話者照合システムに対する合成音声による詐称

益子　貴史[†]　　徳田　恵一[††]　　小林　隆夫[†]

Imposture against a Speaker Verification System Using Synthetic Speech

Takashi MASUKO[†], Keiichi TOKUDA[††], and Takao KOBAYASHI[†]

あらまし　本論文では，話者照合システムに対する合成音声を用いた詐称について検討している．従来，発声すべきテキストがそのつどシステムから指定されるテキスト指定型話者照合システムを用いれば，テープレコーダ等による録音音声を用いた詐称を防ぐことができるとされてきた．しかし，テキスト合成方式の進歩により，任意のテキストに対して多様な声質の音声を合成できるようになってきたことから，合成音声による詐称に関する検討も必要と考えられる．そこで，HMM に基づくテキスト指定型話者照合システムに対し，少量の学習データで合成音声の声質を目標話者の声質に変換できる HMM に基づく音声合成システムを用いて，合成音声による詐称を行った．登録話者による 3 文章の発声を用いて音声合成システムを学習することにより，合成音声に対する等誤り率が 30%以上となり，合成音声による詐称が十分可能であること示した．

キーワード　話者照合，セキュリティ，音声合成，HMM

1. まえがき

コンピュータやネットワークの普及・高度化に伴い，様々な分野に電子化の波が押し寄せている．電話やインターネットを使ったバンキング，ショッピング，情報検索・情報提供，リモートアクセスなどのサービスはそのほんの一例である．これらのサービスの実用化においては，セキュリティの確保が重要な鍵となっており，パスワードや電子署名に利用する暗号理論の研究や，暗号の安全性を検証する暗号解析・攻撃に関する研究などが盛んに行われている．

一方，音声認識や話者認識の研究の発展とともに，人間のコミュニケーションにおいて身近な存在である音声を本人確認手段として用いることが考えられ，実用化に向けて数多くの研究が行われている [1]〜[3]．実際，安全性の高い話者照合技術が確立されれば，電話を用いたバンキングやショッピングなどのほかにも，建物や部屋の入口における音声キーなどへの応用も考えられ，有用性は高いと考えられる．

[†] 東京工業大学大学院総合理工学研究科，横浜市
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama-shi, 226–8502 Japan
[††] 名古屋工業大学知能情報システム学科，名古屋市
Faculty of Engineering, Nagoya Institute of Technology, Nagoya-shi, 466–8555 Japan

当然，話者照合においても，他の本人確認手段と同様，セキュリティの確保ができるかどうかが重要な問題となる．これまでの研究では，発声すべきテキストが任意であるテキスト独立型やあらかじめテキストが決められているテキスト依存型に対し，認識時にシステムからテキストがそのつど指定されるテキスト指定型によれば，テープレコーダなどで録音された音声による詐称の危険性は少ないといわれている．また，話者正規化法の導入により，他の話者による詐称を受理しにくい手法の報告もされている [4]〜[6].

しかしながら，これまで声質変換による詐称 [7], [8] については検討が始められたものの，合成音声を用いた詐称に関する検討はあまり行われていない．最近のテキスト音声合成方式の進歩により，自然性の高い音声 [9] や，音声単位の統計的モデル化に基づいた多様な声質での音声の合成 [10]〜[12] が，任意のテキストに対して比較的容易に実現できるようになってきたことから，今後はこれまであまり考慮されていなかった合成音声を用いた詐称あるいは攻撃に対して，話者照合システムが安全か否かを検討することが必要になると考えられる．

このような観点から，本論文では現在提案されている一般的な話者照合システムに対して，合成音声を用いた詐称の可能性を検討している．現在，多くの話者

照合方式が隠れマルコフモデル（HMM）に基づいていることから，本論文でも HMM に基づく話者照合システムを対象とする．そして，音声合成手法としては，データベースから自動的に学習することができ，また少量の適応データを用いて容易に合成音声の声質を変換できる [11], [12] ことから，筆者らが先に提案した HMM に基づく音声合成 [10] を用い，これにより生成される合成音声による詐称の可能性を検討する．

具体的な詐称の手順としては，

（1） なりすます相手の音声を入手する

（2） 音声合成システムを適応する

（3） 合成音声を再生して照合システムに入力する

となる．なりすます相手による音声を得る方法としては，例えば，相手に直接会って話しかけ音声を録音する，また電話音声を用いた話者照合システムの場合には，なりすます相手に電話をかけて音声を録音する，などが考えられる．この際，会話の内容は何でもよく，例えば道を尋ねる，間違い電話のふりをするなどにより，数文章程度の音声が得られるような場合を想定している．

2. 話者照合システムに対する合成音声による詐称

図 1 に，話者照合システムに対する合成音声による詐称の様子を示す．

テキスト指定型話者照合システムでは，照合時にまず適当なテキストを生成して話者に提示する．次に，入力音声から特徴パラメータを抽出して申告話者のモデルに対するゆう度を計算し，しきい値と比較することにより受理または棄却の判断を行う．また，発話内容が指定したテキストに対応するかを判定するテキスト照合も行い，話者照合及びテキスト照合の両方で受理された場合のみ入力音声が申告話者のものであると判定する．

合成音声を用いて詐称を行う場合は，まず，詐称しようとする登録話者の音声データが入手できたとして，それに基づいて音声合成システムに用いる音声単位のモデルをあらかじめ学習しておく．そして詐称の際には，当該話者として話者照合システムに申告し，提示されたテキストに対応する音声を合成して話者照合システムへ入力する．話者照合システムでは，通常の話者による音声と同様に，合成音声から特徴パラメータを求めてゆう度を計算し，受理または棄却の判断を行う．

2.1 話者照合システム

本論文では，録音音声に対してロバストであること，また，多くの話者照合システムが HMM に基づいていることから，HMM に基づくテキスト指定型話者照合システムを用いる．

まず学習用音声データから特徴パラメータを求め，各話者ごとに特定話者音素 HMM を学習する．また，全登録話者の音声データを用いて，不特定話者音素

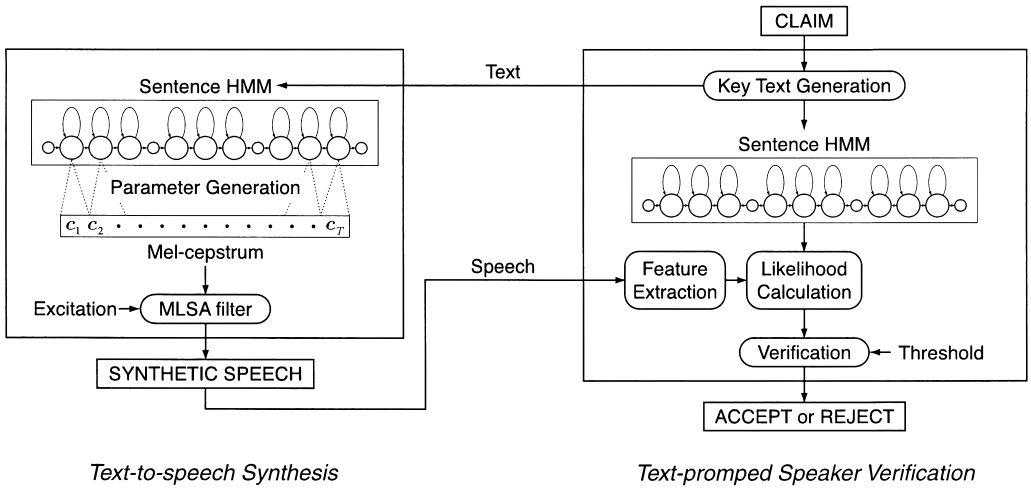


図 1 話者照合システムに対する合成音声による詐称

Fig. 1 Imposture using synthetic speech against an HMM-based speaker verification system.

HMM を学習する．その後，学習データを用いて各話者ごとにしきい値を設定する．

話者照合時には，システムが指定したテキストを音素列に変換し，これに従って音素 HMM を連結して文 HMM を構成する．次に，入力音声から求められた特徴パラメータの申告話者の文 HMM に対するゆう度を計算する．

ゆう度はテキストや発声時期により大きなばらつきがあるため，ゆう度の正規化を行う必要がある．ゆう度正規化法としては，ゆう度比検定の考えに基づく正規化法 [4]，コホート正規化 [5] などが提案されているが，本研究では次式で表される全登録話者から作成した不特定話者モデルを用いて正規化を行う方法 [6] を用いた．

$$L_s(\boldsymbol{O}) = \frac{1}{T}\left(\log P(\boldsymbol{O}|\lambda_s) - \log P(\boldsymbol{O}|\lambda_{all})\right) \quad (1)$$

ここで，$L_s(\boldsymbol{O})$ は話者 s に対する入力音声 $\boldsymbol{O}$ の正規化した対数ゆう度，T は入力音声のフレーム数，λ_s，λ_{all} はそれぞれ話者 s 及び不特定話者の文 HMM である．ただし，実際の計算には $P(\boldsymbol{O}|\lambda)$ の代わりにビタビパス上でのゆう度を用いた．

2.2 音声合成システム

本研究で用いる HMM に基づく音声合成システムは，学習部と合成部の二つの部分から構成される．学習部では，まず音声データベースからメルケプストラム分析 [13] によりメルケプストラムを求める．更に動的特徴量である Δ，Δ^2 パラメータを計算し，静的特徴量と合わせて特徴ベクトルとし，音素 HMM を学習する．HMM の学習後，学習データに対する Viterbi アラインメントにより HMM の各状態の継続長のヒストグラムを求め，これをガウス分布で近似して各状態の状態継続長分布とする．

合成部では，まず合成したい任意のテキストを音素列に変換する．この音素列に従って前述の音素 HMM を接続し，与えられたテキストに対応する一つの文 HMM をつくる．この文 HMM から，ゆう度最大化基準に基づくパラメータ生成アルゴリズム [14] によりメルケプストラム系列を生成し，MLSA フィルタ [15], [16] を用いて合成音声を得る．

HMM からのゆう度最大化基準に基づくパラメータ生成アルゴリズム [14] では，連続出力分布 HMM のパラメータセット λ からある状態遷移系列 $\boldsymbol{Q}$ に沿って観測される長さ T の出力ベクトル系列 $\boldsymbol{O} = [\boldsymbol{o}_1', \boldsymbol{o}_2', \ldots, \boldsymbol{o}_T']'$ を，$P(\boldsymbol{Q}, \boldsymbol{O}|\lambda, T)$ を最大化するように生成している．ここで時刻 t における出力ベクトル $\boldsymbol{o}_t$ は静的特徴 $\boldsymbol{c}_t$ 及び動的特徴 $\Delta\boldsymbol{c}_t$，$\Delta^2\boldsymbol{c}_t$ により，$\boldsymbol{o}_t = [\boldsymbol{c}_t', \Delta\boldsymbol{c}_t', \Delta^2\boldsymbol{c}_t']'$ と表される．動的特徴量は一般に，$\Delta^{(0)}\boldsymbol{c}_t = \boldsymbol{c}_t$，$\Delta^{(1)}\boldsymbol{c}_t = \Delta\boldsymbol{c}_t$，$\Delta^{(2)}\boldsymbol{c}_t = \Delta^2\boldsymbol{c}_t$ とおけば

$$\Delta^{(n)}\boldsymbol{c}_t = \sum_{\tau=-L^{(n)}}^{L^{(n)}} w^{(n)}(\tau)\boldsymbol{c}_{t+\tau}, \quad n = 0, 1, 2 \quad (2)$$

と表すことができる．ただし，$L^{(0)} = 0$ ，$w^{(0)} = 1$ とする．HMM の状態遷移系列 $\boldsymbol{Q} = (q_1, q_2, \ldots, q_T)$ が既知の場合には，$\boldsymbol{Q}$ に沿って出力されるパラメータ系列 $\boldsymbol{O}$ は与えられた HMM λ に対し，$P(\boldsymbol{O}|\boldsymbol{Q}, \lambda, T)$ を最大化することにより得られ，このときのパラメータ系列 $\boldsymbol{C} = [\boldsymbol{c}_1', \boldsymbol{c}_2', \ldots, \boldsymbol{c}_T']'$ は

$$\partial \log P(\boldsymbol{O}|\boldsymbol{Q}, \lambda, T)/\partial \boldsymbol{C} = \boldsymbol{0} \quad (3)$$

として与えられる線形連立方程式を解くことにより求められる．

3. 実験条件

3.1 音声データベース

音声データは ATR 日本語音声データベースの男性話者 20 名による音韻バランス文を用いた．サンプリング周波数は 10 kHz である．話者 10 名を話者照合システムの登録話者とし，詐称者としては本人以外の登録話者及び登録外話者を用いた．各話者 150 文章のデータを 50 文章ずつ A，B，C の 3 セットに分け，A セットを話者照合システムの学習データ，B セットを音声合成システムの学習データ，C セットをテストデータとした．データベースに付属するラベルデータに基づき，無音を含めて 48 種類の音素でラベル付けした．なお，話者照合システム，音声合成システムともに同一の音素セットを用いた．

3.2 話者照合システム

一般に，対象とする話者照合システムがどのようなスペクトルパラメータを用いているかという情報は得られないのが普通である．そこで本実験では，音声合成システムにおいてスペクトルパラメータとしてメルケプストラムを用いていることから，話者照合ではこれと異なる LPC ケプストラムを用いたシステムを想定した．フレーム周期 5 ms，窓長 25.6 ms のブラックマン窓を用い，15 次 LPC 分析により得られた 0〜15 次 LPC ケプストラム，及びその Δ，Δ^2 パラメータを特徴ベクトルとした．

HMM は音素単位の 3 状態 left-to-right モデルで，各状態の出力分布は 1，2，または 3 混合対角共分散ガウス分布とした．特定話者モデルは各話者 50 文章ずつ，また不特定話者モデルは全登録話者の学習データ 500 文章を用い，EM アルゴリズムにより学習した．

照合時の正規化対数ゆう度に対するしきい値は話者独立とし，学習データに対する本人棄却率（FRR）と詐称者受理率（FAR）が等しくなるように設定した．ただし，図 2（3 混合モデルの場合）に示すように，FRR = FAR = 0 となる区間（灰色の領域）が存在する場合には，その区間の中心をしきい値とした．ここで，オープンデータを用いた場合，適切なしきい値は学習データで設定したしきい値より低くなる傾向にある [17] ため，FRR = FAR = 0 となる区間の最大値を用いることは現実的ではない．

なお，今回の実験では入力音声に対するテキスト照合は行っていない．

3.3 音声合成システム

音声合成システムでは，フレーム周期 5 ms，窓長 25.6 ms のブラックマン窓を用い，15 次メルケプストラム分析により得られた 0〜15 次メルケプストラム，及びその Δ，Δ^2 パラメータを特徴ベクトルとした．

HMM は音素単位で前後の音韻環境を考慮しない monophone モデルとした．各音素 HMM は 2，3，または 4 状態単一対角共分散ガウス分布 left-to-right モデルであり，各状態はガウス分布で近似された状態継続長分布をもつ．登録外話者の学習データ 500 文章を用いて学習した不特定話者モデルを初期モデルとし，登録話者による発声 1，3，5，または 50 文章を用いて EM アルゴリズムにより学習した．状態継続長分布は，学習データに対する Viterbi アラインメントにより得られた各状態の継続長のヒストグラムをガウス分布で近似することにより求めた．ここで，登録話者の音声データに出現しない音素については，不特定話者モデルをそのまま用いた．音声合成時の各状態の継続長はそれぞれの状態の継続長分布の平均値とした．

合成音声の音源は，話者照合システムでピッチなどの音源情報を利用しないことから，白色雑音とした．

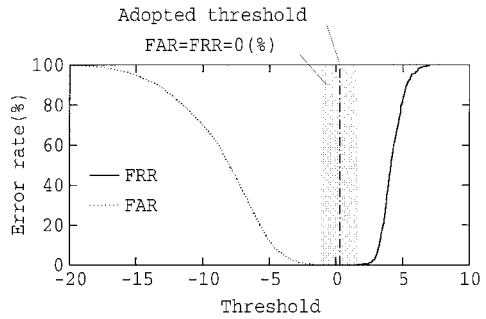


図 2　学習データに対する本人棄却率と詐称者受理率
Fig. 2　False rejection and acceptance rates as functions of the values of the decision threshold (training data).

4. 結果と考察

4.1 話者照合システムの基本性能

表 1 に，本実験で用いた話者照合システムの基本性能を示す．FRR は 6%以上であるのに対して FAR は 0%となっており，システムは FRR よりも FAR を低く抑えるようにしきい値が設定されている，つまり，本人が棄却されやすい代わりに詐称されにくいように設定されていることがわかる．なお，テストデータに対する等誤り率（EER）は混合数によらず 1%以下となっている．

4.2 合成音声による詐称

4.2.1 学習データが十分にある場合

まず，詐称対象話者の音声データが十分に入手可能な場合を仮定し，登録話者の発声 50 文章を学習データとして用いた場合について検討する．ただし，合成に用いるモデルの学習データの発声内容は，照合に用いるモデルの学習データとは異なることに注意しておく．

表 2 に合成音声に対する受理率を，表 3 に登録話者による発声との等誤り率を示す．この結果，合成音声に対する受理率は 77%以上，また EER についても

表 1　話者照合システムの基本性能
Table 1　Baseline performance of speaker verification systems.

	Verification		
	1-mix	2-mix	3-mix
FRR (%)	6.8	8.2	9.6
FAR (%)	0.0	0.0	0.0
EER (%)	1.0	1.0	0.8

表 2　学習データが十分にある場合の合成音声に対する受理率 (%)
Table 2　Acceptance rates (%) for synthetic speech with sufficient training data.

Synthesis		Verification		
state	data	1-mix	2-mix	3-mix
2	50	88.0	79.8	77.8
3	50	89.2	86.4	79.0
4	50	89.2	87.0	80.4

32%以上となっている．

話者照合システムについては，3 混合のモデルを用いた場合に最も合成音声の受理率が低くなっている．これは，照合に用いる HMM の学習データが多く，混合数が多い方がより精密に各話者のパラメータの分布をモデル化できるためであると考えられる．

4. 2. 2 少量の学習データを用いた場合

実際に詐称を行う場合，登録話者の多量の音声を入手することは困難であると考えられる．そこで，ここではより現実的に，対象とする話者の少量の音声データのみが入手できた場合を想定し，詐称の可能性について検討する．

表 3 学習データが十分にある場合の合成音声に対する等誤り率（%）

Table 3 Equal error rates (%) for synthetic speech with sufficient training data.

Synthesis		Verification		
state	data	1-mix	2-mix	3-mix
2	50	53.8	37.2	32.0
3	50	57.0	43.2	38.2
4	50	57.0	44.6	41.8

表 4 少量の学習データを用いた場合の合成音声に対する受理率（%）

Table 4 Acceptance rates (%) for synthetic speech with a small amount of training data.

Synthesis		Verification		
state	data	1-mix	2-mix	3-mix
2	1	74.0	66.2	63.6
	3	88.2	82.8	75.6
	5	88.2	85.0	78.8
3	1	75.0	70.2	66.6
	3	89.0	85.8	80.0
	5	89.2	86.8	84.6
4	1	76.5	74.5	78.0
	3	88.4	85.4	79.0
	5	89.0	86.6	83.2

表 5 少量の学習データを用いた場合の合成音声に対する等誤り率（%）

Table 5 Equal error rates (%) for synthetic speech with a small amount of training data.

Synthesis		Verification		
state	data	1-mix	2-mix	3-mix
2	1	46.8	31.4	26.8
	3	50.8	31.2	28.2
	5	53.0	37.7	32.2
3	1	47.2	34.4	30.2
	3	54.2	38.4	33.8
	5	56.0	42.4	38.2
4	1	53.7	38.0	33.9
	3	54.0	40.0	36.0
	5	57.2	44.4	41.2

学習データを 1，3，または 5 文章とした場合の合成音声に対する受理率及び等誤り率を表 4，表 5 に示す．わずか 1 文章で学習した場合でも，合成音声に対する受理率は 63%以上となり，5 文章で学習した場合には学習データが十分にある場合とほぼ同等の結果となった．また EER についても，音声合成システムで 3 状態以上のモデルを用いた場合には 30%以上となっている．

図 3 にしきい値に対する FRR 及び FAR の変化を，また図 4 に EER が (a) 最も高い話者と (b) 最も低い

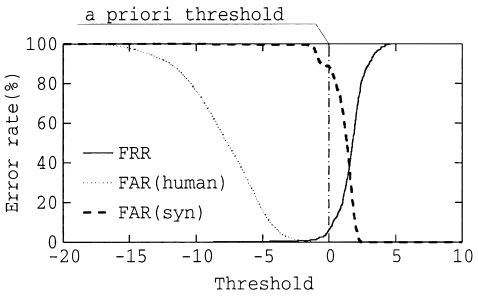
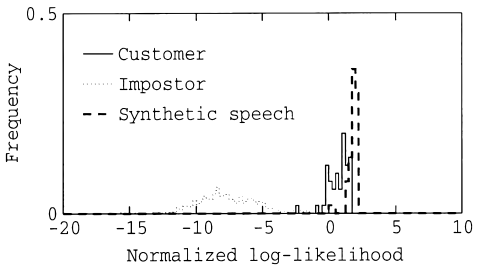
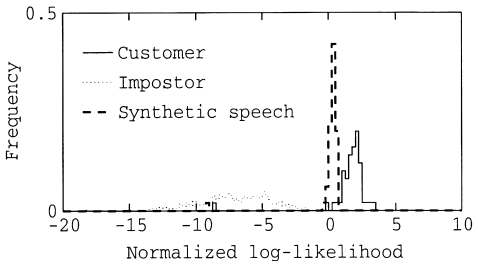


図 3 本人棄却率と詐称者及び合成音声の受理率

Fig. 3 False rejection and acceptance rates as functions of the values of the decision threshold (test data).



(a) The speaker of the highest EER.



(b) The speaker of the lowest EER.

図 4 話者別のゆう度分布

Fig. 4 Distributions of normalized log-likelihood for different speakers.

話者のテストデータと合成音声に対する正規化対数ゆう度の分布を示す．図 3，図 4 はともに話者照合システムには 3 混合のモデル，音声合成システムには 5 文章で学習した 3 状態のモデルを用いた場合である．図 3 中，実線は本人棄却率，細い点線は人間の詐称者受理率，太い点線は合成音声の受理率，1 点鎖線は事前に定められたしきい値を表している．また，図 4 中の実線は登録話者の音声，細い点線は詐称者の音声，太い点線は合成音声に対する正規化対数ゆう度の分布を表している．

図 3 より，自然発声に対する FAR を十分抑制するようにしきい値を調整した場合でも，合成音声に対しては FAR が高い値を示すことがわかる．これに対し，合成音声に対する FAR を十分抑制するようにしきい値を調整した場合には，自然発声に対する FRR が非常に大きくなり，十分な照合性能が得られないことがわかる．

また，図 3 で合成音声に対する FAR の曲線の傾きが自然発声に比べて急であること，及び図 4 より，合成音声に対する正規化対数ゆう度は自然発声に比べて狭い区間に分布していることがわかる．ただし，分布の位置は話者によるばらつきが大きく，話者によっては自然発声よりも高い位置に分布している．なお，話者別の EER は，最も高い話者では 86%，最も低い話者では 8%となっている．

4.2.3 動的特徴量を用いない場合

パラメータ生成時に動的特徴量を用いずに，静的特徴量のみを用いて合成した音声に対する受理率及び等誤り率を表 6，表 7 に示す．この場合，HMM の各状態の継続区間のスペクトルパラメータとしてそれぞれの状態のメルケプストラムの平均値を用いて合成することと等価である．平均値のみを用いた場合には，合成音声から求められた Δ，Δ^2 パラメータは実際の発声と大きく異なると考えられる．それにもかかわらず，受理率及び等誤り率は合成時に動的特徴量を用いた場合とほとんど同じ結果となった．

4.3 考　　察

今回の実験では，話者照合システムと音声合成システムで異なるスペクトルパラメータを用いているとはいえ，サンプリング周波数等，分析条件が一致している点も多く，必ずしも現実的な条件とはいえない．しかし，音声合成システムを 1 文章程度のわずかなデータで学習し，励振源として白色雑音を用いているなど，人間が聞けば明らかに合成音声であることがわかる場合でも受理率や等誤り率が高く，更に HMM の出力分布の平均のみを用いて合成した音声でも受理率，等誤り率が高いことがわかった．また，合成音声に対するゆう度の分布は自然発声に対するゆう度の分布と重なっており，話者によっては自然発声よりも高い位置に分布していることから，しきい値の変更のみでは合成音声を効率的に棄却することは困難である．

表 6　動的特徴量を用いない場合の合成音声に対する受理率（%）

Table 6　Acceptance rates (%) for synthetic speech without dynamic features.

Synthesis		Verification		
state	data	1-mix	2-mix	3-mix
2	3	88.4	84.2	75.4
3	3	88.6	85.8	80.4
4	3	87.6	84.7	79.6

表 7　動的特徴量を用いない場合の合成音声に対する等誤り率（%）

Table 7　Equal error rates (%) for synthetic speech without dynamic features.

Synthesis		Verification		
state	data	1-mix	2-mix	3-mix
2	3	46.8	31.6	29.0
3	3	46.6	35.0	31.0
4	3	48.4	38.2	32.2

これらの結果から，合成音声による詐称は十分可能であり，話者照合システムで合成音声を棄却するために何らかの対策をする必要があると考えられる．この際，前節の結果から，話者照合システムで動的特徴量を用いることは，合成音声を棄却するためには必ずしも有効ではないことが予想される．

なお，本実験では合成音声を再生せずに音声データをそのまま話者照合システムへの入力としているが，実際の話者照合システムに対して合成音声を入力するためには，スピーカで音声を再生する必要がある．この場合には再生時の出力系の特性も考慮する必要があるが，あらかじめ出力系の特性を計測してその逆特性をもつフィルタで補正することにより，出力系の特性による影響を抑えることができると考えられる．

5. むすび

テキスト指定型話者照合システムに対して HMM に基づく音声合成システムから生成された合成音声による詐称の可能性の検討を行った．音声合成システムをわずかなデータで学習し，励振源として白色雑音を用いているなど，人間が聞けば明らかに合成音声である

ことがわかる場合でも高い割合で受理されることを示し，話者照合システムでは，自然発声に対する照合性能を上げることばかりではなく，合成音声を棄却するための何らかの対策が必要であることを示した．

今回の実験では合成音声の励振源として白色雑音を用いていたため，ピッチ [18] や残差 [19] を考慮することにより，本実験で用いた合成音声の受理率を低下させることができると考えられる．しかし，ピッチや残差をモデル化できればこれらの情報をもった音声を合成することも可能であると考えられるため，照合時にこれらの情報を考慮することは問題の本質的な解決策にはならないと考えられる．したがって，今後，自然発声と合成音声の判別手法の検討など，合成音声を用いた詐称に対してロバストな話者照合システムについて検討を行う必要がある．また，話者による受理率の違いの検討，他の枠組みの話者照合システムや音声合成システムについての検討，実際の話者照合システムに対して詐称する際の音声の入出力系の特性による影響の検討なども今後の課題である．

謝辞　研究を進めるにあたり，実験に御協力を頂いた東京工業大学大学院修士課程一ツ松孝文氏（現株式会社デンソー）に感謝します．本研究の一部は文部省科学研究費補助金（課題番号 11750311）によった．

文　献

[1] S. Furui, “An overview of speaker recognition technology,” in Automatic Speech and Speaker Recognition, eds. C.-H. Lee, F.K. Soong, and K.K. Paliwal, Ch.2, Kluwer Academic Publisher, 1996.

[2] 松井知子, “HMM による話者認識,” 信学技報, SP95-111, Jan. 1996.

[3] 内部利明，黒岩眞吾，樋口宜男，“数字を用いた話者照合方式の検討,” 信学技報，SP98-68, Oct. 1998.

[4] A. Higgins, L. Bahler, and J. Porter, “Speaker verification using randomized phrase prompting,” Digital Signal Processing, vol.1, pp.89–106, 1991.

[5] A.E. Rosenberg, J. Delong, C.H. Lee, B.H. Juang, and F.K. Soong, “The use of cohort normalized scores for speaker verification,” Proc. ICSLP-92, pp.599–602, Oct. 1992.

[6] 松井知子，古井貞熙，“テキスト指定形話者認識のための事後確率に基づくゆう度正規化法,” 音講論集，vol.1, no.1-7-20, pp.639–640, Oct. 1993.

[7] D. Genoud and G. Chollet, “Speech pre-processing against intentional imposture in speaker recognition,” Proc. ICSLP-98, pp.105–108, Dec. 1998.

[8] B.L. Pellom and J.H.L. Hansen, “An experimental study of speaker verification sensitivity to computer voice-altered imposters,” Proc. ICASSP-99, pp.837–840, March 1999.

[9] ニック キャンベル，アラン ブラック，“CHATR: 自然音声波形接続型任意音声合成システム,” 信学技報，SP96-7, May 1996.

[10] 益子貴史，徳田恵一，小林隆夫，今井　聖，“動的特徴を用いた HMM に基づく音声合成,” 信学論（D-II），vol.J79-D-II, no.12, pp.2184–2190, Dec. 1996.

[11] T. Masuko, K. Tokuda, T. Kobayashi, and S. Imai, “Voice characteristics conversion for HMM-based speech synthesis system,” Proc. ICASSP97, pp.1611–1614, April 1997.

[12] 田村正統，益子貴史，徳田恵一，小林隆夫，“HMM 音声合成に基づく声質変換における話者適応手法の検討,” 音講論集，vol.1, no.2-P-13, pp.319–320, March 1998.

[13] 徳田恵一，小林隆夫，深田俊明，斎藤博徳，今井　聖，“メルケプストラムをパラメータとする音声のスペクトル推定,” 信学論（A），vol.J74-A, no.8, pp.1240–1248, Aug. 1991.

[14] 徳田恵一，益子貴史，小林隆夫，今井　聖，“動的特徴を用いた HMM からの音声パラメータ生成アルゴリズム,” 音響誌，vol.53, no.3, pp.192–200, March 1997.

[15] 今井　聖，住田一男，古市千枝子，“音声合成のためのメル対数スペクトル近似（MLSA）フィルタ,” 信学論（A），vol.J66-A, no.2, pp.122–129, Feb. 1983.

[16] T. Fukada, K. Tokuda, T. Kobayashi, and S. Imai, “An adaptive algorithm for mel-cepstral analysis of speech,” Proc. ICASSP92, pp.137–140, March 1992.

[17] 松井知子，西谷　隆，古井貞熙，“話者照合におけるモデルとしきい値の更新法,” 信学論（D-II），vol.J81-D-II, no.2, pp.268–276, Feb. 1998.

[18] T. Matsui and S. Furui, “Text-independent speaker recognition using vocal tract and pitch information,” Proc. ICSLP90, pp.137–140, 1990.

[19] K.P. Markov and S. Nakagawa, “Speaker recognition using LPC-residual and pitch,” J. Acoust. Soc. Jpn. (E), vol.20, no.4, pp.281–291, 1999.

（平成 12 年 2 月 25 日受付，6 月 30 日再受付）

益子　貴史　（正員）

平 5 東工大・工・情報卒．平 7 同大大学院博士前期課程了（知能科学専攻）．同年東工大精密工学研究所助手．現在東工大大学院総合理工学研究科物理情報システム創造専攻助手．音声分析・合成・認識，マルチモーダルインタフェースの研究に従事．日本音響学会，IEEE，ISCA 各会員．

徳田　恵一　（正員）

昭 59 名工大・工・電子卒．平 1 東工大大学院博士課程了．同年東工大電気電子工学科助手．平 8 名工大知能情報システム学科助教授．工博．音声分析・合成・符号化・認識，ディジタル信号処理，マルチモーダルインタフェースの研究に従事．日本音響学会，情報処理学会，人工知能学会，IEEE 各会員．

小林　隆夫　（正員）

昭 52 東工大・工・電気卒．昭 57 同大大学院博士課程了．同年東工大精密工学研究所助手．同助教授を経て現在東工大大学院総合理工学研究科物理情報システム創造専攻教授．工博．ディジタルフィルタ，音声分析・合成・符号化・認識，マルチモーダルインタフェースの研究に従事．日本音響学会，IEEE，ISCA 各会員．