口蓋のどこに

最も近づくかを「唇」「歯茎」「硬口蓋」「軟口蓋」「声門」等であらわしたものである。「有声／無声」は声帯振動の有無を表し、「調音方法」は「閉鎖」「摩擦」「はじき」「わたり」といった発声器官内の空気の流れや音声の共鳴の仕方を変化させる方法を表すものである。

母音は、声帯の振動を伴った呼気が妨害を受けずに共振して起こる有声音である。主に下記の三つの動きによって声道の容積と形を変えることで音質が変化する⁽⁴¹⁾⁽⁴²⁾。

唇の丸め

唇を丸める（突き出す）ことで声道の容積を変え、音質を変える。唇の丸めがあるものを円唇母音、ないものを非円唇母音という。

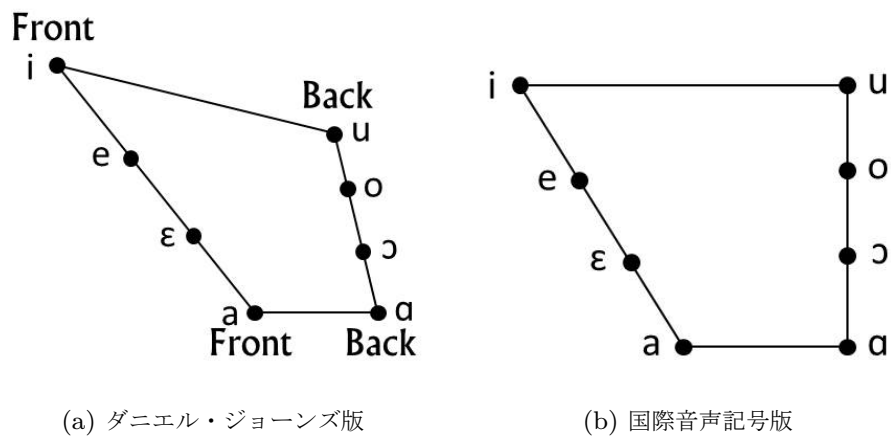
舌の前後位置

舌の最高点が硬口蓋付近の母音を前舌母音、軟口蓋付近の母音を後舌母音、その中間を中舌母音という。

舌の高さ

舌の高さは口の開き具合（開口度）に対応する。国際音声記号では高さは4段階あり、狭母音、半狭母音、半広母音、広母音と呼ばれる。日本語母音の場合、高さは3段階で十分なため狭母音を「高」、半狭母音と半広母音を「中」、広母音を「低」とすることが多い。

言語によらず、人間が出せる母音の最も外側に位置する18の母音を定めたものを基本母音という。基本母音はイギリスの音声学学者ダニエル・ジョーンズによって定められたもので、言語によらず母音を聞いたときにその音が母音図のどこの音か判断する基準として使う。18の基本母音のうち、最初の8つを第一次基本母音と呼び、図5.3aで表される。残りの10の基本母音は第二次基本母音と呼ぶ。現在ではこの基本母音図は簡略化され、国際音声記号 (International Phonetic Alphabet, IPA) として図5.3bのような形に定められている。この台形は下辺:左辺:上辺の比が2:3:4になるのが正式な形である。

図 5.3: 第一次基本母音図⁽⁴¹⁾

日本語の母音は「あ、い、う、え、お」の5種類であり、舌の位置と唇の丸め（開口度）の二つの情報で表すことができる．表 5.1 に母音の素性行列を示す．

	a	i	u	e	o
舌の位置	中	前	後	前	後
開口度	大	低	低	中	中

表 5.1: 母音の素性行列

5.2.3 子音の調音素性

子音は口の中で空気の流れを妨害することによって発する音であり，数多くの種類が存在するため図 5.4 のように体系的に分類される．

子音はまず「阻害音」と「共鳴音」に分けられる．阻害音は調音器官で閉鎖や狭めにより気流を妨げ調音された音でいわゆる子音らしい子音となる．共鳴音は声道による気流の妨げが少ないため母音に近い特性を持つ．

子音の調音素性は母音同様，「調音位置」「有声／無声」「調音方法」の他，拗音，鼻音，口蓋化といった素性がある

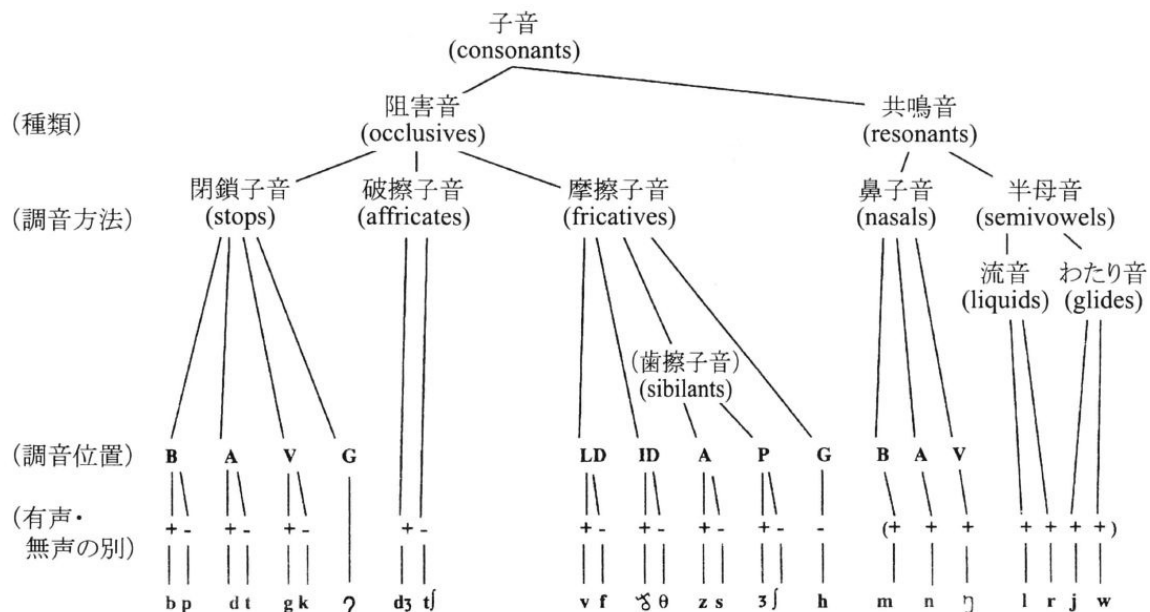


図 5.4: 子音の分類

5.2.4 阻害音

阻害音は調音方法の違いによって「閉鎖子音」, 「破擦子音」, 「摩擦子音」に分類される。

閉鎖子音

呼気の流れをいったん完全に止めて勢いよく吐き出すことで発声するため破裂子音とも呼ばれる。/p/, /b/は発声時に両唇が閉じられるため両唇閉鎖音という。/t/, /d/は舌と歯茎で閉鎖が作られるため歯茎閉鎖音という。/k/, /g/は奥舌と軟口蓋で閉鎖が作られるので軟口蓋閉鎖音という。閉鎖子音はさらに調音位置および有声／無声によって細かく分類される。

破擦子音

破擦子音は閉鎖子音と同様呼気の流れをいったん完全に止めた後、開放するときには摩擦音へと変化する。/tʃ/と/tʃ/は舌先を歯茎の裏につけて呼気を止め、摩擦しながら吐き出すことから歯茎硬口蓋破擦音という。

摩擦子音

呼気の流れを止めるのではなく、声道を狭め呼気の流れを妨げることで発声する摩擦性の雑音を摩擦子音という。/s/, /z/は舌先を歯茎に近づけ、こすって息を出すことから歯茎摩擦音と言う。/ʃ/, /ʒ/は舌先を広く上顎（硬口蓋）に近づけて摩擦を作る口蓋摩擦音になる。舌や歯茎を使わない摩擦子音に唇歯摩擦音（/f/, /v/）

や歯間摩擦音 ($/\theta/$, $/\delta/$), 声門摩擦音 ($/h/$) がある.

5.2.5 共鳴音

共鳴音は調音方法の違いによって「鼻子音」と「半母音」に分類される.

鼻子音

鼻子音は呼気が鼻腔を通して作られる音で有声音である. 鼻子音は口腔内のもっとも強く閉鎖されている位置の違いで3種類に分けられる. 両唇を結んで鼻から息を出す音を両唇鼻音 ($/m/$) という. また, $/n/$ のように歯茎に舌先をあてて閉鎖するのは歯茎鼻音, $/ŋ/$ は舌が最も奥の軟口蓋に接触するので軟口蓋鼻音という.

半母音

半母音はさらに流音とわたり音に分類される. 流音は $/r/$, $/l/$ で舌先を歯茎に近づけたり接触させて発生する. $/r/$ を歯茎半母音, $/l/$ を歯茎単顫動音という. わたり音は $/w/$, $/j/$ のように狭めが緩いため響きが良く母音に近い. 前の音または次の音へのつなぎの音として使用されるため, わたり音という. $/w/$ は両唇わたり音といい唇の丸めが強い. $/j/$ は舌の位置を硬口蓋近くまで持ち上げるので硬口蓋わたり音という.

5.3 聴取実験による合成音声の評価

まず母音および子音の聴取実験を行った。

使用する音声データはモーラから子音毎に1つずつ選定した単音節32個を用いた。また入力音声は機能障害者5の気道音声，上唇の体内伝導音からヒューリスティック手法で生成した合成音声，そして健常者1の気道音声の3種類とした。

被験者は日本語を母国語とする正常な聴力を有する22歳～29歳の日本人成人9名（男性5名，女性4名）である。実験は音声刺激をヘッドフォンで両耳提示し，被験者は聴取した音と同じまたはもっとも近い単音節をPC上に表示されたモーラ表（図5.5）から選択した。また，本実験では調音素性ごとに検討するため，実験を母音のみ提示するもの（図5.5a）と拗音/j/を含む子音のみ提示するもの（図5.5b）の2種類に分けた。

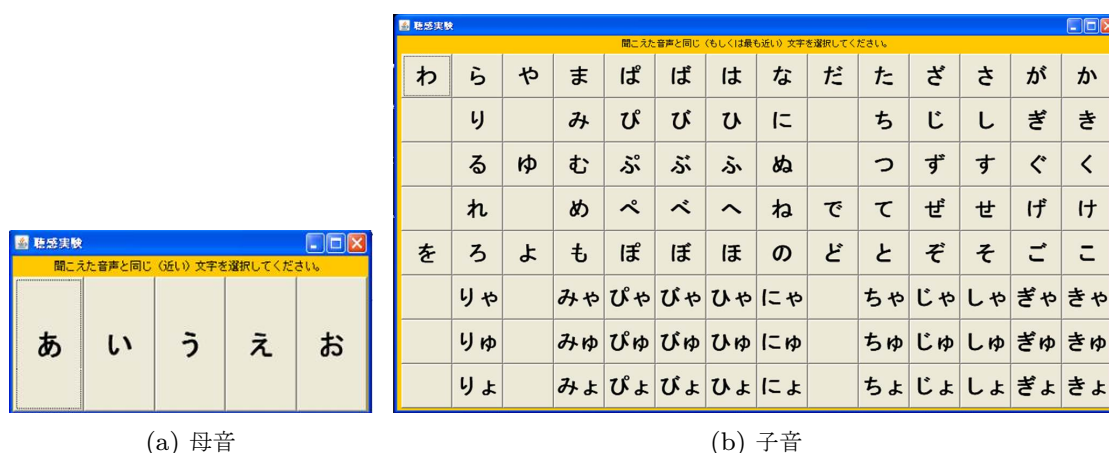


図 5.5: 聴取実験に用いた回答画面

表 5.2 に，健常者，機能障害者5気道音声および合成音声を提示した結果を示す。

各表の行は呈示された母音を表しており，各列は回答された母音を表している。

3種類の音声で比較すると，健常者1気道音声と合成音声では正答率は100%になっている。それに対して，機能障害者5気道音声では正答率はわずか44.4%であり，提示された母音を半数以上誤認識した結果となっている。障害者音声として使用した音源は構音障害者の気道音声であり，調音できないという音声の特徴が原因となって混同（聞き間違い）が多発したと考えられる。それに対し，合成音声では一切の混同が生じていない。合成音声は母音を生成する調音素性を完全に再現できているためと考えられる。機能障害者5の気道音声に注目すると，母音/i/と/u/の間で多数の混同が生じていることが分かる。そこで，障害者原音声において母音の混同が生じた原因を特定するため，混同が起こる音素

はそれを構成する調音素性が正しく伝達されていないと仮定して Sequential information analysis (SINFA)⁽³⁸⁾⁽³⁹⁾⁽⁴⁰⁾ を用いて調音素性分析を行った.

	a	i	u	e	o
a	100	0	0	0	0
i	0	92	0	0	0
u	0	0	101	0	0
e	0	0	0	89	0
o	0	0	0	0	95

正答率：100%

(a) 健常者1気道音声

	a	i	u	e	o
a	103	0	0	0	3
i	6	11	73	2	4
u	3	12	36	29	3
e	15	3	18	59	1
o	45	15	32	1	3

正答率：44.4%

(b) 機能障害者5気道音声

	a	i	u	e	o
a	102	0	0	0	0
i	0	90	0	0	0
u	0	0	102	0	0
e	0	0	0	80	0
o	0	0	0	0	85

正答率：100%

(c) 合成音声（ヒューリスティック）

表 5.2: 母音提示聴取実験結果

5.4 SINFA による母音の調音素性伝達率解析

SINFA は聴取実験の結果である混同行列と、解析したい調音素性を定義した素性行列の2つの行列を用いることで、定義した調音素性の伝達率を解析することができる。本実験では、混同行列として表 5.2b, 表 5.2c を使用し、素性行列として表 5.3 を使用した。

	a	i	u	e	o
舌の位置	中	前	後	前	後
開口度	大	低	低	中	中

表 5.3: SINFA で解析する母音の素性行列

SINFA による解析結果を表 5.4 に示す.

調音素性分析では, 定義した調音素性ごとの伝達量および伝達率が得られる. これら素性ごとの伝達量を合計した情報伝達量と, その他の要素を含む全伝達量の合計である総情報伝達量から, 定義した調音素性がどの程度伝達されているかが関与率として出力される.

素性	機能障害者 5 気道音声		合成音声 (ヒューリスティック)	
	伝達量	伝達率	伝達量	伝達率
舌の位置	0.045	7.4%	1.541	100.0%
開口度	0.477	31.0%	0.774	100.0%
情報伝達量	0.522		2.315	
総情報伝達量	0.723		2.316	
関与率	72.2%		100.0%	

表 5.4: 母音の SINFA 解析結果

この素性伝達の指標である関与率をみると, 表 5.4 より機能障害者 5 気道音声では 72.2% であるが, 合成音声では 100% である. この結果からも本システムの合成音声は母音の素性情報が正しく伝達されていることが確認できた. また, 機能障害者 5 気道音声では「舌の前後」情報はわずか 7.4%, 「開口度」情報も 31% しか伝達されていない. 対して合成音声では, 両情報の伝達率は 100% となっており, 本システムの利用で「舌の前後」, 「開口度」の情報を完全に再現できることが確認できた. 特に機能障害者 5 は, 事故により舌と顎の一部を損失しており, 通常の発声では舌を使う調音は全くできない. また, 開口度も制限されているため, それらの素性情報も伝達されにくいことが予想されたが, SINFA による解析で予想が正しいことが確認できた.

5.5 子音の調音素性伝達率解析

表 5.5 に, それぞれ健常者気道音声, 障害者気道音声, 合成音声による子音提示実験の結果を示す. 各表の行は提示された子音を表しており, 各列は回答された子音を表している. 表中最終列 "mis" は被験者のミスにより選択肢以外を回答した回数を示している. これらの表より健常者においては正答率 93.2% と混同は生じていないが, 障害者気道音声では 5.9%, 合成音声では 7.4% と多数の混同が生じていることがわかる.

母音と同様に表 5.5b および表 5.5c の混同行列を SINFA で解析を行った. 解析に使用した素性行列⁽⁴¹⁾⁽⁴³⁾を表 5.6 に示す. 表 5.7 に SINFA による解析結果を示す. 比較の

ために健常者の結果も載せている。

素性	健常者気道音声		機能障害者 5 気道音声		合成音声	
	伝達量	伝達率	伝達量	伝達率	伝達量	伝達率
口蓋化	0.67	95.70%	0.012	1.40%	-	-
鼻音	0.446	100.00%	-	-	-	-
破裂	0.142	100.00%	0.068	9.90%	0.016	1.60%
有声	0.641	92.40%	0.434	43.60%	0.191	19.20%
調音位置	1.163	90.40%	0.044	3.10%	0.092	7.40%
調音方法	0.476	89.00%	0.011	3.10%	0.037	3.00%
拗音	0.799	98.30%	-	-	0.022	5.30%
情報伝達量	4.336		0.568		0.358	
総情報伝達量	4.420		1.002		0.882	
関与率	98.10%		56.70%		40.60%	

表 5.7: 子音の SINFA 解析結果

表 5.7 より，調音素性の伝達率はおおむね障害者気道音声の方が合成音声よりも高く，関与率においても障害者気道音声の方が高い．また，解析結果の中で一部結果が出力されていない項目があるが，これは解析の途中で，混同パターンに全く関係ないと判断され削除されたためである．

以上より子音については，本システムで生成した合成音声では原気道音声と比較してその認識率を高めることはできなかった．しかし，聴取実験の正答率では合成音声の方が高いという結果であった．

SINFA による調音素性解析は，定義した素性のみを解析するもので，当然ながら素性行列で定義していない素性の伝達率は確認できない．障害者原音声では，情報の総伝達量 1.002 に対して，定義した素性の情報伝達量は 0.568 であり，56.7% の情報しか特定できていない．同様に合成音声では，情報の総伝達量 0.882 に対して定義した素性の情報伝達量は 0.358 であり，わずか 40.6% である．

これらから今回定義した素性行列では障害者原音声と合成音声の混同を捉えることができなかったことが確認できる．引き続き解析に使う素性行列について検討する必要がある．

	b	bj	ch	d	f	g	gj	h	hi	ji	k	kj	m	mj	n	nj	p	pj	r	rj	s	sh	t	ts	w	y	z	mis
b	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bj	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ch	0	0	97	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
f	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0	0	0	0	0	0
g	0	0	0	0	0	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
gj	0	0	0	0	0	0	41	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0
h	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0
hi	0	0	0	0	0	0	0	0	103	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ji	0	0	0	0	0	0	0	0	0	114	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
kj	0	0	1	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m	0	0	0	0	0	0	0	0	0	0	0	0	61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mj	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	7	0	0	0	0	0	0	0	0	0	0	0	0
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	58	1	0	0	0	0	0	0	0	0	0	0	0	0
nj	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0
p	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	0	0	0	0	0	0	0	0	0	0	0
pj	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0
r	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	58	0	0	0	0	0	0	0	0	0
rj	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	0	0
s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0
sh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	111	0	0	0	0	0	0
t	0	0	0	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0
ts	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	43	0	0	0	0
w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	43	0	0	0
y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	8	37	0	0
z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	46	0
mis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

正答率：93.2%

(a) 健常者1気道音声

	b	bj	ch	d	f	g	gj	h	hi	ji	k	kj	m	mj	n	nj	p	pj	r	rj	s	sh	t	ts	w	y	z	mis
b	1	0	0	1	0	4	0	0	0	0	0	0	8	2	22	2	0	0	4	0	0	0	0	0	1	1	0	0
bj	1	1	0	0	0	0	0	0	0	0	0	0	13	2	24	3	0	2	1	0	0	0	0	0	1	2	0	0
ch	8	5	0	1	8	2	1	5	0	0	1	0	6	1	10	0	47	1	5	0	1	0	14	0	1	0	0	0
d	4	0	0	1	0	3	1	1	0	2	1	0	18	3	33	7	1	0	3	0	0	0	0	0	4	0	0	0
f	3	0	0	1	12	3	0	2	0	0	1	0	7	0	1	0	17	1	5	0	0	0	1	1	0	2	0	0
g	7	0	0	3	0	2	0	7	0	0	0	0	9	0	19	0	5	2	3	0	0	0	2	0	6	1	0	0
gj	1	2	0	1	0	0	0	1	0	0	0	0	11	8	26	3	3	3	3	2	0	0	0	0	1	1	0	0
h	1	0	0	0	0	1	0	12	0	0	0	0	1	0	0	0	30	0	1	0	1	0	2	0	1	0	0	0
hi	1	0	0	0	31	1	0	47	3	0	1	0	3	0	4	0	9	0	0	0	0	0	6	0	1	3	0	0
ji	4	9	0	0	3	4	0	0	1	0	3	0	22	2	42	4	2	1	4	1	0	0	0	0	2	6	0	0
k	1	0	0	0	0	0	0	14	0	0	3	0	0	0	1	0	32	0	2	0	1	0	1	0	1	0	0	0
kj	1	0	0	0	0	1	0	6	0	0	3	0	1	0	0	0	43	0	1	0	0	0	2	0	0	0	0	0
m	2	1	0	0	0	2	1	1	0	0	0	0	10	3	16	7	0	0	6	0	0	0	0	0	2	1	0	0
mj	0	0	0	1	0	2	0	0	3	0	0	0	5	11	19	8	0	0	8	3	0	0	0	0	1	7	0	0
n	1	2	0	2	0	1	0	0	0	0	0	0	17	1	21	1	0	1	7	0	0	0	0	0	6	0	0	0
nj	6	2	0	1	0	7	0	0	0	0	2	0	24	10	44	2	3	0	9	0	0	0	0	0	1	2	1	1
p	2	0	0	0	1	0	0	30	0	0	4	0	5	0	1	0	6	1	1	0	0	0	0	0	5	0	0	0
pj	2	0	0	2	0	0	0	8	0	0	1	0	0	0	1	0	19	2	0	0	1	0	12	0	0	0	1	0
r	4	6	0	1	0	3	0	1	0	0	0	0	2	1	16	1	0	1	11	1	0	0	1	0	1	10	0	0
rj	0	4	0	3	0	0	0	0	0	0	0	0	2	4	22	2	1	1	6	0	0	0	0	0	0	1	0	0
s	2	0	0	0	0	0	0	5	0	0	0	0	1	0	0	0	37	0	0	0	0	0	13	0	2	0	0	0
sh	4	0	0	2	10	0	0	4	1	0	5	0	6	0	4	2	38	7	3	0	0	0	14	1	0	5	0	0
t	4	0	0	0	0	1	0	31	0	0	6	0	0	0	2	0	9	0	1	0	2	0	4	0	2	0	0	0
ts	0	0	0	0	20	0	0	1	0	0	0	0	7	0	16	2	4	0	0	0	0	0	0	0	0	2	0	0
w	9	5	0	0	0	1	1	2	0	0	0	0	3	4	9	1	2	3	2	2	0	0	1	0	4	5	0	1
y	0	6	0	0	0	1	0	0	0	0	0	0	12	6	19	2	0	0	2	1	0	0	0	0	2	0	0	0
z	1	1	0	0	0	1	0	1	0	0	0	0	3	6	25	2	1	1	4	1	0	0	0	0	1	1	0	0
mis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

正答率：5.9%

(b) 機能障害者5気道音声

	b	bj	ch	d	f	g	gj	h	hi	ji	k	kj	m	mj	n	nj	p	pj	r	rj	s	sh	t	ts	w	y	z	mis
b	8	0	0	1	0	5	0	6	0	0	4	0	5	0	13	0	1	0	2	0	0	0	1	0	14	1	0	0
bj	5	7	0	0	0	0	0	1	0	0	0	0	4	6	10	5	2	0	3	2	0	0	0	0	4	5	0	1
ch	7	0	1	3	0	18	0	0	3	5	7	0	0	0	1	6	27	0	15	0	1	0	14	0	1	0	1	0
d	5	0	0	2	0	1	0	6	1	0	1	0	3	0	5	2	1	0	16	1	0	0	0	0	4	7	1	0
f	16	0	0	0	4	3	0	0	0	0	5	0	6	0	6	0	6	0	2	0	0	0	0	0	0	1	0	0
g	14	0	0	1	0	0	0	7	0	0	4	0	3	0	2	0	4	0	1	0	0	0	6	0	10	1	1	0
gj	16	3	0	1	1	3	0	0	0	0	0	0	6	2	4	0	9	1	13	3	0	0	1	0	2	3	0	0
h	0	0	0	1	0	0	0	13	0	0	4	0	1	0	0	0	20	0	0	0	0	0	5	0	5	1	0	0
hi	5	0	1	0	0	5	0	27	7	0	25	0	3	0	0	4	12	0	11	0	1	3	0	0	3	1	0	0
ji	16	7	4	1	0	6	0	2	5	4	9	0	6	0	9	10	7	0	17	6	0	4	2	0	1	4	2	1
k	1	0	0	0	0	0	0	2	0	0	3	0	1	0	0	0	43	0	0	0	3	0	7	0	2	0	0	0
kj	2	0	0	0	0	2	0	3	0	0	26	0	1	0	1	0	16	0	1	0	2	0	2	0	8	0	0	0
m	8	3	0	2	0	1	0	6	0	0	3	0	9	2	9	1	3	1	4	0	2	0	6	0	9	3	1	0
mj	5	2	0	3	0	0	2	4	0	0	2	0	9	2	8	2	1	0	10	8	0	0	1	0	1	8	0	0
n	4	2	0	1	0	2	0	2	0	0	0	0	4	5	12	1	1	0	6	2	0	0	1	0	7	2	0	0
nj	13	3	1	1	0	17	0	2	3	1	2	0	16	2	17	12	2	1	10	0	0	0	1	0	1	4	2	0
p	1	0	0	0	0	0	0	18	0	0	8	0	0	0	2	0	14	0	0	0	1	0	9	0	5	1	1	0
pj	0	1	0	1	0	0	0	0	0	0	2	1	1	0	0	0	35	1	1	0	1	0	19	0	1	1	0	0
r	11	1	0	2	0	5	0	0	0	0	0	0	2	1	3	2	2	1	6	1	0	0	4	0	2	4	0	0
rj	12	2	0	1	0	1	1	3	0	0	2	0	7	0	5	0	1	1	12	1	0	0	2	0	0	2	1	1
s	1	0	0	0	0	0	0	6	0	0	25	0	1	0	0	0	16	0	0	0	0	0	2	0	5	1	0	0
sh	1	0	4	0	0	16	0	6	7	6	25	0	1	0	0	4	19	0	9	0	0	3	8	0	2	1	0	0
t	1	0	0	0	0	0	0	6	0	0	12	0	1	1	1	0	7	0	2	0	5	0	2	0	10	0	0	0
ts	12	0	0	0	2	1	0	0	0	0	8	0	7	0	5	2	3	0	7	0	0	0	0	8	0	0	2	0
w	7	1	1	0	0	3	0	4	0	0	4	0	2	0	4	0	4	0	3	0	1	0	2	0	9	1	0	1
y	17	5	0	1	0	1	0	2	0	0	0	0	5	2	11	3	3	1	5	2	0	0	0	0	0	5	0	0
z	12	1	0	0	0	2	0	0	0	0	0	0	2	0	2	0	2	0	8	0	0	0	0	1	4	0	0	0
mis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

正答率：7.4%

(c) 合成音声（ヒューリスティック）

表 5.5: 子音提示聴取実験結果

	b	bj	ch	d	f	g	gj	h	hi	ji	k	kj	m	mj
有声	+	+	-	+	-	+	+	-	-	+	-	-	+	+
調音方法	閉鎖	閉鎖	摩擦	閉鎖	摩擦	閉鎖	閉鎖	摩擦	摩擦	摩擦	閉鎖	閉鎖	閉鎖	閉鎖
調音位置	唇	唇	歯茎	歯茎	唇	口蓋	口蓋	声門	口蓋	歯茎	口蓋	口蓋	唇	唇
口蓋化	-	+	+	-	-	-	+	-	-	-	-	+	-	+
拗音	-	+	-	-	-	-	+	-	-	-	-	+	-	+
鼻音	-	-	-	-	-	-	-	-	-	-	-	-	+	+
破裂	+	+	-	+	-	+	+	-	-	+	+	+	-	-
	n	nj	p	pj	r	rj	s	sh	t	ts	w	y	z	
有声	+	+	-	-	+	+	-	-	-	-	+	+	+	
調音方法	閉鎖	閉鎖	閉鎖	閉鎖	はじき	はじき	摩擦	摩擦	閉鎖	摩擦	わたり	わたり	摩擦	
調音位置	歯茎	歯茎	唇	唇	歯茎	歯茎	歯茎	口蓋	歯茎	歯茎	唇	口蓋	歯茎	
口蓋化	-	-	-	+	-	+	-	+	-	-	-	-	-	
拗音	-	+	-	+	-	+	-	-	-	-	-	-	-	
鼻音	+	-	-	-	-	-	-	-	-	-	-	-	-	
破裂	-	-	-	+	+	-	-	-	-	+	+	-	-	

表 5.6: SINFA で解析する子音の素性行列

第 6 章

結言

本研究では，発声機能障害者を支援するための発声機能障害者支援システムで使用する音質明瞭化手法の開発と評価を行った。

まず線形予測分析を用いた音声合成手法を提案した．提案手法を用いて障害者の体内伝導音から合成音声を作成すると，区間推定さえ正確にできれば健常者音声と同等の音声認識率を得られた．特に，体内伝導音で大きく減衰してしまう 2KHz 以上の高周波成分が合成音声で回復できていることも確認できた．さらに連続サブワード認識による自動区間推定手法を用いた合成音声でも 60% 以上の音声認識率を実現し，区間推定自動化の可能性も確認した．

次に提案手法で生成した合成音声を聴取実験および調音素性分析により主観的に評価した．母音の聴取実験の結果，合成音声では一切の混同なく正確に聴取できた．さらに調音素性分析の結果，母音においては構音障害者気道音声に本来存在しない調音素性を合成音声で再現できた．

また，本システムの評価用に声帯ポリープ，声帯結束，片側声帯麻痺，声帯摘出，構音障害といった貴重な音声および体内伝導音の音声データベースを作成することができた．

線形予測分析による音声の合成と有効性評価

上唇，声帯付近，乳様突起の各体内伝導音から音質明瞭化の対象をどれにするか音声認識を用いて検討を行った．その結果，最も音声に音響的特徴に近い⁽³⁴⁾ 上唇の体内伝導音を本研究の音質明瞭化対象とすることとした．

次に音質明瞭化手法として線形予測（LPC）分析を用いた音声合成手法の有効性を検討した．

ヒューリスティックな手法で音節区間を推定した場合，つまり，区間推定が正しくできた場合は障害者体内伝導音から生成した合成音声の音声認識率が健常者気道音声とほぼ同

等になることが確認できた。また、この合成音声は構音障害者において本来欠落している2KHz以上の高域部分を回復したことも確認できた。

連続サブワード認識 (CSR) による自動区間推定で作成した合成音声も認識率が60%以上まで改善され、区間推定自動化の可能性も確認できた。

聴取実験および調音組成分析による合成音声の評価

次に、LPC分析を用いて作成した合成音声の評価するために聴取実験および調音素性分析の二つの主観評価を行った。

母音については障害者気道音声で多数生じていた混同が、聴取実験の結果、合成音声では無くなっていた。この結果を受けて調音素性分析を行った結果、障害者気道音声では母音の調音素性である「舌の位置」「開口度」どちらの伝達率も大きく低下していた。障害者気道音声では音声が不明瞭であったり、構音に問題があって母音の素性がうまく伝わらないため、混同が多発した。合成音声の聴取実験、調音素性分析の結果から本研究の提案手法は全ての調音素性を完全再現できることが確認できた。

子音については合成音声障害者気道音声よりも調音素性の伝達率は低いものの、聴取実験の正答数は多いという結果になった。調音素性分析の結果では今回使用した子音の素性要素のいくつかは分析に無関係と判断された。つまり今回定義した素性行列では障害者気道音声と合成音声の混同をとらえきれなかったということである。引き続き、分析に使用する適切な素性行列について検討していく。

今後の予定

検討の結果、発声機能障害者支援システムで作成した合成音声は本来、原音声に含まれていない調音情報を再現でき、明瞭性の向上した合成音になっていることを確認できた。引き続き、子音の明瞭度改善を検討し、これらの明瞭化手法を発声機能障害者支援システムへ導入していく予定である。具体的には音質明瞭化手法を確立し、スマートフォンなどのモバイル端末に実装していく。最終的にはモバイル端末で気軽に障害者体内伝導音から障害前の自然な肉声をリアルタイムに再現できるシステムへと発展させ、聴覚障害における補聴器のように、発声機能障害における標準の支援ツールとして実現を目指していく。

また、今回作成した貴重な音声データベースを活用しさらに研究を進めることで発声障害のための支援技術開発を加速させていきたい。

本研究で提案した音声合成手法は障害者支援だけでなく他の分野への展開も計画中である。例えば、ベビーシッターが母親の声を再現することで赤ん坊に安心感を与えるといった育児や介護など福祉の分野にも展開が可能である。また、未来を描いた映画に出てくる

ような人間と機械のマンマシンインタフェースなどロボティクスや人間工学分野への応用も検討する。

本研究の成果を用いて機械に発声機能を与え，機械をより知的で身近に感じられるような展開を検討していく。

謝辞

本研究に関して終始ご指導ご鞭撻を頂きました石光俊介教授に心より感謝致します。

小林康秀教授，藤坂尚登教授には，御多忙中であるにも関わらず本論文の副査をご快諾頂きました事を心より感謝致します。また，様々な事に関してご助言頂きました中山仁史助教授，高橋雄三助教授に感謝致します。

学外の研究活動において，様々なご助言をいただきました日本大学の堀畑聡教授に感謝致します。また，音声収録にご協力頂いた障害者の皆様には，障害のため話すことも困難な中，快くご協力頂き心より感謝致します。

同研究室の名越隼人氏，山中貴弘氏，齋藤愛さんには研究について多くのご助言及びご協力を頂き感謝致します。最後に，一緒に研究に励んできた研究室の皆様に感謝致します。

参考文献

- [1] 内閣府, "障害者白書平成 23 年版",
(<http://www8.cao.go.jp/shougai/whitepaper/h23hakusho/zenbun/index.html>),
(参照日 2015 年 10 月 21 日).
- [2] 厚生労働省, "平成 23 年生活のしづらさなどに関する調査",
(http://www.mhlw.go.jp/toukei/list/seikatsu_chousa_c.html), (参照日 2015 年
10 月 21 日).
- [3] 厚生労働省, "平成 18 年身体障害児・者実態調査結果",
(<http://www.mhlw.go.jp/toukei/saikin/hw/shintai/06/>), (参照日 2015 年
12 月 20 日).
- [4] WHO, "GLOBOCAN 2012 : Estimated Cancer Incidence, Mortality and
Prevalence Worldwide in 2012", (<http://globocan.iarc.fr/Default.aspx>), (参照日
2015 年 12 月 20 日).
- [5] NATIONAL CANCER INSTITUTE, "SEER Stat Fact Sheets: Esophageal Can-
cer", (<http://seer.cancer.gov/statfacts/html/esoph.html>), (参照日 2015 年 12 月
20 日).
- [6] CANCER RESEARCH UK, "Oesophageal cancer incidence statistics",
([http://www.cancerresearchuk.org/content/oesophageal-cancer-incidence-
statistics](http://www.cancerresearchuk.org/content/oesophageal-cancer-incidence-statistics)), (参照日 2015 年 12 月 20 日).
- [7] goo ヘルスケア, "構音障害・発声障害の症状や原因・診断と治療方法",
(<http://health.goo.ne.jp/medical/10F10300>), (参照日 2015 年 10 月 22 日).
- [8] 国立がんセンターがん対策情報センター, "がん情報サービス 集計表のダウンロー
ド", (<http://ganjoho.jp/professional/statistics/statistics.html>), (参照日 2015 年 4
月 19 日).
- [9] 国立がんセンターがん対策情報センター, "がん情報サービス 下咽頭がん",
(<http://ganjoho.jp/public/cancer/hypopharynx/>), (参照日 2015 年 12 月 26 日).
- [10] 株式会社 電制, "電気式人工喉頭 ユアトーン II・UB",

- (<http://www.dencom.co.jp/product/yourtone/yt2ub.html>), (参照日 2015 年 10 月 21 日).
- [11] Hayusa.FK, "おしゃべりサポーター", (<https://play.google.com/store/apps/details?id=com.fuku.voiceagent>), (参照日 2015 年 10 月 21 日).
- [12] Cotyledon Technology Co., Ltd., "自分の声再現ソリューション VSC", (<http://www.cotyledon-technology.co.jp/package-business/vsc/>), (参照日 2015 年 12 月 26 日).
- [13] 株式会社 ウォンツ, "自分の声ソフトウェア ボイスター", (<http://www.wantsinc.jp/voistar/>), (参照日 2015 年 12 月 26 日).
- [14] A.B.Kain, J.P.Hosom, X.Niu, J.P.H.van Santen, M.Fried-Oken and J.Staehely, "Improving the intelligibility of dysarthric speech", Speech comun. Vol.49, pp.743-759 (2007).
- [15] 森口, 戸田, 佐野, 佐藤, ニュービッグ, サクティ, 中村, "無喉頭音声から通常音声へのリアルタイム声質変換処理の DSP への実装", 音講論, 1-2-2, 217-218 (2012).
- [16] 中山仁史, 石光俊介, 中川誠司, "加速度差分を用いた体内伝導音の明瞭化", 電子情報通信学会 信学技報 Vol.108 No.69, pp.39-44, (2008).
- [17] 小田康平, 石光俊介, 中山仁史, "発声機能障害者支援システムの構築", 第 10 回 IEEE 広島支部学生シンポジウム, CDROM pp74-77 (2008).
- [18] H.Nagoshi, S.Ishimitsu, K.Oda, K.Makiyama and S.Horihata, "Development for speech support system using body-conducted speech", Forum Acousticum 2011, CDROM pp.95-99 (2011).
- [19] F.Itakura and S.Saito, "An analysis-synthesis telephony based on the maximum likelihood method", Proceeding of the 6th International Commission for Acoustics, Vol.3, C-5-5 (1968).
- [20] B.S.Atal and M.R.Schroeder, "Predictive coding of Speech Signals", Proceeding of the 6th International Commission for Acoustics, Vol.3, C-5-4 (1968).
- [21] H.Nagoshi, S.Ishimitsu, K.Oda, K.Makiyama and S.Horihata, "Improving the Clarity of Body-Conducted Speech Using Linear Predictive Coefficients", ICIC EXPRESS LETTERS, PartB:Applications Vol.2, No.3, pp.609-614 (2011).
- [22] 石光俊介, 北風裕教, 土伏悌之, 柳川博文, 福島学, 体内伝導音認識システム構築のための基礎的検討, 社団法人電子情報通信学会 信学技報, EA2002-34, pp.25-28 (2002).
- [23] 鹿野清宏, 河原達也, 山本幹雄, 伊藤 克亘, 武田一哉, 情報処理学会, "IT Text 音声認識システム", オーム社, pp.1-16 (2001).

- [24] 嵯峨山, "応用音響学 C 線形予測分析 全極型音声モデル化", (http://ocw.u-tokyo.ac.jp/lecture_files/engin_01/3/notes/notes_ja.zip), (参照日 2015 年 10 月 22 日).
- [25] S.Ishimitsu, H.Kitakaze, Y.Tsuchibushi, H.Yanagawa and M.Fukushima, "A noise-robust speech recognition system making use of body-conducted signals", *Acoust. Sci. & Tech.* 25, 2, pp.166-169 (2004).
- [26] 荒木雅弘, "フリーソフトで作る音声認識システム", 森北出版, pp.103-197 (2007).
- [27] 李晃伸, "汎用大語彙連続音声認識エンジン Julius / Julian rev.3.3", (<http://julius.sourceforge.jp/book/Julius-book-3.3-ja.pdf>), (参照日 2015 年 4 月 19 日) (2002).
- [28] 青木直史, "ディジタル・サウンド処理入門", CQ 出版, pp.87-126 (2006).
- [29] 小田康平, 石光俊介, 中山仁史, 牧山清, 堀畑聡, "発声機能障害者のための音声認識システムの評価", 日本機械学会中国四国支部 第 48 期総会・講演会講演論文集, pp.359-360 (2010).
- [30] S.Itahashi, "A noise database and Japanese common speech data corpus", *J. Acoust. Soc. Jpn*, Vol.47, No.12, pp.951-953 (1991).
- [31] 阿部匡伸, 匂坂芳典, 梅田哲夫, 桑原尚夫, "研究用日本語音声データベース利用解説書 (連続音声データ編)", ATR 自動翻訳電話研究所, TR-I-0166 (1990).
- [32] Y.Saito, H.Yanagawa, S.Ishimitsu, K.Kamura and M.Fukushima, "Improvement of the speech sound quality of the vibration pick up microphone for speech recognition under noisy environment", *Proceeding of the Autumn Meeting of Acoust. Soc. Jpn.*, Vol.1, pp.691-692 (2001).
- [33] 伊福部達, "音の福祉工学", pp.165-190 (1997).
- [34] Y.Saito, H.Yanagawa, S.Ishimitsu, K.Kamura and M.Fukushima, "Improvement of the speech sound quality of the vibration pick up microphone for speech recognition under noisy environment", *Proceeding of the Autumn Meeting of Acoust. Soc. Jpn.*, Vol.1, pp.691-692 (2001).
- [35] 向當一洋, 谷口秀次, 小泉 卓也, "サブワード単位離散単語認識システムの話者依存性の改善", 電子情報通信学会技術研究報告. SP, 音声 98(178), pp.15-21 (1998).
- [36] T.Koizumi, A.Fukuyama, M.Mori and S.Taniguchi, "Speech Recognition Based on Subword Units", *Proceedings of the 3rd Joint Meeting of Acoust. Soc. Am and Acoust. Soc. Jpn*, pp.1129-1134 (1996).
- [37] 板橋秀一, 赤羽誠, 石川泰, 大河内正明, 粕谷英樹, 桑原尚夫, 田中和世, 新田恒雄, 矢頭隆, 渡辺隆夫, "音声工学", 森北出版, pp.6-10 (2005).
- [38] 籠宮隆之, 中川誠司, "骨導超音波補聴器装用時の分節音異聴パターンに対する調音素

- 性を用いた分析", 日本音響学会秋季研究発表会講演論文集, pp559-560 (2011).
- [39] M.D.Wang, "SINFA: Multivariate uncertainty analysis for confusion matrices", Behavior Research Methods & Instrumentation, Vol.8, Issue.5, pp.471-472 (1976).
- [40] SINFA analyzing software, "FIX-Feature Information Xfer", UCL Division of Psychology & Language Sciences (<http://www.phon.ucl.ac.uk/resource/software.html>), (参照日 2015 年 4 月 19 日).
- [41] 猪塚元, 猪塚恵美子, "日本語の音声入門-解説と演習 (日本語教師トレーニングマニュアル)", バベルプレス, pp.2-15, pp.132-133 (2008).
- [42] 猪塚元, 猪塚恵美子, "日本語教育能力検定試験音声パーフェクト対策", アルク, pp.6-14 (2007).
- [43] ジャック・ライアルズ, "音声知覚の基礎", 海文堂, p.4 (2003).
- [44] A.B.Kain, J.P.Hosom, N.Xiaochuan, J.P.H.van Santen, M.Fried-Oken and J.Staehely, "Improving the intelligibility of dysarthric speech", Speech Communication, Vol.49, pp.743-759 (2007).
- [45] K.Nakamura, T.Toda, H.Saruwatari and K.Shikano, "Speaking aid system for total laryngectomees using voice conversion of body transmitted artificial speech", Proceedings of the Interspeech, pp.1395-1398 (2006).
- [46] 中山仁史, 石光俊介, 小田康平, "体内伝導音認識を用いた発声機能障害者支援システムの基礎検討", 日本音響学会 2007 年秋季研究発表会講演論文集, CDROM pp.715-716 (2007).
- [47] 末永康仁, "メディア情報処理 (新インターユニバーシティ)", オーム社, pp.22-23 (2008).

発表文献

査読付きジャーナル論文

- [48] 福井 和敏, 石光 俊介, 名越 隼人, 山中 貴弘, "体内伝導音のサブワード単位線形予測による音質明瞭化と調音素性分析を用いた評価", 日本感性工学会論文誌, 第 15 号 1 号 (2016 論文採択決定).
- [49] Takahiro Yamanaka, Shunsuke Ishimitsu, Kazutoshi Fukui, "Study of Improving Sound Quality in Support System for Speech Impaired", ICIC EXPRESS LETTERS, PartB:Applications Vol.5, No.2, pp.595-600 (2014).
- [50] Shunsuke Ishimitsu, Kenji Takami, Koji Sakamoto, Kazutoshi Fukui, "Study on the Contribution of Intake Noise Using Complex Time-Time Analysis and Subjective Evaluation", International Journal of Wavelets, Multiresolution and Information Processing, Volume: 8, Issue: 4, pp.609-625 (2010).

査読付き国際会議論文

- [51] S.Ishimitsu, T.Yamanaka, M.Saito, K.Fukui, "CONSTRUCTION OF SUPPORT SYSTEM FOR SPEECH IMPAIRED USING BODY-CONDUCTED SPEECH RECOGNITION", Forum Acousticum 2014,CDROM Regular Session R20D-1 6pages (2014).
- [52] M.Saito, S.Ishimitu, T.Yamanaka, K.Fukui, "Construction of A Speech Support System : Sound Quality Evaluation", The Proceedings of The 21th International Congress on Sound and Vibration, CDROM paper number160, 7 pages (2014).
- [53] Kazutoshi Fukui, Shunsuke Ishimitsu, Takahiro Yamanaka, Hyato nagoshi, "Improvement of Sound Quality Using Support System for Speech Disorders", The Proceedings of The 1st IEEE Global Conference on Consumer Electronics 2012, pp.39-43 (2012).
- [54] Masashi Nakayama, Shunsuke Ishimitsu, Hayato Nagoshi, Seiji Nakagawa,

- Kazutoshi Fukui, "Fundamental research on a body-conducted speech microphone using an Optical Fiber Bragg Grating for high magnetic field and noisy environments", Inter-noise 2011 Sound Environment as a Global Issue, 2011 September 4 through 7 Osaka, Japan, Proceedings CDROM 4pages (2011).
- [55] M.Nakayama, S.Ishimitsu, H.Nagoshi, S.Nakagawa, K.Fukui, "Body-conducted speech microphone using an optical fiber bragg grating for high magnetic field and noisy environments", Forum Acusticum 2011, Aalborg-Denmark, 27 June-01 July, 2011 CDROM: R003-Communication Acoustics pp.101-104 (2011)

学会発表論文

- [56] 福井 和敏, 石光 俊介, 名越 隼人, 山中 貴弘, "体内伝導音のサブワード単位線形予測による音質明瞭化と調音素性分析を用いた評価", 第 17 回 日本感性工学会大会, USB A44_TJSKE-D-15-00039 (2015).
- [57] 石光 俊介, 福井 和敏, 山中 貴弘, 齋藤 愛, 中山 仁史, "発声障がい者支援システムの構築", 第 57 回自動制御連合講演会, プログラム・概要集 CDROM 1D05-5 5pages (2014).
- [58] 山中 貴弘, 石光 俊介, 福井 和敏, "発声機能障がい者支援システムの構築—音質改善のための検討—", 日本音響学会 2014 年春季研究発表会 講演論文集 1-P5-1 pp.813-814 (2014).
- [59] Takahiro Yamanaka, Shunsuke Ishimitsu, Kazutoshi Fukui, "Study of Improving Sound Quality in Support System for Speech Impaired", The Proceedings of Eighth International Conference on Innovative Computing, Information and Control (2013).
- [60] Kazutoshi Fukui, Shunsuke Ishimitsu, Hayato Nagoshi, Takahiro Yamanaka, Seiji Nakagawa, Takayuki Kagomiya, "Speech Support System Using Body-conducted Speech For disorders And Its Evaluation", CUTTING EDGE LARYNGOLOGY (2013).
- [61] 山中 貴弘, 石光 俊介, 福井 和敏, "発声機能障がい者支援システムの音質改善法の検討", 第 15 回 IEEE 広島支部学生シンポジウム, USB pp430-433 (2013).
- [62] 山中 貴弘, 石光 俊介, 福井 和敏, "発声機能障がい者支援システムの音質改善に関する検討", Dynamics & Design Conference 2013, USB 論文集, No.423 6pages (2013).
- [63] 藤井 美奈, 石光 俊介, 福井 和敏, "発音機能障がい者のためのスマートフォンによる支援システムの開発", 日本音響学会 2013 年春季研究発表会 講演論文集, 1-P-42, pp.919-920 (2013).

- [64] 福井 和敏, 石光 俊介, 名越 隼人, 山中 貴弘, "発声機能障害者支援システムのための音質明瞭化手法の開発", Dynamics & Design Conference 2012, アブストラクト集 No.607 (2012).
- [65] 山中 貴弘, 石光 俊介, 名越 隼人, 福井 和敏, "発声機能障害者支援システムの音声認識による評価", Dynamics & Design Conference 2012, アブストラクト集 No.608 (2012).
- [66] 山中 貴弘, 石光 俊介, 名越 隼人, 福井 和敏, "構音障害者のための発声支援システムの構築～音声認識による音声評価～", 日本音響学会 2012 年春季研究発表会, 講演論文集 2-Q-25 pp.959-960 (2012).
- [67] 名越 隼人, 石光 俊介, 山中 貴弘, 福井 和敏, 籠宮隆之, 中川誠司, "構音障害者のための発声支援システムの構築～調音素性による音声評価～", 日本音響学会 2012 年春季研究発表会, 講演論文集, 1-Q-16 pp.859-860 (2012).
- [68] 中山 仁史, 石光 俊介, 名越 隼人, 中川 誠司, 福井 和敏, "高磁場・高騒音下で頑健な骨伝導光マイクに関する基礎検討", 日本音響学会 2010 年秋季研究発表会 講演論文集 CDROM pp.725-726 (2010).