

「口腔」を中心に数値シミュレーションと言語進化を考える



若 者

野 崎 一 徳*

Thinking about the spirit of compassion and information science,
centering on "teeth"

Key Words : Supercomputing, Grid computing, Sibilant

我々自身を知ることは誰しもが興味をもつことであり、古くから人はなぜ話すのか、いつ頃から話すようになったのかという議論がなされてきた。フランス音声学会では言語の起源を議論することを禁じたことがあるほどだ。エイステン(1996)は、言語の起源を論ずる時、大部分が主観的な主張であるがゆえに、頑固さが生まれ、それによって受容性が損なわれることにより、このテーマ自体が悪評を得る事となったと論じている。実際、タイムマシンでも開発されない限り原始人がどのような発音で言語を話していたのかを知ることは不可能である。

話は少しそれるが、自動車産業というのはその時代における科学技術の平均的な姿を表しているように思える。最近では自動運転の実現が間近に迫っているという大きな転換点があるように思う。私は以前から自動車レースに興味を持ち、特にそのボディーデザインに強い関心を持ち続けていた。1980年台のフォーミュラーワン(F1)は魅力的な形状をしていたが、先頭や最後尾に設置されたウイングという整流し車体を地面に押し付けるための装置の形状は直線状で平面的であった。ところが、2000年台に入るとF1のウイングのデザインは大きく様変わりをし、曲線的、曲面的である種生物的な形状となった。そこには数値流体力学(CFD)を用いた計算機支援設計(CAD)による革新があった。CFD

はそれまでも当時のパソコンで実行することは出来たが、時速300キロメートルを超えるF1では車体周囲の空気の乱れ(乱流)を考慮する必要があった。乱流というのは発達するとこまかな渦の集合になるので、その空間的構造を把握する必要が出てくる。そのためには解析対象の空間を細かく区切って、区切った空間を1単位として物理法則を満たすようにし、全体を表現する手法を主に用いる。そのなかで特に、物体と空気が触れる部分に最も速度の勾配、つまり物体壁面で速度ゼロと物体近傍で時速300キロメートルになる領域における急激な速度変化を正確にこの手法で補完するには、壁面近傍をより細かく区切るしかない。これにより最終的には対象空間を数千万から数億の要素に区切ることになる。その結果数値計算に必要なメモリ容量は数百ギガバイト程度となり、さらにメモリにある全情報を一気に演算機に送り込む際に必要な最大通信容量とその速度も重要となる。演算機においてはワンクロックで一度に多くの四則演算が行えることが性能に直結する。これらの計算機技術が向上することにより、より複雑な形状に対するCFDが実現し、その結果CADによって導き出されたある種生物的な形状を構築するに至った。このような流れは今をときめく機械学習でもみられ、現在自動車自動運転に用いられている目の部分、特に物体認識において主に用いられている方法では、カメラからの映像を無数に区切られた空間に分割し、どのように分割された枠組みに物体外形が最も該当するかを、学習された重みに基づいて確率を算出する。そこでは先のCFDで用いられた計算機テクノロジーと似た概念が用いられており、より汎用性と市場性が考慮された計算機が搭載されている。

CFDにおいても機械学習による物体認識においても、より良い精度を得るためには観測データもし



* Kazunori NOZAKI

1974年8月生まれ
大阪大学大学院 情報科学研究科 マルチメディアデータ工学専攻博士後期課程(2009年)

現在、大阪大学 歯学部 附属病院 医療情報室 病院准教授 博士(歯学、情報科学) 言語・発話・情報歯学

TEL : 06-6879-2427

E-mail : knozaki@dent.osaka-u.ac.jp

くは理論解析値との比較をし、それとの誤差を最小化する必要がある。つまりデータを尤も妥当に説明できる計算モデルを見つけ出す作業が必要になる。そのような計算モデルが得られ、ある条件をみれば外挿、すなわち観測されている領域外についても予測することが可能となる。これは数値シミュレーションの範疇となる。

話を元に戻すと、タイムマシンでも開発されない限り言語の起源を含む古代人の発話の様式は分からない、もしくは主観的主張にとどまるという問題提起に対して、現在の我々が出来ることがある。それは、現代において観測可能な発話に関する情報を収集し、収集されたデータを尤も妥当な範囲で説明できる理論モデルをいくつか構築することである。もしも理論モデルが構築できたとすれば、それを用いて時間的な外挿を行い、数値シミュレーションによって現在から過去を逆推定することも不可能ではないと思われる。

私は、高校を卒業し歯学部に入學、卒業し歯科医師となった後に、大学院において総義歯やスポーツ用マウスガードを製作するうちに、前歯の裏側周辺の形状と摩擦音の発音しづらさの関係に興味を持った。それは前歯の裏側の形状がF1のウィング形状を想起させたからである。そこで、その時の指導教授であった大阪大学大学院歯学研究科の前田芳信先生に大阪大学サイバーメディアセンターにある当時世界で16番目に高速であったスーパーコンピュータ SX-5 (日本電気株式会社) を利用したいとお願いし、技術専門職員の方の手厚いサポートもあって、おそらく当時世界初であったかもしれないが歯茎摩擦音 /s/ の数値乱流シミュレーションを行った(図1)。大学院を卒業後、やはりスーパーコンピュータを自由自在に利用したいと考え、当時バイオグリッドプロジェクトを主催しておられたサイバーメディアセンターの下條真司教授の研究室に就職した。やはり歯学部を出て歯科医師になったからには、社会に役立つ技術開発を行いたいという思いも強かったことから、歯科医院においていつでもスーパーコンピュータが利用できる環境が実現するようなシステムプラットフォームが作れないかと、個人的にデントグリッドプロジェクトを始めた。現在までそれは継続中である。さて、サイバーメディアセンターでは超大規模シミュレーション(約7千万要素)

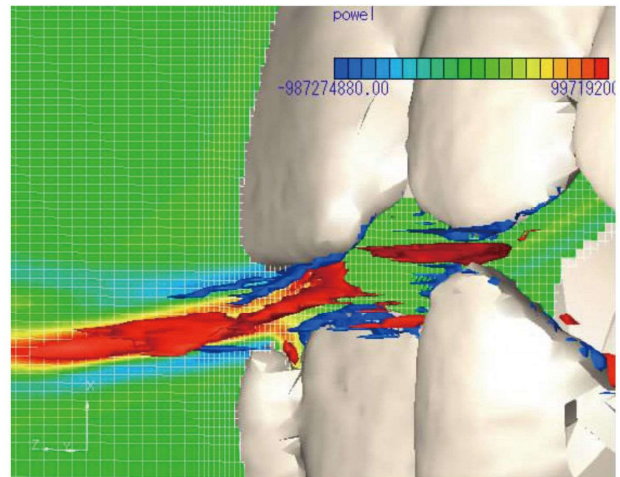


図1 歯列模型を用いた乱流シミュレーションにより、摩擦音発音時の音源を可視化した。前歯の裏側から切端にかけて音源が分布していることが分かる (Nozaki et al., 2005)。

の実現に打ち込んだ。当時(2005年頃、乱流から生じる音(空力音)の数値シミュレーションは困難を極めており、特に空力音源を流れそのものから算出し、音の疎密波を生み出す過程まで計算するためには、音源自体の計算精度に問題が山積していた。そこで東京大学生産技術研究所の加藤千幸教授が開発し一般に公開されていた乱流解析ソフトウェア(FrontFlow / Blue)を用いて、大規模数値シミュレーションを行うことでこの問題に取り組んだ。後に加藤先生の研究室にお世話になることになるが、その際、加藤先生が「歯医者なのに7千万(要素)?」と仰られたとお聞きしたことは、私の中の誇りとなっていることは言うまでもない。その後、大阪大学臨床医工学融合研究教育センターが立ち上がり、グローバルCOEプログラムで特任教員を募集していると聞き、以前から医工学を志している者として参加しない選択肢はないと思い応募した。そこではこれまであまり出会えていなかった、本気で医工学を推進し発展させて来られた先生方に出会うことが出来た。特に、基礎工学研究科の和田成生教授には機械工学的な考え方と論文の書き方について多くを学ばせて頂いた。そこでは結果的に今年(平成30年)に学位を取得することになった学生と一緒に摩擦音に関する研究を徹底的に行なった。結果として、歯科系、音響系[6]、流体系にそれぞれ成果を論文としてまとめることが出来た。途中、ポスト「京」重点分野2のプログラムに参加する幸運に

恵まれ、「京」を用いた計算規模的にハイエンドな数値シミュレーションを実施できる機会を得た。現在のところ3億数千万要素程度あれば摩擦音の数値乱流シミュレーションは精度良く行えるというところまで分かってきている。

さらに話を戻すと、原始人の発話機能を調べる際に、現代科学者がしていることは、まず解剖学的な形態を化石骨等を調べ、現代人との違いを見つけることだ。肺から唇にかけての三次元的な空間形状（声道）の違いから、発話能力を推定している。類人猿とホモ・サピエンスでは口腔領域において、発話時の空気の流れ方向の長さ、すなわち前後方向に違いがある。ネアンデルタール人とホモ・サピエンスの間にも同様の違いが存在する。考古学研究では、流れ方向の口腔の長さが短くなり、喉頭が下降し、それによって咽頭付近の空間が上下に伸びたことで、舌の居場所が狭くなったため、その中心部付近で屈折し折りたたまれた状態になりましたと推測されている。このように折りたたまれた状態になったことによって、舌はより機敏に自由に動くようになった（Bastir et al., 2011）と考えられている。恐らく、機敏に動くには動作の開始時点において力学的に均衡がとれた状態でありつつも揺れている必要があるのだと思われる。

言葉を発生させるしくみについて研究する調音音声学の分野では、声道の形態である声帯と舌と上顎の間の狭窄の位置、そしてそれらの位置での声道断面積に代表される特徴の値が分かれば、そのパラメータからどのような音色の音が発生するかを予測できることが知られている。ホモ・サピエンスは、声道の形態を多様且つ機敏に変化させることで音素の組み合わせを生み出し、言語における単語から文節まで音による符号化を実現させ、意味情報の交換を行うことが出来るようになったと思われる。

調音音声学では、母音と子音でその発生機構が異なることが知られている。母音の音源は声帯にあるが、子音には声帯が音源になる有声子音とそうでない無声子音がある。「さ」行や「は」行などの無声子音の音源の正確な場所は、最近ようやくわかってきた。1990年代にアメリカのマサチューセッツ工科大学が、簡単な木製の狭窄された円管と円管出口付近に設置した三日月状の障害板（歯に模した板）で構成された機械モデルで、「さ」行に近い音を発

生することを示した（Shadle, 1990）。一方で、前歯が欠けたり抜けたりすれば、「さ」行が話しにくいことは歯科医師の間でも古くから知られており、入歯等の制作では「さ」行を特別に意識した歯並びの方法が提案されている（Pound, 1951）。

このようなことから我々は、「さ」行を発音するには歯の裏の形態が重要な要素なのではないか、と考え、当時モータースポーツ分野で発達していた数値流体シミュレーションを口腔内での調音に取り入れた。その結果、上下前歯の隙間付近に流体から発生する音源（空力音源）がある可能性が示唆された（Nozaki et al., 2005）。

研究を進める中で、複雑な三次元形状である歯並び（歯列）を正確に考慮しなければ「さ」行を発音することは出来ないのか、という疑問が生じた。実際には腹話術師や歯科矯正が必要な患者も、「さ」行に近い音を出すことができる。では、どこまで歯列が異なっても同じ「さ」行と聞き取ることができるのだろうか。この疑問を解消するため、ホモ・サピエンス（著者）の実際の声道の特徴量（声道データにどのような特徴があるかを数値化したもの）として中舌部、最狭窄部、下顎前歯後部、上顎前歯部、口唇部の5箇所断面積を算出し、図2に示すような単純な機械モデルを構築した。この機械モデルは、舌を模擬した構造物を流れ方向に移動させることで「さ」行の音を模擬できる（Yoshinaga et al., 2017）。本当の声道とはたった5点での断面積でしか一致させていない。この結果から、母音に限らず子音も首の長さ、顔の形、顎の大きさ、そして歯の形が人によってそれぞれ違っていても、おそらく、我々が「さ」行と聞き取ることのできる声を出すことが出来るのであろうことが想像できた。

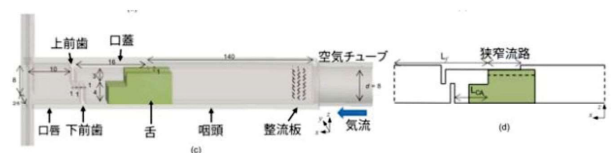


図2 「さ」行の「す」と「し」について、それぞれ複雑な口腔形状を主要な5点での声道断面積を参考に構築した矩形モデルで、実形状の気道モデルと同様の音響スペクトルを発生させることに成功した（Yoshinaga et al., 2017の図を一部改変）。

幼児（1歳～6歳）では、「○○○○マンはきみシャー」とか「おじいシャン」とか「おばあチャン」と発音することが多い。これは日本語の /s/ の構音は出来ないが /sh/ の構音は出来ることを感じさせる。/s/ と /sh/ の物理的なちがいについては、我々の研究チームが実験と数値流体音響シミュレーションの結果から音源における違いと共鳴腔の形状の違いが絡み合っていることを明らかにした (Yoshinaga et al., 2017)。注目すべきは、舌体を模した形状を下顎前歯に密着させるか否かで明らかな音色（音響スペクトル特性）に違いが生じたことである。幼児はなぜ舌を下顎前歯に押し当てジェット気流を上顎前歯に向けて流せないのか、未だ不明である。これらの研究を進めていくうちに、もしも摩擦音の理論モデルが構築できたとすれば、それを用いて時間的な外挿を行い、数値シミュレーションによって現在から過去を逆推定することも不可能ではないと思われる。それによって、例えば、ネアンデルタール人は舌部分の移動を容易には実現できなかったであろうとか、だから遥かに単純な言語体系しか持っていなかったのだろうかを、物理に基づいた仮説として提示することが出来ると考えている。

ヒトがなぜ話せるか、その一部を物理的な観点から解き明かすためには、おそらく口腔と咽頭空間の体積比が大きな要因と思われるが、さらに複雑な言

葉を話すためには、舌運動の言語機能に関する発達が不可欠だ。今後、舌運動の物理的詳細について、様々な分野の研究者と連携しながら調査を進める予定である。

参考文献

- 1) Aitchison J., The seed of speech: Language origin and evolution, Cambridge University Press, Cambridge (1996)
- 2) Bastir, M., et al., Evolution of the base of the brain in highly encephalized human species, Nature Communications, Vol.2, pp.588 (2011)
- 3) Shadle CH., The acoustics of fricative consonants, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 194 (1985)
- 4) Pound E. Esthetic dentures and their phonetic values, J Prosthet Dent, Vol.1, pp. 98-111 (1951)
- 5) Nozaki K. et al., The first grid for the oral and maxillofacial region and its application for speech analysis, Methods Inf Med, Vol.44, No.2, pp.253-256 (2005)
- 6) Yoshinaga T., et al., Effects of tongue position in the simplified vocal tract model of Japanese sibilant fricatives /s/ and /ʃ/, J Acoust Soc Am, Vol. 141, No.3, EL314, (2017)

