〈一般研究課題〉　実用的な音環境における音声認識のための雑音除去に関する研究

助 成 研 究 者　名城大学　佐川　雄二

実用的な音環境における音声認識のための雑音除去に関する研究

佐川　雄二，旭　健作，杉江　昇
（名城大学）

1. はじめに

近年，音声認識技術の向上によりカーナビゲーションシステムなどの一部では音声入力による車内機器の操作が可能となっている．しかし，現状では，通常の室内のような静かな環境におけるほど高い音声認識率の実現には至ってない．

雑音に対処するため，マイクロホンアレー[1]やスペクトルサブトラクション（SS）[2]に代表される雑音信号成分除去法に関する研究が盛んに行われている．これらの雑音抑圧技術は，自動車内のような騒音下での音声による機器の操作を，より確実に行うために助けになると考えられる．複数のマイクロホンを用いるマイクロホンアレーによる方法は，マイク全体で指向性を形成し雑音の発生源の方向の感度を下げることと等価である．しかし，この方法では，次のような問題点がある，最近の車載機器は，ますます小型化されており，多数のマイクロホンを取り付けることが構造上難しい場合が多い．この問題は単一のマイクロホンを用いる方式においては生じない．従来から盛んに研究されているSS法は，比較的定常な騒音には有効であるが，非定常性の強い騒音環境下では不十分である．また，このような環境は，雑音のパワー変動に加えてスペクトルも多様である．このため，雑音スペクトルの引きすぎにより生じる雑音の問題が残存している．

本手法では，自動車内での雑音を抑圧することを目的とし，雑音源の方向が特定できない環境下において単一のマイクロホンのみを用いる雑音抑圧法について検討を行う．周波数領域での有声音特有のスペクトル微細構造を，スペクトログラムと，その画像処理結果から抽出し，有声音区間を検出する．本手法では，対象を有声音区間に限定した．それは，高騒音環境下では無声音は雑音に埋もれてしまうため，無声音を含めた区間検出方法では，誤検出が避けられず，有声音区間を確実に検出する方法が実用的であると考えたためである．

2. スペクトログラム

スペクトログラムとは，高速フーリエ変換（以下，FFT）により分析した周波数情報を，縦軸に周波数，横軸に時間をとり，それぞれの点におけるパワーの強弱を濃淡によりプロットしたものである．その例を図１に示す．スペクトログラムは音声解析によく用いられている．

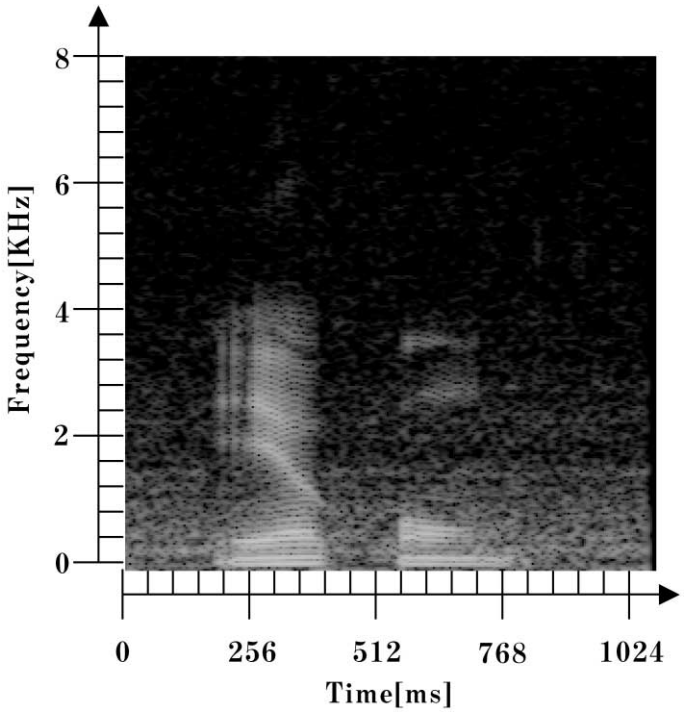


図1　雑音重畳音声のスペクトログラム

図１は，有声音の音声データによるスペクトログラムであるが，ここから，有声音特有の周波数構造である，調波構造がわかる．調波構造とは，一定の周波数間隔で幾つもの成分が存在する構造である．

3. 処理手法

本手法では，単一のマイクロホンにより集音された音声をA/D変換を行いPCMのディジタルデータにする．そして，FFTを施し，スペクトログラムを作成して，解析に必要な時間分のデータを一時的に蓄積する．蓄積されたスペクトログラムに対して細線化を行い，パターンマッチングにより有声音特有の調波構造の含まれる区間を抽出し音声区間とする．検出された音声区間に対しさらに詳細な調波構造の探索を行うために細線化時に用いるパワーの閾値を下げて再度パターンマッチングを行う．このようにして，雑音に埋もれてしまいそうな調波構造の抽出を行い，音声の品質を向上させている．抽出された，音声の周波数成分のみを通過するようなバンドパスフィルタを構成して，雑音の中から，音声を分離する．

3.1. 画像処理

3.1.1. 細線化

スペクトログラムに対して周波数構造を抽出するために細線化処理を行う．通常の画像に適用される細線化アルゴリズムは，処理に時間がかかり，２値化処理が必要となる．ここでは，スペクトログラムの特徴を利用して細線化と２値化を一度に行う方法を提案する．

図２に，ある時刻でのスペクトログラムの周波数方向への振幅の一部分を示す．周波数成分で特徴的な点は振幅の極大点であるため，周波数方向に２次微分（差分）を行い，極大点を抽出する．

$$(b(f+2)-b(f+1))-(b(f+1)-b(f))<0$$

ただし，$b(f)$はスペクトルの振幅，fは周波数．

そして，あらかじめ定めた閾値より大きい場合を音声と判断して，その点を白画素とし，その他の点を黒画素とすることにより２値化を行う．図３に，図１の極大値細線化結果を示す．

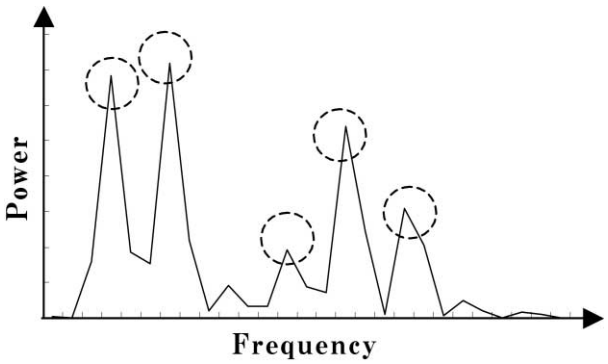


図2　スペクトルの一部

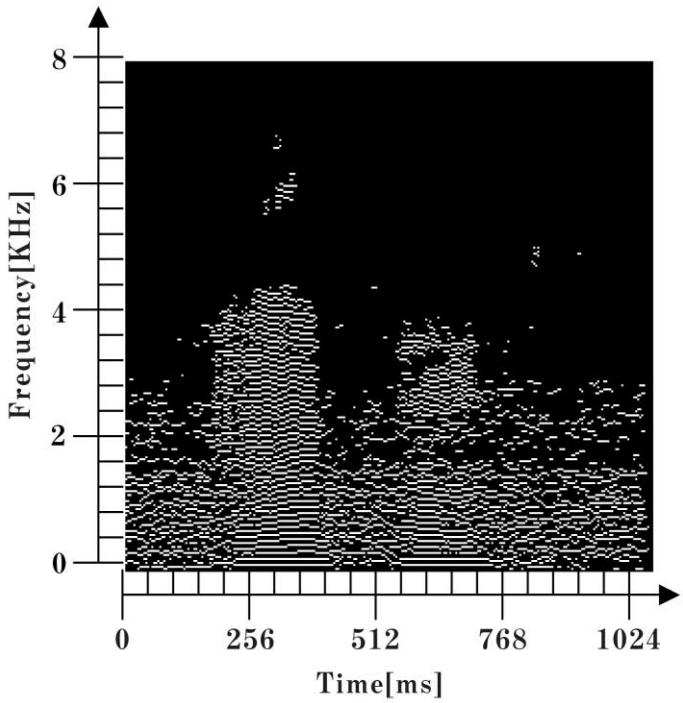


図3　細線化スペクトログラムの例

3.1.2. 画素の追跡，ラベリング

細線化された画像に対して，画素の時間方向の連続性を調べるために画素の追跡，ラベリング処理を行う．雑音の周波数成分は，同一周波数での持続時間が比較的短く周波数成分の構成が大きく変動する場合が多いのに対して，音声は比較的長時間，同一の周波数成分が継続して観察される．このことから，同一ラベルを持つ画素が一定個数以上あれば，音声の周波数成分候補であるとして残し，それ以外の画素を削除する．

次に同一音源からの周波数成分は，ほぼ同じ時刻に立ち上がることから，立ち上がり時刻をラベルにして，それぞれの周波数成分に対してラベルの付け直しを行う．こうすることで，その周波数成分の立ち上がり時刻がわかるようになり，ある時刻付近で立ち上がる成分を，同一音源からと判定することが出来る．

3.2. フィルタリング

以上の判定により残された周波数成分のみを通過するバンドパスフィルタを構成し雑音の重畳した音声に対してフィルタリングを行い音声のみを分離，抽出する．今回は512次のFIR型フィルターを使用した．

4. 実験方法

4.1. 実験条件

従来から用いられているSS法と今回の提案手法との比較実験を行った．

音声データは20代男性話者２名が発声した音声を用いた．また，雑音として，白色雑音，電子協雑音データベースの走行車（2000ccクラス）から２種類（加速時，定常走行時の他車とすれ違い），筆者が録音した走行軽自動車内音の計4種類の雑音を用いた．雑音の重畳条件としては，音声$s(t)$全体のRMSパワーS_{pow}，雑音$v(t)$全体のRMSパワーV_{pow}求め，式（１）のようにそれぞれの振幅比が所望のSNR[dB]となるように雑音$v(t)$の振幅を伸張させ音声$s(t)$に加算することにより生成した．

$$y(t) = s(t) + \frac{S_{pow}}{10^{SNR/20} V_{pow}} v(t) \qquad (1)$$

それぞれの音声に対してSNRが10, 5, 0dBとなるよう雑音を重畳して実験を行った．

音声波形から分析に用いるスペクトログラムの作成には表１の条件を設定した．

標本化周波数	16,000[Hz]
FFT 次数	512
分析窓	Hamming 窓
分析フレーム長	32[ms]
フレーム周期	4[ms]

表1　スペクトログラムの分析条件

画像処理時の画素の追跡を行い短時間の成分の除去を行ったかが，今回は，48ms以下を除去した．

SS法の分析条件を，表２に示した．

分析フレーム長	16[ms]
フレーム周期	2[ms]
学習係数	0.95

表2　SS法の分析条件

4.2. 正規化相互相関

今回，雑音除去性能の検討には，式（２）に示す正規化相互相関を使用した．

$$R = \frac{\sum_{t}^{T_{\max}} (f(t) - \bar{f})(g(t) - \bar{g})}{\sqrt{\sum_{t}^{T_{\max}} (f(t) - \bar{f})^2} \sqrt{\sum_{t}^{T_{\max}} (g(t) - \bar{g})^2}} \qquad (2)$$

$$= \frac{\sigma_{fg}^2}{\sqrt{\sigma_f^2 \sigma_g^2}}$$

ただし，

$f(t)$：信号１，　　$\bar{f}$：信号１の平均値

$g(t)$：信号２，　　$\bar{g}$：信号２の平均値

σ_f^2：信号１の分散，　　σ_g^2：信号２の分散

σ_{fg}^2：信号１，２の共分散

今回は，式（２）で得られた値に100を乗算して百分率で表示した．

5. 実験結果

図４，５，６に各音声に対して，SNRを0, 5, 10[dB]となるように雑音を重畳しSS法と提案手法適用後の音声信号と，雑音重畳前の音声信号との相関値を示す．

それぞれの図において雑音の種類は以下のとおりである．

・A,B=電子協雑音データベースの走行車

（2000ccクラス）から２種類

（A=加速時，B=定常走行時，他車とすれ違い）

・C=筆者が録音した走行軽自動車内音

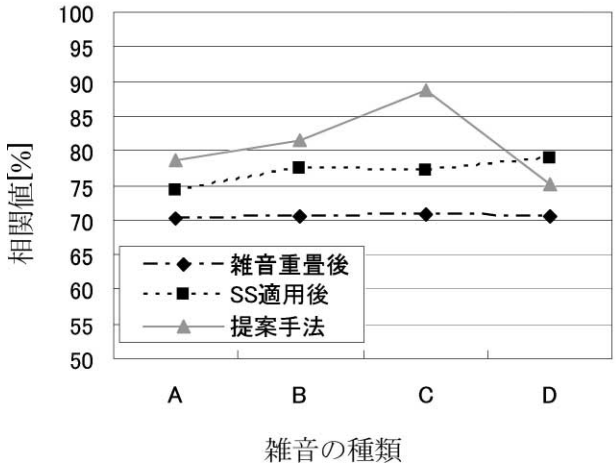
・D=白色雑音

SNRが0[dB]においては，提案手法がSS法よりも高い相関値を示していることがわかる．

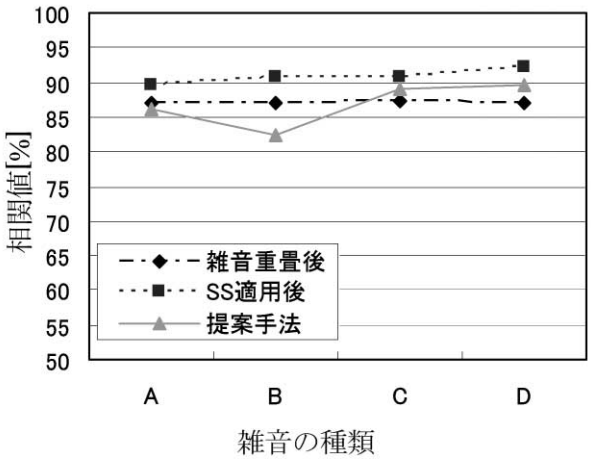
しかし，SNRが大きくなるにしたがって本手法よりもSS法の相関値が大きくなっていることが分かる．

SNRが0, 5[dB]において，雑音が白色雑音の場合，他の種類の雑音の場合に比べ相関値が低下しているが，これは，白色雑音では，音声の周波数構造と類似の構造を誤検出するため，雑音が残存したためと考えられる．

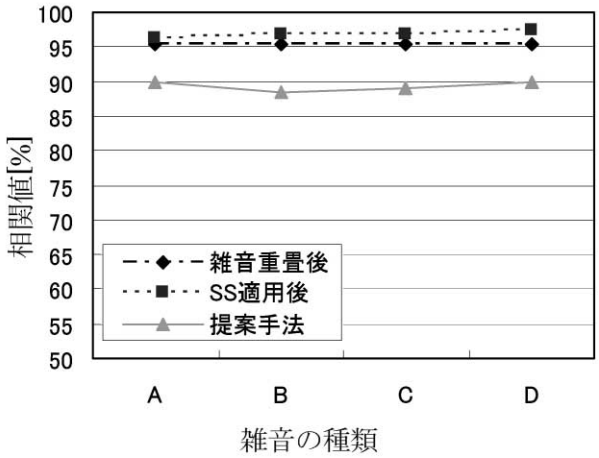
以上の結果から本手法は低SNR時に有効であることが分かった．



(a)　エアコン1　(SNR:0[dB])

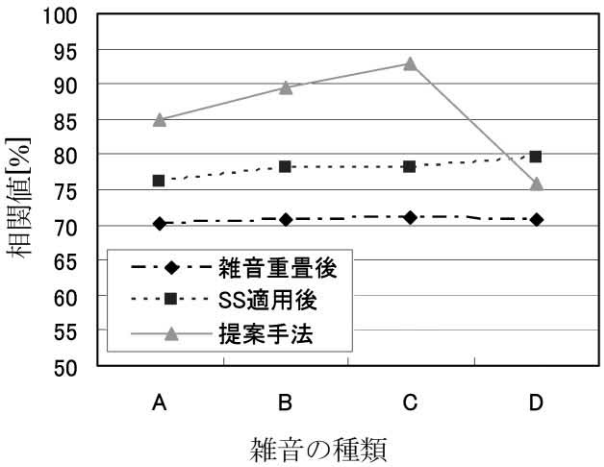


(b)　エアコン1　(SNR:5[dB])

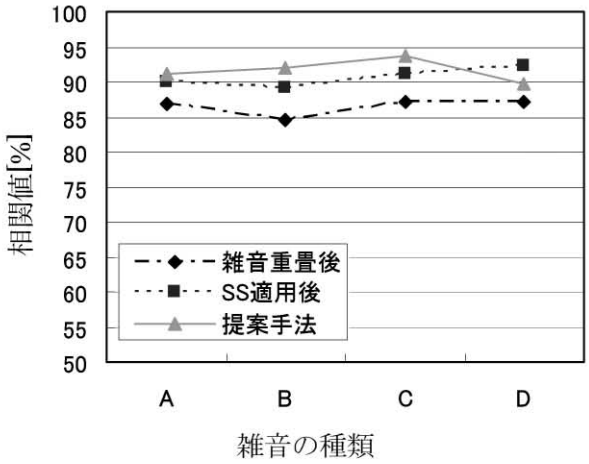


(c)　エアコン1　(SNR:10[dB])

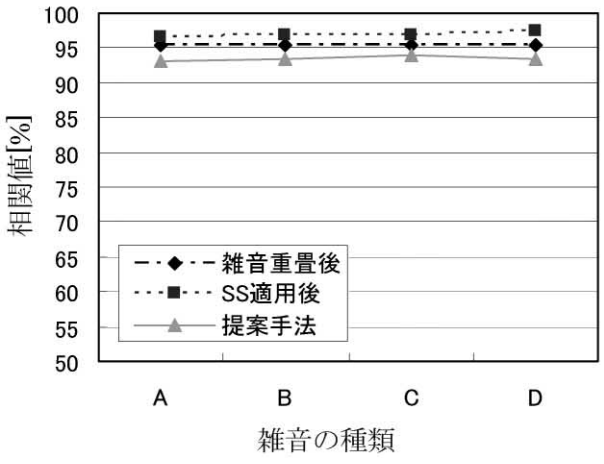
図4　男性話者1による“エアコン”音声の場合



(a)　エアコン２　(SNR: 0[dB])

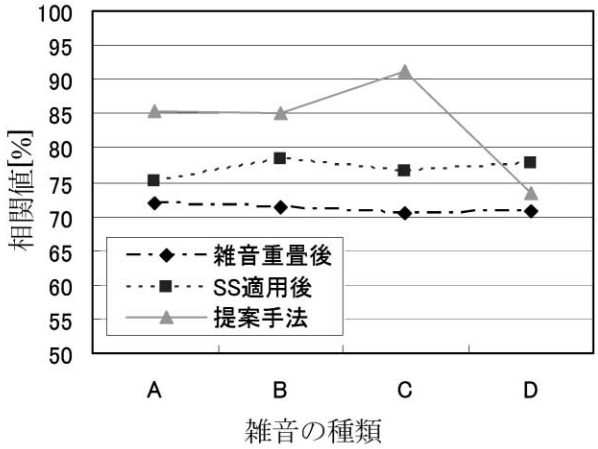


(b)　エアコン２　(SNR:5[dB])

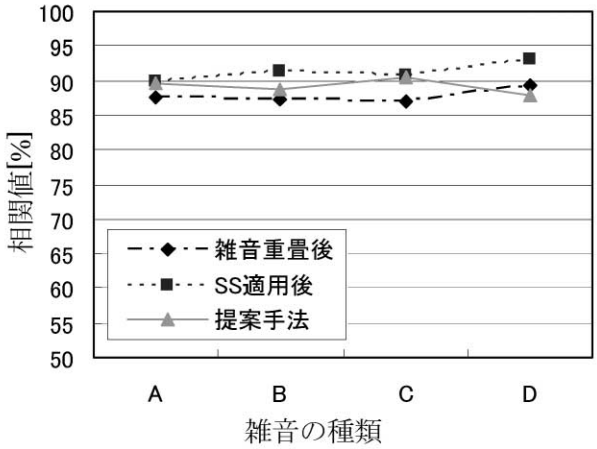


(c)　エアコン２　(SNR:10[dB])

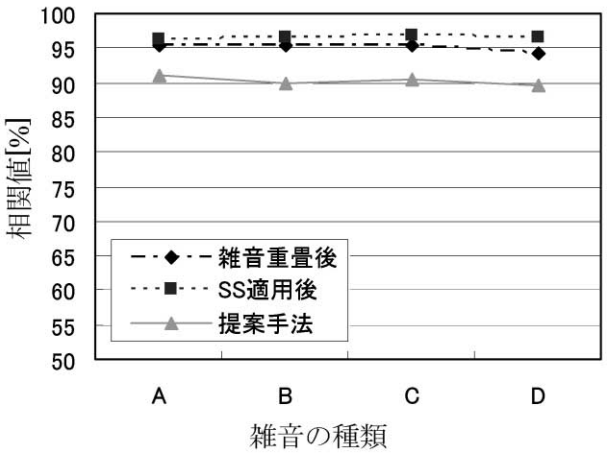
図５　男性話者２による“エアコン”音声の場合



(a)　問われるところだ　(SNR:0[dB])



(b)　問われるところだ　(SNR:5[dB])



(c)　問われるところだ　(SNR:10[dB])

図6　男性話者2による“問われるところだ”音声の場合

6. まとめ

スペクトログラムに対して画像処理を適応することにより音声信号の特徴的な周波数構造の抽出を行い，雑音が重畳した音声から，雑音を抑圧し音声信号を抽出する方法を提案した．

本手法は，低SNR時にSS法に比べ元音声との相関が高く，車内雑音の抑圧に有効であることを確認した．

今後の課題として，さらに音声成分の抽出を行う画像処理法の改良を行い，雑音抑圧能力の向上を図りたいと考えている．また，音声認識を用いた認識率の評価を行いたいと考えている．

謝辞

本研究の一部は，日比科学技術研究助成金の援助により行われた．

参考文献

[1] J.L. Flanagan, J.D. Johnston, R. Zahn, and G.W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," J. Acoust. Soc. Am., vol.78, no.5, pp.1508-1518, Nov. 1985.

[2] 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.

[3] 中川聖一，“音声認識研究の動向”信学論（D-Ⅱ），vol.J83-D-Ⅱ，No.2，pp.433-457，Feb. 2000.

[4] 蜷川和弘，梅山貴士，鈴木賢治，杉江昇，“画像処理を用いた周波数領域での混合母音音声の分離，”電学論C, Vol.121, No.12, pp.1866-1874, Dec., 2001.