ヒューマンインタフェース 50−10
（１９９３．９．１７）

リアルタイム話速変換を用いた対話補助方式の提案

禰寝 義人　在塚 俊之

日立製作所 中央研究所
〒185 東京都国分寺市東恋ヶ窪 1-280

あらまし　高齢者／難聴者における聞こえの衰えを補助することを目的として開発した、話速を変換して「ゆっくり」と音声を聞くためのリアルタイム音声信号処理を、「対話」において使用するためのヒューマンインタフェースを提案する。本方式は、ＤＳＰ１個を用いたポータブルシステムを携帯し、装置上に設けたボタン操作で話速変換の制御を行なう方式である。本方式により、従来ラジオ音声などの「一方的に与えられる音声」を対象としていたシステムを、対話音声にも利用することが可能になる。模擬システムによる予備評価では、発話タイミングを左右するような使い勝手の上でのいくつかの問題点が抽出された。

和文キーワード　高齢者、難聴者、話速変換、対話、ヒューマンインタフェース

An aid for conversation with real-time speech-rate conversion.

Yoshito Nejime　Toshiyuki Aritsuka

Hitachi Ltd., Central Research Laboratory
Kokubunji, Tokyo 185, Japan

Abstract　A human-interfacing technique to use a real-time speech-rate converter as an aid for interactive conversation is proposed. The speech-rate conversion was originally designed to aid hearing-impaired elders by reducing speaking rate of non-conversational speech. A hand-held speech processor with control buttons realized a way to aid user's listening in spoken-dialogue. Some difficulties which affect to speaking timing has been pointed out by preliminary evaluation using a quasi-system.

英文 Key Words　elders, hearing-impaired, speech-rate conversion, spoken-dialogue, human-interface

1.はじめに

一般に高齢者における難聴の場合、聞こえの劣化は単に聴覚末梢系での聴力損失のみならず、上位中枢における短期記憶能力の衰えや、言語処理能力の衰えにも強く依存していると考えられる。これに対して最近、信号処理を用いて音声の話速を遅くすることで、聞こえの衰えを補助しようとする試みが行なわれている[1]。

一方、筆者らは先に、高齢者や難聴者向けの音声加工方式を検討するのためのツールとして、DSP(ディジタルシグナルプロセッサ)1個を用いたポータブルシステムを開発した[2]。このシステムにおいては、リアルタイム処理による音声の周波数特性加工に加え、高齢者の聞こえを補助することを目的として、メモリ録音を利用して直前に聞いた音声を瞬時に聞き直す方式や、話速を変化させる方式を実現した。

さらに信号処理方式の改良を行ない、音声をピッチ単位で伸長する手法を用いた話速変換方式を開発し、あたかも「ゆっくり」話しているように音声を加工することで、聞こえの衰えを補助する方式を提案した[3]。難聴者による評価検査の結果、時間分解能の低下した被験者では話速変換された音声の評価が高くなることを確認した[3]。

これまでの報告では「連続して一方的に与えられる音声」を入力として想定し、実時間からのズレを生じさせながら次々と音声を遅くして、聞こえの衰えを補助する方法を提案してきた。このような利用方法としては、ラジオ等の放送音声やテープレコーダ等に録音された音声を本システムで加工し「ゆっくり」と聞く方法が考えられる。

しかし、従来の補聴器が入力音声の種類に関係なく使用できることを考えると、本システムも日常のあらゆる音声に利用できることが望ましい。特に「対話」において本システムを「通訳装置」のように利用することができれば、難聴者の聞こえを補助する場合のみならず、健聴者が外国語を聞き取る場合にも、本システムを使用できる可能性がある。

また、本システムでは話速変換処理をリアルタイムで行なうので、予め録音された音声ではなく、聞こえた音声をその場ですぐに加工するような場合に利用して、初めてそのリアルタイム性を生かすことができる。

そこで今回は、本システムのリアルタイム性およびポータブル性を生かし、「対話」において話速変換を利用することを目的として従来システムの改良を行なった。本報告では、リアルタイム話速変換を対話に使用するためのインタフェースの提案と、その予備的な評価結果について述べる。

2.話速変換システムの構成

リアルタイムで話速変換処理を実現するため信号処理方式および、それを実行するためのハードウエアの構成については既に報告した[2][3][4]。ここではその概要および従来からの変更点を述べる。

（1）話速変換処理の概要

本システムの話速変換処理方式は、入力音声を48msecの時間長を持つ2つのフレームバッファを交互に用いてフレーム毎に時間領域で波形加工し、バッファメモリに蓄積しながら出力する方式である。

一般に音声波形を加工する場合には歪みが伴い、これによる明瞭度低下が問題となる。このためできるだけ波形加工を施さずに話速の低下を感じさせることが望ましい。そこで本方式では、図1に示したようにフレーム毎のパワーがしきい値Thより大きい場合にのみ、そのフレーム内のデータに対して図2に示す波形伸長処理を施すようにした。これにより、子音明瞭度を左右する音声の立ち上がり／下がりの部分が閾値処理により加工されず、原音の情報が保存されるようになっている。

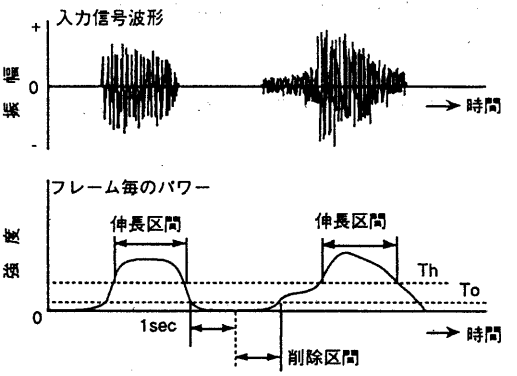

図1　閾値処理による伸長および削除区間の決定

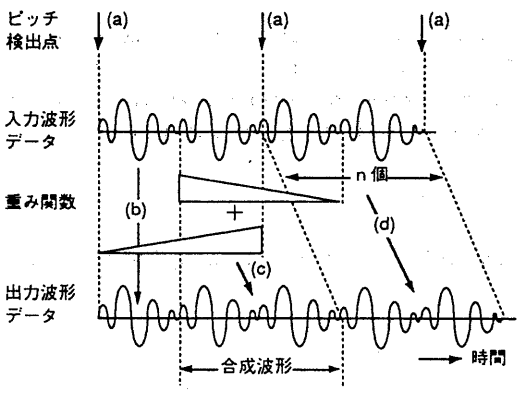

図2　波形伸長処理

図２の波形伸長処理は、Malar等[5]の波形伸長アルゴリズムをベースに開発した。ここでは、(a)自己相関法によるピッチ検出、(b)１ピッチ分の波形データ転送、(c)０から１に変化する２ピッチ幅の三角形重み関数を用いた波形合成、(d)ｎピッチ分の原波形データ転送、の４つの手順を繰り返し行う。処理(d)のｎを0,1,2とすることで、各々３種の伸長率ｅ＝1.50(２→３ピッチ)、1.33(３→４ピッチ)、1.25(４→５ピッチ)が得られる。本方式によれば、いかなる点からでもクリック音を生ずることなく合成波形を挿入することができ、また窓関数の幅が伸長率によって変わらないため、合成波形部分の音質劣化が伸長率に依存しない。

また(a)のピッチ検出は、(b)および(d)の各々の処理の前で行ないその時間間隔は高々２ピッチであるため、細かいピッチの変動にも十分追従することが可能である。なお今回は抽出ピッチの精度向上を目的として、音声データに低域フィルタリングと間引き処理を行なった結果に対して、自己相関処理を行なうように改良を施した。

一方、伸長された出力波形データはリングバッファメモリに記憶されるが、これには実時間との遅れを吸収できるだけの容量が必要である。本方式では、このメモリ容量の削減を目的として、池沢ら[6]の主観評価実験の結果に基づき、図１に示したように閾値Toによって無音区間を検出し、１秒以上の無音区間部分を削除しながらメモリへの書き込みを行なっている。

さらに本システムでは、話速変換後のデータを出力する際に、４バンドの帯域別振幅加工（マルチチャネルコンプレッション）を用いた周波数特性加工も施しており[2]、時間特性と周波数特性の両方を同時にリアルタイムで加工するシステムとなっている。

（２）しきい値の自動更新機能の追加

上記の話速変換処理においては、しきい値ThとToの選択が問題となる。これまでの検討では、ラジオのニュース音声を利用してしきい値の検討を行なって来た。その結果、健聴者の場合にはThを入力音声のフレーム毎パワーのピーク値の40%以下に設定すれば、話速変換の効果を感じることが可能であることがわかっている[4]。しかし対話に本方式を用いる場合、音声のパワーレベルの変動が大きいことが予想される。

そこでThはフレーム毎のパワーのピーク値の20%、Toは10%で決定されるように固定し、DSPの持つタイマー割り込みを利用して、１０秒毎にピーク値の更新を行なうことで、しきい値を自動的に入力レベルに対応させる機能を追加した。なお、これらのしきい値で加工を行なっても、健聴者では子音明瞭度に影響を受けないことを確認している。

（３）話速変換を実行するハードウエア

図３に本システムのスピーチプロセッサ回路のブロック図を示す。ＤＳＰには高速性の面からＴＩ社のTMS320C30(1machine cycle = 60nsec)を採用している。これにより話速変換による時間特性加工と帯域別振幅加工による周波数特性加工の両方を同時にリアルタイムで実現している。

なお、１フレーム(48msec)分の音声信号データの話速変換処理に必要なDSPの演算量は、ピッチ抽出方法の変更により従来より増え、最大で約150000 (machine cycle)である。したがって話速変換のみを行なう場合には300(nsec)程度のプロセッサでリアルタイム処理が実現可能となる。

A/DおよびD/A変換を含むAIC(アナログインタフェース回路)にはTI社のTLC32044を用いている。A/DおよびD/A変換のサンプリング間隔は75μsecに、また各フィルタのカットオフ周波数は5.1KHzに設定した。

上記DSPは16Mwordの外部メモリ空間を持つが、本回路には1Mword分のDRAMが搭載されており、２つの入力フレームバッファと出力リングバッファに使用されている。なお対話の場合、１回の発話が数分以上に及ぶことは少ないので、出力リングバッファの容量は従来の連続的に話速変換を行なう場合に比べて少なくてもよい。

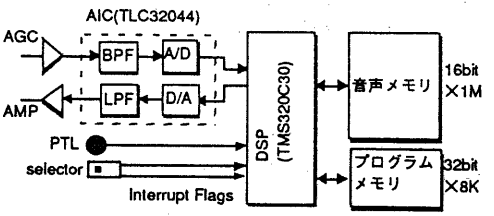

図３　スピーチプロセッサ回路

３．話速変換された音声の各種評価

ここでは、上記システムを用いて話速変換を施した音声に対し、これまでに行なったいくつかの評価検査の結果について簡単に述べる。

なお、以下の評価では話速変換処理の効果を評価することを目的としたため、周波数特性加工は特性が変化しないパラメータを用いて行なった。

（１）難聴者による評価

日本語の３音節の単語を用いた行なった評価については既に報告した[3]。その結果、単語の了解度（聞き取りの正確さ）は話速変換によってほとんど変化しなかった。しかし一対比較による「聞きやすさ」の評価においては、２音分離検査で時間分解能の低い難聴者ほど、話

速変換した音声を聞きやすいと判断する傾向が得られた。図4にこの結果を示す。図の横軸は1kHz純音の2音分離試験により得られた被験者の時間分解能、縦軸は各話速の「聞きやすさ」の評価値（最高で50点）である。これによれば、時間分解能が40～50msec以上の難聴者において、話速変換された音声の相対的な「聞きやすさ」の評価が高くなることがわかる。

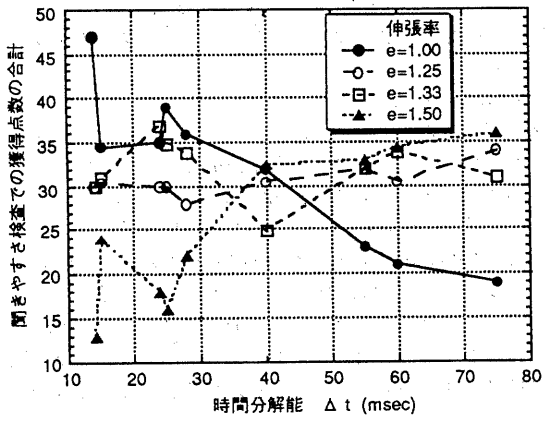

図4　話速変換した音声の「聞きやすさ」と時間分解能の関係

（2）高齢健聴者による評価

難聴者に対して行なったものと同じの評価を、高齢健聴者（65～70才）により行なった。

この年齢の前期高齢者層では、時間分解能の低下はあまり顕著ではなく、被験者全員50msec以下であった。このため難聴者で話速変換の効果が確認できた一対比較による「聞きやすさ」評価においても、話速変換の顕著な効果を確認することはできなかった。これより本システムの適用対象は、高齢者の中でも難聴度の高くなる後期高齢者層のみに限られることが示唆される。

（3）健聴者による外国語音声の評価

上記難聴者による評価結果から、話速変換の効果は時間分解能に関係して現われることが予想される。

一方、健聴者でも普段利用することの少ない外国語の聞き取りの場合には、言語処理に要する時間が増加し、文章理解における時間分解能は低下する。このため話速変換による聞き取り補助の効果が得られる可能性が高い。実際「ゆっくり」発話した外国語の方が、楽に聞き取れるように感じることは日常良く経験する。

そこで3種類の英語音声ソース（英会話教材、FENニュース、講演会録音）に話速変換処理を施し、40～50代の健聴者20名に聞かせ、アンケート形式で聞きやすさの主観評価を行なった。なお、被験者は全員海外出張等で英語を必要とした経験を持つ男性である。

図5にその評価結果を示す。図の白丸の中の数字は平均値を、エラーバーは偏差を示す。解答は「原音」を「普通（0点）」とした相対的な印象として答えてもらった。

同図から伸長率1.25あるいは1.33倍の場合が最も聞きやすいと判断され、1.50倍まで伸長した場合には、むしろ聞きにくく感じるようになることがわかる。また全体の評価傾向は音声ソースの違いにあまり依存しないものの、背景雑音の多かった講演会録音では、1.33倍および1.50倍の音声の評価が、他のソースに比較してやや高くなった。

この結果は、難聴者や高齢者における母国語聞き取りの場合の補助効果を直接示すものではない。しかし本方式の話速変換が、言語処理レベルの時間分解能低下も補助できる可能性を示唆するもの、と考えられる。

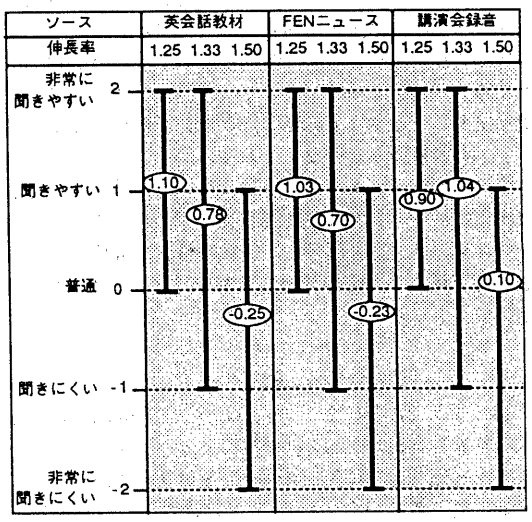

図5　健聴者による話速変換された英語音声の「聞きやすさ」評価の結果

4.対話向けインタフェースの提案

次に、従来システムを対話に応用するための方法について検討した結果について述べる。

（1）対話における話速変換の利用イメージ

対話においてリアルタイム話速変換を利用する場合のイメージを図6に示す。使用者は、スピーチプロセッサを手に持ち装置上のボタンを操作しながら音声を聞く。

音声は両耳用ヘッドホンを用いて装置から出力される音声のみを聞き、直接耳に入って来る音との混同がないようにする。

また使用者は、予め話速変換による時間遅れが生じることを相手に了解してもらい対話を行なう。時間遅れの大きさは一回の発話の長さや話速変換の伸長率に依存する。しかし伸長率は最大でも1.5倍であり、波形のパワーがThを越える部分のみが伸長されることや、無音区間の削除を行なうことから、時間遅れの大きさは前の発話の長さの1／2をこえることはない。

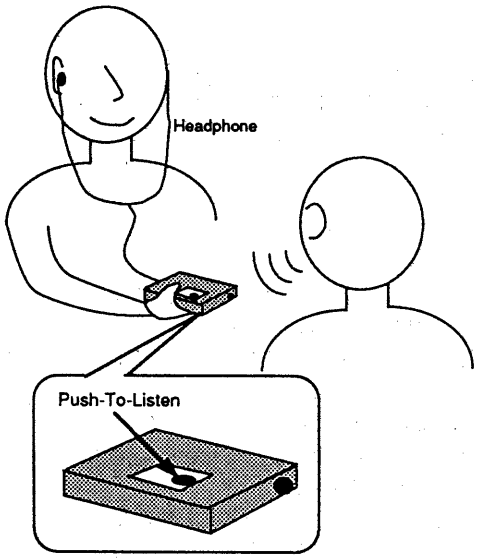

図6　対話における話速変換システムの利用イメージ

（2）対話のためのインタフェース

図6のような使用形態を取った場合、自分の声がシステムを通じてフィードバックされて来る。対話において自分が話す時には話速変換を行なう必要はない。逆に時間遅れを伴なう自分の声がフィードバックされると、殆どの場合発話を続けることが不可能となる。

このため対話において話速変換を利用する場合には、相手の声を聞く時にだけ話速変換処理が動作し、自分が発話する時には通常話速で出力されるような工夫が不可欠である。

本システムでは、PTL(Push-To-Listen)ボタンを装置上に設け、ボタンが押されている間だけ話速変換処理が行なわれ、押されていないときは入力音声がそのまま出力されるようにした。使用者は相手の音声を聞くときだけボタンを押し、自分が話すときにはボタンを離して対話を行なう。音圧レベルや複数のマイクロホンを用いて自動的に自分と相手の発話を区別する方法も考えられるが、本検討では動作の安定性の観点から、ボタン操作を利用することにした。

また押していたPTLボタンを離した時点での動作としては、すぐに現時点の音声が出力されるようにした。したがって、相手の話速変換された音声を全て聞かずにボタンを離すと、その時点で相手の発話内容はクリアされるようになっている。

また話速変換の伸長率は独立したスイッチで設定され、PTLボタンが押された時点で選択されている伸長率で処理を行なうようにした。PTLボタンと伸長率セレクタを交互に操作できるように配置すれば、発話単位での話速制御を行なうことも可能である。

（3）ハードおよびソフトウエアの改良

上記インタフェースを実現するため、従来のハードウエアに対しPLTボタン等を設けると共に、ソフトウエアの変更を行なった。

PTLボタンと伸長率セレクタは、DSPに複数用意されている外部割り込みフラグ用端子に接続し、話速変換処理の実行と伸長率の変更を制御するようにした。図7にプログラムフローを示す。

周波数特性加工は、話速変換の有無に係わらず実行されるように、75μsecごとに行なわれるA/D,D/A変換割り込み動作の内部に入れ、加工するデータをPTLフラグによって切り替えるようにした。即ちPTLボタンがONの間は、話速変換の出力データを処理し、PTLボタンOFFの時は直前のA/D変換データを処理する。

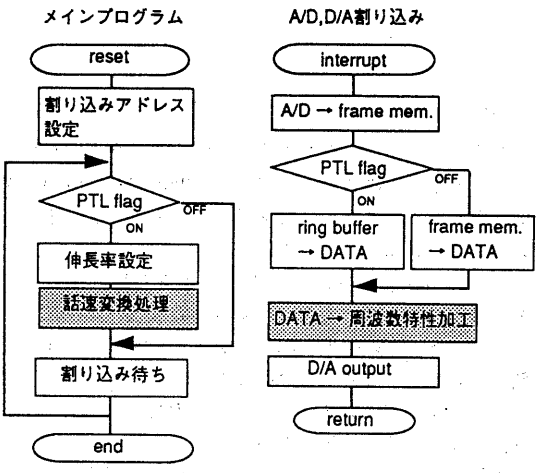

図7　プログラムフロー

5.健聴者によるインタフェースの予備評価

難聴者や高齢者による評価に先立ち、健聴者4名にPTLボタンを使用した話速変換動作を行なってもらい、その使用感を報告してもらった。ただしこの評価で使用したシステムは、従来のDSPボード上にPTLボタンのみを付加した模擬システムである。評価後に指摘された主な問題点は以下の3点である。

（1）PTLボタンを押すタイミングが取りにくい。
　相手の発話開始と共にボタンを押したいが、そのタイミングが取れないので、こちらの発話終了でPTLボタンを押さざるを得ない。

（2）こちらが聞き終わったことが相手にわからない
　対話は必ずしもが交互に行なわれることはなく、相手が続けて2回以上発話するケースもある。この場合には相手はいつ2回目の発話を開始して良いかわからない。

（3）途中でPTLボタンを離すと話がわからなくなる
　PTLボタンを誤って離すと、その段階で現在の時点に飛んでしまうため話がわからなくなる。

以上の指摘された問題点は、すべて発話タイミングを左右するものであり、特に（3）の点は必要な情報が欠落する点で重要である。PTLボタンを離した時に、いきなり現在の時点に飛ばず、残っているメモリ内の情報を全て呈示してから現在の時点に戻るための工夫が必要である。

また（1）および（2）の点に関しては、顔の表情や手振りなど、普段対話で用いている他のモダリティーによりカバーできるものと考えるが、今後実際に手に持って使用できるプロトタイプを作製し、再評価する予定である。

6.おわりに

本報告では、これまでに開発した話速変換方式を対話に応用するための方法として、PTL(Push-To-Listen)ボタンを用いたインタフェースを提案し、それを実現するためのシステム構成について述べた。

また、話速変換された音声に対する各種評価の結果と、上記インタフェースに対する予備評価の結果についても述べた。その結果、本インタフェースの使い勝手に関していくつかの問題点が指摘された。これらの問題点については、本インタフェースを備えたプロトタイプの試作に反映させる予定である。

本報告のシステムは、高齢者／難聴者の聞こえの衰えを補助することを第一の目的としているため、対象となる方々による評価が不可欠である。話速変換音声に対する評価としては、特に実際の音声を聞き取る場合に重要な、文章了解度に対する定量的な評価を次に行なう予定である。

一方、インタフェースの使い勝手等の評価に関しては、実際に手に持ってPTLボタン操作ができるプロトタイプを作製し、本システムの適用対象となる方々でも、簡単に操作できるかどうかという点についても確認する予定である。

さらに、本システムが外国語聞き取りのための補助装置として有効であるかという点にについても、文章了解度などの定量的な評価を進めて行きたい。

＜謝辞＞

本システムの難聴者による評価に関してご協力頂いた、北海道大学・電子科学研究所の伊福部達教授、今村俊樹氏に感謝いたします。

＜参考文献＞

[1]中村、他「高品質リアルタイム話速変換システム」信学技報　SP92-55 (1992.9).

[2]櫛渡、他「高齢者向け音声加工を行なうポータブルDSPシステムの開発」信学技報　SP92-54,HC92-31(1992.9).

[3]櫛渡、他「難聴者による話速変換方式の評価」信学技報SP92-150 (1993.3).

[4]櫛渡、他「ポータブルDSPシステムを用いた話速変換方式の検討」日本音響学会講演論文集1-7-6 (1993.3)

[5]S.Malar "Time-domain algorithms for harmonic bandwidth reduction &time scaling of speech signals",IEEE Trans. Acoust., Speech, Signal Processing. Vol. ASSP-27, No.2(1979)

[6]池沢、他「話速変換に伴う時間伸張を吸収するための一方法」信学技報 SP92-56 (1992.9)