

複合補聴処理における話速変換処理*

張 丹[†] 坂田 聡^{††} 上田 裕市[†]

([†]熊本大学大学院自然科学研究科 ^{††}熊本大学工学部)

1. まえがき

言葉を話し耳で聞くという行動は、人がより快適なライフスタイルを築いていく上で、重要な手段の一つである。しかし、聴覚系に何らかの原因による機能劣化が生じると、その音声言語も情報伝達手段とはなりにくく、日常生活に支障をきたす原因となる。健聴者でも加齢により内耳以降の神経系や中枢系の経年劣化を生じ、速い話速に追従できないといった症状が現れてくる。こういった難聴は、老人性難聴と呼ばれる。本研究では様々な聴覚障害者に適用できる複合補聴処理方式として分析・合成系にケプストラム法を用い、周波数次元での補償に加え、声質変換、話速変換を一括して処理可能な複合補聴処理システム^[1]の実現を目的としている。本稿では、先に提案したピッチ変換(声質変換)及びスペクトル圧縮処理から成る複合補聴処理の評価^[2]に加え、更に、付加機能として話速変換処理について報告する。

2. 複合補聴処理の概要

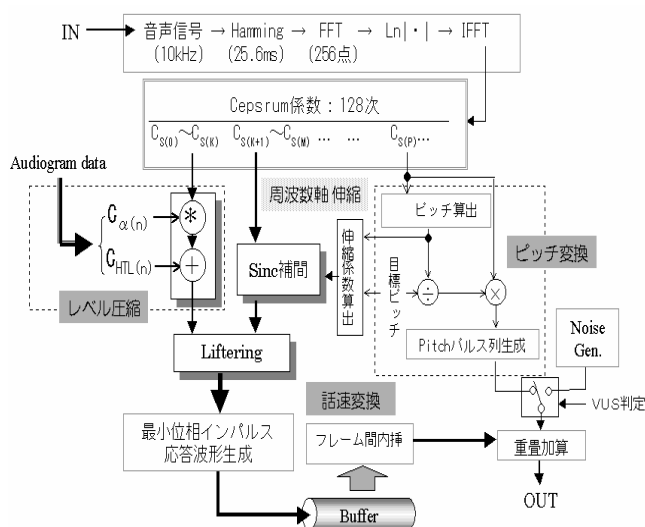


Fig.1 Block diagram of the compound hearing aid system

図1に複合補聴処理の概要を示す。基本的にはケプストラム法に基づく分析合成方式であり、フレーム単位で抽出されるピッチ周期(音源特性)や

最小位相インパルス応答(擬似声道特性)の加工による声質変換やスペクトル圧縮を行い、合成時のパラメータ群の時間補間による話速変換を実現するものである。

3. 話速変換処理

3.1 原理

本手法では合成の際、生成したピッチパルス列に応じて、最小位相インパルス応答を畳み込み演算を行っている。そこで、遅速を与えるために、分析部のWRITEヘッドにおいてフレーム毎に抽出されたインパルス応答及びピッチ周期をフレームバッファに格納し、合成部のREADヘッドをWRITEヘッドとは独立に制御することで話速変換処理を行っている。図2に話速変換の原理図を示す。

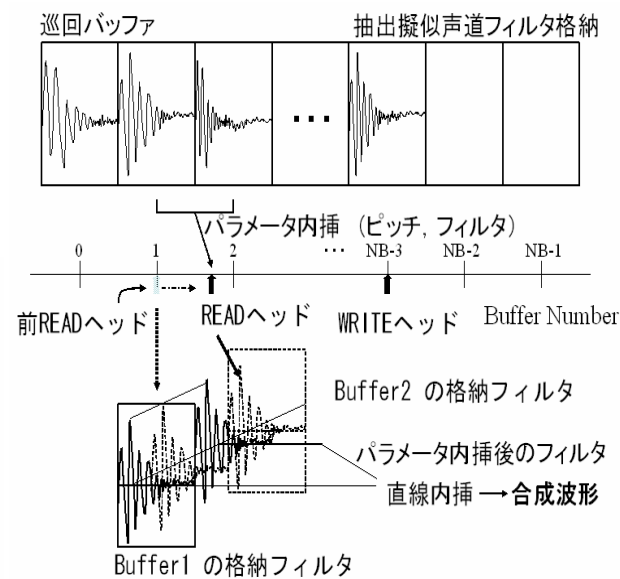


Fig.2 Principle of speech rate control procedure

尚、巡回バッファメモリ長NBフレームは半フレーム(10kHz, 128 サンプル)毎に読み出されるので、最大時間長

$$t_{\max} = NB \cdot 128 / 10 \cdot 10^3 = 12.8NB \text{ [ms]}$$

分の信号を格納でき、同時に最大遅れ時間 t_{\max} を与えることができる。

*Speech rate control procedure for the compound hearing aid processing

音声の各区間(V: 有聲, U: 無聲, S: 無音)に応じた処理を以下に述べ、具体的な話速変換処理の概念を図3に示す。

3.1.1 有声音区間

READ ヘッドの進み方を任意の速度係数 V_{spd} により決定し、その位置のインパルス応答及びピッチ周期をフレーム間の直線内挿により求める (図3 (a)) のときの速度係数と話速変化の関係は次の通りである。

速度係数 $[V_{spd}]$ $\left\{ \begin{array}{l} =1 \text{ 話速変換 なし} \\ >1 \text{ 話速変換 速} \\ <1 \text{ 話速変換 遅} \end{array} \right.$

3.1.2 無声音区間・無音区間

話速の変化には影響しないと考えられるので、Read・Writeのヘッドの間隔を一定にする (図3(b): 速度を1にする)。また、有声音区間において話速変換処理を行った際、入力音声と出力音声の時間的なズレが生じる事が考えられる。そこで、無音区間において時間吸収処理 (ある一定の無音区間が続くとその後の無音部を読み飛ばす) を行い、ズレが生じないように考慮している (図(c))。

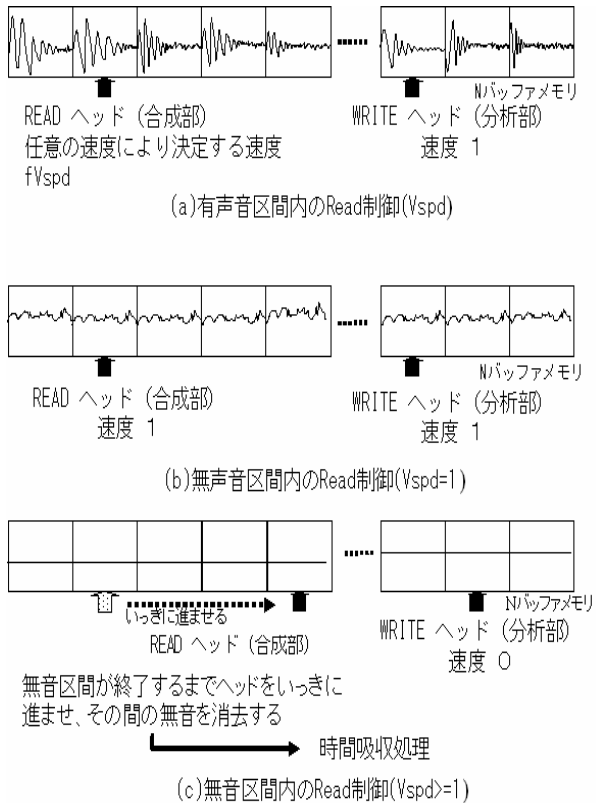
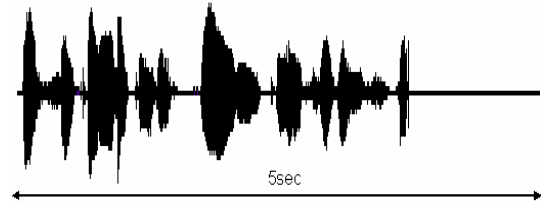
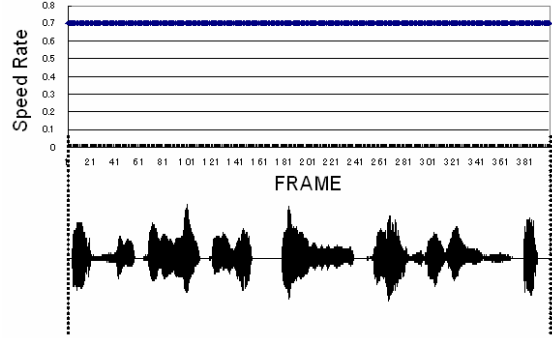


Fig.3 Speech rate control in the three kinds of segments

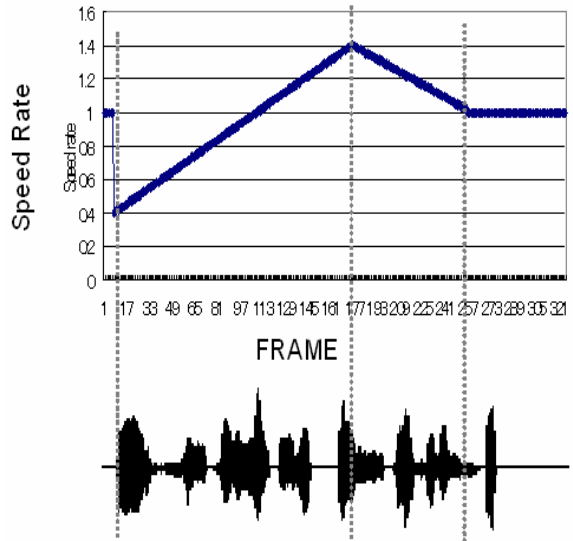
/iQsyuHkaNbakari nyuHyoHkuosyuzaisi ta/
 「一週間 ばかり ニューヨークを 取材した」



(a) Original speech signal ($V_{spd}=1.0$)



(b) Speed rate and its resynthesized speech in the fixed rate speed control ($V_{spd}=0.7$)



(c) Speed rate variation and its resynthesized speech in the preset initial rate speed control (initial $V_{spd}=0.4$)

Fig.4 An example of the groval speech rate control

3.2 話速変換処理方式

本話速変換処理では、二つの方式: 大局変換と局所変換を実装する。

3.2.1 大局変換

一定の速度係数を与え、有声音部のみで遅らせ、無声部・無音部は速度係数 1 とするという方式であり、2通りの変換モードを有する。

① 定速度変換

合成信号の有声音部を定速度率で遅らせる変換である。老人性難聴者(60代~80代)に対する評価実験により、6mora/sec が最も高い評価との報告がある^[3]。また、ニュース音声程度では話速 8.5mora/sec であるので、話速変換係数 V_{spd} 最低値は $0.7(6/8.5=0.70)$ に設定する。図4に入力音声「一週間ばかりニューヨークを取材した」(図4(a))の波形、定速度($V_{spd}=0.7$)で処理音声(図4(b))の波形を示す。

② 初速設定変換

入力音声と処理音声との時間的ずれが次第に大きくなっていき、処理に問題が生じるため、入力音声の文頭有声音開始時の初速($V_{spd}<1$)を設定し、徐々に NORMAL($V_{spd}=1$)に近づけ、遅延に伴う遅れを回復するために加速($V_{spd}>1$)し、実時間に達したら元の NORMAL($V_{spd}=1$)に戻るといふ処理である。初期速度 0.4 に設定した場合処理音声波形は図4(c)に示す。

3.2.2 局所話速変換

局所話速変換処理では、有声音部全体に対し話速変換を行うのではなく、特にエネルギーが弱く聞き取りが困難な無音破裂音に着目し、この音素直前無音区間を挿入し(破裂音がそれに先行する音素から受ける影響 forward masking を軽減する目的)、同時に直後の有声音(母音)のレベルを下げると共に、速度を遅らせる(後続母音が先行破裂音に及ぼす影響 backward masking を軽減する目的)方式である。

入力音声は「一週間ばかりニューヨークを取材した」を用い、本試料の中には4つの無音破裂音(“一週間”の/k/、“ばかり”の/k/、“ニューヨーク”の/k/と“取材した”の/t/)が存在する。VUS(有音・無音・無音)判定の結果は図5のようになる。

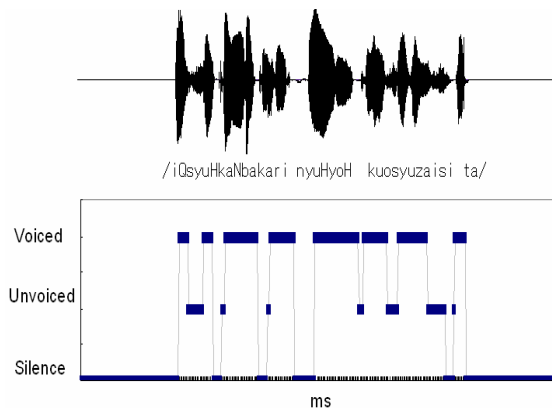


Fig.5 Result of VUS decision

無音(Silence)挿入フレーム数 mark1 及び無声音後の有声音部分の遅延・振幅減衰用フレーム数 mark2 の説明図は図6である。図のように、4つのところに5フレーム(mark1)の無音を破裂音の前に挿入し、直後の有声音(母音)のレベルを下げると共に、速度も遅らせることも分かる。

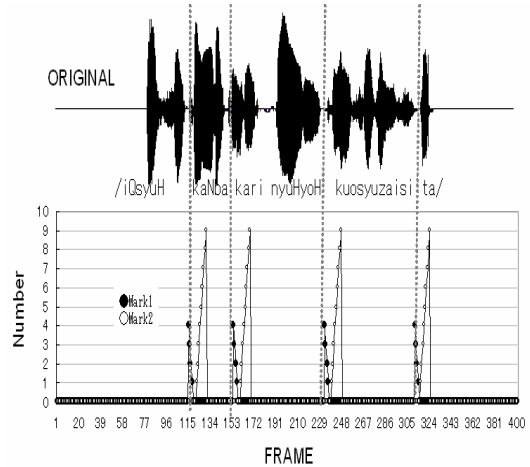


Fig.6 Variation of control parameters (Mark1, Mark2) in the local speed rate control

話速係数 V_{spd} の変化と変換された音声と入力音声の波形は図7ようになる。図に示すように有声音部分の速度を遅らせることが分かる。

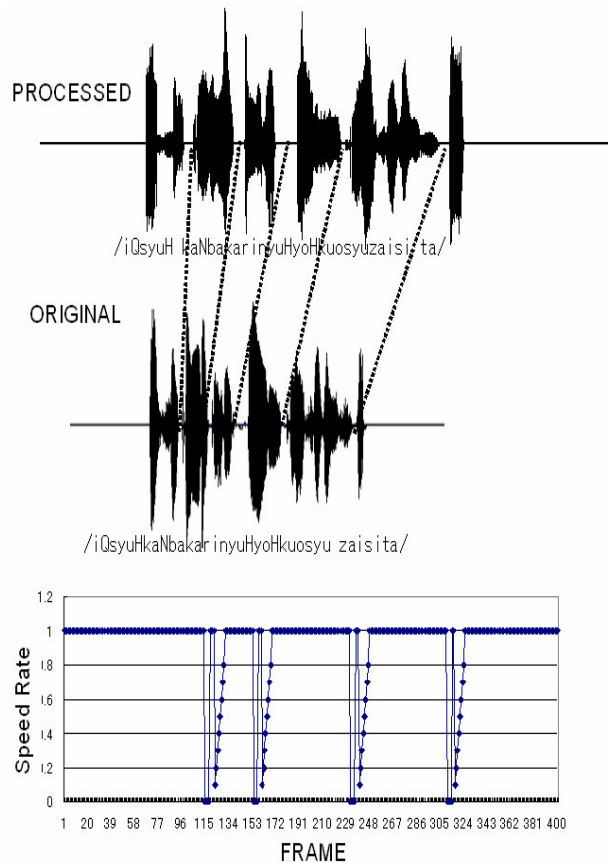


Fig.7 A result of the local speech rate control

部分的(試料の「一週間ばかりニューヨークを取材した」の/kanNba/部分)拡大すると,無音の挿入と有声音部分の遅延とレベルの低下の様子が確認できる.図8に示す.

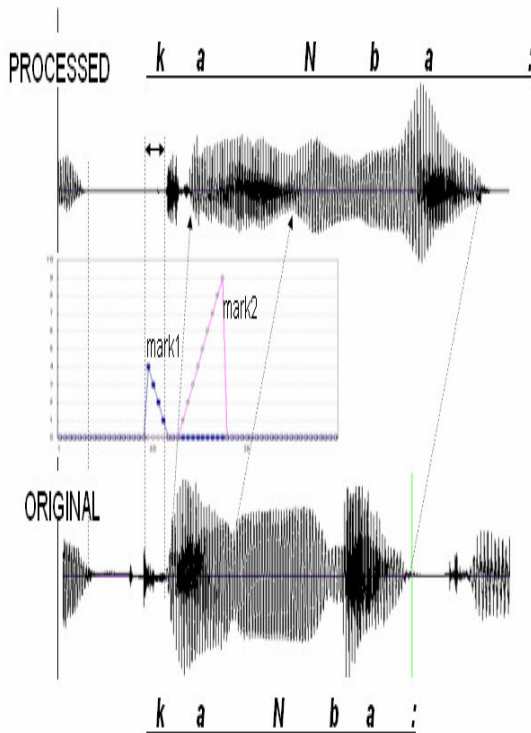


Fig.8 Pause insertion and delay of voiced segment in the local speech rate control

本複合補聴処理方式は分析部で抽出したインパルス応答・ピッチパルス列等のパラメータを一度バッファに格納し,合成部においてそれらのパラメータを使用し合成音声を作成する.その際,話速変換処理ではバッファから取り出す速度を可変とし,分析部でのバッファへの格納する速度と独立に制御することにより処理を行うわけだが,話速を遅らせるとバッファ内での分析部(READ)と合成部(WRITE)のヘッドアドレスに差が生じる.このままの状態では話速を遅らせていくと,バッファメモリ数に限度があるのでヘッドアドレスの逆転現象が生じてしまう.そこで,局所的な話速変換処理では話速を遅らせる毎に,その直後話速を速くすることにより残バッファメモリ数の改善処理をおこなう.READ ADDRESS と WRITE ADDRESS の関係を図9に示す,現在,バッファメモリサイズは350であり, $350 \times 12.8[\text{ms}] = 4.48[\text{sec}]$ の長さの音声のバッファリング即ち最大遅延量が可能である.

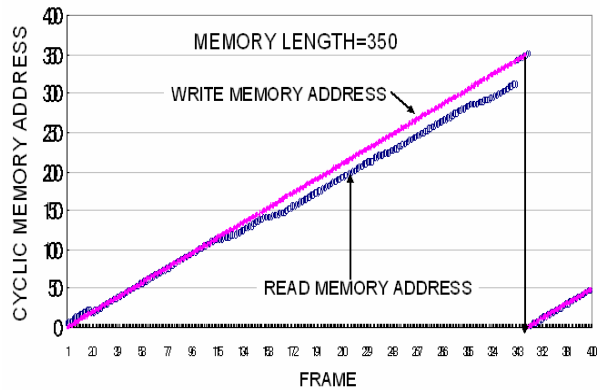


Fig.9 Variation of write memory address and read memory address in the local speech rate control

4. まとめ

本研究では,聴覚障害者に対する補聴手段として分析・合成系にケプストラム法を用い,スペクトル圧縮処理及びピッチ変換処理等の複数の処理を同時に又は選択的に行える複合補聴処理方式を提案した.またその付加機能として話速変換処理を提案し,その評価を行った.話速変換処理では十分な話速変化が確認することができた.これより,今回提案した手法は,難聴者に対してより有効な補償を行うことができるものだと考えられる.

今後の課題として難聴者による評価実験を行い,複合補聴処理方式としての効果の確認を行う予定である.

参考文献

- [1] Yuichi UEDA, Jun NISHIMURA, Takashi IKEDA and Akira WATANABE: "A compound hearing aid processing based on speech analysis and synthesis", proceedings of WESTPRAC VII, pp. 361-364, 2000
- [2] 張 丹, 藤本 雅樹, 坂田 聡, 上田 裕市: "複合補聴処理のためのフィッティング方式", 日本音響学会秋季講演論文集, 3-5-5, 2004年9月
- [3] 今井 篤, 他: "高齢者を対象とした話速変換音声の評価実験", 日本音響学会講演論文集, 1-2-22, 1999年3月