論文 / 著書情報
Article / Book Information

論題(和文)	話者認識の国際動向
Title	
著者(和文)	越仲孝文, 篠田浩一
Author	Takafumi Koshinaka, Koichi Shinoda
出典(和文)	日本音響学会誌, vol. 69, no. 7, pp. 342-348
Journal/Book name	, vol. 69, no. 7, pp. 342-348
発行日 / Issue date	2013, 7

日本音響学会誌 69 巻 7 号 (2013), pp. 342–348

小特集—話者認識に関する研究の動向—

話者認識の国際動向*

越仲孝文 (NEC)**・篠田浩一 (東工大)***

43.72.Fx, Pf

1. はじめに

本稿では，話者認識技術の研究開発に関する国際動向を通して，最近の技術トレンドを概観する。この分野では，1990 年代に提案されたガウス混合モデル (Gaussian Mixture Model; GMM) に基づく方法が基礎となり，機械学習に基づく識別や分析の手法が近年積極的に導入されてきた。本稿ではまず，GMM 以降の技術の変遷について簡単に述べる。各国の研究機関の動きは国際会議及び評価ワークショップにみることができるので，次にはこれらのイベントと，それらに参加する常連研究機関を紹介する。特に，この分野で重要な役割を担う，米国立標準技術研究所 (National Institute of Standards and Technology; NIST) による評価ワークショップ Speaker Recognition Evaluation (SRE) について詳しく述べる。商用化の事例についても言及したい。

2. 技術の変遷

まず，近年のこの分野でどのような技術的変遷があったかを簡単に振り返る。技術の詳細については後続のより踏み込んだ解説 (本特集，『話者認識で用いる機械学習』及び『話者認識におけるロバストネス』) を参照されたい。

2.1 GMM に基づくアプローチ

1990 年代半ば Reynolds ら [1] は，音声から得られる特徴ベクトル系列の生成過程がガウス混合分布に従うとするガウス混合モデル (GMM) による方法を提案した。特徴ベクトルの分布が時刻や発話内容によらず話者のみに依存するという思い切った仮定を置いているにもかかわらず，話者性をよく表現できるこの方法は，以後のテキスト独立話者認識の一つの標準となった。

特に，不特定話者の平均的な音声のモデル，すなわちユニバーサル背景モデル (Universal Background Model; UBM) を導入し，UBM を基点とした最大事後確率 (Maximum *A posteriori* Probability; MAP) 学習により目的話者の音声のモデルを構成する話者照合の方法は GMM-UBM として広く知られている [2]。GMM-UBM では，未知話者 X の音声が既知話者 A のものであるか否かを，話者 A のモデルに関する尤度と UBM に関する尤度の差分によって検定する (図–1)。

2.2 サポートベクトルマシンの登場

2000 年代に入ると，サポートベクトルマシン (Support Vector Machine; SVM) に代表される識別指向な学習機械が話者認識にも採り入れられた。Campbell ら [3] は，GMM を構成する各ガウス分布の平均ベクトルをすべて連結した高次元ベクトル (GMM スーパベクトル) を話者の特徴量とし，入力音声がある話者のものか否かを SVM で識別する方法 (GMM-SVM) を提案した (図–2)。

この方法は GMM-UBM を上回る性能を示した。音声がある話者のものか否かを判断する話者照合問題は，もともと 2 クラス識別問題を得意とする SVM と親和性が高く，GMM-SVM は話者照合の標準的な方法として受け入れられた。

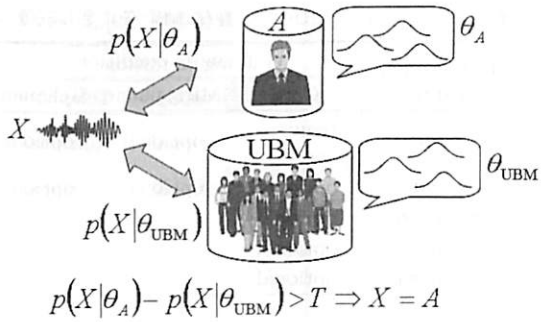
2.3 因子分析によるモデル化

2000 年代後半になると，回線 (チャネル) の違いによる音声の変動，ひいては同一話者でもその時々で音声が変動するという話者内変動 (within-speaker variability) の問題を緩和する方法が提案された。大成功を収めたこの方法は接合因子分析 (Joint Factor Analysis; JFA) と呼ばれ [4]，因子分析の手法に基づき高次元の GMM スーパベク

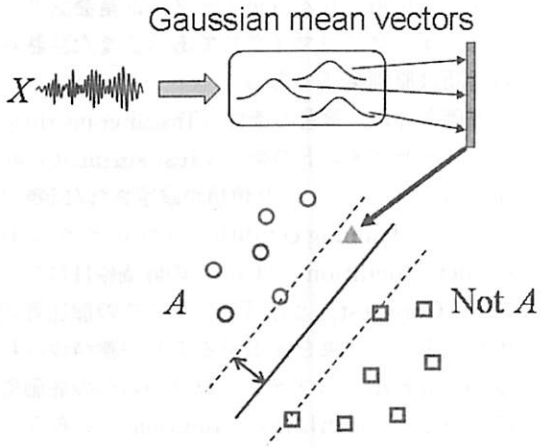
* International trends of speaker recognition technology.

** Takafumi Koshinaka (Information and Media Processing Laboratories, NEC Corporation, Kawasaki, 211–8666) e-mail: koshinak@ap.jp.nec.com

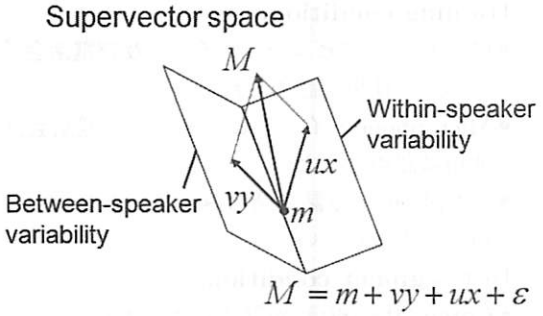
*** Koichi Shinoda (Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Tokyo, 152–8552)



図–1 GMM-UBM



図–2 GMM スーパベクトルと GMM-SVM



図–3 Joint factor analysis (JFA)

トル M を低次元の因子 y（話者性を表す）及び x（話者以外の回線特性などを表す）の二つに分解する（図–3）。話者認識においては話者性（between-speaker variability）のみが重要であるから，y を用いて話者の照合や識別を行う。

更に近年は，JFA を簡略化した i-vector と呼ばれる方法 [5] が，デファクト標準といえるほど普及している。i-vector では，話者性の因子とそれ以外の因子を区別せずに因子分析を行い（total variability），後段で線形判別分析（Linear Discriminant Analysis; LDA）のような別の分析手法を使って話者性を抽出する。このように簡便な手法で抽出した因子が話者認識において非常に有効であることが知られている。また，SVM のような“高級な”識別器を用いずとも，コサイン類似度のような簡単な尺度で高い認識精度が得られるなど，興味深い展開をみせている [6]。

3. 国際会議

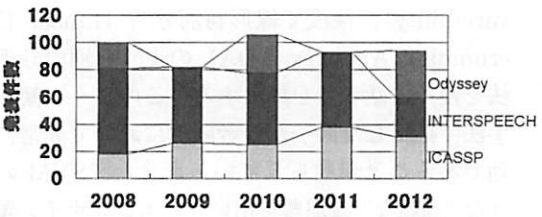
話者認識の最新の研究成果が発表される国際会議を紹介する。主要国際会議は次の三つ：

- ICASSP[1]（IEEE International Conference on Acoustics, Speech, and Signal Processing）：IEEE 主催。音響・音声・信号処理の分野では最も大規模かつ高水準なカンファレンス。年 1 回，春に開催される。一般論文の採択率は例年 50%前後で，2013 年は 53%。
- INTERSPEECH[2]（Annual Conference of the International Speech Communication Association）：International Speech Communication Association（ISCA）主催。音声情報処理に関する主要国際会議で，同じく年 1 回，秋に開催される。一般論文の採択率は例年 50～60%程度で，2012 年は 52%。
- Odyssey[3]（Speaker and Language Recognition Workshop）：ISCA の Speaker and Language Characterization Special Interest Group（SpLC SIG）主催。隔年で，後述の評価ワークショップ NIST SRE と同年に開催される。話者認識と言語認識をテーマとした研究発表がなされる。

これらの国際会議で 2008～2012 年の 5 年間に発表された，話者認識に関する論文件数を図–4 に示す。この 5 年間，発表件数は安定して 100 件前後あり，依然としてこの分野に対する研究者の興味は強い。

近年の研究事例は，JFA や i-vector，SVM，あるいは最近パターン認識の種々の領域で流行中のディープラーニングなどの機械学習手法に関する基礎的な研究から，携帯端末への組み込みを想定した

[1] http://www.icassp2013.com/
[2] http://www.interspeech2013.org/
[3] http://www.odyssey2012.org/



図–4 主要国際会議における関連論文発表件数

高速化や省メモリ化などの応用研究まで，幅広い。

4. 評価ワークショップ

4.1 NIST SRE の概要

話者認識の性能を評価する大規模な競争型評価ワークショップ（コンテスト）として，米国立標準技術研究所（NIST）が主催する NIST Speaker Recognition Evaluation（SRE）がある[4]。1996 年より定期的に開催され，産官学の多くの研究機関が参加している。

当初（1996～2006）は毎年開催されていたが，近年は隔年開催となっている。しなしながら，隔年となった 2008 とその次の 2010 は，連続して過去最多の参加機関数（2008 は 46，2010 は 58）を記録しており，最近の 2012 でも，2010 と同数の参加があったようである。NIST SRE はこの分野の活性化に大きく寄与している。図–5 では，国別の参加機関数を過去 3 回にわたって計数している。米国，中国，フランス，スペインで全体の約半分を占める。参加機関の種別については，米カーネギー・メロン大（CMU）やマサチューセッツ工科大（MIT）などの大学と，米 SRI，仏 LIMSI-CNRS，シンガポール I2R などの政府系研究機関からの参加が非常に多いが，IBM のような大手ベンダや，スペインの Agnitio などのベンチャ企業からの参加もある。我が国からは，東工大が 2010 以降参加している [7, 8]。

NIST SRE の参加者は，共通のデータセットと評価条件の下で，テキスト独立話者照合，すなわち不特定の内容について話された音声が申告話者本人のものであるか否かを判定する課題の精度を競う。データセットや評価条件は不変ではなく，毎回定義される。以降では最新の 2012 で実施された評価内容 [9] を紹介する。

[4]http://www.nist.gov/itl/iad/mig/sre.cfm

表–1 2012 NIST SRE の学習/評価条件（[9] より転載）

Test segment condition	Training condition		
	Core	Microphone	Telephone
Core	required (Core test)	optional	optional
Extended	optional	optional	optional
Summed	optional		
Known	optional		
Unknown	optional		

4.2 2012 NIST SRE

評価に使用される音声データは，電話会話とインタビュー会話（マイク）である。また，話される言語は原則英語である。

評価条件は，学習の条件（Training condition）とテストセグメントの条件（Test segment condition）の組み合わせで，9 種類が設定された（表–1）。このうち，Training condition = “Core” かつ Test segment condition = “Core” の評価条件はコアテスト（Core test）と呼ばれ，すべての参加者が評価を実施して結果を提出することが義務づけられる（required）。コアテスト以外の八つの評価条件については，提出は任意（optional）である。

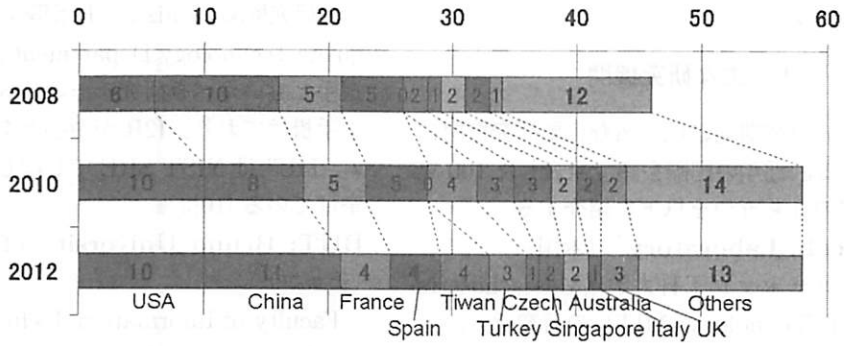
学習及びテストセグメントの条件はそれぞれ次のとおり。

Training condition:

- Core…インタビュー（マイク）及び電話会話のすべての使用が認められる。
- Microphone…インタビューのみ。電話会話の使用は認められない。
- Telephone…電話会話のみ。インタビューの使用は認められない。

Test segment condition:

- Core…20～160 秒の電話会話又はインタビュー（マイク）。
- Extended…テストセグメント（音声データ）は Core と同一だが，学習データ内のより多くの話者と照合を試行する大規模テスト。
- Summed…電話会話又はインタビューの 2 話者の音声が 1 チャネルに記録されている。2 話者のいずれかに目的話者が存在するか否かを判断する話者検出のテスト。
- Known…試行する照合は Extended と同一だが，すべての非目的話者（詐称者）を既知と仮定する。



図–5 NIST SRE 国別参加研究機関数 (2008–2010)

表–2 コアテストの 5 種類の Common condition

	w/o added noise	w/ added noise	in a noisy environment
Interview speech	1	3	—
Phone call speech	2	4	5

- Unknown…Known とは逆に，すべての非目的話者を未知と仮定する。

2010 NIST SRE との対比でみると，2010 では約 10 秒の短時間音声で学習や照合を行う条件が設定されていたが，2012 では廃止された。また，詐称者が既知 (Known) か未知 (Unknown) かという観点は，2012 で初めて採り入れられた。このように，評価条件は毎回変更される。データセットも毎回変わるので，回を跨いだ性能比較は原則意味をなさない。

“必須種目” であるコアテストについて更に詳しくみていこう。コアテストは，Common condition 1〜5 と呼ばれる 5 種類の条件に細分化されており，そのおのおのについて各参加者の成績が算出される。五つの Common condition について表–2 にまとめる。

Common condition 3〜5 は雑音環境を想定した評価になっている。このように実環境を想定した評価は 2012 で初めて導入された。Common condition 3〜4 では，クリーンな音声に雑音を重畳する。重畳する雑音は 2 種類で，一つは空調機器などの機械音，もう一つは多数の話者の音声とされている。

次に，性能の良し悪しを測る評価指標について述べる。一般的な話者照合の評価では，本人を他人と誤る割合である誤棄却率 $P_{\mathrm{Miss|Target}}$ と，他人を本人と誤る割合である誤受理率 $P_{\mathrm{FalseAlarm|NonTarget}}$ が基本となる。誤受理と誤棄却はトレードオフ関係にあるので，$P_{\mathrm{Miss|Target}} = P_{\mathrm{FalseAlarm|NonTarget}}$ となるようなシステムの動作点での誤り率で性能を測る。この誤り率は，等誤り率 (Equal Error Rate; EER) と呼ばれる重要な指標である。

NIST SRE でも，誤受理率と誤棄却率が基本となるのは同じだが，両者を対等に扱う EER ではなく，一方に重みを付けたコスト関数 C_{Det} を用いる。例えばコアテストでは，C_{Det} は次のように定義される。

$$C_{\mathrm{Det}} = C_{\mathrm{Miss}} \times P_{\mathrm{Miss|Target}} \times P_{\mathrm{Target}} + C_{\mathrm{FalseAlarm}} \times P_{\mathrm{FalseAlarm|NonTarget}} \times (1 - P_{\mathrm{Target}})$$

ここに C_{Miss} と $C_{\mathrm{FalseAlarm}}$ は，それぞれ誤受理率と誤棄却率の単位コスト，P_{Target} は照合対象の音声セグメントに目的話者が存在する事前確率である。2012 のコアテストでは $C_{\mathrm{Miss}} = C_{\mathrm{FalseAlarm}} = 1$，$P_{\mathrm{Target}} = 0.01$ 又は 0.001 と設定される。すなわち，誤棄却よりも誤受理が重視される。

最終的には，すべて受理あるいはすべて棄却と判断した場合のシステム性能 (最悪値) が 1 となるように正規化したコスト関数 C_{Norm} で性能を測る。

$$C_{\mathrm{Norm}} = \frac{C_{\mathrm{Det}}}{\min[C_{\mathrm{Miss}} \times P_{\mathrm{Target}}, C_{\mathrm{FalseAlarm}} \times (1 - P_{\mathrm{Target}})]}$$

コアテストでトップクラスのシステムは，C_{Norm} で 0.2 程度，EER で 2%程度の成績をマークしている。近年はシステムの大規模化が進んでおり，5〜10 個程度の話者照合システムを組み合わせて更に高精度を追求するケースが多い。システム開発も，複数の研究機関による混成チームによって

行われることが多い。

5. 主な研究機関

話者認識の研究開発に取り組む主要研究機関について述べる。近年の国際会議で発表した 100 を超える機関のうち五つを以下に紹介する。

MIT Lincoln Laboratory[5] (米国)

米マサチューセッツ工科大学 (Massachusetts Institute of Technology; MIT) の研究所で，主に国家安全保障に関する研究に取り組んでいる。GMM によって各話者の音声をモデル化する方法を 1990 年代に提唱した。特に，不特定話者のユニバーサル背景モデル (UBM) をベースとして，事後確率最大 (MAP) 学習によって各話者のモデルを構成する GMM-UBM は，単純だが有効性の高い方法として広く用いられた (2.1 節参照)。また，GMM の全ガウス分布の平均ベクトルを連結した GMM スーパベクトルを話者の特徴量として，サポートベクトルマシン (SVM) により本人か否かを判断する GMM-SVM も彼らの提案による (2.2 節参照)。

CRIM: Centre de Recherche Informatique de Montréal[6] (カナダ)

ケベック州の政府系研究機関。主要な研究領域の一つに音声情報処理を掲げている。回線特性などに起因する音声の変動 (within-speaker variability) は，話者認識の障害となる。因子分析の考え方に基づいてこの種の変動要因を定式化する手法である接合因子分析 (Joint Factor Analysis; JFA) は，CRIM の音声チームを率いる Patrick Kenny の提案による。JFA 及びその後継手法である i-vector は，目下この分野の多くの研究者に使われており，影響力は非常に大きい (2.3 節参照)。今現在の最有力研究機関の一つ。後述の Agnitio, BUT との協力により開発されたシステム (三者の頭文字から "ABC" と名付けられた) は，NIST SRE でトップクラスの成績をマークしている。

I2R: Institute for Infocomm Research[7] (シンガポール)

シンガポール科学技術研究庁 (Agency for Science, Technology, and Research; A*STAR) 傘下の研究機関。Haizhou Li が率いる Human Language Technology Department は，アジアの代表選手ともいうべき研究チーム。音声情報処理を広く手掛けており，I2R が中心となった多国籍チーム "I4U" は NIST SRE でしばしば優秀な成績を挙げている [10]。

BUT: Bruno University of Technology[8] (チェコ)

Faculty of Information Technology (FIT) の音声処理グループは，Matlab で書かれた JFA のデモプログラムを公開している [11]。JFA Cookbook と呼ばれるこのプログラムは多くの研究者に利用されており，JFA Cookbook を使って NIST SRE に参戦する研究機関も非常に多い。

Agnitio[9] (スペイン/南アフリカ)

スペインの Universidad Autónoma de Madrid からのスピンオフで誕生した大学発ベンチャ。次節で紹介するように科学捜査などの用途に供する製品を扱っている。研究活動では，同社のチーフサイエンティスト Niko Brummer が，話者認識システム等が出力する任意のスコアを対数尤度比と見なせるように正規化 (calibrate) するツールキット BOSARIS Toolkit を公開している [12]。このツールにより，多数のサブシステムを統合した大規模システムを容易に構築することができる。NIST SRE に参加するすべてのシステムがこのツールキットを利用している。

6. 商用化状況

本節では，話者認識技術の商用化の状況と国内外ベンダの製品事例について紹介する。本稿では特に，個人認証系，分析系という主要な二つの応用について述べる。

6.1 個人認証系

応用例として，電話を介した金融取引や商取引で，本人確認を音声で行う，あるいは，重要な施設や情報へのアクセスを人物ごとに制御するなど，幅広い用途が考えられる。しかしながら，この種の用途ではたいてい認識誤りが致命的であり，おのずと高い認識精度が要求される。また，音声というメディアの利便性が活きるためには，数秒程度の短いフレーズで本人確認できることが必要で

[5] http://www.ll.mit.edu/
[6] http://www.crim.ca/
[7] http://hlt.i2r.a-star.edu.sg/
[8] http://speech.fit.vutbr.cz/
[9] http://www.agnitio-corp.com/

ある。更に，スマートフォンなどの携帯端末の普及に伴い，最近は屋外等の雑音下での使用にも耐える頑健性が求められている。

そのため，個人認証系の応用では，発声内容を限定しないテキスト独立型だけでなく，テキスト依存型の手法を用いて技術的な不足をカバーするケースが多い。幸いなことに個人認証の場合，ユーザ（話者）はシステムに対して原則協力的であり，発声内容を指定されることを特段重い負担とは考えない。近年，研究はテキスト独立一色といってよく，テキスト依存に関する論文は少数派だが，少なくとも個人認証に関しては，テキスト依存の方がユーザのニーズに応えている。研究者はこの現実にもう少し目を向けてもよいのではなかろうか？

以下，誌面の制約もあり多くは記せないが，個人認証系の製品を有するベンダを紹介する。

Nuance[10]

音声認識ではよく知られた最大手ベンダ。話者認識製品も有する。VocalPassword は，音声自動応答による銀行取引（テレホンバンキング）などで実績のあるサーバ型話者照合システム。テキスト独立/依存/指定に対応可能。FreeSpeech はコールセンタオペレータの会話から顧客を認証するテキスト独立型のシステム。クレーマなどのブラックリストチェックにも適用可能。

Baidu-I2R

5 節でも紹介した I2R が昨秋，中国インターネット検索大手 Baidu との共同研究によりスマートフォン（Lenovo A586）向けの話者照合機能を実現[11]。テキスト依存型の手法により 2 秒程度の発声で個人を認証し，端末のロック解除を行う。

アニモ[12]

富士通からのスピンオフで誕生した日本のベンチャ企業。VoicePassport は，テキスト独立/テキスト依存，サーバ/組込み，電話/マイクなど種々の運用形態に対応可能な話者照合ミドルウェア。

6.2 分 析 系

大量の音声データ（人間が聴ききれないほどの）を分析して，人物に関する有用な情報を抽出する。例えば警察の科学捜査が重要な応用の一つである。古くは 1963 年の「吉展ちゃん誘拐殺人事件」で，身代金要求の電話の声と容疑者の音声の同一性を判断する音声鑑定の重要性がクローズアップされたが（本特集，『法科学分野における話者認識の動向』参照），捜査過程で得られた音声試料を音声データベースに照会して容疑者を絞り込むといった応用は，コンピュータが最も得意とするところであろう。

このような応用は，現在のところ警察や防衛など，特殊な場面に限られる。しかし同様の技術はインターネットを流通する膨大な動画像データの検索などにも転用できるので，産業上の利用可能性は小さくはない。またこの種の応用では，最終的な結果の解釈を人間が行うので，多少の認識誤りがあっても，人間の作業を代替して人間の負担を軽減できれば十分有用である。なお，分析系の応用は，個人認証系とは逆にユーザが協力的ではない，もしくは分析システムの存在を認識していないので，発声内容を制約することができない。従ってテキスト独立型の手法を適用することになる。

以下，分析系の製品の幾つかを紹介する。

Agnitio

5 節でも紹介した大学発ベンチャ。ASIS は話者照合による検索機能を備えた音声データベースシステム。BATVOX は音声試料の一対一照合を行う音声鑑定支援システム。BS^3 は軍事用途を想定した音声分析システムで，人物の行動監視などに用いられる。

NEC

国内では NEC が，音声分析系の製品の提供を最近開始した。本製品は，電話等の音声を大規模データベースと照合し，音声から人物特定を行うための分析支援システムである。

7. おわりに

本稿では，話者認識技術の国際動向というテーマで，近年の主な技術変遷と，国際会議や NIST による評価ワークショップ（SRE）が重要な役割を果たしている状況を述べた。また，国際的な場で中心的な働きをする研究機関の幾つかを紹介した。更に，国内外における話者認識技術の商用化事例についても簡単に触れた。

さて，ここで我が国に目を向けると，評価ワークショップ NIST SRE への参加は，2010 以降の東工

[10] http://www.nuance.com/

[11] http://www.a-star.edu.sg/Media/News/PressReleases/tabid/828/articleType/ArticleView/articleId/1747/Default.aspx

[12] http://www.animo.co.jp/

大が唯一のケースである。話者認識の分野での現状の日本のプレゼンスは，残念ながらさほど高くない。しかしながら，1990 年代，MIT の Reynolds ら [1] とほぼ同時期に，GMM を用いた話者照合に関する先駆的な研究が Matsui と Furui[13] によりなされて以来，決して少なくない数の独創的研究が日本から出ている（最近では [14–16] など）。また，近年は GMM や SVM はもちろんのこと，JFA でもフリーのソフトウェアが公開されており (5 章参照)，NIST SRE は敷居の高いものではなくなってきている。この分野で，我が国から世界への情報発信が今後増えることを期待したい。

謝　辞

執筆に際して有益な助言をいただいた東京工業大学の Sangeeta Biswas 女史，NEC の大西祥史主任研究員，谷真宏主任に感謝の意を表する。

文　献

[1] D.A. Reynolds and R.C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech Audio Process.*, 3, 72–83 (1995).
[2] D.A. Reynolds, T.F. Quatieri and R.B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Process.*, 10, 19–41 (2000).
[3] W.M. Campbell, D.E. Sturim and D.A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Process. Lett.*, 13, 308–311 (2006).
[4] P. Kenny, G. Boulianne, P. Ouellet and P. Dumouchel, "Speaker and session variability in GMM-based speaker verification," *IEEE Trans. Audio Speech Lang. Process.*, 15, 1448–1460 (2007).
[5] N. Dehak, P.J. Kenny, R. Dehak, P. Dumouchel and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio Speech Lang. Process.*, 19, 788–798 (2011).
[6] A. Kanagasundaram, R. Vogt, D.B. Dean, S. Sridharan and M.W. Mason, "Weighted LDA techniques for i-vector based speaker verification," *Proc. ICASSP 2012*, pp. 4781–4784 (2012).
[7] M. Ferras, S. Biswas, K. Shinoda and S. Furui, "NIST SRE 2010: Tokyo Tech Speaker Recognition," *Proc. NIST Speaker Recognition Evaluation (SRE) Workshop* (2010).
[8] S. Biswas, J. Rohdin and K. Shinoda, "Tokyo Tech Speaker Recognition," *Proc. NIST Speaker Recognition Evaluation (SRE) Workshop* (2012).
[9] The NIST Year 2012 Speaker Recognition Evaluation Plan http://www.nist.gov/itl/iad/mig/upload/NIST_SRE12_evalplan-v17-r1.pdf
[10] H. Li, B. Ma, K.-A. Lee, H. Sun, D. Zhu, K. C. Sim, C. You, R. Tong, I. Karkkainen, C.-L. Huang, V. Pervouchine, W. Guo, Y. Li, L. Dai, M. Nosratighods, T. Tharmarajah, J. Epps, E. Ambikairajah, E.-S. Chng, T. Schultz and Q. Jin, "The I4U system in NIST 2008 speaker recognition evaluation," *Proc. ICASSP 2009*, pp. 4201–4204 (2009).
[11] Joint Factor Analysis Matlab Demo http://speech.fit.vutbr.cz/software/joint-factor-analysis-matlab-demo
[12] BOSARIS Toolkit https://sites.google.com/site/bosaristoolkit/home
[13] T. Matsui and S. Furui, "Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMM's," *IEEE Trans. Speech Audio Process.*, 2, 456–459 (1994).
[14] T. Ogawa, H. Hino, N. Murata and T. Kobayashi, "Speaker verification robust to intra-speaker variation using multiple kernel learning based on conditional entropy minimization," *Proc. INTERSPEECH 2011*, pp. 2741–2744 (2011).
[15] S. Biswas, M. Ferras, K. Shinoda and S. Furui, "Acoustic forest for SMAP-based speaker verification," *Proc. INTERSPEECH 2011*, pp. 2377–2380 (2011).
[16] M. Nishida and S. Yamamoto, "Speaker clustering based on non-negative matrix factorization," *Proc. INTERSPEECH 2011*, pp. 949–952 (2011).

越仲　孝文

平 3 京大・工・航空卒，平 5 同大大学院修士課程了。同年 NEC 入社。現在，NEC 情報・メディアプロセッシング研究所主幹研究員。統計的パターン認識，機械学習の研究に興味を持つ。平 12 電子情報通信学会学術奨励賞。本会，電子情報通信学会各会員。工博。

篠田　浩一

昭 62 東大・理・物理卒，平元同大大学院修士課程了。同年 NEC 入社。平 9〜10 米ルーセント・テクノロジー社ベル研究所客員研究員。平 14 東大。現在，東工大大学院教授。統計的パターン認識，音声認識，映像認識，ヒューマンインタフェース等の研究に興味を持つ。平 9 本会粟屋潔学術奨励賞，平 10 電子情報通信学会論文賞。本会，電子情報通信学会，情報処理学会，人工知能学会，IEEE，ACM 各会員。工博。