

適応型話速変換方式の開発と視覚障害者支援への応用に関する研究

著者	今井 篤
学位授与年月日	2014-03-24
URL	http://hdl.handle.net/2261/57925

論文の内容の要旨

論文題目 適応型話速変換方式の開発と
視覚障害者支援への応用に関する研究

氏 名 今井 篤

本論文は、電子媒体に記録された音声を意図的に高速再生して聴取する際に、従来の一様に時間伸縮する話速変換方式に比べて、同じ再生速度でありながら発話内容が聞き取りやすい適応型の話速変換方式を提案し、その視覚障害者のための情報取得支援技術としての応用と有用性について述べたものである。

文字の読み取りが困難な弱視者や全盲者は、情報取得の大部分を音声に頼っており、より多くの情報に効率的にアクセスするためには、発話内容を短時間で効率よく把握できる技術が求められる。特にビジネスとして多くの文書を音声で読んでいる人の中には、現状の音声再生器がサポートする上限の3倍速度では飽き足らず、さらに高速に聞きたいという要求も少なからず存在する。しかし、音声は時系列情報であるがゆえに斜め読みのように俯瞰的に情報をとらえることが難しく、再生速度をできるだけ上げて聞く以外に有効な代替手段がないのが現状である。アナログ時代の可変速再生は、声の高さが変化して非常に聞きづらいものであったが、今日ではそのような問題は解消されており、音質については大きく改善されている。しかし、現状の線形に時間伸縮する話速変換技術では、高速に再生すればするほど発話全体が一様に聞きづらくなり、聴取者それぞれの語音識別速度の能力限界を超えたところでまったく聞き取れなくなる。この再生方法の問題点は、冗長な情報も聞き取るべき情報も同じ重みで再生されることであり、意図した情報を音声コンテンツの中から探し出して聞くような“プル型”の聞き方にしても、また、受け身的に情報を概観するような“プッシュ型”の聞き方にしても、そのいずれ

にしても、音声が高速に再生された場合に、聞き取るべきポイントが浮かび上がって聞こえてこないということにある。そこで、聴感的に冗長な部分と、聞き取るべき部分を自動で判別して、冗長な部分の再生時間を聞くべき部分に再分配することを意図した適応型の話速変換技術を提案するものである。

高速音声の好ましい聴取条件に関しては、視覚障害者がどこまで高速な音声聞き取れるのかという観点での報告がされているが、発話内容がほぼ100%聞き取ることができるとされる2倍速前後の話速でも、筆者らによる視覚障害者へのアンケート調査では、「聞きづらい」、「疲れる」といった声が多く寄せられており、聞き取れることが必ずしも快適な聴取条件にはならないということが示されている。このような現状に対し、斜め読みのように効率的に情報の取得ができ、かつ身体的な負荷の少ない快適な高速音声の聴取手段を実用化することが、本研究の最終的な目標である。

以下、本研究の内容を章ごとに整理する

第1章においては、視覚障害者による活字情報の取得手段として、録音図書とスクリーンリーダーを取り上げ、それぞれの利用の現状と課題を述べた。その課題を踏まえ、本論文の目的である、高速に再生した音声聞き取りやすい適応型話速変換技術の開発に取り組む意義、および音声聴取補助技術としての新規性、有用性を述べた。

第2章においては、話速変換技術に関連する先行研究を概観した上で、現状の高速音声の再生方法に関する問題点を考察した。また、本論文で提案する適応型話速変換技術の基礎となる、放送の早口をゆっくりと聞きやすくすることを目的として開発された高齢者向けの話速変換技術についてレビューし、提案手法につながる基礎知見をまとめた。

第3章においては、本研究で提案する適応型話速変換技術を実現するための、基本的な音声信号処理技術、および本技術を実用に供することを目的として、必ずしも音響的に理想的な条件ではない場合でも、安定的に動作するための信号処理上の工夫について述べた。

第4章においては、視覚障害者への情報保障の観点から、晴眼者による斜め読みを相当の条件として、高速音声の提示条件について調査した結果を述べた。視覚障害者の多くは音声聴取の際に効率を優先しており、聞きづらさや疲労を感じながらも、聴力の限界付近の速さで聞いている人が多いという実態を示した。また、黙読の代替手段としての相当性において、どこまでの速さの聞き取りを目指すかなど、アルゴリズム設計に関する要求条件を明らかにした。

第5章においては、視覚障害者の高速音声聴取を目的として開発した2つの適応型話速変換技術について述べた。提案する2つの方式とは、①逐次処理による適応型話速変換方式と、②事前分析型の適応型話速変換方式であり、前者は、既存の機器への実装が容易な方式であり、音声コンテンツがあれば直ちに動作させることができる。後者は予め再生するコンテンツの音響的な特徴を分析しておき、その結果を再生の際に参照する

ことで、再生時点よりも未来の情報も加味した動作の最適化ができる。方式の原理的な利点のほか、外国語を含む多様な発話にも柔軟に対応できることを示した。

第6章においては、視覚障害者を対象に、従来法である一様に時間伸縮する話速変換と、提案法である適応型話速変換による、同一時間長での聞きやすさ、および内容理解に関する評価実験を行なった。いずれの方式も、従来法に比べて高速な音声聴取において効果的にはたらくことを示した。

第7章においては、英語とドイツ語について、それぞれネイティブスピーカーである視覚障害者を対象に従来法と提案法の比較実験を行ない、高速音声の聞きやすさの向上に関して、提案法の有効性を示した。この実験結果により、音響情報のみを手がかりとする適応型話速変換技術の、多言語への展開についての可能性を示した。

第8章においては、研究成果の実用化展開について述べた。PC上で動作するマルチメディア DAISY のシミュレータや、スマートフォン上で動作するアプリケーションを実際に作成し、提案技術の実用化への具体的な道筋を示した。

ここで、第8章までに得られた主な結論をまとめる。

1. 高速再生の上限値を与えるため、高速な黙読相当の速さを実験的に調べた結果、現状の録音図書再生器の再生レンジ（～3倍速）が、情報受容の時間効率の観点で、ほぼ同等であることが分かった。しかし、一方で、比較的高齢の多くの視覚障害者が、2倍速を超えると聞き辛く、特に長時間聴取においては疲労が大きいと感じていることが分かった。具体的な数値目標として、上記実験から音声の提示速度の必要条件は2.9倍速（＝約3倍速：1250モーラ毎分）、また、実験で得られた黙読速度の分散を考慮して、十分条件を3～3.5倍速とした。
2. 逐次処理による適応型話速変換技術では、未来を加味した情報の再配分ができないため、話速変換処理に伴って音声の遅れが一定時間以上蓄積した場合には、音声区間の削除も許容するアルゴリズムとした。削除の指標を、発話の冒頭から一定時間Tが経過した以降で、基本周波数とパワーの移動平均値が所定の閾値を下回る場合とし、この条件の場合は削除することを許容した。
3. 高速再生時の休止区間の短縮許容限に関する実験結果に基づき、再生速度が上がっても休止区間は30ms以下にはしないこととした。これを下回ると、隣接する音素が知覚的に融合してしまい非常に聞き辛くなる。音声を高速に聴取するという目的においては、最低限必要な休止区間であり、より長時間の聴取を行う場合には、時間が許す限り長いほうが望ましい。
4. 事前分析型の話速変換技術において、基本周波数とパワーの変化の両方の特徴に基づいた倍率関数を生成するが、音声を高速に変換する際には、言語によって両者の好ましい寄与率に違いがある傾向が示された。高速再生の聞き取りがしやすいという観点での評価実験によると、日本語は、基本周波数の変化に重みを置くことが好ましく、英語・ドイツ語はパワーの変化に重みを置くことが好ましいという傾向が

示された。各言語において、実験の評価が最も高かった寄与率を適用した高速話速変換アルゴリズムを構築し、それぞれの言語で高速音声の聞きやすさに関する評価実験を行なった。その結果、日本語だけでなく、英語、ドイツ語のいずれの言語においても、従来の一様に時間伸縮する話速変換方式に対する有効性が確認され、本アルゴリズムの音響パラメータの適応化による外国語適応の可能性が示唆された。

5. 視覚障害者を対象にした実証実験の結果、2つの提案方式（逐次方式、事前分析方式）のいずれもが、聞きやすさおよび内容理解の観点で、従来法に比べて優位であることが示された。被験者の平均年齢が高かったこともあり、2倍速での評価値をベースラインと考えると、情報保障の観点で斜め読み相当の最高速度とした3倍速程度までの聞き取りの劣化は従来法に比べて明らかに小さく、かつ、いずれの方式も従来法の評価値に比べて20ポイント以上高い値を示した。また、ビジネス用途などで、より高速な音声を求める人の指標とした3.5倍速においても、いずれの方式も従来法に比べて優位な結果になったが、特に、事前処理方式では、内容理解度で約50%を維持しており（従来法では約30%）、高い有効性が認められた。

第9章においては、今後の課題を整理した。本論文で提案した適応型話速変換アルゴリズムは、自然発話の韻律を利用したものであり、視覚障害者への情報保障技術という観点では、録音図書同様に利用者が多い合成音声の研究対象から除外されていた。これに対し、本研究で得られた知見に加え言語情報の利用による今後の対応について展望した。また、認知科学的な観点から、高速音声を聴取する際の手掛かりとして優先される要因とそれを阻害する要因について考察し、本論文の結論と合わせて、より効果的でストレスの少ない高速音声の再生手段の実現の可能性についても展望した。