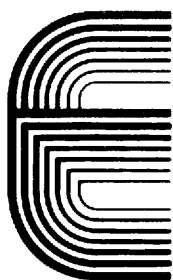


論文 / 著書情報
Article / Book Information

論題(和文)	音声の個人性情報と話者認識
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会誌, Vol. 75, No. 6, pp. 631-632
Citation(English)	, Vol. 75, No. 6, pp. 631-632
発行日 / Pub. date	1992, 6
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1992 Institute of Electronics, Information and Communication Engineers.



音声の個人性情報と話者認識

教養のページ

古井 貞 熙

古井貞熙：正員 NTT ヒューマンインタフェース研究所

Individuality Information in Speech Waves and Speaker Recognition. By Sadaaki FURUI, Member (NTT Human Interface Laboratories, Musashino-shi, 180 Japan).

人が話した音声の中には、伝えようとする言葉の内容（言語的情報）だけでなく、だれが話しているかという情報（個人性情報）、話している人の感情（情緒性情報）などさまざまな情報が含まれている¹⁾。この中で、言語的情報が最も基本的で重要であることはいうまでもないが、他の情報も我々の生活の中で重要な働きをしている。もし音声で、個人差や感情を伝えることができない活字のようなものであったら、我々の生活がいかに不便で味気ないものになってしまうか、容易に想像できる。特に個人性情報は、我々が意識しないで利用している重要な情報で、例えば、テレビやラジオの討論会などで、画面を見ていなくてもだれが発言しているかがわかったり、電話がかかってきたときに相手が名乗らなくても、親しい人ならだれからかかってきたのかすぐわかるのも、声が人によって異なるからである。

それでは、音声の個人差、音声に含まれる個人性情報とは何であろうか？このような一見単純な問いを、筆者が音声の研究分野に飛び込んで最初の研究テーマに選んでから、もう20年以上になる。この問い自身は、人間の音声生成の原理や音声知覚の原理に結びついた科学的なものであるが、音声を用いて発声者（話者）を判定する話者認識をはじめ、音声認識において声の個人差に自動的に適応する話者適応、音声合成において声の質を変化させる声質変換などの技術と密接に関連している。昔から、音声のスペクトルの時間パターンを濃淡の図形に表し

たものを声紋とよんで、これを人が目で見ただれの声かを判定するというのがしばしば行われてきたが、これには判断する人の主観に結果が大きく左右されるという問題がある。ここでは、話者認識という言葉で、工学的かつ自動的な手段で行う場合に限って用いることにする。

我々人間は、相手がどんな言葉を話しても、言葉によらない共通した声の個性があると感じていると思われる。そのような個人性情報の抽出法として、10 ms 程度の短時間ごとの音声スペクトルを、10～30 s 程度の文章全体について平均した長時間平均スペクトルを用いる方法がある。種々の音素のスペクトルを平均すれば、音素によらない平均的な個人性特徴が浮き出てくるのではないかという発想である。長時間平均スペクトルを対数変数変換したのちフーリエ逆変換したケプストラムに、共分散を用いた重みをかけあわせることによって、マクロなスペクトル形状などに含まれる、言葉によらない個人差を抽出できることがわかっている²⁾。発声内容を限定しない話者認識は、科学的興味からも工学的難しさからも依然として面白い研究テーマであり、最近、ベクトル量子化の方法によって工学的に大きな進歩が得られたが³⁾、その科学的意味に関しては説明が難しく、進歩はまだこれからという段階である。

一方実用的には、発声する言葉をあらかじめ決めておいてもよいから、話者を自動的に認識する精度を高くしてほしいというニーズがある。音声を、本人であるか否かを判断する一種

の鍵として用いる用途の場合には、「開けゴマ」のようにしゃべる言葉をあらかじめ決めておいても差しつかえのない場合が多いからである。人間の聴覚器官に関する研究によれば、人間の聴覚は音の変化に対して敏感であることがわかっていて、そこで話者の認識においても音の変化に関する情報を用いたら、うまくいくのではないかと考えた。そこで発声内容を限定した話者認識において、ケプストラムの動的特徴を、それまで音声認識で用いられてきたDPマッチングの方法に組み合わせて用いることを試みた。動的特徴の表現法としては、90 msの区間におけるケプストラムの時系列の、1次と2次の多項式展開係数を用いた。この方法は極めて高い話者認識精度を示し⁽³⁾、10年以上たった現在でも、発声内容を限定した話者認識方法の基本になっている。なおケプストラムの多項式展開係数を用いる方法は、音声認識にも有効であることがわかり⁽⁴⁾、現在ではデルタケプストラム法として、世界中の主な音声認識システムのほとんどで用いられている。

話者認識の難しさは、同じ人の同じ言葉でも、発声するときによってその物理的特徴が変化して認識精度が低下してしまうという点にあり、これが指紋と大きく異なる点である。同じ日に繰り返し発声した音声の変動は極めてわずかであるが、数か月たつとかなり大きな変化を生じる。更に風邪をひいたり、電話を通したりすれば、変化は一層大きくなる。このため変化を自動的にキャンセルする方法、各個人の標準パターンを逐次更新する方法などが研究されているが⁽⁵⁾、まだ解決されたとはいえない。

音声認識においてだれの声でも認識できるようにする、いわゆる不特定話者の音声認識技術は、現在の重要な課題の一つである⁽⁶⁾。音声から個人性情報を取り出すことができれば、それをキャンセルすることによって、個人差に起因する認識誤りを減らすことができるのではないかと、考えるのは自然であるが、それほど単純ではない。話者が認識できることと、その情報を音声から取り除くことの間にはまだかなり

ギャップがある。不特定話者音声認識の実現のためには、できるだけ多くの人の音声データを集め、それらを包含できるような音声認識モデルの構築につとめ、それでも対応できない個人差に話者適応機能で対処するというのが、工学的には望ましいと思われる。音声認識系は、発声者の個人差だけでなく、周囲の雑音やマイクロホン、伝送系などの特性の変化などにも対応できなければならないので、変化への適応機能を持つ必要があるのはもちろんであるが、基本的に予想できないことには対応できないと考えれば、むしろできるだけ予想できることはあらかじめモデルの中に確率の変動として入れておいた方がよいのではないかと考える。

音声の個人性情報は、つきつめれば人の数だけあるわけで、その中から普遍的な原理を見つけるのは極めて難しい課題である。冒頭でも述べたように、この研究課題は、音声のあらゆる分野と関連を持っており、この解明は極めて大きな波及効果を生む可能性を秘めている。それだけ奥が深い研究分野であるといえる。この分野に興味と関心を持って下さる方が増えてくれることを大いに期待している。

文 献

- (1) 古井貞照：“デジタル音声処理”，東海大学出版会（1985）。
- (2) Matsui T. and Furui S. : "A text-independent speaker recognition method robust against utterance variations", Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Toronto, S 6.3 (1991).
- (3) Furui S. : "Cepstral analysis technique for automatic speaker verification", IEEE Trans. Acoust., Speech, Signal Processing, ASSP-29, 2, pp.254-272 (1981).
- (4) Furui S. : "Speaker-independent isolated word recognition using dynamic features of speech spectrum", IEEE Trans. Acoust., Speech, Signal Processing, ASSP-34, 1, pp.52-59 (1986).
- (5) 古井貞照：“不特定話者音声の認識”，信学誌，73，12，pp.1286-1290（1990-12）。



ふるい ただおき
古井 貞照（正員）

昭43 東大・工・計数卒，昭45 同大学院修士課程了。同年電電公社電気通信研究所入社。以後、同研究所において、音声認識、話者認識、音声知覚などの研究に従事。昭53～54 年ベル研究所客員研究員。現在、NTT HI 研究所古井特別研究室長。工博、米沢賞、論文賞、著述賞、音響学会論文賞、科学技術庁長官賞、IEEE 論文賞各受賞。