

前向き後向き差分関数とフィルタバンクを利用した音声信号の雑音低減

國枝 伸行[†] 島村 徹也[†] 鈴木 誠史[†]

Noise Reduction of Speech by Using Forward and Backward Difference Function and Bandpass Filters

Nobuyuki KUNIEDA[†], Tetsuya SHIMAMURA[†], and Jouji SUZUKI[†]

あらまし 我々は相関関数を利用した雑音低減方式について検討している。先に提案した SPAC や SPAD は 2 乗ひずみや高調波ひずみを生じ、明瞭性を低下させる欠点があった。また前向き後向き差分関数は、ランダム雑音に埋もれた単一正弦波をこうしたひずみを生じることなく強調できる。しかしながら、この関数を音声のような複合波に適用するとひずみが生じる欠点があった。本論文では、前向き後向き差分関数を音声強調に利用するため、フィルタバンクを利用した手法を提案する。本方式では、音声の調波構造からなることに着目する。すなわち、音声をフィルタバンクを利用して複数の帯域制限信号に分解し、それぞれの信号に対して前向き後向き差分関数によって高調波を強調する。強調した高調波を再合成することによって、音声強調が実現できる。本論文では、まず前向き後向き差分関数が帯域制限信号に対しても有効であることを示す。そして、帯域幅が 100 Hz のフィルタバンクを構成した提案法により、疑似音声の SN 比を 6~7 dB 改善できることを計算機シミュレーションによって求めた。本方式の実際の音声に対する効果を調べた結果、スペクトル包絡のピークを強調できることを確認できた。試聴実験の結果からは、提案法が SPAC や SPAD で生じるひずみを抑え、雑音低減できるという結果が得られた。

キーワード 前向き後向き差分関数, フィルタバンク, SN 比改善, 音声強調

1. ま え が き

音声通信を行う場合、伝送過程で雑音が混入したりひずみが生じて、しばしば SN 比が低下する。また、我々を取り巻く環境にはさまざまな音が溢れており、携帯電話や移動電話などの普及した今日では、こうした周辺雑音が音声通信において大きな障害となっている。

雑音によって品質劣化した音声を改善する技術は、これまでに数多く提案されている [1]~[16]。まず、古くからのウィナーフィルタや適応フィルタなどのさまざまなフィルタリングによるもの [2],[3] がある。また、スペクトル上で雑音を取り除くスペクトルサブトラクション法 [4],[5] や音声の調波構造を利用したコムフィルタ [6],[7] は、数多くの改良や検討が行われ

てきた。他にも AR モデルや正弦波モデルのパラメータを推定して音声を復元する手法 [8]~[11]、相関関数を利用して雑音低減する手法 [12],[13]、隠れマルコフモデル (HMM) を利用する手法 [14] など、その原理は実にさまざまである。

近年、音声認識の実用化に向けて、雑音対策の研究が活発に進められている [17]。しかしながら音声認識を目的とした研究では、認識処理部における距離尺度などの検討を行うものが多く、音声の復元を必ずしも目的としないものも多い。

これまでの音声強調 (Speech Enhancement) 技術には、それぞれさまざまな制約や欠点があり、その目的である全体的な品質の改善、明瞭度の向上、聞き手の疲労軽減などのうち、ごく限られた改善効果しか期待できない。例えば、スペクトルサブトラクション法やパラメータ推定による手法では、雑音特性の推定が必要であることが多く、コムフィルタでは正確なピッチ周期の検出が課題であった。

[†] 埼玉大学工学部情報システム工学科, 浦和市
Faculty of Engineering, Saitama University, Urawa-shi, 338
Japan

我々は相関処理を利用した音声強調方式として、自己相関関数を利用した音声処理方式 (SPAC) [12] や差分関数を利用した音声処理方式 (SPAD) [13] を提案してきた。これらの手法は、他の手法に比べて大幅な雑音レベルの低減が実現できるが、2乗ひずみや高調波ひずみが生じるために明瞭度が減少する欠点があった。その一方で、単一正弦波をひずみなく強調する前向き後向き差分関数を定義し、優れた雑音低減効果をもつことを示してきた [18]。しかしながら、この関数を音声のような複合波に適用するとひずみが生じる欠点を有していた。

本論文では、こうした2乗ひずみや高調波ひずみを抑えるために前向き後向き差分関数を音声強調に利用する手法を提案する。本方式では、音声が高調波構造をもつことに着目する。まず、フィルタバンクを用いて音声を各高調波が1本ずつ含まれる複数の帯域制限信号に分解する。そして、各帯域制限信号に対し前向き後向き差分関数によって、各高調波を強調する。こうして得られた信号を再合成すれば音声強調が実現できる。

本方式は、調波成分を強調するという意味においては、コムフィルタと同様である。しかしコムフィルタは、逐次変化する基本周波数を抽出し、高調波の周波数に対応する信号成分を強調するものである。そのため、基本周波数の抽出精度と雑音低減効果が結びついている。本方式では、信号の周波数は既知である必要がないという利点をもつ。また、フィルタバンクを用いた音声強調法としては古くに McAulay ら [15], [16] が試みている。しかしながら、これらの手法は各高調波の強調を目的としているのではなく、ウィーナーフィルタやスペクトルサブトラクション法のように各帯域における出力を変化させて雑音低減を実現するものであり、本手法とは本質的に異なるものである。

本方式では、前向き後向き差分関数が周期信号を強調するという原理に基づいているため、厳密な基本周波数の推定や雑音の統計的性質を調べる必要がないというメリットがある。

2. 提案法の原理

ここでは、まず前向き後向き差分関数が単一正弦波であればひずみなく雑音低減が行えることを示す。次に、この関数を音声のような複合波に対しても雑音低減が実現できるようにするため、フィルタバンクを利用した新しい音声強調方式を提案する。

2.1 前向き後向き差分関数の性質

時系列信号 $x(i)$ の前向き後向き差分関数を次式のように定義する。

$$d(j) = \frac{1}{N} \sum_{i=0}^{N-1} |2x(i) - x(i+j) - x(i-j)| \quad (1)$$

$$j = 0, 1, \dots, M.$$

今、入力信号が

$$x(i) = A \sin(\omega i + \theta) \quad (2)$$

で示される単一正弦波であるとき、式 (2) で計算される前向き後向き差分関数は

$$d(j) = \frac{4A}{\pi} \{1 - \cos(\omega j)\} \quad (3)$$

となる。この $d(j)$ は、振幅軸を $4A/\pi$ だけ移動すれば、入力信号と同じ周波数の正弦波として扱うことができる。そこで、入力波形 $x(i)$ の代わりに $d(j)$ を出力波形として切り出して接続していくことにする。こうして得られた信号は、位相は異なるものの、振幅が入力信号に対して線形に変化したものとして扱うことができる。

2.2 正弦波強調の原理

次に、ランダム雑音が重畳している信号を

$$x(i) = s(i) + w(i) \quad (4)$$

として表すものとする。ここで $s(i)$, $w(i)$ はそれぞれ正弦波信号と雑音を示す。この雑音の重畳した信号に対して式 (2) で示される関数をそのまま計算すると正弦波にひずみが生じる [18]。そこで、実際には次式のように変形した関数を用いる。

$$d'(j) = \frac{1}{N} \sum_{i=0}^{N-1} |2x(i) - x(i+j) - x(i-j)| + \alpha \cdot \text{Sign}(S_{ij}) \quad j = 0, 1, \dots, M \quad (5)$$

ここで、 S_{ij} は絶対値内の信号成分を表し、

$$S_{ij} = 2s(i) - s(i+j) - s(i-j) \quad (6)$$

である。このときの雑音成分は

$$W_{ij} = 2w(i) - w(i+j) - w(i-j) \quad (7)$$

で表される。また、

$$\text{Sign}(S_{ij}) = \begin{cases} 1 & S_{ij} \geq 0 \\ -1 & S_{ij} < 0 \end{cases} \quad (8)$$

で定義される。αは雑音によるひずみを解決するための定数であり、 W_{ij} の最大値を W_{max} として、

$$\alpha > |W_{max}| \tag{9}$$

を満たすものとする。

雑音を重ねている場合には、式(8)を直接求めることはできない。しかしながら、入力信号が雑音の重畳していない単一正弦波 $s(i)$ であるとき、

$$\text{Sign}(S_{ij}) = \text{Sign}(s(i)) = \text{Sign}(x(i)) \tag{10}$$

が成立する。そこで、 $\text{Sign}(S_{ij})$ のより確からしい値を得るために、ここでは3値の移動平均を利用し、

$$\text{Sign}(S_{ij}) = \text{Sign}(x(i-1) + x(i) + x(i+1)) \tag{11}$$

として計算することにする。

式(5)によって求めた前向き後向き差分関数 $d'(j)$ の様子を図1に示す。図1を見てわかるように、ランダム雑音成分が $d'(j)$ の $j=0$ 付近に集中し、そこから離れた部分では雑音低減されている。従って、入力信号 $x(i)$ の代わりに $j=0$ から離れた部分の波形 $d'(j)$ を出力すれば雑音低減された正弦波が得られる。

実際の処理においては、計算した前向き後向き差分関数の最小値から波形の切出し単位を設定し、オーバーラップさせて接続する。また、計算された $d'(j)$ はすべて正值であるため、出力する前に波形の平均値を計算し、その平均値が0となるように振幅軸を移動させる処理を行う。

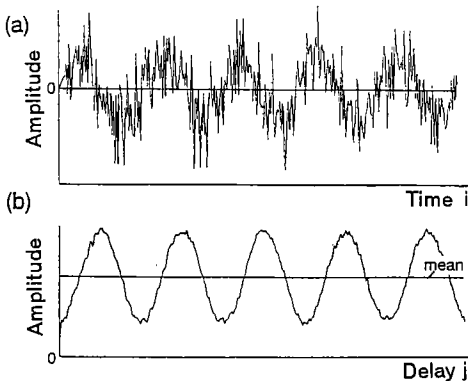


図1 前向き後向き差分関数による雑音低減の例
Fig.1 Noise reduction by forward and backward difference function.

2.3 フィルタバンクを利用した提案法の原理

よく知られているように、音声は調波構造をもっている。すなわち、適当なフィルタバンクを用いることによって、音声を複数の単一正弦波信号に分解できる。今、雑音で劣化した音声をフィルタバンクを用いて一つの高調波が含まれる帯域制限信号に分解したとする。帯域制限された各信号に、前向き後向き差分関数を適用すれば、各高調波は強調される。このときSPACやSPADで生じた2乗ひずみや高調波ひずみは生じない。そこで、強調されたすべての高調波を再び合成すれば音声強調が実現できる。これが提案法の原理である。

前向き後向き差分関数は、入力信号の振幅に対して線形に変化するため、各高調波のバランスは保たれる。また、位相は保存されないが、人間の聴覚は位相の違いに鈍感であるから音として聞いた場合の影響は少ない。

図2に提案法のシステム構成図を示す。まず、音声信号をフィルタバンクを用いて K 個の帯域制限信号に分解する。各帯域制限信号に対して、前向き後向き差分関数を計算し、周期信号である各高調波を強調する。音声波形の切出し単位 T は

$$D(j) = \sum_{n=0}^{K-1} d_n(j) \tag{12}$$

の最小値($j=15\sim 200$)によって設定する。ここで $d_n(j)$ は、 n 番目の帯域における前向き後向き差分関数を示す。このとき、 T が音声の基本周期あるいはその倍数に一致していればひずみは生じない。また、無声音やポーズ区間では周期性が見られないが、このときも同様に最小値 T を求めて切出し単位とする。

本方式では、フィルタバンクの帯域幅を固定している。そのため、高調波が1本も入らない帯域もあるが、その場合にも雑音レベルは低減される。更に、高調波がフィルタバンクの境界にくることもあるが、この場合はフィルタバンクによって二つの帯域信号に分割され、それぞれ強調される。従って、これらを再び合成して出力したときの振幅値が、入力に一致するようにあらかじめフィルタバンクを構成しておけばよい。

3. 雑音低減効果

ここでは、まず前向き後向き差分関数が帯域制限信号に対してどの程度の雑音低減が実現できるのかについて述べる。その後、フィルタバンクを構成した提案法による雑音低減特性について検討する。

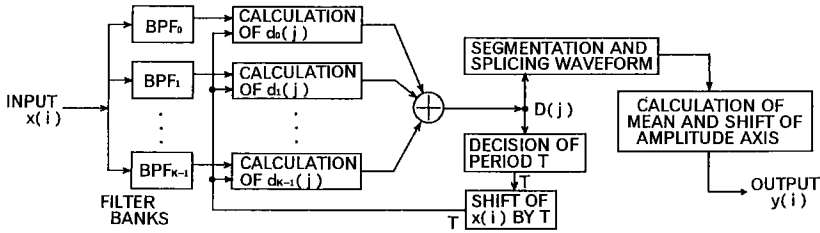


図2 提案法の流れ図
Fig. 2 Flowchart of proposed method.

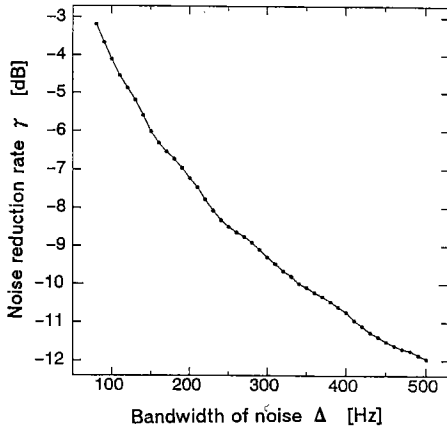


図3 帯域制限雑音の帯域幅と雑音低減比
Fig. 3 Noise reduction rate for band-limited noise.

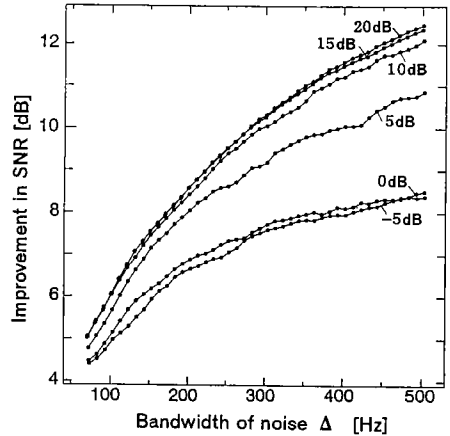


図4 帯域制限雑音を重畳した正弦波信号のSN比改善度
Fig. 4 Improvement in SNR for sinusoidal waveform degraded by band-limited noise.

3.1 帯域制限雑音のレベル低減

まず、前向き後向き差分関数が帯域制限雑音をどの程度レベル低減できるのかを計算機シミュレーションによって求めた (50 波形の平均によって算出)。入力雑音電力を P_{in} 、前向き後向き差分関数計算後の出力雑音電力を P_{out} としたときの雑音低減比 γ を次式で定義する。

$$\gamma = 10 \log\{P_{out}/P_{in}\} \quad (13)$$

まず、帯域幅を Δ Hz に制限した白色雑音に対する雑音低減比 γ を求めた結果を図3に示す。この図を見ると、帯域幅の広い雑音ほど雑音低減効果が大きくなるが、帯域幅 100 Hz の雑音に対しても約 -4 dB のレベル低減ができることがわかる。

次に、正弦波信号 $s(i)$ に帯域制限雑音を重畳したものを入力信号として考える。このとき、入力信号と出力信号のSN比をそれぞれ SNR_{in} 、 SNR_{out} とし

て、SN比改善度を次式で定義する。

$$SNR_{imp} = SNR_{out} - SNR_{in} \quad (14)$$

ここで、出力信号 $\hat{s}(i)$ のSN比は

$$SNR_{out} = 10 \log \frac{\sum s^2(i)}{\sum (\hat{s}(i) - s(i))^2} \quad (15)$$

によって計算する。雑音の帯域幅 Δ とSN比改善度の関係を入力信号のSN比ごとに計算した様子を図4に示す (正弦波の周波数は 500 Hz)。この図から、前向き後向き差分関数によって、帯域幅 100 Hz の雑音が重畳した正弦波のSN比を $5 \sim 6$ dB 改善できることがわかる。また、この改善度は入力信号のSN比が 5 dB 以下では小さくなっている。

以上の結果から、帯域制限信号の雑音レベル低減に対しても前向き後向き差分関数が有効であることがわかる。

3.2 フィルタバンク利用による雑音低減効果

次にフィルタバンクを構成した提案法による雑音低

表1 フィルタバンクの帯域幅と単一正弦波のSN比改善度 (dB)
Table 1 The relation between improvement in SNR (dB) of single sinusoidal wave and bandwidth of bandpass filters.

		信号周波数 (Hz)						Ave.
		75	200	500	1000	2000	3000	
帯域幅 Δ (Hz)	75	6.42	6.39	6.42	6.46	6.45	6.46	6.43
	100	7.10	7.06	7.13	7.00	7.08	7.05	7.07
	125	7.54	7.56	7.77	7.64	7.73	7.66	7.65
	150	7.99	8.01	8.03	7.91	7.95	8.11	8.00
	175	8.49	8.40	8.38	8.44	8.49	8.59	8.47
	200	8.90	9.20	8.95	9.09	9.17	9.16	9.08

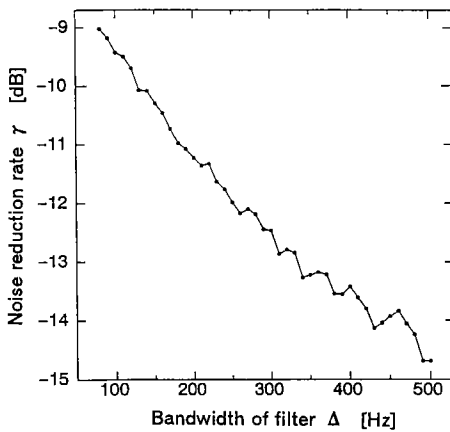


図5 提案法による雑音低減比
Fig.5 Noise reduction rate of proposed method.

減効果を前節と同様にして調べた。白色雑音を入力した場合のフィルタの帯域幅 Δ と雑音低減比 γ の関係を図5に示す。ここでは、提案法のフィルタはすべて 3.4 kHz までに帯域制限している。この図によると、今回提案している方法によっても帯域幅 100 Hz のフィルタバンクによって、約 -9.5 dB 程度の雑音レベル低減が行えることがわかる。また、単一正弦波信号を入力したときの SN 比改善度を表1に示す ($SNR_{in} = 10$ dB)。この表からも、帯域幅 100 Hz の提案法によって正弦波の SN 比を約 7 dB 改善できるという結果が得られた。以上のように、フィルタバンクと前向き後向き差分関数を利用することによって、雑音低減ができることがわかる。

次に複合波に対する効果を調べる。2.1 で見たように前向き後向き差分関数を計算すると、信号の位相がすべてそろう。そこで、式(14)で求められる SN 比改善度を計算するために、次式で示されるような疑似音

表2 提案法の疑似音声に対する SN 比改善度 (dB)
Table 2 Improvement in SNR of pseudo speech.

		基本周波数 (Hz)			
		150	200	250	300
帯域幅 Δ (Hz)	75	7.00	5.91	5.17	6.41
		6.36	5.87	4.94	6.27
	100	7.10	7.98	6.13	7.47
		6.61	7.24	5.77	7.23
	125	7.32	7.25	7.42	7.19
		7.08	7.16	6.47	7.19
150	---	8.04	7.09	9.14	
	---	7.65	6.63	8.34	
175	---	8.27	7.11	8.31	
	---	8.11	6.73	8.29	
200	---	---	7.71	9.13	
	---	---	7.24	8.76	

上段：ケース1、下段：ケース2

声を用意する。

$$s(i) = - \sum_n a_n \cos(n\omega i) \quad (16)$$

このとき $\omega/2\pi$ が基本周波数に相当する。また高調波は、最大 3.2 kHz までとした。各高調波の振幅 a_n の値は、高調波の次数 n にかかわらず一定値の場合 (ケース1) と、周波数特性が音声と同様の -10 dB/oct で減衰する場合 (ケース2) の両方について調べた。SN 比改善効果の結果を表2に示す ($SNR_{in} = 10$ dB)。疑似音声の基本周波数によって若干の変動はあるものの、いずれの場合も帯域幅が 100 Hz で約 6~8 dB の SN 比が改善できることがわかる。また、振幅特性が -10 dB/oct で減衰するケース2についても、ケース1にやや劣るものの同程度の改善が実現できる。このことから、実音声に対しても SN 比を約 6~7 dB 改善することが期待できる。

4. 音声強調への応用

本章では、提案法が実音声に対してどの程度効果が

表3 分析条件
Table 3 Datasets for analysis.

標準化周波数	10kHz
帯域制限	3.4kHz
AR次数 q	16次
フレーム長	25.6ms
フレーム窓	ハミング窓
フレーム周期	10ms

あるのかについて述べる。

4.1 フィルタバンクの構成

提案法によって雑音低減を実現するためには、フィルタバンクによって分けられたそれぞれの帯域内に正弦波が2本以上含まれないことが条件となる。本研究では、音声の基本周波数が100 Hz以下になることがまれであることを考慮して帯域幅を100 Hzに設定した。付加雑音は白色雑音とし、フィルタは501次FIRフィルタを用いている。本実験における分析条件は表3のとおりである。

4.2 実音声に対する雑音低減効果

まず、男性および女性の発声した約3秒の長母音/a/について提案法および先に開発したSPAC, SPADによって処理を行った。SN比が5 dB, 15 dBの女声について、各方式で処理したスペクトル包絡の様子を図6, 図7に示す。これらの図を見ると、SPACやSPADではスペクトル包絡のピークが過度に強調されているのに対し、提案法では原音声に近い形でピークが強調されている。その一方で図6のSN比5 dBの場合には、提案法では2.5 kHz付近の改善度が良くないことに気づく。これは、この付近の帯域のSN比が局所的に低くなり、前向き後向き差分関数による強調効果が小さかったものと考えられる。

このときの様子を表したのが図8である。この図は、SN比5 dBの音声について、音声と雑音の対数パワースペクトルの大きさを示している。図の下方には、このときのフィルタバンクの各帯域におけるSN比が示してある。図2に示すようなフィルタバンクにおける k 番目の帯域のSN比 SNR_k は、その帯域内における音声と雑音の対数パワースペクトルをそれぞれ P_{ks} , P_{kn} として、

$$SNR_k = P_{ks} - P_{kn} [dB] \tag{17}$$

によって求められる。

図8によると、1 kHz付近と2.5 kHz付近の帯域信号ではSN比で約30 dBもの差があり、図4で見たよ

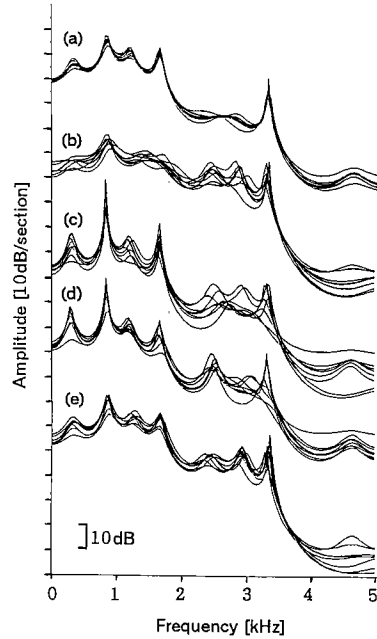


図6 女声/a/のスペクトル包絡
Fig.6 Spectral envelopes of /a/ (female).
(a) Noiseless speech. (b) Degraded speech (SNR = 5dB). (c) Enhanced speech by SPAC. (d) Enhanced speech by SPAD. (e) Enhanced speech by proposed method.

うに高調波の強調効果に差が出ることがわかる。特に、2.5 kHz付近の帯域ではSN比が-5 dBになり、信号の強調効果が低下する。一方SPACやSPADでは、広い帯域のまま処理するためこのような差が生じることはないが、2乗ひずみや高調波ひずみが生じる。

次に男女各3名の発声した約2~3秒の連続音声(データベース: "Multi-Lingual Speech Database for Telephonometry 1994") に対して、提案法による雑音低減を試みた。男声(「黒いドレスの...」)のスペクトル改善の様子を例を図9に示す。

提案法と先に開発したSPACとSPADとの雑音低減効果を定量的に評価するため、LPCケプストラム距離を計算した。LPCケプストラム距離(CD)は、雑音の重畳していない音声と各処理後の音声のLPCケプストラム定数を c_n , c'_n として、次式で定義する。

$$CD = D_b \sqrt{2 \sum_{n=1}^q (c_n - c'_n)^2} \tag{18}$$

ここで、 $D_b = 10/\ln 10$ である。また、SPACおよ

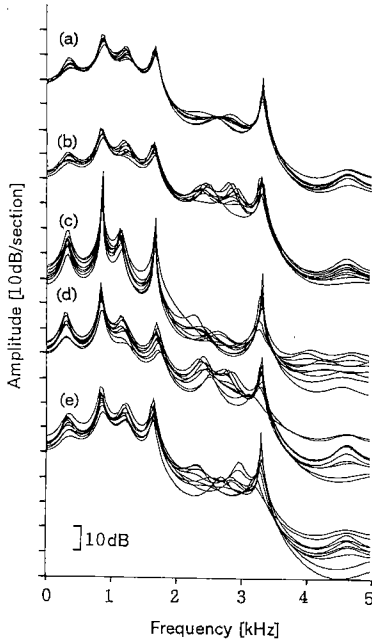


図7 女声/a/のスペクトル包絡
Fig.7 Spectral envelopes of /a/ (female).
(a) Noiseless speech. (b) Degraded speech (SNR = 15 dB). (c) Enhanced speech by SPAC. (d) Enhanced speech by SPAD. (e) Enhanced speech by proposed method.

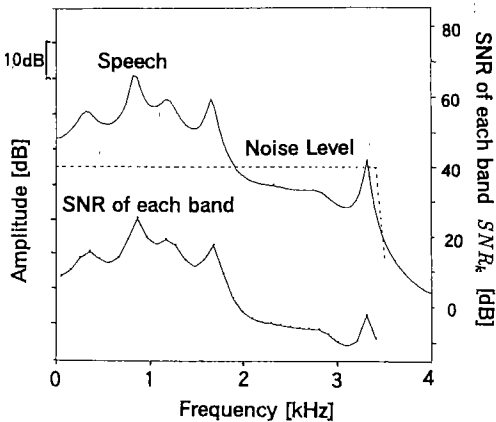


図8 SN比5 dBの音声のスペクトルと各帯域信号のSN比
Fig.8 Log spectrum of speech and the SNR of each band.

び SPAD いずれの場合にも処理後に再び帯域制限 (3.4 kHz) して、処理によって生じた高い周波数領域のひずみを取り除いて距離を計算した。

長母音についての LPC ケプストラム距離の改善結

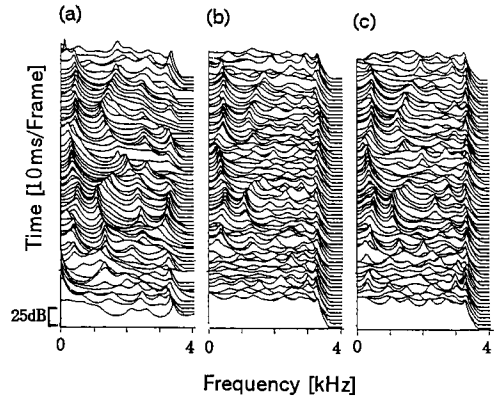


図9 連続音声のスペクトル
Fig.9 Spectral envelopes of continuative speech.
(a) Noiseless speech. (b) Degraded speech (10 dB). (c) Enhanced speech by proposed method.

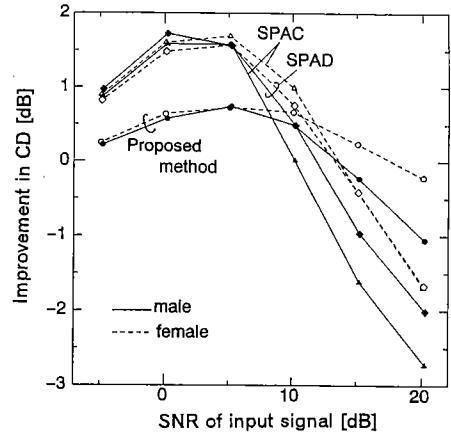


図10 LPC ケプストラム距離の改善度 (母音)
Fig.10 Improvement in LPC cepstrum distance (vowel).

果を図10に、連続音声についての改善結果を図11に示す。これを見ると、長母音、連続音声いずれの場合にも SN比 15 dB 以上の高 SN比においては、SPAC や SPAD よりも提案法の方が改善されている。これは、高 SN比音声に対しては、SPAC や SPAD の処理によってそれぞれ 2 乗ひずみや高調波ひずみの影響が大きくなるのに対し、提案法ではそうしたひずみの影響が少ないことを意味する。しかしながら、提案法による改善度の大きさは全体的に SPAC や SPAD に比べて小さくなっている。これは図6、図7で見たように、スペクトル包絡の谷の部分の改善度が小さいためである。実際に処理音を聞いてみても、高い周波数領

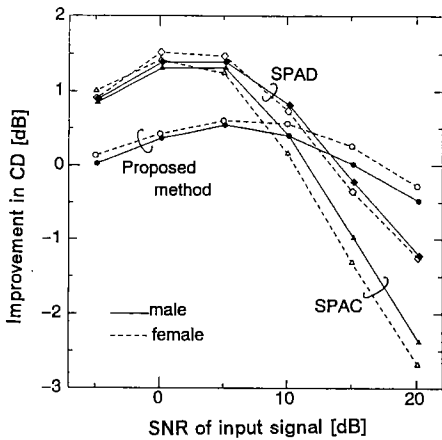


図 11 LPC ケプストラム距離の改善度 (連続音声)
Fig.11 Improvement in LPC cepstrum distance (continuative speech).

域の雑音が残ってしまっている。

提案法における改善の特徴を見ると、女声の方が男声より良くなっている。これは、強調すべき高調波の数が少ないためと考えられる。また、連続音声に対する効果よりも長母音の方が改善度が大きくなっている。これは、時間変化が少ない長母音の方が相関関数の計算による平滑化の影響が少なかったことと、連続音声においては調波構造の崩れる無声音などで効果が低下したことによるものと思われる。従って、前向き後向き差分関数を利用する提案法は、時間変化のゆっくりした音声の方が効果があると考えられる。

4.3 試聴実験

提案法による雑音低減効果を調べるために、対比較による試聴実験を行った。音声資料は 4.2 で使用した男女各 3 名の発声した連続音声を用い、未処理音声と提案法、SPAC、SPAD の 3 種類の手法で処理したものを比較した。被験者は 5 名、繰返し数は 4 回、SN 比は 5 dB、10 dB、15 dB とした。また、それぞれ実験前に原音を 2 回ずつ呈示した。

まず、提案法によるひずみの改善を調べるために評価基準を「どちらが原音に近いか」として対比較実験を行った。実験結果からプレファレンススコアを計算し、サーストンの心理尺度 (ケース V) に変換したものを図 12 に示す。サーストンの心理尺度では、数直線の右側ほど好ましいと判断したことになる。この図をみると、提案法は SPAC や SPAD よりもひずみが小さく、極めて原音に近いという結果が得られたことがわかる。

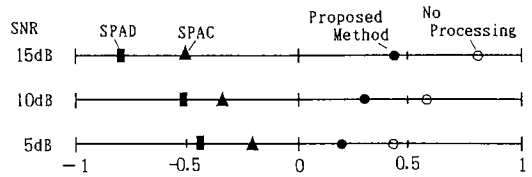


図 12 試聴実験の結果 (評価基準:「どちらが原音に近いか」)
Fig.12 Results of hearing test. Criterion is "which is close to original speech".

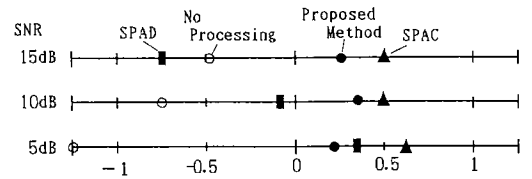


図 13 試聴実験の結果 (評価基準:「どちらが聞き取りやすいか」)
Fig.13 Results of hearing test. Criterion is "which is prefer to hearing".

次に、評価基準を「雑音低減を考慮して、どちらが聞き取りやすいか」として対比較試験を行った。その結果をサーストンの心理尺度に変換したものを図 13 に示す。この結果から、提案法により雑音劣化音声の聞き取りやすさを向上させることができたことがわかる。特に 15~10 dB において、提案法は SPAC の次に聞き取りやすいと判断された。これは提案法と SPAC を比較した場合に、SPAC ではひずみはあるが雑音が少なくて聞き取りやすく、提案法では雑音低減効果は劣っているが明瞭で聞き取りやすいと判断したためである。

以上の二つの試聴実験から、提案法によって雑音劣化音声原音に近いままで雑音低減できることが示された。

5. むすび

本論文では、単一正弦波を強調する前向き後向き差分関数とフィルタバンクを組み合わせることによって新しい音声強調方式を提案した。まず前向き後向き差分関数が 100 Hz 程度に帯域制限された信号の SN 比改善に対しても有効であることを示した。そして、フィルタバンクと前向き後向き差分関数を組み合わせた提案法の雑音低減特性を調べた。その結果、帯域幅 100 Hz のフィルタバンクを利用した本方式によって、疑似音声に対しても 6~7 dB 程度の SN 比改善ができ

ることが計算機シミュレーションによって示された。また実際の音声についても、提案法によってスペクトル包絡のピーク強調やLPCケプストラム距離を改善できることが確かめられた。一方の欠点としては、比較的低音声成分の少ない高い周波数領域では局所的にSN比が低くなり、改善効果が小さくなることがわかった。

試聴実験の結果からは、SPACやSPADで問題となっていたひずみを小さくし、原音に近い音声を復元できることが示された。また、聞き取りやすさも向上させることができたが、SPACにやや劣る結果となった。しかし、SPACではひずみによって原音とは異なる音になるという結果も得られており、総合的には提案法の方が品質を向上できたと言える。

本手法のSN比改善度は6~7dBである。この数字は、SPACやSPADの単一正弦波に対するSN比改善度6~20dBに劣っている[18]。しかしながら、従来の手法[2],[5],[10],[11]におけるSN比改善度3~6dBに劣るものではなく、本手法は有効であると言える。また、前向き後向き差分関数を利用するメリットとしては、帯域内のランダム雑音に埋もれた周期信号を強調するため、音声の厳密な基本周波数の抽出や雑音の性質などの情報を必要としないことがある。こうした特性を活かすためにも、ひずみを少なく、かつ雑音低減効果を大きくするための改良が今後の課題として残される。

今回は、フィルタの帯域幅を100Hzに固定して音声に対する効果を検討してきた。しかし、前向き後向き差分関数は帯域幅が広いほどSN比改善効果が大きいことから、できるだけ帯域幅を広くすることが望ましい。そのための手法としては、基本周波数の推定を行い、この情報をフィルタの帯域幅の設定に利用して性能を向上させることが考えられる。こうしたフィルタの設定についての検討や明瞭度試験などのより定量的な評価を行うことが今後の課題として挙げられる。

文 献

- [1] J.S. Lim, ed., "Speech Enhancement," Prentice-Hall, Englewood Cliffs, 1983.
- [2] M.R. Sambur, "Adaptive noise canceling for speech signals," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-26, no.5, pp.419-423, Oct. 1978.
- [3] W.A. Harrison, J.S. Lim, and E. Singer, "A new application of adaptive noise cancellation," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-34, no.1, pp.21-27, Feb. 1986.
- [4] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-27, no.2, pp.113-120, April 1979.
- [5] L. Hoy, B. Burns, D. Soldan, and R. Yarlagadda, "Noise suppression methods on speech applications," Proc. IEEE Int. Conf. Acoust. Speech and Signal Process., 24.5, pp.1133-1136, 1983.
- [6] 長瀬裕実, 小林 勉, "雑音を加わった音声の分析合成における合成音声の品質改善," 信学論 (A), vol.J64-A, no.9, pp.723-730, Sept. 1981.
- [7] J.S. Lim, A.V. Oppenheim, and L.D. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-26, no.4, pp.354-358, Aug. 1978.
- [8] J.S. Lim and A.V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-26, no.3, pp.197-210, June 1978.
- [9] J.H.L. Hansen and M.A. Clements, "Constrained iterative speech enhancement with application to speech recognition," IEEE Trans on signal processing, vol.SP-39, no.4, pp.795-805, April 1991.
- [10] 小畑秀文, "最尤推定法に基づく音声波形の復元," 音響誌, vol.50, no.10, pp.809-815, Oct. 1994.
- [11] 菅野俊夫, 小林隆夫, 今井 聖, "雑音劣化音声の一般化ケプストラムモデルにおける事前情報の利用," 信学論 (A), vol.J77-A, no.7, pp.945-953, July 1994.
- [12] 高杉敏男, 鈴木誠史, 田中良二, "自己相関関数を利用した音声処理方式 (SPAC) の機能と基本特性," 信学論 (A), vol.J62-A, no.3, pp.175-182, March 1979.
- [13] 國枝伸行, 島村徹也, 鈴木誠史, 八嶋弘幸, "差分関数を利用した音声処理方式-SPAD," 信学論 (A), vol.J75-A, no.11, pp.1769-1772, Nov. 1992.
- [14] Y. Epharaim, "A bayesian estimation approach for speech enhancement using hidden Markov models," IEEE Trans. on Signal Processing, vol.40, no.4, pp.725-735, June 1992.
- [15] R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. Acoust., Speech & Signal Processing, vol.ASSP-28, no.2, pp.137-145, April 1980.
- [16] 市田多加雄, 坂井利之, "フィルタ群を用いた雑音除去," 信学技報, EA80-63, Jan. 1981.
- [17] 松本 弘, "実環境における音声認識の現状と課題," 平5春音講論, 1-4-10, pp.515-518, March 1993.
- [18] 國枝伸行, 島村徹也, 鈴木誠史, "前向き後向き差分関数による単一正弦波信号の強調," 信学論 (A), vol.J77-A, no.9, pp.1307-1311, Sept. 1994.

(平成7年4月27日受付, 9月11日再受付)



國枝 伸行 (学生員)

平4 埼玉大・工・情報卒。平6 同大大学院博士前期課程了。現在、同大大学院博士後期課程に在学中。音声情報処理に関する研究、特に音声強調や雑音に埋もれた信号の処理に関する研究を進めている。日本音響学会、IEEE 各会員。



島村 徹也 (正員)

昭61 慶大・理工・電気卒。平3 同大大学院博士課程了。工博。同年埼玉大・工・情報助手。スペクトル解析および適応信号処理に関する研究に従事。計測自動制御学会、日本音響学会、IEEE 各会員。



鈴木 誠史 (正員)

昭30 東工大・工・電気卒。同年郵政省電波研究所(現、通信総合研究所)入所。昭42 同通信機器部音声研究室長、昭57 同情報処理部長、昭60 同総合通信部長、昭63 通信総合研究所長。平1 埼玉大・工・教授、現在に至る。昭41~42、MIT 客員研究員。

この間、音声合成、音声認識、通信品質の改善、音声伝送方式などの音声情報処理、通信システムの研究に従事。工博。アメリカ音響学会、日本音響学会、テレビジョン学会、情報処理学会、IEEE 各会員。