

吃音者の発声におけるピッチ制御の聴覚フィードバック特性

佐藤 裕* 森 浩一* 福島 康弘* **

Control Characteristics of Voice Pitch by Auditory Feedback in Persons Who Stutter

Yutaka SATO*, Koichi MORI* and Yasuhiro FUKUSHIMA* **

Functional abnormality in auditory feedback has been hypothesized to be one of the causes of stuttering, a speech dysfluency. We investigated this possibility by a transformed auditory feedback method (TAF; Kawahara, 1993), which is a method for measuring the effect of auditory feedback onto voice pitch (F0) using parametric perturbations inserted into the feedback loop. The results showed that the stutterers' cross-correlation functions between the F0s of auditory feedback and uttered voice differed from those of non-stutterers. Furthermore, the control characteristics of voice pitch were analyzed in the first and middle parts of phonation separately. The cross-correlation functions between the F0s of auditory feedback and uttered voice differed significantly between the first and middle parts of phonation in nonstutterers. However, the control characteristics for those who stutter were diverse, suggesting that different etiologies or pathophysiology are involved among those who stutter.

キーワード：変換聴覚フィードバック, ピッチ制御, 吃音, F0(基本周波数)

1. はじめに

発話と聴覚が相互に関係している証拠として、Lombard効果[1]や遅延聴覚フィードバック (Delayed Auditory Feedback: DAF)[2]が挙げられる。Lombard効果は、騒音下での発話により自分の声がよく聞こえない環境では、静かな環境に比較して、声が大きくなり、有声音の部分が長くなる等、音声に変化する効果のことである。DAFとは、自分の発話を時間的に(50~200 ms程度)遅らせたものを聴覚にフィードバックさせることであり、吃音に似た症状が誘発されることが知られており、これによる乱れた発話を人工吃音と呼んでいる。また、発話の障害である吃音が聴覚フィードバックに関連していることが知られている。DAFを

吃音者に適用することで吃音がかえって改善する[3]ことや、雑音(白色雑音等)を用いて聴覚遮断した時に吃音が減少するマスキング効果[4]等から吃音者は聴覚フィードバックに障害があるのではないかという説が提唱された(Servo Theory)[5]。その他、音声のピッチをオクターブシフトさせて聴覚にフィードバックさせる手法(Frequency Altered Feedback: FAF)[6]も吃音を減少させる。これらの手法は、吃音の改善にはある程度有効であるが、さらに分析的に検討しようとする、定量性が不十分である。

河原は発声における音声知覚から生成への作用を定量的に明らかにすることを目的とし、聴覚フィードバックされる音声の周波数に摂動を加えた人工的なフィー

* 国立身体障害者リハビリテーションセンター研究所
感覚機能系障害研究部

** 玉川大学工学部

* Department of Rehabilitation for Sensory Functions,
Research Institute of National Rehabilitation Center
for Persons with Disabilities

** Tamagawa University Faculty of Engineering

ドバック、すなわち変換聴覚フィードバック (Transformed Auditory Feedback, TAF) により、発声された音声の基本周波数 (F0) に対する聴覚の影響を測定した [7]。そして、基本周波数の変動に対しては約150 msの遅れを伴う補償方向の応答が発声時に働いていると報告している [8,9]。

DAFの効果が吃音者と非吃音者で異なるのは、音声のフィードバック制御特性に違いがある可能性が考えられる。本研究では、TAFを用いてこれを検討したので報告する。

2. 方法

2. 1. 被験者

非吃音被験者は特別なボイストレーニングを受けていない日本語話者10名 (男性7名、女性3名、平均29.7歳) である。絶対音感保持者が2名 (女性2名) 含まれていたが、統計上差を認めなかったため単一群として検討した。

吃音者は11名 (男性8名、女性3名、平均30.0歳) であり、治療施設およびセルフヘルプグループより実験協力者を募集し、実験の前に説明を行い書面にて同意を得た。この研究は国立身体障害者リハビリテーションセンター倫理委員会の承認を得ている。

2. 2. 装置

実験系は河原が用いたTAFの手法 [10] に準じた。マイク (SM58, SHURE) からヘッドフォン (ST-12M, ASHIDA SOUND) に至る人工的な音響フィードバック系に周波数変換装置 (SE70, BOSS) を挿入し、変換量をMIDIコントローラ (Power Macintosh 9600/300, アップルコンピュータ) により制御することで、フィードバック音声に1/2半音以下の周波数の摂動を加えた。また、自身の気導音や骨導音による影響を取り除くため、密閉型のヘッドフォンを用いると共に、ノイズ発生装置 (1405, B&K) により、約80dB(A) のピンクノイズを加え自声を遮蔽した (図1)。

摂動のための制御信号には、擬似白色信号であるM系列信号を用いた。M系列信号は、時間ずれのある自己相関が0になる特徴があり、システムの応答特性を測定する目的でよく使われる。実際に与えた制御信号は、周期31のM系列を8倍にオーバーサンプリングした後、8 Hzのローパスフィルターで帯域制限して作成したMIDIデータ (河原英紀教授提供) を、ピッチバンド信号として1/128秒毎に送出した。この摂動信号のくり返し周期は2秒である (図2)。フィードバック音声とそのときに発声された音声は、44.1kHzでサン

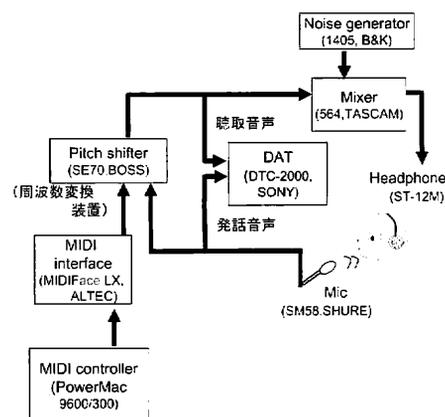


図1 実験装置構成図

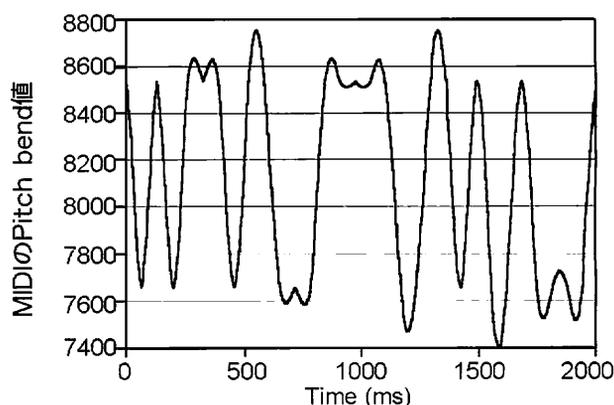


図2 用いられた摂動信号

プリングされてそれぞれDAT (Digital Audio Tape Deck DTC-2000ES, Sony) の左右チャンネルに同時に取り込まれた。

2. 3. 手続き

被験者には、母音/a/を約1分間にわたって息継ぎをしながら発声するよう教示した。約10秒毎に息継ぎの合図を実験者が示した。発声する声の高さ (ピッチ) は自分の発しやすいものにし、できるだけ一定にするよう教示した。

2. 4. データ処理

データ処理は河原の方法 [10] をもとに、若干の変更を加えて以下のステップで行った。

- (1) F0の抽出: DATに記録されたデータを左右チャンネルの同期を失わないように発声部分を含めて連続データとして取り出し、5 ms毎に25 msのフレーム長でF0を抽出した (Windows版音声録聞見, Datel)。
- (2) 同期加算による平均化: F0のデータを左右チャンネルの同じ位置から2秒毎に切り取り、摂動信号の周期性を利用しそれぞれのチャンネル毎に約1分間の

データすべてに渡り同期加算を行った。ただし、F0の抽出に関して2秒ずつ切り取った際にいずれかのチャンネルにF0が0(無音)を含む回は加算から除外した。同期加算の結果、400ポイントの信号系列となる。

(3) 相互相関の計算：河原の原法[10]では、使用したM系列と発話音声のF0の相互相関を計算し、相互相関が最大値になる時点時刻を時刻0と定義している。本研究では、制御特性の遅延時間の絶対値を評価しやすいように発話音声とフィードバック音声のそれぞれについて同期加算平均を行い、両者のF0の周期的相互相関を計算した。

また、非吃音者では発声開始付近とそれ以降で特性が異なる可能性が認められたので、さらに詳細な検討のため、発声開始期と持続中のフィードバック特性を比較した。上記2のステップにおいて、それぞれの持続発声中で2秒毎に区切られた最初のを発声開始期 (first part, F) のデータとして採用し、他の区間は発声持続中 (middle part, M) のデータとし(図3)、両条件毎にF0の相互相関を算出した。なお、発声開始からFの区間までの時間は発話毎にランダムであり、吃音者・非吃音者間で有意差はなかった(吃音者：平均=1.07s, SD=0.21s; 非吃音者：1.08s, SD=0.21s; p=0.90)。

3. 結果

3.1. 摂動信号あり・なしの比較

図4はフィードバック音声に周波数摂動を加えていない場合の相互相関を、図5はフィードバック音声に周波数摂動を加えたTAF法の場合の周期相互相関を代

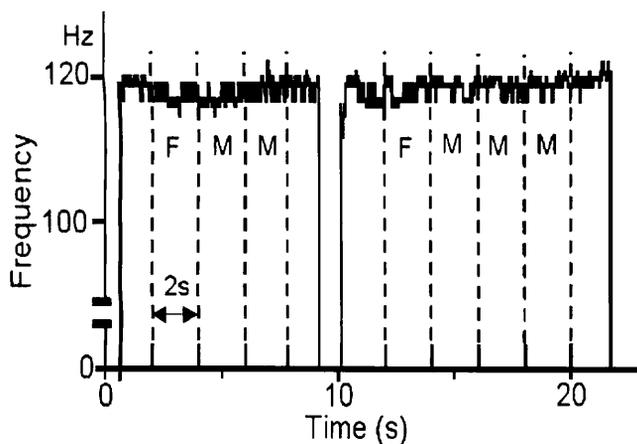


図3 発声F0の軌跡とデータ採取法

1 被験者の発話データのF0の0~約25秒を图示。0 Hzになっている部分は息継ぎ。摂動周期2秒毎に切り取り、データとして採用可能な区間について、各約10秒の発声中、最初の区間をfirst (F)、それ以降をmiddle (M) partとして採用。

表例について示す。図4では、発話音声は約10 msの遅延以外は無処理で聴覚フィードバックになっているため、グラフはほぼ発話音声の周期自己相関を示していることになる。また、1000 msを中心にほぼ対称の特性になっているが、これは連続音声のF0を2秒毎に区切って加算した上で周期相関を求めているため、通常の自己相関に折り返しが重畳しているものと考えられる。図4と図5の比較から、発話音声にM系列のピッチシフトをかけることによりフィードバック音声は十分に白色化されており、フィードバック音声から発話への影響を観察しうると考えられる。また、図4は1000 ms付近で0に近づいており、これにより十分に短い潜時では折り返しの影響が小さいと考えられるので、潜時500 ms以内の特性についての結果を検討した。

3.2. 発話全体 (F・Mを含む) の比較

非吃音者と吃音者1名ずつの発話音声とフィードバック音声のF0相互相関の代表例を図5と6に示す。縦軸は相関係数を、横軸はフィードバック音声から発話音声への遅延時間を示している。両者とも1990 msに最

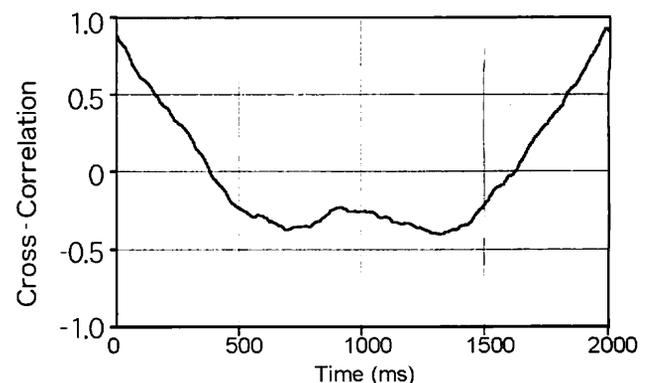


図4 非吃音者の発話音声とフィードバック音声の周期相関 (周波数変調なし)

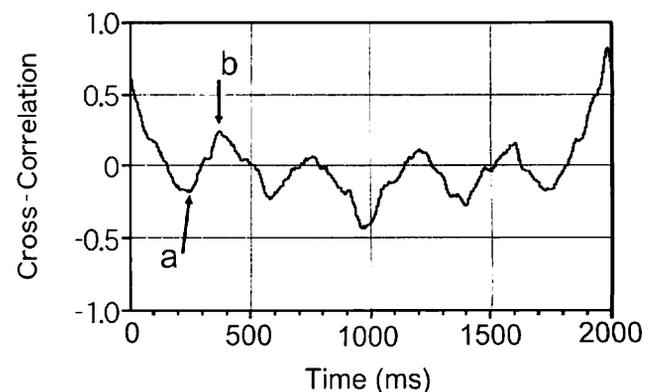


図5 発話音声とフィードバック音声のF0周期相関 (非吃音者)

大のピークがあるが、これは周波数変調装置の処理遅延のため、聴覚フィードバック信号が発声音声から10 msあまり遅れているため、2 sの周期相関であることを考慮すると、本来は-10 msである。

図5にa, bと示したように、相互相関係数が極大・極小を示す位置に着目し、0 msに近い方から2つの主要なピークの差をみると、非吃音者に比較して吃音者はこれが小さいという特徴がみられる。このピーク差について、500 ms以内にピークがみられない者（非吃音者2名、吃音者3名）を除いて両群で比較すると、非吃音者に比べて、吃音者で有意に小さくなっていた（図7）。

3. 3. 発話開始付近と持続発話中の比較

典型的な非吃音者1名（図5と同一被験者）の発声音声とフィードバック音声のF0の相互相関を、発声開始付近と持続発話中にわけて分析した結果を図8に示す。相互相関係数が0 msに近い方から最初に示す極小・極大のピーク差をみると、発声開始時に比べ持続発話中の方が大きくなっている。10名中、500 ms以内にはっきりした極小・極大のピークを認めない2名を除く非吃音者8名全員で、発声開始付近より、発

声持続中でピーク差が大きくなっていた（図9）。

図10に典型的な吃音者1名（図6と同一被験者）の発声開始付近と持続発話中における発声音声とフィードバック音声のF0の相互相関を示す。発声開始期には極大・極小の差がほぼなく、変曲点のようにも見える。図8に比して発声持続中もピーク差が比較的小さい。他の吃音者の結果も合わせてまとめると、発声開始付近と発声持続中の両方ともに500 ms以内に極大・極小が出現した吃音者は11名中5名のみで、出現率は吃音者でやや低かった（ $p=0.104$ ）。極大・極小が評価できる5名では発声開始付近と発声持続中でピーク差の関係は一定せず、有意差はなかった。発声開始付近と発声持続中の両方ともに500 ms以内に極大・極小が出現した非吃音者（8名）と吃音者（5名）の持続発話中のピーク差を比較すると、後者の方が小さい傾向がみられた（図11）。

また、残りの6名中5名では、発声開始時、持続中の少なくともどちらかのピークが500 ms以内に出現しておらず、他の1名は発声開始付近のデータが解析不能であった。吃音者の結果はばらつきが大きかった。

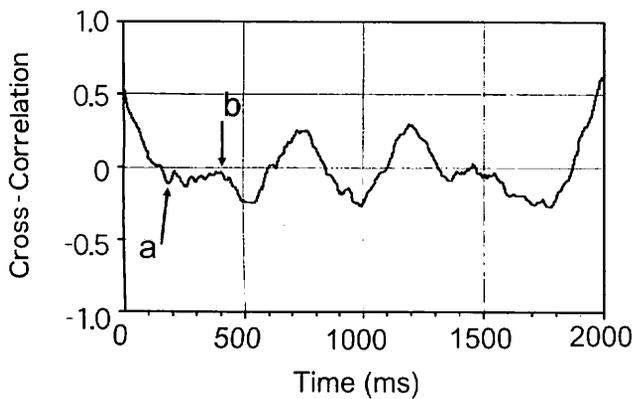


図6 F0周期相関（吃音者）

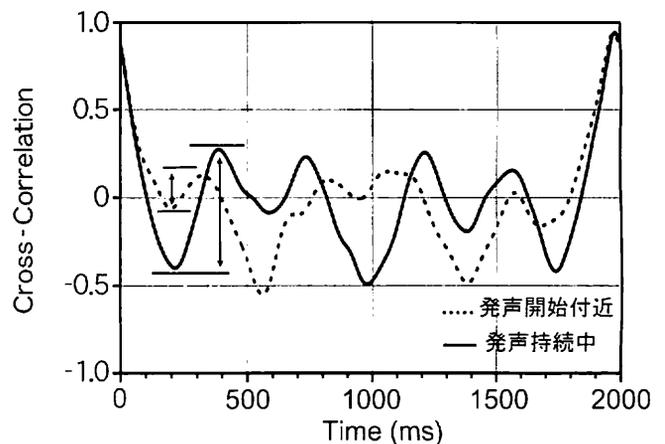


図8 発声開始・持続中のF0周期相関（非吃音者）

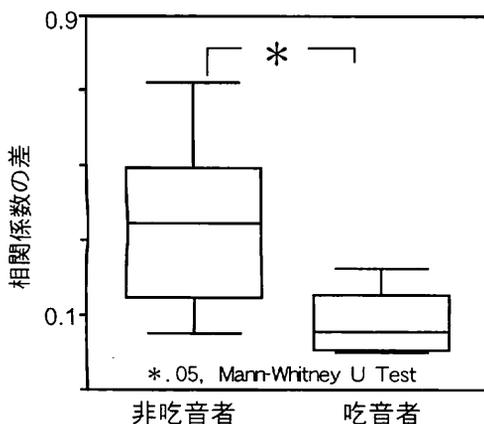


図7 相関係数 (b-a) の差

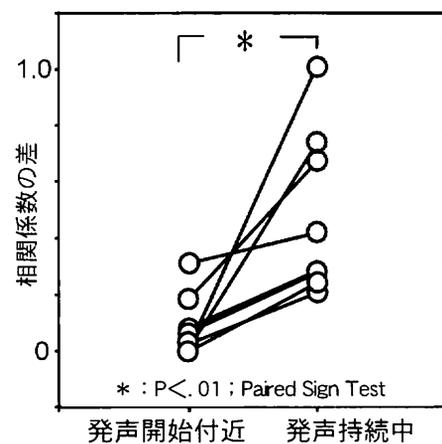


図9 発声開始・持続中での相関係数のピーク差（非吃音者）

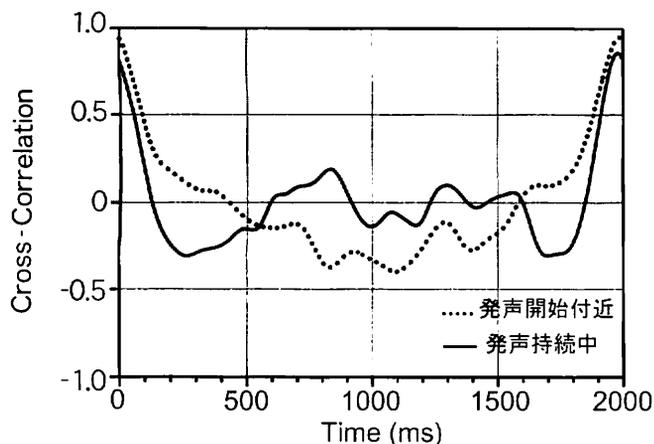


図10 発声開始・持続中のF0周期相関 (吃音者)

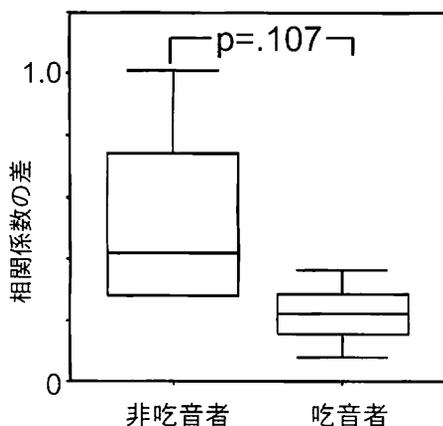


図11 持続発話中の相関係数のピーク差

3. 4. 持続発話中の解析

持続発話中のデータのみで、500 ms以内に極大・極小が出現した者は非吃音者で10名中9名であり、吃音者では11名中6名であり、両群におけるピークの出現する者の比率は有意傾向があった ($p=0.072$)。上記の非吃音者9名と吃音者6名の持続発話中のピーク差は有意差がなかった ($p=0.157$)。

4. 考察

本研究では、TAF[7]の変法を用いて発話音声とフィードバック音声との相関を算出し吃音者と非吃音者とを比較した。

河原らの原法[7]は、発話音声と振動原信号との周期相関を算出し、時間0の原点としては周期的相互相関の最大値を与える位置を採用しており、必ずしも時間軸の絶対精度が保証されない。実際、図5に示すように、 $a \cdot b$ で示されたピーク以降にもピークが存在しており、振動原信号との相関においても、どのピークが

最大となるかは一定しない場合があると思われ、時間0が一定しない可能性がある。今回の変法は、同時記録した発話音声とフィードバックされる音声のF0間で周期的相互相関をとっており、聴覚フィードバックから発声までの制御の遅延時間は正確に求まる。その一方、周波数変調装置の処理時間の遅延に一致して自己相関によるピークが負の領域 (ないし折り返されて周期の最後付近) に必ず生じることと、これ以外の音声の自己相関も結果に混入することが、フィードバックの特性を純粹に抽出するには問題となる。しかしながら、図4と図5を比較すると、自己相関の鋭いピークは0 ms付近以外ではほとんど認められないことがわかる。また、 -10 msのピーク潜時が一定であることは、この変法の時間精度が良好であることの傍証となる。今回の変法は、時間精度を保証することにより、聴覚フィードバックの時間特性が問題となる吃音者への応用に適していると考えられた。

本研究で用いた手法により算出される相関は、フィードバック音声の聴覚入力による発話音声への制御特性を評価した、フィードバック機構のインパルス応答と自己相関が含まれている。また、今回は500 ms以内の相関の極大・極小のピーク差に着目した。これにより、自己相関による折り返しを含む遅い成分を解析から除外し、人工吃音を起こしやすい遅延 (150–300 ms) や治療に使われる遅延時間 (50–250 ms) を含む。聴覚フィードバックに重要なフィードバックの自動制御機構を捉えるためである。

本手法により吃音者と非吃音者とのフィードバック機構の差異を捉えられることを示した。さらに発声開始付近と持続発話中で発話音声とフィードバック音声との相関を比較することで、両群間でフィードバック機構が本質的に異なっている可能性を見出した。

非吃音者と比較して、吃音者の相互相関は0 msから最初に見られる極大・極小の差が小さくなっており、500 ms以内の速いフィードバック特性の時間方向の変化が小さいことが示され、吃音者では聴覚による発声周波数制御機構の働きが弱いことが考えられる。このことから、比較的速いフィードバック機構の差異が両群間で存在することを示唆している。これは、吃音者においてDAFに対する反応が非吃音者と異なることを説明する可能性がある。すなわち、非吃音者では200 ms遅れた発声を聴取すると、本来は聴取後200 msで補正されるF0が逆方向に振れることになり発話が安定しない。吃音者ではこのような速い制御が弱いためDAFにおいても人工吃が生じないと考えられる。

非吃音者では、500 ms以内の相関係数の極小・極

大間の差をみると、発話開始付近に比べ持続発声中でこの差が有意に大きかった。発声が続くこと（今回の実験からは発声開始より3-4s以降）で、よりダイナミックなフィードバック機構が働いていることが示された。これは、発声開始付近と発声持続中では発声に対する喉頭等の発話器官の安定性の面から、フィードバック特性が異なることが考えられる。すなわち、起声時には呼気流の変化が大きくなることや[11]、持続発声の母音であっても発声の開始や停止部分では波形の振幅や基本周期が乱れる[12]ために、発声開始付近では強いフィードバック制御がかけにくいのかもかもしれない。本手法では、発話開始の区間（F）の開始時刻は1回の発声毎に異なっており、発話が始まってから最高4秒後近くの発話もFに含まれる場合もある。それに関わらず発話開始の区間と持続発声中の相関係数に差異がみられた。このことから、発話と摂動信号を完全に同期させることが可能であれば、本研究による結果よりさらにFの区間の相関係数が低くなっていた可能性がある。

吃音者の発声開始付近と持続発話中の結果は、500ms以内に極大・極小が出現する者としめない者が存在し、一定の傾向をつかむことは困難であった。また、発話持続中に限った場合でも極大・極小のピークが出現する者の割合は非吃音者と異なる傾向にあった。このことから、吃音者のフィードバック機構の問題が一樣でないと考えられる。吃音者の発話の異常に関しては、種々の計測がすでになされている。構音器官[13]と喉頭[14]の両者において、吃音症状発生時に異常な運動が見られており、さらに、おのおのの異常に加え、これらの間のタイミングも異常とされている[13]。今回は喉頭の運動そのものではなく、その制御において、明らかな吃音症状の生じていない持続母音の発話においても、吃音者では非吃音者と異なるパターンがあることを、F0の計測によって定量的に明らかにしたと言える。

聴覚フィードバックのループには、聴覚と発話器官とそれらをつなぐ経路が含まれるため、これらのどの部分の異常がそれぞれどの程度関与しているのかは、今回の解析ではまだ明らかではない。Mysak[5]がフィードバックの問題はスピーチ回路機構の様々な部分で生じると考えたように、それらの多様性が吃音者の結果にばらつきをもたらした可能性があり、この方法は将来的には吃音の異なる病態生理を分離する一手法となる可能性がある。

5. まとめ

TAFを用い、聴覚フィードバックが発話に影響する効果をF0の相互周期相関にて求め、それが吃音者と非吃音者とで異なるパターンを示すことを見出した。また、非吃音者では発声の開始付近・持続中でピッチ制御が異なるものの、吃音者はばらつきが大きく一定した傾向を示さなかった。

実験に用いたMIDIデータならびに数々の技術資料を提供していただいた和歌山大学 河原英紀教授に深く感謝いたします。この研究の一部は、厚生科学研究「吃音の病態解明と医学的評価および検査法の確立のための研究（H14-こころ-001）」と「吃音の病態解明と検査法の確立および受療機会に関する研究（15130801）」の補助を受けて行われた。

文 献

- 1) Lane, H. L. and B. Tranel: The Lombard Sign and the Role of Hearing in Speech. *J Speech Hear Res.* 14(4), 677-709 (1971).
- 2) Lee, B. S.: Effects of Delayed Speech Feedback. *J Acoust Soc Am.* 22(6), 824-826 (1950).
- 3) Adamczyk, B.: Use of instruments for the production of artificial feedback in the treatment of stuttering. *Folia Phoniat.* 11, 216-218 (1959).
- 4) Cherry, C. and B. Sayers: Experiments upon the total inhibition of stammering by external control, and some clinical results. *J Psychosom Res.* 1(4), 233-246 (1956).
- 5) Mysak, E.: Servo Theory and stuttering. *J Speech Hear Disord.* 25(2), 188-195 (1960).
- 6) Ingham, R. J., R. J. Moglia, P. Frank, J. C. Ingham and A. K. Cordes: Experimental investigation of the effects of frequency-altered auditory feedback on the speech of adults who stutter. *J Speech Lang Hear Res.* 40(2), 361-372 (1997).
- 7) 河原英紀：変換聴覚フィードバックによる音声生成・知覚相互作用の検討。音響学会聴覚研究会資料。H-93-24, 152-158 (1993).
- 8) 平山和彦, 河原英紀：音声基本周波数の揺らぎに対する聴覚フィードバック条件の影響について。信学技報。H-94-52, 260-268 (1994)
- 9) Kawahara, H.: Effects of Natural Auditory Feedback on Fundamental Frequency Control. *Proc 3rd Int Conf on Spoken Language Processing.* PC-ICSLP. Yokohama, 1994-09, ICSLP, 1399-1402.

- The Secretariat of the Acoustical Society of Japan, Tokyo (1994).
- 10) 河原英紀：音声知覚・生成相互作用の伝達特性について. 音響学会聴覚研究会資料. H-95-35, 223-226 (1995).
 - 11) Koike, Y. and H. von Lenden: Pathologic Vocal Initiation. *Ann Otol*, 78(1), 138-148 (1969).
 - 12) 桐谷滋：声の音響分析. 声の検査法 (日本音声言語医学会編), 133, 医歯薬出版株式会社, 東京 (1994).
 - 13) Hutchinson, J. M. and K. L. Watkin: Jaw mechanics during release of the stuttering moment. *J Commun Disord*. 9(4), 269-279 (1976).
 - 14) Conture, E. G., G. N. McCall and D. Brewer: Laryngeal behavior during stuttering. *J Speech Hear Res*. 20(4), 661-668 (1977).