

# フォルマント分解信号のピーク制御による 音声強調方式の検討\*

濱川 智哉<sup>†</sup> 坂田 聡<sup>††</sup> 渡邊 亮<sup>†††</sup> 上田 裕市<sup>†</sup>

(<sup>†</sup>熊本大学大学院自然科学研究科 <sup>††</sup>熊本大学工学部 <sup>†††</sup>熊本県立技術短期大学校)

## 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<sup>[1]</sup>が開発され、音声画像化<sup>[2]</sup>や単共振分解補聴処理<sup>[3]</sup>などに用いられている。本稿では、逆フィルタ制御法において得られる単共振成分の応用として、ピークレベル制御や先鋭化により、音声のフォルマント成分を独立に強調し、音声の明瞭度を改善する手法を提案する。

## 2. フォルマント強調処理

### 2.1 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パラメータである。しかし、雑音環境下において雑音レベルがフォルマント成分のレベルより高ければ、マスキングにより音声聴取が困難になる。逆に、フォルマントレベルを雑音レベルよりも高くすることができれば、音韻性の復元が期待できる。

逆フィルタ制御法では、音声信号から各フォルマントに対応する擬似単共振信号を取り出し、その単共振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、背景雑音の周波数特性を考慮し、各単共振信号のピークレベルを独立に操作することで、明瞭度を改善した音声の再合成を行う。また、ピークレベルの制御に加えて、単共振成分のNBPFフィルタリング(後述)によりフォルマント帯域幅を狭めることで、フォルマントピーク近傍の成分のみを強調する。

本手法により、雑音環境下で放出される拡声器などの間接音声の明瞭度を改善することができると考えられる。また、フィルタリングを用いたフォルマント近傍帯域の間接的強調ではなく、フォルマント周波数を抽出し、そのフォルマントレベルのみをピンポイントで制御対象とすることで強調特性の改善が期待できる。提案する音声強調の概念を Fig.1 に示す。

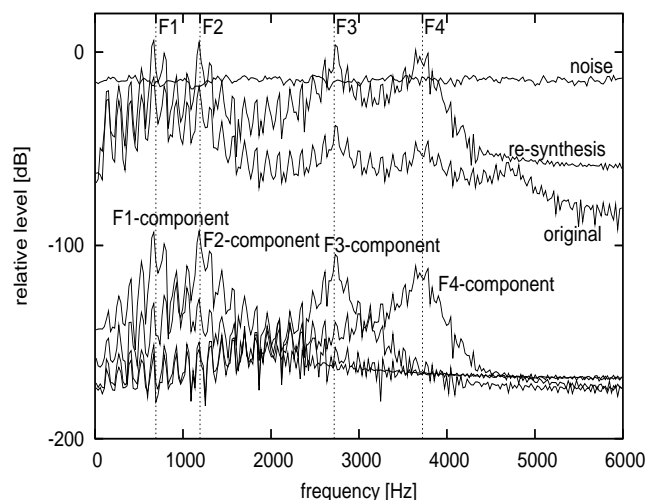


Fig. 1: Principle of the proposed speech emphasis method based on formants decomposition.

### 2.2 NBPFによるフォルマントピークレベルの推定

フォルマントピークレベルを制御することで音声強調を実現するには、ピークレベルの正確な推定が必要になる。フォルマントピークレベルの推定には、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の伝達関数を以下に示す。

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

ここで、 $\omega_i = 2\pi f_i / f_s$ 、 $\gamma$ 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(=12kHz)、 $f_i$ 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。

逆フィルタ制御法により第1から第4までのフォルマント周波数が抽出されると、それぞれのフォルマント周波数で共振する単共振信号が得られる。そこで、各単共振信号において、対応するフォルマント周波数を $f_i$ に設定することでフォルマントピークのレベルが推定できる。

\* A study of the speech emphasis method based on a formant peak control.

By Tomoya Hamakawa<sup>†</sup>, Tadashi Sakata<sup>††</sup>, Akira Watanabe<sup>†††</sup> and Yuichi Ueda<sup>†</sup>

(<sup>†</sup>Graduate School of Science and Technology, Kumamoto University <sup>††</sup>Faculty of Engineering, Kumamoto University <sup>†††</sup>Kumamoto Prefectural College of Technology)

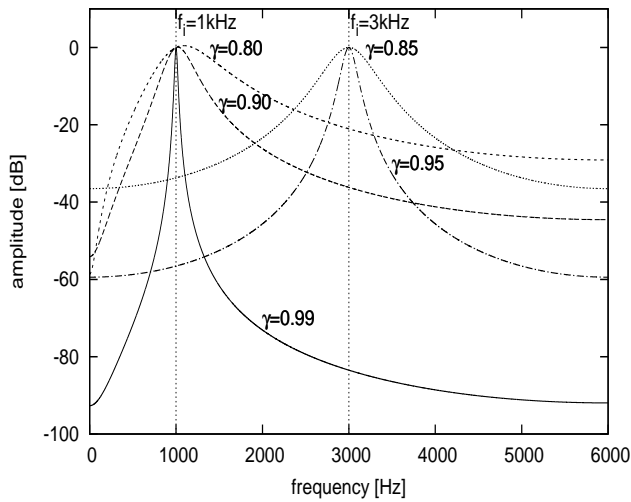


Fig. 2: Examples of NBPFs' frequency responses.

### 2.3 NBPF 特性の制御

NBPF によって正確なフォルマントピークレベルを推定するには、式 (1) におけるパラメータ  $\gamma$  の決定法が重要となる。

そこで、フォルマント周波数とその帯域幅の関係式として知られている Fant の理論式を用いる。

$$BW_i = BW_0 \left( 1 + \frac{F_i^2}{6 \times 10^6} \right) \quad [\text{Hz}] \quad (2)$$

ここで、 $F_i$  はフォルマント周波数 [Hz]、 $BW_i$  はその帯域幅 [Hz] であり、 $BW_0$  は男声の場合 50Hz、女声の場合 80Hz に設定する。式 (2) より、以下の式を得る。

$$F_i = \sqrt{6 \times 10^6 \left( \frac{BW_i}{BW_0} - 1 \right)} \quad (3)$$

次に、フォルマント周波数からピークレベル推定に最適なパラメータ  $\gamma$  を決定するために、以下のように多項式近似できると仮定する。

$$\gamma = a_0 + a_1 F_i + a_2 F_i^2 + a_3 F_i^3 + \dots \quad (4)$$

NBPF の周波数特性 (Fig.2) より、任意の  $\gamma$  における帯域幅  $BW$  が実測でき、 $BW$  と式 (3)、式 (4) から  $F_i - \gamma$  関数の実測値が得られる (Fig.3 'x', 'x')。その実測値を基に、重回帰分析により次の実験式を得る。

$$\gamma = 0.978 + 1.34 \times 10^{-6} F_i - 4.38 \times 10^{-9} F_i^2 + 1.57 \times 10^{-13} F_i^3 \quad (\text{for male}) \quad (5)$$

$$= 0.964 + 3.10 \times 10^{-6} F_i - 7.53 \times 10^{-9} F_i^2 + 4.06 \times 10^{-13} F_i^3 \quad (\text{for female}) \quad (6)$$

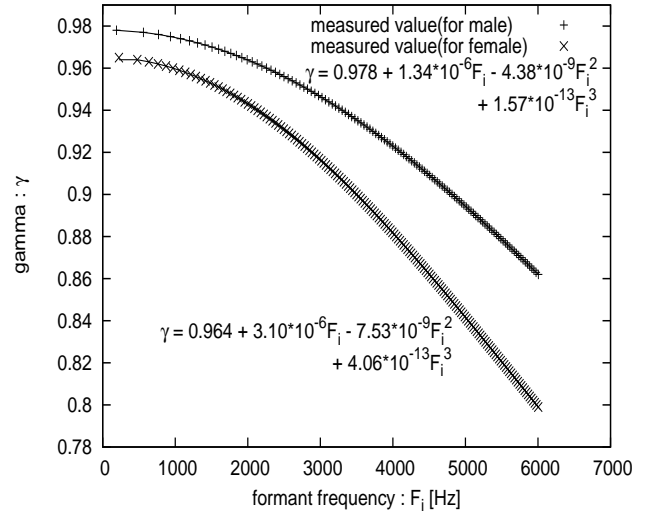


Fig. 3: Optimal gamma values( $\gamma$ ) for NBPF characteristics and those regression curves.

### 2.4 音声強調処理ブロック

提案する音声強調手法のブロック図を Fig.4 に示す。

まず、逆フィルタ制御法により得られた擬似単共振信号とフォルマント周波数から、NBPF の通過帯域幅を決定するパラメータ  $\gamma_{nA}$  を導出し (式 (5) (6))、 $\gamma_{nA}$  を用いて原信号のフォルマントピーク  $AF_n$  と単共振信号のフォルマントピーク  $RF_n$  を推定する。事前に測定される背景雑音の周波数特性からフォルマント位置での雑音レベル成分  $C_n$  を導出しておき、 $AF_n$ 、 $RF_n$ 、 $C_n$  より、単共振信号のピークを強調するためのゲイン  $G_n$  を次式から導出する。

$$G_{rn} = AF_n / RF_n \quad (7)$$

$$G_{fn} = (\Delta AF + C_n) / RF_n \quad (8)$$

$$(\Delta AF = 6 \text{ dB/oct} \cdot \log_2(F_n / F_1))$$

ここで、 $G_{rn}$  は原音声再現する場合、 $G_{fn}$  は背景雑音を考慮する場合に用いるゲインであり、 $\Delta AF$  は積分処理で減衰するエネルギーである。一方、NBPF<sub>nB</sub> によりフォルマント近傍の周波数成分を強調し、急峻化された単共振波成分に算出したゲイン  $G_n$  を乗ずることで、フォルマントピークを強調・制御した単共振波を得る。4つの単共振信号において強調処理を施した後、重畳加算する。再合成音声は逆フィルタ制御法の前処理によって高域強調されているので、逆特性として積分処理をした後に最終的な強調音声を得る。

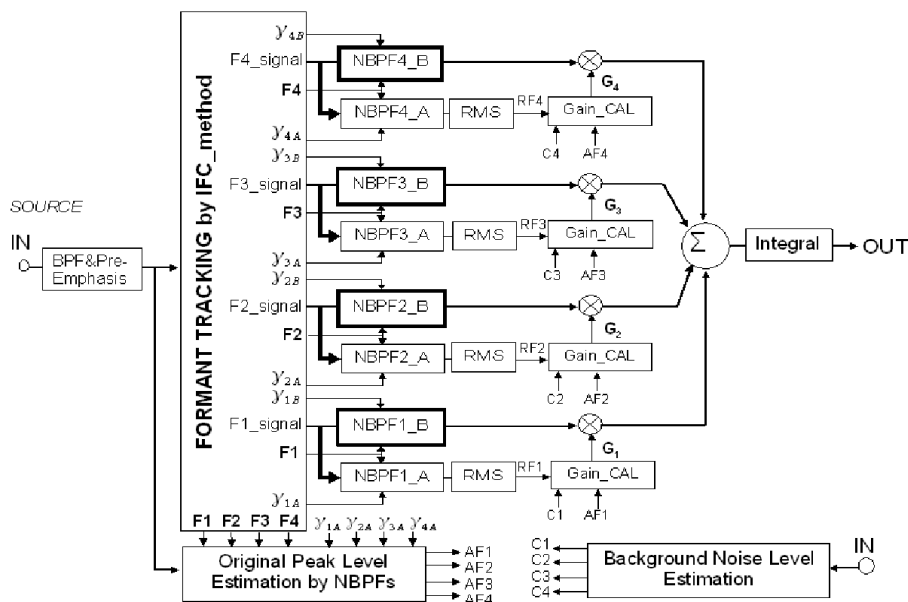


Fig. 4: Block diagram of the proposed speech emphasis system.

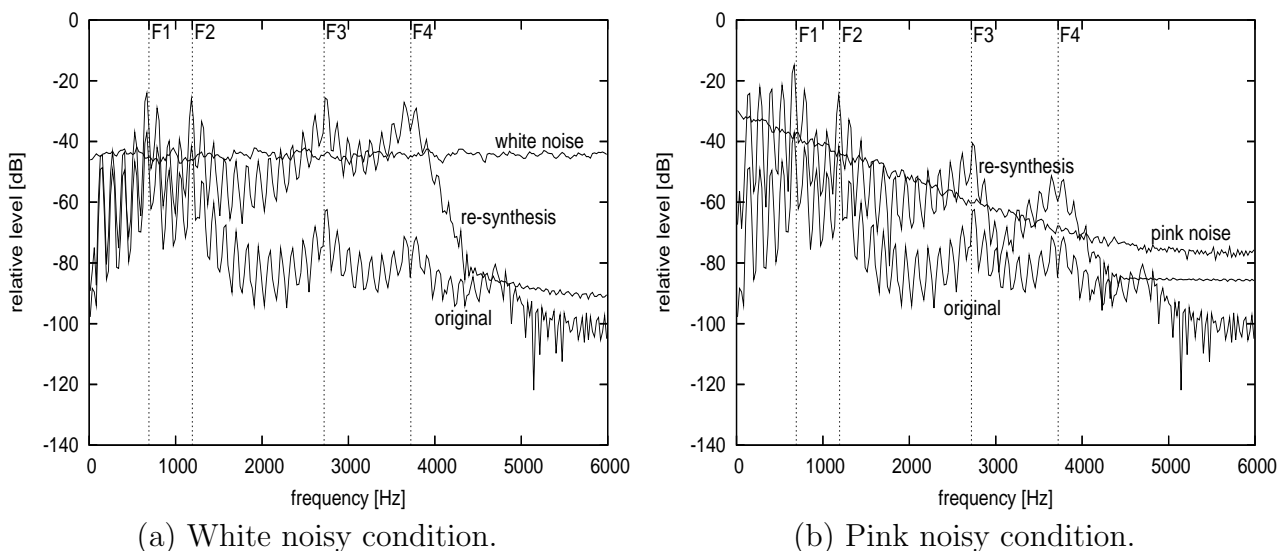


Fig. 5: Examples of the original and re-synthesized speech spectra by a formant peak level control under the background noisy conditions.

### 3. フォルマント強調処理例

#### 3.1 ピークレベル制御

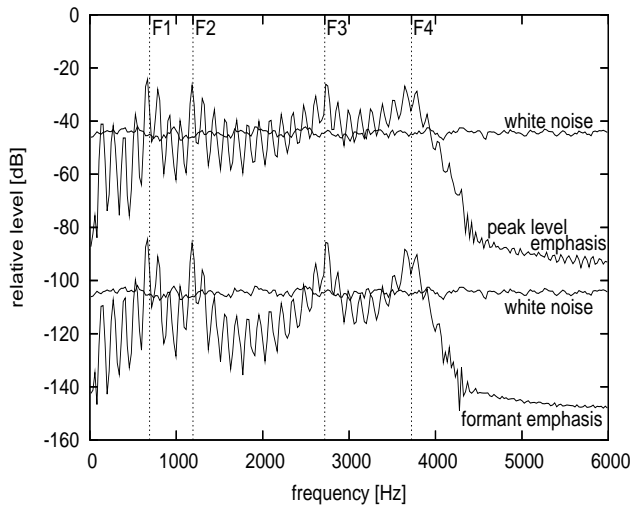
提案手法では、各フォルマント単共振信号のピークレベルをそれぞれ独立に操作することで、背景雑音の周波数特性に合わせてフォルマントを強調できる。つまり、白色雑音のように全ての周波数帯域においてエネルギーを持つ背景雑音であれば、再合成音声のフォルマントピーク包絡は全体的にほぼフラットになり、低い周波数帯域においてエネルギーが強いピンクノイズでは低次フォルマントのピークレベルが強調されることになる。

ホワイトノイズ、及びピンクノイズを背景雑音とした場合の制御例を Fig.5 (a), (b) に示す。

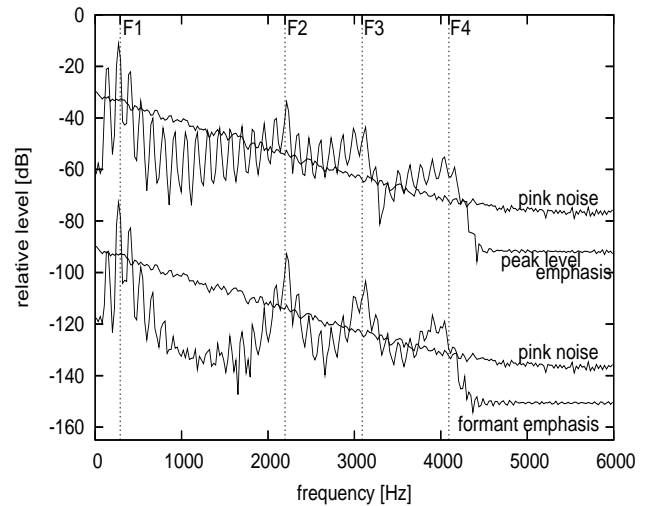
原音声（合成音声）と比較すると、全てのフォルマント周波数において、背景雑音のスペクトルよりも高い位置にピークレベルを制御できていることが分かる。これにより、背景雑音に強い再合成音声を得られると推測できる。

#### 3.2 フォルマント帯域幅制御

更に本手法では、NBPF ( Fig.4  $NBPF_{nB}$  ) によってフォルマント単共振信号をフィルタリングすることで音声を強調する。これにより、フォルマント帯域幅を狭めて擬似単共振信号を急峻化し、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。また、この帯域幅制御は強調音声の品質向上にも有効で



(a) White noisy condition.



(b) Pink noisy condition.

Fig. 6: Examples of the peak emphasized spectra and the formant sharpened spectra.

あると考えられる．逆フィルタ制御法から得られる擬似単共振信号には，フォルマントピークがはっきりしないものや，高次フォルマントの擬似単共振信号において低域エネルギーが強いものがある．これは，逆フィルタ制御法の目的がフォルマント周波数推定であり，その途中で生成される擬似単共振波群が十分な単共振特性を有していなくても高精度のフォルマント抽出が可能だからである．このような単共振信号に対してNBPFのフィルタリング処理を施すことにより，対応するフォルマント近傍の強調が可能となり，音韻性の高い音声再合成が期待できる．

Fig.6 (a), (b) にフォルマント帯域幅の制御例を示す．雑音として，(a) はホワイトノイズ，(b) はピンクノイズを用いている．フォルマント強調をしない再合成音声と比較すると，フォルマントピーク間の帯域でエネルギーが減衰し，ピーク近傍で急峻化していることが分かる．

#### 4. まとめ

本稿では，逆フィルタ制御法において生成されるフォルマント単共振信号のピークレベルを制御し，雑音環境下で聴き取り易い音声を再合成する音声強調手法を提案した．

フォルマントピークレベルの推定にはNBPFを用い，通過帯域幅を決定するパラメータ $\gamma$ は，Fantの理論式，及び最小二乗誤差による重回帰分析で導出された $F - \gamma$ 関数の近似曲線から求める．これにより，各フォルマント周波数に最

適な $\gamma$ を決定することができる．

提案手法は，各フォルマント単共振信号のピークレベルを独立に操作することで，背景雑音の周波数特性に応じて音声を強調することができる．また，フォルマント周波数の近傍帯域をBPFなどで強調する手法と比較すると，フォルマント周波数レベルのみを直接的に操作することで強調特性の改善も期待できる．更に，フォルマント単共振波をフィルタリングして急峻化することで，フォルマントピーク近傍の周波数成分のみを強調する．これにより，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．

#### 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, "ホルマント情報を伝達する聴覚障害補償システムの開発と音声の認知に関する研究", 熊本大学大学院自然科学研究科博士論文, 2002.