非音響ノイズを用いた話者照合における耐雑音性の改善*

☆中西亮介，塩田さやか，貴家仁志 (首都大)

1　はじめに

実環境における話者照合システムの構築を考える際，雑音の影響を考える必要がある．雑音の種類やSNRが事前に分かっていれば効率的に耐雑音性を向上することが可能であるが，事前にそれらを把握することは困難である．そこで，本研究では非音響ノイズが持つ話者性と耐雑音性に着目した．非音響ノイズとは話者が話す際に無意識に発生させてしまう呼気によって発声するノイズを指す．本稿では，非音響ノイズが含まれた話者モデルを用いて話者照合実験を行った結果，環境雑音の種類やSNRに依存することなく照合率が改善したことを報告する．

2　非音響ノイズの話者性と耐雑音性

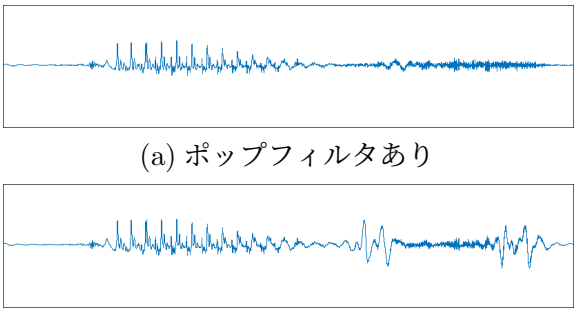
2.1　非音響ノイズ

マイクに向かって話しかける際，息継ぎやリップノイズといった話者自身が無意識的に発生させてしまうノイズが含まれることがある．このノイズは音響的特徴を含まないため非音響ノイズと定義する．本稿では，非音響ノイズの中でも特に破裂音や促音など息の吹かれ方によって収録時に発生するマイク内の振動板のぶれによって起こるノイズに注目する．

従来の音声収録では，非音響ノイズや背景雑音，風等の影響を軽減するためにマイクにポップフィルタを使用することが多かった．本研究では非音響ノイズをあえて収録するためにポップフィルタを使用しない状況で音声収録を行った．図1は同じマイクでポップフィルタありなしそれぞれの実際に音声を収録した際の波形である．図1(b)の波形のとおり，ポップフィルタがない場合には非音響ノイズの影響を受けて波形に歪みが発生していることがわかる．

2.2　話者性と耐雑音性

非音響ノイズを含む話者モデルは話者識別性能を向上させることが報告されている[1]．これは，非音響ノイズの入り方が話者ごとに異なるため音響的特徴を有していなくても話者モデル内の分布としては個人性を表現することが可能であるからである．一方，非音響ノイズはノイズ成分を表現しているともいえるため耐雑音性が向上することも期待できる．特に非音響ノイズは低周波成分に現れるため，話者モデルに含ませることで環境雑音の低周波成分を吸収することが考えられる．雑音除去手法のひとつとして入力音声の雑音やSNRを合わせることで耐雑音性を向上させることがあった．しかし，非音響ノイズを用いることでこれまでのようなあらかじめ雑音に対して処理を行うことやノイズ推定をすることなく，識別性能が向上できればその有用性は非常に高くなる．



(a) ポップフィルタあり

(b) ポップフィルタなし

図1　ポップフィルタありとなしのヘッドセットマイクで音声を同時に収録した際の音声波形

3　実験

非音響ノイズの耐雑音性について調査するためにUBM-GMMを用いた話者照合実験を行った．

3.1　データベース

非音響ノイズを含むデータベースとしてヘッドセットマイク（SHURE SM10A-CN）を2本用意し，1本はポップフィルタを装着，1本はポップフィルタを装着しない状態で同時に収録された音声を使用した．口からマイクまでの距離は3センチとした．収録文章はJNASデータベースから音素バランスを考慮した全話者共通の50文章と話者ごとにランダムに選択した50文章で，各話者合計100文章ずつ収録した．

複数の異なる雑音での性能を比較するために，電子協騒音データベース[2]の中から走行自動車内（1500 ccクラス）と展示会場（ブース内）の2種類の雑音を選び，SNRを0dBから30dBまで5dBきざみで雑音重畳をした．

3.2　実験条件

実験条件等を表1に示す．学習データおよびテストデータが同種の雑音・同SNRの場合をMC（マッチドコンディション）とした．また，使用

*Improvement of noise robustness using non-acoustical noise for speaker verification by NAKANISHI Ryôsuke, SHIOTA Sayaka, KIYA Hitoshi (Tokyo Metropolitan University)

表 1　実験条件

学習データ （話者依存モデル）	70 文章 × 17 名 （計 1190 文章）
テストデータ	30 文章 × 17 名 （計 510 文章）
UBM 用データベース	JNAS（女性のみ）
UBM 学習データ	23657 文章
GMM 混合数	1024
サンプリング周波数	16 kHz
フレーム長	25 msec
フレームシフト	10 msec
特徴量	MFCC 19 次+ Δ + ΔΔ

したデータはヘッドセットマイク (H) のポップフィルタあり，なし (1,0) が存在するため，表記としてはマッチドコンディションでポップフィルタありの場合 MC(H1)，ポップフィルタなしの場合 MC(H0) とした．ポップフィルタなしでテストデータにのみ雑音を重畳した場合を提案法とし，それぞれの照合スコアを算出した．照合スコアを Z-Norm[3] によって正規化し，その等価エラー率（EER）の比較を行った．Z-Norm による正規化式は以下のとおりである．

$$S_{\mathrm{ZN}}(X,U,I) = \frac{S(X,U,I) - \mu_U}{\sigma_U} \qquad (1)$$

ここで，$S(X,U,I)$ は入力データ X とユーザモデル U，および詐称者モデル I から求めた照合スコアである．詐称者音声に対する照合スコアの平均および分散をそれぞれ μ_U，σ_U とした．

3.3　実験結果

走行自動車内雑音（1500 cc クラス）を重畳した結果を図 2 に，展示会場（ブース内）の雑音を重畳した結果を図 3 に示す．ここで，clean は雑音を重畳しない場合を表す．MC(H1) と提案法を比較すると，走行自動車内および展示会場 (ブース内) どちらの雑音においても，全 SNR において MC(H1) よりも提案法の方が EER が低い結果となった．また，提案法と同じく非音響ノイズを含む MC(H0) と提案法を比較しても提案法の方が EER が低くなっている．このことから非音響ノイズは話者性を含むだけでなく，雑音に頑健なモデルになっていることがわかる．さらに図 2，図 3 で傾向が同じであることから，非音響ノイズを含むモデルは雑音の種類にも依存せず識別性能を向上させることができることがわかる．一方，MC(H0) と MC(H1) の EER を比較すると，特に自動車内雑音下においては差が非常に小さく，展示会場（ブース内）においても改善の幅は小さい．これは学習データにも雑音を重畳させると，非音響ノイズのノイズ成分が他

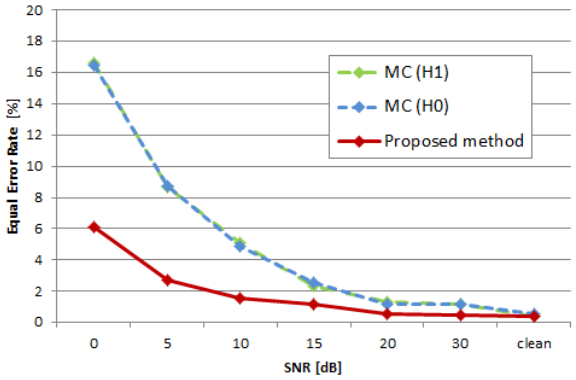


図 2　走行自動車内雑音を重畳した場合

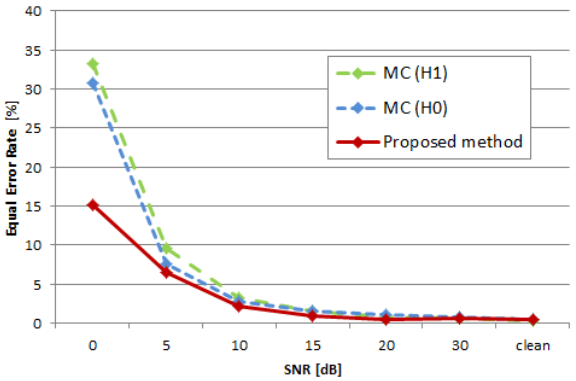


図 3　展示会場（ブース内）雑音を重畳した場合

の雑音と同じ分布になってしまい話者性を表現しつつ雑音を吸収する分布としての効果が弱まってしまうことが考えられる．以上の結果から，非音響ノイズをモデルに含ませるだけで環境雑音の種類や SNR に依存しにくいノイズロバストな話者モデルを構築できることがわかった．

4　おわりに

本稿では，非音響ノイズを含む話者モデルを用いて雑音環境下での話者照合実験を行い，非音響ノイズの与える影響について調査した．実験結果より，非音響ノイズを含む話者モデルは環境雑音の種類や SNR に影響を受けることなく話者性および耐雑音性を向上させることがわかった．今後は，耐雑音性についてより詳細な調査や大規模な話者照合実験を行うことなどが考えられる．

参考文献

[1] 塩田ら，“ 非音響ノイズを用いた話者照合の検討，”日本音響学会秋季大会，pp.88–89，2014.
[2] 電子協騒音データベース，http://research.nii.ac.jp/src/JEIDA-NOISE.html.
[3] C.Barras and J.L.Gauvain,”Feature and score normalization for speaker verification of cellular data,” in Proc. ICASSP’03, pp.49–52，2003.