

音声の構造的表象に基づく音声認知と外国語発音学習に対する一考察

峯松 信明[†] 西村多寿子^{††} 櫻庭 京子^{†††}

[†] 東京大学大学院新領域創成科学研究科 〒277-8561 千葉県柏市柏の葉 5-1-5

^{††} 東京大学大学院医学系研究科 〒113-0033 東京都文京区本郷 7-3-1

^{†††} 清瀬市障害者福祉センター 〒204-0013 東京都清瀬市上清戸 1-16-62

E-mail: [†]mine@gavo.t.u-tokyo.ac.jp, ^{††}nt-tazuko@ams.odn.ne.jp, ^{†††}sakuraba@mtd.biglobe.ne.jp

あらまし 音声コミュニケーションには、話者・環境・聴取者に起因する、多様な音響歪みが不可避免的に混入する。その一方で幼児は、母親・父親の音声という非常に偏った音環境の下で音声言語を獲得する。これは、音響的に極端に偏った音声資料の聴取を通して、多様な音響歪みに対処する術を獲得することを意味する。何故このようなことが可能なのだろうか？本稿ではまず、筆者らが提唱する話者不変の音声の構造的表象である音響的普遍構造が、非線形変換を含む、幅広い変換群に対して一切不変であることを示す。次に、階名としての音韻を考えることで、上記の「多様性の謎」を考察する。その中で、パラドックスとも言われる失読症者や自閉症者の音声認知が、至極当然の現象として考察可能であることを示す。最後に、文字列という「一次元配列としての音声視覚化技術」をインターフェイスとして新たな言語と接すること、即ち、現状の外国語学習の形態について、一つの考察を行なう。

キーワード 音響的普遍構造, 変換不変, 相対音感, 言語獲得, 失読症, 自閉症, 外国語学習

Consideration on foreign language pronunciation learning with respect to structural recognition of speech

Nobuaki MINEMATSU[†], Tazuko NISHIMURA^{††}, and Kyoko SAKURABA^{†††}

[†] Grad. School of Frontier Sciences, Univ. of Tokyo, 5-1-5, Kashiwanoha Kashiwa, Chiba, 277-8561 Japan

^{††} Grad. School of Medicine, Univ. of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0031 Japan

^{†††} Kiyose-shi Welfare Center for the Handicapped, 1-16-62, Kamikiyoto, Kiyose-shi, Tokyo, 204-0013 Japan

E-mail: [†]mine@gavo.t.u-tokyo.ac.jp, ^{††}nt-tazuko@ams.odn.ne.jp, ^{†††}sakuraba@mtd.biglobe.ne.jp

Abstract In speech communication, various acoustic distortions are inevitably involved by speakers, channels, and hearers. However, infants acquire a spoken language mainly with speech samples of their mothers and fathers. They can easily solve the variability problem only with a remarkably biased speech corpus. Why and how is it possible? To answer this hard question, we already proposed a speaker-invariant structural representation of speech, called acoustic universal structure. In this report, the proposed representation is shown mathematically to be invariant with more general transformations including non-linear ones. Then, based on this representation, it will be shown that the speech recognition processes of dyslexics and autistics, often viewed as paradox, could be taken for granted. Finally, considering the proposed solution, a discussion is done about whether or not the one-dimensional visualization of speech, i.e., written language, is valid enough in foreign language learning.

Key words Acoustic universal structure, transformation invariance, relative sense of sounds, language acquisition, dyslexia, autism, foreign language learning

1. はじめに

大部分が「母親と父親の音声」という、音響的に非常に偏った音声資料の提示を通して、幼児は音声言語、即ち、音声に不可避免的に混入する多様な音響歪み（変動）への対処法を獲得する。一方、音響音声学／音声工学は個々の音素の音響モデルを、数千・数万という話者の音声を集めて作ることで多様性問題の解決を図ってきた。両者の差異はどこにあるのだろうか？

話者 A の音声を書き起こす。話者 B の音声を書き起こす。この時、話者 A によって発せられたある音響事象を「あ」という記号で表記し、話者 B によって発せられたある音響事象も「あ」という記号で表記する。当然、両事象に物理的等価性は保証されない。物理的に異なる音響事象を、同一の表記を用いて書き起こす訳である。なぜ話者毎に「あ」という記号の変種を用意することなく、「あ」と表記できるのだろうか？

提示された曲を、階名を用いて「ドレミ」として書き起こ

す。曲が階名として聞こえてくる聴取者は、その曲を移調しても、書き起こされる「ドレミ」列は変わらない。相対音感者である^(注1)。いくら「ド」の音高が変わろうとも、「ド」として表記する。第一著者は絶対音感を持っており、この階名での書き起こしが全くもって理解できない一人である。異なる音高に同一の音ラベルを振ることなど、全く理解不能である。

異なる話者間で「あ」の同一性が感覚できない人がいるのだろうか？感覚できない「機械」が、限られた話者の音声から構築された（特定話者）音声認識器である。そして、感覚できない「人」として、一部の自閉症者がいる^[1]。優れた絶対音感を持つ率が、健常者に比べて有意に高い自閉症者の中には、「ハ」の音（固定ドとしてのドレミの場合は「ド」の音）で始まらない「カエルの歌」を、それと認めない者もいる^[2]。

相対音感者による階名による書き起こしは、音階の構造（全全半全全半という音高遷移の枠組み）をメロディーの中に感覚し、例えば長調の場合、主音をドとして認識し、同様に、上主音、中音、下主音を、レミファ、として認識する。即ち、音列の流れを通して、全体的なメロディー構造（音楽学では、これを「横の構造」と言う）の認知が先に起こり、それに基づいて個々の要素音の（他音群との関係によって定まる）機能的・相対的価値を認識する訳である。その結果、要素音の絶対的物理特性とは全く独立に、個々の要素音が同定されることになる。物理的には全く異なる二音が同一の機能的価値を有した時、両者が同一音として「聞こえる」ことになる^{[3]、[4]}。

幼児は、極端に限られた音声の多様性に接することで、広範囲に渡る音声の多様性に対処できるようになる。発達心理学によれば、「幼児の音声言語獲得は、分節音の獲得の前に語全体の音形・語ゲシュタルトの獲得から始まる」とされている^{[5]、[6]}。個々の音韻意識が定着するのは小学校入学以降であり、それまでは「しりとり」に難を示す児童もいる^[7]。即ち幼児の音声コミュニケーション（例えば音声生成）は、個々の音韻（モーラ）を一つ一つ音に変換する形では（少なくとも意識の上では）行なわれていない。日本語には母音が5つあり、それが/あ/い/う/え/お/であることを知る以前に、幼児は両親と音声コミュニケーションを行ない、自己主張までする。

全体的なメロディー構造を通して、移調された曲同士の同一性を感覚し、更には、個々の要素音の（階名としての）同一性を感覚する。幼児の言語獲得も同様に、語全体の音形の獲得から開始される、とした場合に、この語全体の音響表象が、音楽同様、音声の移調（話者の違いなど）による多様性問題を解く鍵になるのだろうか？従来より筆者らは、非言語的要因による音響変動に対して簡素な数学モデルを考え、この問題を解決してきた^{[8]~[13]}。本稿ではこれを一般化し、非常に広範囲な変換においても、移調不変な構造的表象が存在することを数学的に示す。そして、自閉症・失読症者の音声認知を考察し、更に、話者不変の音シンボルである文字列を通して行なわれる現在の外国語学習の形態について、一つの検討を行なう。

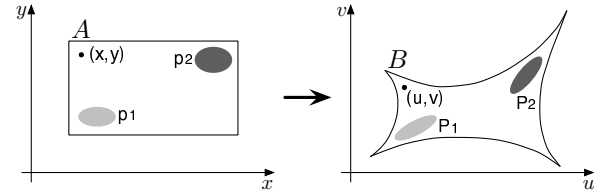


図1 一対一対応関係を有する二つの空間 A と B

2. 非言語的音響変動不変の音声の構造的表象

2.1 2つの空間における頑健な不変量

図1に示す様な、二つの空間 A と B を考える。両者には一対一の対応関係があり、空間 A のある点は空間 B の対応点へ写像され、逆もまた成立する。但し、その写像関数は明示的には与えられていない。以下、一般性を失わない範囲で2次元空間を用いて説明する。空間 A, B の点を (x, y) , (u, v) とし、両者の対応付け（変数変換）を一般的に下記の様に示す。

$$x = x(u, v), \quad y = y(u, v) \quad (1)$$

空間 A における事象を考える。但し、全ての事象は点ではなく、確率密度分布関数として存在するものとする。この時、A の分布 p は、下記式によって B の分布 P へ写像される。

$$P(u, v) = p(x(u, v), y(u, v)) |J(u, v)| \quad (2)$$

$J(u, v)$ はヤコビアンである。ヤコビアンを用いた積分変数変換を考えると、バタチャリヤ距離の空間写像不変性が導かれる。

$$BD(p_1, p_2) \quad (3)$$

$$= -\ln \iint \sqrt{p_1(x, y) p_2(x, y)} dx dy \quad (4)$$

$$= -\ln \iint \sqrt{q_1(u, v) q_2(u, v)} |J(u, v)| du dv \quad (5)$$

$$= -\ln \iint \sqrt{q_1(u, v) |J|} \sqrt{q_2(u, v) |J|} du dv \quad (6)$$

$$= -\ln \iint \sqrt{P_1(u, v) P_2(u, v)} du dv \quad (7)$$

$$= BD(P_1, P_2) \quad (8)$$

$q_i(u, v) \equiv p_i(x(u, v), y(u, v))$ である。上記式は、空間 A のバタチャリヤ距離と、空間 B のそれとの等価性を保証する。写像関数やヤコビアンを直接的に求めなくても、分布が正しく推定できれば、距離は不変である。この性質は、式(1)の空間対応付けに対して、強い制約を求めない（ヤコビアンによる変数変換が可能であればよい）ため、非線形変換を含む、幅広い変換群に対して距離不変となる。この性質は、カルバックライブラ距離、ヘリンジャ距離などでも成立する一般的性質である。

2.2 不変事象間距離から普遍的に存在する不変構造へ

ユークリッド空間に存在する n 点からなる幾何学構造は、 nC_2 個だけ存在する二点間距離を全て求めれば、（鏡像の曖昧性を無視すれば）その幾何学構造を一意に規定することになる。即ち、距離行列は幾何学構造を規定することになる。前節での証明により、空間に存在する N 個の分布群によって張られる距離行

(注1)：なお、階名の書き起こしが出来ない（ハミングしかできない）相対音感者もいる。この場合、彼らは「言語化が困難な」相対音感者である。

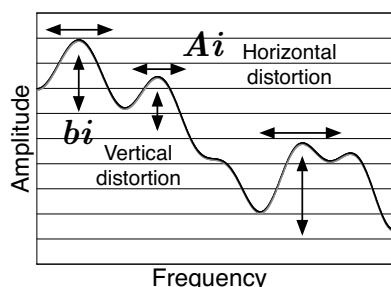


図2 スpekトルの水平・垂直歪みと一次変換

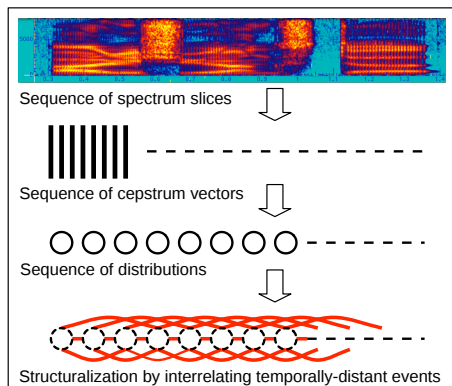


図3 音事象間の差のみを抽出して構成される不変構造

列、即ち、幾何学構造は、空間写像に対して一切不変となる。空間 A, B を二話者の音声音響空間とすれば、話者非依存の普遍的な不変構造が存在することになる（音響的普遍構造）。

この数学的性質は、非常に強力な枠組みとなると考えられる。従来筆者らは、英語学習者における英語母音群構造に対して、構造解析を行ってきた^{[12], [13]}。話者／マイクなどの不可避的な音響歪みを除去し、外国語話のみを構造歪みとして抽出することが目的であった。しかし、上記に示すように、本来の構造不変性は、非常に広い変換群において成立するため、学習者空間と教師空間との間に一対一対応があれば、外国語話を超えて、構造の不変性・同一性を約束することになる。しかしこの場合、空間 A でのガウス性分布が、空間 B では非ガウス性の分布として変換されるなど、分布形状の極端な歪みを生じることが予想される。例えば、変換後もガウス分布となるという制約の下で本数学的性質を使う等、積極的かつ妥当な制約導入によって、本性質は有効利用されると考える。逆に言えば、分布の推定を正確に行なうことができれば、それほど頑健な不変構造が、数学的には普遍的に存在する、ということである。

3. 音声の構造的表象に対する様々な実験的検討

従来より筆者らは、非言語的要因による音響変動をケプストラムの一次変換としてモデル化して議論してきた^{[8], [9]}。これは、特にスペクトルの水平方向歪みが $c' = Ac$ として、垂直方向の歪みが $c' = c + b$ として記述できることによる（図2参照）。この場合、ガウス分布はガウス分布へと変換されることになる。これに対して図3に示す様に、絶対項を全て捨象し、音事象間差異のみを求めることで、話者／マイク不変の構造的表象が得られる。実際に、孤立発声された5母音系列をタスクとした音声認識では（語彙数120の孤立単語認識に相当）、LPFなどの前処理が必要ではあったが、一人の話者の音声で不特定

話者音声認識が可能であることを示した^{[10], [11]}。この実験では、4,130人の話者の音声から構築されたHMMよりも高い頑健性を示した^(注2)。話者性を消去する、という方法論は、外国語発音学習支援にも応用されている。学習者と教師の発音を、体格／性別／年齢などを無視した形で、直接的・構造的に比較可能となっている^{[12], [13]}。発音ポートフォリオの提案、効率的学習のための教示生成、更には学習者分類などを検討している。

4. 幼児は親の声の何を模倣しているのか？

幼児は親の声の模倣を通して音声言語を獲得すると言われている（音声模倣）^[6]。しかし幼児は、親の声そのものを模倣しようとはしていない。太くて低い声を出そうと努力している幼児はいない。音韻意識が未熟である彼らは、個々のモーラ（話者非依存の音シンボル）を一つずつ生成する、という術は少なくとも意識的には不可能である。となると彼らは親の声の何を模倣しているのだろうか？九官鳥による音声模倣では、話者性までも真似ることが知られている^[15]。優秀な九官鳥は、その音声模倣を聞いただけで飼主が分かるが、どんなに優秀な幼児の音声模倣を聞いても、飼主（親）の同定は不可能である。音響的な音声模倣と言語的な音声模倣は何が違うのだろうか？

「幼児の音声言語獲得は、分節音の獲得の前に語全体の音形・語ゲシュタルトの獲得から始まる」との主張が正しければ^(注3)、この語ゲシュタルトの音響的実体には、非言語的情報は含まれないはずである。筆者らはこの語ゲシュタルトの音響的定義について、多くの発達心理学・言語獲得研究者に問いかけてみたが^[16]、残念ながら、明確な答えは得られなかった。

近年の脳科学の進歩により、聴覚音声学の議論は、蝸牛から、聴覚皮質のモデリングに移行しつつある^[17]。脳科学における多くの知見は偶然によって齎されている^{[18], [19]}。交通事故や、一部の医師の不適切な処置が原因で、不幸にも脳損傷を負った患者を通して多くの知見が得られている。動物実験においても、偶発的な刺激提示によって重要な知見が得られている。前頭葉、海馬、扁桃体の機能、更には、ミラーニューロンなどは良い例である。脳は研究者の机上の議論を超えた処理を行なっている、と解釈することもできる^[19]。さて、聴覚皮質モデリングであるが、視覚皮質のような定説が存在する状況には無いが、幾つか興味深い主張がある。まず「音声の言語的情報と非言語的情報（話者の情報）は分離されて処理されている」との主張である^{[17], [20], [21]}。特に[21]では、音楽と音声とを対比し、音声の言語情報は、音声の動き（speech motions）によって伝搬されると主張している。音楽で言えばメロディーである。一方「話者の同定は音楽で言う楽器の同定に相当し、それは時不変の情報として処理される」と主張している。図3に示した音声の構造的表象は、音声を「音の運動」と考え、その運動（コントラスト）成分のみを抽出する形となっている。即ち音声から「音

(注2)：連続発声を対象とした分布列推定方法がまだ確立できていないため、孤立発声母音系列という、人工的なタスクを用いた。

(注3)：なお、日本人乳児が[r]と[l]を弁別できることが知られているが、弁別ができるのであって、[r]を/r/として同定している訳では無い。同定能力の獲得の前に、まず、弁別・区別、即ち差異の知覚が可能になることは重要である。

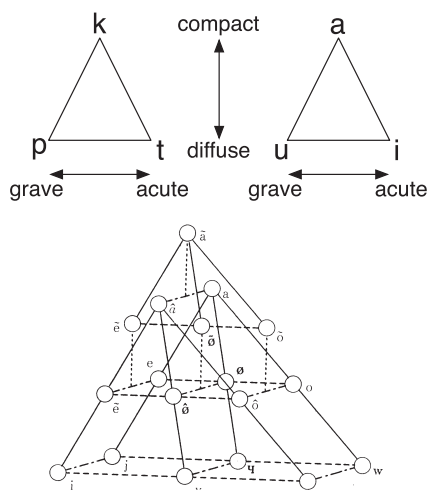


図4 母音・子音三角形^[24]と仏語母音群構造^[25]

であること」を一切捨てた物理表象である。何が動いているのかは不明である。「動きだけを抽出した時に、話者／年齢／性別を超えた頑健な不変表現が数学的に入手できる。それこそ言語である」と主張するのが音響的普遍構造である。

近代言語学の祖ソシュールが一世紀以上も前に興味深い主張をしている^[22]。The important thing in the word is not the sound alone but the phonic differences that make it possible to distinguish this word from all others. 即ち、音ではなく、音的差異の重要性を説いている。差異を捉えることで単語が同定できる、との主張である。彼はまた Language is a system of conceptual differences and phonic differences. と主張している。「言語＝差異・動きのシステム」である。分布間差異を集めたものが頑健な不変構造を成し、それをういた語同定が可能である。この不変構造こそ語ゲシュタルトではないだろうか。

音声伝搬の情報は、言語／パラ言語／非言語情報と分類される。各情報を担う音響量に着眼すると、言語及び非言語情報は声道情報となるため、スペクトル包絡に相当し、パラ言語情報は音源情報となるため、 F_0 、パワー、継続長に相当する。即ちソース・フィルタの分離である。幼児の音声模倣は、親の声からまず非言語情報を分離すると考えられる。音声＝[言語＋パラ言語]＋[非言語]という枠組みである。しかし、音声科学・工学が構築した枠組みは、音声＝[言語＋非言語]＋[パラ言語]という枠組みである。調音音声学の価値観に基づけば、声道と音源を分離する、自然かつ妥当な枠組みである。しかし、音声コミュニケーションの観点から考えると、この枠組みでは幼児の音声模倣問題は解けない。非言語情報を頑健に分離する術が無いからである。人と音声との遭遇は聴取であって、生成ではない。しかし、科学は音声と生成を通して遭遇した、というのは歴史的事実である^[23]。音声科学が実験科学である以上、それは時代の技術的制約の下で議論を重ねなければならない。聴覚音声学は観測技術の未熟さから、調音・音響音声学と比較して、その進展が遅れざるを得なかった。脳科学の進展によって「話されたもの」としてではなく「耳に届くもの」として音声に焦点が当たった時に、従来の枠組みでは想像できない情報処理を脳が行っていたとしても何ら不思議ではない。

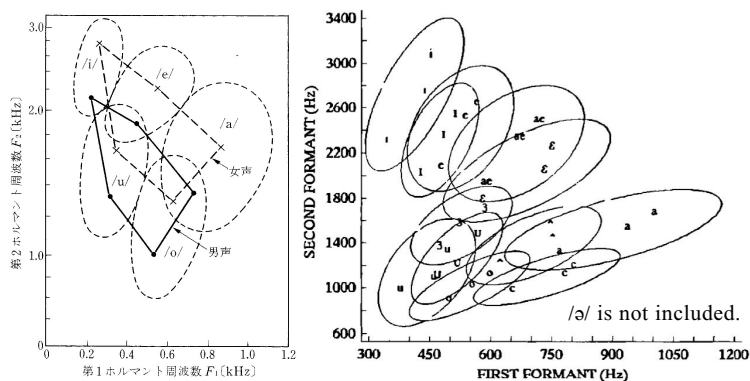


図5 日本語5母音と米語12母音の F_1/F_2 図^{[31], [32]}

ソシュールの phonic differences という言葉はやがて、ヤコブソンの弁別素性、即ち構造音韻論へと引き継がれる。音そのものではなく、音/x/と音/y/はどう違うのか、その違いを定性的に表現するために弁別素性が使われた(図4参照)。つまり、音素群が成す外部構造の議論である^[25]。やがて、弁別素性はそれが束となって音素を表象する、即ち音素の内部構造の議論に使われるようになる^[24]。この素性の束としての音素は音楽の「和音」をメタファーとして生まれた^[26]。和音＝音素、音符＝素性、である。音楽学では、音楽には横の構造(メロディー構造)と縦の構造(ハーモニー構造)があると説く。和音は縦の構造であり、ヤコブソンの素性による内部構造の議論は音楽の縦構造を発端としている。一方、筆者らが提唱する音声の構造的表象は当然のことながら、音楽の横構造に相当する。縦と横の構造、どちらが音楽にとってより本質的かと言えば、当然、横構造である。和音の無い音楽はあれど、メロディーの無い音楽は存在しない。筆者らの知る限り、音声工学において横構造の議論が皆無であり、縦構造の議論^{[27]~[29]}が多い理由は、音声の表記方法に起因すると考える。音声(話者非依存)の音シンボル列として表記し、音声をシンボルに対応させて区切れれば、その時点で横構造は消失してしまう。例えば Bloomfield は、Writing is not language, but merely a way of recording language by visible marks. と述べているが、文字言語は本来音声言語の副産物であり、5 万年以上の音声活動を通して人類が造り出した「音声言語の視覚化技術」でしかない^(注4)。即ち「結果」であって「原因」ではない。しかしその「結果」を「前提・原因」として捉え、音声の物理現象を切り刻んで来た、という歴史は否めない。そもそも科学は要素還元主義の上に構築され、その枠組みの限界が指摘されたのはごく最近である^[30]。

5. 音韻の意識は音声言語運用に必要なのか？

提案している音声表象は音声ストリームをメロディとして捉え、その横構造を頑健な不変項として導出している。孤立発声母音の系列という人工的なタスクではあるが、個々の音事象の絶対的物理特性は一切用いずに、個々の音事象の同定は一切行わずに、単語の同定が可能であることを示した^[10]。音楽の場合、横構造を通して各音事象の機能的・相対的価値を感覚

(注4)：文字起源は象形文字であるため、本来文字は意味の視覚化技術であって、音声の視覚化技術ではない。表音文字は単なる借り物技術でしかない。

し、「ドレミ」が聞こえてくる。音声の音韻知覚も同様の枠組みとして捉えられないだろうか？音声の相対音感である。

音楽の場合、如何なる鍵盤（即ち物理特性）も「ド」になれる。しかし、音声の場合、例えば F_1/F_2 図において任意の点の音を「あ」と知覚できるか、と考えれば、それは困難である。**図5**に示すように、日本語の場合、男女を考慮しても凡そ5母音は分離している。音韻とその物理実体との対応が凡そ一対一、絶対音感的である。しかしフォルマント周波数が発声者の声道長に依存していることを考えれば、例えば目玉親父やウルトラマンが日本語母音を発声した場合を考えれば、これらの分布群は大きな重なりを呈することになる。一対一対応の崩壊である。このような場合、音の絶対量に基盤を置く処理系は機能せず、音と音の相対量に基盤を置いてこそ、頑健な処理系が期待できる。アニメの世界では音声は相対音感的でなければならない。

アニメの世界を想像しなくても、相対音感の世界を創成することは可能である。母音の数を増やせばよい。**図5**には米語12母音の F_1/F_2 図についても示している。成人男性・女性・子供（10～12歳）139名の/h V d/から得られた結果である^[32]。なお、音質が容易に変動する/ə/はこの図には含まれていない。これだけの重なりは、複数話者のデータを同時に表示するから生じるのであり、話者別に示せば当然重ならない。この事実を顧みずに、母音毎に、複数話者データに対して物理的な絶対量を統計的にモデル化しても、母音認識は困難となる。日本語は絶対音感的、米語は相対音感的なのだろうか？何れの場合も、個々の音はシンボル化される。音楽の場合でもドレミは階名としても（移動ド）、音名としても（固定ド）使われている。

相対音感者の多くは、言語化できない相対音感者である。メロディーの記述を、音名／階名で行なうのではなく、「ラ〜ラ」即ちハミングで行なう相対音感者である。彼らも主音は認知しており^[3]、音楽の横構造を認知しているが、主音に対して「ド」を対応させることが困難である。そもそも音高（基本周波数）と「ドレミ」という声（スペクトル包絡）とは無関係であり、これを恣意的に結びつけたのが階名である^(注5)。音声に対して「言語化できない相対音感者」とはどのような存在になるのだろうか？曲を「ラ〜ラ」として再生する際に頻繁に移調されることを考えれば、ある話者の発声を移調して再生することは、「繰り返し発声」に相当する。一方、曲を「ドレミ」に落とす作業はどうなるであろうか？スペクトル特性とは全く関係の無い、「声」に対して恣意的に関連付けられた「モノ」を考えれば明らかなように、それは「(表音)文字」に落とす作業となる。以上の考察から得られる帰結は「相対音感的な音認知が不可避免的に要求される言語の場合、文字の読み書きに困難を覚える人が多い」となるが、こんな考察、意味があるのだろうか？

第一著者は、このような無意味かもしれない考察の最中に失読症 (dyslexia) を知った。「頭が良いのに、何故か本が読めない」方々である^{[33]~[35]}。具体的な症状は様々であるが、共通項として存在する症状が音韻意識が希薄、即ち、単語音声に対して、それを個々の音に分割したり、個々の音が連結して単語音

声になる、ということを感じることが困難な方々である。幼児の音声認知をそのまま引きずっている方々である^[33]。個々の音をカテゴリとして知覚するのが困難である一方で、異音の区別は健常者よりも成績が良い^[35]。これは [r] を /r/ というカテゴリとして同定できないが、[r] と [l] が区別できることに相当する。米国では約20%の人が失読症である^[33]。政治家、作家、起業家、学者にも失読症者はおり、グラハム・ベルもその一人である。彼に音声認識・合成器を作らせても、音シンボル列と音声間の変換技術など作ら（れ）なかったはずである。「そんなモノの上に言語は出来ていない」と主張したであろう。幼児の言語獲得は、彼らの認知能力の未熟さが、個々の音韻を意識させないのだろうか？無意識下では音韻を操作しているのだろうか？或いは、個々の音韻意識は音声言語運用に不要なのだろうか？「音声 ↔ シンボル列変換」に難を示す多数の音声ユーザの存在を、音声科学・工学者はどう考えるべきなのだろうか？

6. 究極の音声絶対音感者と音声言語

極端な絶対音感を持つ奏者は、オーケストラ／ホールが変わる度に十分な耳慣らしが必要となる。基準音がオケ／ホールによって、数Hz異なるからである。参照パターンとして絶対項を持ってしまうと（例えば、基準音＝440Hz）、環境の変化に対して柔軟に対応できなくなる。音声の極端な絶対音感者は二話者の「おはよう」という発話の同一性の認知が困難になると考えられるが、自閉症者の中に、母親の音声のみ言語メッセージになる者がいる^[1]。自閉症は端的に「関係の病」と言われるように^[36]、入行情報の整理整頓が困難であり、個々の要素的事象を丹念に記憶する。電話番号、住所など互いに無関係なものを膨大に記憶する一方で、物事の因果関係や複数の刺激群が成すパターンの抽出、事象の抽象化に困難を示す。そのため、目の錯覚などが起き難い。顔の要素的特徴を覚える一方で、顔を見て表情や話者が同定できない。優れた音感を持ち、絶対音感者が多い。一言で言えば、ゲシュタルト知覚が困難である^[37]。

第一著者にとって「ドレミ」とは音名であるため「曲が階名として聞こえる」という事実は想像を絶する。「ソ」が「ド」と聞こえる、というのは「え」が「あ」と聞こえる、というのに等しい。勘違いか錯覚の類いではないか、とさえ考えることもある^(注6)。極端な音声の絶対音感を持つと考えられる自閉症者にとって、物理的に異なる特性を持つ話者Aの音と話者Bの音を「同一音」として認知する健常者の感覚こそ、想像を絶するものであると推測する。勘違いか錯覚の類いではないか、と主張しても不思議ではないが、彼らは口を開かない。何故なら、極端な絶対音感を持つ自閉症者は、音声言語を持たないからである。二話者の「おはよう」の同一性が認知できなければ、音声言語が破綻するのは自明である。音声言語は、ある種の勘違い・錯覚の上に構築されている、と考察することもできる。自閉症者の中には、文字言語を通して言語コミュニケーションを開始する場合もある^{[1]・[38]}。音は全て聞こえているにも拘らず、聞こえ過ぎるが故に、文字（図形）言語が第一言語となる。

(注5)：「ドレミ」という命名は僧侶の名前の第一音節から来ている。

(注6)：実際には「ド」の意味が異なるので、勘違いでも錯覚でも無いのだが。

音響空間を N^3 の部分空間に分け、各々の独立性を仮定してモデル化するのが triphone である。環境が変わる度に耳慣らしを行なうのが音声認識装置である。フレーム問題に端を発してロボットと自閉症児との類似性が議論されているが^[39]、音声認識装置も自閉症と類似している。言語障害の無い自閉症をアスペルガー症候群と言うが、彼らの一見奇妙な言語活動は、音声対話システムのそれと類似していることを指摘しておく^[40]。

7. 音声の一次元視覚化技術の習得と外国語学習

多くの場合、外国語を学習する学生は、母国語音声を読む／書く能力を習得済みである^(注7)。即ち、音声の一次元的視覚化技術を習得した方々である。音声学や音声工学では第4節で述べたように、音声の物理現象を、音シンボル表記を基に頻繁に切り刻む。これは、音声に含まれる言語情報の視覚化技術として、文字表記の物理的妥当性を前提にした議論である。「切り刻まれた音声区間に存在する物理現象が、各々単独で、対応する音韻を聴取者に知覚させる」という仮定を是とした議論である。当然音シンボルに対応する物理現象は多様を極め、「数万という話者から物理現象を集め、分布としてモデル化する」という枠組みに落ち着く。それが現在の音声科学・工学の姿である。

本稿では、これとは別の形で音声を議論する枠組みを提供している。音声の一次元的視覚化技術には、そもそも物理的かつ致命的なバグがあるのでは、という主張である。そして提案する枠組みでは、言語獲得における「多様性の謎」は謎でなくなる。バグがあれば、そのバグに「音声言語の正規ユーザ」が悩んでいても、何ら不思議ではない（失読症）。バグがあれば、音シンボルの物理的対応物の不変性を信じて、その物理対応物を絶対的に記憶する方々が音声言語運用に悩んでいても、何ら不思議ではない（自閉症）。音声は運ぶ情報のうち、言語情報のみを物理的に表象する場合、それは1〜3次元の貧相な空間で表象できるのだろうか？本来音声の言語情報は視覚情報（貧相な空間で表象できる物理実体）ではない。この表象が本質的に困難であった場合、文字はその時点で落第となる。一般的に人間は、4次元以上の空間認知が困難だからである。興味深いことに失読症者は、その空間認知能力が高いことで知られている^[41]。

音楽における五線譜は、階名（機能的価値）よりも音名（物理的価値）の表象を目的とした技術である。音楽教育では、機能的価値の感覚こそ第一と考え、五線譜とは異なる表記法を用いた音楽教育を実践する教育者もいる^[4]。文字の場合、音名よりも階名としての意味合いが強い。文字（音韻）は心的表象である。心的表象であれば、それは物理実体を持たない、即ち、（物理的には）錯覚や勘違いの産物、と解釈できる^[42]。それを視覚化したのが文字であるが、文字そのものには、その心的表象の認知に至る物理的プロセスの記述が欠落している。外国語学習において必要なのは、音楽教育者の一部が模索しているように、そのプロセスを明らかにし、そのプロセスを自然と誘発する視覚刺激としての文字を開発することでは無いだろうか？

(注7)：音声言語は生得的であるが、文字言語はあくまでも学習によって習得される能力である。文字の無い言語は多数存在する。

8. ま と め

音響の普遍構造が頑健な変換不変性を有することを数学的に示し、相対音感としての音声認知を通して、言語獲得、失読症、自閉症を考察した。また、外国語学習における音声の一次元視覚化技術としての文字に焦点を当て、これについても考察した。なお、本考察がこれら障害の全容を網羅している訳ではなく、例えば失読症と自閉症の合併例が存在するのも事実である^[43]。しかし筆者らは「音声・シンボル列」変換を物理的前提とする従来の音声研究の方法論を懐疑的に受け止めている。この前提は公理ではなく、単なる、一つの後付けの解釈でしかない。

文 献

- [1] 東田他、この地球にすんでいる僕の仲間たちへ、エスコアール出版社 (2005)
- [2] 奥平、自閉症の息子ダダくん 11 の不思議、小学館 (2006)
- [3] 谷口、音は心の中で音楽になる、北大路書房 (2003)
- [4] 東川、読譜力ー「移動ド」教育システムに学ぶ、春秋社 (2005)
- [5] 加藤、コミュニケーション障害学, 20, 2, pp.84–85 (2003)
- [6] 早川、月刊言語, 35, 9, pp.62–67 (2006)
- [7] 原、コミュニケーション障害学, 20, 2, pp.98–102 (2003)
- [8] 峯松他、信学技報, SP2005-12, pp.1-8 (2005)
- [9] 峯松他、信学技報, SP2005-131, pp.121-126 (2005)
- [10] 村上他、信学技報, SP2005-14, pp.13-18 (2005)
- [11] 村上他、信学技報, SP2005-130, pp.115-120 (2005)
- [12] 朝川他、信学技報, SP2005-24, pp.25-30 (2005)
- [13] 朝川他、信学技報, SP2005-156, pp.37-42 (2006)
- [14] 峯松他、信学技報, SP2004-27, pp.47-52 (2004)
- [15] 宮本、音を作る・音を見る、森北出版 (1995)
- [16] N. Minematsu, *et al.*, “Universal and invariant representation of speech,” Proc. Int. Conf. Infant Study (2006)
- [17] 柏野、月刊言語, 33, 9, pp.102–107 (2004)
- [18] M. Spitzer, 脳・回路網の中の精神、新曜社 (2001)
- [19] 茂木、心を生みだす脳のシステム、日本放送出版協会 (2001)
- [20] K. S. Scott *et al.*, Trends in Neurosci., 26, pp.100–107 (2003)
- [21] P. Belin *et al.*, Nature Neurosci., 3, 10, pp.965–966 (2000)
- [22] F. D. Saussure, Course in general linguistics, McGraw-Hill Humanities/Social Sciences/Language (1965)
- [23] 前川、音声研究, 8, 3, pp.35–40 (2004)
- [24] R. Jakobson *et al.*, Preliminaries to speech analysis, MIT Press, Cambridge, MA (1952)
- [25] R. Jakobson *et al.*, Notes on the French phonemic pattern, Hunter, N.Y. (1949)
- [26] S. E. Blache, The acquisition of distinctive features, Univ. Park Press (1978)
- [27] K. N. Stevens, J. Phonetics, 17, p.3–45 (1989)
- [28] L. Deng *et al.*, Speech Comm., 33, 2–3, pp.93–111 (1997)
- [29] M. Ostendorf, Proc. ASRU, pp.79–84 (1999)
- [30] M. M. Waldrop, 複雑系、新潮社 (2000)
- [31] R. K. Potter *et al.*, JASA, 22, 6, pp.807–820 (1950)
- [32] J. Hillenbrand *et al.*, JASA, 97, 5, pp.3099–3111 (1995)
- [33] S. Shaywitz, 読み書き障害（ディスレクシア）のすべて〜頭はいいのに本が読めない〜、PHP 研究所 (2006)
- [34] 石井、科学技術政策研究所・科学技術動向 45, pp.13–24 (2004)
- [35] W. Serniclaes *et al.*, Cognition, 98, pp.B35–B44 (2005)
- [36] 酒木、自閉症の子どもたち、PHP 研究所 (2001)
- [37] U. Frith, 自閉症の謎を解き明かす、東京書籍 (1991)
- [38] R. Martin, 自閉症児イアン物語、草思社 (2001)
- [39] 渡部、鉄腕アトムと晋平君、ミネルヴァ書房 (1998)
- [40] 泉、僕の妻はエイリアン、新潮社 (2005)
- [41] B. Hornsby, 読み書き障害の克服、協同医書出版社 (1995)
- [42] A. J. Lotto *et al.*, Chicago Linguistic Society, 35, pp.191–204 (2000)
- [43] 月文他、自閉症者からの紹介状、明石書店 (2006)