Vol.2009-SLP-79 No.27
2009/12/22

順位統計量を用いた話者照合のためのコホート話者選択方法

岡本 悠[†1] 柘植 覚[†2]
堀内 靖雄[†1] 黒岩 眞吾[†1]

本論文では，順位統計量を用いた話者照合手法を紹介する．さらに，順位統計量を用いた話者照合手法における照合コストを下げるためのコホート話者の選択方法について提案する．コホート話者は申告者の音声に対してシステムに登録された不特定多数の話者モデル（GMM）との尤度の順位を基準に作成する．評価実験として，科学警察研究所が構築した大規模話者骨導音声データベースに収録されている男性 283 名の気導音声を用いて実験を行った．従来手法では，全話者 283 名による順位統計量で算出した minDCF が 0.0092 に対して，提案手法では平均 57 名の順位統計量で 0.0098，101 名の順位統計量で 0.0094 という同等の性能を達成した．また，照合スコアとして T-norm を用いた場合の minDCF が 0.0154 だった．

Using Cohort Speaker for Text-Independent Speaker Verification with Rank-Estimator

Haruka Okamoto,[†1] Satoru Tshge ,[†2]
Yasuo Horiuchi[†1] and Shingo Kuroiwa[†1]

In this paper, we introduce a novel speaker verification method which determines whether a claimer is accepted or rejected by the rank of the claimer in a large number of speaker models instead of score normalization, such as T-norm and Z-norm. The method has advantages over the standard T-norm in speaker verification accuracy. However, it needs much computation time as well as T-norm that needs calculating likelihoods for many cohort models. Hence, we also discuss the speed-up the method that selects cohort speakers for each target speaker in the training stage. This data driven approach can significantly reduce computation time resulting in faster speaker verification decision. We conducted text-independent speaker verification experiments using large-scale Japanese speaker recognition evaluation corpus constructed by National Research Institute of Police Science. From the corpus, we used utterances collected from 283 Japanese males. As results, the proposed method whose the number of cohort speaker is 57 achieved an minDCF of 0.0098, while using 282 speakers as cohort speaker obtained 0.0092 and T-norm obtained 0.0154.

1. 序　　論

話者照合では，ユーザがシステムに対して音声を入力すると共に，自分が誰であるかを申告話者として ID 番号を入力する等の方法で申告する．そして，入力音声が申告話者の発声であるか否かを判定する[1)]．この判定には音声ベクトルが入力されたとき，申告話者モデルに対する対数尤度を求めるとともに，背景話者（詐称者）に対する対数尤度も算出し，それらの尤度比を用いる事で閾値判定を行う[2)]．しかし，人間の音声は発話内容や環境，発声時期等の影響を受けるので入力の度にばらつき，再現性がなく不安定である．そのため，同じ話者でも算出される尤度比もばらついてしまうことが知られている[3)]．音声の特性に起因するこの問題は，話者照合における閾値設定を困難なものとしてきた．

この問題に対し，申告話者の尤度を，背景話者の尤度を用いて統計的に正規化することによって尤度のばらつきを抑制するためのスコア正規化手法が提案された．正規化は式（1）に従う．

$$\tilde{S}_{norm}(X) = \frac{P(X|m_{Tar}) - \mu_\lambda}{\sigma_\lambda} \quad (1)$$

式（1）内の $\tilde{S}_{norm}(X)$ は正規化後の照合スコア，$P(X|m_{Tar})$ は入力音声 X と申告話者 m_{Tar} との対数尤度を示す．詐称者の平均 μ_λ，分散 σ_λ は，Zero-normalization[4)]（Z-norm），Test-normalization[5)]（T-norm）といった，代表的な手法によって異なる．これらの手法は，申告話者モデルに対する詐称者の尤度は正規分布に従うという仮定に基づいている．

Z-norm は申告話者モデルに対して，不特定多数の詐称者音声との間で尤度を計算し，各尤度の平均と分散を事前に求めておく．この平均と分散を用いて照合時に出力される $P(X|m_{Tar})$ を正規化する．

一方，T-norm はテスト音声が入力されたときに目標話者モデルに対する尤度を求めると

†1 千葉大学大学院融合科学研究科
Chiba University
†2 徳島大学工学部
The University of Tokushima

ともに，不特定多数の背景話者モデル（T-norm モデル）に対する尤度を計算して各尤度の平均と分散を求める．この平均と分散を用いて $P(X|m_{Tar})$ を正規化する．

一般に，Z-norm の正規化項である平均と分散は学習音声を用いて計算する．そのため，テスト音声に含まれる不安定さが考慮されないので，照合精度が低くなる．ただし，Z-norm で用いる平均・分散と分散は学習音声を用いるために学習の段階で求めておくことができるため，テスト音声の目標話者に対する尤度のみを計算すれば良く，照合に必要な計算コストは目標話者ただ一人ですむ．これに対し，T-norm では正規化項である平均と分散は照合時のテスト音声を用いて計算している．そのため，学習音声を利用した Z-norm よりもテスト音声の音響的な変動に取り入れている分，背景話者から得られる分布をより正確に表すことができるため，照合精度が高い．ただし，照合時に多数の T-norm モデルと尤度を計算しなければならないため，計算コストが T-norm モデルの数だけ高くなる．

最近の話者照合の研究では，T-norm の照合精度を向上させつつ，計算コストを削減するために，目標話者の照合に効率の良い話者であるコホート話者[6] を選択する手法として AT-norm[7]，KL-T-norm[8]，SMC-T-norm[9] が提案されている．これらの手法はいずれもコホート話者を選択後，T-norm スコアを求めて話者照合を行う．

AT-norm は，事前に目標話者毎に T-norm モデルの中からコホート話者を選択する．この選択方法はまず，不特定多数の音声を用意しておき，目標話者と各 T-norm モデルそれぞれとに対して尤度を求める．次に，話者モデル毎に求められた不特定多数の音声に対する尤度を成分としたベクトルとする．そして，目標話者のベクトルと各 T-norm モデルの尤度ベクトルを，市街地距離（マンハッタン距離）によって比較し，目標話者と距離が近い k 人の話者を選択する．

KL-T-norm も AT-norm と同様に，コホート話者を目標話者毎に不特定多数の音声に対する尤度のベクトルを比較して選択する．しかし，市街地距離を用いた AT-norm とは異なり，カルバックライブラー距離（KL 距離:Kullback-Leibler distance ）によって尤度ベクトルを比較し，目標話者と距離が近い k 人の話者を選択する．

SMC-T-norm は KL 距離をベースとした K-means 法用いたクラスタリングにより，コホート話者を選択する手法である．クラスターは話者モデル間で互いに似ているモデルを集めたことを意味する．AT-norm，KT-norm は k 人というある決められた数のコホート話者を選択するのに対し，SMC-T-norm は目標話者毎に各クラスターの人数によって，照合に適切な数のコホート話者を選択することができる．コホート話者数に自由度を持たせて T-norm の照合スコアを計算できる点で，上述の二つの手法とは性質が異なる．

以上のような統計的手法による照合スコアである T-norm を用いた閾値設定に関する研究がされてきたが，喜多らは T-norm ではなく順位情報を用いた話者照合手法[10)11)] を提案した．順位情報とは，テスト音声とシステム内の目標話者と多数の話者モデルとの間で尤度を算出し，順位付けを行うことで得る順位そのものである．順位付けによって得られた目標話者の順位情報が上位 R 位以内に入っていれば，発話者を目標話者として受理，入っていなければ詐称者として棄却するという手法を提案した．順位情報を用いた話者照合は，様々な研究の土台となった T-norm よりも性能が高いことが示されている．しかしながら，喜多らの研究では，背景話者にどれほどの人数が必要なのかが吟味されておらず，順位付けには，モデル数が多いほど計算コストが多くなるという問題がある．

そこで本稿では，順位情報を用いた話者照合に必要な背景話者としてコホート話者を利用する手法を提案する．そして，科学警察研究所が構築した大規模話者骨導音声データベースを利用した話者照合実験で背景話者及び，照合精度，計算コストについて評価する．また，本論文では，順位情報はロバスト統計学の順位統計量[12)] として定義されていることに基づき，順位統計量と定義しなおす．

以降の第２章では，順位統計量を用いた話者照合手法について説明する．第３章では，背景話者にコホート話者を利用した，順位統計量を用いた話者照合手法について説明する．第４章では，評価実験について説明し，結果および，考察を述べる．最後に第５章では，本稿の結論を述べる．

2. 順位統計量を用いた話者照合手法

本章では，従来手法として順位統計量を用いた話者照合手法について説明する．図１にこの話者照合手法の流れを示す．

まず，目標話者モデル m_{Tar} と，全 N 人の背景話者モデル $m_1, m_2, m_3, \ldots, m_N$ を用意し，そのそれぞれに対してテスト音声 X が入力されたときの尤度 $P_{Tar}(X|m_{Tar}), P_1(X|m_1), \ldots, P_N(X|m_N)$ を計算する．

次に，求めた全ての各尤度を降順にソーティングし，モデル毎の順位統計量（尤度の順位）$R_{Tar}, R_1, \ldots, R_N$ を得る．そして，目標話者モデルの順位統計量 R_{Tar} が，ユーザが定めた順位閾値 R_{th} と等しいか，上位ならば，申告者を受理し，下位ならば棄却する，という手法である．

しかしながら，喜多らの順位統計量を用いた照合手法は，背景話者の数だけ尤度の計算が必要という問題点がある．これは背景話者が多いほど，計算コストが増加していくことを意

Vol.2009-SLP-79 No.27
2009/12/22

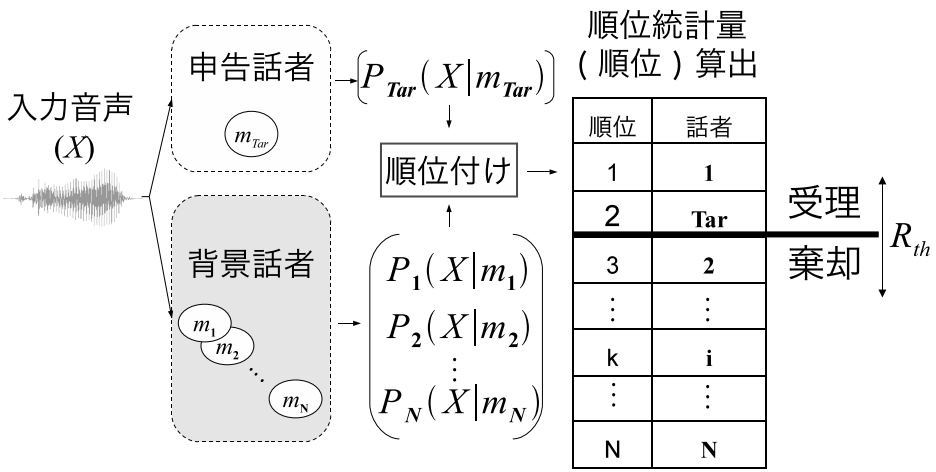

図 1 順位統計量による話者照合手法

味する．それにも関わらず，実際の照合では，目標話者の尤度の順位と順位閾値までの上位数人の尤度の順位だけで発話者の受理，棄却を決定している．以上の理由から従来手法では，順位閾値以下の背景話者モデルに対しては尤度を計算しても照合に有効利用されない場合があると考えられる．

そこで次章では，背景話者全てではなく，目標話者毎に定めた照合精度に寄与が見込めるコホート話者のみを用いた上で順位統計量を算出し，話者照合に利用する手法を提案する．

3. コホート話者を用いた順位統計量による話者照合方法

本章では，従来手法よりも照合にかかる計算コストを削減するために，背景話者として目標話者毎のコホート話者を用いた順位統計量による話者照合手法を提案する．始めに，コホート話者の選択方法について説明する．従来の T-norm をベースにしたスコアを利用する話者照合手法[7)8)9)] は背景話者（あるいはコホート話者）に対する尤度の分布を用いる事で正確な照合を行おうとするものであった．そのため，分布を表現するための平均と分散を求める必要があり，背景話者との尤度の計算が必須となる．それらに対して順位統計量を用いた話者照合手法では，順位を求めるために必要な数人の尤度を計算すればよく，背景話者に対する計算コストを削減しやすい．そこで，提案手法では順位統計量を用いるためのコホート話者を選択する．提案手法で用いるコホート話者は図 2 の流れに従って求める．

まず，目標話者の学習音声 T と全 N 人の背景話者モデル $m_1, m_2, \ldots, m_N$ を用意し，各話者毎の尤度 $P_1(T|m_1), P_2(T|m_2), \ldots, P_N(T|m_N)$ を計算する．

次に，求めた全ての尤度を降順にソーティングし，モデル毎の順位統計量（尤度の順位）$R_1, R_2, \ldots, R_N$ を得る．そして，上位 r 位までの背景話者モデルを目標話者のコホート話者モデル $S = \{m_1, m_2, \ldots, m_r\}$ として照合で利用する．

学習音声 T が $n \geq 2$ 発話ある場合，まず，各学習音声 $T_1, T_2, \ldots, T_n$ それぞれにおける，背景話者モデルに対する尤度を求めることで，上位 r 位までのコホート話者モデル $S_1, S_2, \ldots, S_n$ を選出する．そして，各学習音声毎に選出された全てのコホート話者モデルを和集合（$S_1 \cup S_2 \cup \ldots \cup S_n$）によって統合し，照合で利用する．このとき，コホート話者数は r よりも多くなり得る．ここで得られるコホート話者数を新たに r' と定義する．

次に，以上の手法で求めたコホート話者を用いた場合の順位統計量による話者照合手法の流れを図 3 に示す．提案手法では照合の際に目標話者だけでなく，同時に目標話者毎に予め選択されたコホート話者を背景話者として利用する．残りの手順は従来手法と同様である．まず，テスト音声が入力された際，目標話者モデルとコホート話者モデルに対する尤度を求め，モデル毎の順位統計量を得る．そして，目標話者の順位統計量が，順位閾値 R_{th} よりも等しいか，上位ならば，申告者を受理し，以下ならば棄却する，という手法である．

この提案手法により，尤度の計算コストを（コホート話者数）/（全話者数）に削減できる．

4. 評価実験

4.1 実験条件

本節では，評価実験の条件を説明する．実験には，科学警察研究所により整備された，『大規模話者骨導音声データベース』[13)14)] を使用した．このデータベースに収録されている音声の中で，気導音声として収録された男性話者 283 名の ATR 音素バランス文 50 文のうち 1 時期分の音声を用いた．各話者は同時期に 2 回ずつ同じ発話セットを発話しており，学習音声として 1 回目の発話のうち 5 文を利用した．話者モデルには GMM（Gaussian Mixture Model）[16)] を用い，混合数は 96 で学習した．コホート話者の選択にも，学習音声と同じ音声を利用した．これらの音声から特徴パラメータを表 1 の音響分析条件に従って抽出した．本実験では HTK（Hidden Markov Model Toolkit）を用いて求めた特徴量（12 次元 MFCC+1 次元対数パワー +12 次元デルタ MFCC+1 次元デルタ対数パワー）を CMS 処理[15)] を行った後に学習，評価した．

Vol.2009-SLP-79 No.27
2009/12/22

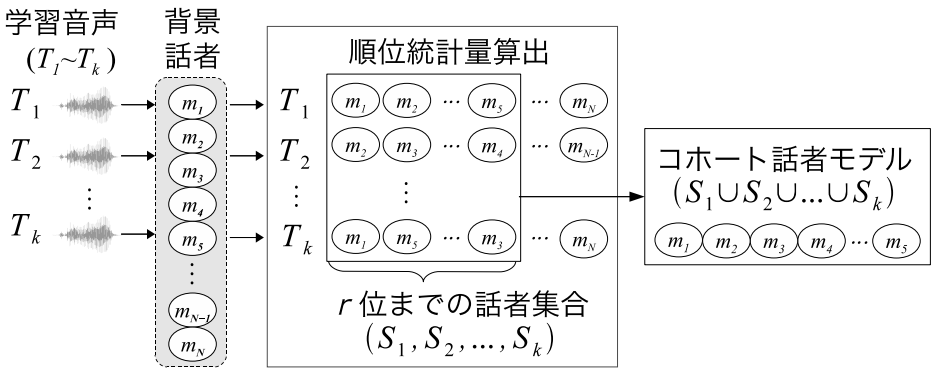

図 2 コホート話者の選択方法

表 1 分析条件

サンプリング周波数	16kHz
高域強調	0.97
フレーム長	25ms
フレーム周期	10ms
窓タイプ	ハミング窓
特徴パラメータ（CMS 処理後）	12 次元 MFCC 1 次元対数パワー 12 次元デルタ MFCC 1 次元デルタ対数パワー

評価には，各時期の 2 回目に発話された学習音声とは異なる発声内容の音声 45 文を用いた．これにより試行回数は本人の場合で 12,735 回（283 人× 45 発話），詐称者の場合で 3,591,270 回（283 人× 282 詐称者× 45 発話）となる．なお，詐称者の発声はオープンな実験となるよう，詐称者本人のモデルは除いた．

4.2 評価実験

評価実験として，提案手法と従来手法の順位統計量を用いた話者照合精度，及び T-norm を用いた場合の話者照合精度を比較した．背景話者については，従来手法と T-norm では，目標話者を除く 282 名全てを利用した．提案手法では，5 文ある学習音声の 1 文毎に，目標話者を含めた全 283 名の 1/16（$r = 18$），1/10（$r = 27$），1/8（$r = 35$），1/4（$r = 70$），にあたる順位 r までを占める上位の話者を選択し，和集合をとることでコホート話者を決定した．このため，実際に用いたコホート話者は発話毎に選択した人数よりも多く，申告話者毎に異なる．照合精度の比較には，DET（Detection Error Trade-Off）曲線と NIST 評価でよく用いられる式（2）で定義される最小決定コスト関数（minDCF）[17] を用いた．

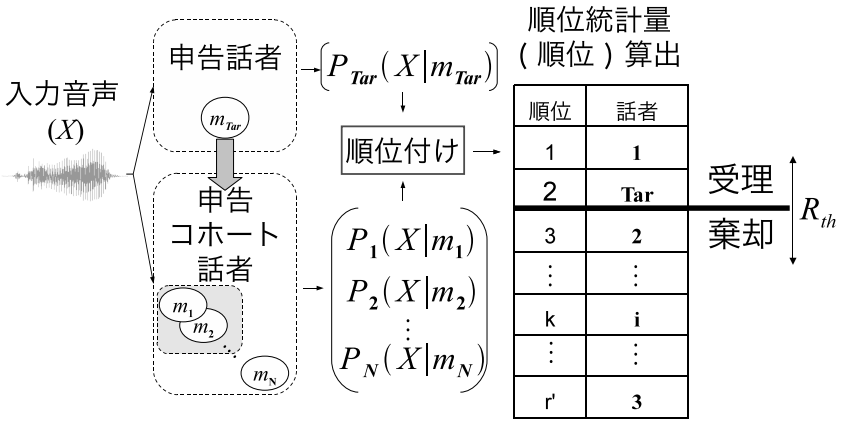

図 3 コホート話者を用いた順位統計量による話者照合手法

$$minDCF = min\{0.1P_{fr}(\theta) + 0.99P_{fa}(\theta)\} \quad (2)$$

ただし，θ は閾値，$P_{fr}(\theta)$ と $P_{fa}(\theta)$ は，それぞれ閾値 θ に対する本人誤棄却率と詐称者誤受理率である．

4.3 実験結果

提案手法と従来手法，及び T-norm の評価実験結果を図 4，表 2，表 3 に示す．図 4 の縦軸は本人誤棄却率（FRR:False Rejecton Rate）を，横軸は詐称者受理率（FAR:False Acceptance Rate）を示し，曲線は各背景話者数における DET 曲線を示し，図中の曲線「cohort283」は従来手法を意味する．順位統計量を用いた手法では，順位閾値 R_{th} により離散的に FRR と FAR が求まり，図中の左側の点から $R_{th} = 1, R_{th} = 2, R_{th} = 3, \ldots$ となる．表 2 は背景話者数毎に得た minDCF と minDCF を得た FRR と FAR を示す．minDCF を得た順位閾値はいずれも 1 位のときである．表 3 は r によって目標話者毎に選ばれたコホート話者数の平均値と，その平均値の全人数に対する割合を示す．

まず，照合精度の観点で評価する．順位統計量を利用した照合手法の曲線同士をそれぞれ比較する．図 4 より，全背景話者を利用して得られた曲線と比べて，背景話者が少ないほど，曲線は右にシフトする．しかしながら，$r = 70$，$r = 35$ による曲線は全背景話者を利用した場合の曲線とほぼ同等と判断できる．表 2 より，minDCF にも大きな差はない．また，

Vol.2009-SLP-79 No.27
2009/12/22

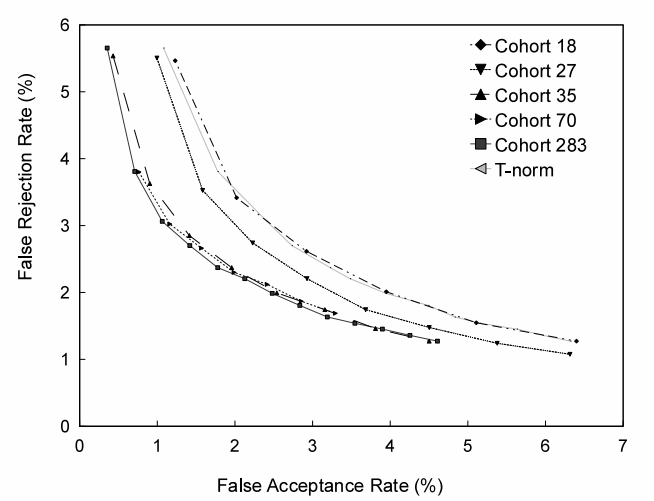

図 4 R 位の詐称者受理率と本人棄却率

$r = 1/16$ の曲線が T-norm と同等の曲線を示し，その他の r における曲線は全て T-norm よりも精度が高いことが，図 4，表 2 からわかる．

次に，照合コストの観点で評価する．表 3 より，各 r における実際に増加したコホート話者の平均値から，それぞれ計算対象となるコスト（人数）が減少する様子が分かる．

4.4 考　　察

評価実験の結果により，照合精度・コストの二つ観点から，順位統計量を用いた話者照合で背景話者が全人数を用いた場合と同等の照合精度を達成するために必要なコホート話者数は，全人数の約 1/5 で十分になる．順位統計量を求めるためのコホート話者数が，計算コストの削減量に直結するため，計算コストが 1/5 ですむ事がわかる．提案手法は計算コストを削減しつつ，従来手法と同等の精度を発揮した．T-norm と同等の照合精度を，提案手法で達成するために必要な人数は，全人数の約 1/10 であることがわかる．

しかしながら問題点として，まず表 2 より，背景話者が少ないほど FRR は低いものの，FAR が高く，背景話者が多いほどそれらは逆の傾向を示す．次に，提案手法の閾値が順位なので T-norm に比べて細かい調整ができないことも問題点として上げられる．本稿では，前者の点に焦点をあてて，FAR を重視する minDCF において，提案手法は従来手法にわずかに劣る要因を考察する．

表 2 r に対する minDCF

r	対人数比	FRR(%)	FAR(%)	minDCF
18	1/16	5.40	0.55	0.0108
27	1/12	5.50	0.46	0.0101
35	1/8	5.54	0.43	0.0098
70	1/4	5.64	0.38	0.0094
282	1/1	5.65	0.35	0.0092
282	T-norm	7.50	0.80	0.0154

表 3 実際のコホート話者数の平均値

r	話者数平均値	計算コスト
18	33	11.7%
27	46	16.3%
35	57	20.2%
70	101	35.8%
282	282	100.0%

まず，話者モデルを多次元空間（話者空間と定義する）中のベクトルとして考える．簡便的に 2 次元上の話者モデルの分布模式図を図 5 に示す．提案したコホート話者の選定方法では学習音声毎に，尤度の大小のみで背景話者からコホート話者を選択するため，目標話者の近傍には存在するものの，次元毎のモデルの近さが考慮されない．このため，あるいくつかの特定の次元から見て距離が近いがために尤度が高くなる場合で選ばれたコホート話者に対し，尤度が高くなる要因でない次元に，テスト音声が入ってきたとき，コホート話者は有効でない．

このような観点から表 2 をとらえ直す．FRR は従来手法では，話者空間に多数の背景話者が存在するため，コホート話者では補いきれない次元軸で目標話者よりも近い背景話者が存在するので FRR は上がる．一方，提案手法では，コホート話者のみで照合を行うために FRR は下がる．FAR においても同様に考える事ができる．従来手法では，多数の背景話者モデルにより，詐称者の発声は背景話者モデルと尤度が高くなりやすいため，FAR は低くなりえる．一方，提案手法では，話者空間においてコホート話者では補いきれない次元軸に詐称者の入力音声が入ると，目標話者に対する尤度が高くなる可能性が上がるため，FAR が高くなりえる．

以上のような理由から，第 1 章で述べた T-norm モデルの選定方法に関する諸研究のように，順位統計量を用いた話者照合手法においても，目標話者毎に適切な背景話者を選択する必要が伺える．この手法における，適切な背景話者とは，申告者が本人か詐称者かを選り分ける空間を規定できる十分有効な背景話者である．この背景話者を選択する事ができれば，照合精度を向上させつつ，計算コストを削減する事ができると考えられる．

5. 結　　論

本論文では，背景話者にコホート話者を利用した，順位統計量を用いた話者照合手法につ

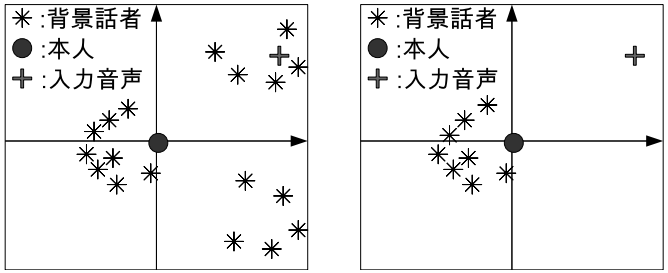

図 5 2 次元話者空間上の話者モデルの分布模式図（左：背景話者全て，右：コホート話者のみ）

いて提案した．本研究ではコホート話者を，目標話者の学習音声毎に多数の背景話者との尤度を求め，上位に来た話者の和集合をとることで選択した．大規模話者骨導音声データベースを用いた評価実験の結果，提案手法では，全話者の約 1/5 にあたるコホート話者数で全話者を背景話者とした場合の精度と同等の精度を発揮することがわかった．また，T-norm と同等の精度を達成するには，全話者の約 1/10 のコホート話者で十分であった．これらのコホート話者を用いることで，従来手法と比較して約 80%の計算コストの削減を達成し，T-norm と比較して 90%のコスト削減を達成した．

今後は，話者空間上の話者モデルの分布を考慮したコホート話者の選択方法を検討するとともに，順位統計量と T-norm をサポートベクトルマシンで組み合わせた照合手法も検討する．

謝辞 本研究には科学警察研究所との共同研究により大規模話者骨導音声データベースを利用させていただいた．本研究の一部は科学研究費補助金（基盤研究（B）21300060，若手研究（B）19700172）の補助を受けて行った．

参 考 文 献

1) Reynolds, D.A., Campbell, W.M., “Text-Independent Speaker Recognition,” in Benesty, J., Sondhi, M., Huang, Y., (Eds.), *Springer Handbook of Speech Processing*, Springer, pp.763–784, 2008.
2) K. P. Li and J. E. Por ter, “ Normalizations and selection of speech segments for speaker recognition scoring, ” Proc. IEEE. Int. Conf. Acoustics, Speech and Signal Processing, vol. 1, pp. 595-598, April 1988.
3) Reynolds, D.A., Campbell, W.M., “ Text-Independent Speaker Recognition, ” in Benesty, J., Sondhi, M., Huang, Y., (Eds.), Springer Handbook of Speech Processing, Springer, pp.763-784, 2008.
4) Li, K.-P., Porter, J.E., “Normalization and selection of speech segments for speaker recognition scoring,” ICASSP, pp.595–598, Apr. 1988.
5) Auckenthaler, R., Carey, M., Lloyd-Thomas, H., “Score Normalization for text-independent speaker verification system,” Digital Signal Processing, Vol.10, pp.42–54, Jan. 2000.
6) Rosenberg, Aaron, E. DeLong, Joel. Lee, Chin-Hui. Juang, Biing-Hwang. Soong, Frank K., ”The use of cohort normalized scores for speaker verification”, ICSLP, pp. 599-602, 1992.
7) Sturim, D.E., Reynolds, D.A., “Speaker Adaptive Cohort Selection for T-norm in Text-Independent Speaker Verification,” ICASSP, pp.741–744, Mar. 2005.
8) D. Ramos-Castro, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Or tega-Garcia, “ Speaker verification using speaker- and test- dependent fast score normalization, ” Pattern Recognition Letters, vol. 28, pp. 90?98, Jan. 2007.
9) Ravulakollu, K., Apsingekar, V.R., De Leon, P.L., “Efficient speaker verification system using speaker model clustering for T and Z normalizations,” ICCST, pp.56–62, Oct. 2008.
10) 喜多 雅彦，黒岩 眞吾，柘植 覚，蒔苗 久則， 長内 隆，鎌田 敏明， 谷本 益巳，土屋 誠司，福見 稔，任 福継,“大規模話者骨導音声データベースを用いたテキスト独立型話者照合実験 “，情報処理学会研究報告. SLP, 音声言語情報処理 2007(129) pp.183-188 20071220
11) 喜多 雅彦, 柘植 覚, 黒岩 眞吾, 任 福継， “多数の話者モデル内での順位情報を用いた話者照合 “, 言語・音声理解と対話処理研究会, Vol.2007, No.A701, pp.7-12, 2007 年 7 月.
12) P.J.Huber.， E.M.Ronchetti “Robust Statistics” Wiley Series in Probability and Statistics, 第 2 版, John Wiley Sons (2009).
13) Tsuge, S., Osanai, T., Makinae, H., Kamada, T., Fukumi, M., Kuroiwa, S., “Combination method of Bone-conduction Speech and Air-conduction Speech for Speaker Recognition,” Interspeech, pp.1929–1932, Sep. 2008.
14) 蒔苗 久則，長内 隆，鎌田 敏明，谷本 益巳，” 大規模話者骨導音声データベースの構築 ‘” と予備的な解析電子情報通信学会技術研究報告. SP, 音声 107(165) pp.97-102 20070719
15) Furui, S., “Cepstral analysis technique for automatic speaker verification,” IEEE Trans. Acoust. Speech Signal Processing, Vol.ASSP-29, pp.254–272, Apr. 1981.
16) Reynolds, D.A., Quantieri, F.T., Dunn, R.B., “Speaker Verification Using Adapted Gaussian Mixture Models,” Digital Signal Processing, Vol. 10, pp. 19–41, Jan. 2000.
17) The 2006 NIST Speaker recognition evaluation plan, http://www.itl.nist.gov/iad/mig/tests/sre/2006/sre-06 _evalplan-v9.pdf, 2006.