人工内耳のためのハイブリッド型音声符号化における有声/無声切替え方式の改良*

生山 雅人† 錦戸 暖† 豊島 広紀†† 佐藤 正幸††† 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 †††熊本県立技術短期大学校)

1. はじめに

人工内耳システムは，聴神経を構成する3万とも4万とも言われる神経線維を約20個の電極で代用して音声の伝達を行うため，聞こえる音に制限があり，不快な音として認識されてしまうことがある．本研究では，先に開発された逆フィルタ制御(IFC)法[1]を用いてフォルマント情報を伝達するFPEAK(Formant Peak)方式と，一般的に用いられているスペクトル情報を伝達するSPEAK(Spectrum Peak)方式を有声/無声によって切替えるハイブリッド型のスピーチプロセッサ[2] (図1)を用いて高品質の音声伝達を行なってきた．

しかし，有声/無声の誤判別や方式の切替え時点で滑らかさが損なわれてしまうことがあった．そこで，有声/無声判別に基づく符号化方式の切替えではなく，ハイブリッド方式の利点を残した改良を行い，音声に滑らかさを持たせる新たな符号化方式を提案する．

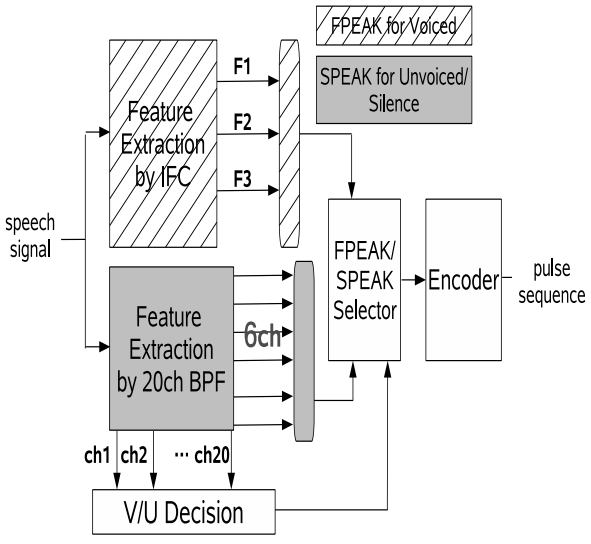

図 1: ハイブリッド方式のブロック図

2. ハイブリッド型音声符号化

2.1 SPEAK 方式

SPEAK方式によってフォルマントの存在しない無声音の特徴量抽出を行う．分析条件は，サンプリング周波数が12[kHz]，フレーム長が20[ms]，フレームシフトが10[ms] である．入力音声は，1フレームごとに20個のBPF群により，20帯域に分割される．遮断周波数はメルスケールで等間隔に設定される．各帯信号に対して全波整流を行い，遮断周波数400[Hz]のLPFに通して得られた包絡信号の実効値を求めその電極(チャネル)の振幅値とする．チャネルごとに求められた振幅値を大きい順に6つ選択し，その電極番号と振幅値が符号化パラメータとなる．図2にSPEAK方式による符号化例を示す．

2.2 FPEAK 方式

FPEAK方式によってフォルマントの存在する有声音の特徴量抽出を行う．FPEAK方式は本研究で提案する符号化法であり，IFC法によりフォルマント周波数の推定，抽出を行う．抽出された第1フォルマント周波数(F1)，第2フォルマント周波数(F2)が存在する周波数帯域のチャネル信号を符号化する．このときの振幅は各フォルマント周波数成分が存在する帯域の実効値となる．図3にFPEAK方式による符号化例を示す．左図の2つの矢印がIFC法によって抽出されたフォルマント周波数を表している．

【FPEAK方式の改良モード】

音声ピッチ周期に同期して刺激パルス列を伝達する現行の符号化法では20チャネル中，男声で6チャネル，女声では3チャネル程度が

∗ Improvement of voiced/unvoiced switching method in a hybrid type of speech processor for C.I.

By Masato Ikiyama†, Dan Nishikido†, Hiroki Toyoshima††, Masayuki Sato†††,
Tadashi Sakata† and Yuichi Ueda†
(†Graduate School of Science and Technology, Kumamoto University ††Faculty of Engineering, Kumamoto University †††Kumamoto Prefectural College of Technology)

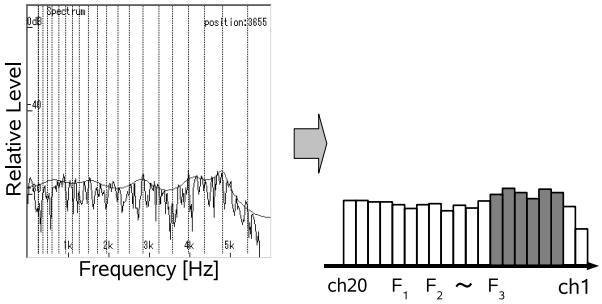

図 2: SPEAK 符号化方式例 (男声：/s/)

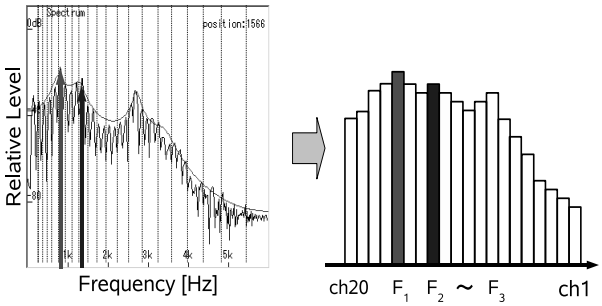

図 3: FPEAK 符号化方式例 (男声：/a/)

符号化チャネル数の限界である．そこで，最大 6 チャネルという条件下で，① フォルマントチャネルの分割 (1 つのフォルマントを複数のチャネルで表現する)，② 個人性を表すとされている第 3 フォルマント周波数 (F3) の付加の 2 点を考慮すると，図 4 の 6 つのモードが考えられる．これまでの研究において (5)F3+2CH-FPEAK と (6)F123-FPEAK 方式が有効であることがわかっている．

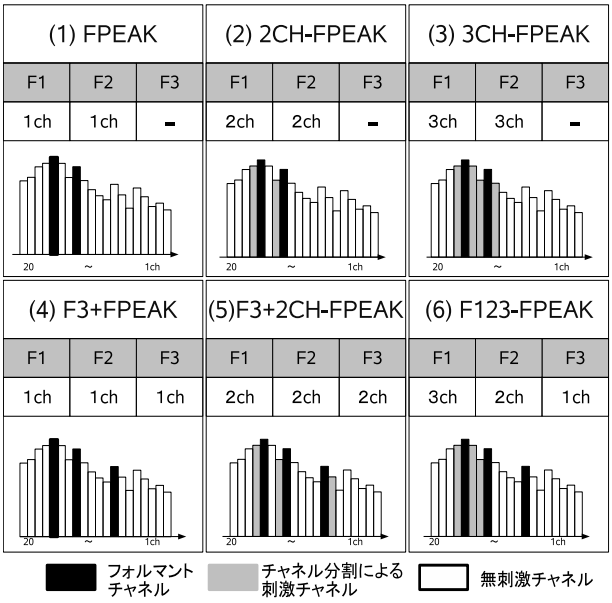

図 4: FPEAK 方式 (6ch 符号化) の符号化モード

2.3 ハイブリッド方式での問題点

現在，有声/無声判別を行い，FPEAK 方式と SPEAK 方式とをフレーム単位で切替えるハイブリッド型の符号化方式を採用している．有声/無声判別には線形判別関数を用いており，その関数の決定には，男女各 1 名の発話による VCV62 音節，計 124 試料を用いた [3]．しかし，少ない発話者による VCV 音節のみを用いて判別関数を決定しているため，連続音声中のフレームでは誤判別のために，有声音フレームに対して SPEAK 方式の符号化を行いフォルマント情報を正しく伝達できないことがある．また，FPEAK 方式と SPEAK 方式が切替わるフレームで選択チャネル位置に急激な変化が起こり，音声の滑らかさが損なわれることがある．

3. ハイブリッド方式の改良

有声/無声の判別を行わずに，有声から無声，無声から有声への切替えが滑らかで，ハイブリッド方式と同様に有声部ではフォルマント情報を伝達することができる新たな方式を提案する．

3.1 重みつき符号化

図 5 は新たな符号化方式のブロック図である．入力音声から SPEAK 方式に基づく 20 チャネル分の振幅値を得る．同時に IFC 法によるフォルマント周波数の推定，抽出を行い，フォルマントが存在する 3 つのチャネル以外の 17 個のチャネルのレベルには重み係数 w(0.0 ～1.0) を乗ずる．その後，重み付けされた 20 チャネル信号の中で大きい方から 6 チャネル分を選択し符号化する．

3.2 符号化処理例

図 6 に重みつき符号化例を示す．w =1.0 のときは BPF 群出力から得られた振幅値のまま，大きい振幅値の 6 チャネルが選択されるので，従来の SPEAK 方式と同じ符号化になる．w =0.5 のときはフォルマントチャネル以外の振幅値が減少した状態で 6 つのチャネルが選択される．w =0.0 のときはフォルマントチャネル以外の振幅値は 0 となり，フォルマントチャネルを用いる FPEAK 方式と等価で

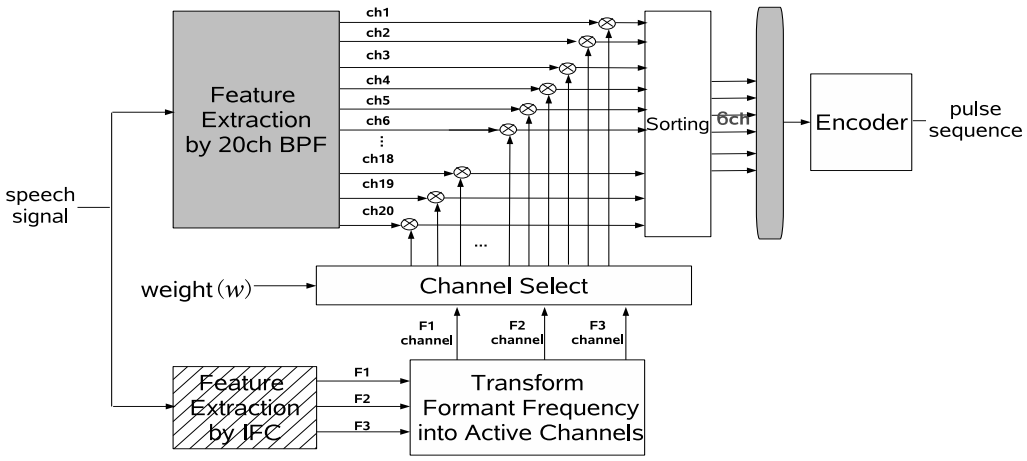

図 5: 改良方式のブロック図

ある．ただし，図 4(4) の F3+FPEAK 方式と同じ符号化であり，このとき選択されるのは 3 チャネルのみで，その他の重みで符号化した音声や図 4(5)，(6) よりも劣化したものとなる．

3.3 評価実験

新たに提案した重みつき符号化方式を評価するために，ケプストラム距離を比較した．w =1.0，0.5，0.1 として符号化した再合成音声，従来のハイブリッド方式によって符号化した再合成音声の 4 種類について，各音声と原音声のケプストラム距離をフレームごとに求め，12 次元ユークリッド距離で表現し，比較した．音声試料として，男声話者 2 名の 20 個の有意味単語，計 40 単語を用いた．

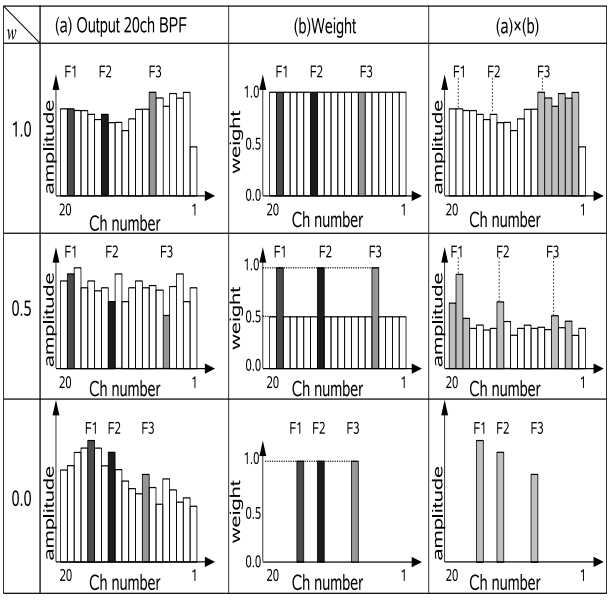

図 6: 重みつき符号化処理例

(1) 重みの違いの影響

図 7 は試料/daidokoro/についての，各符号化におけるケプストラム距離のフレーム変化を表しており，図 8 は図 7 の有声音フレーム，無声音フレーム，及び無音を除いた全フレームのケプストラム距離の平均を表している．例にあげている/daidokoro/のように，有声音フレームが多い試料のほとんどは w =1.0 で符号化した音声のケプストラム距離が最大となり，重みが小さくなるにつれケプストラム距離も小さくなる傾向が見てとれる．無声音フレームが多い試料は，w =1.0 で符号化したときのケプストラム距離が最小になるものもあったが，多くの試料で有声音フレームが多い単語と同様の傾向が見られた．また，全試料の平均である図 9 を見ても w =0.1 で平均距離は最小となり，最も原音声に近い符号化であることがわかる．

(2) ハイブリッド方式との比較

重みつき符号化で最も原音声に近づいた w =0.1 のときの符号化とハイブリッド方式のケプストラム距離を比較すると，ほとんどの試料においてハイブリッド方式のケプストラム距離が小さくなった．しかし，図 9 の平均を見ると両者の差は小さい．このことは，有声/無声判別を用いずに，図 5 のように FPEAK での非アクティブチャネルへの重みづけ処理のみで，従来のハイブリッド方式に近い結果が得られることを表している．

さらに，重みつき符号化ではフォルマントチャネル振幅値は SPEAK での値をそのまま

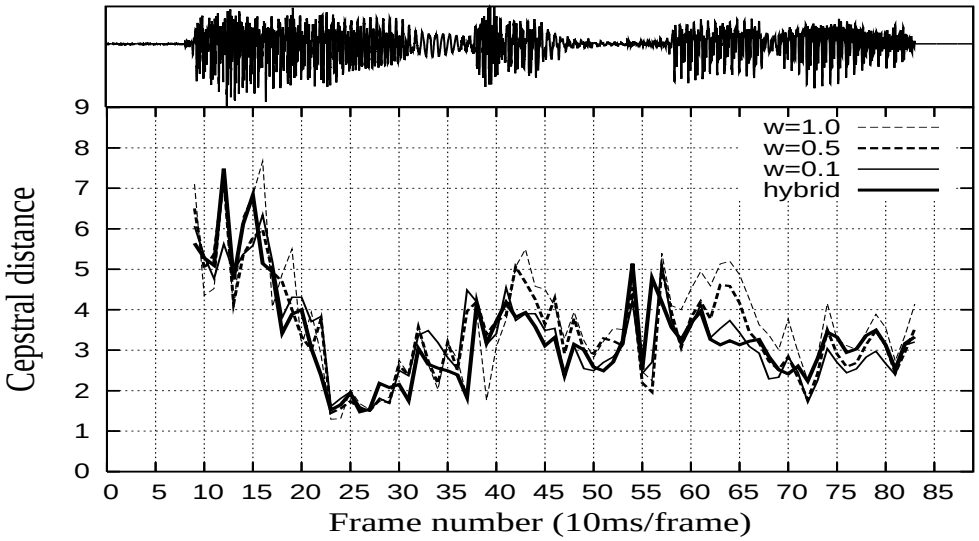

図 7: /daidokoro/のケプストラム距離

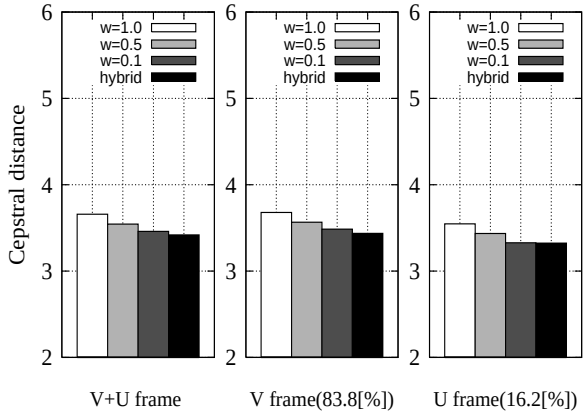

図 8: /daidokoro/のケプストラム距離の平均

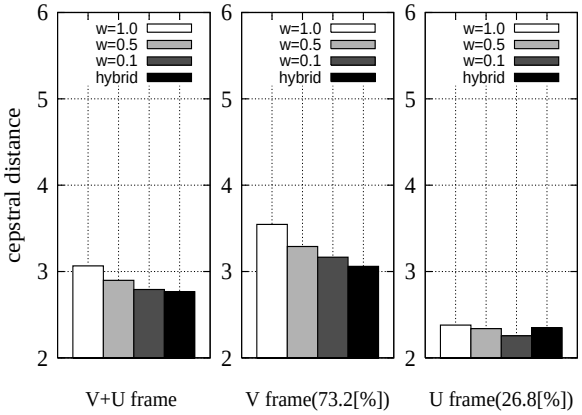

図 9: 全試料のケプストラム距離の平均

用いるため，無声音フレームでもフォルマントチャネルが選択されやすくなり，有声音フレームから無声音フレームに切替わるとき，あるいは逆のとき，急激なチャネル変化とフレーム間のレベル変化が抑えられることになる．そのため，従来のハイブリッド方式で滑らかさが損なわれていた音声を実際に聴取すると，重みつき符号化処理では改善されている傾向があった．

4. まとめ

本稿では，SPEAK 方式と FPEAK 方式を有声/無声によって切替えるハイブリッド型のスピーチプロセッサの問題点である，誤判別や方式の切替わり時点で滑らかさの欠如を解決するために，重みつき符号化方式を提案した．また，ケプストラム距離により評価した中では，新たな方式の最適な重み係数は 0.1 程度で，ハイブリッド方式に近い結果を得ることができ，有声/無声の切替わり時点で滑らかさを持たせることができた．

今後の課題として，より原音声に近い符号化を実現するためにチャネルの重みを一定にせず，従来の FPEAK 方式で良いとされていた，F3+2CH-FPEAK や F123-FPEAK 方式を模擬するような重み付けを行い，聴取実験を含めた評価を行なう必要がある．

参考文献

[1] Akira Watanabe, “ Formant Estimation Method Using Inverse-Filter Control ”, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 9, NO. 4, pp. 317-326, MAY 2001.

[2] M.Sato,T.Sakata,A.Watanabe, Y.Ueda, ”Formant Peak Stimulating Method based on Phantom Sensation for Cochlear Implant System,” Proc. of WESPAC IX,CD-ROM,hu-2-5-370,2006-6

[3] 上田裕市，西口直宏，坂田聡，佐藤正幸，渡邉亮，”人工内耳用ハイブリッド音声処理方式とその聴覚評価” 日本音響学会秋季研究発表会講演論文集 3-5-2, pp.479-480，2004.09