

孤立音を聞いて音韻同定できる能力は音声言語運用に必要か？

峯松信明, 西村多寿子, 朝川智, 喬宇 (東大), 櫻庭京子 (清瀬市福祉センタ)

1 はじめに

一つの言語には数十種類の音素がある。しかし音素の音的実体は前後の音素環境などによって変形し、異音と呼ばれる。音素と比較して種類数も多く、より具体的な音的現象に対応している。しかし奇妙なことに、これら音的事象を記号を用いて記す場合、性別、年齢、収録・伝送機器特性などの非言語的要因による音の変形は一切無視される。その音響的変形が幾ら大きくても、である。音声認識の音響モデリングは、凡そ、異音に相当する音事象をモデル化しているが、「非言語的変形の無視」を実装するために、数千～数十万人の話者から、様々な環境で収録した音サンプル群を統計的に分布としてモデル化している。本稿では、「非言語的変形の無視」の実装は、集めることではなく、音事象間の差異を捉えることで可能となることを数学的に示し、極めて少数の話者の音声で、不特定話者音声認識が可能であることを示す。提案する枠組みでは（音的要素ではなく）音的差異に着眼し、差異の集合で規定される全体的な音的構造をモデル化する。構築したシステムは孤立音を音韻として同定することは出来ない。構造が無いからである。一方人間は孤立音に対する音韻同定は通常可能であるが、それが困難となる状況は容易に作り出すことができる。また、音声コミュニケーションに難は無いが、発話を音韻列として認知困難な方々は数多く存在する（発達性ディスレクシア）。本稿では、孤立音を聞いて音韻同定する能力は、音声言語運用の必要条件では無いとの主張を行なう。

2 「集めること」は必要なのか？

35 万人の話者の声を用いて構築された各異音の統計モデルに基づく音声認識エンジンが市販されている⁽¹⁾。一方人間を眺めた場合、35 万人の声を聞いて初めて頑健な音声情報処理が可能となった個体は恐らく存在しない。子供の言語発達を考えた場合、乳児の聞く声の大半は母親、父親の声であり、自らが話せるようになると、聞く声の半分は自らの声となる (speech chain)。話者バランスのとれた大規模音声コーパスを要求する現在の音響モデリング技術は、些か不可思議な技術体系である。

現在の音声認識技術は、音響モデリングに関して言えば、与えられた「音」を学習し、「音」のモデルを構築し、新たに与えられた「音」が、以前与えられた「音 (カテゴリ)」の何れかを推定する技術である。音声認識技術における「音」そのものをモデル化する枠組みは、音響音声学が、単音 [a] 或いは音素/a/を物理的に規定する場合に、その音的実体（フォルマント周波数など）を用いることに起因する。声道長などの要因によりこれらは変動するため、結局「音を集める」方策を採択してきた。

幼児の言語獲得は音声模倣という言葉で表現される⁽²⁾。他の霊長類には無い行動であると、指摘されている⁽³⁾。しかしこの時、幼児は「音」を学習しない。両親の「声」を模倣しようとししない。一方九官鳥は「声」を真似る。車、ドア、動物の鳴き声など「音」を真似る。優秀な九官鳥は聞けば飼い主が分かる⁽⁴⁾が、どんなに優秀な幼児を聞いても、父親は当てられない。九官鳥は「音」を学習し、（恐らく「音」のモデルを構築し）以前聞いた「音」に反応し、その「音」を鳴管を使って生成する。幼児の模倣を「音声」模倣と呼ぶならば、九官鳥の模倣は「声」模倣である。「声」は音そのものである。では、「音声」とは音の何を指すのだろうか？本節では以下、音声と声を特に区別して記述する。

現在の音響モデリングは、幼児と九官鳥、何れに似ているだろうか？間違いなく、九官鳥、である。「音声」認識とは名ばかりで、実は「声」認識というのが、その実体である。「声」は如何にしたら「音声」となるのか、音声工学（音声科学も含まれるであろう）が導いた方程式は下記である。

$$\text{音声} = \sum_{\text{話者, マイク, 場所}} \text{声} \quad (\text{即ち「集めることで音声となる」との主張である})$$

しかし、幼児は 35 万人の話者の「声」、及びその書き起こしテキストがなくても、電話越しのお婆ちゃんと会話し始める。しかも、電話声を模倣しようとせず、自らの「声」で会話する。

「声」を音事象の連続体（ストリーム）と考えた場合、個々の音事象は、話者・収録機器の違いで当然変形を被る。音韻意識が未熟な幼児は、両親の「声」を仮名列に落とし、個々の仮名を音にして出すという技は使えない。変形を被った音ストリームに対し、彼らが模倣しているのは「音」ではない。彼らが模倣している、音ストリーム内に符号化されているコンテンツ（つまり音声）を直接的にモデル化したいのであれば、音そのもの（つまり声）に対する音響モデリング技術は甚だ不適切である。母親の「おはよう」を模倣しても、父親の「おはよう」を模倣しても、幼児の同じ「オハヨウ」となる。結局、この三者に共通して存在する話者不変の音響現象をモデル化する必要性が生じる。



図 1: 移調前後のメロディー (ハ長調とト長調)

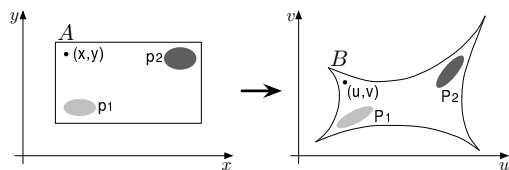
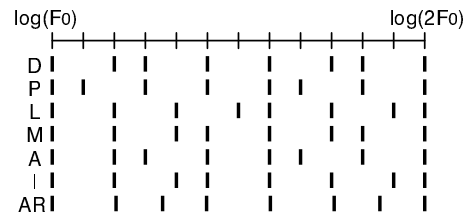


図 4: 一対一対応関係を有する二つの空間 A と B



D=Dorian, P=Phrygian, L=Lydian, M=Mixolydian
A=Aeolian, I=Ionian, AR=Arabian

図 2: 6 種類の古典的教会音階とアラビア音階

3 「音」の何をモデル化すべきなのか？

声を音韻列として認知困難な幼児が、如何にして母親の「おはよう」と父親の「おはよう」の（類似性ではなく）同一性を感覚するのだろうか？二つの発声の同一性を検討する場合、各発声を音韻列として表象し、音韻表象としての比較を通して初めて同一性が保証されるとすれば、聞き手の音韻意識が希薄な場合（幼児や発達性ディスレクシア）、「おはよう」と「おはよう」の意識的な同一性認知は困難となるはずである。しかし、（一部の自閉症者を除き）そのような事例を筆者らは知らない。

発達心理学は「幼児は単語全体の語形・音形（語ゲシュタルト⁽²⁾）を獲得し、その後、個々の分節音を獲得する」と主張する⁽⁵⁾。上記はこの話者不変と思しき「語ゲシュタルト」の音響的定義を問うている。音韻表象を用いずに同一性を保証する方法論を問うている。筆者らの一部は言語獲得研究者に広く問いかけたが⁽⁶⁾、適切な回答は無い。「惑星」の定義を求めずに議論を繰り返した天文学と同じである、との意見も得た。その物理的存在は議論せず、存在を仮定した議論が繰り返されている。

そもそも、二音の同一性を感覚するのに、両者の物理的同一性は必要なのだろうか？人間は他の霊長類と異なり⁽⁷⁾、（ある環境下では）物理的に異なる二音の同一性を感覚できる。相対音感である。

3.1 調不変のドレミ同定 ～言語化可能な相対音感～

図 1 に示す二つの曲（上曲を移調したものが下曲）をドレミに落とすよう依頼した場合、どのような反応が考えられるだろうか。返答は三通りある。「初めはソーミソドー、次がレーシレソー」と答えた場合、その人は絶対音感者であり、この場合ドレミは音名である。「両方ともソーミソドー」と答えたとすれば、その人は言語化可能な相対音感者であり、この場合ドレミは階名である。「ラーララーとしか歌えません」となった場合、その人は、言語化できない相対音感者である。

言語化可能な相対音感に着眼する。この場合、調を幾ら変えても「ソーミソドー、と聞こえてきます」と彼らは主張する¹。彼らは、何故、音高の異なる音を「ド、と内なる声が聞こえる」と主張する程に、その同一性を感覚するのだろうか？この認知プロセスの必要条件の一つとして、調不変の音階構造（音配置構造）がある⁽⁸⁾。西洋音楽（平均律）では、1 オクターブ ($\log(F_0)$ から $\log(2F_0)$ に渡る音高帯域) を 12 個の音程に区分する (12 半音)。 $\log(F_0)$ が第 1 音であれば、 $\log(2F_0)$ は第 13 音となる。長調と呼ばれる音階は、1 オクターブを「全全半全全全半」という音程に区分して 8 音を配置する。これが「ドレミファソラシド」である。上記音程が満たされさえすれば、各音の絶対的な音高に意味はない。個々の音には機能名があり、第 1 音＝主音、第 3 音＝中音、第 5 音＝属音、などと呼ばれ、ドミソ、はそのニックネームである。これが階名の定義である。彼らはこの音の機能・価値を感覚して、ドレミが聞こえてくるのである。移調したところで音配置構造は不変であるため、ドレミ列は変わらない。長調の曲は、オクターブ等価性を前提にすれば、原則的に上記 8 音で構成されている。極端な場合を考えると、メロディーの中の任意の 2 音が、三全音を音程（音高差）として持つ場合、その 2 音に対して「ファとシ（或いはシとファ）」が聞こえてくる⁽⁹⁾。調不変の音高差異に基づいて要素音の同定を行なうのが、言語化できる相対音感者である。彼らが超頑健な要素音同定を行なえるのは、個々の音の絶対的な物理特性など、（少なくとも意識的には）記憶しないからである。

さて、この音配置構造が崩れるとどうなるのだろうか？古典的教会音楽には、種々の音階がある。図 2 のイオニア音階、エオリア音階が現代音楽の長調、短調として生き延びている。これらの音階では 12 半音の原則は守られており、5 全音と 2 半音の配置の違いとなっている。さらに 12 半音の原則までも壊すとどうなるだろうか？図 2 にはアラビア音階も示している。12 半音では表現できない音が要求されるため、通常のピアノでは再生できない。西洋音楽をアラビア音階で再生した場合、言語

¹ 声を出さずに「ソーミソドー」と発声した時と全く同一と思われる感覚が自動的に再生される、と言う主張である。

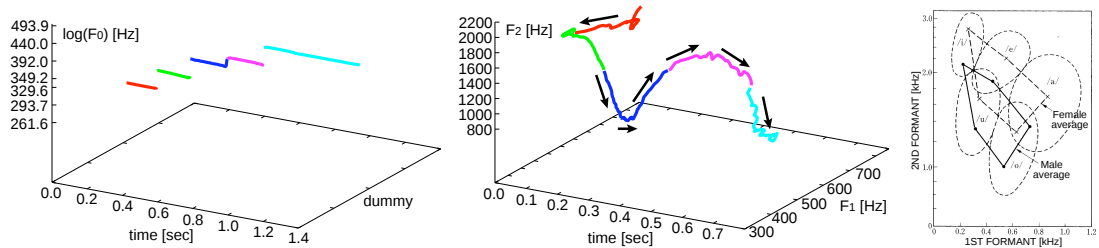


図 3: F_0 の動的変化としての CDEFG と音色の動的変化としての/aieuo/, 及び、日本語母音図

化できる相対音感者は「ドレミが聞こえてくるところと、聞こえて来ないところがある」という反応を示す。彼らの言語化は、音配置の様子に依存し、個々の音の音高には全く無関係に行なわれる。しかし逆に、孤立音の言語化は不可能である。メロディーという全体像があって初めて要素音のシンボル化が可能となる。音シンボルが先にあって、それらを並べてメロディーが構成されるのではない。

3.2 音高の動的変化と音色の動的変化

主旋律（メロディー）のみを対象とすれば、音楽は F_0 （ピッチ）の動的変化パターンである。音声として母音列のみを対象とすれば、これは音色の動的変化パターンである。母音の生成は声道（音響管）の共鳴現象であり、これは、管楽器における音生成と物理的には等価である。即ち「あいうえお」の違いは、声道形状の変化による共鳴現象の変化である。音楽学では音色はしばしば「基本音及び各倍音に対するエネルギー分布（配分）」として定義されるが、これはスペクトル包絡と同値である。結局、音色を表現するための最も簡素な物理パラメータはフォルマント周波数となり、ここでは F_1 と F_2 を考える（十数次元のケプストラム係数を考えても下記の議論は成立する）。なお母音同様、複数の管楽器を F_1 - F_2 平面上にプロットし、音色配置を示す場合もある⁽⁸⁾。図 3 に F_0 の動的変化としての CDEFG、及び音色の動的変化としての/aieuo/を示す。前者を移調しても、この動的パターンは上下に移動するだけであり、階名同定が要求する音群配置は不変である。一方/aieuo/の動的パターンであるが、日本語母音図（図 3）に示すように、音響音声学では、 F_1 - F_2 平面で男声の母音構造を移動すると女声の母音構造に重なると言われる。このような単純な写像で変換できれば、母音構造の話者不変性は容易に実現できるが（即ち二次元の移調＝平行移動）、厳密にはこのような単純な写像で変換できる訳では無い。音声合成の話者変換技術は、話者 A の音響空間と話者 B の音響空間との対応付け（写像）を精密に定義することで実装されるが、音群構造の不変性は、この両空間における不変構造を要求する。逆に言えば、線形・非線形を問わずあらゆる写像に対して不変なる構造が定義できれば、「音色の相対音感」は議論可能となる。なお、三角形は三辺の長さを規定すればその形状が一意に定まるように、 N 角形の場合、全ての二点間距離（距離行列）を規定すればその形状は一意に定まる。即ち、不変なる構造は、不変なる差異（群）の存在を証明することで立証される。

4 非言語的音響変動不変の音声の構造的表象

4.1 2つの空間における頑健な不変量

図 4 に示す、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。空間内の全ての点の写像が一つの式で表現される必要も無い。さて、空間 A の分布 p_i は空間 B の分布へと写像されるが、これを P_i とする。この時、次の等式が常に成立する⁽¹⁰⁾。

$$\iint_A \sqrt{p_1(x,y)p_2(x,y)} dx dy \equiv \iint_B \sqrt{P_1(u,v)P_2(u,v)} du dv$$

この量に対して $-\log$ をとったものがバタチャリヤ距離（分布間距離の一つ）である。結局、バタチャリヤ距離は任意の二空間（話者）間で常に等しい。この距離（差異）不変性は、空間写像の種類に依らず、また、カルバックライブラ距離などでも成立する一般的性質である（頑健な不変性）。

4.2 不変事象間距離から普遍的に存在する不変構造へ

頑健に変換不変な距離尺度を用いて、ある発話を変換不変的に表象することを考える。図 5 に示すように、音声ストリームを分布系列へと変換した後に（系列長 $= N$ ）、時間的に離れているものも含め、全ての二分布間距離を求めて $N \times N$ の距離行列として表象する。この時、個々の音響事象の絶対的な物理特性は全て捨象する。距離行列は一つの幾何学的構造を規定するが、この構造が変換不変と

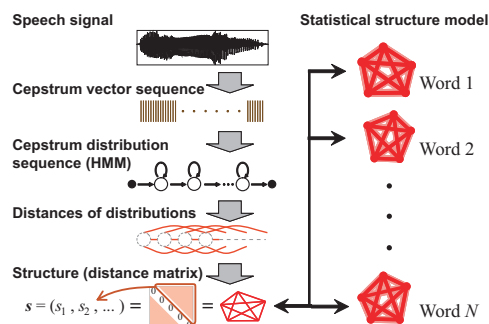
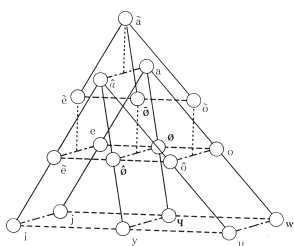
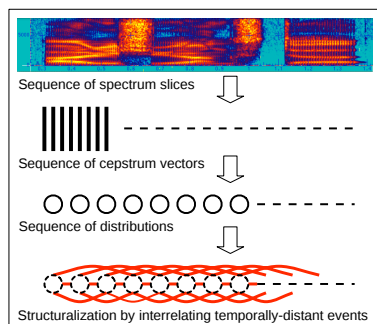


図 5: 話者に不変な構造的表象

図 6: ヤコブソンの音韻構造

図 7: 構造的な音声認識

なる。この構造は、例えば $m+1$ 次元の音響パラメータ時空間に存在する音色の動的変化パターンを分布系列化し、各分布を m 次元空間へと射影して得られる分布群が成す構造である。図 6 にヤコブソンによる仏語の母音・準母音構造を示す⁽¹¹⁾。構造音韻論では、このような構造が話者に依らず観測されることを主張するが、筆者らが提唱する音声表象は構造音韻論の物理的・数学的解釈である。ヤコブソンを更に遡れば、「言語は概念的差異と音的差異の系である」と主張したソシュールにまで行き着くことになる。筆者らが提案している音声の構造的表象は、彼等の議論とは全く独立に、純粋に数学的、物理的に「音声から非言語的要因による変形成分を消失させる」ことを目的として提案された枠組みであるが、結果的に約一世紀ほど前の古典的音声論と合致していることは非常に興味深い。

5 音的実体を全く使用しない構造的音声認識

5.1 連続母音系列発声をタスクとした音声認識

音声の音的実体を一切捨象し「音的差異の系」として発声を物理的に表象する方法を用いた音声認識を検討した。日本語五母音を入れ替えて構成される連続母音系列発声（語彙数 120 であるため、PP=120 の孤立単語認識となる）を対象語彙として検討した⁽¹²⁾。

図7にその枠組みを示す。入力音声を構造化（距離行列化）し、統計的にモデル化された構造的テンプレートと照合する。この際、片方の構造を回転及び平行移動して両構造を合わせて照合する。提案する構造的表象は変換不変性を有するため、任意の変換関数は、幾何学構造に対して回転或いは平行移動として作用する。例えば、声道長の差異（周波数ウォーピング）は回転として、音響機器特性の差異（伝達関数の掛け算）は平行移動となる²。回転・平行移動後の音響スコアは二つの距離行列のみを用いて計算される。即ち、音の実体を全く用いずに、入力された語を同定する。

男女計 8 名に 120 単語を 5 回ずつ発声させ、これを用いて 120 単語の統計的構造モデルを作成した。これとは異なる男女 8 名に同様の発声を依頼し、評価データとした (合計 4,800 発声)。結果を表 8 に示す。学習話者 260 名, 4,130 名の不特定話者 HMM の結果も示す。単語単位, 母音単位の両性能において, HMM(4,130) を越える結果が得られた。スペクトル包絡など, 音的実体に関する物理量を一切用いず音色の動きのみを捉えることで, 連続発声中の母音のほぼ全てが少ない学習話者数で同定できて「しまう」事実は驚嘆に値する。声に含まれる言語 (音韻) 情報は, 音的実体ではなく, 音色の動きとして符号化されている, と解釈すべきである。「音韻情報は音的実体によって伝搬される」と主張すれば, 音は話者によって変わるため, 35 万人もの話者を集める方法論に陥る。

5.2 音高に対する相対処理／音色に対する絶対処理

男女が同一歌詞の歌を歌った時、音高の動的パターンには絶対的な違いがある。男声は低く、女声は高い。これは男性の声帯が長く、固く、重たいために声帯振動周期が長くなるためである。純粹に物理的な要因のために男女間の音高差は生じる。よって、両者の動的パターンの同一性を論じる場合、絶対的な音高知覚は役に立たない。極端な絶対音感者は、移調前後で同一性認知が有意に遅れる⁽¹³⁾。

その男女が同一歌詞を読み上げた場合、音色の動的パターンには絶対的な違いがある。男声は太く、女声は細い。これは男性の声道長が長いがために、共鳴周波数が低くなるためである。純粋に物理的な要因のために男女間の音色差は生じる。よって、両者の動的パターンの同一性を論じる場合、絶対的な音色知覚は役に立たない、と記したいところであるが、筆者らの知る限り、全ての音声科学・工学の議論は音色に対しては絶対的な処理系を常に指向・構築してきた。筆者らは、この両者の

²ケプストラム空間では、周波数ウォーピングは行列 A の掛け算、伝達関数の掛け算はベクトル b の足し算となるため、最も簡素な話者変換は線形変換 $c' = Ac + b$ となる。この時、 $\times A$ が回転、 $+b$ が平行移動となる。

図 8: 構造的音声認識結果 [%]			
	HMM	HMM	提案手法
学習話者数	4,130	260	8
単語単位	97.4	82.1	98.3
母音単位	98.8	90.4	99.3

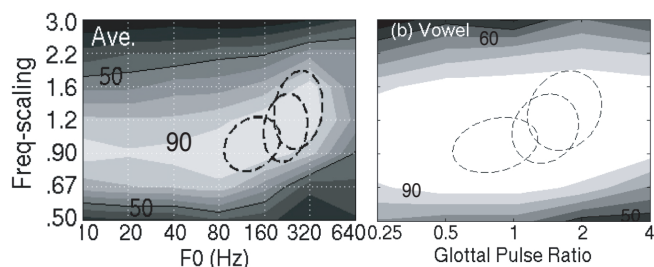


図 9: 孤立母音及び連続音声中の母音同定

隔たりに強い不自然さ(恣意性)を感覚している。何故、音色に対しては絶対音感ばかりを議論してきたのだろうか? 答えは簡単である。孤立音[あ]を聞いて、それを音韻/あ/であると同定できるからである。これは完全な絶対音感であり、音楽の階名同定とは完全に異なる。この絶対音感を抛り所として、数十万人の音声から統計モデルを構築してきた。ならば問うてみたい。「孤立音[あ]を聞いて、音韻/あ/であると同定できる能力は、音声言語の運用に果たして必要なのか?」と。

6 母音は音名なのか、階名なのか?

図 3 に示すように日本語の場合、話者の違いを考えても、母音間の重なりは大きく無い。しかし、この重なりは容易に増加できる。フォルマント周波数は声道長の関数であるため、巨人/小人の声を合成すればよい。通常の領域から外れた孤立母音に対する同定は可能なのだろうか? もしそれが困難であるにも拘わらず、音の連続ストリームの中にある母音の同定は容易に行なわれる場合、これこそ、音色に関する階名認知として解釈することができる。

先行研究にその答えを見ることができる⁽¹⁴⁾。図 9 左が孤立母音に対する同定率、右が無意味 4 モーラ単語の中の母音同定率である。縦軸の値 y に対して、 $170/y[\text{cm}]$ が凡そ話者の身長となる。また、右図の横軸の値 x に対して、 $160x[\text{Hz}]$ が基本周波数である。即ち、様々な身長・基本周波数の音声に対する、孤立母音の同定、及び無意味モーラ列中の母音同定の正解率である。図中点線の楕円が 3 つあるが、これは、実在する男性、女性、子供の領域を示す。全ての提示音声は STRAIGHT による分析合成音声である。孤立母音提示時(絶対的音認知時)は、実際に人間が存在する領域では 90% を越えるが、それを越え始めると同定率は下がり、例えば $65[\text{cm}]$ の小人となると、 $160[\text{Hz}]$ の音声で同定率は約 20% となる。これはチャンスレベルであり、母音同定は全く不可能の状態になる。

一方、無意味連続モーラ列中に母音が置かれると同定率が上昇する。 $65[\text{cm}]$ の小人ですら約 60% の正解率を呈する。提示単語が有意味語や親密語であれば、正解率は更に上昇するだろう。上記したように、筆者らはこの結果を音色に対する階名認知として考えている。再度問うてみたい。孤立音を聞いて音韻同定できたとして、それは音声言語運用と関係あるのだろうか?

言語化できない相対音感者(ラー音感者)は次の要求に難儀する。「次に提示されるメロディーの三番目の音を覚えてください。その後、別のメロディーが提示されます。同一音が出てきたら手を上げてください」音のシンボル(音名/階名)化が出来なければ、この問いは困難である。同様に「次に提示される音声の三番目の音を覚えて下さい。その後別の音声提示されます。同一音が出てきたら手を上げて下さい」という問いに難儀するのが発達性ディスレクシアであり、欧米には数多く存在する。音声を音韻(音シンボル)列として認知することが困難であり、その結果、文字の読み書きに苦勞する。語ゲシュタルトに基づく認知プロセスを引きずり、個々の分節音をシンボル認知することが困難である⁽¹⁵⁾。米国では、音声 CD を添付しなければ法的に教科書を発行できないほど「読めない」子が多い。これらの事実を省みた時に、音声ストリームを音シンボル列として認知する能力、孤立音を音シンボルとして同定する能力は、そもそも、音声言語運用の必要条件なのだろうか? 第一義的に必要なのは、母親の「おはよう」と父親の「おはよう」に同一のコンテンツが乗って(符号化されて)いると認知する能力であり、その認知に、発声を音韻列として表象する能力は不要である。必要なのは発声を「音的差異の系」として捉える能力であると筆者らは考えている。

音高に対する極端な絶対音感を持つと、移調前後で曲の同一性認知が遅れる⁽¹³⁾。同様に、音色に対する極端な絶対音感を持つと「おはよう」と「おはよう」の同一性認知が困難となるが、自閉症者の一部にその症状は観測される⁽¹⁶⁾。当然、音声言語(コミュニケーション)は成立しない。彼らの中には、音声模倣ではなく、声模倣を楽しむ者もいる⁽¹⁷⁾。音的実体に着眼すれば、音声模倣は異常であり、声模倣こそ自然である。しかしこの場合も、残念ながら、音声言語は存在しない。

7 音響音声学は意味のある議論をしてきたのか？

第3.1節において「メロディーという全体像があって初めて要素音のシンボル（階名）化が可能となる。シンボルを並べてメロディーが構成されるのではない」と書いた。前節の聴取実験は、「音声ストリームという全体像があって初めて要素音のシンボル化が可能となる。シンボルを並べて音声ストリームが構成されるのではない」ことを示唆する。全体が先にあるのか、要素が先にあるのか。言語音群を系（システム）として捉え、各音の（他音群との差異を通して定義される）相対的価値を議論するのが音韻論であり、個々の音を個別に観測し、その絶対的価値を議論するのが音声学である。となれば、（音響）音声学は果たして正しいのだろうか、という問いすら生まれてくる。

言語化できる（階名同定できる）相対音感者が時々犯す勘違いがある。「全ての長調の曲がハ長調として聞こえる」という主張である。彼らは調不変のドレミ知覚を行なうが、「ド」と聞こえた部位がピアノのC音の鍵盤を弾いている（即ち物理的に全く同一の音である）と解釈した時に、この勘違いが生まれる。音響音声学では、男女が読み上げた同一文に対して音韻「あ」を感覚した各部位には、同一の（類似した）物理現象が存在するという仮定で議論を繰り返して来た。前者の勘違いについては音高の絶対音感者がその誤りを指摘してくれる。後者も勘違いであるとすれば、それを音色の絶対音感者が指摘してくれることが期待されるが、残念ながら彼等は音声言語を有していない。

第3.1節において、音階における音配置構造のバリエーションを示した。では、図3に示した母音配置構造に対するバリエーションを考えた場合、これは、何に対応するのだろうか？周知のように、これは欧米圏における方言である。幼児の音声模倣を思考実験と共に再考する。一卵性双生児を出産直後に親が離婚して、父親、母親が一人ずつ養育する場合を考える。10年後、この双子の声は（どれほど父親、母親のことを愛していたとしても）片方がより太く、他方がより細くなることは無いだろう。彼らは声（音）を模倣する訳ではない。10年後彼らの発音は、一つの例外を除いて、非常に類似しているだろう。その例外とは、両親が異なる方言話者であった場合である。この場合、例えばappleの最初の母音/æ/が、双子の間で異なることは容易に想像できる。同一方言話者の男女の/æ/の違いは、共鳴周波数の違いである。異なる方言話者の男女の/æ/の違いも、共鳴周波数の違いである。前者は発音に影響せず、後者は影響する。何故か？結局「幼児が模倣するのは音ではなく、（音的差異の集合として規定される）音群の体系である」との説明が最も妥当かつ簡潔である。もし、両方の家庭で九官鳥が飼育されていれば、彼等は「音」を模倣するため、,, などの議論はもはや不要だろう。

8 まとめ

九官鳥は提示された「声」から何を学び、何を模倣するのか？幼児は、提示された「声」から何を学び、何を模倣するのか？そして両者の違いは何なのか？これを考えた場合、「音」だけを対象とした議論は、本質を欠く恐れが多分にある。古典的な言語学の場において「音ではなく音的差異」の議論があった。にも拘わらず、35万人から音声を集めるまでして「音」に拘る音声工学（残念ながら、音声科学も含まれるとするのが筆者らの意見である）という分野がある。これは多くの（恐らく全ての）研究者が、波形やスペクトルといった音声の音的実体、即ち「声」に目・耳を奪われて、言語学における議論を物理学的・数学的に探求することを怠った結果であろう。なお、本稿で述べた音声の構造的表象は、言語学の差異論とは全く独立に導かれた方法論であることを再度強調しておく。

参考文献

- [1] <http://tepia.or.jp/archive/12th/pdf/viavoice.pdf>
- [2] 早川, 月刊言語, 35, 9, pp.62-67 (2006)
- [3] W. Gruhn, "The audio-visual system in sound perception and learning of language and music," Int. Conf. on Language and Music as Cognitive Systems (2007)
- [4] 宮本, 音を作る・音を見る, 森北出版 (1995)
- [5] 加藤, コミュニケーション障害学, 20, 2, pp.84-85 (2003)
- [6] N. Minematsu *et al.*, "Universal and invariant representation of speech," Proc. Int. Conf. Infant Study (2006)
<http://www.gavo.t.u-tokyo.ac.jp/~mine/paper/PDF/2006/ICIS.t2006-6.pdf>
- [7] D. J. Levitin *et al.*, Trends in Cognitive Sciences, vol.9, no.1, pp.26-33 (2005)
- [8] 谷口, 音は心の中で音楽になる, 北大路書房 (2003)
- [9] 東川, 読譜力ー「移動ド」教育システムに学ぶ, 春秋社 (2005)
- [10] 峯松他, 春音講論, 1-P-12, pp.147-148 (2007)
- [11] R. Jakobson *et al.*, Notes on the French phonemic pattern, Hunter, N.Y. (1949)
- [12] S. Asakawa *et al.*, "Automatic recognition of connected vowels only using speaker-invariant representation of speech dynamics," Proc. InterSpeech (2007, accepted)
- [13] 宮崎, 日本音響学会誌, vol.60, no.11, pp.682-688 (2004)
- [14] 青木他, 秋音講論, 2-P-6, pp.373-374 (2004) / 林他, 春音講論, 2-Q-27, pp.473-474 (2007)
- [15] 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)
- [16] 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)
- [17] 深見, ひろしくんの本, vol.5, 中川書店 (2006)