ヒューマンインタフェース 44－6
（1992. 9. 10）

高品質リアルタイム話速変換システム

中村 章，清山信正，池沢 龍，都木 徹，宮坂栄一

NHK 放送技術研究所

内容梗概

従来の補聴器は入力音の周波数帯域の一部，もしくは全体を増幅する拡声器であり，伝音系難聴者にとっては有効であるが，聴覚経路全般，特に内耳や聴覚中枢系で機能劣化した高齢者にとっては必ずしも有効でない場合が多い．一般に加齢により，伝音系の機能劣化の他に，聴覚中枢系の機能劣化が生じる．その結果，(1)語音識別速度の低下，(2)信号音と背景音との分離能力低下等の症状が見られる．そこで今回，症状(1)を補い，高齢者にも良好な音声放送サービスを行うことを目的とし，早口で話された音声をピッチや個人性を保ったまま話速のみを，ゆっくりした音声に高品質かつリアルタイムで変換できる補聴システムを試作した．このシステムの特徴は従来の補聴器とは全く異なり，聴覚中枢系の劣化に伴う「聞こえ難さ」を補償することにある．

Real time voice speed converting system without impairment in quality.

Akira Nakamura, Nobumasa Seiyama, Ryou Ikezawa, Tohru Takagi and Eiichi Miyasaka

NHK Science & Technical Research Laboratories

1-10-11, Kinuta, Setagaya-ku, Tokyou 157, Japan

This paper presents a new hearing aid system intended to compensate degradation of perceptual functions in the central auditory pathways which can be found typically for elderly people, while conventional hearing aid systems are effective only for conductive hearing impairments. Typical types of such functional degradation are (1) degradation of rate of processing speed for identification and that of effective information capacity, (2) degradation of ability for extraction of a focusing signal from background noises (music, environmental noise etc.). The new system is designed to compensate the first item; that is, it is able to convert speech more slower at a rate of voice speed to what you want, for yourself, on real time, with invariance in pitch and without impairment in quality.

1 はじめに

総務庁が発表した1989年の高齢者人口推計によると，65歳以上の人口は1,429万人で総人口の11.6%を占めている．厚生省人口問題研究所の推計[1]では今後増加を続け，2000年には16.3%，2020年には現在の2倍以上の23.6%に達する勢いで，急速に高齢化社会に突入していくことが予想されている．30年後には放送視聴者の4人に1人は高齢者という勘定である．娯楽や生涯教育をテレビやラジオに求め，生活時間のかなりの部分を，放送メディアとともに過ごす高齢者がますます多くなってくる．

加齢によって身体機能や生理的機能は一様に変化するわけではないが，一般的にだれしも低下の一途をたどる．特に，聴覚や視覚などの感覚器官の衰退が顕著にあらわれてくる．現時点において，高齢者（60歳以上）の男性で約1/3，女性で約1/4の人たちが何らかの音の聞こえの悪さを訴えているという報告[2]がある．今後，高齢者にも良好な放送サービス*)を行うことが急務の課題となる．

2 高齢者の聴覚特性

従来の補聴システム（補聴器）は入力音の周波数帯域の一部，もしくは全体を増幅する拡声器であり，伝音系難聴者にとってはある程度有効であるが，聴覚経路全般，特に内耳や聴覚中枢系で機能劣化した高齢者にとっては必ずしも有効でない場合が多い．というのは，聴覚中枢系の機能が劣化すると，その結果，(1)語音識別速度，(2)信号音と背景音との分離能力低下等が生じるためである．

ここでは(1)に着目し，高齢者を含む各年代層の被験者を対象に，最もききやすい発話速度，及び間（ま）の関係を求め，この知見をもとに，各受聴者に最適な発話速度，及び間（ま）を制御できる補聴システムを試作した．このシステムの特徴は従来の補聴器とは異なり，聴覚中枢系の劣化に伴う「聞こえ難さ」を補償することにある．

3 話速変換方式の原理[3]

話速変換方式のブロックダイアグラムを図1に示す．

入力音声をブロック(a)で無音区間，無声区間，有声区間に分割する．無音区間の延長を行うことにより間（ま）を制御する．無声区間は発話者の個人性，及び音韻性を保つために加工しない．有声区間はブロック(7)でピッチ周期を自己相関等により抽出し，ブロック(e)でピッチ区間の分割を行い，ブロック(f)で補間して延長する．ブロック(b)，(f)，及び(c)から得られた無音，有声，無声区間をもとにブロック(g)で合成し，話速を変換する．図2に原音声波形データと話速を変化した波形を示す．

通常，このような処理を施すと合成音

*) 例えば，アメリカでは聴覚障害者のために，すでにテレビニュースの字幕スーパーが義務づけられているし，HDTV（ハイビジョン）に対してもその音声チャンネルに，聴覚障害者用の専用チャンネルを設けるべきだとの提案がアメリカ，カナダ，スウェーデン，イギリス等から出されている．

声の品質が劣化してしまい，明瞭度も低下する．そのため健聴者が聴取しても，その歪のために聞き難くなってしまい，変換パラメータの効果が現われてこないことがよくある．本方式の特長の1つは，この変換音声の品質が極めて良いことにある．

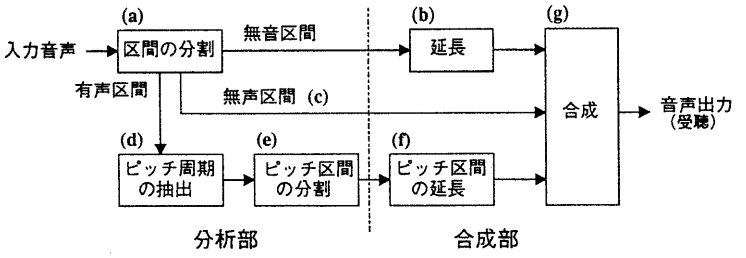

図1 話速変換方式のブロックダイアグラム

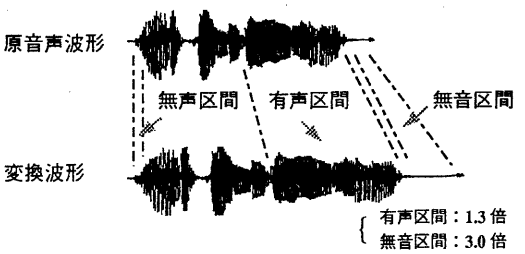

図2 話速変換波形

4 評定実験[4]

4.1 刺激音

熟練したアナウンサー男女10名が発声した同一文のニュース音声のうち，発話速度が最も速い男性発話者（約9.4モーラ／秒，長さ11秒）を原刺激として用いた．刺激に用いたニュース文を**表1**に示す．

森田村への企業誘致は，これで23社になりますが，女性だけを採用する企業が多く，村では今後，男性も雇用する企業の誘致に力を入れることにしています．

表1 刺激に用いたニュース文

4.2 実験方法

有声区間，及び無音区間の延長比率の変化幅を予備実験により適切と考えられる比率（**表2**）に変えた13のニュース音声，及び原音声の合わせて14のニュース音声を刺激とし，『聞きやすさ』，及び『落ち着き度合い（せわしさ）』の指標について5段階（1,2,…,5）の絶対評価を行った．通常，この様な文章を刺激として評定実験を行う場合，その内容を書き取らせたり，質問するなどして了解度を求める場合が多い．しかし，内容を理解できたかどうかは各被験者の経験と知識

刺激番号	延長区間	
	有声（倍）	無声（倍）
1	1.0	1.0
2	1.2	1.0
3	1.4	1.0
4	1.0	1.4
5	1.0	2.0
6	1.0	3.0
7	1.2	1.2
8	1.2	1.4
9	1.2	2.0
10	1.2	3.0
11	1.4	1.2
12	1.4	1.4
13	1.4	2.0
14	1.4	3.0

表2 有声・無音区間の延長比率

に依存し，今回のようなバックグラウンドが異なる被験者を対象とする場合に，取扱いが困難となる．従って，抽象的ではあるが『聞きやすさ』と『落ち着き度合い（せわしさ）』を評価の指標として用いた．

被験者は20歳代10名，30歳代8名，40歳代2名，50歳代3名，60歳代6名，70歳以上2名の合計31名である．刺激提示方法はスピーカー再生，提示音圧レベルは約70dB(A)であり，実験は1～40回繰り返した．

5 実験結果及び検討

5.1 評定実験

実験結果の代表的な20歳代の例を図3，図4に，60歳以上の例を図5，図6に示す．評価の平均値を濃淡（濃：高い評価，淡：低い評価）で示す．各被験者毎に多少のばらつきが生じているが全体として以下の傾向が見られる．

(1)健聴な20歳代の被験者では有声区間を1.2倍，無音区間を1.2倍延長した方が聞きやすい．

(2)60歳代以上の被験者では原音声よりも有声区間を1.2～1.4倍，無音区間を1.2～2.0倍延長した方が聞きやすい．

(3)若年者，高齢者とも無音区間を3.0倍延長して，間（ま）があきすぎると逆に聞きにくくなる傾向がある．

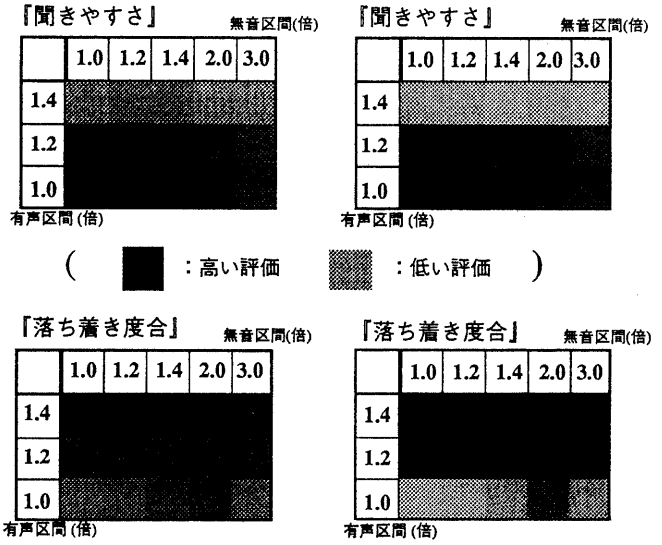

図3 5段階評定結果（20歳代の評定者A）

図4 5段階評定結果（20歳代の評定者B）

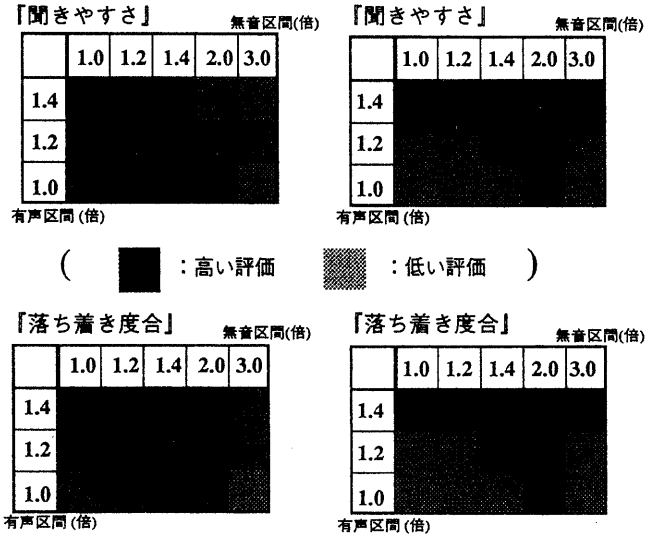

図5 5段階評定結果（60歳代の評定者C）

図6 5段階評定結果（70歳代の評定者D）

5.2 『聞きやすさ』と『落ち着き度合い』との相関

発話速度，及び間（ま）が及ぼす『聞きやすさ』と『落ち着き度合い（せわしさ）』との関係を年代別に検討するため，これらの指標間の個人別相関係数を求めた．年代別に分け，表3に示す．

50歳代以下の被験者では『聞きやすさ』と『落ち着き度合い』との相関傾向にばらつきが生じているが，60歳以上の被験者では高い相関を示していることが分かる．このことは高齢者の場合，発話速度，及び間（ま）を延長することにより，落ち着いて聞きやすいと評価しているのに反し，若年者では落ち着いていることが必ずしも聞きやすいとは評価していないことを示している．

5.3 実験の検討

今回の実験の結果で，高齢者にとっては『聞きやすさ』と『落ち着き度合い』には高い相関があり，若年者に比べて，『聞きやすさ』の要因は発話速度と間（ま）の延長にあることが分かった．今回の実験から，発話速度と間（ま）を延長する手法が，語音識別臨界速度や短期記憶容量の低下を補償し，最終的に音声の分析や単語の認識（統語），あるいは文の意味把握（統辞）等の能力低下を補うことが可能と推察できる．

6 高品質リアルタイム話速変換システム

上述の知見をもとに，発話者の個人性

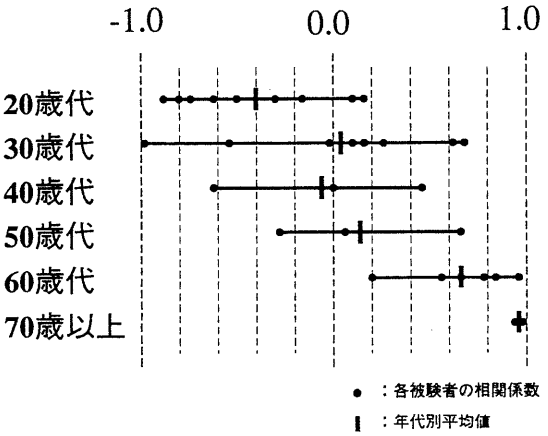

表3 年代別相関係数

やピッチを保ったまま，変換音声の品質劣化を抑え，各受聴者が聞き取りやすい速度に自らリアルタイムで変換できる話速変換型補聴システムを試作[6]した．

7 ハードウェア構成

3節で示した話速変換方式のアルゴリズムを実時間で処理するために，システムのCPU部を8個のトランスピュータモジュール（動作周波数:30MHz，DATA-BUS:32bit，公称CPU性能:15.0Mips，FPU性能:2.3MFlops，リンク数:4，リンク転送速度:20Mbps）をパイプライン接続して構成し，入出力部に16ビットのA/D，D/A変換器，及び音声の有声区間，無音区間を延長するためのパラメータ入力用ボリュームとして8ビットのロータリエンコーダを用いた．

システムの構成図を図7に示し，各トランスピュータでの処理内容を以下に概説する．各々のブロックが個々のトランスピュータに相当している．

各ブロックにおける処理波形の一部を

図8に示す.

(1)入力音声（**図8-a**）は16bit，16kHzサンプルし，入力用バッファに蓄える.

(2)固定フレーム長で短時間パワー（**図8-b**），零交差数を計算し（**図8-c**），分析区間の分割を行う．パワー，零交差数および自己相関から，各フレーム毎に無音，無声，有声の判定を行う．同時に，分析窓幅およびデシメーション倍率を変えて自己相関を計算し，最適なデシメーション倍率を求める.

(3)前段で求めたデシメーション倍率に従い，波形のデシメーション（**図8-d**）を行う.

(4)デシメーション波形の自己相関（**図8-e**）から，各フレーム毎に平均ピッチ周波数を算出する．半ピッチ，倍ピッチなどのピッチ抽出誤り対策として，ピッチを求める際に自己相関のピークからいくつかの候補を求めて，閾値によりピッチを決定する.

(5)デシメーション波形に，各フレーム毎に求めたデシメーション倍率と平均ピッチ周波数に合わせたLPF（**図8-f**）をかけ，この波形のピークからピッチ周期でピッチの開始点およびピッチ数を求め，ピッチスケールを構成する.

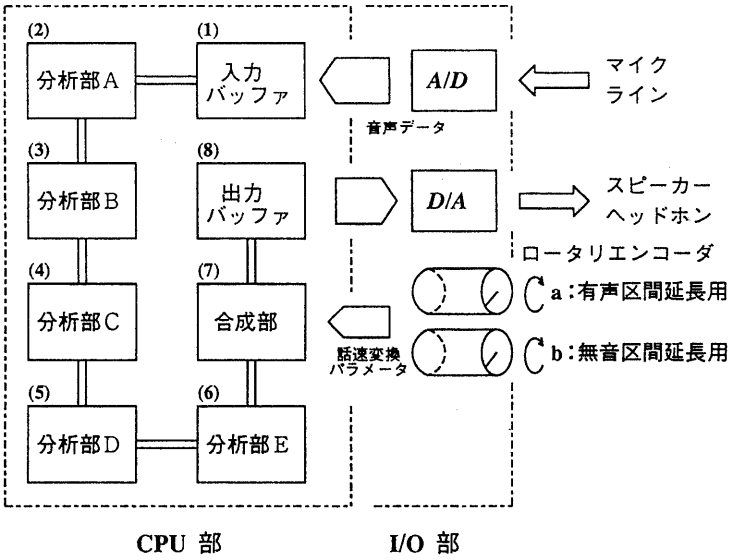

図7 話速変換システムの構成図

原波形の有声区間の中心（定常部）から時間的に前後に，波形の1ピッチ区間内の最大振幅の直前の零交差がピッチ開始点となるように，ピッチスケールに合わせて最適なピッチ開始点を定める（**図8-g**）.

(6)最終的に分割後のデータのつじつまが合うように各区間の微調整を行い，無音，無声，有声区間の開始点，終了点ピッチ数およびピッチ開始点等のパラメータを次段の合成部へ転送する.

(7)ロータリエンコーダから入力された話速変換パラメータ（有声区間と無声区間の延長比率）をもとに，有声区間，及び無音区間を延長して話速を変換した音声を合成する.

(8)出力用バッファに蓄えられたデータを逐次，D/Aして音声出力する.

このシステムは前述したように，音声

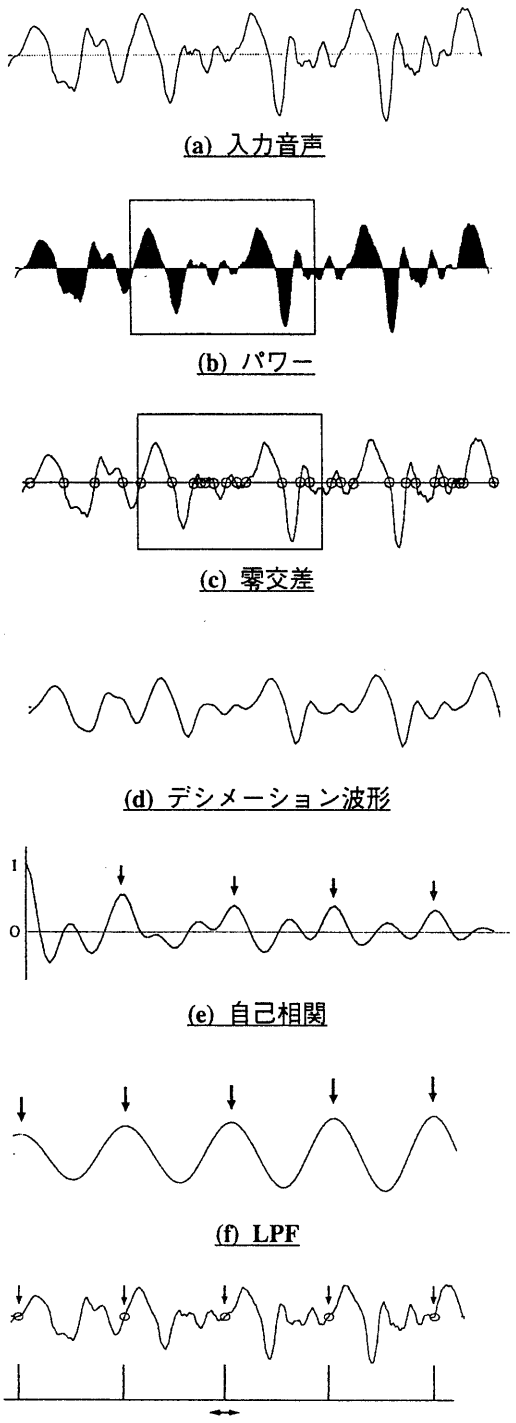

図8 各ブロックにおける処理波形

のピッチ周期を自己相関等により抽出し，波形レベルで適宜，補間を行うことにより，(a)話者の声の高さを変えずに，(b)声の質も保ったまま高品質に，(c)受聴者が自分の好みの速度に合わせて，実時間で話速を変えることができる．また，高齢者が自分の好みの速度にリアルタイムで変換できる点も大きな特徴である．

7 むすび

各年代層の被験者を対象に，最も聞きやすい発話速度，及び間（ま）の関係を求め，発話者の個人性やピッチを保ったまま，変換音声の品質劣化を抑え，受聴者の聞き取りやすい好みの速度にリアルタイムで変換できる話速変換型補聴システムについて報告した．

本話速変換方式では無音，無声，有声の判別を行い独立に制御しているが，有声区間と母音の区別を行っていないため，発話者によっては変換音声に「もたつき感」が生じる場合が考えられる．今後，有声子音と母音を独立に制御するなど，更に高品質化を図る予定である．

2節で，加齢により聴覚中枢系の機能が劣化すると，その結果，語音識別速度の低下のみならず，信号音と背景音との分離能力等が低下することを示した．広い意味で，高齢者のための補聴システムを考える場合，語音識別速度の低下を補う話速変換型補聴システムでは不十分であり，信号音と背景音との分離能力低下を補うことも重要である．現在，信号音と背景音との分離方法について検討している．また，このような補聴システムを使

用する場合，高齢者にとって操作性に関するヒューマンインターフェイス部が重要である．そこで現在，話速変換パラメータ入力用ボリュームを操作せずに，各受聴者に最良の発話速度へ自動的にフィッティングする「自動フィッティングシステム」についても検討している．

冒頭でも示したように，我国は急速に高齢化社会に向かって進んでいる．娯楽や生涯教育をテレビやラジオに求める高齢者の期待に応えるためにも，このような聴覚補助システムの研究がさらに重要になってくるであろう．

最後に，このシステムはネイティブスピードで発声された外国語を，聴取しやすい速度に変換できるなど，幅広い応用が考えられる．

参考文献

[1] 厚生省編：簡易生命表(1988)

[2] 岡本：老人性難聴と補聴器，日本医師会誌，Vol.101,No.5(1990)

[3] 都木ほか：電子情報通信学会誌 A,Vol.J73A,No.3,387-396(1990)

[4] 中村ほか：日本音響学会秋季大会,381-382(1991)

[5] 佐藤ほか：高齢者の母音識別臨界速度，Audiology Japan, No.31, 737-743(1988)

[6] 中村ほか：日本音響学会春期大会,329-330(1992)