リアルタイム話速変換装置

Real-Time Speech Rate Converter

UDC 534.85 : 681.846

小林雅哉	Masaya Kobayashi	情報通信事業本部 端末機器事業部 第１技術部
待寺侯大郎	Kohtarou Machidera	情報通信事業本部 端末機器事業部 第１技術部
明戸正人	Masato Aketo	情報通信事業本部 端末機器事業部 第１技術部
黒田政廣	Masahiro Kuroda	研究所 開発部

1 まえがき

高齢者は，テレビやラジオ番組中の音声が早口で，聞き取りにくいと感じる人が少なくない。また，電話機，カラオケ等のオーディオ機器に音声処理（合成，伸張・圧縮，音程変換等）機能を持つ装置が増加している。これらの装置は，音声を符号化し，内部処理後に復号化して出力している。日本放送協会（以下NHK）殿では，符号化したデータを用いて音声伸張を行い，早口で話された音声を，声の質をそのままに，話し手があたかも「ゆっくり」話したように変換できる話速変換技術の研究開発を進めている。この技術を搭載し，実用化に向けて基本性能の向上と，携帯性を考慮したサイズの製品を実現した（図1参照)。

図1　外観
External view of speech rate converter

2 開発方針

2.1 特徴

①非常にゆっくりとした話速に変換しても，変換音声に不自然さが少ない。

②変換音声の高域特性の改善により，子音が明瞭である。

③音楽や効果音などの背景音に影響されない安定した話速変換が可能である。

④話速変換された音声と映像との「ズレ」に伴う違和感が少ない。

⑤ポータブルタイプであるため，持ち運びが可能である。

2.2 据置タイプとの比較

据置タイプとは，ＮＨＫ殿により開発された話速変換装置である。

据置タイプの装置と比較して体積比で約1/4，重量比で約1/5とする。また，信号処理に特化した構造を持つDSPプロセッサを採用し，より高品質な音声を再現する。据置タイプとの比較を表1に示す。

表1　本装置（ポータブルタイプ）と据置タイプの比較
Comparison list

		本装置（ポータブルタイプ）	据置タイプ
ハードウェア	プロセッサ	DSP	RISC
	メモリ	32MByte	16MByte
	サンプリング周波数	32kHz	16kHz
	量子化ビット数	16bit	16bit
	バッテリ動作	可	不可
外観	質量	約320g	約1.5kg
	サイズ	132(W)×83(D)×32(H)mm	180(W)×130(D)×65(H)mm

（注）据置タイプのサイズ，質量は，日本工業新聞（平成9年4月10日発行）から抜粋

2.3　機能

基本的な機能を以下に示す。

①高品質な音声を再現するため，量子化ビット数16ビット，サンプリング周波数32kHz（据置タイプの2倍）とする。

②リアルタイム性を損なわないように，入力音声に対する話速変換後の出力音声の最大遅延時間は100ms以下とする。

③連続蓄積時間（音声を内部メモリに蓄積可能な時間）は8分以上とする。

④2種類の変換モード（一様伸張／時間吸収）を選択可能とする。一様伸張モード（以下ラジオモードという）では，有声区間および無音区間倍率をそれぞれ1.0～2.0倍まで伸張可能とする。時間吸収モード（以下テレビモードという）では，有声区間先頭倍率を1.0～1.6倍まで伸張可能とする。設定範囲は7段階で，遅れ解消機能を持つ。また，モードや設定を変更する場合に出力音声が即応する機能を持つ。

⑤AGC回路（Automatic Gain Control）を搭載する。

⑥赤外線リモコンにより制御可能とする。

2.4　バッテリ

バッテリによる連続動作は1時間以上とする。定格700 mAh以上で，本体ケースサイズを考慮した形状（角形または円筒形）とする。

2.5　システム構成

表2にシステム構成を示す。

表2　システム構成
Construction of speech rate conversion system

項番	名　称	個数	備　考
1	本体部	1	
2	拡張部	1	充電機能，音声入出力用ピンジャックを搭載
3	リモコン	1	電源ON/OFF，モード設定等
4	ACアダプタ	1	

本体部と拡張部に機能分割する。本体部には，電源，設定ボタン類，音声入出力端子（ミニジャック）等を具備し，拡張部には，充電機能，音声入出力端子（ミニジャック，ピンジャック）を搭載する。基本的に本体部は，携帯することを目的とする。そこで，本体部の外形サイズを小型・薄型で，丸みを帯びた形状とし，軽量化を図る。拡張部は，本体部と接続したときに違和感がなく，操作性の良い形状とする。

2.6　ソフトウェア

リアルタイムかつ高品質な話速変換を行えるよう，高速処理可能なソフトウェア構成とする。

検証にはシミュレータを使用し，演算時間・遅延時間の推定を行う。

3　設計の要点

3.1　ハードウェア

ハードウェアの構成を図2に，仕様を表3に示す。本開発では，まず本体サイズを決定した。そのため，内部の空間を効率良く使用するため，メインボードとパネルボードにプリント基板を分割した。さらに，メインボードの実装効率向上およびROM交換の作業性向上のために，メモリボードを分離した。

表3　仕様
Specification

項　目		仕　様	備　考
プロセッサ		32bit浮動小数点DSP	
メモリ		32MByteDRAM	
音声サンプリング周波数		32kHz	
音声記憶データ		16bit	
音声チャネル		1	モノラル
最大出力レベル		2.7$V_{p\text{-}p}$	
ヘッドフォン端子		1	モノラル
外部入力端子		1	モノラル
バッテリ		ニッケル・水素蓄電池	
電池寿命		約1時間	
状態表示LED		18個	
電源アダプタ	入力電圧	AC101V ± 10V，50/60Hz	
	定格出力電圧	9V	
	公称出力容量	10.8W	
	外形寸法	98(W)×57(D)×50(H)mm	
	質量	約600g	
拡張部入力電圧		AC101V ± 10V，50/60Hz	
消費電力		約3W	
外形寸法	本体	132(W)×83(D)×32(H)mm	
	拡張部	162(W)×115(D)×43(H)mm	
質量	本体	約320g	電源アダプタ含まず
	拡張部	約360g	電源ケーブル含まず
動作環境条件	温度	5～35℃	
	湿度	20～85％	ただし結露しないこと

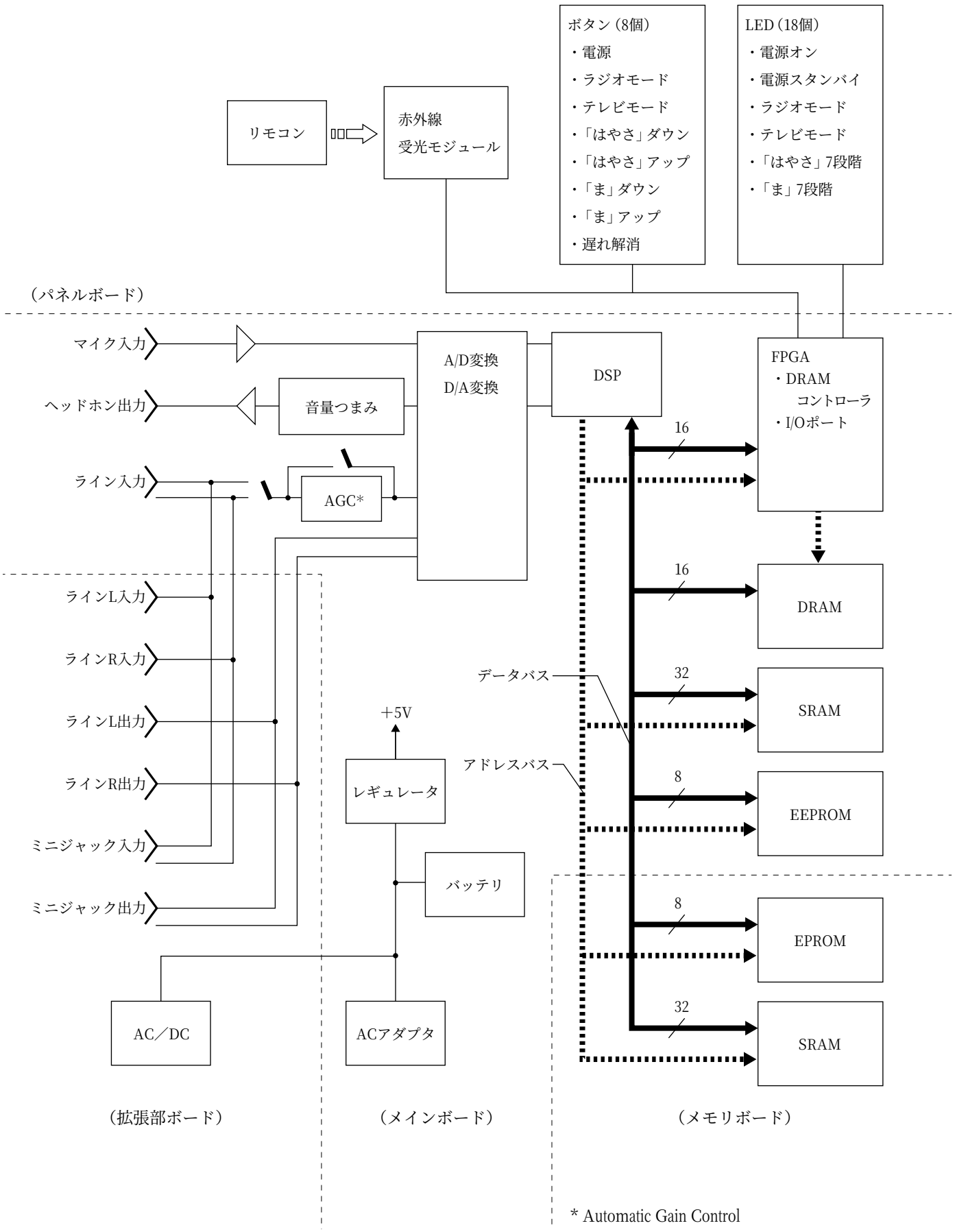

図2　ハードウェアブロック構成図
Block diagram of hardware design

3.1.1　メインボード

（1）プロセッサ

次の①～　の条件を満足するADSP21060プロセッサ（アナログデバイセズ社製）を選定した。

①浮動小数点DSP

②高速

③マルチプロセッサ対応（DSPを並列で接続可能であること）

④アクセス可能な外部メモリ空間が64 MByte以上

⑤開発環境（Cコンパイラ，アセンブラICE等）

⑥シミュレータ（コンピュータ上でのソフトウェア動作が可能で，仮想的なDSP動作のシミュレーションを行うことが可能）

変換精度を重視し，また主にC言語で開発することを想定して浮動小数点DSPを選択した。自然な再生音声とするため演算処理が多く，1個のDSPではリアルタイムな変換が不可能である場合に備えて，同じDSPを並列に接続可能であることを条件とした。音声データは8分以上蓄積可能とするため，外部メモリ空間は64MByte以上とした。

（2）音声入出力部

マイク入力端子，LINE入力端子または拡張部からの音声入力信号のA/D変換，話速変換，音声データ蓄積，D/A変換を行い，ヘッドホン／ライン出力端子または拡張部（LINE出力端子）へ音声信号を出力する。

（3）FPGA

DRAMコントローラおよびパネルボード制御用I/Oポートを FPGA（Field Programmable Gate Array）で構成することにより部品点数の削減，および同一サイズでDRAMの大容量化を行った。

JTAG（Joint Test Action Group; IEEE1149.1）ポートによるオンボード書換え型を使用することにより，FPGA変更作業の簡素化を図った。また，DSPに対応する高速動作のEEPROM型を使用した。

（4）その他

本体下面にディップスイッチを備え，LINE入力端子のLチャネル，Rチャネルの選択およびAGCのオン/オフを設定可能とした。

安全性と約1時間の連続バッテリ動作を可能にするため，ニッケル水素（Ni-MH）電池を搭載した。

3.1.2　メモリボード

音声蓄積用メモリは，大容量が見込まれるため，DRAM（32 MByte）を使用した。変換用変数領域メモリとしてSRAM，設定内容保持用としてEEPROM，BOOT用としてEPROMを使用した。

また，DSPプログラムはDSP内部のSRAMに格納しているが，ソフトウェアの拡張性を考慮し，SRAMを増設可能とした。

3.1.3　パネルボード

タクトスイッチ，LED，および赤外線受光部で構成し，ボタンまたはリモコンにより設定した動作状態を18個のLEDにて表示する。

「はやさ」UPまたはDOWNボタンを押下することにより，ラジオモードでは有声区間の伸張倍率を，テレビモードでは時間吸収開始カーブ倍率を7段階で設定可能とする。「ま」UPまたはDOWNボタンを押下することにより，ラジオモードでは無音区間の伸張倍率を7段階で設定可能とした。

3.1.4　拡張部ボード

AC/DCコンバータ，LINE入力端子，LINE出力端子，ミニジャック入力端子，ミニジャック出力端子で構成した。

3.2　ソフトウェア

プログラム構造を図3に示す。

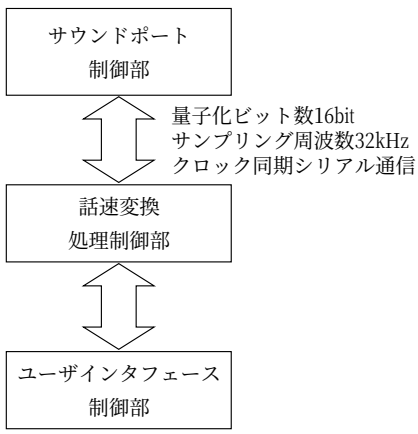

図3　ソフトウェア構造図
Block diagram of software design

3.2.1　処理機能

高速処理に対応するために，以下の処理を行った。

（1）並列演算処理

DSPは32ビットIEEE浮動小数点フォーマットを扱い，終始この精度を保持することにより，演算途中における丸め込

み誤差を少なくした。また，DSPの機能である1サイクル同時処理（2つの処理を同時に実行）を行うことにより，演算処理の高速化を図った。

（2）ローディング（BOOT）機能

電源投入後，プログラムを外部のEPROMからDSPの内部SRAMへロードして，プログラムを実行する。EPROMよりアクセススピードが速い内部SRAMでプログラムを実行することにより，高速処理可能とした。

（3）ライブラリ使用

DSPの最適なライブラリを使用することにより，自己相関処理等の高速化を図った。

（4）EEPROM制御

動作モード（ラジオモード，テレビモード），および各モード時の設定データをEEPROMに保持する。通常，EEPROMの書込みには時間がかかるが，書込み中でも話速変換処理時間等に影響のでないように，書き込み確認を分離したソフト構成にした。

（5）シミュレーション

サンプリング周波数32kHzでの入力音声に対する出力音声の遅延時間が100 ms以下となることをシミュレーションにより検証を行った。シミュレーションによる結果は，45.2msの遅延時間であり，1個のDSPで今回の仕様を満足できると判断した。

3.2.2　動作モード

（1）ラジオモード

ユーザの設定した有声区間倍率（1.0～2.0），無音区間倍率（1.0～2.0）でそれぞれ一様に音声データを伸張して，「ゆっくり」にする。音声を一様に伸張していくので，徐々に原音声と比較して遅延が大きくなる。出力される音声の遅延が余りにも大きくなった場合は，蓄積した音声データをクリアして遅延を解消できるようにした。

（2）テレビモード

ユーザの設定した有声区間先頭倍率（1.0～1.6）で音声の最初の部分を伸張して徐々に倍率を低くし，無音区間を削除することにより，音声の伸張に伴う遅延を吸収する。よって，次の音声の話し始めを合わせることができる。

3.2.3　A/D・D/A変換

クロック同期シリアルによる送受信によって制御する。本LSIは，マイク等から入力された音声をA/D変換し，DSPにより音声処理された音声データをD/A変換して出力する。

3.3　機構

3.3.1　本体部

（1）外観

外形サイズは132(W)×83(D)×32(H)mm，質量は約320 g（電源アダプタ含まず）で，従来装置の外形サイズ180(W)×130(D)×65(H)mm，質量約1.5kgに比べ，小型・薄型・軽量とした。この外形サイズとするために，ケース板厚を2mm，メインボード厚さを0.8mm，プリント基板の上面電気部品高さを最大7mm（電源部除く），下面電気部品高さを最大4mmとした。また，メモリボード厚さを0.6mmとし，スタッキング接続とした。

また，出力端子をヘッドホン／ライン共用し，ケース内に納めた。

バッテリの選定も，小型・軽量化には重要である。外形寸法は85(W)×50(D)×14.5(H)mmで，質量は約160 gのニッケル水素電池を使用した。

本装置の外形サイズ132(W)×83(D)mmは，美観を与える黄金分割比の約1:1.6である。

本体の色彩は，近年のオーディオ機器等で流行のシャンパンゴールドとした。また，質感を出すため，および指紋，汚れ等を目立たなくするために，ケース全体にヘアライン処理を施した。

（2）構造

本体部前面には，設定状態表示用のLED，赤外線受光部を配置した（図4参照）。

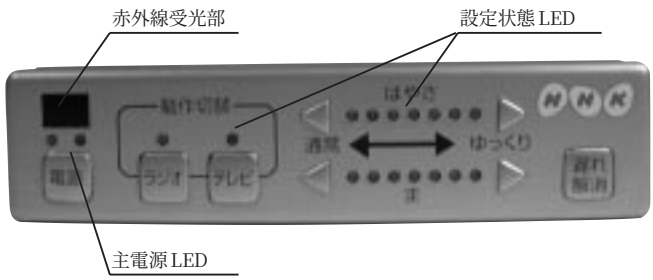

図4　本体部正面図
Front panel of main unit

背面には，電源アダプタ端子，リセットスイッチを配置した（図5参照）。側面には，ライン入力端子，マイク入力端子，ヘッドホン／ライン出力端子，音量つまみを配置した（図6

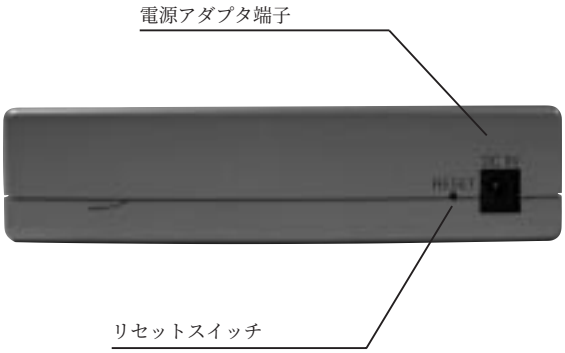

図5　本体部背面図
Rear view of main unit

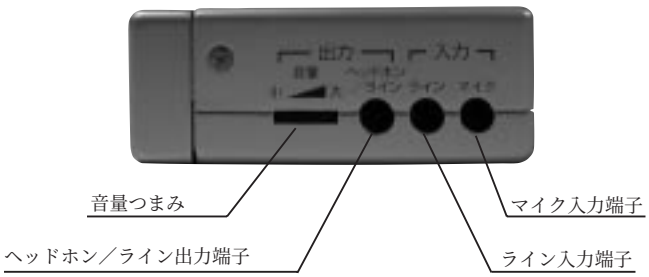

図6　本体部側面図
Side view of main unit

参照)。下面には，ディップスイッチ，拡張部との接続コネクタを配置した。この接続コネクタは，拡張部との接続時にのみ必要となるので，本体部単体で使用した時に，外観を損なわないようにフタを設けた。また，拡張部と接続するための位置決め用の凹部を設けた。

ケースの材質は，難燃ABS（Acrylonitrile-butadiene-styrene）を使用した。

設定状態表示用LEDは，ケースから突出させることにより，視認性を良くした。特に主電源のLEDには，緑色（動作中）と赤色（待機中）の２色を使用し，状態を容易に確認できるようにした。

表示文字（3.5mm）は，高齢者にも見やすいように大きくした。

操作ボタンは，キートップ＋タクトスイッチをパネルボードで押さえる構造とした。キートップは操作性を良くするために大きく（10×10mm），タクトスイッチは本体部を小型にするために薄型（3.1mm）で，クリック感のあるものを使用した。

図7　拡張部背面図
Rear view of extended unit

赤外線受光部のフィルタには，外来光による誤動作をなくすため，特定の赤外線（近赤外線波長　700nm以上）を透過するメタクリル樹脂を使用した。

3.3.2　拡張部

外形サイズは，162(W)×115(D)×43(H)mmで，質量は約360ｇ（電源ケーブル含まず）の安定感のあるものとした。ユーザの配線作業を容易に行えるように，背面に電源ケーブル，音声入力端子，音声出力端子を集中的に配置した（図7参照)。本体部と接続したときに，操作ボタン等の操作性を考慮して，拡張部前面の形状を傾斜させた。本体部と接続したときは，拡張部から電源を供給するので，誤って本体部の電源アダプタを付けたまま接続できないように，本体部の電源アダプタ端子部を隠す形状とした。本体部と同様，接続するための位置決め用の凸部を設けた。また，拡張部は据え置き型なので，操作時の滑り止め用のゴム足を設けた。

4　変換の効果と性能評価

4.1　話速変換音声の評価

ゆっくりとした話速に変換する場合，アナログ的にテープレコーダの回転数を遅くする方法と，ディジタル的には再生時サンプリング周波数を小さくする方法等がある。しかし，これらの方法ではピッチ周波数が低下したり，有声音・無声音・無音が一様に時間軸上で伸張されるため，間延びした音声になり，声質が変化し明瞭度も損なわれてしまう。

本装置では，不自然さが少なく，子音が明瞭な音声であることを確認した。また，テレビモード時に，画面の話し始めとの違和感が少ないことを確認した。

参考文献（2）の評価によると，老人性難聴者に対する伸張率は，アナウンサーの平均的な話速（450～570モーラ*／分）に対して，有声区間を約1.3～1.4倍にすることが適当であると示唆されている。

*1モーラは短母音を含む1音節に相当する。

4.2　音声の遅延（原音声からの話し始めの遅延）

映像と同期してリアルタイムに話速変換するためには，演算による遅延を最小限としなければならない。音声の遅延については，リップシンク（映像とそれに伴う音声のタイミングを一致させる）に関する規格がITU-R（国際電気通信連合無線通信部門）で検討されている。この遅延に関しての検討も報告されており，それによると，音声が映像より遅れる場合，検知限（遅延と認識する限界）は122msで，許容限は182msである。本装置の実遅延時間は，60.98±14.27 ms（平均±3σ）である（一例を図8に示す）。ただし，この遅延時間は息継ぎ（無音部分）後の話し始めの遅延である。本装置は，息継ぎ時の無音区間を削除することで，音声の伸張に伴う遅延を吸収する。したがって，話し終わり時には口と音声にズレ（遅れ）が生じる。また，息継ぎがない場合や短い場合は，遅延を吸収できないことがある。

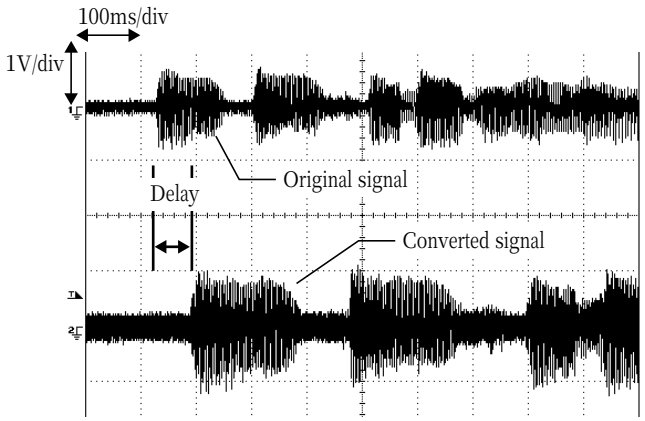

図8　変換波形の原波形からの遅延（例：テレビモード1.6倍伸張時）
Delay between converted signal and original one (e.g. 1.6 times expansion at TV mode)

4.3　ラジオモード時の効果

有声区間倍率2.0倍（はやさ最大），無音区間倍率2.0倍（ま最大）での効果を図9に示す。

発声開始から原波形と変換波形を比較すると，変換波形（有声区間1820ms＋無音区間440ms＋有声区間1840ms）がすべて約2倍に伸張されていることが分かる。このモードでは，無音を吸収しないため，話速変換によって伸張された分だけ，原波形から遅延が生じている。

4.4　テレビモード時の効果

有声区間先頭倍率1.6倍（はやさ最大）の効果を図10に示す。変換波形の無音区間（1440ms）が1360ms削除されており，遅延が吸収されている。同じく有声区間（1380ms）は約1.53倍に伸張され，「ゆっくり」とした音声に変換されていることが分かる。

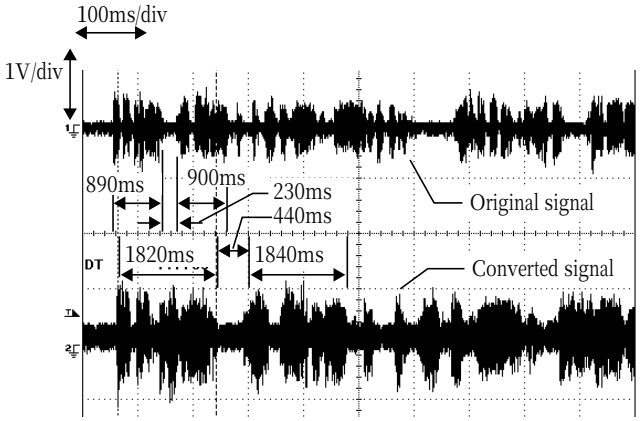

図9　ラジオモード時の効果
（例：有声区間倍率2.0倍，無音区間倍率2.0倍）
Effect of radio mode (e.g. 2.0 times 'expansion of a voiced sound portion, the same of silence portion)

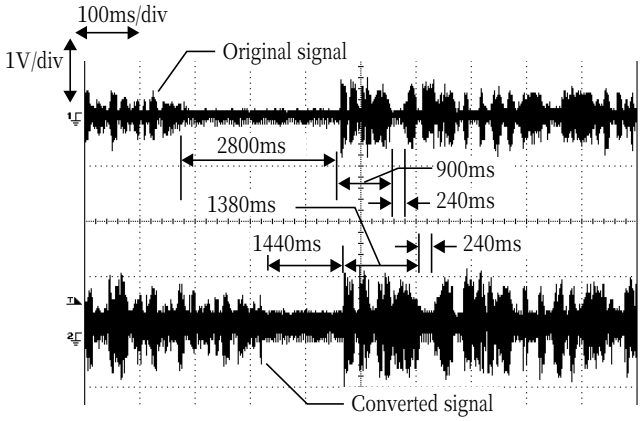

図10　テレビモード（時間吸収モード）時の効果（例：1.6倍伸張時）
Effect of TV mode (e.g. 1.6 times expansion)

4.5　AGCの効果

入出力特性を図11に示す。入力振幅2.8Vで出力振幅1.25Vとなり，以降定常状態となっていることを確認した。これにより，過大入力での音声の歪みを防止する。

5　原理解説

磁気記録または電子化の手段で記録された音声を，速度を変えて再生する場合，いくつかの方法が考えられる。例えば磁気記録された音声の場合は録音時の速度と異なる再生速度

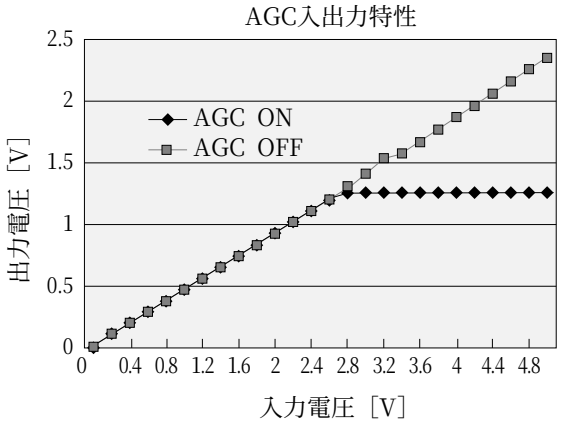

図11　AGC入出力特性
Input-output characteristic of automatic gain control

で再生すれば，音声の速度を早くも遅くもできる。

また，音声が標本化によって電子化されている場合は，再生時の周波数を標本化時の周波数と異なる周波数にすれば，前述と同様の原理で音声の速度を早くも遅くもできる。

ただし，上の原理に基づく話速変換では話速と同時に音声のピッチ（音程）も変わってしまい，用途によってはこれが問題になる。また，原理的に再生音声が不自然になり，著しく音声の明瞭度が低下してしまうという問題がある。

一般に，音声の話速変換に関して言えば話速そのものは変わっても，音声のピッチそのものは変わって欲しくないという要求がある。再生音声のピッチを変えずに話速を変化させる方法として，音声の無音域を 時間的に短縮したり伸長することによって話速を変換する方法がある。ただし，この方法は原理的に話速を早める方向に限界がある。一般的なスピーチ，アナウンス，会話での音声では無音域の全体に対する割合 (パーセンテージ)に限界があり，また，この方法のもうひとつの 問題は再生音声が不自然になることである。

以下に述べる音声の話速変換法は，以上の問題点をすべて克服した， 極めて応用範囲の広い方法である。

人間の声(一部の動物の鳴き音も同様)は肺からの空気の流れと声帯の振動とで生成される。音声には声帯の振動を伴う音声と，声帯の振動を伴わない音声とがある。簡単な分類では前者を有声音，後者を無声音と呼んでいる。ここではひとまず有声音のみを考える。音は言うまでもなく空気の粗密波である。したがって，音声は結局のところ空気の粗密波をつくり出す物理現象である。肺からの空気の流れによって声帯が振動し，その音が喉を通して伝わり，口や鼻を通して外に放出される。これが音声である。

次はスピーチについて考える。ここで言うスピーチとは人間の発する声，または言葉を意味する。スピーチは母音と子音の組み合わせで成り立っている。そして前述の有声音が母音に相当し，無声音が子音に相当する。したがって，スピーチは有声音と無声音とで構成される。話速変換法は基本的に有声音の音響学的特徴を利用してスピーチ の速度を圧縮（Compress）または伸長（Expand）する技術である。

単純な母音を連続的に発して，その音をマイクで拾い，それをオシロスコープで波形を観察すると，似たようなパターンが一定周期で繰り返すのを確認することができる。これが音声のピッチ構造と言われるものであり，基本周波数に相当する。しかし，前述の波形を注意深く観察してみると，繰り返し波形のそれぞれは非常に似ていても，微妙に異なっていることが確認される。これは意図的に一定の音程で母音を発したにもかかわらず生体的な限界から微妙な変化が生じてしまったからである。まして実際に言葉を話すときは音節の変化と同時に母音も激しく変化するので上に述べたことが顕著に現れる。

言葉とは，母音と子音を交互に発しながらピッチも複雑に変化する音響現象である。しかし，声の変化は結局，生体運動の変化であるから，変化時間にはおのずと限界がある。これまでの音声に関する研究成果によればこの時間は5msから20 msのオーダーであることが知られている。つまり5msから20 msの短い時間では，音声のピッチは殆んど変化しないことになる。これが波形観測で認められた繰り返し波形の特徴である。

有声音が同じような波形の繰り返しであるとすれば，繰り返し波形のいくつかを削除したり，あるいは追加したりしても音のピッチに変化は生じないが，スピーチの持続時間が短くなったり，長くなったりするので，結果として話速が変化することになる。つまり繰り返し波形を削除した場合には話速は高速になり，また逆に繰り返し波形を追加した場合には話速は低速になる。

TDHS（Time Domain Harmonics Scaling）はこの原理を用いて音声の圧縮または伸長を行う。TDHSはもともと音声符号化における情報圧縮技術として考えられたが，原理的に話速変換に応用できる。

有声音の繰り返し波形の時間長をピッチと呼び，またこれの逆数を音声の基本周波数と呼ぶ。さらに，音声のピッチを検出する技術をピッチ検出と呼ぶ。ピッチ検出は，音声信号処理において極めて重要な技術である。TDHSは音声のピッ

チ構造を利用した情報圧縮技術であるので，ピッチ検出の精度が重要になる。これの検出精度の低下が直ちに処理後の音質に影響するからである。

ピッチ検出について詳細を述べることは本稿の範囲を越えるので 割愛するが，若干の説明を加える。ピッチ構造をもつ波形，つまり有声音の波形を目で観察しても理解できるが，ピッチ周期を見つけることは必ずしも容易ではない。ましてこれをコンピューターやDSP（Digital Signal Processor）で行うには高度なアルゴリズムが必要である。よく問題になるのは，正しい周期のピッチに対して整数倍のピッチに間違えることである。このような場合，検出可能なピッチの範囲を経験則に基づいて決めておくとよい。また，標本化した信号をそのまま使わずに800Hz相当のカットオフ周波数を持つLPF（Low Pass Filter）でフィルタリングしてピッチ検出することもよく行われる。これはピッチ検出に不要な信号の高域成分を除去してピッチ検出精度を高めるためである。幸い，ピッチを整数倍のピッチに間違えるのは，TDHSアルゴリズムに限ればそれほど重要な問題にならない。例えば，正しいピッチに対して間違って２倍のピッチを検出したとしても，始めからその周期でピッチ構造ができていたと考えれば話速の圧縮または伸長で問題にはならないからである。

音声には有声音と無声音とがあることはすでに述べた。ピッチ検出は有声音に対してはうまく動作するが，無声音に対してはうまく動作しない。もともと無声音にはピッチ構造が存在しないからである。無声音は声帯の振動を伴わない音声なので，雑音に近い性質をもっていることと，相対的に高域成分が多いスペクトラム分布を示すことが知られている。無声音のこの性質のため，ピッチ検出精度の低下が処理後の音質に及ぼす影響は少ない。

TDHSの性能，つまり処理後の音質を大きく左右するのはむしろ周期波形の削除，または付加の操作で生じる歪みである。滑らかに変化するピッチ周期に対して波形素片（１ピッチ分の波形）を強制的に削除したり，挿入したりする理由による。

TDHSアルゴリズムでは上の問題点を解決するのに内挿（interpolation）と外挿（extrapolation）の技術を使う。つまり，削除または挿入する場所と相前後する波形を加味して，ピッチ素片の挿入削除を行い，削除または挿入によって生じる不連続性の歪みを軽減する。具体的にはTDHS伸長アルゴリズム(つまり話速の低減)の場合，N<Mなる条件でＮ個のピッチ素片を基にＭ個のピッチ素片を計算で求めて，結果的にピッチ素片を増加させる。このとき新規に挿入されるピッチ素片を，時間的に前後するピッチ素片を加味して求める。つまり，新規に挿入されるピッチ素片の前半の波形を時間的に先行するピッチ素片で近似し，一方，同波形の後半の波形を時間的に後行するピッチ素片で近似することによって，結果として滑らかに挿入が行われるようにする。

一方，TDHS圧縮アルゴリズム(つまり話速の加速)の場合，N>Mなる条件でＮ個のピッチ素片を基にＭ個のピッチ素片を計算で求めて，結果的にピッチ素片を低減させる。このとき単純にピッチ素片の削除を行うとすでに述べた理由で歪みが発生するので，削除するピッチ素片に相前後するピッチ素片の波形も加工をして，結果的に滑らかにピッチが変化するようにする。つまり，削除の対象となるピッチ素片の前のピッチ素片の後半の波形を削除の対象になっているピッチ素片の前半の波形に近似し，一方，削除の対象となるピッチ素片の後ろに続くピッチ素片の波形を削除の対象になっているピッチ素片の後半の波形に近似する。

以上がTDHSアルゴリズムを使用した話速変換の簡単な原理である。

6 むすび

話速変換技術は，人に優しい技術（バリアフリーの一環）として，本装置のような単体製品，および電話機やテレビ，ビデオ等への組み込み機器として需要が期待される。その場合，①話速変換による音声の聞きやすさ，②低価格，③使いやすさ，④ポータブル性が重視されると思われる。

今後は，更にスピーカ内蔵，ステレオ対応等の機能追加，および機器の小型化／ユーザビリティの向上を図る。

謝　辞

本開発を進めるにあたり，音声データの提供や話速変換の評価，問題点の指摘等をいただいたNHK放送技術研究所ヒューマンサイエンスの宮坂栄一部長，安藤彰男主任研究員，都木徹主任研究員，清山信正研究員他研究所の方々，および大学入試センター小野博教授に深く感謝いたします。

参考文献

1）中村他：“リアルタイム話速変換受聴システム”，NHK技研R&D No.35(1995)

2）小野他：“ 高齢者を対象とした放送の聞き取りの改善 ”, Audiology Japan Vol. 36 No. 5, 1993

3）MALAH：“ Time-Domain Algorithms for Harmonic bandwidth Reduction and Time Scaling of Speech Signals ”, IEEE TRANS. Acoust., Speech, Signal Processing, Vol. ASSP-27, No.2, 1979

4）CHROCHIERE他：“ Real-Time Speech Coding ”, IEEE TRANS. COMM., Vol. COM-30, No.4, 1982

5）MALAH他：“ Performance of Transform and Subband Coding System Combined with Harmonic Scaling of Speech ”, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-29, No. 2, 1981

6）COX他：“ Real-Time Implementation of Time Domain Harmonic Scaling of Speech for Rate Modification and Coding ”, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-31, No.1, 1983

7）岡田他：“ テレビの映像と音声の相対時間差に関する検討 ”, TV学会年次大会(1996)