

逆フィルタ制御法を用いた フォルマント分解型音声強調方式の検討*

針尾 尚太[†] 立川 加奈絵^{††} 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. はじめに

音声の零交差周波数によって制御される逆フィルタ群を用いたフォルマント推定手法として、逆フィルタ制御法^[1]が開発され、音声画像化^[2]や単共振分解型補聴処理^[3]などに用いられている。本研究は、逆フィルタ制御法において副次的に得られるフォルマント成分信号の応用として、ピークレベル制御や先鋭化を行ったフォルマント成分信号から音声信号を再合成することで、雑音環境下に放射される音声の明瞭度の向上を目的としている。

2. フォルマント強調処理

2.1 フォルマント分解・再合成原理

フォルマント周波数は、音韻性知覚において最も重要な音声特徴量である。しかし、雑音環境下ではフォルマント成分のマスキングにより音声聴取が困難になる。そこで、雑音レベルに応じたフォルマントレベル制御を行う事で、音韻性の復元が期待できる。

逆フィルタ制御法では、音声信号をフォルマント成分信号に分解し、その成分信号から

フォルマント周波数を推定している。提案手法ではこのフォルマント成分信号のピークレベルを推定し、雑音レベルを考慮したピークレベル制御を行う。さらに、フォルマント成分信号にNBPFによる帯域幅制御を行い、各フォルマント成分信号のピークレベル近傍を急峻化させる事により、音声強調を実現する。Fig.1に本稿で提案する手法の音声強調の概念を示す。原音声の状態では各フォルマントのピークレベルは雑音によりマスキングされているが、強調処理を施した再合成音声のピークレベルは雑音レベルよりも高くなっている事がわかる。

2.2 フォルマントピークレベル推定

フォルマントピークレベルを制御し音声強調を行う上で、ピークレベルの正確な推定が必要になる。本稿ではこのピークレベル推定に狭帯域通過特性を有し、2次のIIR型フィルタで実現されるNBPF(Notch Band-Pass Filter)を用いている。NBPFの伝達関数を式(1)に示す。

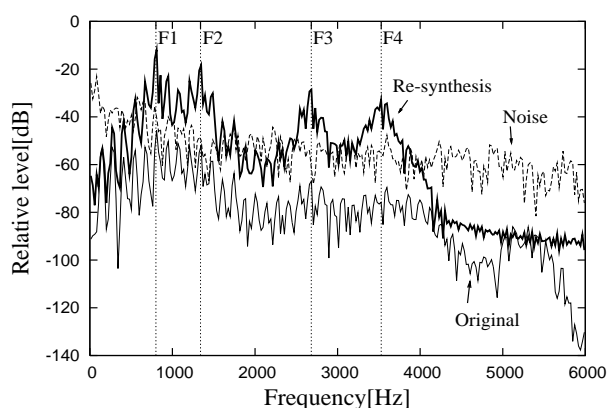


Fig. 1: Principle of the proposed speech emphasis method based on formants decomposition.

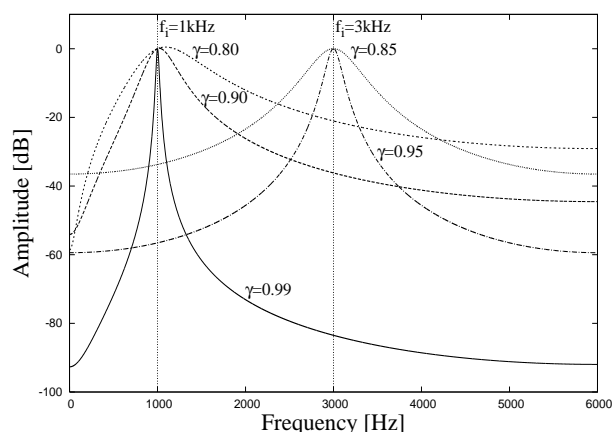


Fig. 2: Examples of NBPFs' frequency responses.

* A study of the formant resolution type speech emphasis method using Inverse-Filter Control.

By Shota Hario[†], Kanae Tatsukawa^{††}, Tadashi Sakata[†] and Yuichi Ueda[†]

([†]Graduate School of Science and Technology, Kumamoto University ^{††}Faculty of Engineering, Kumamoto University)

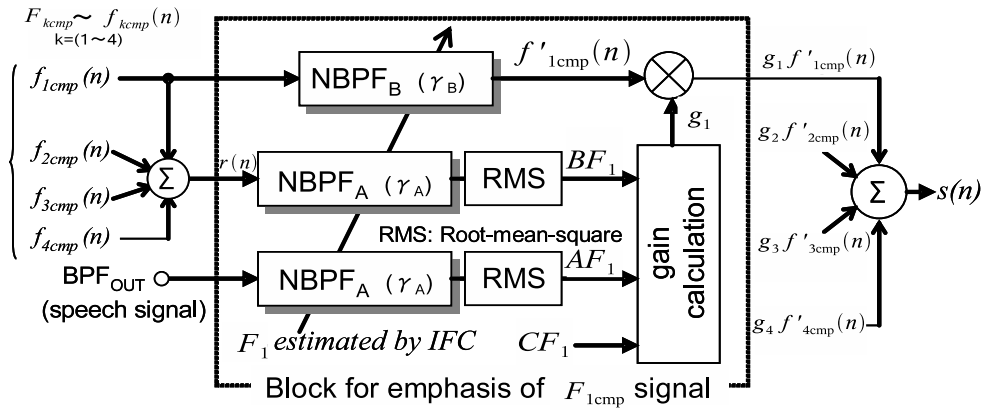


Fig. 3: Block diagram for emphasizing F_1 component signal in the proposed speech emphasis system.

$$H(z) = 1 - \frac{1 - 2 \cos \omega_i z^{-1} + z^{-2}}{1 - 2\gamma \cos \omega_i z^{-1} + \gamma^2 z^{-2}} \quad (1)$$

ここで, $\omega_i = 2\pi f_i / f_s$ である. γ は通過帯域幅を決定するパラメータ, f_s はサンプリング周波数 (= 12[kHz]), f_i は通過帯域の中心周波数である. NBPF の周波数特性例を Fig.2 に示す.

NBPF の中心周波数 f_i をフォルマント周波数に合わせることで, 当該フォルマント成分信号のフォルマントピークレベルが推定できる.

2.3 フォルマント制御による音声強調処理

Fig.3 に提案した音声強調手法の例として, 第1フォルマント成分信号の処理ブロック図を示す. まず, 逆フィルタ制御法の前処理として帯域制限が施された原音声 (BPF_{OUT}) から $NBPF_A(\gamma_A)$ を用いて原音声のフォルマントピークレベル AF_1 を推定する. 次に逆フィルタ制御法から得られた各フォルマント成分信号 ($f_{1cmp} \sim f_{4cmp}$) を再合成した $r(n)$ より, 同じ $NBPF_A(\gamma_A)$ を用いて再合成信号での第1フォルマントピークレベル BF_1 を推定する. 同様に, 別システムで入力される背景雑音の第1フォルマント位置でのピークレベル CF_1 を $NBPF_A(\gamma_A)$ を用いて推定する. 得られた AF_1, BF_1, CF_1 より, 第1フォルマント成分信号に乗じるゲイン g_1 が決定される. 計算された g_1 は, $NBPF_B(\gamma_B)$ によって帯域幅制御が施された第1フォルマント成分信号 (f'_{1cmp})

に乗じられる.

以上の処理が各フォルマント成分信号毎に行われ, 強調処理された第1から第4までのフォルマント成分信号の総和 $s(n)$ が各フレーム毎に出力される. 以下に各処理について詳しく説明する.

(1) フォルマントレベル制御

提案手法では, 各フォルマント成分信号のピークレベルを独立に制御し, 背景雑音の特性を考慮した強調処理が可能である. 再合成信号 $r(n)$ の第 k フォルマント成分信号のピークレベル BF_k を強調するためのゲイン g_k を以下の式から導出する.

$$g_k = 10^{G_k/20} \quad (2)$$

$$G_{ks} = AF_k - BF_k [dB] \quad (3)$$

$$G_{kv} = AF'_k + CF_k - BF_k [dB] \quad (4)$$

$$(AF'_k = (L_{max} - CF_k) AF_k / L_{max})$$

ここで, 式 (3) は原音声のピークレベル AF_k に各フォルマント成分信号のピークレベル BF_k を合わせる場合 (原信号の復元), 式 (4) は背景雑音のピークレベルを考慮したフォルマントレベル制御を行う場合に用いる.

式 (4) のみを使用してフォルマントレベル制御を行う場合, 非音声部分も一様にフォルマントレベルが上昇するため, 再合成音声に歪みが生じてしまう. そこで, この2つのゲイン計算式を組み合わせるハイブリッド利得制御方式を採用する. Fig.4 に本方式のブロック図を示す.

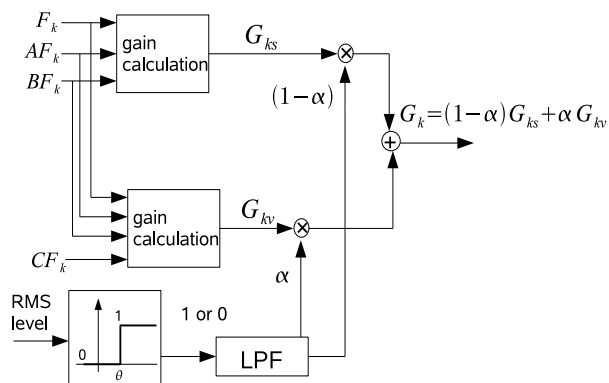


Fig. 4: Hybrid gain calculate system

まず、逆フィルタ制御法を用いてフォルマント抽出を行った際に得られる各フレーム毎の実効値をある閾値レベルで2値化する。その時系列信号を1次のIIR-LPFに通すことにより0~1の α が決定される。 α と $(1-\alpha)$ は2つのゲイン計算式 G_{kv} と G_{ks} それぞれに乘じられ、その和がフォルマントレベル制御に用いるゲイン G_k となる。

(2) フォルマント帯域幅制御

更に本手法では、フォルマント成分信号をNBPF(Fig.3 NBPF_B)に通すことでフォルマント帯域幅を狭めてフォルマント成分信号を急峻化し、フォルマント近傍のエネルギーのみを強調した再合成音声を得ることができる。

(3) 強調処理例

Fig.5はSNR(-5[dB])条件で、強調音声、原音声それぞれに背景雑音としてBubble Noiseを付加したスペクトルである。強調音声は、上述したハイブリッド利得制御方式を用いて、無雑音時の暗騒音レベルが雑音付加時の背景雑音レベルに対応(一致)するように、各フォルマントレベルを圧縮増幅(強調)するフォルマントレベル制御と、フォルマント帯域幅制御を第1から第4フォルマント分析信号に施し、再合成した。提案手法を用いる事でフォルマント付近のパワーが急峻化され、背景雑音によってマスキングされていた原音声のピークレベルのみが上昇している事がわかる。

Fig.6はフォルマントレベルの散布図で、例として男声1名のVCV音節62個(計1769フレーム)に同様の強調処理を施し、ピークレベルが処理前と処理後でどのように変化した

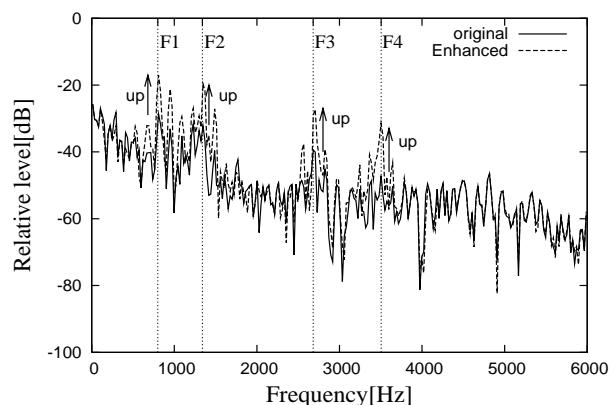
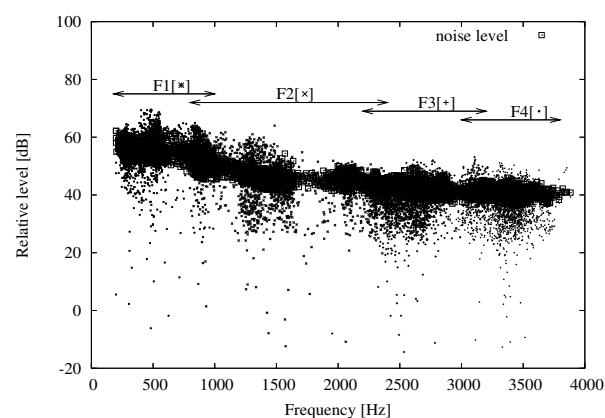
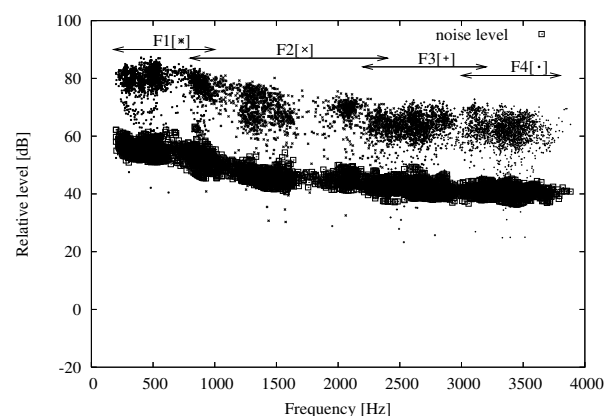


Fig. 5: Example of the original and re-synthesized speech spectra by formant peak level control and formant bandwidth control.



(a) Distributions of original formant peak level.



(b) Distributions of emphasized formant peak level.

Fig. 6: Distributions of formant peak level before(a) and after(b) emphasis.

かを表している。(a)は原音声のピークレベル AF_1 から AF_4 までと背景雑音のレベルをプロットし、(b)は各フォルマント成分信号のピークレベル BF_1 から BF_4 に、計算され

た G_k を足し合わせたレベルと、背景雑音のレベルをプロットしている。音声と背景雑音の SNR は約-5[dB] で、想定した背景雑音は Bubble Noise である。フォルマントレベル制御により、フォルマントレベルは設定したダイナミックレンジ ($L_{max}=100$ [dB]) に収まるように制御され、背景雑音によりほとんどのピークレベルがマスキングされていた処理前 (a) から、処理後 (b) に示すように多数のピークレベルが圧縮増幅している事がわかる。

3. 再合成音声の主観的評価

提案手法で再合成した音声、実際に雑音環境下で聴取した際に原音声より明瞭度が向上しているかを予備的聴取実験から検討した。使用した音声試料は男声 1 名の VCV 音節 62 個の原音声、強調処理を行った再合成音声のそれぞれに Bubble Noise を付加した音声である。実験手順はヘッドホンからランダムに呈示される 62 個の音声を 1 度だけ聞いてもらい、書き取りテストを行う。

Fig.7 に示すのは被験者 2 名 (男性 1 名、女性 1 名) の各 SNR での単語正答率の平均を示している。低 SNR 時に原音声に比べ、提案手法で再合成した音声の方が高いスコアを示している事がわかる。また高 SNR においては再合成音の歪みのため、原音声よりわずかにスコアが低下する傾向が見られる。

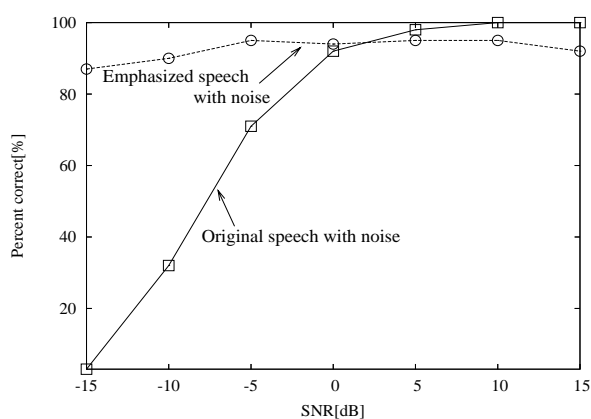


Fig. 7: Result of percent articulation test with VCV syllable before and after emphasis process.

4. まとめ

本稿では、逆フィルタ制御法において副次的に得られるフォルマント成分信号のピーク

レベルを制御し、雑音環境下での明瞭度が向上した聞き取り易い音声を再合成する音声強調手法を提案した。

提案手法では、フォルマントピークレベルの推定に NBPF を用い、推定されたピークレベルに対してフォルマントレベル制御を行う事でマスキングされないレベルまでピークレベルを向上させることができる。フォルマントレベル制御は 2 つのゲイン導出式を用いるハイブリッド方式を採用する。これにより、無音の部分を強調することなく、歪みが少ない再合成音声を得ることができる。

また、NBPF による帯域幅制御を加えることにより、フォルマント成分信号をフィルタリングして急峻化し、フォルマント近傍の周波数成分のみを強調する。これによって、十分な単共振特性を有さない信号であっても精度の高い再合成が可能となり、音韻性の向上も見込める。

本手法から得られた再合成音声を用いた聴取実験では、低 SNR において原音声よりも明瞭度の向上が見られ、提案手法の有用性が確かめられた。

謝辞

本研究において、有益な御指導と助言を賜りました渡邊亮 博士に心から感謝致します。

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] A.Watanabe, S.Tomishige, M.Nakatake, "Speech Visualization by Integrating Features for the Hearing Impaired", IEEE Trans. Speech and Audio Processing, Vol. 8, No. 4, pp. 454-466, 2000
- [3] 池田隆, 上田裕市, 渡邊亮, "高品質音声を伝達する単共振分解型補聴器", 日本音響学会誌 57 巻 5 号, pp.326-336, 2001-5
- [4] Y.Ueda, T.Hamakawa, S.Hario, T.Sakata, A.Watanabe, "Speech Emphasis Method Based on Peak Level Control of Formant Component Signal", Proc. of WESPAC IX, CD-ROM, pr-2-1-340, 2006