声とは、言葉とは、何か
——音声研究を通して考えること

東京大学大学院工学系研究科教授
峯松 信明

プロフィール／みねまつ・のぶあき
1990年 東京大学工学部卒業、95年東京大学大学院工学系研究科にて博士（工学）を取得。95年より豊橋技術科学大学に勤務し、2000年より東京大学に戻る。現在、東京大学大学院工学系研究科電気系工学専攻教授。音声科学から音声工学に至るまで、幅広い観点から音声コミュニケーションに関する研究に従事。特に音声技術を使った語学教育に関する造詣が深く、2009年よりOJADの開発を手がけている。

声とは何か、言葉とは何か。この根源的なテーマに応えてくださったのは、音声工学の第一人者である峯松信明先生。機械に音声を認識させる・合成させる、その研究を通して対極に見えてきたものとは何でしょうか。それはヒトの持つ不思議な能力——言葉と記憶、ヒトは言葉を操作しながら、実は言葉によって記憶を操作されている——その謎に科学の目で迫ります。

30数年ぶりの同窓会でのできごと

昨年Facebookを発端として30数年ぶりの中学同窓会が行われた。変わらぬ友の笑顔を通して、そして彼らとの会話を通して、30年ぶりの記憶が鮮明に蘇った。覚えていることすら忘れていた記憶が、である。同様の経験をされた方々も多いと思う。と同時に、「私はあと何を、頭の引き出しにしまっているのだろう？」と感じた方も少なくないはずだ。

脳科学の中で記憶研究はまだまだ未解決の部分が多いと聞く。あるものを聞く、見る、触る、嗅ぐ、味わう。五感を通して感覚した刺激は電気信号として脳に届く。ある刺激を記憶するとは、その電気信号によって脳の神経細胞群の一部が変容することを意味する。しかし、ある特定の記憶は特定の一つの神経細胞が担うのか、あるいは、ある特定の神経細胞群の同期発火が担うのか、対立する二つの仮説の間で科学者が議論を交わしている(1)。

記憶の神経生理学的謎解きは彼らに任せるとして、もう少し身近な記憶の不思議を考えたい。卒業以来30数年ぶりに会った友人の顔を見て「あの子だ」と分かる。中学の時の彼女の顔と今の顔を重ねても、(失礼かもしれないが）当然重ならない。さまざまな部位は変容し、皺、しみ、化粧…これ以上は必要ないだろう。でも「あの子だ」と分かる。私は彼女の顔を覚えている。でも、顔の何を覚えているのだろう？　私にとって、30年前の彼女の顔と今の顔との共通項とは何なのだろう？

母親が髪を切っただけで、あるいは、風邪声になっただけで「母親がいなくなった」と叫ぶ子供たちがいる。自閉症児に時として見られる光景である。自閉症の方々は、見聞きした刺激を正確に記憶する能力に長けていると聞く(2)。その一方、健常者なら許容する違い・ズレが許容できず、例えば「コージー冨田の物真似は分かるが、コロッケの物真似は何が面白いのか分らない」ということも起きる(3)。デジカメやボイスレコーダーのような正確な記憶術しか持ち合わせていなければ、30数年ぶりの再会は楽しめない。

『あの横顔を見てピンときた、という記憶の再生』と『修学旅行でKYがバスガイド泣かせたの覚えてる？と言われて思い出すバスガイドの泣き顔』の違いを考えたい。前者は私の頭の中にある彼女の横顔（の何か）と、目の前に出現した横顔（の何か）が一致した、ということなのだろう。後者は、KYがバスガイドを泣かせた記録映像を見ながら思い出したわけではない。つまり、その場に居合わせた時に見た視覚刺激が再現されて思い出したわけではない。友人のことば（声、音、空気振動）に対して、私の鼓膜が、聴覚器官が、そして、脳が反応し記憶が蘇るのである。視覚を通して蓄えたさまざまな記憶が、聴覚の刺激で蘇る。当たり前のように思えるかもしれない。しかし、これは当たり前ではない。

音声言語は他者の記憶を操作する

「いぬ」という声を聞くと、４本足の愛玩動物を思い出すが、この動物と、「いぬ」や「dog」という音とは本来無関係であり、前者に後者の音を意識的に結びつけることが「ことばを覚える」こと、である。「ことばを覚える」ということをしなければ、聴覚刺激によるさまざま

外国人が好きな響きの日本語　「〜しなければなりません」　理由「むずかしい発音が言えるようになった喜びを感じます」（中国系シンガポール、60代男性）

な記憶の再生は起こらない。これは『犬の鳴き声を聞くと、吠える犬の視覚イメージが頭の中に蘇る』というものとは異なる。鳴き声と吠える視覚イメージは、その場で、同時に見聞きしたものである。

「ある出来事」が起きた時、それに不可避的に付随して起きた刺激以外の、「ある出来事」と本来無関係の聴覚刺激によってその出来事の記憶が蘇り、時として記憶が操作される。私はこれが音声言語の基本的な機能だと考えている。音声言語の基本的機能としてコミュニケーションや思考という言葉をしばしば聞く。しかし、他者の記憶を検索したり（させたり）、操作したり、書き換えたり、上書きする、これがコミュニケーションであり、自己の記憶を対象とすれば、それが思考である、と考えることもできる。

動物実験などで「ある音を聞かせて、ボタンを押させる」訓練を行う（「音」と「ボタン押し」を連合させる）場合、何百回と訓練して連合できるようになる。子どもが指を差しながら「あれ、なあに？」と聞いてきた場合、「キャベツ」と一度教えれば、それだけで「キャベツ」という音と、薄緑色の葉の球状集合体（視覚刺激）は結びつけられる。

どうして音が「（本来それとは無関係のさまざまな）記憶」と、こうもやすやすと結びついてしまうのか？　何故、個々の事物に音でもって名前を割り当てることができるのか？　読者は「人間は言語を持っているから」と答えるかもしれないが、でもそれは「何故カレーが好きなのですか？」という問いに「好きだから」と答えていることと変わらない。私は大学の学部を卒業して以来、20年以上音声研究の場に身を置いてきたが、これが、最も根源的な問いの一つであると考えている。言葉の研究は、記憶の謎解きが不可欠である。

「ある出来事」と無関係だった刺激に対して、突如としてその出来事との間に強い結びつきが生まれ、その刺激を知覚すると、必ずその出来事の記憶が蘇る。研究者の中では共感覚と言語の起源を結びつけて検討している例もあるが[4]、もう少し身近な例でこのような事例を考えると、催眠術・催眠療法が類似した現象に思えてくる。催眠後、ある人の顔を見ると「初恋に似た淡い気持ち」に心が満たされる、といった催眠はテレビでよく見かける。そのような気持ちに満たされずにその人の顔を見られなくなる。「カレーライス」という音を「カレーライス」を思い出さずに聞くことができないように。

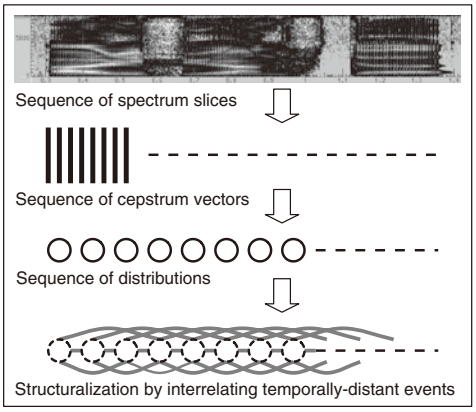
催眠は、催眠術師や医療関係者が相手の記憶の有り様を操作していることになるのだろう。と考えると、言語とは互いに相手の記憶を操作しあい、時には、相手の記憶を乗っ取るという操作を意味することになる。仮にある人が「言語のない状態」から「ある状態」に瞬時的に変容することがあるとすれば[5]、言語を有する状態に「恐怖」を抱くかもしれない。

ある出来事（吠える視覚イメージ）の記憶を想起させる際に、それと視覚的に類似した刺激や、その出来事と共起した別の出来事（鳴き声）が必要だった状態から、その出来事と無関係の刺激、しかも、他者が制御可能な刺激によってその出来事の記憶が再生させられる自分、自身が意図しないものまで思い出してしまう自分。初めての言語体験がこのようなものであれば、恐怖体験となるのではないだろうか。他者に乗っ取られる自分がそこにいるからである。

ソーシャル・ブレインという言葉がある[6]。他者の脳（当然複数の脳）との関わり・インターラクションを通して、個々の脳の機能を捉える考え方である。貴方が好む・好まないに関わらず、我々人間は、他者から操作され、他者を操作する存在である。群れを成して生息する動物も同様であるが、言語を有している人間は、動物と比較して遥かに他者の脳を操作することに長けている。教育、もその一つに過ぎない。チンパンジーやボノボが道具を使うことが観測されているが、道具の使い方を子どもに教えることはない[7]。

多様な声に潜む共通項をあぶり出す

話をもう少し、私の専門である音声の物理的側面に向けたい。最近、iPhoneの「Siri」や、ドコモ・スマホの「しゃべってコンシェル」など、音声対話システムが搭載されているスマホが人気となっている。従来ボタンを押して操作していた機械を「XYに電話をかけて」という音声で操作し、機械からも「携帯ですか？自宅ですか？」と音声で返答が来る。言わばバーチャル秘書のような機



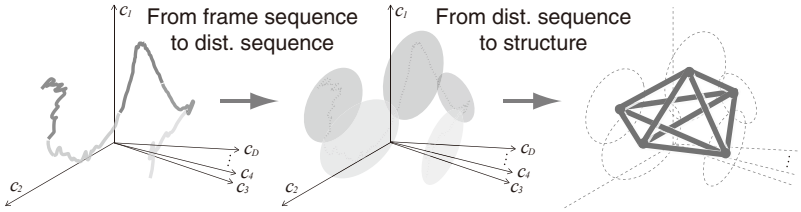


図1：声色の相対音感。音声の構造的表象と呼んでいる。この図についてより詳細を知りたい場合は、論文(12)をご参照下さい。

能である。

私の専門は「機械に音声言語を操る機能を授けること」を目標とする音声工学と呼ばれる分野である。鉄腕アトム、アナライザー[8]、ドラえもんなど、アニメの世界ではお馴染の、あの機能である。本稿の冒頭で30年の隔たりを越えて存在する（と思われる）顔の共通項について述べた。音声の物理的側面を日頃眺めていると、似たようなことは、30年という時間をかけずとも頻繁に遭遇する。

「XYに電話をかけて」と小学校入学前の女の子が言う、バスケ部の背の高い男子高校生が言う、30年ぶりに同窓会で会った女性が言う、来年退職を控えた年輩男性が言う。それぞれ音は異なる。声というのは「声道」という楽器を使って生まれる楽器音である。当然、声道の長さ、形状は人によって異なるため、声は話者によって異なる（逆に一卵性双生児の声は似てくる）。だから声を聞けば「あ、アイツが来た」と分かる。

つまり、Aさんの「電話をかけて」とB君の「電話をかけて」は音としては大きく異なる。しかし「誰がしゃべった」という視点で捉えれば異なる二つの刺激も、「何をしゃべった？」という視点で捉えれば「同じ」と判断する。30年という年月を考えるまでもなく、複数の話者の声を物理的に考えると、声の多様性と不変性という相反する二側面に気付かされる。と同時に、多様性の中に潜む共通項、音声の中の共通項は一体何なのだろう？と考えさせられる。

同様のことは、同一曲をハミングさせても起こる。子どもや女性の声帯は短く、軽い。なので、男性と比べて同じ呼気を生成しても、子どもや女性の声帯振動は早く、その結果、彼らの声は「より高く」なる。歌い手によって、より高いハミング、より低いハミング、となり、結局、音としては異なってくる。でも、聞けば同じメロディーであると分かる。音の高さそのものではなく、高さがどう変化したのか、この変化の様子の同一性から同じメロディーであると分かる。相対音感と呼ばれる。実は動物にはこの能力がないため、キーを上下したメロディーが同じものであると判断できない[9]。

一般には絶対音感を持つ人は音感が優れていると言われるが、単に、原始的な能力を後生大事に持ち合わせているだけである。鼓膜のすぐ裏側についている聴覚器官は、音の絶対的な高さ（単位はHz、ヘルツ）に基づいた処理が行われる[10]。つまり、誰でも動物同様、絶対音感を有しているのである。ただ、絶対音感に基づいた処理の結果を意識的に把握できるかどうか、の違いである。後生大事に動物的な能力を持ち合わせていれば、絶対音感者と言われるだけである。なお、相対音感が延びてこないと、音楽の鑑賞に支障を来すこともあるようだ[11]。

話者が違えば声の高さも異なるが、上記したように、少女の声、年輩男性の声など、声色（音色）も変わってくる。つまり、声帯の長さ・重さの違いは、声の高さを上下させ、声道の形状・長さの違いは、声の声色を多種多様にする。であれば、声色の多様性の中に潜む共通項も、音高の相対音感のように、声色の相対音感という考え方で説明できないのだろうか？

実はここ数年、峯松研究室で検討しているテーマの一つがこれである。多種多様な声の変形の中に潜む共通項をあぶり出す技術である。どのようにして共通項（不変項）を導くのか、というのは数学の力を借りる必要があるため、ここでは「声色の相対音感」とだけ紹介しておく。興味のある方は論文を参照していただきたい[12、図1参照]。

「言語がない状態」から「ある状態」へ

私は時として、「言語がない状態」から「ある状態」へ

外国人が好きな響きの日本語 **「ございます」** 理由「in Spanish “gozar” means “to enjoy”, and “y m←.A←Nas” means “and more”. So, “gozaimasu” sounds to us like “he/she enjoys, and more!”」(スペイン、30代男性)

図2：賢馬ハンス。当時の人は「本当に計算できる」と信じていた。やがて、トリックが判明したが、飼い主が意図的にやらせていたわけではなさそうだ。

の変化とは如何なるものか？を考えることがあり、既に思考の一つをお披露目した。このような変化は進化の過程の中で一度起こっている。サルの先祖から分かれ、ヒトになった時である。自閉症者かつ動物学者であるテンプル・グランディンは自閉症者には動物に多く見られる情報処理が色濃く残っている様子を指摘している[13, 14]。ヒト（健常者）特有の情報処理として、情報の抽象化、汎化能力を挙げている。確かに、自閉症者には絶対音感者が多い。母親以外の声だと言葉の意味をとるのに苦労する、なども自閉症に見られる例である[15]。デジカメやボイスレコーダーのような記憶術ばかりとなれば、抽象化、汎化は難しいはずだ。

私は受取った刺激から不要なものをそぎ落とし、本質的な骨格だけを抽出する能力を機械に授けようとしているわけだが、逆に言えば、「Siri」も、「しゃべってコンシェル」もそのような能力はまだ十分には身に付けていない。身に付けているのは、そのような能力を持っているように上手に見せかける技術、と言えばよいだろうか。

20世紀初頭ドイツに「計算ができる馬」がいた[16, 図2参照]。「３＋２」と見せると、蹄で地面を蹴り始め、５回で止める。新聞は「計算ができる馬」とはやし立てたが、詳細な調査の結果そのトリックが判明した。解答数だけ地面を蹴ると出題者や観衆が息を飲む。この雰囲気の微妙な空気の変化を察知して、彼は止めていた。馬は計算していない。「計算していた」と判断したのは人の方である。

「Siri」や「しゃべってコンシェル」は「言葉ができる機械」なのだろうか、消費者が勝手にそう判断しているだけなのだろうか？　ヒトとサルの違い、先天的障がいにより音声言語の獲得に困難を示す子供たちと健常者との違い、を考えながら「言葉のない状態からある状態」への変化を機械の上で実装することを考えると、巷に流通している音声対話システムの見え方が変わってくる。機械の「言葉使い」は、まだまだ自然な「言葉使い」じゃないなと。

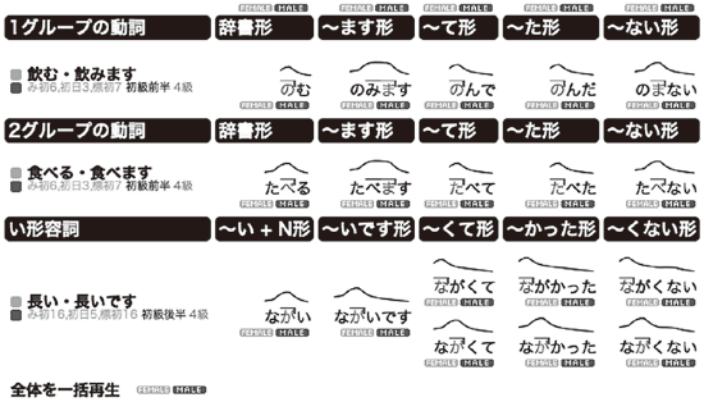
母語話者に聞き取りやすい発声を——韻律トレーニングのすすめ

最近、興味深い音声サンプルに遭遇した。中国語を母語とする日本語学習者の日本語音声である。アニメのアフレコが流行っており、日本のアニメに自身の声を入れて練習していた。その様子を見せてくれる、というので、彼らの声入りアニメを複数の日本人と一緒に鑑賞した[17]。彼らの第一声を聞いた直後、「何て言ってた？」日本人は互いに顔を見合わせた。近くにいた（恐らくそのアニメを初めて見たであろう）中国人日本語教師に、「何て言ってたか分りますか？」と聞くと、「はい」と言われた。学習者がアニメキャラクターとして意図した記憶・状況が、その先生の脳裏に蘇ったらしい。

私は「日本人英語を、日本人と会話したことがない米国人に聴取させて書き取らせる」というデータ取りを行ったことがある[18]。確かに（本人は正しく発声できたと思っても）日本人英語は聞き取ってもらえないことが多い。同様のことが起きていたのでは、と想像する。音声は年齢、性別、体格、さまざまな要因によって変容する。しかし許容される変容、されない変容があり、非母語話者の音声は時として後者となる。

日本語を勉強する場合、多くの場合聞き手は日本人となるから、母語話者にとって聞き取りやすい（記憶が操作されやすい）発声が望まれる。そのような発声術を効率的に身に付けるための一つの方法として、韻律トレーニングがある。単語を一つ一つ発声するのではなく、意味の区切りを意識してフレーズとイントネーションを構成し、フレーズ区切りには明確にポーズを置く、という

(A)



(B)

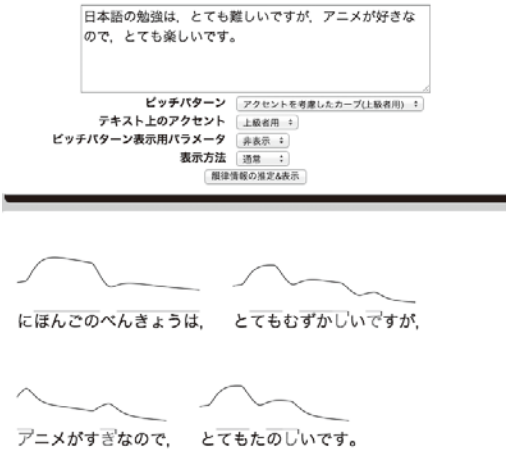


図3：Online Japanese Accent Dictionary, OJAD。用言の活用に伴うアクセント変形を示したり(A)、任意の文に対して適切なアクセント位置やピッチパターンを表示します(B)。

ものである[19]。

トレーニング前後の音声を聞くと、聞き取りやすさの変化に驚かされる（恐らく学習者自身はその変化に気付いていないのではないか）。日本語の場合、単語アクセントがピッチアクセント[20]であるため、イントネーション（ピッチパターン）はアクセントにも影響される。より自然な日本語発声を目指す場合は、アクセントにも幾分注意を払う必要がある。

このような観点から、現在、韻律教育を支援するwebシステムを開発している[21]。Online Japanese Accent Dictionary（OJAD図3参照）と呼んでおり、アクセントのみならず、イントネーション教育も踏まえて各種モジュールを開発している。興味のある方は「OJAD」でGoogle検索して欲しい。恐らく、トップに表示されるはずである。

「声とは、言葉とは、何か」というお題をいただいて、この一年経験したことを振り返りながら、私なりの意見を徒然なるままに書かせていただいた。この記事が、読者の頭の中にある、これまで連合したことのない記憶と記憶の結びつきに貢献できたとすれば幸いである。

「！」と感覚したとすれば、私の言葉が貴方を乗っ取った、のかもしれない。言葉、あな恐ろしや。

参考文献と解説

（1）T. グリューター、“再燃する「おばあさん細胞」論争”、別冊日経サイエンス154、2006

（2）U. フリス、“自閉症の謎を解き明かす”、東京書籍、2005

（3）峯松信明、“声の物理的多様性とその認知的不変性 ～音声認識技術と自閉症の類似性～”、コミュニケーションとリハビリテーションの現象学研究会、2010（アスペルガー症候群である綾屋五月さん（“発達障害者当事者研究”医学書院、執筆者）との対談より）なお、コージー冨田は模倣相手をそっくりそのまま真似る。コロッケはデフォルメして真似てくる。

（4）V. S. ラマチャンドラン、E. M. ハバード、“数字に色を見る人たち ～共感覚から脳を探る～”、日経サイエンス８月号、2003

（5）このような現象を直接記述したわけではないが、下記の本は非常に示唆的である。言語のない状態からある状態への変化を、当事者が語っている。S. シャラー、“言葉のない世界に生きた男”、晶文社、1993

（6）開一夫、長谷川寿一、“ソーシャルブレインズ ～自己と他者を認知する脳～”、東京大学出版社、2009

（7）明和政子、“霊長類から人類を読み解く なぜ「まね」をするのか”、河出書房新書、2003

（8）“宇宙戦艦ヤマト”に登場したロボット。このアニメ、小・中学時代にハマったのだが、そのリメーク版が４月からテレビで始まっている。39年ぶりのリメークであるが、2199年（この年にヤマトは旅立つ）までに人類はアナライザーを作れるのだろうか？

（9）M. D. Hauser, et al., “The evolution of the music faculty：a comparative perspective”, Nature neurosciences, 6, 663-668, 2003

（10）中川聖一、東倉洋一、鹿野清宏、“音声・聴覚と神経回路網モデル”、オーム社、1990

（11）最相葉月、“絶対音感”、新潮社、2006

（12）峯松信明他、“音声に含まれる言語的情報を非言語的情報から音響的に分離して抽出する手法の提案 ～人間らしい音声情報処理の実現に向けた一検討～”、電子情報通信学会論文誌、J94-D、１、12-26、2011

（13）T. グランディン、“動物感覚 ～アニマル・マインドを読み解く～”、日本放送出版協会、2006

（14）最相葉月、“ビヨンド・エジソン（第六章：言葉の不思議を探求する ～音声工学者・峯松信明と動物科学者テンプル・グランディンの自閉症報告）”、ポプラ社、2012

（15）東田直樹、東田美紀、“この地球にすんでいる僕の仲間たちへ ～12歳の僕が知っている自閉の世界～”、エスコアール、2005

（16）賢馬ハンス、http://ja.wikipedia.org/wiki/賢馬ハンス

（17）ICJLE2011のバンケットで行なわれた中国人日本語学習者によるアニメ・アフレコのデモ。同様のことを感じた日本語母語話者は多かったのではないだろうか？

（18）N. Minematsu, et al., “Measurement of objective intelligibility of Japanese accented English by using ERJ（English Read by Japanese）database”, Proc. INTERSPEECH, 1481-1484, 2011
http://www.gavo.t.u-tokyo.ac.jp/~ogaki/

（19）中川千恵子、許舜貞、中村則子、“さらに進んだスピーチ・プレゼンのための日本語発音練習帳”、ひつじ書房、2009

（20）多くの言語では、単語の一部をその他の部分より目立たせるアクセントと呼ばれる現象が観測される。但し、音声のどの側面を使って“目立ち”を作るのかは言語依存である。英語の場合、強さを制御して強勢・弱勢を作る。日本語の場合は、高さ（ピッチ）を制御して、高／低を作る。飴と雨が良い例であるが、方角の東、西、南、北と人名の東、西、南、北などもアクセントを考える良い例となっている。

（21）峯松信明他、“日本語韻律教育の支援を目的としたオンラインアクセント辞書と読み上げチューターの開発”、電子情報通信学会音声研究会資料、SP2012-115、pp.1-６、2013

（担当：大谷、松田）