科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ６ 月　 ５ 日現在

機関番号：１２６１２
研究種目：基盤研究(C)
研究期間：2011～2013
課題番号：２３５００１４７

研究課題名（和文）フレキシブルな時間軸による音声再生システム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 speech reproduction system using flexible time-axis and speech database for researchers

研究代表者
高橋　弘太（Takahashi, Kota）

電気通信大学・情報理工学（系）研究科・准教授

研究者番号：１０１８８００５

交付決定額（研究期間全体）：（直接経費）　　3,900,000円、（間接経費）　　1,170,000円

研究成果の概要（和文）：本課題では，フレキシブルな時間軸による再生機の研究の中の要素技術である話速推定の技術や，話速変換のための信号処理技術を体系的に研究する．当該研究期間中には，ソフトウエアによる研究環境を拡充し，信号処理を処理要素ごとに分解して記述し，それを連結することで処理が行えるようにした．FPGAによるシステムについては，Xilinx社の評価ボード上に再生システムの雛形を実装した．また，本研究課題では，その研究に用いるためだけでなく，全国の研究者が無償で利用できるように,音声データベースの構築も行っている．研究期間終了時において，2216文の読み上げを擁するデータベースとなっており，すでに公開を行っている．

研究成果の概要（英文）：Speaking rate estimation techniques for a reproduction system using flexible time-axis and signal processing techniques for a speaking rate conversion are studied. In the study period, the software research environment is expanded. In this expanded environment. signal processing is decomposed and described for every processing element, and a new signal processing method can be made by connecting elements. About the system with FPGA, the proposed method of a reproduction system was implemented on the evaluation board of Xilinx. Moreover, in this research, construction of a special speech database is also performed so that researchers all over the country can use free. At the time of the end of this academic study period, the database has 2216 sentences, and has already opened to the public.

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：話速変換　音声データベース　話速推定

１．研究開始当初の背景

本研究の背景には、現代人がアクセスする視聴覚コンテンツの劇的変化がある。すなわち、近年のテレビ放送の多チャネル化とインターネットによる番組配信により、人々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ＨＤＤ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。ある研究機関の調査によれば、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を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分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

以上の認識をスタート地点として、我々は効率的視聴のための新しい方法を研究している。特に、この課題では、音声を短時間でいかに効率的に聞かせるかということについて、その具体的方法を提案し、さらにその提案手法を実現するハードウエアを試作する。

また、このような研究のためには、話速の推定技術を確立する必要があるが、研究の材料となる音声データベース（すなわち、話速を正確に制御して収録された音声データベース）が存在しないため、この研究の中でそのデータベースも構築し、あわせてそれを公開することによって、国内の音声研究者に役立ててもらおうというのも本研究の重要な目的である。このデータベースを利用してもらうことによって、話速推定や効率的再生の研究が、音声研究の一分野として花開くことをめざしている。

３．研究の方法

本研究の内容は多岐にわたるため、本報告書では、以下の３つに絞ってその内容を報告する。

（１）第一は、音声をより聞かせやすくするための研究である。一般に、効率的な再生を行おうとするとき、ピッチを変化させずに、いかに時間軸だけを変更するかに焦点が絞られるが、本課題による研究の前半において、時間軸を変更するだけでは音声に不自然感が生じてしまい、聴き取りについて支障が出ることがわかってきた。そこで、本課題による研究の後半においては、あえて音声のピッチを積極的に改変することで、聴取者により聴き取り易い音声を提示する方法を研究した。研究手法としては、比較的速い話速で収録した音声を、（例えば高齢者用に）遅い話速で提示する場面を考え、話速変換によって生成した音声と、もともと遅い話速で収録した音声の比較を行い、どのような差異があるかを解析した。そして、その結果にもとづいて、より自然感を増すことのできる定時法について提案をまとめあげた。この研究は時間軸のフレキシブル性にあわせて、ピッチについてもフレキシブルに調整するという方向での研究の発展であると言える。

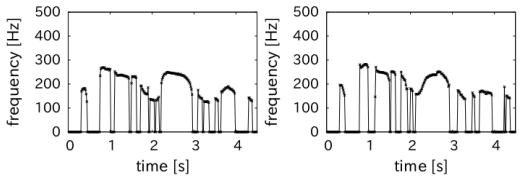
（２）次に、フレキシブルな効率的視聴を実現するためのハードウエア上の実装についても研究した。Ｘｉｌｉｎｘ社の中規模ＦＰＧＡ上に我々が提案する効率的視聴アルゴリズムを実装し、ＦＰＧＡ素子の使用率などを評価することで、ハードウエア上での実現が現実的なものであるかどうかを調べるという手法である。

（３）最後に、製作して公開するデータベースを構築する方法について述べる。話者としてアナウンサーなどプロの発話者を選び、プロダクションと交渉する。その際、研究用データベースであるため、録ったデータを誰でも利用できるように公開することを説明し了承を得る必要がある。次に、台本を作り、また話速を管理して発話してもらうための独自のシステムを用意して、録音に臨む。録音データは、リップノイズと呼ばれる唇の触れ合う音が混入するため、録音後に、これを手作業で取り除く。この取り除き作業には多大な時間がかかるが、良質なデータベースとするためには避けられない作業である。また、話速が正確にコントロールされているかの確認をし、さらに音量の正規化を行って、文章ごとに正確にファイルに切り落として、初めて公開データとなる。

４．研究成果

（１）音声のピッチまで含めてフレキシブルな変更を行う、より聴き取りやすい音声の提示法については、以下の成果を出すことができた。

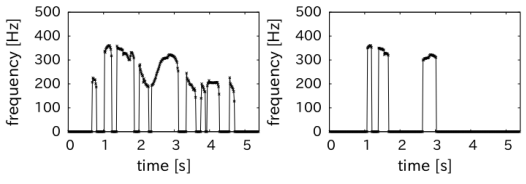
まず比較的速い話速と、遅い話速での発声に関して、ピッチの変化がどのように違うかについて、詳しく解析を行った。



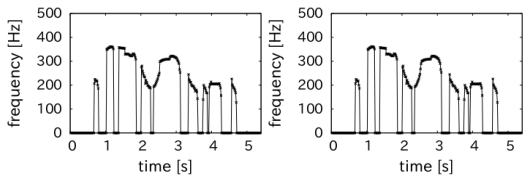
(a) 話速：5.00 mora/sec (自然)　(b) 話速：5.00 mora/sec (人工)

上の図は、ひとつの音声について、元々遅い話速で発声したもの（左の(a)）と、速い話速を人工的な話速変換アルゴリズムで処理して生成したもの（右の(b)）を比較したものである。人工的な話速変換では、ピッチの平坦部が失われており、これが不自然さの原因のひとつであることがわかった。

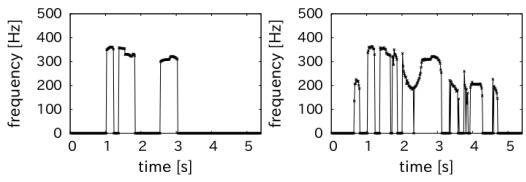
そこで、以下の手順で平坦部の長さを延長するアルゴリズムを考案した。



(a) ピッチ変換前　(b) (a) の平坦部分



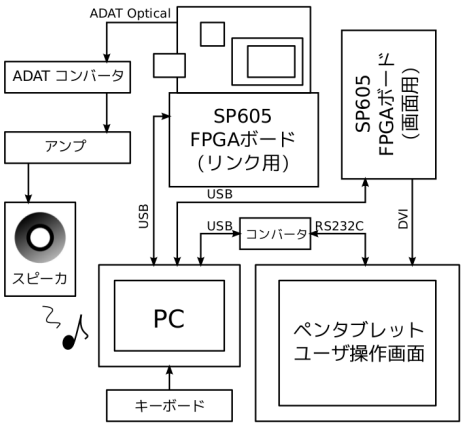
(c) 平坦部分を増加　(d) 平坦部分を滑らかに



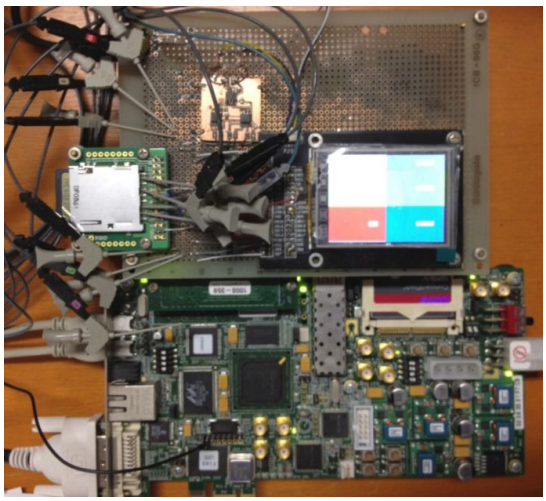
(e) 変換先の平坦部分　(f) ピッチ変換後

この手法を導入したことによって、より聴き取り易い音声を生成することができるようになった。（詳細は文献①を参照）

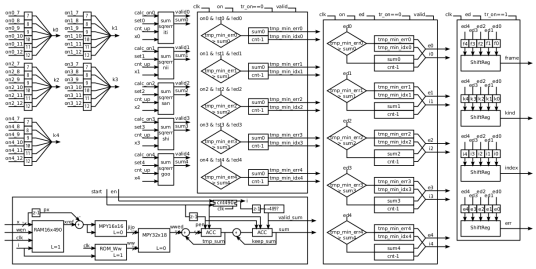
（２）ＦＰＧＡによる実現については、以下の図に示すシステムを構築した（文献②）。



実験中のシステムの写真を以下に示す。



フレキシブルな時間軸で再生するために、音声の再生時刻を時々刻々変更するのに用いるポインタ情報の生成部を一例と示すと以下のようになる。



このような複雑なブロックを ハードウエア記述言語 Verilog で記述し、論理合成して配置配線を行った。

実装には、Ｘｉｌｉｎｘ社のＦＰＧＡの中でも中規模の集積率である XC6SLX45T を用いた。中規模のデバイスであっても、フレキシブルな時間軸での再生システムを１デバイス内に収められることが実証できた。ＦＰＧＡ内での素子使用数を以下に示す。

		REG	LUT	IO	R	B	D	P
必要数	画面用	5,701	5,784	104	35	9	3	2
	リンク再生	16,651	16,911	35	32	7	3	1
	リンク生成	1,350	1,961	80	5	1	25	0
	リンク用	17,981	18,872	115	37	8	28	1
リソース	LX45T	54,576	27,288	296	116	16	58	4
	LX25T	30,064	15,032	250	52	16	38	2

注) R, B, D, P はそれぞれ RAM, BUFG, DSP, PLL を表す

（３）音声データベースについては、毎年、収録と編集を重ねて、データを追加しつつ公開している。

現在、次ページの左に示すように２３０６文章を公開している。平成２５年度に収録した「カーナビの音声を模擬したデータ」と「一文字違うだけの類似度の高い文章」について

は、現在公開している話速以外の話速でも収録が完了しているので、編集が済み次第、順次公開していく予定である。

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
	オリジナル原稿 74 文	4話速	248
	オリジナル原稿 74 文	4話速	296
アナウンサー	カーナビ音声等 40 文	1話速	40
	一文字違い文章 64 文	1話速	64
合計収録文章数			2306

音声データベースは、ＳＲＶ－ＤＢと名付けている。その公開ページを右図に示す。話速ごとに整理されているだけでなく、どのようなデータがデータベース中に存在しているかについても一目瞭然となるようにページのレイアウトも工夫してある。また、一文章づつ試し聴きすることに対応すると同時に、データセットごと zip 形式で一括ダウンロードできるようにもなっている。その他、録音環境（機材や手法など）についても詳細に記述してある。

　またデータだけでなく、話速を正確に制御して話者に発声を行ってもらい、それをリアルタイムでチェックするためのツール（ReCoK5）についても、ソフトウエアを公開しており、望むならば外部の研究者がこのような話速管理データを製作することもできるようになっている。

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）

〔学会発表〕（計 ２ 件）

①井上愛梨,高橋弘太, 時間伸長音声のための局所的ピッチ変換規則の検討, 電子情報通信学会,応用音響研究会, 112(478), 19-24, 株式会社 KDDI 研究所, 平成２５年３月１２日

②鈴木達弘,高橋弘太、長時間音声を聴くための時間節約技術とその FPGA 実装, 電子情報通信学会マルチメディア情報ハイディング・エンリッチメント研究会, 112(467), 47-52, ＡＴＲ（京都府）, 平成２５年２月２８日

〔産業財産権〕

○取得状況（計 ３ 件）

名称：再生装置
発明者：高橋弘太, 政木康生
権利者：電気通信大学, 船井電機
種類：特許
番号：特許 第 5093648 号
取得年月日： 平成２４年９月２８日
国内外の別：国内

名称：REPRODUCING APPARATUS
発明者：Kota Takahashi, Yasuo Masaki
権利者：The University of Electro-Communications,Funai Electric Co., Ltd.
種類：特許
番号：US8165888 B2
取得年月日：平成２４年４月２４日
国内外の別：国外（米国）

名称：REPRODUCING APPARATUS
発明者：Kota Takahashi, Yasuo Masaki
権利者：The University of Electro-Communications,Funai Electric Co., Ltd.
種類：特許
番号：US8165459 B2
取得年月日：平成２４年４月２４日
国内外の別：国外（米国）

〔その他〕

ホームページによるデータベースの提供

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、研究者から学生まで全データを無料でダウンロードして自由に研究に役立て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・情報理工学研究科・准教授
研究者番号：１００８８００５