HOKKAIDO UNIVERSITY

Title	自然な声を表出できる人工喉頭の研究とその展開
Author(s)	伊福部, 達; 上見, 憲弘
Citation	電子科学研究, 5, 23-29
Issue Date	1998-01
Doc URL	https://hdl.handle.net/2115/24404
Type	departmental bulletin paper
File Information	5_P23-29.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

自然な声を表出できる人工喉頭の研究とその展開

感覚情報研究分野　伊福部達，上見憲弘

喉頭癌などで喉頭を摘出すると，声帯音源を失うため声を生成できなくなる。そのための代用発声法が色々提案されているが，喉に振動子をあてて口腔内に音を送る人工喉頭が多く使われている。しかし，この方法は声質が悪く，発声音の自然性が極めて低いという決定的な欠点があった。我々の研究室では約10年にわたり，人工喉頭の音源波形を改良したり抑揚を出せるように工夫し，より自然な音声を表出できる人工喉頭を追究してきた。最近，抑揚を出せる人工喉頭については実用器が完成し，この実用器で講義を行っている者も出てくるなど，その有用性がほぼ確定した。喉頭摘出者団体からの強い要望の基に来年度から製品化することが決定した。さらに本稿では，声質の自然性を向上させる研究は高品質音声合成に活かされつつあることを述べとともに，将来的な埋込み型人工喉頭を想定して筋電位を用いた抑揚の制御法について説明する。

1　はじめに

喉頭癌などで喉頭摘出手術を受けた人達は，音声器官のうち音源となる声帯を失うため，そのままでは音声を生成することが困難になる。このような喉頭摘出者の多くは高齢者であるため，これから到来する高齢社会では急増することは必至である[1]。これらの人達は，喉頭摘出後にも比較的残されている声道や舌などの構音器官に，失った声帯音源の代わりになる音を送り込むことにより発声を行っている。

現在，このような代用発声法は，器具を使わない方法と人工喉頭と呼ばれる器具を使う方法の2つに大きく分けることができる[2]-[4]。器具を使わない方法には，図1 (a)の食道の一部にためた空気を利用する食道発声法や，図1 (b)の気管と食道間に開けた細い孔からの呼気を利用するT-Eシャント法などがある。一方，器具を用いる人工喉頭には，図1(c)のように，リードの音をパイプにより口腔内に送り込む笛式人工喉頭や，図1(d)のように，振動音を口腔を通じてあるいは経皮的に声道内に伝える電気式人工喉頭などがある。

これらの内，食道発声法が最も普及してきたが，訓練のために長期間を必要とし，また，体力消耗などで高齢者には不向きであるなどの問題もある。一方，人工喉頭は比較的修得が容易であるが，その音源が健常者の声帯音源と大きく異なるため，生成される音声は不自然になる。我々は，食道発声の修得に成功しなかった人達や，喉頭摘出直後の人達に利用できるような人工喉頭の開発研究に取り組んできた。とくに，従来の課題であった人工喉頭音声に自然性を付与する方法について色々な角度から調べてきた[5]-[7]。また，このような音声

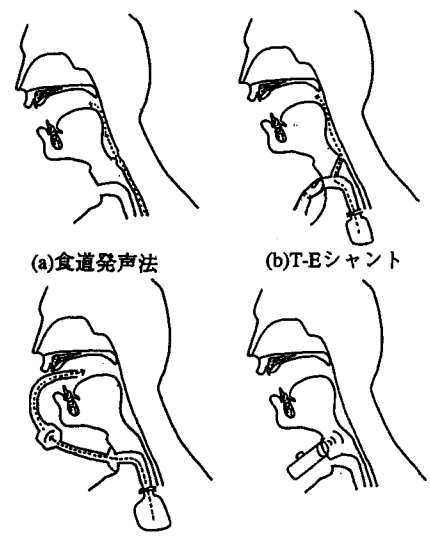


図1　従来の代用発声法

の自然性を向上させる研究は人工喉頭音声の改良のみならず，音声合成などのより汎用的な分野にも適用できる[8]。本報告では，人工喉頭音声の自然性の向上のために行ってきた研究とその展開について述べる。

2 自然性を備えた人工喉頭音声を作り出す方法

人工喉頭は声帯音源の代わりとなる音源を作り出すための装置であることは前節でのべた。この装置による音声の自然性をあげるための方法として，(1) 人工喉頭音源波形自体の改良，(2) 人工喉頭音源へのアクセント・イントネーションの付加，の2点について検討を行ってきた。これらの検討の結果について述べる。

2.1 人工喉頭音源波形の改良に関する研究

現在の電気人工喉頭の音声が不自然な理由として考えられることの一つには，その音源波形が健常者のものと大きく異なっていることがあげられる。また，人工喉頭は音声の中でも主に母音を作り出すのに用いられる装置である。そこで，我々は母音の自然性を決める要因を探り，その結果を人工喉頭に応用することを検討してきた。定常な母音は，ほぼ一様な波形がピッチ周期ごとに繰り返された時間構造をもっている。しかし，その形は時間とともに微妙に変化している。われわれは，このような波形の微妙な変化つまり波形ゆらぎが音声の自然性にどのように関与しているかを心理物理実験を用いて調べてきた。本節では，まずその実験に触れ，つぎにその結果を人工喉頭装置で評価した結果について述べる。さらにこのような研究は人工喉頭音声の音質向上のみならず，より高品質な合成音声を生成するのに活かすことができる。本節の最後に，人工喉頭研究の展開としてこのような側面から検討を行った結果についても触れたい。

2.1.1 音声の波形ゆらぎと人工喉頭音声について

2.1.1-a ゆらぎの周波数成分と自然性の関係

ゆらぎを含んだ波形をフーリエ変換することにより，周波数領域におけるゆらぎと自然性の関係を探った。自然性を保つには，母音から切り出した時と同じ順序のピッチ波形が32個程度あれば良いということを以前の実験結果から得ている[5]。そこで，正しい順序で並べられた32個のピッチ波形から構成される5母音（ゆらぎ有）を用意し，まず，それぞれ32ピッチ分をまとめてフーリエ変換した。1つのピッチ波形は256点で構成されるようにし，FFTは8192点で行った。

実験では，波形ゆらぎによって生じている高調波成分の側波を一度全て除いて線スペクトルにした後，少しずつ側波を戻していくときに自然性がどのように変化するかを調べた。側波は，各高調波成分の周波数に近い側波から，上側波，下側波をペアにして両側で1つずつ戻して行くようにした。つまり，試験音1番は，側波を全て除いた線スペクトルに，試験音8番は各高調波成分の両側に7つずつの側波を戻したものになっている。そして，各試験音を4回繰り返して接続し，被験者に聴かせた。

上述の方法で作成した8つの試験音を用いて自然性の一対比較を行った。つまり，試験音全ての組み合わせ（56通り）をランダムに呈示し，被験者には二つの音のうちどちらがより人の声らしいかを応答させた。そして，自然に聞こえた試験音には1点を，どちらとも区別ができないときには両方の音に0.5点を与えた。その総得点によって自然性を判断した。

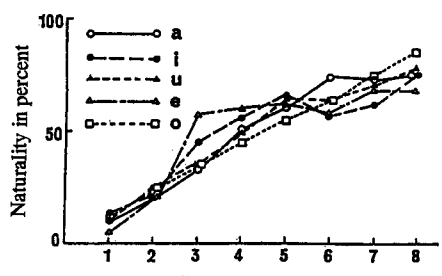


図2 ゆらぎを取り除いたあとに各高調波成分に側波を順に付け足して行ったときの音声の自然性の変化

3回の平均を5母音について求めた結果を図2に示す。縦軸は各試験音の総得点の割合であり，100%とは他の全ての試験音より自然性が高いと判断されたことを意味する。図から，少しの側波を戻しただけで自然性が向上していき，側波を4つずつ戻したあたり（試験音5に相当）で自然性の上昇がゆるやかになる傾向が見られた。今回の実験では，周波数分析のきざみ幅を約4.5 Hzとしているので，高調波成分の周波数を中心におよそ±18 Hzの範囲内にある側波が母音の自然性に特に深く関わっているものと考えている。聴覚ではこのようなゆらぎを手がかりに自然性の認知を行っているものと予想される。

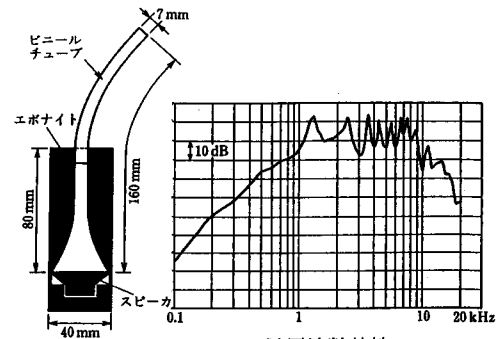
2.1.1-b　パイプ挿入式電気人工喉頭による評価

このようなゆらぎを持つ波形を電気式人工喉頭の音源として使用することによって，音質の改善が可能かどうかの評価を行った。ここでは，歪みの少ない波形を送り込める可能性があること等の理由から，図3（a）に示すようなパイプ挿入式の電気人工喉頭を試作し用いた。

実験では，健常者1名（23歳男性）の口腔に人工喉頭のパイプを挿入し，口形のみを変化させて発声した音源の異なる4種類の母音を用いて，一対比較法により自然性の評価を行った。その音源は，(a) ローパスフィルタ（-12 dB/oct，カットオフ周波数 90 Hz）にインパルス列を入力して作成した波形，(b) 母音/a/から線形予測分析によって32ピッチ分の残差信号を求め，残差信号から側波を除去して作成した波形，(c) 各高調波成分の周波数に ±18 Hz 幅の側波だけを付与した残差信号から作成した波形，(d) 各高調波成分の側波をすべて残した残差信号から作成した波形，である。被験者は4名である。

結果を図4に示す。グラフの横軸は音源波形の種類を，縦軸は各音源波形が自然性の一対比較によって得た総得点の割合を表している。実験から，(c) と (d) の自然性が高く，(a) と (b) が低いという結果が得られた。このことから，高調波成分に ±18 Hz 幅の側波だけを残した音声音源波形を用いることによって，人工喉頭音声の自然性が向上すると言える。

ただし，実際に口腔に送り込まれている波形は図3(b) に示すような人工喉頭の共振特性のためにゆがんでしまっており，入力した音源波形とはかなり違った波形となってしまっている。この問題は，パイプ挿入式の構造上避けて通れないものであるが，このような場合でも十分自然性を高められることを確認している。



(a) 概略図　　(b)周波数特性

図3　パイプ挿入式電気人工喉頭

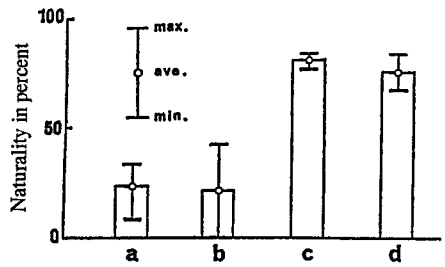


図4　音源を変えることにより得られた4種類の人工喉頭音声の自然性の比較

2.1.2　音源波形ゆらぎと合成音の自然性について

前項までの結果は音声合成においてもゆらぎの付加により高品質の合成音が得られることを示唆している。そこで，本項では人工喉頭研究の展開として波形ゆらぎ成分の統計的性質を調べ，その結果を合成音声で検証した結果について述べる。

2.1.2-a　波形ゆらぎの統計的性質

まず，波形ゆらぎがどのような統計的性質を持っているかを調べた。解析に用いた音声資料は喉頭に疾患のない22～26才の男性被験者10名よりサ

ンプリングした正常持続発声母音/a/である。波形ゆらぎのピッチ同期的な性質を調べるため，ピッチ周期内の相対位置が等しい点の振幅をピッチ同期的に128点抽出し，その周波数特性を求めた。同様の解析を一人の被験者について10個の音声資料について行い周波数領域において加算平均したところ，512個のピッチ同期的波形ゆらぎの周波数特性は全て $1/f^{\beta}$ 特性を示すことが明らかになった。

図5は512個のピッチ同期的波形ゆらぎの周波数特性を $1/f^{\beta}$ 特性として近似するときの β の値について全被験者の平均と標準偏差を示した。結果として，ピッチ同期的波形ゆらぎの周波数特性を $1/f^{\beta}$ 特性で，$0.5 < \beta < 1.5$ 程度としてモデリングすることの妥当性が示唆された。次に，波形ゆらぎのピッチ周期内における特性を調べるため，各ピッチ周期に対応する波形ゆらぎについて周波数特性を求め，周波数領域において128ピッチ周期分の加算平均をおこなった。結果として，ピッチ周期内の波形ゆらぎの周波数特性もまた $1/f^{\beta}$ 特性で，$2.5 < \beta < 3$ 程度であることが明らかになった。

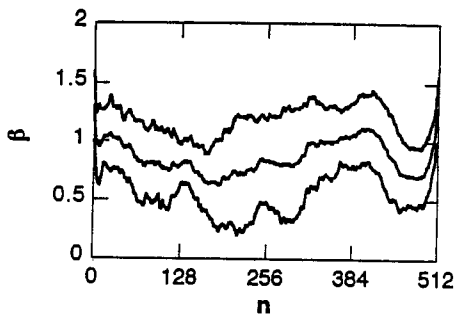


図5　512個のピッチ同期的波形ゆらぎの周波数特性を1／f $^{\beta}$特性として近似するときのβの値の平均と標準偏差

2.1.2-b 波形ゆらぎの生成と知覚

本稿で提案する波形ゆらぎの生成法ではピッチ同期的特性については，FFTに基づくfractional Brownian motion(fBm)法によって $1/f^{\beta}$ 特性の信号を生成することによって実現し，ピッチ周期内の波形ゆらぎの特性については中点変位法によるfBm法により $1/f^{\beta}$ 特性の信号を生成することによって実現した[8]。次に，音響心理実験のために，波形ゆらぎを考慮した呈示刺激音を合成した。呈示刺激音は持続時間 1 sec の母音/a/でありPARCOR分析合成により作成している。

聴取者は音声研究に従事しない22～26才の20名である。各刺激音は防音室内において被験者より 1.5 m 離れたラウドスピーカにより音圧レベル65 dB で呈示された。

一対比較法の系列刺激として用いた刺激音の波形ゆらぎは，A）基準刺激音（原音）と同じ，B）提案する波形ゆらぎ，C）-6 dB/oct 特性の雑音，D）波形ゆらぎなし，の四種類である。図6に，各刺激音グループの基準刺激音に対する類似度の平均値と範囲を示した。

図から刺激音グループA，Bについては類似度が極めて高いが，刺激音グループC，Dについては類似度が極端に低くなることがわかった。刺激音グループCは粗造性を強調した音質，刺激音グループDはバズ音質に知覚されるため，どちらの刺激音グループも基準刺激音とは異なると知覚したとする内観報告をほぼ全被験者から得ている。言うまでもなく，波形ゆらぎが音声の自然性に大きく貢献するという知見は，人工喉頭に活かされるばかりでなく，高品質音声合成の研究へも発展している。

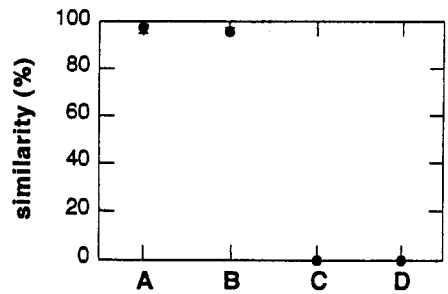


図6　波形のゆらぎ方が違う4種類の音声と原音声との類似度の評価
A）原音声と同じ，B）提案する波形ゆらぎ，
C）-6d/oct特性の雑音，D）波形ゆらぎなし

2.2　イントネーション制御に関する研究

人工喉頭音源に波形ゆらぎを付与すると自然性が向上することがわかった。ただし，この改良だけでは人工喉頭音声は単調な声になる。本項では人工喉頭研究のもう一つの側面である，人工喉頭音源にイントネーションをつける方策について述べる。

イントネーションやアクセントは自然な声を実現するためには重要な要素である。一方，これらには言葉の意味を明確に伝えるという役割と，話者の意志や感情を伝えるという重要な役割がある。つまり，話者自身がそれらを制御できることが必要である。また，喉頭摘出者に残存する音声器官の一部を利用してイントネーションを作り出せれば，訓練は短期間ですむであろう。

以上のことから，我々は実用により近い方法として，喉頭摘出者の呼気を利用する方法と，将来的に埋め込み型人工喉頭の制御手段として，発声に関与する筋の一つである胸骨舌骨筋を用いてイントネーションを制御する方法の２つについて検討してきた。その結果について報告する。

2.2.1　呼気圧を用いた抑揚制御型人工喉頭

健常者の音声生成時には，呼気量や声帯筋を制御して音声強度やピッチ周波数の変化を作り出しており，呼気の制御は自然な音声を生成する上で重要な役割を果たしている。喉頭摘出者は声帯筋をも切除している場合が多いことから，呼気量は音声強度やピッチ周波数を制御する信号として活用され得る。そこで，我々は喉頭摘出者の残された機能をできるだけ活用する意味で呼気圧によりイントーネーションを制御できる人工喉頭の実用化研究をおこなってきた。このような人工喉頭で問題となるのは，呼気を呼気圧に変換するための気流抵抗の値と呼気圧からイントネーションに変換するときの関数である[6]。パラメータの一つである呼気圧対周波数変換関数を求めた時の実験装置と結果を図７，図８に示す。

実験装置は，図７左のような 1.呼気圧検出部，2.呼気圧－周波数変換部，3.振動子の３つの要素からなる。被験者は本人工喉頭の使用法に精通した49才男性1名である。変換関数は呼気圧を P，声の高さを f として，$f[\mathrm{Hz}] = k \times (P[\mathrm{cmH_2O}] - 1) + 60$ とし，傾き k の値を $12, 25, 50, 66.7, 100$ Hz/cmH$_2$O の５種とした。まず，被験者はそれぞれの傾きに設定した装置で音声/aoiumi/を発声し，その時の制御のしやすさを図７左下の表に丸を付け示した。また，それぞれの傾きから得られた５つの音声で，一対比較（図７右）をおこない，声の自然さを評価した。聴取

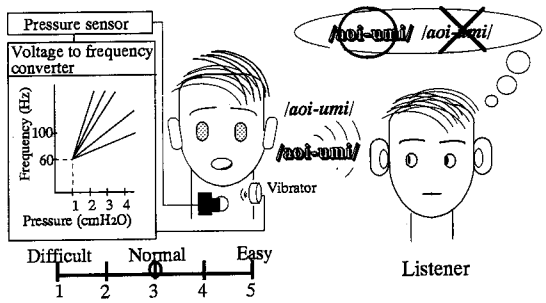


図７　呼気圧によるイントネーション制御型人工喉頭の評価方法

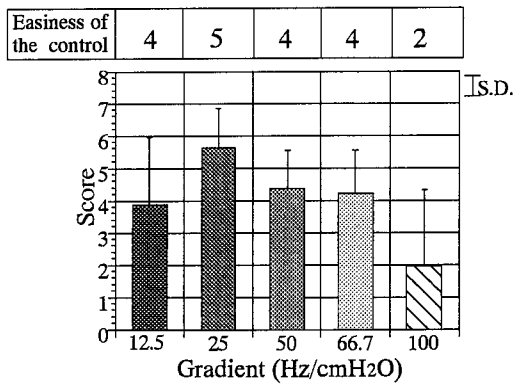


図８　制御のしやすさと人工喉頭音声の自然さの評価

被験者は健聴者６名である。

図８の横軸は関数の傾きを表す。また，縦軸は音声の自然さを表し，点数が高いほど自然に聞こえたことを示す。また図の上の表内の数字は１を「制御しづらい」，３を「普通」，５を「制御しやすい」としたときの値である。自然性および制御のしやすさから判断すると傾き 25 Hz/cmH$_2$O が適している。同様の実験を気流抵抗値についても行うことにより，ある程度本人工喉頭に適したパラメータを見つけることができている。本人工喉頭で発声を行った時のイントネーションのパターンを図９に示す。この人工喉頭は現在実用化の段階にあり，来年の４月には製造販売する予定である。図10に実用型人工喉頭使用時の概観を示す。

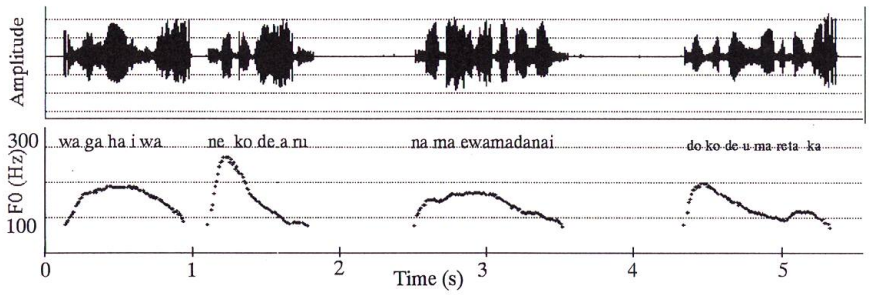


図9　本人工喉頭で発声を行った時のイントネーションのパターン
（小説「吾輩は猫である」の一節）

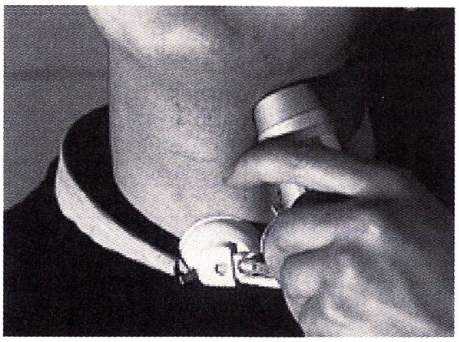


図10　イントネーション制御型人工喉頭
使用時の概観

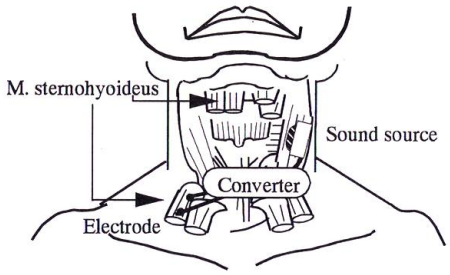


図11　埋め込み型人工喉頭の概念図

2.3　埋込み式人工喉頭のための筋電位による声高さ制御

理想的な電気式人工喉頭とは，意図したときに自由に声が出せて，声の強さや高さの制御ができ，良好な音質を持ち，日常生活に支障がないものである。このような観点から，将来的に埋込み形電気人工喉頭を想定した研究も行っている[7]。その制御手段として，喉頭摘出時に残すことのできる発声筋の節電位を用いる方法について検討した結果について述べる。

喉頭摘出時に残せ得る発声に関与する筋として，胸骨から舌骨に付着している胸骨舌骨筋（sternohyoideus，以下SHと略す）がある。SHは舌骨を引き下げる筋であり，特に発声においては，通常より低い声を出すときに活動するといわれている[9]。この筋を舌骨付着部で切断したまま保存すれば，電気式人工喉頭の制御にも利用することができるであろう。本人工喉頭の概念図を図11に示す。いままでに，健常者のSH付近の筋電位を用いて音源のオンオフ制御や振動周波数すなわち声高さの制御実験を行なっている。また，喉頭摘出者でも，簡単な制御実験を行っている。

まず，健常者の筋電位による声高さの制御能力について述べる。実験では，電子ピアノのド(C3)，レ(D3)，ミ(E3)，ファ(F3)の音を呈示音として被験者に聞かせた後，その呈示音と同じ音程で，調音器官を/i/の発声と同じ構えにして，試作の電気式人工喉頭により音を出すように指示した。ちなみにこの人工喉頭では，SHの機能を考慮して，筋電位の振幅が小さくなるに従って振動周波数が高くなるように設定してある。

図12に，呈示した音の高低に人工喉頭の声の高さを追従させたときのピッチ周波数と，そのときの正規化されたSHの筋電位の大きさを示した。また，試作電気式人工喉頭を用いた発話でも，個人によるばらつきは大きかったものの，平均85～143 Hzまで声高さを変化させることができた。

つぎに，喉頭全摘出術においてSHを舌骨下で切断し保存した被験者に，喉頭摘出以降8ヶ月後

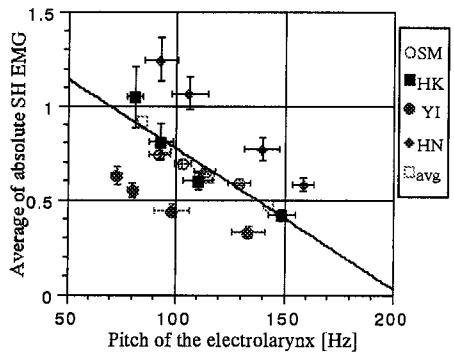


図12 C3(ド)からF3(ファ)までのピッチ変化の指示に対する 試作人工喉頭の音声のピッチとその時の胸骨舌骨筋の筋電位

に，SHで制御する試作電気式人工喉頭を用いて/a/の発声を行わせたところ，不安定であったが一応の発声はできたことを確認している。SHを残すことができれば，その筋電位の制御でイントネーションをつけることのできる可能性が示された。

現在，埋込み形電気式人工喉頭の一番の難題は小形大出力の発音体を開発し，それをどこに埋め込むかである。今後，圧電形スピーカなどの進歩により，この難題も解決すると考えている。

3 おわりに

本稿では私共が行ってきた人工喉頭音声の自然性を向上させる研究について述べた。このような研究を行うためには，まず，人がどのようにして自然な声を生成し認知しているかを知らなければならない。この基礎研究の成果は障害者のための補助装置を実現する人工喉頭に活かされたのは言うまでもないが，より一般的な合成音声の自然性を向上させる研究などに結びついてきた。人工喉頭については，今後より自然で使いやすい装置設計のために現場の意見をとりいれ，普及のための努力を続けたいと考えている。

なお，本研究を推進した八幡英子（現：歯学部助手），橋場参生（現：北海道工業試験場研究員），関惠貞，青木直史（現：ＤＣ１）に感謝の意を表す。

[参考文献]

[1] 馬谷克則, 鶴田至宏, 吉野邦俊, 宮原裕, 佐藤武男: 日本気管食道科学会会報, 36(3), 261 (1985)

[2] 福田宏之: 日本音響学会誌,44(2),130 (1988).

[3] 高橋宏明: 音声言語医学,28(2),132 (1987)

[4] Electroacoustic Analysis and Enhancement of Alaryngeal Speech, edited by A.Sekey, Charles C Thomas, Springfield, Ill.(1982)

[5] 伊福部達, 橋場参生, 松島純一: 日本音響学会誌, 47(12), 903 (1991)

[6] 上見憲弘, 伊福部達, 高橋誠, 松島純一: 医用電子と生体工学, 33(1), 7 (1995)

[7] 関惠貞, 高橋誠, 西澤典子, 西澤伸志, 伊福部達, 犬山征夫: 医用電子と生体工学, 32(4), 297 (1994)

[8] 青木直史, 伊福部達: 音響学会聴覚研究会資料, H97-43, 1 (1997)

[9] Hirose, H: Haskins Status Report, SR25-26, 73 (1971)