

機関番号：12612

研究種目：基盤研究(C)

研究期間：2008年度～2010年度

課題番号：20500111

研究課題名(和文) フレキシブルな時間軸を持つ高効率音声再生法の研究と研究者用音声データベースの作成

研究課題名(英文) A study on efficient playing system with flexible time axis and speech database for researchers of speech

研究代表者

高橋 弘太 (TAKAHASHI KOTA)

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：10088005

研究成果の概要(和文)：人間の聴覚特性に基づいて、最も聞き取り易い再生速度で音声再生する手法について研究した。この研究では、第一に、話速が聴覚にあたえる影響の定量的研究、第二にその研究成果に基づいた再生速度決定アルゴリズムの研究とそのアルゴリズムを実証するためのシステム実装と実験の2つが大きな柱である。さらに、第三の柱として、この研究のために自前で製作する話速バリエーション型音声データベースをインターネットで公開し、音声分野の研究者に利用してもらい話速推定研究を啓蒙することがあげられる。3年間の期間で、これらを計画どおり実施した。

研究成果の概要(英文)：A method for reproducing a speech by the optimum reproduction speed has been studied based on hearing characteristics. This study consists of two subjects, one is quantitative analysis of relationship between speaking rate and recognition rate of speech, and another is implementation and experiments of reproducing system. The speaking rate variation speech database (SRV-DB) was also constructed and opened to the public for the researchers. These subjects have been completed according to the schedule.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,100,000	330,000	1,430,000
2009年度	1,200,000	360,000	1,560,000
2010年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：音声データベース、

## 1. 研究開始当初の背景

近年のテレビ放送の多チャンネル化とインターネットによる番組配信により、現代人が手にする視聴覚コンテンツ量は爆発的に増大している。一方、コンテンツを収録するH

DDレコーダ等の記憶容量も着実に増加し、結果として現代人がレコーダに録り溜めているコンテンツ数は著しく増大している。野村総研の2006年の調査によれば、HDDレコーダに録り溜めてある番組の平均数は

37.3 番組であり、これは前年に比較して4割の伸びであるという。さらに、録ったものの視聴が追いつかず削除されてしまう番組数の増加はさらに多く、前年比6割増であったことが報告されている。このような背景のもと、コンテンツの内容を短い時間で可能な限り大量に把握できる技術の開発が望まれている。この要請に対して、現時点で実用化されている方法は、映像部分のカットやシーンの切れ目を根拠にした画像のサムネイル表示による選択的視聴、スポーツ番組での歓声の盛り上がり根拠とし重要シーンのみを自動選択させる視聴などがある。しかし、どちらも番組の一部しか聞くことができず、漏れが生じてしまう。

番組の内容を全て提示するためには、再生速度を上げて再生する以外ない。映像部分は再生速度が高くても内容把握が可能であるが、音声部分は2倍速を超えると、ほとんど聞き取れなくなってしまう。無音部をつめることで大幅な効率化が達成できる場合もあるが、報道番組やニュースなどでは、無音部がほとんど存在しないので、この方法での改善は期待できない。

## 2. 研究の目的

以上の問題意識のもと、再生速度を一定速に保つのではなく、番組視聴者（以下、ユーザと呼ぶ）の操作や、視聴覚コンテンツにおける原音声（以下、ソース音声と呼ぶ）の話速に応じて、再生速度を最適に設定する方法を提案した。

このとき、すなわち、操作と話速に応じて再生速度を適宜調整することで最高の効率が得られるわけだが、どのようなアルゴリズムで速度を決定すればよいのかという問題（再生速度曲線の最適化問題）は、未だ解かれていなかった。

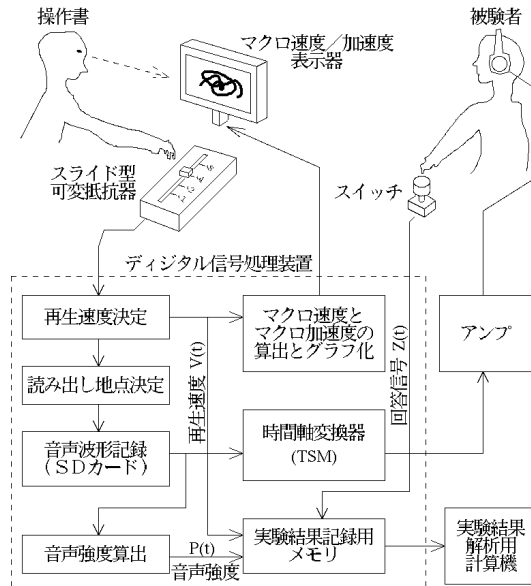
以上の問題意識に基づき、本研究の第一の目的を、聞き取りに関する人間の聴覚特性を定量化し、その特性を境界条件とした再生速度決定のアルゴリズム開発においた。

ところで、音声再生速度が随時変化する提示法に対する聞き取りの可能性に対する学術的な議論は今までなかった。本研究では、その部分を実験データに基づいて定量的に調べることで、学術的な研究分野となるようにしたい。また、この研究過程で得られる特殊な音声データをWebで研究者に公開することで、この分野に入ってくる研究者の便

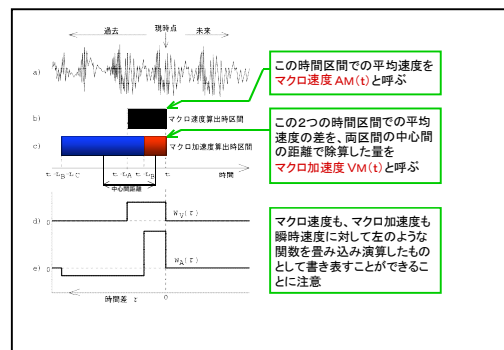
をはかり、研究分野として確立させたいと考えた。このデータベースの構築と公開が、本課題の第二の目的である。

## 3. 研究の方法

(1) 話速が人間の聴覚に与える影響については、以下の図に示すようなシステムを構築して実験を行った。



このシステムにより、被験者に対し、さまざまな話速変化で音を提示することができる。被験者は、聞き取り不可となった時間区間、スイッチを押すことで、そのことをシステムに伝える。システムは、速度変化、音量、スイッチ位置などを記録する。これらの結果を解析することで、われわれが提唱する、マクロ速度、マクロ加速度による指標の妥当性を検証した。マクロ加速度、マクロ速度の概念を下図に示す。



## (2) 可変速の音再生システム

話者が、さまざまな速度で話したとき、その話速を自動推定し、あらかじめ設定した話速に変換して再生するシステム、parrot をノートパソコンの上に実装した。話速の推定アルゴリズムとして、独自の方法である、2帯域



このようなシステムを実装し、評価実験を重ねた結果、十分実用になることが確かめられ、この分野の研究の必要性を内外にアピールできたと考えている。

(3) 話速バリエーション型音声データベース (SRV-DB) の製作

現在公開している音声の全体像を以下に示す。文章数は、全体で1658に上る。現在編集集中の収録データもあるので、この数はさらに大きくなる予定である。

発話者種別	原稿	話速	文章数
アナウンサー	読売新聞コラム 14文	6話速	84
	読売新聞コラム 17文	6話速	102
	読売新聞コラム 13文	6話速	78
	ATRによる 25文	3話速	225
一般人	ATRによる 25文	5話速	625
声優	オリジナル原稿 62文	4話速	248
	オリジナル原稿 74文	4話速	296

また、公開ページを以下に示す。文章ごとにダウンロードできるほか、全体を一括でもダウンロードできるようするなど、利用者の便をはかっている。

話速バリエーション型音声データベース SRV-DB

このページは、SRV-DB のダウンロードページです。案内の各ファイルは、Microsoft WAVE形式 (wav ファイル) にて提供しております。また、一括ダウンロード用のファイルは、ZIP形式にて圧縮しております。音声データは、PCM 44,100 Hz 16bit で収録しております。チャンネル数は、基本的にモノラル(Mon)ですが、一部のものはステレオとなっております。

1. 楽語のプロフェッショナルによるATR 25文の読み上げ

	話者名: PF01	話者名: PF01	話者名: PF01	同時音一括ダウンロード
自然な発音 (自由発音)	ダウンロード	ダウンロード	ダウンロード	この音もダウンロード
4.78 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この音もダウンロード
8.88 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この音もダウンロード
9.51 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この音もダウンロード
11.31 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この音もダウンロード
13.45 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この音もダウンロード
同時音一括ダウンロード	この音もダウンロード	この音もダウンロード	この音もダウンロード	一括ダウンロード (約49MB)

2. 本研究室の所属メンバーによるATR 25文の読み上げ

	話者名: AM01	話者名: AM01	話者名: AM02	話者名: AM03	一括ダウンロード
4.78 [モーラ/秒]	↓選択	↓選択	↓選択	↓選択	音ダウンロード
8.88 [モーラ/秒]	↓選択	↓選択	↓選択	↓選択	音ダウンロード
9.51 [モーラ/秒]	↓選択	↓選択	↓選択	↓選択	音ダウンロード
11.31 [モーラ/秒]	↓選択	↓選択	↓選択	↓選択	音ダウンロード
13.45 [モーラ/秒]	↓選択	↓選択	↓選択	↓選択	音ダウンロード
同時音一括ダウンロード	この音もダウンロード	この音もダウンロード	この音もダウンロード	この音もダウンロード	一括ダウンロード (約129MB)

3. 楽語のプロフェッショナルによるATR 25文の読み上げ

	話者名: PF01	話者名: PF01	話者名: PF01	同時音一括ダウンロード
5.00 [モーラ/秒]	↓選択	↓選択	↓選択	この音もダウンロード
8.88 [モーラ/秒]	↓選択	↓選択	↓選択	この音もダウンロード
11.88 [モーラ/秒]	↓選択	↓選択	↓選択	この音もダウンロード
同時音一括ダウンロード	この音もダウンロード	この音もダウンロード	この音もダウンロード	一括ダウンロード (約49MB)

4. 声優によるオリジナル原稿の読み上げ (台詞を連続してストーリーにしたもの)

	モナラル	ステレオ	男性のみ	女性のみ	一括ダウンロード
自然な発音 (7.58)	ダウンロード	ダウンロード	ダウンロード	ダウンロード	音ダウンロード
4.78 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	ダウンロード	音ダウンロード
8.88 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	ダウンロード	音ダウンロード
11.31 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	ダウンロード	音ダウンロード
一括ダウンロード	この音もダウンロード	この音もダウンロード	この音もダウンロード	この音もダウンロード	一括ダウンロード (約219MB)

5. 声優によるオリジナル原稿の読み上げ (台詞ごとにファイル分割した形式)

	話者名: PF01	話者名: PF01	同時音一括ダウンロード
自然な発音 (7.58)	↓選択	↓選択	この音もダウンロード
4.78 [モーラ/秒]	↓選択	↓選択	この音もダウンロード
8.88 [モーラ/秒]	↓選択	↓選択	この音もダウンロード
11.31 [モーラ/秒]	↓選択	↓選択	この音もダウンロード
同時音一括ダウンロード	この音もダウンロード	この音もダウンロード	一括ダウンロード (約219MB)

以上が、本研究の成果である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕 (計1件)

- (1) 高橋弘太、デジタルオーディオデータのサンプル欠けの位置推定と修復、電子情報通信学会論文集、査読有り、Vol. J94-D, No. 7, 2011年。

〔学会発表〕 (計4件)

- (1) 葛木圭悟、高橋弘太、視聴覚コンテンツにおける音声の高品位な時間軸伸縮法、電子情報通信学会技術報告、コミュニケーションクオリティ研究会、CQ2011-03、長崎県やすらぎ伊王島、2011年3月8日
- (2) 吉原 亨、高橋弘太、音声の効率的な再生を目的とした最適な再生速度の決定法、電子情報通信学会技術研究報告、スマートインフォメディアシステム研究会、109巻、202号、95-100、広島大学、2009年9月25日
- (3) 高橋弘太、葛木圭悟、吉原亨、話速管理機能を持った原稿提示収録システム (ReCoK5) と話速バリエーション型音声データベース (SRV-DB) の公開について、第10回音声言語シンポジウム、227-232、早稲田大学 (東京)、2008年12月2日
- (4) 吉原亨、葛木圭悟、高橋弘太：音声の高速再生のための話速推定法と高速発話時の特性解析～話速バリエーション型データベースの活用例～、電子情報通信学会技術研究報告、第10回音声言語シンポジウム、233-238、2008年12月2日

[その他]

この研究費を使い、収録、編集して公開している話速バリエーション型音声データベース（略称：SRV-DB）は、以下のアドレスからアクセスすることで、全データを無料でダウンロードすることが可能である。

<http://www.it.ice.uec.ac.jp/SRV-DB/>

## 6. 研究組織

### (1) 研究代表者

高橋 弘太 (TAKAHASHI KOTA)

電気通信大学・大学院情報理工学研究科・  
准教授

研究者番号：10088005