E-046

持続時間を補正した高齢者音声認識

Elderly Speech Recognition using Normalized Duration

伊藤 健哉† 大津 圭一郎† 佐藤 正樹† 鎌田 直希† 畑岡 信夫†（東北工業大学）
Kenya Ito Keiichiro Ootsu Masaki Sato Naoki Kamata Nobuo Hataoka

1. はじめに

現在の音声認識技術は、成人の音声を対象に開発されており、高齢者の音声モデルには対応していない。しかし、近年の日本では、高齢化が進んでおり、高齢者音声にも対応した認識システムの構築が重要な課題となっている[1]。現在、大語彙連続音声認識を利用したシステムの普及に伴い、さまざまな用途で、音声認識が利用されつつある。高齢者音声データベースから音響モデルを作成することによって、認識性能の改善が見込まれるが、現状では語彙数やデータ量が十分でなく、より汎用の大語彙連続音声認識に適用できるかどうかの検証は十分ではない。前回、我々は高齢者音声認識では沸き出しが多発して認識率が劣化している事を報告した[2]。今回は、高齢者音声の持続時間に着目し、持続時間を補正（短縮）して、連続音声認識の評価を行い、認識率の向上が図れたので詳細を報告する。

2. 大語彙音声認識について

高齢者音声を成人と同じように認識する為にはどうすれば良いかを目標として、高齢者の音声認識を行なった。高齢者音声認識実験を行うにあたって、大語彙音声認識エンジン Julian を使用した[3]。Julian は HP 上で公開されており、無償で入手が可能である。今回使用したのは、Julius/Julian 内の Julian-kit である。Julian はあらかじめ文法を設定しておく必要があるが、決められている文章の認識率が高くなる。なぜなら、登録してある単語から、ネットワークで次の単語を探すためである。文法は、単語の組み合わせで構成されており、各単語はネットワークで繋がっている。つまり、認識実験に使う単語がわかっていれば、Julius よりも Julian のほうが認識実験には向いているといえる。高齢者の音声データは、音声資源コンソーシアム(NII)から購入したデータを使用する。Julian の初期設定では、比較的静かな環境でパソコンマイク入力を用いて、丁寧に発話されていることを前提としている。さらに性能をチューンし、別の環境に適用するには、jconf ファイルなどの設定ファイルで Julius のオプションを指定する必要がある。そのためには、Julius の動作原理、さらには音声認識の基礎原理の理解が必要である。

3. 実験方法

3.1 実験装置

表 1. 実験装置機器

実験装置機器	メーカー
turbo linux Endeavor	EPSON
Windows XP	DELL
spwave	フリーソフト

3.2 実験方法

①Linux で大語彙音声認識 Julian を用いて、高齢者と成人音声を評価し、認識率を求めた。
②認識率の悪い高齢者話者 4 人を選び、さらに認識率の悪い長文 10 文ずつを評価対象にした。
③Spwave を使用し、成人音声と高齢者音声の波形と持続時間を比較する。高齢者音声の持続時間を成人音声の持続時間に近づけるよう無音区間の削除を行った。
④無音区間を削除した高齢者音声データを認識させ、成人音声と元の高齢者音声の認識結果と比較した。

4. 実験結果

表 2 と図 1 は音声の持続時間の補正前と補正後の結果を示している。不要である無音区間の除去を行うと平均して 0.5 秒程の短縮となった。一般に高齢者では成人の音声と比較して発声長が長い事が分かる。

表 3 と図 2 は無音部を除去して持続時間を補正した結果の認識率である。無音区間の削除を行った結果、全体として 20%から 30%の認識率の向上を確認できた。補正後の高齢者音声は成人の音声とほぼ同じ認識率を持つ音声もあり、持続時間を成人音声に近づける事により認識率の向上を確認できた。

表 2. 持続時間 [s]

	r2f01 女	r2m01 男	r4f01 女	r4m01 男	成人
無音区間削除前	5.923	4.743	6.438	5.102	4.413
無音区間削除後	5.394	4.566	5.801	4.769	

表 3. 認識率[%]

	r2f01 女	r2m01 男	r4f01 女	r4m01 男	成人
無音区間削除前	45.63	64.39	47.95	67.84	81.79
無音区間削除後	75.47	88.79	67.87	84.00	

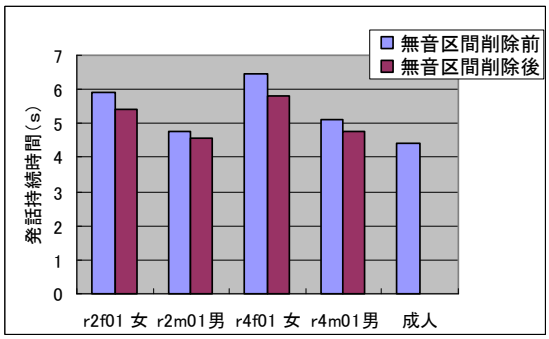

図 1. 補正後の発話持続時間の差

†東北工業大学
〒982-8577 仙台市太白区八木山 香澄町 35-1

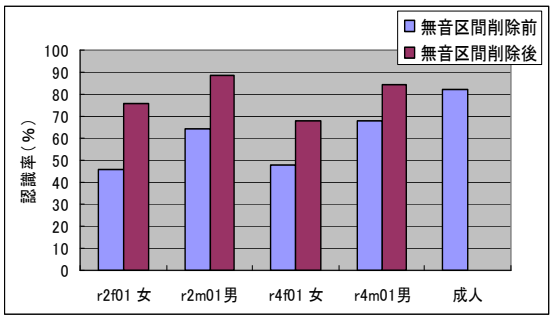

図 2.　無音区間削除前後の認識率

5.　考察

5.1 音声データと認識結果

各音声のデータを図３～図５に示す。認識結果は次のようになっている。

1）読み上げ文
「新築祝いを買いたいんですけど、二万円くらいでよいのはないですか」
2）認識結果
・成人音声
「新築祝いを買いたいんですけど、二万円 くらいでよいのはないですか」

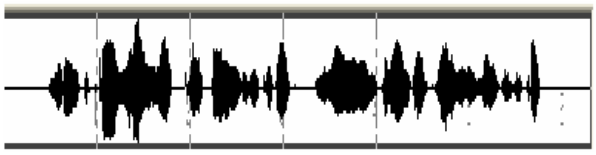

図 3.　成人音声

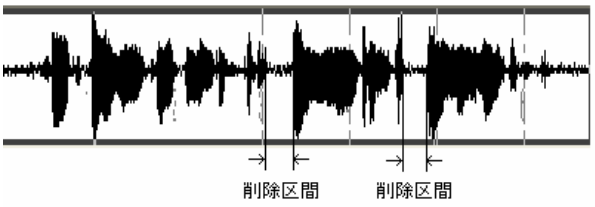

図 4.　高齢者音声（無音区間削除前）

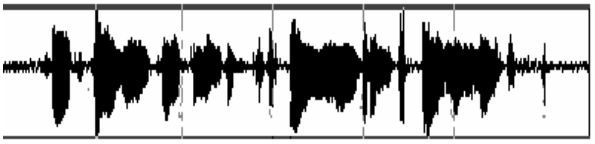

図 5.　高齢者音声（無音区間削除後）

・高齢者音声（無音区間削除前）
「ニンジンをあるですいいで出て多いいいのはないですか」

・高齢者音声(無音区間削除後)
「新築祝いの買いたいんですけど、二万円ぐらいでよいのはないですか」

5.2 持続時間長に関して

高齢者は、成人に比べて音声の持続時間が長い為、単語の沸き出しや置換などによる誤認識が多く、認識率が低下する原因の一つとなっている。高齢者の場合、思考しながらゆっくりと発声すると考えられ、特に複文の場合に文章間に長い無音区間が発生する事が多い。この結果、音声に沸き出しが多発している。

解決策としては、高齢者の音声の特徴に合わせ、ネットワーク文法で長い無音部が必ず入るようにする事も考えられるが、今回試みた無音部の削除により大きい効果を得る事ができ、現実的な処理であると考えられる。

6.　まとめ

高齢者の音声から無音部の削除を行い、持続時間を補正した結果、湧き出しが減り、成人に近い認識率の向上を確認する事ができた。

しかし、高齢者の音声には今回行った持続時間の長さの他にも、フォルマントの欠落などの問題があるので、これを今後の課題としたい。これから更に高齢化が進む社会で、高齢者の音声認識率を上げていく為に高齢者の音声モデルの更なる解析を進め、より精度の高い音声認識に繋がる研究をしたい。

参考文献

[1]井ノ上直己他, 高齢者用HMMによる認識実験,電子情報通信学会 2000 年総合大会講演文集 ｐｐ.193(2000 年)

[2]佐藤孝志他、モバイル社会を目指した高齢者音声認識、平成 20 度東北地区若手研究者研究発表会　講演資料 pp.23-24（2008 年）

[3]大語彙音声認識 Julius/Julian
http://julius.sourceforge.jp/