

## 話速変換技術・音声変換技術の放送および関連ビジネスへの応用

都木徹<sup>†1</sup> 今井篤<sup>†2</sup> 清山信正<sup>†2</sup> 世木寛之<sup>†2</sup> 田高礼子<sup>†2</sup> 田澤直幸<sup>†1</sup> 岩鼻幸男<sup>†1</sup>

放送において音声言語は視聴者に情報を伝達する重要な役割を担っており、音声信号処理を用いて、早口が苦手なお年寄りの聞き取りを支援することや、アナウンサー等の話し手のスキルアップのための訓練装置や語学学習システムなどの技術開発が期待されている。我々は、話速変換や音声変換の研究開発を行い、話速変換技術は、テレビやラジオ受信機の音声聴取補助機能や、インターネットの話速が選べるラジオニュースサービス、スマートフォンの語学学習アプリなどに応用した。また、音声変換の一部として開発したイントネーションやアクセントの分析・変換技術は、PCを用いた語学学習や発声練習を目的に、語学教育番組内や市販の語学学習・発声訓練ソフトウェアとして実用化した。本稿では、それらの技術的特徴と実用化の経緯について述べる。

### Applications to Broadcasting Services and Related Businesses Using Speech Rate Conversion Technology and/or Voice Conversion Technology

TOHRU TAKAGI<sup>†1</sup> ATSUSHI IMAI<sup>†2</sup> NOBUMASA SEIYAMA<sup>†2</sup>  
HIROYUKI SEGI<sup>†2</sup> REIKO TAKO<sup>†2</sup> NAOYUKI TAZAWA<sup>†1</sup> YUKIO IWAHANA<sup>†1</sup>

Spoken language plays an important role to transmit broadcasting information to many viewers. Speech signal processing is expected to assist elderly people who are hard to understand rapid speech and/or to develop any tools for improving announcer's skill and language learning effect. We have researched "Speech Rate Conversion" and "Voice Conversion". The speech rate conversion technology therefore has built into TV and radio receivers and has applied to new internet radio news service capable of selecting its speech rates and a language learning application for smartphones. Furthermore, analysis and conversion technologies for intonation or accent which are partial functions of voice conversion have applied to a language learning TV program and voice training tools with personal computer. This paper introduces their technical features and the ways how they were implemented.

#### 1. はじめに

放送法には NHK の目的として、公共の福祉のために、あまねく全国に放送サービスを行き届かせることが謳われている<sup>\*1</sup>。通常は、これは受信環境において地域格差を生じさせないことと理解されているが、より広く解釈すれば、個々の視聴者の年齢の違いや障がいの有無においても、極力情報格差を生じさせないということでもある。

放送各局においては、番組制作の際、映像や音声を最適に調整し、視聴者にとって見やすく聞きやすい番組として届ける努力をしている。また、視聴覚障がい者のためには、字幕放送や解説放送など、放送システムそのものに補完サービスが可能な機能を付加して対応している。インターネットも放送の補完サービスとして利用している。

放送の音声による情報伝達においては、アナウンサーやキャスター、俳優・声優の存在が必須であり、彼らの音声表現能力が極めて高いため、一般の多くの健聴な視聴者に

とっては何の問題もなく受け入れられているものと考えられる。しかし、加齢や障がいに伴う情報受容能力の低下や、視聴者の個人的な好みまでも考慮すると、実際には万人に受け入れられることは難しいと思われる。

このような状況に対応するには、従来の不特定多数の視聴者を対象とした番組制作の音作りや受信機の音響再生技術だけではなく、音声に特化した信号処理も積極的に応用した新しい放送音声技術とこれを利用したサービスの提供が有効である。さらに、それらの音声信号処理技術が、番組の演出により広がりを持たせたり、放送そのものだけでなく、その他のビジネスにも新たな可能性を与えることができれば、一層好ましいことである。

本稿では、デジタル音声信号処理による、話速変換技術や音声変換の機能の一部であるイントネーションやアクセントの分析・変換技術の開発と、メーカー等と協力して実施した放送や関連ビジネスへの応用について述べる。

話速変換技術においては、早口の聞き取りが苦手なお年寄りの視聴者のためにアナウンサーなどの音声をゆっくりと聞き取りやすくする機能をテレビやラジオ受信機に内蔵したり、過去1日分のラジオニュースをインターネットを介して話速を選んで聞いていただけるサービスを提供したり、最近ではラジオの語学講座のコンテンツを好みの速さで聞くことができるスマートフォンアプリを配信している。

<sup>†1</sup> NHK エンジニアリングサービス  
NHK Engineering Services, Inc.

<sup>†2</sup> NHK 放送技術研究所  
NHK Science and Technology Research Laboratories

\*1 放送法第十五条 協会は、公共の福祉のために、あまねく日本全国において受信できるように豊かで、かつ、良い放送番組による国内基幹放送を行うとともに、放送及びその受信の進歩発達に必要な業務を行い、あわせて国際放送及び協会国際衛星放送を行うことを目的とする。

イントネーションやアクセントの分析・変換技術に関しては、アナウンサーなどの話し手が適切な音声表現を訓練するための支援システムや、テレビの中国語会話において声調を学習するシステムとして利用したり、これらの機能をそのまま市販の発声訓練ソフトや中国語学習ソフトに組み込んだ商品化を行ってきた。

## 2. 話速変換技術

### 2.1 話速変換方式

「話速変換」は声の高さや質を保ちながら、任意の倍率で音声の速度を変換する技術である。1970年代後半には音声の効率的な伝送を目的に TDHS (Time Domain Harmonic Scaling) アルゴリズム<sup>1)</sup>が開発された。その後、品質の向上<sup>2)</sup>や難聴者や高齢者の補聴<sup>3,4)</sup>への応用などを目指し、さまざまな方式が開発されている。

#### (1) 話速変換の基本原則

アナログテープで可変速再生をしたり、デジタル録音再生時に入出力で異なる標準化周波数を用いれば話速は変化する。しかし、同時に声の高さや質も大きく変わってしまう。声の高さや質を維持するためには、音声波形の基本周期を保存する必要がある。

そこで、できるだけ基本周期を正確に抽出<sup>5)</sup>し、基本周期ごとの波形を編集する方式を採用した。その例を図1に示す。この図では、話速を20%だけ速くしたい場合には、5回の基本周期のうち1回を間引き、20%遅くしたい場合には、5回の基本周期に1回繰り返すことを行えばよいことを示している。

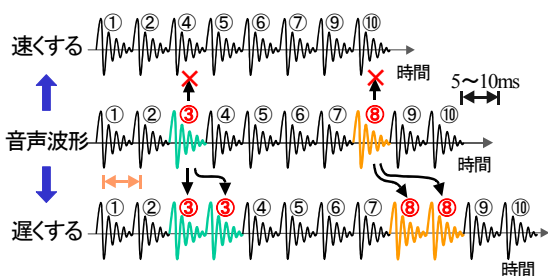


図1 話速変換のための波形の挿入/間引き

#### (2) 適応的話速変換

例えば、音声のどの部分も一様に話速をゆっくりにすると、テレビやラジオの音声に適用した場合、番組の時間枠を越えてしまう。また、電話に応用した場合、相互に会話のタイミングをとることが著しく難しくなる。

そこで、音声の話し始めをゆっくりにし、次の息つきまでに次第に元のはやさに戻すとともに、途中、声が高くなった部分は、一時的にゆっくりにする。そして、“ま”は違和感のない程度に短くして実時間に追いつく。この操作を“ま”で区切られる文単位で繰り返すことで、声のある部分は一切欠落せず、かつ全体として長さが延びていないに

もかかわらず、原音声より“ゆっくりに”聞こえるようになる<sup>6)</sup>。これを模式的に示したのが図2である。

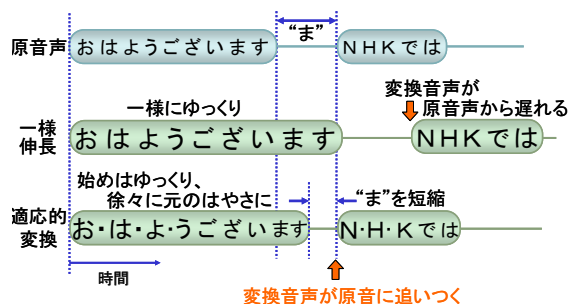


図2 一様伸長と適応的話速変換の動作の模式図

しかし、“ま”が短い場合など、発声の仕方によっては、実時間からの遅れが蓄積してしまうので、その場合は直後の話し始めのゆっくりの度合を減らしたり、早めに元のはやさに戻すとといった操作を加える<sup>7)</sup>。このように、適応的話速変換を行った音声波形の実例を図3に示す。

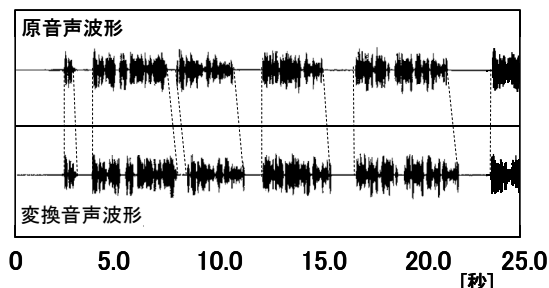


図3 原音声波形(上段)と

適応的話速変換方式で変換した音声波形(下段)

### 2.2 話速変換技術の応用

#### (1) ラジオ・テレビへの応用

適応的話速変換の効果を、60歳以上の被験者約280名により評価したところ、原音声との比較において、60、70歳代の約80%、80歳代以上でも約50%の被験者が、“変換前よりゆっくりに聞こえる”と判定した。加えて絶対的な印象判断で、“聞きやすい”、“落ち着いている”、“自然”などの項目が多く支持され、高齢者に対する有効性が確認できた。

そこで、大手音響機器メーカー<sup>\*2)</sup>と共同で、ラジオやテレビへの導入を積極的に進めた<sup>8)</sup>。基本周波数の抽出精度低下など音質への影響を最小限に抑えたアルゴリズムの簡略化を実施し、固定小数点演算でのDSP化をはかり、まず2002年末にラジオでの製品化を果たした。その後、さらにアルゴリズムの効率化を行い、より低コストでのDSP化を実現して2004年6月にテレビに内蔵した製品も発売された。

これらは、ユニバーサルデザインをコンセプトとした商品として雑誌等でも話題となり、特にラジオの売り上げは

\*2) 日本ビクター株式会社 (当時)  
 現在、当該製品は販売されていない。

順調に推移した。高齢者はもとより、外国語の放送をゆっくり聞いたり、逆に外国出身の方が日本語放送をゆっくり聞くためにも有効である。



図4 話速変換機能を内蔵したテレビ・ラジオ

## (2) インターネットでのラジオニュースサービス

2004年3月末から、NHKのホームページ（NHKオンライン）上で、NHKラジオ第1放送の放送済みのニュースを好みの話速で聞くことができるオンデマンドサービスを開始した（図5）。

通常の話速に加え、1.2倍の遅延伸長の「ゆっくり」音声と、放送時間の0.6倍の時間に適応的話速変換で短縮された「はやい」音声を選ぶことができる。高速音声は一緒に0.6倍にするよりも聞き取りやすい。これによって、従来の放送の枠組みでは実現できなかった話速を選んで番組を楽しむという新しいサービスが可能になった。

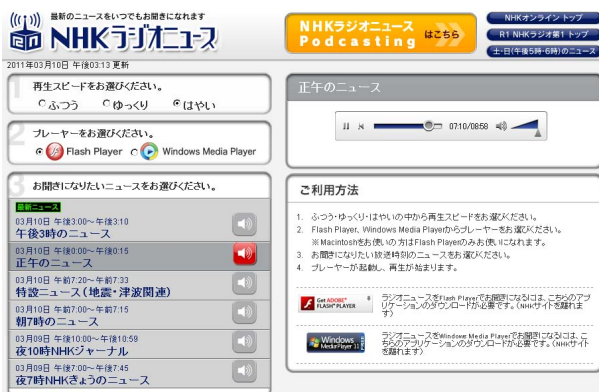


図5 NHKオンラインの話速が選べる「ラジオニュース」

話速変換音声の選択割合は、約3割で、そのうち「はやい」と「ゆっくり」が選択される割合はほぼ同率である<sup>9)</sup>。

「ゆっくり」な音声は、高齢の聴取者のもとより、日本語を勉強する外国人の方々にも利用されていると考えている。また、「はやい」音声は、昼休みや深夜の帰宅後に、実際の番組の6割の短かい時間で聞くことができる。また、後で述べるように、視覚障がい者の方々にとっても、情報を効率的に取得できる極めて有効な機能である。

## (3) ラジオの自動音声放送の時間調整

NHKでは、2010年3月末より、ラジオ第2放送の番組「株式市況」を合成音声で放送している。1億円未満のすべての株価を自然な声質とアクセントで合成する方式<sup>10)</sup>を開発し、東証一部上場の約830の銘柄の終値と前日比を

アナウンサーに代わって自動で読み上げ可能とした。

合成音声を放送で利用するには、十分な品質を確保することに加え、放送時間枠に正確に納めることが必須となる。そのため、この「株式市況自動読み上げシステム」では、話速変換を利用して、残りの時間と銘柄数を考慮しながら各銘柄の読み上げ速度を常に微調整し、秒単位で予定時刻にちょうど終了するための工夫がなされている。

すべての銘柄を読み上げるのに与えられる時間は40分余りであるが、従来の担当アナウンサーには、800を超える銘柄の読み上げをこの時間枠内に納める高度な技術があった。本システムでは、そういったベテランアナウンサーのノウハウも含めて、自動化に成功したと考えている。

## (4) スマートフォンでの応用

現在、広く普及しているスマートフォンで話速変換技術を利用できれば、さまざまな応用が考えられ、より多くの方々にその有効性を確かめていただくと考えられる。そこで、iOSに対応した話速変換ライブラリを作成し、iPhoneアプリとして、(株)NHK出版の「語学プレーヤー」（図6）に導入した。すでにApp Storeで配布されている。

0.5倍速～3.0倍速の間で可変速再生ができ、ラジオ英会話等の語学学習コンテンツを、聞き手の習熟度に応じて自由な速度で聞くことができる。適応的話速変換により、高速再生でも内容が聞き取りやすい。さらに、Podcastingやオーディオブックなどのコンテンツにも対応しており、便利に活用できる。

語学学習では0.8倍速から1.2倍速程度での利用が多い。さまざまなコンテンツが再生できるため、4倍速以上の機能拡張を望む声も多い。

## (5) 視覚障がい者の情報取得の支援技術として<sup>\*3</sup>

視覚障がい者は音声から多くの情報を得ている。書籍や雑誌などの活字情報については、朗読などによって作成された「録音図書」を聴取することが多い。しかし音声は時系列情報であるため文字情報の読み取りにくらべて文章の全体像をつかみにくく、すべての音声情報を聞き終えるまでには膨大な時間を要するため、音声の再生速度を速くして聞くことを希望する人も多い。



図6 iPhoneアプリの話速変換機能付「語学プレーヤー」（NHK出版のHPより）

\*3 本研究開発は、独立行政法人情報通信研究機構「平成20年・21年度高齢者・障害者向け通信・放送サービス充実研究開発助成金」「平成22年・23年度高齢者・チャレンジド向け通信・放送サービス研究開発助成金」により実施したものである。

ところが、従来の録音図書再生装置やソフトウェアは高速再生に特化した聞きやすさについては特に考慮されておらず、2倍速を超えると聴きづらくなるため、思うように聴取速度を上げられないのが現状である。そこで、適応的な話速変換を適用することで、従来の高速再生方式と同じ速度倍率ならば、より楽に聞き取ることができ、結果として従来より高速での聴取が可能となるよう、研究開発を行なっている。

視覚障がい者の方々に、録音図書やラジオの朗読番組・ニュース番組を2倍速から3倍速に変換した音声の評価実験を行っていただいた結果、その約8割が、一様に短縮した場合よりも適応的な話速変換による高速再生を聞き取りやすいと回答した。

具体的な応用の対象としては、視覚障がい者が利用するDAISY<sup>\*4</sup>規格の録音図書再生装置や、パソコン版プレーヤー、iPhone/iPad向けアプリケーションへの適用を目指している。図7は、試作した0.5～5.0倍速の間で可変速再生が可能なパソコン版プレーヤーの画面である。



図7 試作したDAISY規格のパソコン版プレーヤー

### 3. 音声変換技術

#### 3.1 イントネーションやアクセントの分析・変換

本稿で「音声変換」と呼ぶのは、男女声の違いや個人ごとの声色・響きの違いなど、いわゆる“声質”に加え、韻律を構成する要素である“イントネーション”、“アクセント”、2章で扱った“話速”などを独立にまたはいくつか組み合わせる技術の総称とする。変換を行う際には、まず対象とする音声を分析し、さまざまな要素に相当する物理的特徴量を抽出したのち、各特徴量に所望の変更を加え、再度合成するという手順を踏むことが多い。

この技術に関しては、これまでにさまざまな方式<sup>11)-14)</sup>

が開発されており、比較的処理が軽く実用システムに導入しやすい方式から、処理は重いが高品質な変換音声を得られる方式など、それぞれに特徴がある。

#### (1) 利用分野

音声変換の利用目的は、テキストを音声に変換する音声合成や、感情などの音声表現の変換、音声知覚の基礎研究などさまざまである。

そのような中で、近年、語学学習への応用も盛んである。このコンピュータ支援型語学学習(CALL: Computer Assisted Language Learning)は、音声認識なども利用し、発音・韻律の学習やリスニング、会話、読解支援などに関するものが開発されている<sup>15)</sup>。

我々も、3.2や3.3で述べるように、文献12の音声変換技術を用いて、日本語など声の高さの高低の変化が意味理解に大きく影響する言語の、イントネーションやアクセントを学習するCALLシステムの開発を行ってきた。

#### (2) イントネーションやアクセントの変換のポイント

図8にイントネーションやアクセントを変換する際の基本的な操作手順を示す。分析された自然音声のピッチ軌跡(音声の基本周波数の時間変化パターン)には、微少なゆらぎ成分と緩やかに大きく変化する成分とが混合している。図ではわかりやすいようにゆらぎ成分が全区間に現れているが、実際は、発声開始・終了や、発声中の子音部分などにおいて、声帯の振動が安定しない場合に見られることが多い。この微少なゆらぎはイントネーションやアクセントの知覚には寄与しない。

そこで本来のピッチ軌跡に対して、カットオフ周波数5～10Hz程度で平滑化したものをイントネーションやアクセント成分と考え、ゆらぎ成分と分離する。平滑化したピッチ軌跡を希望の形に変形し、ゆらぎ成分はそのまま保存する。変形された平滑化ピッチ軌跡とゆらぎ成分を加えることにより、実際の変換音声のピッチ軌跡を生成する。このような手順をとることで、入力音声のピッチ軌跡の各部分に対する変化量を必要最小限に抑えることができ、変換音声の品質を向上させることができる。

なお、この変換の対象としているのは、声帯の振動を伴った音声区間(これを「有声音区間」と呼ぶ)のみである。

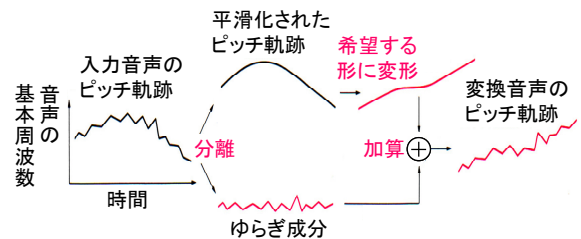


図8 ピッチ軌跡変換のための操作手順

\*4 Digital Accessible Information SYstem: デジタル録音図書の国際標準規格。視覚障がい者が利用する録音図書はこの規格に則って制作されているものが多い。規格の開発・維持は、国際非営利法人DAISYコンソーシアムが行っている。

### 3.2 中国語の声調学習システムへの応用

NHK では、3.1 の(2)で示したピッチ軌跡の変換手法を応用して、声調(中国語のアクセント)教育を支援する CALL システム<sup>16)</sup>を開発し、2000 年度から 2002 年度までの 3 年間、教育テレビの語学番組「中国語会話」で使用した。なお、番組内では、この CALL システムを「声調式号」(2000-2001 年度)、「声調参考」(2002 年度)と命名した。図 9 に声調参考の画面の一例を示す。

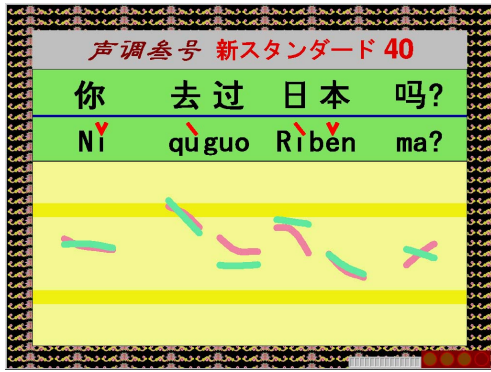


図 9 “声調参考”の画面の一例

模範音声(ピンク)と学習者の音声(水色)の両者のピッチ軌跡を比較表示し、学習を視覚的に支援するとともに、学習者の声質で模範音声のピッチ軌跡を持った変換音声(矯正音声)を聞くことができ、聴覚的にも新しい学習が可能である。

#### (1) 声調参考の仕組み

図 10 にシステム全体の構成を示す。システムは全て Windows パソコン上に構築されており、機能は主に以下の a)~c)の 3 つの部分から構成される。

- a) 中国人レギュラー出演者が発声した音声(以後、“モデル音声”と呼ぶ)を基に、予めピッチ軌跡を抽出して作成しておく声調学習用データベース(スタンダード 40 データベース)
- b) 日本人出演者が発声する音声(以後、“ユーザー音声”と呼ぶ)のピッチ軌跡をその場で抽出して、モデル音声のものと比較表示する。
- c) ユーザー音声の基本周波数、パワー、音の時間長を、モデル音声に合わせて時々刻々変換した合成音を生成する。

番組講師の指導のもと、開発で留意した点としては、①声調を理解しやすいピッチ軌跡の表示法として、平滑化だけでなく、中国語では音節毎に声調が付与されることから 1 音節ごとに区切ることにした。②モデル音声とユーザー音声の声調の比較を容易とするため時間方向及び声の高さ方向で正規化を行い、同じ音節で重なるようにした。③ユーザー音声のピッチ抽出等の分析誤りを、モデル音声との位置照合により自動補正するようにしたことなどである。

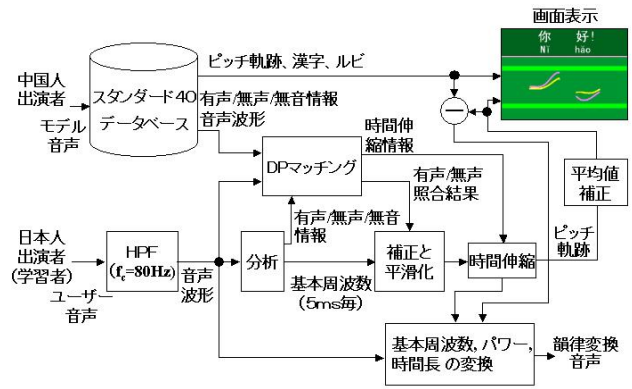


図 10 “声調参考”のシステム構成

#### (2) 商品化

“声調参考”は番組中で安定に動作し、番組の演出および CALL システムとしての効果も番組関係者・視聴者から評価された。これを受け、中国語や韓国語の語学学習システムを幅広く手がけているソフト開発・販売メーカーへの技術移転が実現し、“声調参考”の声調分析・表示機能を導入した製品<sup>\*5</sup>が 2005 年に販売開始され、現在に至っている。

### 3.3 アナウンストレーニングの支援システムへの応用

日本語の規範を NHK の表現者、特にアナウンサーに求める声が強い。アナウンサーは研修をはじめ、現場での日常の実務を通じて、正確な表現、発音を目指してトレーニングを続けているが、より効率的で充実したトレーニング法が求められている。

そうした背景の中、我々はアナウンサーとともに、音声表現育成トレーニングの方法に関して、コンピューターを利用して、より科学的かつビジュアルにトレーニングできるシステムの研究を進めてきた。

#### (1) アナウンスクリニックシステムの開発

そしてある程度研究が進んだ 2001 年度から、アナウンス室が中心となり、パソコンのソフトウェアとして、放送現場で音声表現に関わる職員・スタッフ等が、ひとりで発音トレーニングを行なうことができる“アナウンスクリニックシステム”の開発を開始した。

このシステムにおいては、イントネーションやアクセントの分析・変換機能だけでなく、母音の無声化の点検機能や五母音の発声訓練機能も付加した。

図 11 は、同システムの画面の一例であるが、模範音声・学習者音声の波形の下のバーは、有声(水色)／無声(ピンク・緑)／無音(青)を表しており、母音の無声化を確認することができる。赤の実線は模範音声ピッチ軌跡、青の実線は学習者のピッチ軌跡である。最下断は、各人のパワー(声の強さ)の軌跡である。

図 12 は、五母音の発声訓練の画面であり、横軸に第 1

\*5 株式会社高電社の“ChineseWriter”シリーズ

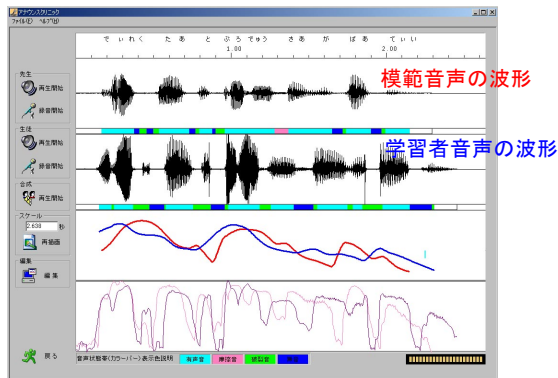


図 11 “アナウンスクリニック”の画面の一例

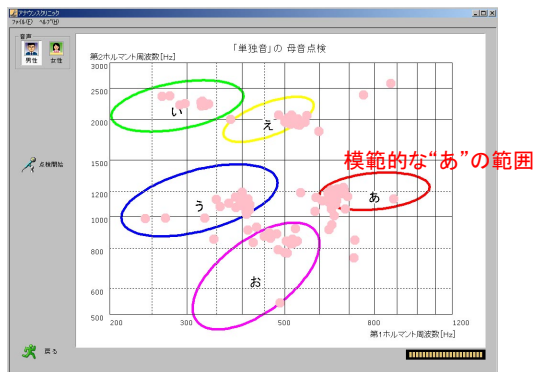


図 12 五母音の発声訓練の画面

ホルマント周波数、縦軸に第2ホルマント周波数を取り、楕円は模範的な各母音の範囲を示している。学習者が発声した母音をリアルタイムで分析して、ピンクの○で表示するようになっていて、学習者は○が楕円の中に入るよう口の構えや舌の位置を調整して練習する。

(2) 成果の還元

アナウンスクリニックシステム開発の成果を、社会に還元する施策として、NHK 出版と連携して商品化を図った。

まず、多くの音声表現者の必携の書である、「NHK 日本語発音アクセント辞典」について、2002年にCD-ROM版を発行した(図13左、アクセントトレーニングの画面)。さらにアナウンスクリニックシステムとほぼ同様の機能を搭載した、CD-ROMブック「NHK アナウンス実践トレーニング」(図13右)が2005年に発行された。どちらもパソコンにインストールして、発声練習が可能となっている。



図 13 アナウンストレーニング教材としての商品化

4. おわりに

話速変換技術や音声変換技術を利用した、音声情報受容を支援するシステムや、語学や発声のトレーニングシステムの開発と応用について述べた。

実用化のためには、技術者としての視点だけでなく、アナウンサーを含む番組制作者や、高齢者・視聴覚障がい者など、ユーザーとの密接な連携が必要である。今後とも音声信号処理技術のシーズと音声表現を駆使する放送分野のニーズとをマッチさせる研究開発に努力していきたい。

**謝辞** 視覚障がい者向けの高速話速変換の研究にご協力いただいている慶応義塾大学 SFC 研究所の鳥原信一氏、主観評価実験にご協力いただいた横浜市視覚障害者福祉協会の方々に感謝申し上げます。

参考文献

- 1) D. Malah: Time-Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals, IEEE Trans. Acoust. Speech and Sig. Proc., Vol.ASSP-27, No.2 (1979)
- 2) 森田直孝, 板倉文忠: 自己相関法による音声の時間軸での伸縮方式とその評価, 信学技報, EA86-5 (1986)
- 3) Y. Nejime, T. Aritsuka, T. Imamura, T. Ifukube, and J. Matsushima: A portable digital speech-rate converter and its evaluation by hearing-impaired listeners, Proc. ICSLP 94, pp.2055-2058 (1994)
- 4) 中村章, 清山信正, 池沢龍, 都木徹, 宮坂栄一: リアルタイム話速変換型受聴システム, 音響誌, Vol.50, No.7, pp.509-520 (1994)
- 5) 都木徹, 清山信正, 宮坂栄一: 複数の窓幅から得られた自己相関関数を用いる音声基本周期抽出法, 信学論, Vol.J80-A, No.9, pp.1341-1350 (1997)
- 6) 池沢龍, 中村章, 清山信正, 都木徹, 宮坂栄一: 話速変換に伴う時間伸長を吸収するための一方法, 信学技報, SP92-56 (1992)
- 7) 今井篤, 池沢 龍, 清山信正, 中村章, 都木徹, 宮坂栄一, 中林克己: ニュース音声を対象にした時間遅れを蓄積しない適応形話速変換方式, 信学論, Vol.J83-A, No.8, pp.935-945 (2000)
- 8) A.Imai, T.Takagi, and H.Takeishi: Development of radio and television receiver with functions to assist hearing of elderly people, IEEE Trans. Consumer Electronics, Vol.51, No.1, pp.268-272 (2005)
- 9) 今井篤, 都木徹, 黒住幸一, 小山隆二, 島津敏雄: 話速変換技術を利用したインターネットでのラジオニュースサービス, 映情学誌, Vol.59, No.2, pp.265-270 (2005)
- 10) 世木寛之, 清山信正, 田高礼子, 都木徹, 大出訓史, 今井篤, 西脇正通, 小山隆二: 高品質な株価音声合成装置の開発とデジタルラジオ放送での試験運用, 映情学誌, Vol.62, No.1, pp.69-76 (2008)
- 11) 阿部匡伸, 田村震一, 桑原尚夫: FFT スペクトルからの Signal Reconstruction による音声変換手法, 信学技報, SP88-48 (1988)
- 12) 都木徹, 梅田哲夫: ピッチ変更時のひずみをスペクトル領域で修正する声質変換方式とその品質の心理評価, 信学論, Vol.J73-A, No.3, pp.387-396 (1990)
- 13) E.Moulines and F.Charpentier: Pitch-Synchronous waveform processing techniques for text-to-speech synthesis using diphones, Speech Communication, Vol.9, Nos.5/6, pp.453-467 (1990)
- 14) 河原英紀: 聴覚の情景分析と高品質音声分析変換合成法 STRAIGHT, 音講論集, 1-2-1, pp.189-192 (1997)
- 15) 中川聖一, 牧野正三, 壇辻正剛: 音声言語処理技術を用いた語学学習システム, 音響誌, Vol.59, No.6, pp.337-344 (2003)
- 16) 都木徹, 服部有希子, 小宮恵, 今井篤, 岸憲史, 伊藤崇之: 韻律の視覚化及び矯正音声を用いた中国語学習システム, 信学論, Vol.J88-D- I, No.2, pp.478-487(2005)