

論文 / 著書情報
Article / Book Information

論題(和文)	統計的特徴と動的特徴による話者認識の比較
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1980年春季講演論文集, , , pp. 543-544
Citation(English)	, , , pp. 543-544
発行日 / Pub. date	1980, 5

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 話者認識の方法は一般に、音声スペクトルの時系列をそのまま用いるものと、その統計量を用いるものにわけることができる。両者にはそれぞれ長所・短所があるが、ここでは、発声内容を限定した話者認識の場合について、両者の特徴量の有効性を比較した。

II 認識方法 図1にシステム全体のブロック図を示す。実験には男性9名が7年間の長期間にわたって繰返し発声した /baŋgo:/ と /namae/ の二種類の単語音声を用いる。音声波は 8kHz, 11bit で A/D 変換し、10ms 毎に 32ms の区間をハミング窓を乗じて切り出す。各区間毎に有聲/無聲の判定を行い、有聲音区間ごみの相関係数の時系列を得る。この相関係数を単語の全区間にわたって平均化し、この値を用いて二次の臨界制動系の逆フィルタを構成する。元の相関係数に畳み込みを行って、この逆フィルタによるスペクトル等化処理を行い、このうち基本周波数と12次までの対数断面積比(LAR)の時系列に変換する。【統計的特徴】: 単語の全区間における各パラメータの平均値、標準偏差および相互相関係数(後二者を相関行列成分とよぶ)を抽出し、これらの成分のうち、話者の分離に特に有効な33成分を選択する。このうち13成分は平均値、20成分は相関行列成分である。選択する成分は、学習サンプルを用いた分散分析によって決めておく。入力音声と標準パターンの特徴ベクトルの重みつき距離にもとづいて認識の判定を行う。⁽¹⁾⁽²⁾ 【動的特徴】: 基本周波数とLARの時系列を用いて、標準パターンとの時間正規化マッチングを行い、最後に対応づけられたときの全区間にわたる平均距離によって認識の判定を行う。このとき対応づけのフレームごみの距離には、統計的特徴の場合と同様に、各成分の変動量にもとづいた重みつき距離を用いる。⁽³⁾

III 実験結果 音声スペクトルの長期間にわたる変動が認識精度に与える影響を調べるため、学習サンプルの収集期間を10日間(短期間学習)と10か月間(長期間学習)の二種類に変え、学習サンプルから入力音声までの時期差を2~3日から5年まで変えて、話者9名による話者識別と話者照合の実験を行った。話者照合の判定の閾値は、事後的に最適値に設定した。

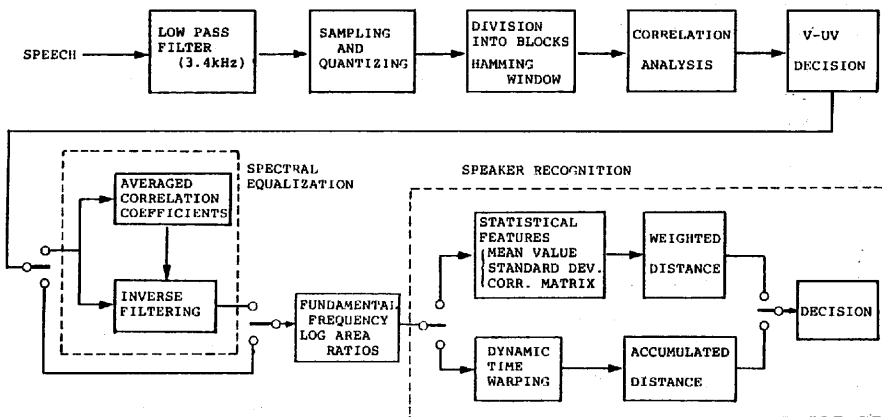


Fig.1. Block diagram of total system including feature extraction and speaker recognition.

*Comparison of Speaker Recognition Systems Using Statistical Features and Dynamic Features.
By Sudaoki Furui (Musashino Electrical Communication Laboratory, NTT)

長期間学習で時期差が3か月のときの二種の特徴による認識結果を図2に示す。統計的
特徴の一部の平均値ベクトルを、動的特徴と組合せて用いる場合、二種の単語を組合せて用い
る場合についても実験を行った。特徴量と組合せる場合は、それぞれの特徴による距離の和に
よって判定を行った。この結果
から、二種の特徴量の有効性
は同程度と見せ、統計的
特徴を用いる方が、動的特徴
の1/10程度の計算量であるので
有利であると言える。両者の特
徴や二種の単語を組合せると、
誤り率が約1/2に低下する。

図3に、時期差を10か月まで
変えたときの話者照合の結果を
示す。短期間学習では、時期
差とともに認識率が大きく低
下し、特に統計的特徴におい
て顕著であるが、時期差が10
か月になると、両者の特徴の差は小さくなる。

図4に、時期差を5年まで変えたときの、二
単語の統計的特徴による認識結果を示す。
短期間学習の場合にスペクトル等化処理の有
効性が顕著に見られる。長期間学習を行って
スペクトル等化処理を施せば、長期間にわたって
比較的安定した認識精度を得ることができる。

[謝辞] 日頃御指導頂く当所野田基礎幹長、小地寛長、
ならびに本研究に御指導頂いた富田東大教授に感謝す。
[文献] の古平・森倉：『信号処理』55-A, 11, p. 717 (1978)
② 古平：『信号処理』57-A, 12, p. 880 (1979) ③ S. Furui:
J. Acoust. Soc. Amer., 64, Suppl. 1, p. S183, (1978)

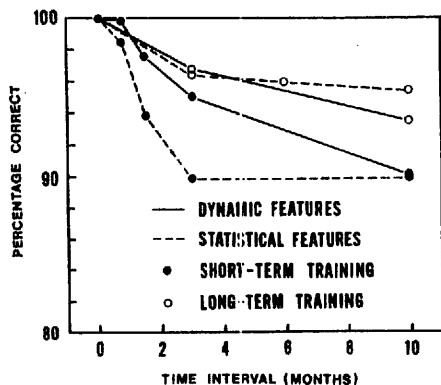


Fig.3. Verification accuracy as a function of interval between training and input utterances.

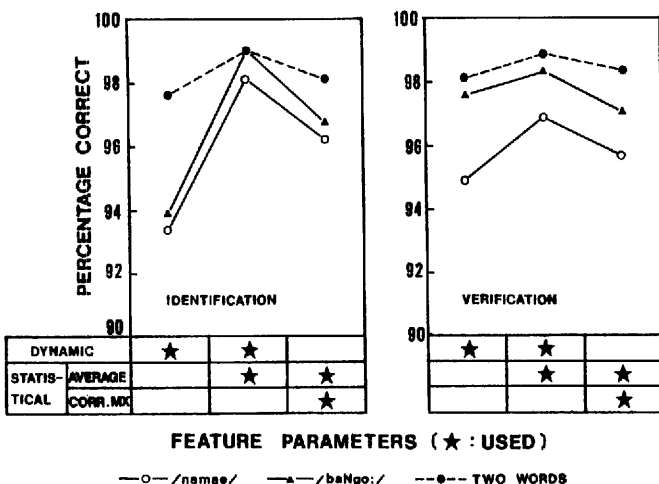


Fig.2. Comparison between accuracies by statistical and/or dynamic features. (Long-term training, three months interval)

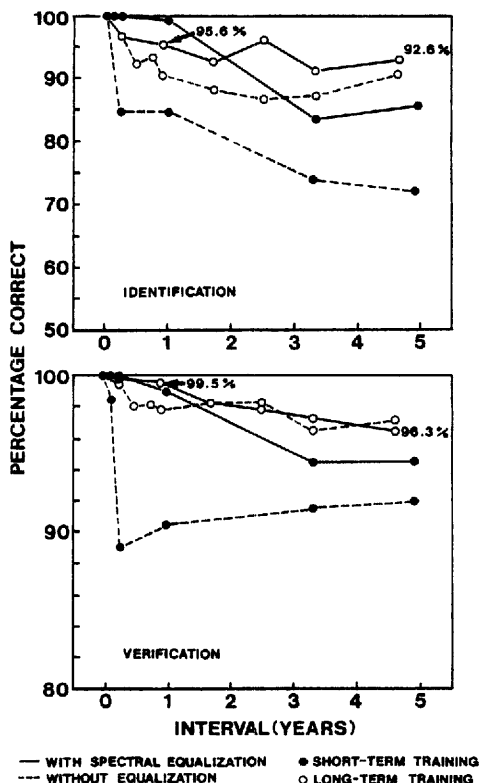


Fig.4. Effectiveness of spectral equalization procedure on recognition by statistical features. (Decision by two words)