

音声情報の魅力



電子技術総合研究所
知能情報部 部長
(財)サウンド技術振興財団 評議員
工学博士

中島隆之

音声研究に入り20年以上になるので、大学時代を含めると音響の分野で30年近く過ごしたことになる。本文では、これらの研究に直接あるいは、間接的に関連して私の頭に引っかかっている問題のモノログである。

音声が届く情報について

人の声は、様々な情報を運んでいる。思っていることを言葉にして伝える（音韻情報の伝達）ことが、その第一の役割である。多くの場合、音声の中には話し手の、個人性、教養、職業、出身地、人柄、健康状態、などの個性情報のほか、聴く相手に対する感情情報（怒り、喜び、容認、好意、反発、媚び、拒絶、落ち着き、傲慢、同情）が含まれている。

これらの情報の存在は、人間同士の対話を電話という機械を通じて行う場合にもよく経験する。音声波形を分析すれば、上記の音韻情報のほか、個性情報や感情情報さえも、抽出できそうである。これらの情報の特徴や特徴の組み合わせが明らかになれば人の声を聞き取る機械（音声タイプライターや音声理解機）のほか、うっとりするような美

声や、ある人の口調を真似てしゃべる機械（音声合成）が実現する。

たかだか3300ヘルツまでの音波成分を伝達する電話によって、上記のほとんど全ての情報が聞こえるので、機械による音声認識や、音声合成は容易に実現しそうに思えるが実は容易ではない。現象は、物理現象で観測可能な音声波形に凝縮しているが、その中の情報の完璧な検出はむずかしく、情報の理解には、背景に人間が持っている膨大な、知識、常識、言語体系が必要らしい。もっとも、音声は、種々の様相と局面があり、一般論で片付けることは危険である。一般に、ひどい雑音中の音声の聞き取りには音声波形はそんなに確かな情報を運んでいない。しかし、母親が幼児にゆっくり、はっきり発声して、言葉を教えているとき、音声の波形に含まれる情報を最大限に活用しているだろう。

人間の聴覚の素晴らしさ

ある種の記号またはパターンから音声を作り出すのを、音声合成と呼ぶ。音声の合成の研究を通じて私が感じているのは、人間の耳の素晴らしさである。不正確な合成法で作った音声は、やはり聴こえも不確かだ人間の耳を満足させられない。たとえば、サとシャは誰か聴いても違う音であり、合成にあたっては細心の注意が求められる。一般に合成音は、その音韻のエッセンス的要素から合成されるので、典型的な音韻であるべきであるが、現実には不完全な音である。したがって、このような音は聴く人によって違ったり、また、同じ人でも時間の経過にしたがって周期的に別の音に聞こえることがある。

音声波形の録音・編集による音声出力は、この種の不安定さがない。そういえば、電子楽器でも、特定の楽器音を合成することが盛んにおこなわれたが、最近は実物の楽器の演奏音を系統的に記憶して再生する形式の楽器が、最も本物らしい素晴らしい音を出している。

人間の耳による打音試験を機械で置き換えることは、まだ、かなり難しい。耳には明らかに聴こえるピリツキ音や、こすれ音についての記述的な分析は、まだまだである。この程度のこと

出来なければ、音声の認識など出来る訳がないではないか？

なお、打音試験に関連して、最近、振動の分析から、ボールベアリングの傷を発見することは、可能になっている。機械による音の分析能力を人間の聴覚に近づけることはどうしても必要である。

最近のインコは話がうまい、という話

先日「話すインコ」がテレビで紹介されていた。おばあさんが、手塩にかけて教え込んだものらしく、「むかし、むかし あるところに、おじいさんとおばあさんが、ありました。……」と長い桃太郎の話を、よどみなく、かなり長くしゃべったのである。その発声はインコのものとは、思えないほど明瞭に私には聴こえた。もしかして、テレビ局が人の声で吹替えたのでは？ と思わせるほどの出来映えであった。従来この種の話では、聞き手の人間が、その話を良く知っているから聞こえるのだと言われているが、少なくともこのインコは、この物語を正確に記憶（無論、意味を理解しているわけではないが）して発声（再生）している。どのような形式でどこに記憶しているのだろうか？ このインコは長い物語を記憶してテープレコーダのように再生するだけでなく、おばあさんと（オーム返しではなく）簡単な対話もするのである。

他人にも理解できたインコの身のうえ話

同じくテレビのニュースか、紹介か忘れたが、インコのはなしの続編である。先のインコかそれとは、全く別のインコか確かではないが、とも角、あるインコが可愛いがられていた家から逃げ出したという。しかし、数日後のある日、「お宅のインコでは、ないか？」と電話がかかり、めでたく帰宅となったという。その行きさつは次のようなものである。県も違う遠い家にそのインコは迷い込んだ。その家の人か捕まえて保護していたある日、そのインコが何やら、盛んにしゃべっている。耳を澄ましてよく聴くうち、住所をしゃべっていることが分かり、ついに、電話番号まで聴き取ることが出来たという。この話が、本当とすれば極めて興味深い。だいたい、九官鳥、インコの声は、その気で聞いて聴くから聞こえるので、その気が

無ければ聞こえない、といわれている。しかし、今回のインコの拾い主は、そのインコがしゃべるとのことさえも、思っていなかったのに、正しい住所だけでなく、電話番号まで正しく聞こえたとは驚くばかりである。無論、インコから住所を聞き出すためには、最近の音声認識機械が行っているのと同じように、「ああ言っているのかな、こうではないかな？」という、仮説を立て、そのつもりで聞いて確認する課程を通して正しい答えに到達したに違いない。しかし、これが成功するためには、やはり、かなりの正確な、発声があったはずである。テレビの報道に作意があるのでは？ などと疑いの目で見るときが無いが、先の、桃太郎の話を聴かされていた私としては、「さもありません」という次第である。

雑音中の音声か「ありあり」と聴こえる話

人間の耳では、一端理解された音は、ありありと聞こえる。華厳の滝の写真の中に死者の首を見た人は、常にありありと、それを見続けるように。これは、心理学でいうゲシュタルトと関連していると思われるが、要するに、人間は与えられた信号をもとに、それを解釈するための環境や知識を総合して「意味あるもの」として理解する能力をもっているのである。与えられた信号が確実であれば、正しい「意味あるもの」を認識するが、あいまいな信号でも、不適當な知識を総合して、間違った結果を「意味あるもの」として、ありありと導くことがある。騒音に埋もれた人の声でも何回も聴くうちに、突然ありありと聴こえてくることがある。一端、聴こえた言葉は、別の言葉に聴きなおすことはむずかしいが、同じ音を別の人に聴かせても同じように聴くとは限らない。

人間が、曖昧な信号に対しても「意味あるもの」として認識する機能をもつことは、先にのべたように、とんでもない誤解を生じさせる可能性がある。にもかかわらずこの機能があるということは、これらの欠点を補って、余りある利点があるに違いない。

それは、素早く危険や敵を見分ける原始的なパターン認識から、高度の音声認識に不可欠な機能であろう。