非定常雑音・時変残響環境下でのパワースペクトログラム領域セミブラインド音声強調

SEMI-BLIND SPECTROGRAM ENHANCEMENT OF SPEECH CORRUPTED BY NON-STATIONARY NOISE IN TIME-VARIANT REVERBERANT ENVIRONMENT

池澤 浩気 [1]　北条 伸克 [2]　エスピ ミケル [2]　亀岡 弘和 [2,3]　嵯峨山 茂樹 [2]
Hiroki Ikezawa　Nobukatsu Hojo　Miquel Espi　Hirokazu Kameoka　Shigeki Sagayama
東京大学工学部 [1] Faculty of Engineering, The University of Tokyo
東京大学大学院情報理工系研究科 [2] Graduate School of Information Science and Technology, The University of Tokyo
NTT コミュニケーション科学基礎研究所 [3] NTT Communication Science Laboratories

1　導入

雑音・残響環境下で音声信号を強調する問題は，原音声信号に未知の雑音と残響が重畳された観測信号から原信号を推定する不良設定の逆問題である．古典的アプローチ（[1][2] 等）では雑音や残響に定常性・時不変性等の強い仮定を置き当該逆問題を定式化することが多かった．これに対し本稿では，音声スペクトログラムに見られる「低ランク構造」(後述) に着目し，音声らしさを最大化するような問題として当該逆問題を定式化し，非定常雑音・時変残響環境下での音声強調問題の解決を目指す．

2　提案手法

時不変な残響環境下では観測信号の複素スペクトログラムは近似的に $y(\omega,t)=\sum_\tau s(\omega,t-\tau)h(\omega,\tau)+n(\omega,t)$ と表現できる [3]．ただし，$s(\omega,t)$, $n(\omega,t)$, $h(\omega,t)$ はそれぞれ原音声信号，加法性雑音，室内伝達関数の時間周波数成分を表し，ω, t は周波数と時刻の添数である．しかし，実環境では音源の移動等により室内伝達関数は時変となる．特に，室内伝達関数の時間周波数成分の位相は，音源-観測点間の音響経路の小さな変化に伴って鋭敏に変化することが予想される．そこで本稿では，音源の移動を考慮した，室内伝達関数の振幅は時不変で位相 $\theta(\omega,t,\tau)$ のみが時変である系 $y(\omega,t)=\sum_\tau s(\omega,t-\tau)|h(\omega,\tau)|\mathrm{e}^{\mathrm{j}\theta(\omega,t,\tau)}+n(\omega,t)$ を仮定する．これを半時変系と呼ぶこととする．さらに原音声信号 $s(\omega,t)$ と雑音 $n(\omega,t)$ が平均が 0 の複素ガウス分布に従い，室内伝達関数の位相 $\theta(\omega,t,\tau)$ が一様分布に従うと仮定すると，モデルパラメータの対数尤度が，観測信号のパワースペクトログラムとパワースペクトログラムモデル $\sum_\tau \Phi_{ss}(\omega,t-\tau)|h(\omega,\tau)|^2+\Phi_{nn}(\omega,t)$ との板倉斎藤距離と定数項・符号を除いて等しくなる．ただし $\Phi_{ss}(\omega,t)$, $\Phi_{nn}(\omega,t)$ は原音声信号，雑音の時刻 t におけるパワースペクトル密度である．ここで，音声は幾種類かの母音や子音，特定範囲内の基本周波数等の限定的な要素で構成されるため，各時刻の音声パワースペクトルが高々 K 個の基底スペクトルの重ね合わせで表されると仮定する．これはパワースペクトログラムの行列表現が低ランクな行列で表現可能（低ランク構造）との仮定に相当する．雑音も PHS の呼出音や空調音等，低ランクであることが多いため音声同様の仮定を置く．以上の仮定より新しいパワースペクトログラムモデルを得る．

$$\sum_\tau \sum_k^K B_k^{(s)}(\omega)G_k^{(s)}(t-\tau)\,|h(\omega,\tau)|^2+\sum_l^L B_l^{(n)}(\omega)G_l^{(n)}(t)$$

$(B_k^{(s)}(\omega),\ G_k^{(s)}(t))$, $(B_l^{(n)}(\omega),\ G_l^{(n)}(t))$ は音声，雑音の基底とその励起である．また，[5] と同様，音声の基底 $B_k^{(s)}(\omega)$ のみ事前学習 [4] により得られているものとする．このモデルにおける先の擬距離最小化問題について，板倉斎藤距離，I ダイバージェンス，2 乗距離のそれぞれの場合に，[3][4] をヒントにした補助関数法による反復最適化アルゴリズムを導出した．

3　動作実験

残響室内（残響時間 1.3s）での移動音声信号（…☆）に PHS の呼出音と背景雑音を断続的に加えたものを観測信号（図 1 左辺）として提案手法の動作実験を行った．事前学習には観測信号と同一話者の音声データベースを用いた．音声，雑音の基底の数はそれぞれ 18, 6 とし，擬距離は I ダイバージェンスを用いた．

原音声信号と推定信号を図 2 に示す．人間の聴覚尺度に近く自動音声認識にも用いられる MFCC について，原音声信号との距離が☆よりも小さくなった．

4　まとめ

ある非定常雑音・時変残響環境下で提案手法により MFCC 距離が改善し一定の雑音・残響除去ができた．

参考文献

[1] P. C. Loizou, *Speech Enhancement: Theory and Practice.*, 2007.
[2] P. A. Naylor, N. D. Gaubitch, *Speech Dereverberation.*, 2010.
[3] H. Kameoka, et al., Proc. ICASSP '09, 45-48, 2009.
[4] M. Nakano, et al., Proc. MLSP '10, 283-288, 2010.
[5] P. Smaragdis, et al., Proc. ICA '07, 414-421, 2007.

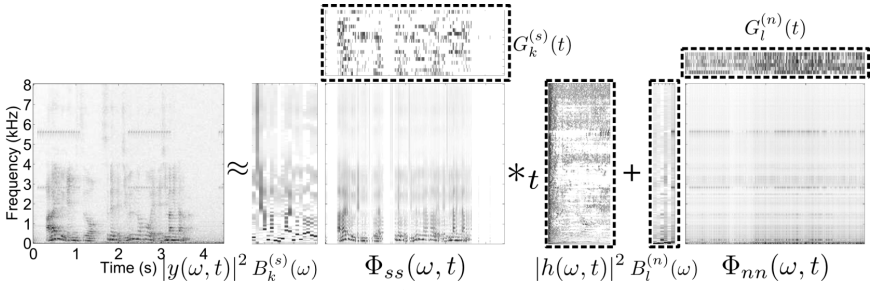

図 1: 観測信号（左辺）と対応するモデル（右辺）．$*_t$ は t についての畳み込みを表わす．点線で囲まれた箇所が推定すべきパラメータ．$B_k^{(s)}[\omega]$ は事前学習により取得．

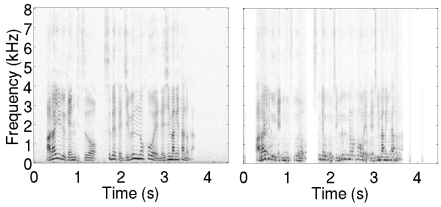

図 2: 原音声信号（左）と提案手法により推定された原音声信号（右）