現象面から見た吃音とその認識
Stuttering, its Phenomena and Recognition

柳田益造（同志社大学工学部）
Masuzo YANAGIDA(Doshisha University)

要旨

言語障害のひとつである吃音について，その症状・要因・治療に関するこれまでの研究を概観し，遅延聴覚フィードバックによる吃音者と非吃音者の発話流暢性への効果の観測から吃音のモデルを仮説として提案している．また，吃音現象を音声認識の立場から分類してその対処法を単語認識と連続音声認識に分けて論じている．

Abstract

Phenomenological description about stuttering is given together with an outline of its supposed causes and clinical procedures. A model of stuttering is proposed based on differences in occurence frequencies of unfluencies between stutterers and normal speakers under Delayed Auditory Feedback conditions. Attempts for recognizing stuttered utterances of words and continuous speech are presented with some performance examples.

1　まえがき

計算機の高速化，音声認識技術の発達などによって，近い将来日常生活においても音声認識による機械操作や各種の自動販売機が普及することが見込まれる．それにより，別の作業をしながらでも音声によって家電機器を操作したり，電話を使って遠隔地から家の風呂の湯を沸かしたりなど便利な使い方ができるようになる．しかしながら，音声認識をする装置が吃音を含む音声をうまく認識できないとなると，人口の約１％と言われる吃音者にとっては融通のきかない機械以外の何物でもなくなる．認識してもらえるまで何度も繰り返し言い直さなければならなくなり，機械そのものが吃音者に精神的な苦痛を与えることになりかねない．そうならないためにも，吃音に限らず他の言語障害を持った人の音声も考慮した音声認識システムを開発する必要がある．

本報告では，２，３，４で吃音の症状・要因・治療に関するこれまでの研究を概観 [1, 2, 3, 4] した後，５で，吃音に対するモデルとして，遅延聴覚フィードバックによる吃音者と非吃音者の発話流暢性への効果の観測に基づいた「確認タイミング遅延モデル」 [5] を述べ，６では吃音の機械認識について，単語発声の場合と連続音声の場合について，それぞれに対応した手法を試みた例を述べる [6]．6.1 では，吃音を含んだ単語音声に対する認識として，ＤＰを吃音に対応させる手法を述べ，6.2 で吃音を含む連続音声から吃音と推定される部分を削除して通常の音声に近い音声にし，それを通常の認識手法で認識する手法について述べる．

2　吃音

2.1　吃音の症状

発話において，発声が詰まったり上ずったりすることを発話の「非流暢性」という．普通の人でもしばし観察されるが，この生起頻度が高く，慢性的なものが「吃音」と呼ばれる言語障害の一つの形態である．吃音が発症して間もない頃は，音節や音素の繰り返しや音が出ないブロックとかブレーク，あるいは音の引き延ばしなどの現象が見られる．これらは吃音の基本的な症状である．

吃音が治癒しないである程度経つと，吃音児（者）は自分が吃音であることを自覚するようになる．そして，吃音を発したことに対する埋め合わせのような言動をとったり，吃音から逃れるあるいは吃音を避ける方策を試み，成功したと感じた方法を学習していくようになる．これによって形成されるのが補償行動・逃避行動・回避行動であり，これらは吃音としては，いわば２次的な症状であると言える [7]．以下，吃音の症状を概説する．

1. 繰り返し
   同じ音を何度か繰り返す場合のことを指す．「と，と，とうもろこし」とか「s:, s:, s:, さくらんぼ」などが例としてあげられる．発話の最初の音に現れる場合が多く，文の途中でも単語の頭に現れることが多い．文節の途中，単語の途中に現れることは少ない．　１つの単語に対する「繰り返し」が，１，２回と回数が少なく，生起頻度の低い場合は「単なる突発的な非流暢性」と判断され，２回以上の繰り返しが見られ，生起頻度が高い場合が「吃音」とみ

なされる．日本語吃音における「繰り返し」の例をTable 1 に挙げておく．繰り返しの単位としては．音素，モーラ，音節などがあり得るが，音節以下であることが吃音によるくり返しの特徴である．
１音節未満でも３音素以上になることもある．例えば，英語では "str, str, str, structure"，中国語では "kaN, kaN, kaN, KaNtong" などがある．日本語では吃音の繰り返し単位がモーラを越えることは少ないが，撥音が前のモーラに付随する場合がある．

表 1: Repetition units.

	repetition unit	example
(a)	phoneme	s- s- saNka
(b)	mora(syllable)	sa- sa- saNka
(c)	over mora	saN- saN- saNka

2. 語頭のブロック・語中のブレイク
頭の中ではこれから言おうとする文章は決めているが，それが声としてなかなか出てこない現象．これが文頭あるいは語頭に生じる場合が「ブロック」，語中に生じるのが「ブレイク」である．
例えば，友人に「どこへ行くの」と尋ねられた返事として「これから○○を買いに行く」と言いたいときに，「こ」の音がなかなか出てこなくて，「・・・・・・・・これから○○を買いに行く」というように，文のはじめに無発声が長い間（ひどいときは，２秒以上）続く症状である．一方，「ブレイク」は言葉の途中で不自然な間延びが起こる場合を指す．促音の後に見られる場合が多い．例としては，「行ってきた」というはずが，「行っ・・・・・・・・てきた」のように「いっ」の後で急に言葉が出なくなってしまう場合などである．

3. 引き伸ばし
単語や文の最初の音（音素，あるいは音節中の母音部分）を伸ばして発音する症状．繰り返しと同じく文の頭もしくは単語の頭に出てくることが多い．
上述の「繰り返し」や「ブロック」を回避するために，故意に最初の音を伸ばして発音することもある．
成人吃音者の場合，ブロックよりも引き伸ばしの方が多い．（吃音者の内観として，繰り返しやブロックに陥るよりも，引き伸ばしで済むならそれにしようという意識がある．）
引き伸ばしの例として，「この寒い中」と言いたいときに，「こ」の音が言いにくいため，「こーの寒い中」のように伸ばしたりすることなどがある．「さー，さ，さ，さーーきに学校いくね．」の様に繰り返しと一緒になって現れることもある．

4. ポーズの短縮
吃音者の発声で，ある文の終了から次の文に移るまでの時間が明らかに短すぎる，と思える場合がある．一つの文の終了後，ほとんどポーズを置かずに次の文の発話を始めるのが「ポーズの短縮」である．吃音者の内観として，「早くしゃべらなければ」という強迫観念のようなものがあり，この症状もそれが影響しているらしい．また，そこまでの段階で発話が遅れてしまっているから，その遅れを取り戻せる機会があれば取り戻しておこうという意識の結果としての，一種の補償行動とも解釈できる．

5. 文終了後の余剰語付加
一つの文を言い終わった後で「・・・・・でした，っと．」や「・・・・・・です，ハイ」のように，余分な言葉を付加する現象がよく観察される．この原因としては，吃音者の内観として，次の４点が挙げられている．

(a) 以後の発話がスムーズになるように，意識的に余剰語をつけて，発話のリズムを整える．
(b) かつて，その言葉を文末に付けることによって，うまく話をすることができた経験（記憶）から，そのような余剰語付加を意識的に繰り返しているうちに，習慣になった．
(c) どんなに途中で非流暢性があっても，最後の言葉がちゃんと言えていれば，吃音者自身として多少なりとも安心感が得られる．
(d) 一つの文の発声を終了できたということを自分で確認し，納得することによって自然発声的に発する．

以上より，これも一種の補償行動といえる．

6. 身体的な随伴行動
これは，吃音発声時，つまり発声しようとしているのに声が出ないというような場面で，早くその状況から逃れようとする心理的状況が身体的な行動として現れる場合である．舌打ち，うなづき，視線を背ける，口元の歪み，顔の引きつり，唇の震え，唇をギュッと閉じる，手足を動かす等がある．これは心理的な困難から早く逃れようとする一種の逃避行動である．

7. 代替表現の検索
吃音者の特徴の一つに，経験的に発音の苦手な単語

（音）がわかっているため，それを避けて別の言い方を探そうとする傾向がある．本項では，そのことについて文献 [5] の第一著者自身の内省に基づいて説明する．

吃音者は，苦手な音の発声を避けるために，これから自分が発話するべき文からそのような音を語頭（まれに語中の場合もある）にもつ単語の使用を，可能ならば避けようとする．つまり，発話するつもりでとりあえず生成した文の中に言いにくいと予測される部分（単語単位であることが多い）があれば，それを別の言い方で代替表現できないかを探す傾向にある．

例１：「あの店で買えば，ポスターがついてくる」の「ポ」の音を発音するのに不安を感じてしまった場合に，別の言い方でほぼ同じ意味になる「あの店で買えば，おまけ特典がついてくる」と置き換える．

例２：「出身地はどこですか」と聞かれたときに「名古屋です」の「な」の発声に不安を感じ，「愛知です」と，多少おかしな言い回しではあるが，言い換えてしまうことがある．

吃音者はこのように，発話内容を自分で変える自由がある場合には，常に言いやすい言葉を探しながら会話をする傾向がある．これに対して，原稿を与えられ，どうしてもその通りに読まなければならない場合，苦手な音の発声をしなければならない事態を事前に察知しても，それを回避することができない．ある音に対して特に苦手意識を持っている吃音者は，朗読になると，普段の会話のときよりも吃音頻度が増すことがある．また一方で，朗読の方が別の言い回しを考えることなく，棒読み状態でいいために読むことに専念でき，吃音頻度が減る吃音者もある．この状況は，吃音の程度や症状によって個人差が大きい．

2.2 吃音の「進展」と「軌道」[2]

吃音はどの言語にも見られ，言語によらず発症率は約５％で，男女比はほぼ３：１と明確な差がある．発症者の約５０％が６歳までに治り，成人期に達するまでに約８０％が自然治癒 (spontaneous recovery) するという統計があるが，残りの約２０％すなわち全体の１％程度の人が吃音を成人期まで持ち越す．成人期まで持ち越すと，一般に治癒は困難であるといわれている．

非吃音者でも発話を繰り返したり，部分的に引っかかって喋る場合があることから，吃音は流暢性の程度の差であるという考え方（連続体仮説：continuity hypothesis by O. Bloodstein [8]）がある．一方，専門家でなくても吃音と非吃音の区別は容易に可能で，吃音が治癒しても特徴的な非流暢性が残ることから，吃音と非吃音は質的な差があるとする考え方（M. E. Wingate [9]）もある．

2.2.1 吃音の「進展」

Bloodstein は 2 〜 16 歳の吃音児 418 名の横断的な観察から，吃音の進展を次の 4 段階に分けた [10]．不思議なことに，この「進展 (development)」には吃音の改善あるいは治癒・消滅の概念が入っていない．

1. 第 1 段階
   通常は正常な発話であるが，場合によって（興奮したり狼狽したりした場合など）に吃る．主たる症状は繰り返しで，ほとんどが文章，文節の最初の音に限られるが，ブレイクは機能語でも起こる．吃音者としての自覚を持たない．
2. 第 2 段階
   症状が慢性的になり，吃音者としての自覚を持つが，自分のスピーチがスムーズでないことをほとんど気にしていない．
3. 第 3 段階
   特定の場面（教室での朗読，店での注文，電話など）に対する反応としての側面を備えてくる．本人には特定の単語や語音がその他のものよりも喋りにくいように思われてくる．
4. 第 4 段階
   吃音に対する予期不安が明瞭に認められる．ことば，語音，場面に対する恐怖がある．語の置き換えや回りくどい言い方がしばしば見られる．話す場面の回避や，恐怖・困惑が感じられる．

2.2.2 吃音の「軌道」

Van Riper [11] は，それまで吃音の進展が上のように不可逆的に捉えられていたことを反省し，44 例の縦断調査（発症から成人期まで）から，吃音をその進展の様態によって次の 4 つの「軌道 (tracks)」に分類した．

1. 軌道１　21 例 /44 例中.
   ２歳６月〜４歳２月で発症．言語発達は正常．よくなったり悪くなったりの「振れ」が大．自然治癒率大．進展すると，いわゆる「吃音の悪循環」（吃音症状→自覚→もがき行動→恐怖→回避）に陥る．
2. 軌道２　11 例 /44 例中.
   早期の言語発達段階から流暢性が欠如．言語発達に遅滞．３〜６歳に発話開始とともに発症．くり返しが主で引き伸ばしやブロックは殆どない．随伴行動なし．進展しても深刻化しない．吃音意識なし．
3. 軌道３　5 例 /44 例中.
   希な症状．発吃が遅い．５〜９歳で突然発症．随伴行動に進展が速く，病的傾向が強い．吃音の悪循環に早くはまりこみ，重い吃音に陥る．

4. 軌道4　4例/44例中.
　4歳以降に突然発吃. 吃音を意識しているが, おおっぴらに吃る. 症状は殆ど変化しない.

3 原因

まだ解明されていないが, 次のような説がある [4].

1. 大脳半球間葛藤説 (S. Orton, 1929)
　左利きを右利きに矯正された吃音者が多かったことから, 矯正による半球間の葛藤が原因とする説. 反例が多いが, 再び左手を使うことによって, 本来の優位半球を働かせるようにして吃音を治療するという考えを生んだ.

2. 失敗経験に基づく葛藤説 (J. Sheehan, 1958)
　話すことに失敗した経験を繰り返したくないために話すことを避けるという説. 具体性に乏しい.

3. 診断起因説 (W. Johnson, 1961)
　幼児期の正常な非流暢性を周囲 (特に母親) によって吃音と診断されたことが契機となって本当に吃音になるとする説. 周囲が神経過敏になるので, 子供自身も自分の話し方を意識し, スムーズに話そうと過度に神経質になり, 逆に吃音になるという説であり, 現在では広く認められている. これは, 学習過程において一般に見られるUカーヴの底の部分に相当する時期に学習環境が変わり, その後の学習の進展に異常を来した結果と考えられる.

4. タイミング障害説 (C. van Riper, 1982)
　発声のための筋肉の連続的な連携動作のタイミング制御がうまく行かないとする説.

5. 大脳の不適切局在化説 (R. D. Kent, 1983)
　大脳における言語機能の局在化が不適切であるために, 言語の理解・生成に要する処理の速度制御の調整が十分に機能していないとする説. タイミング障害説の一つに分類できる.

6. 要求能力不一致説 (G. Andrews, 1983)
　神経系の発達の過程で音声的に正確に話すことを要求されるようになり, まだ十分発達していない段階で過度な要求を出され, 神経系が処理しきれなくなったことが原因とする説.

4 成人吃音の治療法 [1]

吃音に関しては原因が曖昧なまま種々の治療法が試みられている. 成人に対する治療法について述べる.

4.1 行動治療法 (流暢性形成治療法)

行動の変容 (吃音症状の軽減・消滅) を目的とする. 吃音症状が軽減あるいは消滅しても, 原因が解消したという保証はない. 吃音者のしゃべり方が「正常でない」ので, それを「矯正する」という考えが根底にある. 吃音に対する不安や恐怖を消すこととか, 回避行動を減らすなどの心理面は考慮しない. 一部の治療法には, それによる人工的な影響が生じるので, 批判もある.

例1：メトロノーム法
メトロノームによって, 一定リズムでのしゃべり方を獲得させる. 吃音は軽減するが, 感情を乗せられない速度一定のしゃべり方になるという人工的な影響が生じるといわれている. 吃音者は一般にリズムに乗るのが苦手なようで, 場合によると, メトロノームに合わせて喋るということは, 普通の状態よりも喋りにくくなるという印象を持つ吃音者もある.

例2：DAF (Delayed Auditory Feedback) 法 [13]
遅延をかけた自分の声を聞きながら喋ることにより, 発話速度は落ちるが吃音は軽減する. 装着中は, 本人も喋り易く感じ, 客観的にも吃音の生起頻度が減少するが, ゆっくりしたしゃべり方になるという人工的な影響とともに, 機械に頼る習性がつくという弊害がある. ゆっくりした発話になるという悪影響を除くとともに, 機械への依存心を最終的には排除するために, DAF法で遅延時間を徐々に短くするように練習し, 遅延時間が0になってもしゃべれるようにするという方法も提案されている [4].

例3：斉読法 (コーラス・リーディング)
集団で一斉に声をそろえて音読する方法である. 初めは大きい集団での斉読からはじめて, 徐々に小さな集団に移していく. 最終的に治療者と二人になっても吃音が抑制されるところまできたら, 今度は治療者の声を小さくしていく. 治療者の声がハミングする程度にまで小さくなったときでも, まだ吃音が抑制されているなら, 今度は斉読のとき使ったテキストとは別のテキストを読む. この段階でも吃音の生起が抑制されているようであれば, 治療者がテキストを読むのをときどき休止する. そして, この休止時間を徐々に長くしていく. 最終的には一人で音読した場合でも, 吃音の生起が抑制されるようになる.

例4：シャドーイング法 [14]
治療者は吃音児 (者) に文章を見せないで, 文節と文節の間を空けることなく読み続ける. 吃音児 (者) はその音声にできるだけ接近して, 後ろを追うように追唱する方法である. 吃音児 (者) は常に治療者の声に聴覚的集中を保持しながら発語しなければならないため, 自己音声に対する聴覚的フィードバックが遮蔽され, 発語に対する不安が低減されるので, これらの効果を利用して吃音の生起を抑制する方法である.

例5：マスキング法 [14]
ヘッドホンを使用して両耳に高い音圧の白色雑音の低音域 (50 ～ 500Hz) を聞かせ, その間に音読させて自己音声の聴覚フィードバックを遮蔽することによって, 自己音声に対する干渉を不可能にして, 自己音声に対する過度な注意の傾注を抑制

し，発語に対する不安緊張の生起を抑制することによって，流暢に発話ができるようになることを期待するものである．

4.2 心理療法（吃音軽減治療法）

吃音に対する恐怖をなくし，吃音からの回避をなくすことを主たる目標とする．これは，Johnson の「吃音は吃音者が自分のことばに対して持っているある特殊な認知と評価の仕方である」という考え方に基づく．臨床的には，Van Riper が提唱した「吃音の認知→脱自意識化→吃音の軽減→安定化」の 4 段階で構成する方法による．基本的には，「吃音を軽い楽なものにする」ということと，「吃音に対する恐れやその恐れのために生じる回避的な行動を減らす」という 2 点に尽きる．

成人期まで持ち越した自分の吃音を完全に治癒させることは難しいという現実を認識できるので，成人吃音者は一般に自分の意見を言ったり権利を主張する意欲に欠け，「逃げの生活」が身についてしまうことが多い．日本における成人吃音者の自助グループ「言友会」は，吃音症状そのものの改善への努力には否定的で，「吃音を持ちながら前向きの生活をする」ことを目標に掲げている．

5 吃音のモデル

吃音のモデルを考えるということは，吃音の原因を考えることとほとんど等価である．したがって，吃音のモデルは吃音に対する原因の説と同じ数だけあると考えられる．ただし，モデルといっても漠然としたものはその妥当性の確認の方法がなく，実験などによって妥当性が確認できるモデルはそれほど多くない．例えば，Riper によるタイミング障害説 [12] を取り上げてみると，これは発声のための筋肉の連続的な連携動作のタイミング制御がうまく行かないとする説であるが，これを確かめるためにどのような実験をすればよいかは必ずしも明確でなく，まして治療と結びつくものではない．その理由は，調音器官に対する制御指令の内容あるいはその生成機構，学習機構などが明らかになっていないからである．

本稿では，ＤＡＦを使った実験によって妥当性の確認ができ，かつ治療法にも結びつけられる「確認タイミング遅延モデル」について説明する．

5.1 DAF 効果から考えられる「確認タイミング遅延モデル」

発想

自分の声を 0.1 〜 0.2 秒ほど遅らせて自分の耳に戻すと，話し方がぎこちなくなる．これは遅延聴覚フィードバック (Delayed Auditory Feedback, DAF) による人工吃音 (articifial stuttering) と呼ばれる．DAF は非吃音者に対しては発話の事後確認作業を混乱させると考えられる．ところが，吃音者に DAF を適用すると，遅延時間 200 〜 250ms 程度では，かえって発話がしやすくなるという，吃音者自身の印象とともに客観的にも非流暢性が減少することが知られている．このことから，「吃音者は，非吃音者に比べて自分が発したスピーチを確認するタイミングが遅い．」というモデルが考えられる．

普通にフィードバックされる経路による音（骨導音および口から耳への気導音）では，自分が発声した音声は 0.2 〜 0.3ms 程度で返ってくる．非吃音者は，この程度の遅れで返ってくる音を聞くことによって自分の発声結果をうまく確認しているものと考えられる．しかし，吃音者はこの確認のタイミングが 200 〜 250ms 程度後になっているのではないかと考えるのがこのモデルである．

根拠

この根拠としては，次のことが考えられる．自分が吃音者であることを意識している吃音者は言いにくい言葉（単語）を避けるため，常に代替表現を探索しながら喋っている．代替表現を検索する時に，言語中枢であるウェルニッケ領域が活動することになるが，ウェルニッケ領域はまた次に発話すべき文を生成するための活動も行っている．発話運動指令そのものの生成はブローカ領域で行われるが，発声すべき文を生成する作業はウェルニッケ領域で行われる．一般に意識的な作業は逐次的に行われるため，同時に複数は実行できないと考えられており，ウェルニッケ領域での言語処理は逐次的に行われているはずである．従って，次の発話の中に現れる言いにくい単語を避けるために代替表現を探す分だけ，吃音者は非吃音者に比べて次に発話すべき文の生成に，余分な時間をかけなければならないことになる．このための時間を，発話中の可能なところで確保する必要がある．これが，文の境界あるいは文節の境界になるものと考えられる．このために，吃音者は非吃音者が普通に行っているような「骨導，気導によるフィードバック」タイミングで自分の発声を発声直後に確認するという体勢を常時とることができない．吃音者にとっては骨導，気導によるフィードバックのタイミングよりも，ＤＡＦにより遅れてやってくるフィードバックの方が自分の声を確認しやすいのではないかと考えるのである．

演繹

このように考えると，遅延時間を 200 〜 250ms 程度にすると吃音頻度が減るという現象が，数値はともかくとして，概念的には納得できる．ただし，その考え方を推し進めると，吃音者に対してこの最適遅延時間よりもずっと長い遅延時間のフィードバックをかけると，非吃音者

に対して遅延フィードバックをかけたときのような現象が現れるものと考えるのが自然である．

検証方法

したがって，もしこの確認タイミングモデルが正しいなら，吃音者にとっての最適遅延時間（吃音の生起が最も減少するる遅延時間）が，ちょうど非吃音者にとっての遅延なし（あったとしても 0.2 〜 0.3ms 程度）の状態に相当し，最適遅延時間を超えると非吃音者と同じような現象を呈するようになるはずである．

このように考えると，このモデルの正当性は，吃音者と非吃音者の非流暢性の生起頻度を遅延時間の関数として観測し，吃音者の吃音産出頻度の遅延時間特性が非吃音者の遅延時間特性を最適遅延時間だけ時間軸方向に平行移動したような形になっているかどうかによって確かめられる（そうなっていなければモデルが間違っていることになる）．

5.2 非流暢性の遅延時間依存性

吃音者と非吃音者を被験者とし，DAF をかけた場合のスピーチの単位モーラ数あたりの非流暢性の生起頻度を測定した．フィードバックは両耳へとし，そのレベルは各被験者にとって（ヘッドホンを通して遅延をかけて返される音が気導・骨導による通常のフィードバックよりも）十分に大きく聞こえるようにした．遅延時間は，50 〜 500ms は 50ms きざみ，500 〜 1000ms は 100ms きざみでデータを取った．それぞれの遅延時間について，被験者にリラックス状態での発話（普段からよく話をする友人との研究室内での対話）を 1 分間を収録し，非流暢性の生起頻度を調べて 100 モーラあたりの値として正規化を行う．被験者は，吃音者 1 名および非吃音者 3 名で，すべて男子 21 〜 22 歳である．

DAF をかけた場合の 100 モーラあたりの非流暢発話の生起頻度を，遅延時間を横軸として図 1に示す．上が吃音者，下が非吃音者の結果である．吃音者については遅延時間 500ms までは 8 収録，それ以上は 6 収録の，また非吃音者については 3 名の各 1 収録の平均と標準偏差である．

5.3 確認タイミング遅延モデルの妥当性

これらを見ると，非吃音者では，遅延時間が 50ms までは非流暢発話の生起頻度が低いが，150 〜 250ms で最も高くなり，以後は 400ms 辺りまで徐々に減少し，そこから再び増加している．それに対して吃音者については，遅延時間 0 では非流暢発話の生起頻度が高く，遅延時間 50 〜 150ms で低くなり，その後 350ms 程度までの間に徐々に増加し，700ms 辺りで一旦減少して再び増加している．これらの遅延時間特性を比較すると，非吃音者の 50ms 点，150ms 点，400ms 点がそれぞれ吃音者の 150ms 点，350ms 点，700ms 点辺りに相当すると考えることができる．

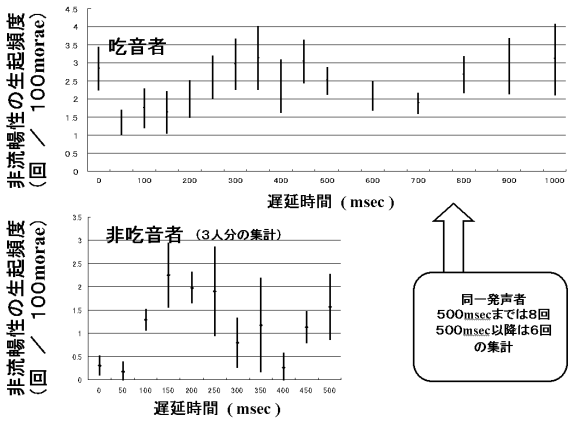

図 1: 非流暢生起頻度の遅延時間依存性

ここで，非吃音者にとっての 400ms という時点は，DAF による影響が少なくなるということから，発話の確認作業から解放される時点であろうと考えると，非吃音者が発話後だいたい 50 〜 150ms あたりから始めて 400ms くらいまでで済ませる確認作業を，吃音者は 150 〜 350ms あたりから 700ms くらいまでかかっていると解釈できる．つまり，吃音者の確認作業は非吃音者のそれより約 100 〜 200ms 程度遅れて始まり，確認作業に要する時間が非吃音者では 400-50=350ms 程度であろうと推測されるのに対して，吃音者ではそれに相当する時間が 700-150=550ms となり，時間がかなり余分にかかっている．これは，発話の済んだ部分に関する確認作業と将来発話すべき文章の中の苦手な音の部分に関する代替表現の検索という余分な作業が入り込むからであろうと思われる．現状ではデータのバランスが悪いので，吃音者と非吃音者の遅延時間軸上の対応点を正確に求めることができないが，今後収録データを充実させてこれを推定する予定である．

6 吃音の認識

吃音者にとって，音声認識装置は「自分の声は吃音なので認識してくれないのではないか」という疑念・不安が先立ち，吃らずに 1 回でうまく発声しないといけないという強迫観念に襲われる恐ろしいシロモノである．しかし，これからの音声認識装置は吃音者にそのような不安や恐怖をいだかせるものであってはならない．筆者らは，吃音の認識について，単語認識には語頭音節を重複させた逆ＤＰによって，また連続音声にはくり返し部分

を削除することによって，吃音にある程度対処できる方法を提案し検証した [6]．以下，その概要を述べる．

6.1 単語認識のための吃音処理

6.1.1 基本的考え方

少数語彙の単語認識にはＤＰを使うのが常套手段であるが，吃音音声に対してＤＰをそのまま適用することはできない．吃音の基本的な症状として「繰り返し」「ブロック・ブレイク」「引き伸ばし」があることを 2.1で述べたが，単語を一つだけ発声する際には，語頭の第１モーラに「繰り返し」が起こる場合がほとんどである．また，ブロック・ブレイクは始点が遅れるだけであるので，通常のＤＰによって対処できる．そこで，語頭のモーラの「繰り返し」への対処に重点を置き，語末から語頭に向かって，時間軸上を逆向きにＤＰを実行させる方法が考えられる．これによって，標準パターンの語末から語頭までの整合がとれれば，入力側の語頭より前に余分な音があっても無視することによって，語頭の繰り返しに対応できることになる．ただし，この方法では認識対象の語に接頭語を付けた語も認識対象に含まれる場合（例：大阪と新大阪）には問題が生じる．

6.1.2 語頭音節重畳逆ＤＰ

逆向きＤＰを実行させるだけでは，繰り返し部分を積極的に使ったことにならず，吃音による繰り返し部分の音響的な整合度がそれほど上がらないので，第１モーラに「繰り返しが何回か起きる可能性が高い」ことを積極的に使い，さらに上述の接頭語の問題もクリアーするために，標準パターンの第１モーラを数段重ねたものを改めて標準パターンとして使う．これを「語頭音節重畳逆ＤＰ（Augmented Backward DP, AB-DP）」と呼んでいる（図 2）．標準パターンの第１モーラを積み重ねた部分では，入力時間方向のコストを低く設定した規則を使ってマッチングを行う．これによって語中のブレイクや特定音素の「引き伸ばし」にも対処できるようになる．

6.1.3 動作の評価

前項で述べた AB-DP の動作を従来の端点制約／フリー前向き DP と比較した．AB-DP では，第１モーラを重ねる回数は 6 回とし，入力音声方向の始端および終端は，認識に使う範囲のうちの最初20 %と最後20 %に制約し，DP 処理の終端は，標準パターンについては第１モーラを重ねた部分の各先頭ならどこでも構わないとした．対象語彙は近鉄の駅名30 で，標準パターンは非吃音者の発声１回である．サンプリングレートは

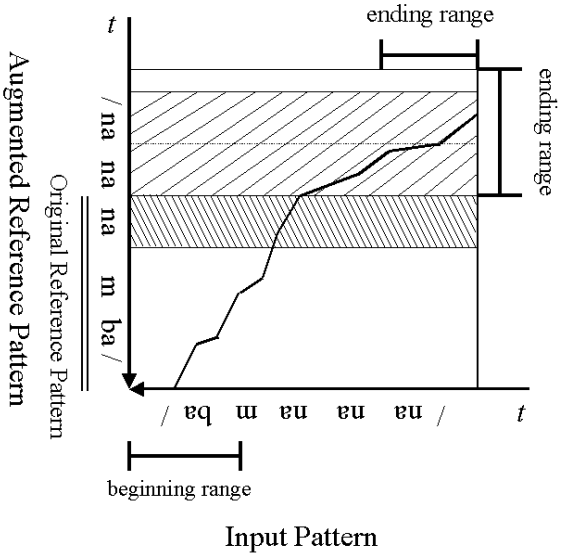

図 2: 語頭音節重畳逆ＤＰ (AB-DP)

16ksamples/sec，距離尺度には 16 次の LPC ケプストラム距離を使い，テストデータは，吃音者１名による「繰り返し型の吃音になった音声」と「繰り返しにはならなかった発声」，標準パターンとは別の非吃音者３人の「普通の音声」を各話者について 100 個ずつ用意し，各方法による認識率を調べた．結果を表 2に示す．

ここでは DP そのものの性能向上は特に図っておらず，標準パターンは非吃音者一人の音声だけなので，非吃音者に対する認識率でさえも最高で 85 %と，それほど良くない．それにも拘わらず，吃音音声に対しては端点フリーにしても 11 %であったところが AB-DP では 58 %と認識率が向上しており，AB-DP の効果が顕著であることを示している．

認識対象語彙の中には，語頭が「新」で始まる駅名が４，その内「新おお」で始まるものが２，「西」で始まるものが３，「新」の有／無の駅名が２対あり，語彙数が 30 ではあるが，単語認識としては難しい設定になっている．非吃音者に対する認識率で端点制約型より端点フリーの方が悪いが，これは端点の探し方がまずかったものと考えられる．なお，標準パターンをマルチテンプレート化するなどの常套手段によって，全体の認識率を向上させることができるものと期待できる．

6.2 連続音声認識のための吃音処理

6.2.1 重複部削除法

連続音声中での吃音は，音の繰り返しとブロック・ブレイク，引き延ばしであるが，時間軸方向の整合で問題

表 2: 従来のDPとAB-DPの動作比較
対象語彙: 近鉄の駅名30
標準パターン＝非吃音者1回
テスト音声=吃音者による吃音，非吃音各100発声

	Conventional DP		Proposed DP	
	Limited Ends	Ends Free	Backward DP	AB-DP
stuttered utterance	0	11	24	58
unstuttered utterance by a stutterer	55	50	44	72
normal utterance	74	56	59	85

になるのは繰り返しである．たとえば，「ホテルのよよ予約をお願いします」など，特に自分が発声を不得意とする音が次に発声しようとする単語の先頭音であるとき(この場合は”よ”)，同じ音の繰り返しが起こる．ここで述べる吃音処理は，隣接する有音区間についてLPCケプストラム距離を用いて局所的なＤＰマッチングを行い，距離が一定値より近ければ，先行部分を吃音による繰り返しとみなして削除するものである．
実際の削除方法は，有音区間を抽出しておいて，DPを用いてその中で同じようなパターンが隣接して続いている部分を吃音部分として検出することになる．同じようなパターンが隣接している状況としては，前後に無音区間がある「独立型」，繰り返しと関係のない部分が前あるいは後ろに続く「前／後方連続型」，前後に続く「前後連続型」などがあり得るので，それぞれの可能性を調べて繰り返し部分を検出する．検出されたら，前の区間を吃音による不要部分として順次削除する．この方法を「重複部削除法」と呼ぶ．この方法により，吃音の約８０％を占める繰り返し型吃音の中の約６０％を削除することができた．繰り返し部分がうまく削除できれば，以後は通常の音声認識で処理できることになる．

6.2.2 動作の評価

重複部削除法によって吃音部分が削除された音声をATRの音声認識システムSPRECで認識させてみた例を述べる．音響モデルには不特定話者モデルを用い，言語モデルには品詞クラスbigramを初期状態として状態数400にまで分離したN-gram [15]を用いた．音響モデルの学習データには吃音者の音声は含まれていない．また言語モデルの学習に用いたデータはATRの自然発話旅行会話データベースで，166対話2519語から構成される．
例として「はい，わわわわかりました．」と発話した吃音音声を入力した場合を説明する．吃音入力を１とすると，それには「わかりました」の前に「わ」の繰り返しがいくつか含まれている．それから吃音の繰り返しを削除したものを２とし，１，２のそれぞれをAPRECに認識させたところ，次のようになった．
1. はい / 気 / は / バッグ / な / の / を / う / が / あ / り ＋まし / た
2. はい / わか＋り＋ま＋した

7 むすび

吃音の症状，原因，治療を概観し，DAFを用いた実験から「吃音者は自分のスピーチの確認をするタイミングが非吃音者よりも遅れている」という「確認タイミング遅延モデル」を述べた．また，吃音音声に対する単語認識として「語頭音節重畳逆ＤＰ」，連続音声認識の対処法として「重複部削除法」を紹介した．

本研究の一部は，同志社大学学術フロンティアプロジェクト「知能情報科学とその応用」の援助を受けた．

参考文献

[1] 内須川洸：「吃音診断学序説」，風間書房，1986.
[2] 水町俊郎：「吃音の問題を構成している心理学的要因に関する研究」，風間書房，1991.
[3] 飯田京子，若菜陽子，長崎勤：「吃音の診断と指導」，学苑社，00.196-198, 1990.
[4] 笹沼澄子監修，伊藤元信編：「成人のコミュニケーション障害」，大修館書店，第４章，1998.
[5] 成川，柳田:「吃音のモデルとその実験的検証」，音声研資，SP98-78, 1998.
[6] 奥村，成川，渡辺，柳田：「音声認識における吃音処理に関する検討」，音声研資，SP99-135, 2000.
[7] C. S. Bluemel:”Primary and secondary stammering”, Quart. J. Speech, Vol.18, pp.187-200, 1932.
[8] O. Bloodstein:”Stuttering and normal nonfluency - A continuity hypothesis-”, British J. of Disorders of Communication, Vol.5, pp.30-39, 1970.
[9] M. E. Wingate:”Stuttering: theory and treatment”, Irvington, NY, 1976.
[10] O. Bloodstein: ”The development of stuttering: III. Theoretical and clinical implications”, J. of Speech and Hearing Disorders, Vol.26, pp.67-82, 1962.
[11] C. van Riper:”The nature of stuttering”, Prentice-Hall, Englewood Cliffs, NJ, 1971.
[12] C. van Riper:”The nature of stuttering. 2nd ed.”, Prentice-Hall, Englewood Cliffs, NJ, 1982.
[13] W. H. Perkins: ”Replacement of stuttering with normal speech”, J. of Speech and Hearing Disorders, Vol.38, pp.283-303, 1973.
[14] C.Cherry and B.McA.Sayers: ”Experiments upon the total inhibition of stammering by external control and some clinical results”, J. Psychosom. Res., 1, pp.233-246. 1956.
[15] 政瀧浩和，松永昭一，匂坂芳典：A続音声認識のための可変長連鎖統計言語モデルC信学技報，SP95-73，1995.