

論文 / 著書情報
Article / Book Information

論題(和文)	電話音声によるオンライン話者照合システム
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1981年春季講演論文集, , , pp. 569-570
Citation(English)	, , , pp. 569-570
発行日 / Pub. date	1981, 5

3-1-25 電話音声によるオンライン話者照合システム*

古井 貞熙 (日本電信電話公社 武蔵野電気通信研究所)

I まえがき 単語音声の統計的特徴を用いた話者認識方式の改良形の原理とその有効性についてはすでに報告した。⁽¹⁾⁽²⁾ ここでは、この方式にもとづき、実際のC-400交換機及び内線を通じた電話音声を対象とするオンライン自動話者照合システムを作成し、多数話者が発声した音声による話者照合実験を行った結果について報告する。

II 特徴抽出と認識方法 特徴抽出と認識方法のブロック図は図1に示す通りで、認識に用いている特徴は、単語音声の有声音域の基本周波数とLPCケプストラムの平均値、標準偏差、相関行列およびフーリエ展開係数である。実際の電話音声に適用するために、音声区間の切り出しとピッチ抽出に特別な考慮を払っている。オンラインシステムのハードウェア構成は図2に示す通りで、各話者の識別番号はダイヤルによって入力され、種々の演算処理は、アレイプロセッサを内蔵したミニコンによって行っている。相関分析およびピッチ抽出までを実時間でを行い、音声区間検出、統計的特徴の抽出、認識のための距離計算にそれぞれ約1.5秒、4.5秒、3秒を要している。表示とオーバヘッドの3秒を加えて、システムの応答時間は約12秒であるが、ハードウェア化を行えば、容易に短縮可能である。

ピッチ抽出は、まずフレーム毎に自己相関関数のピーク位置を検出し、発声終了後にその時系列のヒストグラムをもとめ、これにもとづいて真のピッチの存在範囲を推定し、異常値の修正を行っている。

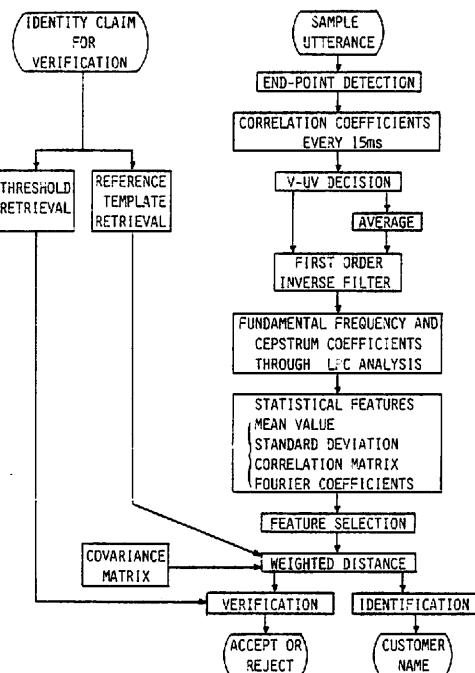


Fig.1. System-operation block diagram.

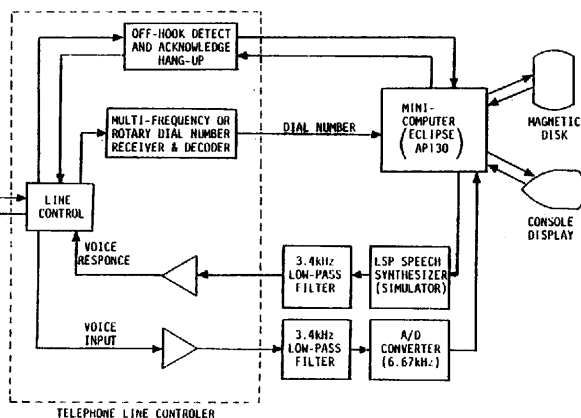


Fig.2. Experimental setup block diagram.

* An on-line speaker verification system for telephone speech.
By Sadaaki FURUI (Musashino Electrical Communication Laboratory, N.T.T.)

III 認識実験の方法 実験に

は男性29名と女性5名が発声した /baŋgo:/, /namae:/, /ko:geN/, /bakuoN/ の4種類の単語音声を用い、4単語の特徴を組合せて話者照合の判定を行った。まず初めに各話者毎に各単語を2回発声させて標準パターンを作成し、続けて最初のテストサンプルを発声してもらった。以後は各話者に、日を変えて週に2~3回程度手近な電話機（いずれも炭素送話器）からシステムを呼び出して、実験を行ってもらった。照合の結果棄却された場合、S/N比が異常に低い場合、発声間違いの場合等には、再発声を依頼した。発声者への各指示は、LSP合成音を用いた音声応答によって自動的に行った。

標準パターンと照合の判定のしきい値は、本人の音声を受理されたときに自動的に過去の履歴に従って更新されるようにした。異なる話者（詐称者）の音声を用いた実験は、オンライン実験で入力された音声を異なる話者の標準パターンと比較することによって行った。

IV 実験結果 図3に、ピッチ修正アルゴリズムの有効性を示す例を示した。左に男性話者の /baŋgo:/, 右に女性話者の /bakuoN/ の結果を、ヒストグラムとあわせて示した。実線(一)が原ピッチパターン、---が自動修正後のパターンである。図4に、話者照合の2種の誤り率(本人の音声を棄却する誤り False Rejection と他人の音声を受理する誤り False Acceptance)の時間的変化(発声回数による変化)を示す。FRはほぼ9回目の発声以降安定しており(約3%)、FAも20回目前後から安定している(約2%)。なおひどい風邪の

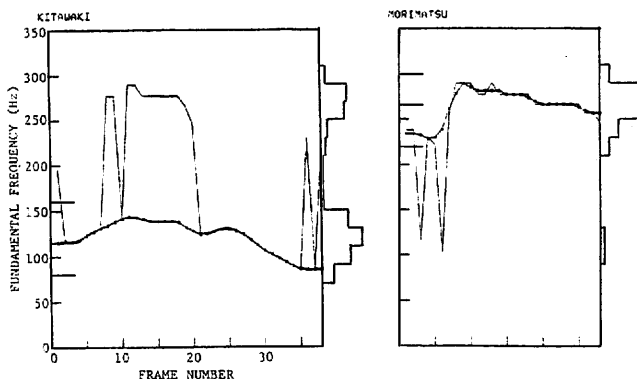


Fig.3. Original and corrected pitch contours.

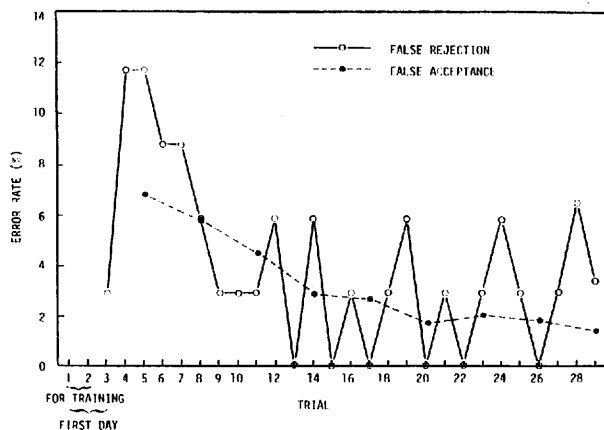


Fig.4. Speaker verification error rates vs. trial number.

ため本人の声が2回以上続けて棄却された場合が、2人について1回ずつあり、これは図4に含んでいない。その他の場合は、本人の声が2回続けて棄却されたことはほとんどない。

また、計算機室のように雑音レベルの高い場所(65~70dB(A))等で発声した場合も、特に問題はなかった。9回目の発声以降のFRの回数は、0回:20名, 1回:10名, 2回:2名, 3回:2名で、FRの無い人が最も多い。

V むすび 電話音声を対象としたオンライン話者照合システムを作成し、実験的に有効性を確認した。現在、誤認識の原因の分析等を進めている。

【謝辞】日頃御指導頂く当所群聊基礎研究部長、坂倉オキ研究室長、ならびに小池前室長に感謝する。

【文献】1) 古井: 音学秋季講義 1-1-20 (昭55-10)

2) 古井: 昭55 電通大 33-6 (昭55-10)