論文

CELPパラメータを用いた話者照合方式

北九州市立大学国際環境工学部　正会員　山崎　恭
早稲田大学理工学部　近藤維資
〃　正会員　小松尚久

〈あらまし〉 ディジタル音声通信で使用される符号化された音声には，音韻性情報と個人性情報が保存されており，これらの情報を適切に処理することにより，符号化された音声情報のみを用いて音声認識あるいは話者認識を実現することが可能になると考えられる．そこで，本稿では携帯電話をはじめとするディジタル音声通信における音声符号化方式との親和性を考慮した新たな話者照合方式を提案する．提案方式では，移動通信システムやIPネットワークで利用されているCELP(Code Excited Linear Prediction：符号励振線形予測)符号化方式により符号化された音声情報を使用して話者照合を行うため，①移動通信システムに適用した場合，照合のために必要な機能の追加を抑えられ，重量やサイズに制限のある端末側での認証に有利である．②調音動作を反映するパラメータを使用することにより，話者の発話内容に依存しないテキスト独立型の話者照合が可能である，といった特徴を有する．実際の音声を用いたシミュレーション実験により提案方式の信頼性を評価した結果，提案方式の有効性が明らかとなった．

キーワード：話者照合，バイオメトリック認証，情報セキュリティ，CELP, LSP

〈**Summary**〉 The encoded speech for digital transmission systems contains semantic information and singular information. Therefore, a speech or speaker recognition system can be realized by using only the encoded speech. In this paper, we propose a speaker verification method based on a speech coding scheme in the digital transmission systems. The proposed method utilizes CELP (Code Excited Linear Prediction) parameters which are used in speech coding schemes for mobile communication systems or IP networks, and verifies a speaker only with the encoded speech. The merits of the proposed method are as follows; ① Speaker verification is easily realized in the current mobile terminals or network systems by adding a little function. ② Since CELP parameters contain a speaker's characteristics of articulation, text-independent speaker verification is realized. The reliability of the proposed method is discussed with some simulation results.

Key words: speaker verification, biometric person authentication, information security, CELP, LSP

1. まえがき

近年，インターネットにみられる情報通信ネットワークの急速な進展に伴い，あらゆる場所にネットワークが存在し，いつでもどこでもユーザあるいはユーザが持つ情報機器がそのネットワークを利用できるというコンセプトをもつ「ユビキタスネットワーク」が注目されている[1]．ユビキタスネットワークは，ブロードバンドネットワーク，モバイル通信インフラ，放送型通信など様々な構成要素からなり，今後，我々の社会生活に多大な影響をおよぼすことが予想される．このようなユビキタスネットワーク環境が整備され，本格的に普及する時代が到来すれば，ユーザは様々な情報端末を使用して多様なネットワークサービスを享受することが可能となる．特

"A Speaker Verification Method Using CELP Parameters" by Yasushi YAMAZAKI (Member) (The University of Kitakyushu) Tadashi KONDO and Naohisa KOMATSU (Member) (Waseda University).

に，我が国の通信事情として，平成14年3月の時点における携帯電話の加入数が約7,000万台に達し，そのうち75%以上が携帯電話を利用したインターネットサービスに加入している現状[2]をかんがみると，今後進展するユビキタスネットワーク環境においてユーザがネットワークサービスを利用する場合，携帯電話が最も有力な情報端末の一つになると考えられる．

一方，ネットワークサービスの利用機会の増加に伴い，情報セキュリティ，プライバシー保護の観点から，ユーザの正当性を確認する個人認証の必要性が今後ますます高まってくるものと考えられる．従来，簡便で安価な個人認証手段として，カードやパスワードなどユーザの所有物や知識に基づく個人認証が行われてきたが，現在，従来手法と比較して盗難，紛失，忘失などの危険性が少ないユーザの身体的特徴や特性，すなわちバイオメトリクスに基づく個人認証(以下，バイオメトリック認証)[3]が注目されている．今後，携帯電話から様々なネットワークサービスを利用する機会が増加すれば，携帯電話のユーザが正当なユーザであるか否かを確認する個人認証が必要となるケースが増加するものと考えられる．携帯電話を利用して個人認証を行う場合，現在は主としてパスワードが用いられているが，最近では，端末に搭載した指紋センサやカメラを用いて指紋認証や顔認証を行うバイオメトリック認証に基づく個人認証も試みられている．しかしながら，携帯電話の主要な機能が音声情報の伝送であること，また，音声は我々にとって日常的なコミュニケーションの手段であるとともに，本人を特定する手段としても使用できることを考慮すると，音声は，携帯電話を利用した個人認証に最も適したバイオメトリクスの一つであると考えられる．このように，音声には音声認識に有効な音韻性情報(何を話したかという情報)と話者認識に有効な個人性情報(誰が話したかという情報)が併存するという著しい特徴がある．ところで，携帯電話をはじめとするディジタル音声通信では，音声は符号化されて伝送されるが，符号化された音声にも，音韻性情報と個人性情報が保存されている．従って，これらの情報を適切に処理することにより，符号化された音声情報のみを用いて音声認識あるいは話者認識を実現することが可能になると考えられる．しかしながら，従来の話者認識に関する研究[3),4)]では，符号化された音声情報を対象とした研究は少なく[5]，まだ検討の余地がある．

そこで，本稿では携帯電話をはじめとするディジタル音声通信における音声符号化方式との親和性を考慮した新たな話者照合方式を提案する．提案方式では，移動通信システムやIPネットワークで利用されているCELP (Code Excited Linear Prediction：符号励振線形予測)符号化方式[6]により符号化された音声情報を使用して話者照合を行う．提案する話者照合方式は，以下の特徴を有する．

- 符号化された音声情報のみを用いて話者照合を行うため，端末側，ネットワーク側(センタ側)のいずれでも話者照合を行うことができる．
- 移動通信システムに適用した場合，話者照合のために必要となる機能の追加を抑えることが可能であり，重量やサイズに制限のある端末側での認証に有利である．
- 調音動作を反映するパラメータを使用することにより，話者の発話内容に依存しないテキスト独立型の話者照合が可能である．

2. 話者照合手順

提案する話者照合方式の照合手順について説明する．提案方式は，図1に示すように，CELP符号化方式により符号化を行うCELP符号化ブロック(CELP Coding Block)，符号化された音声を用いて話者照合を行う話者照合ブロック(Speaker Verification Block)により構成される．例えば，現行の携帯電話端末に話者照合機能を付加する場合，CELP符号化ブロックは，端末に実装されているコーデックを使用することにより実現できるため，新たに話者照合ブロックのみを追加すればよい．更に，話者照合ブロックは，個人性情報を事前に登録する登録プロセス(Enrollment Process)，話者照合を実行する照合プロセス(Verification Process)により構成される．

以下，各ブロックの概要について説明する．

2.1 CELP符号化ブロック

CELP符号化ブロックでは，入力音声をCELP符号化方式により符号化し，符号化パラメータを抽出する．CELPは，線形予測(LPC)に基づく合成による分析手法を用いて，励振信号をベクトル量子化する音声符号化方式である．現在，携帯電話の音声符号化方式をはじめとするほとんどの移動体用高能率音声符号化方式において，CELPの枠組みが採用されている．CELPでは，数多くの符号ベクトルごとに線形予測合成フィルタを通して音声を合成し，入力と聴覚的に最も近くなるベクトルを選択してその番号を伝送する．つまり，符号器自体に復号器(合成器)を内蔵し，閉ループを構成して伝送パラメータを決定する[6]．提案方式では，CELP符号化方式の一種で，ITU/TでG.729として標準化されているCS-ACELP(Conjugate Structure Algebraic CELP：共役構造代数CELP)[7]に着目した(図2参照)．CS-

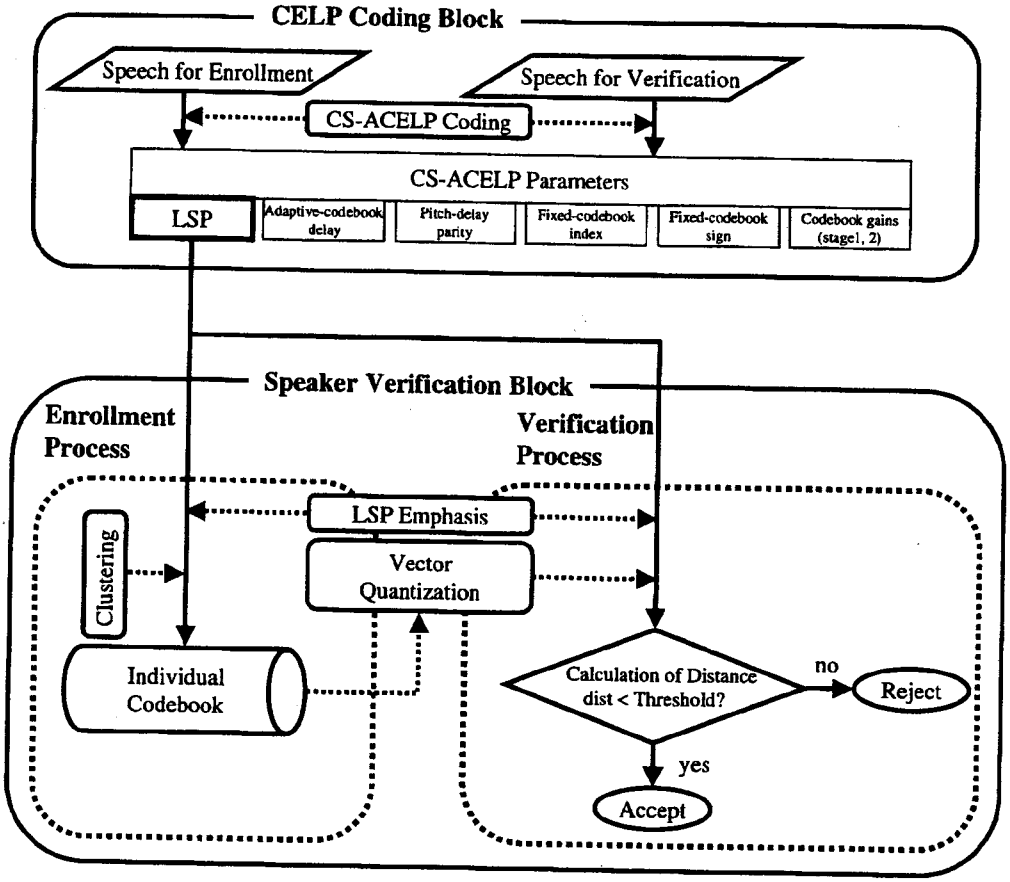

図 1　話者照合方式の概要
Fig. 1　Outline of the proposed speaker verification

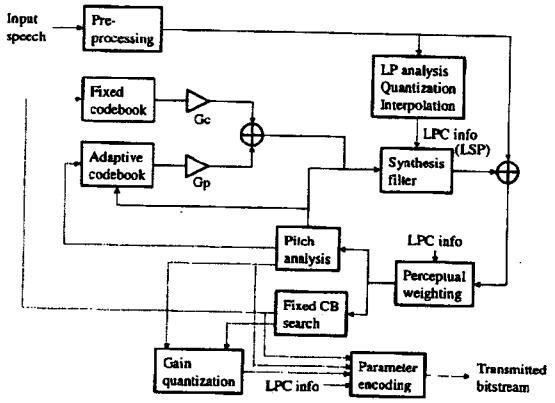

図 2　CS-ACELP 符号化の原理
Fig. 2　Principle of CS-ACELP coding

ACELP は，8 kbit/s の符号化方式でありながら，日本の PHS で採用されている 32 kbit/s の ADPCM (Adaptive Differential PCM：適応差分 PCM) と同等の音声品質をもち，日本のディジタル携帯電話の改良フルレート標準の一方式として採用されているほか，VoIP (Voice over IP) をはじめとする多くのアプリケーションへの適用が期待されている主要な音声符号化方式の一つである．更に，提案方式では，CS-ACELP の符号化パラメータのうち，LSP (Line Spectrum Pair：線スペクトル対) に着目した．LSP は，言語音を発声するために声道の形状を調整する調音に対応するパラメータの一つであり，音声のスペクトル包絡 (スペクトルの概形) との対応が良い．更に，スペクトル包絡は，声の質を反映した物理特徴であり，話者認識に有効な特徴であることが従来の研究により明らかにされている．従って，ここで LSP を使用することにより，話者照合に有効な話者の個人性を抽出できることが期待される．

2.2　話者照合ブロック

話者照合ブロックは，登録プロセスと照合プロセスにより構成される．登録プロセスでは，音韻性情報の影響を軽減するために，十分に長い音声データから登録用の LSP (以下，登録用 LSP) を抽出し，登録用 LSP のクラスタリング (Clustering) により作成されるコードブック (以下，特徴コードブック：Individual Codebook) を話者の個人性情報とする．一方，照合プロセスでは，照合用の LSP (以下，照合用 LSP) を特徴コードブックによ

りベクトル量子化(Vector Quantization)したときの量子化誤差に基づき，テキスト独立型の話者照合を行う．

以下，登録プロセスおよび照合プロセスの概要を述べる．

2.2.1　**登録プロセス**

話者の個人性を強調するため，各話者を対象とし，抽出された登録用LSPの特徴を強調する前処理(LSP強調：LSP Emphasis)を施す．提案方式では，多数の話者のLSPについて，次数ごとにLSPの値を算出し，全登録話者に関するi次のLSPの値の平均値をc-LSP(i)と定義する．一方，特定の話者のLSPについて，次数ごとにLSPの値を算出し，i次のLSPの値の平均値をp-LSP(i)と定義する．ここで，c-LSP(i)とp-LSP(i)の差をLSP偏差と呼ぶ．図3にLSP偏差の例を示す．前処理では，式(1)に示すように，n次のp-LSP(n)の値が，($n-1$)次のp-LSP($n-1$)もしくは($n+1$)次のp-LSP($n+1$)の値を超えない範囲において，LSP偏差を一律にk倍することにより，p-LSP(n)に現れる個人性を強調する．ここで，式(1)におけるp-LSP(n)'は前処理後のp-LSP(n)の値を表す．

$$\text{p-LSP}(n)' = \text{c-LSP}(n) + k(\text{p-LSP}(n) - \text{c-LSP}(n))$$
$$\text{p-LSP}(n-1) < \text{p-LSP}(n)' < \text{p-LSP}(n+1) \quad (1)$$

前処理を施すことにより，LSPというパラメータを通してみたときの平均的な話者の特徴とある特定の話者の特徴との差異が強調され，ひいては特徴空間における話者間の距離が増大し，話者照合の精度が向上することが期待される．

次に，前処理を施した登録用LSPに対してクラスタリングアルゴリズムを適用し，生成された各クラスタ(以後，カテゴリ)の重心位置に関する情報を集めて特徴コードブックを作成する．本稿では，クラスタリングアルゴリズムとして，LBG+splittingアルゴリズム[8]を使用する．特徴コードブックは，スマートカードや端末，あるいはネットワーク上のデータベースに保存しておくことが可能である．例えば，話者照合機能を付加した携帯端末に，特徴コードブックを登録したスマートカードを挿入し，話者照合の結果，正当なユーザであることが判明した場合にのみ携帯端末の使用を許可するといった使用方法が考えられる．

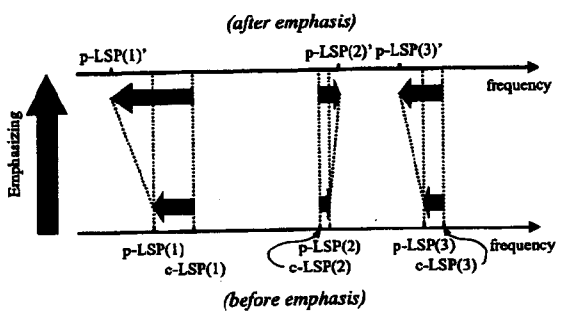

図3　LSP偏差と強調

Fig. 3　LSP deviation and its emphasis

2.2.2　**照合プロセス**

登録プロセスと同様に前処理を施した照合用LSPを特徴コードブックでベクトル量子化し，量子化時の量子化誤差の値とあらかじめ設定したしきい値とを比較して話者を照合する．なお，提案方式では，量子化誤差を算出する際の尺度として，ユークリッド距離を使用する．

3.　信頼性評価実験

3.1　実験の条件

研究用ATR日本語音声データベース[9]を使用して，提案方式の信頼性を評価するシミュレーション実験を行った．シミュレーションの条件を表1に示す．同表の値は，安定した照合特性を得るために必要な値を，予備実験の結果に基づき決定したものである．なお，以下の実験では，登録用音声と照合用音声を置換する交差検証法(cross validation)[10]を適用して信頼性を評価した．

3.2　実験結果

3.2.1　LSP強調の効果

表2は，LSP強調に使用するパラメータkの値を変えたときのFRR(False Rejection Rate：本人拒否率)

表1　シミュレーションの条件

Table 1　Simulation condition

Celp Coding Block	[音源] 研究用ATR日本語音声データベース (連続音声，音素バランス文) [話者数] c-LSP作成用：20名(男女各10名) 話者照合実験用：20名(男女各10名) ※c-LSP作成用の話者と話者照合実験用の話者は異なる [音声長] c-LSP作成用：90秒 話者照合実験用(登録用音声)： 30秒，60秒，90秒 話者照合実験用(照合用音声)： 6秒，12秒，18秒 ※登録用音声と照合用音声の内容は異なる [サンプリング周波数] 8 kHz [カットオフ周波数] 3.1 kHz
Speaker Verification Block	[LSP偏差の強調] k=1，20 [クラスタリングアルゴリズム] LBG+splitting [特徴コードブック] 16カテゴリ

表 2　LSP 強調の効果
Table 2　Effectiveness of LSP emphasis

k	1	20
EER(%)	6.6	3.1

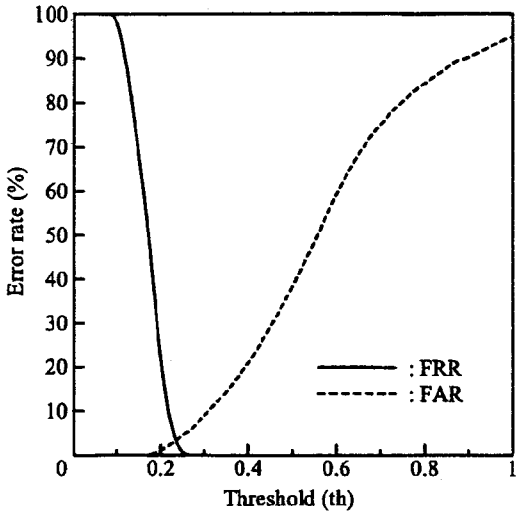

図 4　しきい値と誤り率の関係
Fig. 4　Verification results

およびFAR(False Acceptance Rate：他人受け入れ率)の値が一致するEER(Equal Error Rate：等誤り率)の値を示したものである．本実験では，登録時の音声長を90秒，照合時の音声長を6秒とした．本実験において，k=1はLSP強調を行わない場合に相当し，k=20はLSP強調を行う際，LSP強調の効果が最大となるときのkの値を実験結果から求めたものである．表2の結果より，LSP強調を行わないk=1の場合よりも，LSP強調を行うk=20の場合の方が，EERの値が低くなることがわかる．これらの結果は，LSP強調を行うことにより各話者の個人性が強調され，特徴空間上で話者間の距離が増大することにより照合精度が向上することを示していると考えられる．

3.2.2　話者照合実験結果

図4は，照合時のしきい値と二種類の誤り率FRRおよびFARの関係を示したものである．ここで，FRRとFARの値は，すべての話者についての平均値を表している．本実験では，前述と同様に，登録時の音声長を90秒，照合時の音声長を6秒とした．図4より，FRRとFARの値が一致する点における誤り率EERは3.1%である．また，表3は，異なる音声長に対するEERの値を示したものである．表3より，登録プロセスおよび照合プロセスの双方において，音声長の増大に伴い，EERの値が減少する傾向のあることが確認される．一般に，安定した照合特性を得るには，十分な長さの登録用音声と照合用音声が必要となる．本実験では，登録時の音声長が90秒程度あれば，照合時の音声長を登録時の音声長の15分の1程度に設定しても3%程度の誤り率が達成されている．照合時の音声長は短い方がより実用的であることを考慮すると，本実験結果はこの要求を満足できる可能性があると考えられる．以上の実験結果より，提案方式により，符号化された音声に基づくテキスト独立型の話者照合方式が実現できる可能性のあることが明らかとなった．

表 3　音声長と等誤り率の関係
Table 3　EER for different speech lengths

EER(%)		照合時の音声長		
		6秒	12秒	18秒
登録時の音声長	30秒	3.9	2.8	2.5
	60秒	3.2	2.5	1.9
	90秒	3.1	2.4	1.8

4. むすび

本稿では，CELPパラメータを用いたテキスト独立型の話者照合方式を提案し，シミュレーション実験に基づき提案方式の信頼性を評価した．実験の結果，提案方式により，符号化された音声のみを用いてテキスト独立型の話者照合方式を実現できる可能性が確認された．提案方式は，テキスト独立型の話者照合方式であるため，本方式を携帯電話の利用者確認手段として用いた場合，通話開始時の本人確認のみならず，通話中に本人でないことが確認されれば通話を遮断するといった応用機能も実現可能である．今後，LSPとLSP以外の符号化パラメータを併用したときの効果や雑音環境下における提案方式の信頼性を評価する必要がある．更に，照合時の適切なしきい値の設定方法や話者数が増加したときの性能，また，録音音声を用いたなりすましに対する対策などが今後の検討課題として残されている．

参考文献

1) 塚本昌彦，西尾章治郎，宮原秀夫："ユビキタスネットワーク"，信学誌，Vol. 86, No. 3, pp. 180-185 (2003).
2) 総務省(編)："平成14年版　情報通信白書"，総務省，(2002).
3) A. Jain, R. Bolle, and S. Pankanti: "Biometrics—Personal Identification in Networked Society", Kluwer Academic, (1999).
4) S. Furui: "Recent Advances in Speaker Recognition", in Audio-and Video-based Biometric Person Authentication (AVBPA '97), pp. 237-252, Springer, (1997).
5) T. Mogaki and N. Komatsu: "Text-indicated speaker verification method using PSI-CELP parameters", Proc. of

SPIE, Vol. 3657, pp. 184-193 (1999).
6) 守谷健弘："音声符号化技術", 信学誌, Vol. 84, No. 11, pp. 836-842 (2001).
7) ITU-T: "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)", ITU-T Recommendation G. 729, (1996).
8) Y. Linde, A. Buzo, and R. M. Gray: "An algorithm for vector quantizer design", IEEE Trans. Commun., COM-28, 1, pp. 84-95 (1980).
9) 桑原尚夫, 匂坂芳典, 武田一哉, 阿部匡伸："研究用ATR日本語音声データベースの作成(別冊I連続音声テキスト)", TR-I-0086, ATR自動翻訳電話研究所, (1989).
10) R. O. Duda, P. E. Hart, and D. G. Stork (原著), 尾上守夫(監訳)："パターン識別", 新技術コミュニケーションズ, (2001).

(2003年5月1日受付)

山崎 恭 (正会員)

平5, 早大・理工・電子通信卒. 平10, 同大大学院博士後期課程了. 博士(工学). 平9～平11, 日本学術振興会特別研究員. 平11, 早大・理工・助手. 平13, 北九州市立大学助教授. 現在に至る. 主に, 情報セキュリティとその応用に関する研究に従事. 電子情報通信学会, 情報理論とその応用学会, IEEE Computer Society, ACM各会員.

近藤維資

平14, 早大・理工・電子・情報通信卒. 現在, 同大大学院修士課程在学中. 話者認識に関する研究に従事.

小松尚久 (正会員)

昭54, 早大・理工・電子通信卒. 昭56, 同大大学院修士課程了. 同年, 日本電信電話公社(現, NTT)入社. 昭62, 早大・理工・助手. 平1, 同講師. 平3, 同助教授. 平8, 同教授. 主に, 情報通信システムにおけるセキュリティ, ヒューマン/ネットワークインタフェースの研究に従事. 工学博士. 電子情報通信学会会員. 本学会編集委員長, 将来型テレマティクス検討会(FTS)委員長.