

論文 / 著書情報
Article / Book Information

論題(和文)	単語音声認識における個人差の能率的学習法
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1977年講演論文集, , , pp. 515-516
Citation(English)	, , , pp. 515-516
発行日 / Pub. date	1977, 10

3—1—19 単語音声認識における個人差の能率的学習法*

古井 貞昭 (日本電信電話公社 武蔵野電気通信研究所)

I まえがき 不特定多数の話者が発声した多数の語彙を対象とした単語音声認識において高い認識率を得るためには、発声者が変わる毎に少数の音声サンプルを発声させ、それを用いて個人差の学習を行うのが最も現実的であると考えられる。その方法としては、音韻単位の標準パターンを用いて、学習サンプルからそれらの音韻の全部あるいは一部を学習し、不足分があればそれを推定する方法⁽¹⁾、学習サンプルとして5母音と撥音のみを発声させ、それらにもとづいて平均的な標準パターンを話者に適応化する方法⁽²⁾、母音学習サンプルから個人の特性を声道長と声帯波スペクトルの概形として推定し、標準パターンを話者に適応化する方法⁽³⁾などが提案されているが、まだ必ずしも満足すべき結果は得られていない。

II 学習方法 少数の学習サンプルから標準パターンを作成する際(“部分単語学習”)の問題点としては、①不足する音韻のスペクトルの推定、②学習した音韻のスペクトルに対するcontextの影響の除去、の2点がある。ここでは、この両者を解決する次のような方法を検討した。

標準パターンは音韻単位とし、文献⁽⁴⁾の認識方法を用いる。図1に示すように、学習サンプルから、それに含まれる各音韻の平均スペ

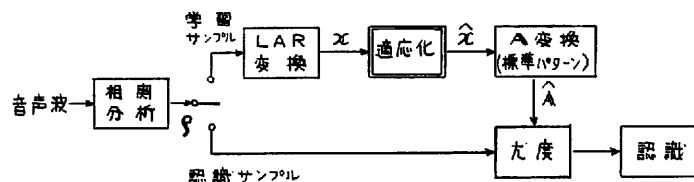


図1. 標準パターンの能率的学習と認識のブロック図

クトルを相関係数によって抽出し、対数断面積比(LAR)に変換する。求まった各音韻のLAR値から、全単語を学習に用いた際(“全単語学習”)に得られる各音韻の標準パターンを推定し、その値に変換する(“適応化”)。LARの領域で処理を行うのは、変換を行ってもパラメータが不安定になることがなく、スペクトル感度ともよく対応しており、個人性情報を表現するパラメータとしてLARが有効であるためである。適応化されたLARは最尤スペクトルパラメータ $\hat{A}$ に変換し、標準パターンとして蓄える。認識の際には、認識サンプルの相関係数と $\hat{A}$ から尤度をもとめ、判定に用いる。

学習サンプルから抽出したN次元LARベクトル x_{jk} (j:音韻, k:話者)から、適応化した $\hat{x}_{ik}$ を得るには、次の式を用いる。

$$\hat{x}_{ik} = \left(r / \sum_{j \in X_L} w_{ij} \right) \sum_{j \in X_L} w_{ij} \Phi_{ij} x_{jk} + (1-r) y_{ik}, \quad i \in X \quad (1)$$

ただし、X: 標準パターンの音韻の集合($a_1, a_2, i_1, i_2, u_1, u_2, e, o, n, N, g, r$ の12種類)

X_L : 学習サンプルから抽出される音韻の集合($X_L \subset X$)

Φ_{ij} : 抽出された音韻jから、全単語学習のときの音韻iのLARを推定する $N \times N$ 行列。話者k以外のサンプルを用いた重回帰分析によってもとめる。

* Efficient Learning Method of Talker Differences in Spoken Word Recognition.
by Sadaaki FURUI (Musashino ECL, NTT)

$$w_{ij} = (1/N) \sum_{l=1}^N R_{ijl}^2, \quad 0 \leq r \leq 1 : \text{重み}$$

R_{ijl} : 音韻 l の成分 i を音韻 j から推定するときの重相関係数

$$y_{ik} = \begin{cases} x_{ik} : i \in X_L \\ x_{ik} : i \notin X_L, i' \in X_L \\ \quad i' \text{ は } i \text{ とペアの音韻 (} a_1-a_2 \text{ など)} \\ x_{ik} : i, i' \notin X_L \end{cases}$$

x_{ik} : 全単語学習のときの話者 k 以外の平均値

Ⅲ 認識実験 男性・女性各30名が発声した10数字音声の認識実験を行った。各話者の学習サンプルとして、10数字のうちの一部(0, 2, 3, 5, 9の5単語または2, 3, 9の3単語)を用い、認識サンプルとして、各数字を10個ずつランダムに居る100サンプルを用いた。学習サンプルに含まれる音韻は図2に示す通りである。なお Φ_{ij} および x_{ik} は、男女別にもとめた。予備実験の結果から、 $r=0.5$ 、 $N=10$ とした。

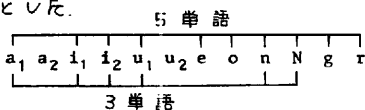


図2. 学習サンプルに含まれる音韻

比較のために、全く学習を行わずに $\hat{x}_{ik} = x_{ik}$ とした場合(学習なし)、 $r=0$ すなわち推定を行わない場合(推定なし)、および全単語学習の場合についても実験を行った。

図3に平均認識誤り率、図4に単語別の結果を示す。図3から、ここで検討した適応化方法を用いれば、男性・女性に共通して、5単語学習で全単語学習と同程度の認識率が得られ、3単語学習でも、学習なしの場合より誤り率が1.5%程度低下することがわかる。推定なしの場合は、3単語学習のように一部の音韻しか抽出されないときに大きな誤りを生ずる。図4を見ると、推定なしに比較して、推定を行うことにより学習単語に含まれなかった4, 6, 7等の誤り率が大きく低下し、全単語学習の場合と類似した分布を示すことがわかる。

Ⅳ むすび 少数の音声サンプルを発声して能率的に個人差学習を行う方法について検討し、その有効性を確認した。今後この方法を多数語彙の認識系に適用して評価する必要があり、現在その準備を進めている。この方法では学習サンプルとして認識対象語彙以外のものを用いることもできる。一方これまでに認識と同時に学習を行う方法⁽⁶⁾⁻⁽⁸⁾なども検討されており、これらの特徴を組合せた改善等についても検討して行きたい。

[謝辞] 日頃御指導頂く当所斎藤研究室長ら及び、熱心に討論して頂く研究室の諸氏に感謝の意を表す。

[文献] (1) 好田・斎藤: 信学研資, PRL72-87 (昭47-12) (2) 坂井・中川・林: 信学研資, EA75-61 (昭51-01) (3) 古井: 音学研資, S75-25 (昭50-11) (4) 好田・橋本・斎藤: 信学論, 55-D, 3, p.186 (昭47-03) (5) 古井・飯倉: 信学論, 55-A, 11, p.717 (昭48-11) (6) 古井: 音学論, 2-2-10 (昭49-10) (7) 古井: 音学研資, S74-17 (昭49-10) (8) 中川・白方・三村・坂井: 音学研資, S77-11 (昭52-06)

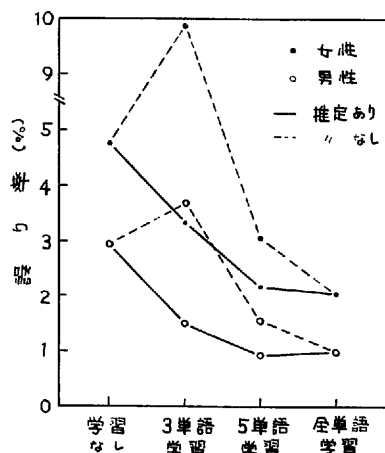


図3. 平均認識誤り率

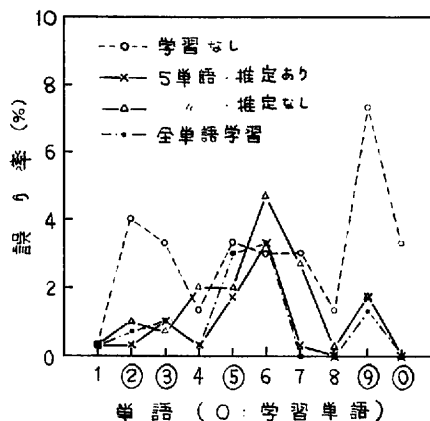


図4. 単語別認識誤り率