

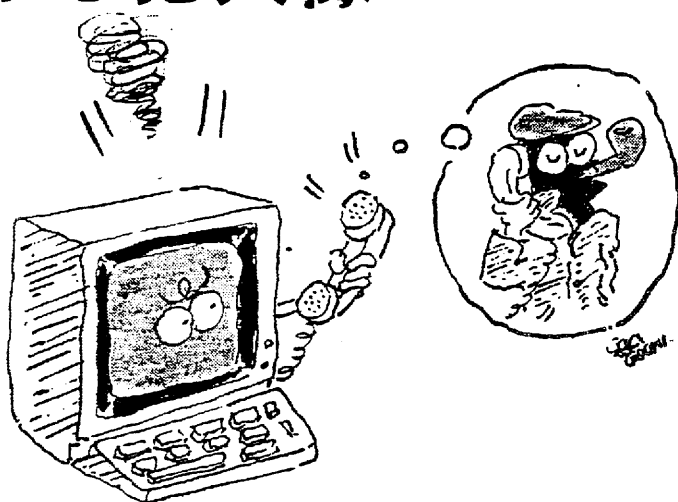
論文 / 著書情報
Article / Book Information

論題	声紋で浮かび上がる犯人像
著者	古井貞熙
出典	科学朝日 立体構成[現場を洗う] 3, 1983, 12月号, pp. 26-29
発行日	1983, 12
承諾番号	25-2865
Note	朝日新聞社に無断で転載することを禁じる。

声紋で浮かび上がる犯人像

古井貞熙

日本電信電話公社／研究開発本部調査役／ふるいさだおき



音声には、話す内容に関する情報のほかに、話している個人に関する情報や喜怒哀楽に関する情報など、種々さまざまな情報が含まれている。電話で話をする時、だれでも受話器を耳に当てるとすぐに相手がだれであるかを判断しようとする。そしてそれが思いがけない電話であっても、名乗るより先に相手がだれであるかを識別できる場合も多い。また知らない人からの電話であれば、男か女か、年齢はどのくらいか、などの判断をする。そして相手の話す内容を聞きとると同時に、うれしそうとか、悲しそうとか、怒っているようだとかの感情的な情報もとらえることができる。

人間が音声を用いて相手を識別する能力は、電話に限らず、非常に多くの場合に用いられている。たとえばガヤガヤした満員電車やエレベーターの中で、相手の顔が見えなくても会話ができるのは、この能力のためであろうし、テレビ討論会でしばらく画面を見れば、カメラがその人をとらえていなくても、だれが話しているのかを容易に判断できる。

周波数成分に手掛かり■

電話を用いた犯罪などに関連して、「声紋」を用いて個人を識別する方法を開発する研究が進められている。「声紋」は声の指紋という意味でつけられた名前だが、指紋と違って直接目で見ることとはできない。声紋を得るには、音声のスペクトル（各周波数成分の大きさ）が時間とともにどのように変化しているかを図に表す（スペクトル分

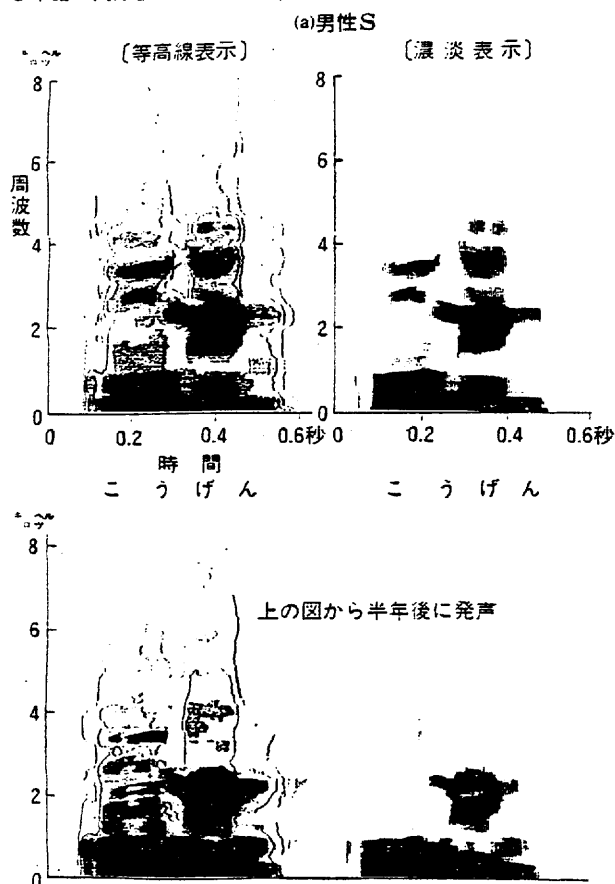
析）。声紋を自動的に描く機械は、1945年にアメリカのベル研究所のポッター博士によって開発された。この機械をサウンドスペクトログラフ、描かれた図形をサウンドスペクトログラムまたはソナグラムと呼んでいる。

サウンドスペクトログラムの表示方法としては、「濃淡表示」と「等高線表示」がある（①）。いずれも横軸が時間を、縦軸が周波数を示し、スペクトルの変化する様子を示しているが、濃淡表示の場合はスペクトルの大きさを濃淡で示すのに対し、等高線表示の場合は、地図の等高線のように、同じスペクトルの大きさを示すところを線で結んで、同じ大きさの部分を同じ濃さで塗りつぶす。「声紋」の名は、この等高線表示の図に由来し、ベル研究所のカースタ博士が、これが指紋に似ていることに着目して名づけたのである。

カースタ博士は、同じ言葉でも人によって声紋が異なることをもとに、初めて音声から個人を識別することを試みた。ところで声紋と指紋とは大きな相違点がある。指紋は生涯、変化しないのに対して、声紋は同じ人が同じ言葉を発声しても、しばらく時間がたってから言い直したり、前後の言葉が異なっていたり、作り声をしたり、電話回線を通ったりすると、変化する点である。このため、声紋を単に直接比較したのでは、必ずしも正しく個人を識別することはできない。

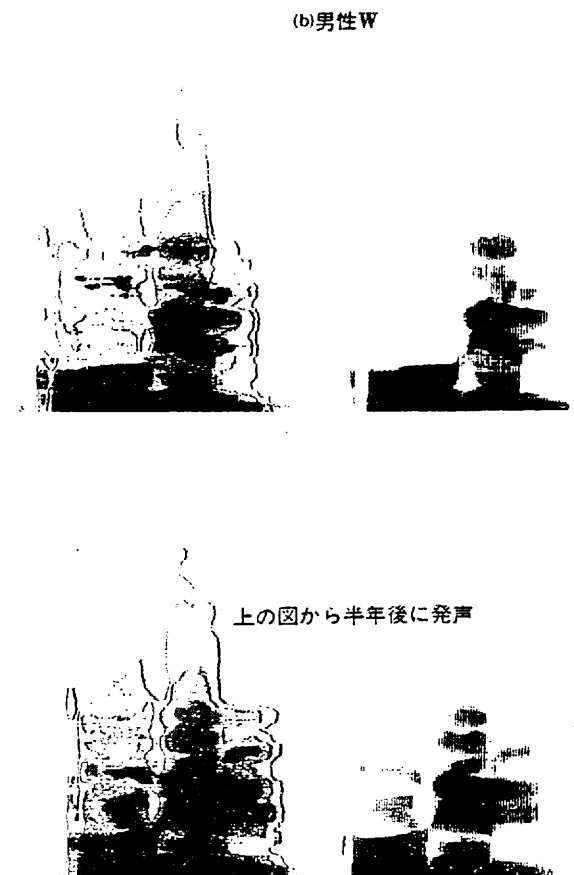
音声波は、肺から出された空気流が声帯と作用し合うことによって生じた声帯波を音源とし、それに口の形に応じて種々の音色が付加されて作られる。このとき声帯の緊張度が強いと高い声にな

① 単語「高原」のサウンドスペクトログラム。



り、緊張度が弱いと低い声になる。声帯の振動周波数のことを基本周波数あるいはピッチと呼ぶ。声帯から唇までのパイプ状の区間は声道と呼ばれているが、舌、あご、唇などを動かすことによって声道の形が変わると、それに伴ってさまざまな共鳴特性、すなわちスペクトルの大きい部分と小さい部分の違いが生じて、「あ」や「い」の音ができる。

従って、音声の特徴はピッチ（声の高さ）を含む声帯波の特性と、声道の共鳴特性に分けて考えることができる。音声に含まれる個人の声の特徴は、これらの発声器官の個人差によるものだけでなく、それらの発声器官をどう動かすか、すなわ



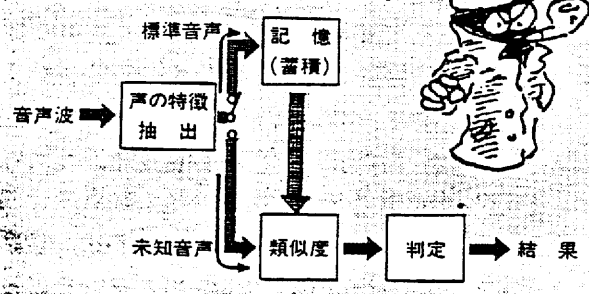
ち話し方の癖にも多く含まれている。いわゆる、しわがれ声・かすれ声は、声帯にでき物が生じたり、声帯が老化したりして、声帯が不規則かつ不完全な振動をするために生じるものである。

さまざまな用途

音声に含まれる個人の特徴を用いて、だれの声かを判定する方法には3種類ある。(1)人間が耳で聞いて判定する、(2)声紋を人間が目で見えて判定する、(3)コンピューターによって自動的に判定する、の3種類である(②)。

いずれの方法の場合も、あらかじめだれの声かが分かっている音声を用いて、それぞれの人の声の特徴を記憶しておき、だれの声かわからない音声、あらかじめ記憶している声のうちのどれか(話者識別)、あるいはある容疑者または登録されている人の声と見なせるか否か(話者照合)を判定する。話者識別と話者照合を総称して話者認識という。人間が耳で聞いて判定する場合は、声の質(音色)、話し方の癖など多くの特徴を総合的に組み合わせてだれの声かを判定しており、物理的にはピッチ、スペクトル、およびその時間変化

② だれの声かを判定するアルゴリズム。





特性が複雑に寄与していることが分かっている。

声帯模写では、声の質は必ずしもまねしようとしている人のものに似ているわけではない。話し方の癖を巧みにまねしたり強調したりして、それらしくしていることが多い。声紋を人間が目で見で判定する場合は、声紋が変化するので、変化しない特徴をうまくとらえて判定することが必要になる。声帯模写の音声进行分析してみると、変えようとしてもなお本人の声の特徴が残っていることが分かる。最近ではコンピューターによって声紋が、より精密かつ正確に描けるようになってきたが、個人の声の特徴を的確にとらえるには、まだかなりの熟練が必要である。

コンピューターによって自動的にだれの声かを判定する方法については、世界各国の研究機関で盛んに研究が行われており、音声波に含まれるどの物理的特徴が有効かが実験的に調べられている。コンピューターによる判定は、犯罪捜査ではなく、主として音声を一種のカギとして用いるサービスの実現のために研究されている。「開けゴマ」と本人が言った時だけ開くカギができれば、家庭用として使えるのはもちろん、コンピューター室や重要機密情報のある場所に、決まった人以外が入り込むのを防ぐことができる。テレフォンショッピングで預金から自動的におカネを払い込む時に、音声だけで本人か否かの判定をすることができれば、カギやクレジットカードを用いる場合のように、落として他人に悪用される心配もなく、余計なものを持たなくてもすむ。

だれの声かを判定する方法の信頼性、すなわち

どの程度の正確さが得られるかは、言うまでもなくその条件の難しさ、すなわちどのような声を対象とするかによって変わる。具体的には、あらかじめ声の特徴を記憶するのに用いる「標準音声」と、判定すべき「未知音声」が、(1)同じ言葉か違う言葉か、(2)同じ条件で録音されたものであるか、一方が電話音声で、一方が生の音声の場合のように異なる条件で録音されたものであるか、(3)発声音がカゼをひいていたり、鼻をつまんだり声を押し殺したりしてわざと声を変えているか、それともふだんの状態で発声しているか、(4)発声時期に数カ月あるいはそれ以上の時期的隔りがあるか否か、によって難しさは大きく異なる。

99%の正答率

音声を一種のカギとして用いる場合は、本人はそのカギがスムーズに開くように、できるだけふだんの調子で、決まり文句をしゃべることが期待されるので、本人が発声する限りは、上述の(1)～(4)のどの条件に照らしても、標準音声と未知音声はかなり類似する。このような場合は研究の対象として比較的取り扱いやすく、コンピューターを用いて実用に近い段階にまで技術が進んでいる。最近、電電公社で開発された方法や、同公社とアメリカのベル研究所が協力して開発した方法によれば、マイクロホンで録音した音声で99%、電話回線を通った音声でも97%程度の確度で、だれの声かを正しく判定できる。

しかし一方、誘かいの身代金要求や脅迫の電話など犯罪捜査の場合は、(1)～(4)のどの条件をとっ

「開けゴマ」

扉の前でひとこと叫ぶと、音もなく開く……という現代版「開けゴマ」装置が、このほど立石電機で試作された。叫んだ人が本人でないのに、開いてしまう誤り（誤照合率）は1%前後。

正式には「話者照合装置」と呼ばれており、装置の前に立った人があらかじめ登録された本人であるかどうか、音声によって確認する（話者照合）機能だけをそなえる装置で、試作機には「扉」はつけられていないが、これが実用化されれば、銀行の自動窓口や特別の資格を持った人以外は出入り禁止といった出入り口に広く利用される。さらに研究が進めば脅迫電話や誘かい犯の身代金要求電話など音声を手がかりになる犯罪捜査に役立つものと、科学警察研究所も注目している。

この装置では、あらかじめ対象者の声が登録されている。6音節の「ワシヤショウゴウ」という言葉が合言葉になる。この音声を低域通過フィルターを通して高周波成分を取り除いた後、ひとつながりのデジタル（数値）信号に変換し、さらにこれをもとに「パーコ



ル係数」と呼ばれる32次元のベクトルにする。ひとつの単語が、クリのイガのように中心から32方向に、いろいろな長さで伸びる矢で表されることになる。

照合を行う場合、利用者はまず併用する暗証カードで身元を明らかにした後、合言葉をとなえればよい。登録時と同じ手続きで同じような矢印の群れがコンピューター内部で作られ、同じ方向を向いている矢印の長さがひとつひとつ比べられて、類似度が算出される。

一定のレベルを決めておき、類似度がそれより大きければ本人、小さければ他人と判断する仕組み。

同一人であれば、類似度は100%になるのが理想的だが、実際はカゼをひいたり、時間をおいたりすると、本人の声であっても「違う」と判断することや、逆に他人なのに「本人だ」と判断することもある。この装置のシステムを開発した日本通信技術の小池恒彦電子装置開発部長らが、

成人男女30人を使って行った実験では、99.9%の照合率だった。

小池さんらは今後、さらに研究を進め、だれか分からないが、何人かの中にその声の持ち主がいるかいないかを調べる「話者識別」の装置も開発したいという。これができると、音声を手がかりにする捜査はさらに一歩前進することになる。

でも難しいケースが多い。異なる言葉に含まれる情報を用いて個人の判定を行うのは難しいので、通常は英語なら I, it, you, 日本語なら「あのう」、「ええと」など、よく会話に出てくる言葉を手掛かりにして、標準音声と未知音声の比較を行う。現段階では、何人かの中に必ず犯人がいる時は、90%以上の的中率で犯人を見極めることができるが、犯人がいるかいないかわからない時は、それだけの的中率を得るのはまだ難しい。

コンピューターと音声で対話をするための音声認識装置の開発は近年急速に進歩し、電話による預金の残高照会や、工場における品質管理データのコンピューターへの入力、空港や配送所での荷物の仕分けなどに実際に用いられるようになってきた。しかし、現在の音声認識装置には、まだいろいろな制約がある。

そのひとつは、あらかじめ使う人の声を登録しておけば、100種類以上の単語をほぼ正しく認識

できるのに対し、だれの声でも認識できるようにすると単語の数が十数単語程度に限られたり、認識装置の規模が大きくなってしまふ、いわゆる「人見知り現象」があることである。

どのような言葉をしゃべっても、その音声からすばやくその人の声の特徴を取り出して、認識装置を自動的にその声に合わせてたり、個人の癖や特徴を除去するようなことができれば、だれの声でも認識できる装置ができるであろう。声紋に象徴される音声の個人差の研究は、このように音声認識の研究と、いわば表裏一体の関係がある。

音声に含まれる個人情報と能率よく取り出して、だれの声であるかを判定したり、音声認識装置をその人に適応させたりする研究は、今後ますます重要になると予想される。そして、電話による犯罪の容疑者の割り出しの根拠として、さらに積極的に用いられるようになるのも、そう遠い将来のことではないと思われる。