韻律に寄与する音響特徴量を用いた聞きやすい高速話速変換技術

Intelligible High-speed Playback Technology Using the Acoustic Features of Speech Prosody

正会員 今井 篤^{†1,†3}, 田澤直幸^{†2}, 岩鼻幸男^{†2}, 正会員 都木 徹^{†1},
 正会員 清山信正^{†1}, 田中敏明^{†3}, 伊福部 達^{†4}
 Atsushi Imai^{†1,†3}, Naoyuki Tazawa^{†2}, Yukio Iwahana^{†2}, Tohru Takagi^{†1},
 Nobumasa Seiyama^{†1}, Toshiaki Tanaka^{†3} and Tohru Ifukube^{†4}

Abstract We have developed an intelligible high-speed speech rate conversion technology using the acoustic feature quantities that contribute to prosody. In contrast to the conventional method, which plays back accelerated speech at the same uniform rate from the beginning to end, our proposed approach varies the playback rate adaptively on the basis of acoustic detection of the position of an utterance and any fluctuations in a speaker's fundamental frequency (F0) and power. In so doing, we hope to make high-speed playback easier to listen to by providing the listener with a "slowed-down" playback effect. Since this approach converts speech rate using just the acoustic features of audio data, it can be applied to not only Japanese but other languages as well. While the algorithm we developed in this study is optimized for the Japanese language, we aim to implement the proposed approach in a wider array of commercial devices and customize the technology to various languages.

キーワード：高速音声，話速変換，速聴，視覚障害者，DAISY，オーディオブック

1. ま え が き

新聞や雑誌の記事をすべて精読するという人はまずいない。内容の興味に応じて、精読したり斜め読みをしたりするのが一般的である。これがオーディオブックなどの活字本を音声化したコンテンツを楽しむ場合にも求められている。通常の再生だけでなく、斜め読みに相当する聞き方も選びたいというものである。このような要望は、音声情報への依存度が高い視覚障害者からは以前より聞かれたが、ここ数年、小説やビジネス書等を音声で聞くオーディオブックの急速な普及に伴って、一般の人からも同様の声が

聞かれるようになってきた。

オーディオブックのような録音音声聞くことを目的に開発された機器には、例えば、視覚障害者の多くが使用するDAISY (Digital Accessible Information System) 規格の録音図書再生器がある¹⁾。これには3倍速程度までの話速変換機能が備わっており、長時間のコンテンツを効率的に聞くための有効な手段となっている。しかし実際のユーザ30名へのアンケート調査(2009年2月15日に神奈川県視覚障害者福祉協会福祉大会にて筆者らが実施)によると、高速再生技術に関して現状で満足だという人は意外と少なく、「点字ならいくらかでも速く読めるが、音声はすぐに聞き取れなくなる」、「一冊聞き終えるのに時間がかかりすぎる。しかし速く再生すると内容がよく聞き取れない」、「高速音声をよく聞いているが、とにかく疲れる。もっと気楽に聞きたい」など、高速音声の聞きづらさに関する不満が多く寄せられた。

合成音声を用いて、視覚障害者の速聴能力を調べた研究によると、1,100 mora/min*では提示内容の90%以上が聞き取れるが、1,400 mora/minあたりでは50%程度しか聞き取れなくなると報告されている²⁾。これを、日本語の一般的な話速である8 mora/secを基準に倍率換算すると、2倍速くらいまでは、ほぼすべてを聞き取ることができるが、

2011年12月22日、冬季大会で発表

2011年12月23日受付、2012年5月10日再受付、2012年6月7日採録

†1 NHK 放送技術研究所

(〒157-8510 世田谷区砧1-10-11, TEL 03-5494-3370)

†2 財団法人NHKエンジニアリングサービス

(〒157-8540 世田谷区砧1-10-11, TEL 03-5494-2400)

†3 東京大学 先端科学技術研究センター

(〒153-8904 目黒区駒場4-6-1, TEL 03-5452-5382)

†4 東京大学 高齢社会総合研究機構

(〒113-8656 文京区本郷7-3-1, TEL 03-5841-1524)

†1 NHK Science and Technology Research Laboratories

(1-10-11, Kinuta, Setagaya-ku, Tokyo, 157-8510, Japan)

†2 NHK Engineering Services, Inc.

(1-10-11, Kinuta, Setagaya-ku, Tokyo, 157-8540, Japan)

†3 Research Center for Advanced Science and Technology, the University of Tokyo

(4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan)

†4 Institute of Gerontology, the University of Tokyo

(7-3-1 Hongo Bunkyo-ku, Tokyo 113-8656, Japan)

* 1 moraは、短母音を含む1音節の長さに相当し、日本語では、ほぼ、かな1字(拗音では2字)がこれに相当する。撥音や促音も含まれる。

3倍速あたりになるとかなり厳しくなるということである。視覚障害者は、一般の人に比べて速聴能力が高いと言われているが、2~3倍速程度が聞きやすくなればうれしいという声は少なくない。また、オーディオブックユーザなど、視覚障害者以外にも高速再生が使いやすいものになることが期待できる。

そこで本稿では、自然発話の音声を対象にした高速音声の聞きやすい適応形の話速変換方式を提案する。一様に話速を上げるのではなく、聞き取りに大切だと思われる部分を相対的にゆっくり再生することを意図している。ここで、「大切と思われる部分」の程度は、言語解析のような知識処理から導かれるものではなく、あくまで発話の音響的な特徴のみを根拠に序列化したものである。知識処理の利用については、音声に対応したテキスト情報の取得や、その解析などに伴って課題が複雑化するため、本稿の検討範囲には含めず、今後の課題とした。高速再生は、予め録音されたものを対象に行うという必然性があることから、事前にコンテンツの中身を解析しておくことができる。そこで、韻律に寄与する音響特徴量を予め分析し、これを付加情報として記述したデータファイルを作成した。これを話速変換の際に参照することで、意図した部分を正確に話速変換できるだけでなく、ファイル全体を俯瞰した上での無駄のない緩急の付与を可能にした。手法の評価は、日常的に高速音声を利用している視覚障害者で行い、その多くが使用しているDAISY規格の録音図書再生装置(DAISYプレーヤ)の高速再生方式を比較対象とした。また、音響情報のみを話速変換の際の手がかりにすることで、多様な発話様式に柔軟に適応できると考え、日本語だけでなく外国語への適応可能性についても検討を行った。その結果、日本語だけでなく、英語についても高い有効性を確認することができた。

2. 高速に再生された自然発話音声を聞きやすくする話速変換方式

高速に音声を再生して効率的に情報を取得するための研究としては、テキスト音声合成を対象とした技術が提案されている。対象テキストを言語解析し、予め設定された言語的な重要度に基づいて、合成前に情報量を減らす方式である³⁾。

一方、本提案方式は合成音声ではなく、自然発話の音声を対象とする。発話の音響的な特徴の変化(韻律)に応じて話速に緩急をつける適応形の話速変換方式である。一般に再生速度が上れば上がるほどコンテンツ全体が聞きづらくなるが、話速の緩急による情報の取捨選択により、聞き取りに大切そうな部分の聞き取りをメリハリつけて確保するというのが動作のねらいである。課題は音声中のどの部分を選択することが高速再生時の聞き取りに有効であるかということである。

適応形の話速変換方式については、高齢者の聴力補助を目的として、放送音声を時間遅れなくゆっくりと感じられ

るように変換する方法がこれまでに提案されている⁴⁾。また、その高速再生応用についても有効性が報告がされている^{5) 6)}。この方式は、無音区間の一定時間までの短縮と固定関数による音声区間の緩急を特徴とするが、音声が入力された時点での逐次処理が前提であるため、未来のデータを加味した話速変換ができないという課題があった。また、適応変換のルールについても、アナウンサ数名のニュース発話の音響的な特徴をモデルに関数化されたものであり、ニュース以外の多様な発話スタイルへの適応については言及されていない。そこで本稿では、関数ではなく発話の特徴に応じて、適応的に緩急を与える適応形話速変換方式を提案する。事前に音声ファイルを分析して、各音声の属性やその出現時刻などの情報をデータ化し、それを参照しながら、指定の再生時間内で高速音声の聞きやすいように音声区間と休止区間を効果的に再配置することを特徴とする。従来法に対する直接的な効果の比較は今後の検討課題とするが、特に高速再生においては、参照データに基づいて音声区間の再生時間を従来法よりも安定的に長く確保できることから、聞きやすさにおいて有利に働くものと考えられる。

2.1 メタデータの作成

韻律に寄与する音響特徴量として、コンテンツ毎に、音声、非音声の分類、また有声音、無声音、無音などの音声属性情報と、基本周波数の抽出、音圧(パワー)の変化を分析して、これらをテキストデータとして保存した。本稿ではこれをメタデータと呼び、話速変換処理の際の参照情報とした。メタデータは事前に分析できる情報であれば何でもよいが、今回は音響情報のみとした。音響情報だけを扱うことで、発話のバリエーションや言語の違い、文語体や口語体などを一元的に扱うことができるからである。

音声分析は、5msごとに、パワー、ゼロ交差数、自己相関分析を組合せて、音声、非音声の判定、また有声音、無声音、無音の分類と基本周波数の抽出を行った⁷⁾。パワーも同じく5ms毎に定義するが、計算は20msのフレーム単位で行い、サンプル値の2乗和をフレーム長で除算して正規化したものを用いた。以上の値をコンテンツ毎にメタデータとして記述した。

2.2 アルゴリズム

メタデータを参照して伸縮率波形を作成する方法を述べる。本アルゴリズムは声の高さと大きさのそれぞれの時間変化に着目し、この二つのパラメータの組合せから適応的な話速変換倍率を導出した。この手法は、発話の音響的な特徴の変化から、発話意図が存在する部分と、その程度を動的に推定して、それに応じた話速の緩急を付与することを意図している。

音声表現に関しては、特に日本語は音調によって発話意図を表現することが知られている。中でも上昇音調は何かの意味があるときに見られるとされており、話のニュアンスや意図、話者の態度などによって多様に変化するとさ

(a) 基本周波数とパワーの変化に基づく適応的な話速変換倍率の導出

(b) (a)で導出した話速変換倍率波形の混合例

図1 予め分析された音響情報に基づく適応的な話速変換倍率関数の導出

A process for developing adaptive speech rate conversion function Using pre-analyzed acoustic information, (a) a process for developing adaptive speech rate conversion function Using pre-analyzed F0 frequency and power contours, (b) a sample of combined speech rate conversion function developed by (a).

れている⁸⁾。また、プロミネンス(卓立)として並列句に対して意図的に音程の差を付けたり、声の強さを変えたりすることがあり、声の高さ、特に上昇音調と、声の強さの組合せでさまざまな発話意図が表現されていると考えることができる。筆者らによるNHKアナウンサーへのインタビュー調査でも、発話意図は声の高さ、強さ、間合いで表現することが多く、特に声の高さを上げる場合は、何らかの伝えるべき情報があるということであった。また、話速変換技術に関する先行研究において、声の高さが特に上昇する部分を局所的にゆっくり再生することで、発話全体の聞きやすさの向上、特にゆっくり感に効果的であることが報告されている⁴⁾。以上から、声の高さが大きく上昇し、かつ大きさが大きい部分の聞き取りを確保することが、高速音声の聞きやすさに効果的に作用すると仮定し、声の高さと大きさの逐次変化を、これを動的に検出する手掛かりとした。次節では、これをもとに話速変換倍率波形を生成する方法を述べる(図1)。

2.3 適応的な話速変換倍率の生成

基本周波数の変化とパワーの変化を話速変換倍率の変化

に見立てて、それらを適当な重みで加算する計算式を作成する。一般的な会話音声においては基本周波数の変動が10Hz以下になるとされている⁹⁾。そこで、2.1節の分析により算出した基本周波数波形をカットオフ周波数10Hzで低域ろ波して平滑化し、一有声音区間の最後の数フレームと次の有声音区間の最初の数フレームのデータを用いて非有声音区間をスプライン補間した。感覚量との対応のため、これを対数変換した波形の中央値を1として振幅を0~2に正規化した。パワーについても同様に、2.1節の分析によって算出した波形をカットオフ周波数6Hzで平滑化した後、対数変換し、その中央値を1として振幅を0~2に正規化した。パワーのカットオフ周波数は、無音区間における値が完全には0にならない程度(つまり倍率関数として用いたときに無音区間の伸縮率が完全に0にならないように)に筆者自身が適当に設定したものである。

以上により求めた、基本周波数に基づく倍率波形を P_t 、パワーに基づく波形を P_o 、 P_t の寄与度を a 、 P_o の寄与度を b とすると、 P_t と P_o の寄与度と動きを総合的に加味して話速変換倍率を与える式を、次のように定義した。

$$P_t^a \times P_o^b \quad (a > 0 \quad b > 0) \quad (1)$$

一例として、 $a=0.5$ 、 $b=0.5$ (P_t の寄与度50%、 P_o の寄与度50%)のときの伸縮率波形は、図1(b)のSr1のようになり、 $a=0.7$ 、 $b=0.3$ のときSr2のようになる。最後に、この混合波形で与えられる伸縮率全体を線形補正して所望の話速変換倍率を導出する。

以上の手続きによって生成された倍率は基本周波数、およびパワーのいずれも高いほどゆっくり変換されるが、逆にいずれも低いような場所には、聴感上削除したように聞こえるほど短縮される部分も出てくる。つまり、一様に変換する従来法よりも高速に再生される部分も発生するということである。

2.4 多様な発話様式への適応

P_t と P_o の混合波形の作り方は2.3節で述べたとおりであるが、それぞれの適切な寄与度について実験的に調べた。予備実験で日本語のほか英語、独語について2~3倍速の高速音声の聞き取りやすさに関して調べたところ、好ましい混合率が言語によって異なる傾向が見られたからである。言語によるイントネーションやアクセントの役割については種々論じられているが、ここでは音声言語学の観点での考察は行わず、 P_t と P_o の音響的な寄与度に関してのみ最適化を行った。

(1) 日本語適応について

両パラメータの寄与率の最適化の根拠は、高速再生が聞きやすくなることにある。まず、日本語について両パラメータの好ましい寄与率についての評価実験を行った。実験音声は、①視覚障害者用DAISY録音図書から1編、②一般向けオーディオブックから2編、③NHKニュースから2編、計5編から作成した。両パラメータの混合割合は、 P_t 側から100%(P_o :0%)、75%(P_o :25%)、50%(P_o :50%)、

図2 PtとPoの好ましい寄与率の調査 (3倍速再生：日本語)
Investigating adequate ratio of Pt and Po for making mixture function (3 times normal speed: in Japanese).

図3 PtとPoの好ましい寄与率の調査 (3倍速再生：フランス語)
Investigating adequate ratio of Pt and Po for making mixture function (3 times normal speed: in French).

25% (Po：75%)，0% (Po：100%) の5水準とした。各実験音声はこの5種類で変換されたが，すべて同じ音声の1種類ではなく，各実験音声の中の異なる場所から5カ所選択して1水準あたり5種類の音声を用意した。時間はいずれも30秒程度である。再生速度は2倍速および3倍速の2種類とした。

実験は，従来法（一様に時間伸縮する方法）による変換音声との一対比較で，従来法に対する聞きやすさを5段階，「1. 非常に聞きづらい」，「2. 聞きづらい」，「3. どちらともいえない」，「4. 聞きやすい」，「5. 非常に聞きやすい」で評価した。被験者は聴力に問題のない大学生5名である。今回の実験で従来法とした話速変換手法は，提案法の話速変換倍率を一定値にしたものであり，音質は提案法とまったく同じである。3種類の実験音声間で特に傾向の差が見られなかったため，すべてのデータを合算集計した。3倍速の実験結果を図2に示す。一評価点の評価回数は25である。

これによると，有意な差ではないが，日本語の高速再生ではPtの寄与率をやや高め ($a=0.75$ $b=0.25$) に設定することが好ましいという傾向が示された。この結果は，2.2節で述べた日本語の言語的特徴を反映していると考えられる。

2倍速の結果は，寄与率間で有意な差が見られなかったため図は割愛するが，内観調査によると，2倍速の場合は，従来法でも内容が十分に聞き取れてしまうため，評価が難しいということであった。しかし，聞きやすさの観点では，寄与率に関わらず提案法が従来法に比べて平均的に高い評価を得ていた。本実験結果は，寄与率を同程度くらいに設定すれば良いということ過ぎないが，システム設計においてはPtの寄与率を高めた上記の値を日本語の高速話速変換時に適用することとした。

(2) 外国語適応について

外国語適応の一例として，フランス語で実験した結果を図3に示す。実験方法は日本語の場合と同様であり，被験者は20歳代のフランス語を母語とするフランス人2名である。実験音声には語学教材の朗読本を用いた。被験者数が

少ないため，試行回数を多くした。一評価点の評価回数は10である。

フランス語の場合はPtとPoの寄与率均等 ($a=0.5$ ， $b=0.5$) が最も好ましいという結果になった。これ以外にも，ドイツ語，英語について，それぞれネイティブスピーカー2名による予備的な調査を行った。実験手続きはフランス語の場合と同じで，用いた実験音声はそれぞれの言語のニュース音声である。PtとPoの寄与率については，他の言語で低評価だった各パラメータの両端を除外して，Pt側から90% (Po：10%)，70% (Po：30%)，50% (Po：50%)，30% (Po：75%)，10% (Po：90%) とした。一評価点の評価回数は4である。その結果，ドイツ語，英語ともに， $a=0.3$ ， $b=0.7$ が総合的に好ましいという傾向が見られた。実験条件が異なるので単純な比較はできないが，これは日本語における好ましいPtとPoの寄与率とは逆の傾向である。このように，言語によって好ましい寄与率の傾向が異なる可能性が示されたが，これは，言語によって，発話意図の表現方法が異なることがその理由の一つと考えられる。例えば，日本語ではイントネーションが，英語ではアクセントや強勢がその手段として使われることが一般に知られているが，今回の実験結果からもその傾向が伺える。

2.5 休止区間の短縮について

高速再生時に最低限必要な休止区間の時間長について調べた。アナウンス発話を対象に文中の休止区間の短縮許容限を調べた先行研究によると，約200msまでは聞こえの自然さが保たれるとされており⁴⁾，この値を目安に高速音声の聞き取りにおいてどのように変化するかを調べた。200msまで休止区間が短縮された音声を，一様に3倍速に話速変換すると約70msとなるため，その前後の0～180msの間で，30ms単位で休止区間を設定した。被験者への指示は，「なるべく短時間に効率的にコンテンツの内容を聞きたいという目的に照らして，音声全体の印象の好ましさを5段階で評価せよ」とした。実験音声はニュース音声と朗読音

図4 高速音声の休止区間の短縮許容限
Allowable limit value of curtailing pause within the fast speech.

声の30秒程度から作成し、それぞれ2倍速、2.5倍速、3倍速の3種類の話速で実験を行った。上述の先行研究との比較のため、話速変換は一樣に伸縮する従来法を用いた。被験者は、2.4節と同じ大学生5名である。

今回調べた3種類の話速、および実験音声間による傾向の違いは見られなかったため、すべてのデータを合算集計して3倍速の結果のみを示した。結果を図4に示す。これによると、今回、実験条件として設定した休止区間の範囲では、30ms以上はいずれの話速ともほぼ一定の評価値となった。一般に、呼気段落の冒頭ではパワーが大きく、末尾に向かって次第に小さくなっていくため、呼気段落間の休止区間が短い場合、小さな声から大きな声に急激に変化した場合に、逆行性の継時マスキングを生ずる可能性がある¹⁰⁾。30msという時間はその影響が排除できる閾値付近に相当しており、本実験結果からは、高速に音声を書くという目的に対しては、最低限、声の切れ目が分かればよいというような結果となった。通常の音声を聞くという行為とはまったく異なるが、斜め読みと精読が異なるように、目的に応じた情報提示の好ましさがあると考えることができる。したがって、実用的な2倍速以上の高速話速変換のシステム設計において、休止区間を30ms以下には短縮しないことを条件とした。

3. 高速話速変換音声の聴取実験

日常的にDAISYプレーヤやスクリーンリーダーの高速再生を日常的に利用している視覚障害者を対象に適応形話速変換方式の効果を確認した。一口に高速音声の聞きやすさを評価するといってもさまざまな観点があるが、本実験では、個々の単語レベルまでの正確な聞き取りは求めず、話の内容がどの程度正確に伝わったかを指標に評価を行った。

3.1 実験音声

外国人向け日本語能力試験¹¹⁾の中の聴解問題を用いた。一人もしくは二人の会話音声をも20秒間ほど聞いた後、内容について正しく説明されている音声を四者択一で回答する

ものである。日本人大学生4名で予備試験を行ったところ、全員がほぼ満点であった。実験は、リスニング問題の部分のみを提案手法によって各種再生倍率で高速再生する以外は、日本語能力検定試験の手續きどおりに実施した。変換倍率は2倍速から5倍速までの6段階に設定し、従来法と提案法でそれぞれ話速変換した。一話速あたり従来法3問、提案手法3問の計6問を用意した。ここでいう従来法とは、DAISYプレーヤと同様に一樣に時間伸縮した音声であり、提案手法の一樣伸縮変換によって作成した。したがって提案法と従来法の音質はまったく同じである。

3.2 方法

実験は一般的な会議室で、スピーカ提示により全員一斉に行った。音声明瞭度については、各被験者に問題がないことを事前に確認した。実験は日本語能力試験と同じ手続きで行うが、被験者が全盲の視覚障害者であるため解答用紙は使わず、別途用意したワイヤレス式の押しボタン回答器を用いた。手を上げるという簡単な手段を用いなかったのは、挙手の際の衣擦れ音が回答者に影響を及ぼすことが予備実験で判ったためである。回答は四者択一で、わからないという回答も認めた。被験者は8名ずつ二つのグループに分け、それぞれ別の実験を行った。内容はいずれも同じだが、1グループ目で従来法によって変換した問題は2グループ目では提案法で変換し、手法間のカウンターバランスをとった。

3.3 被験者

高速音声に慣れている視覚障害者に限定した。各被験者にはDAISYプレーヤやスクリーンリーダーの利用状況について事前にアンケート調査を行い、日常的に高速再生機能を利用していると回答した26名を選んだ。年齢層は40～60歳代で、障害歴は考慮しなかった。

3.4 結果

実験結果を図5に示す。従来法と提案法の再生倍率毎の正解率を折れ線グラフで示したものである。提案法のスコ

図5 提案法と従来法の効果の比較(日本語)
Comparison of proposal method and conventional method (in Japanese).

アは2倍速から5倍速のすべての倍率で従来法を上回った。また棒グラフは、それぞれの方式で「わからない(聞き取れない)」と回答された割合である。この「わからない」も従来法に比べて提案法はその割合が低く、特に高速再生でその差が広がる結果となった。以上から、高速音声に従来法に比べて聞きやすくするという目標については達成したといえることができる。

提案法に関する内観調査では、キーワードが一部聞き取れなかったため試験としては不正解だったが、話の概要はよくわかったという声が多く聞かれた。実験結果で「わからない」の回答率が3.5倍速で不連続に上昇したためその群の回答を調べた結果、ほとんどの被験者が「わからない」と答えた問題が含まれていた。設問の難易度統制については公的な試験問題を利用した段階で特に考慮しなかったことが反省点であるが、同倍率内での従来法と提案法の効果の比較は同条件でできるため、3.5倍速の実験データも落とさずそのまま表示した。

3.5 考察

今回提案した適応形話速変換方式は従来法との比較において有効性を確認できたが、3倍速を超えた場合には正解率が50%を割り込む結果となった。この結果をどう捉えるかだが、比較すべきは、通常の3倍速以上で新聞などを斜め読みして得た情報の解像度である。「回答は間違えたが概要はわかった」という状況を何らかの指標で評価する必要がある。

斜め読み相当の聞き方を期待するということであれば、少なくとも情報の一部欠落は認められるべきであり、その観点で、どう聞きたいのかを実際のユーザを対象に調べた上で、アルゴリズムの最適化を進める必要がある。

今回は音響情報のみを指標にしたが、言語解析情報を用いた情報の取捨選択方法についても現在、実践的に検討を進めている。

4. 外国語での評価

外国語(英語)への適応性について、ネイティブスピーカー21名を対象に評価実験を行ったので、参考として示す。実験は、San Diego (USA) のBlind Centerの協力を得て現地で実施した。詳細な実験は時間の都合上できなかったため、従来法と提案法で話速変換された同じ時間長の音声24セットを一对比較し、「どちらが聞き取りやすいか」を挙手で回答する方法で行った。実験音声は、英語版DAISY図書、ABC、TOEFL、NHKワールドニュースを用いた。PtとPoの寄与率は、2.4節(2)の予備調査の結果から、Ptを30%、Poを70%とし、話速を適応制御した。結果を図6に示す。

評価点は、従来法と提案手法のいずれかが好ましいと選択された割合を示す。いずれも提案法が高い評価を得た。4倍速では無回答が増えたが、3倍速までは高い支持を得たことがわかる。しかし、休止区間が完全になくなっている。単語の意味を取り違えるような欠落があったなどというコ

図6 提案法と従来法の効果の比較(英語)

Comparison of proposal method and conventional method (in English).

メントもあった。日本語同様、どう聞きたいのかを指標に外国語対応についても検討を進めていく。

5. むすび

聞きやすい高速再生を実現する適応形の話速変換方式を提案した。予め分析した韻律に寄与する音響特徴量を手がかりとし、基本周波数とパワーの逐次変化を反映した倍率波形にしたがって話速変換した結果、高速再生時の内容把握において、従来のように時間伸縮手法に比べて高い有効性を確認した。また、倍率波形生成時に基本周波数に基づく倍率波形とパワーに基づく倍率波形の寄与率を調節することで、日本語だけでなく、外国語の変換にも適応できる可能性が示された。最近の国内外でのオーディオブックの急速な普及に伴って、聞きやすい高速音声再生技術に対するニーズは、視覚障害者から一般にも拡大しつつあり、音声を含むさまざまなメディアでの新しい聴取(視聴)形態を提案できればと考えている。

なお、本報告の一部は、NICT(情報通信研究機構)の助成事業の成果に基づくものである。

[文 献]

- 1) Daisy Consortium, <http://www.daisy.org>
- 2) 鳥原: 「「ななめ聞き」システムー統語情報を用いた、視覚障害者のための「速聞き」システムー」, 信学技報, 第5回福祉情報工学研究会(2000)
- 3) 浅川, 高木, 井野, 伊福部: 「視覚障害者への音声提示における最適・最高速度」, ヒューマンインタフェース学論, 7, 1, pp.105-111 (2005)
- 4) 今井, 池沢, 清山, 中村, 都木, 宮坂, 中林: 「ニュース音声を対象にした時間遅れを蓄積しない適応型話速変換方式」, 信学論, J83-A, 8, pp.935-945 (2000)
- 5) 今井, 都木, 小山, 島津: 「話速変換技術を用い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, 59, 2, pp.265-270 (Feb. 2005)
- 6) A. Imai, N. Seiyama, T. Mishima, T. Takagi, E. Miyasaka: "Application of Speech Rate Conversion Technology to Video Editing-Allows up to 5 times normal speed playback while maintaining speech intelligibility-", AES 20th International Conference, pp.96-101 (Oct. 2001)
- 7) 都木, 服部, 小宮, 今井, 岸, 伊藤: 「韻律の視覚化および矯正音声

を用いた中国語学習システム”，信学論，J88-D-II，2，pp.478-487 (Feb. 2005)

- 8) 杉原：“音声表現に見る共通語の韻律理論”，放送研究と調査，pp.76-90 (Apr. 2011)
- 9) 古井：“デジタル音声処理”，東海大学出版会，pp.18 (1992)
- 10) L.L. Elliot: "Backward masking: monotonic and dichotic conditions", J. Acoust. Soc. Am., 34, 8, pp.1108-1115 (1962)
- 11) (財) 日本国際教育支援協会，(独) 国際交流基金：“The Japanese Language Proficiency Test”，(2004～2009の聴解試験)

今井 篤 1989年，埼玉大学工学部電気工学科卒業。同年，NHK入局。名古屋放送局を経て，1992年より，放送技術研究所に勤務。2008年～2011年，(財) NHKエンジニアリングサービスに出向。話速変換，音声合成，音声知覚，音響認知科学，高齢者向け音声処理技術などの研究に従事。正会員。

田澤 直幸 1999年，日本工学院専門学校卒業。同年，(財) NHKエンジニアリングサービスに入社。視覚障害者向け高速話速変換の研究および実用化などに従事。

岩鼻 幸男 1973年，東京電機大学電子工学科卒業。1968年，NHK入局。2006年より，(財) NHKエンジニアリングサービスに勤務。視覚障害者向け高速話速変換の研究および実用化などに従事。

都木 徹 1981年，電気通信大学電気通信研究科修士課程修了。同年，NHK入局。長野放送局を経て，1984年より，放送技術研究所に勤務。2010年より，(財) NHKエンジニアリングサービス出向。音声知覚，声質変換，話速変換，音声合成，字幕制作などの研究に従事。2003年，映像情報メディア学会技術振興賞，2005年，第13回日本音響学会技術開発賞，2008年，第54回大河内記念技術賞受賞。博士(工学)。正会員。

清山 信正 1987年，早稲田大学理工学部電子通信学科卒業。1989年，同大学院修士課程修了。同年，NHK入局，同放送技術研究所に勤務。2007年～2010年，(財) NHKエンジニアリングサービスに出向。現在，放送技術研究所主任研究員。話速変換，声質変換，音声合成，高齢者向け音声信号処理などの研究に従事。正会員。

田中 敏明 1984年，北海道大学医療技術短期大学部理学療法学科卒業。1992年，ニューヨーク大学大学院理学療法学科修士課程病態運動学コース修了(MA) 授与。1992年，札幌医科大学保健医療学部理学療法学科助手。1996年，同大学講師。1999年，同大学助教授。2003年，マサチューセッツ工科大学客員研究員。2006年，札幌市立大学デザイン学部教授。2008年，東京大学先端科学技術研究センター特任教授。博士(工学)。

伊福部 達 1971年，北海道大学大学院修士課程(電子工学) 修了。1989年，北海道大学電子科学研究所教授，2002年，東京大学先端科学技術研究センター教授。北海道大学，東京大学名誉教授。2011年より，高齢社会総合研究機構特任研究員。この間，感覚・コミュニケーション支援のための福祉工学の開拓と産業応用の研究に従事。工学博士。