コンテンツ制作における収録音編集のための音声強調*

中村一文 (法政大学情報科学部), 伊藤克亘 (法政大学情報科学部)

1　まえがき

ドラマ・アニメ・ラジオ等のコンテンツ制作において収録音に対する編集作業は不可欠である．コンテンツ制作を新規・個人で始める人にとって，収録環境を整えることは至難である．この問題を解決する方法として 1 入力で行える音声強調システムが必要であると考えられる．

1 入力音声強調法に変調スペクトルの特性を用いたランニングスペクトルフィルタ (RSF)[1] がある．変調スペクトルはランニングスペクトルをフーリエ変換したもので,「パワーの時間変化に対するスペクトル」が変調スペクトルである．白色雑音のような定常的な雑音はランニングスペクトルでの時間変化が小さいことから変調スペクトルでは低域に成分が集中している．また音声は非定常的な音であり，スペクトルの時間変化が大きいので変調スペクトルが広く分布している．特に音声認識において重要な変調周波数は 1〜16Hz であり [2]，また知覚実験により 17Hz 以下のみを用いても音声の明瞭性にはほとんど影響がない [3] とされている．RSF はこの特性を利用し，各周波数帯にバンドパスもしくはハイパスをかけることにより変調スペクトルにおける音声の重要な成分を強調する手法である．しかし文献 [4] では 0Hz 付近，17Hz 以上の成分にも話者情報が含まれていることを示唆している．また 1〜16Hz にも雑音成分が含まれていることから，音声の明瞭度と雑音除去の両立には音声/雑音領域を決定することが重要である．従来の音声/雑音領域を決定する 1 入力音声強調法に [5] がある．[5] ではランニングスペクトル上で領域を判別し，領域毎にスペクトルサブトラクション (SS) を行っている．しかし，領域の誤判別により雑音の引き残しや音声の引き過ぎが生じる．これはランニングスペクトル上では音声と雑音の違いを明確にすることが困難であるからと考えられる．そこで本論文では変調スペクトル上の音声/雑音領域判別法を提案する．

2　提案方法

図 1 に提案手法の構成を示す．変調スペクトルにおける雑音成分は全体的に分布しているが 1 Ｈｚ以上ではその強さはある程度一定である．また音声成分は全体的に分布しているが高域になるにつれ緩やかに下降している．この差を利用して雑音と音声の判別を行う．具体的には 2 Ｈｚ以上におけるスペクトルの並びから回帰直線を求めて，その傾きから雑音と音声の判別を行う．図 2 に 2 Ｈｚ以上における変調スペクトルの並びを示す．

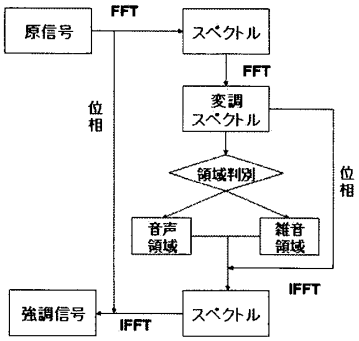

図 1. 提案方法の構成

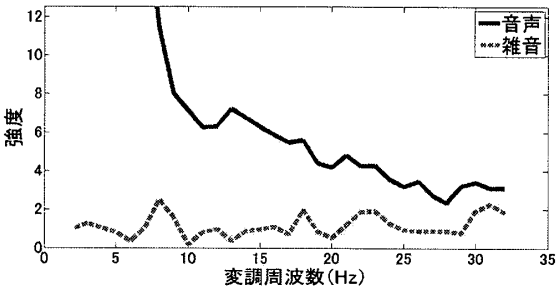

図 2. 2 Ｈｚ以上の変調スペクトルの並び

先に述べた通り，雑音成分の強さはある程度一定であるので回帰直線の傾きが小さい可能性が高い．また音声成分は高域になるにつれ下降しているので傾きが雑音成分の傾きより大きい．しきい値を設定して回帰直線の傾きがしきい値以下なら雑音，それより上の場合は音声と判断する．音声だと判断されれば成分をそのまま残し，雑音だと判断されれば成分を除去する．領域判別のしきい値には以下の式を用いた．

$$Th(i) = \frac{1}{M}\sum_{r=1}^{M} a(r,i) - d$$

$$\text{領域 (r,i)} = \begin{cases} \text{音声領域} & if \quad a(r,i) < Th(i) \\ \text{雑音領域} & if \quad a(r,i) \geq Th(i) \end{cases}$$

ここで i はランニングスペクトルの各周波数帯，r は変調スペクトルでのフレーム番号，a(r,i) は回帰直線の傾きである．M は変調スペクトルでの全フレーム長である．つまり Th(i) は各周波数帯での傾きの平均である．d は音声/雑音領域の分類を正しくおこなうためのパラメータである．平均だけでは雑音の引き残しが発生する可能性が高いため，d によって調節する．

* Single channel Speech Enhancement for Recoreded Audio Contents by KazufumiNakamura. (CIS,HoseiUniversity) et. al.

3 評価実験

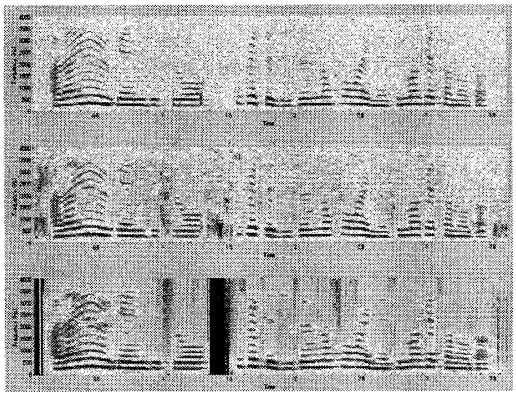

図 3. スペクトログラム (上:処理前, 中:RSF, 下:提案手法)

実験は 16kHz,16bit のサンプリングで録音した音声を用い，男女各 2 名づつに「あらゆる現実をすべて自分のほうへねじ曲げたのだ」という文章を読み上げてもらった．録音した音声に白色雑音・バブルノイズ・自動車の走行音・雨音をそれぞれ付加した．雑音の SNR はそれぞれ 10dB である．ランニングスペクトルは，短時間スペクトルを 512 点 FFT で求め，フレームシフト量を 256 点として計算した．変調スペクトルはランニングスペクトルの時間軌跡に対しフレーム幅を 8 点フレームシフト量を 4 点として計算した．パラメータ d の値は 0.04 とした．図 3 に白色雑音付き音声に対する RSF と提案手法処理後のスペクトログラムを示す．また表 1, 2 に RSF と [5] の SS との比較をした Segmental SNR の改善度と MOS テストの結果をそれぞれ示す．MOS テストは最初にクリーン音声と雑音を付与した音声を聞いてもらい，その後に提案法及び従来法で処理した音声をランダムで聞いてもらう．聞いてもらった後に 5 段階で評価をつけてもらった．被験者は 5 名とした．

表 1. Segmental SNR の改善度

SNR	Method	White	Babble	Rain	Car
10	Proposed	10.37	4.63	5.48	5.50
	SS	3.58	2.78	5.46	5.75
	RSF	3.56	3.05	4.42	5.69

表 2. MOS テスト

SNR	Method	White	Babble	Rain	Car
10	Proposed	3.1	2.7	3.0	2.1
	SS	2.5	2.6	2.8	1.8
	RSF	3.6	3.7	3.6	3.1

図 3 より提案手法は RSF や SS と比べ雑音と思われる部分が少ない．表 1 においてもスコアが大きく改善されており，提案手法が雑音除去性能に優れていることがわかる．しかし表 2 より提案手法は主観評価において結果が芳しくなかった．これは提案手法独特のミュージカルノイズが知覚されるためと考えられる．このミュージカルノイズは持続性のある電子音のようなノイズとなっている．提案手法は変調スペクトルを求めるためにランニングスペクトル上で数フレーム用いる．そのため音声/雑音領域はランニングスペクトル上で点ではなく線となる．そして領域誤判別によって雑音の引き残し，音声の引き過ぎが生じると残った成分が倍音構造を持つようなスペクトル線となり電子音のようになる．分析フレーム幅を短くすればノイズのスペクトル線が短くなり低減されるが，フレーム幅を短くしすぎると音声の非定常性が失われてしまい雑音との判別が難しくなってしまう．

しかし，図 3 において RSF 法及び提案手法のスペクトログラムを見比べると 1kHz 以下の低域において RSF 法は音声成分が欠落している部分が見受けられるのに対し，提案手法は保存できている．また高域になるほど提案手法は引き過ぎにより孤立した線スペクトルが目立つようになるのに対し，RSF 法は雑音は残っているが音声成分も残っている．この高域の音声成分の保存が主観評価の差につながったと考えられる．しかし低域では音声を保存できたことから提案手法のパラメータ d は周波数帯毎に適宜設定することで高域においても低域と同じように判別できると考えられる．実験材料を増やし提案手法のしきい値を決定することが必要となる．

4 あとがき

コンテンツ制作のための音声強調として変調スペクトルにおける音声/雑音領域判別法を提案した．提案手法では音声と雑音の判別として変調スペクトル上でのスペクトル列に対する回帰直線の傾きを利用した．今回，しきい値は手動で決定したが，しきい値を適応的に設定できれば音声を歪ませずに雑音だけ低減することも可能と考えられる．提案手法について SegmentalSNR の改善度と MOS テストによる性能評価を行った．その結果，SegmentalSNR の改善度は従来手法よりスコアが改善されたが，MOS テストは従来手法を下回った．今後は適応的にしきい値を設定する方法を検討していく．

参考文献

[1] 藤岡一馬他，“ランニングスペクトルフィルタリングを用いた音声の雑音低減法”電子情報通信学会論文誌 Vol.J88-D-II No4 (2005), pp. 695-703

[2] 金寺登他，“変調スペクトルの重要な成分のみを選択的に用いた雑音に強い音声認識”電子情報通信学会論文誌 Vol.J84-D-II No7 (2001), pp. 1261-1269

[3] 早坂昇他，“ランニングスペクトルフィルタを用いた雑音にロバストな音声認識”電子情報通信技術研究報告 CAS2003-6,VLD2003-16,DSP2003-36 (2003-06), pp. 31-36

[4] 金寺登他，“音声の変調スペクトル中に含まれる情報の調査-音声認識情報と話者識別情報との比較-”電子情報通信技術研究報告 SP2000-34 (2000), pp. 15-22

[5] 野村行弘他，“雑音量に依存しない音声/雑音領域判別法を利用した音声強調の改良”日本音響学会誌 62 巻 1 号 (2006), pp. 12-22