社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

孤立音 [あ] を聞いて音韻/あ/と同定する能力は音声言語に必要か？

峯松　信明† 　西村多寿子†† 　櫻庭　京子††† 　朝川　智† 　齋藤　大輔†

† 東京大学大学院新領域創成科学研究科，†† 東京大学大学院医学系研究科，††† 清瀬市障害者福祉センター
E-mail:{mine,asakawa,dsk_saito}@gavo.t.u-tokyo.ac.jp, nt-tazuko@ams.odn.ne.jp, sakuraba@mtd.biglobe.ne.jp

あらまし　発達心理学では幼児の言語獲得を「音声模倣」という言葉で表現するが，通常，声（音）を模倣しようとする幼児はいない。一方，九官鳥の「音声模倣」では彼等は声（音）を模倣する。何故，幼児は声（音）を模倣しようとしないのか？音の音色は共鳴特性に支配されるため，音を模倣する場合，親が持つ声道と同様の形状を有する声道が必要となり，結局，親と同じ体格が要求される。よって，物理的に声模倣は不可能である。では，何故，模倣しようと努力しないのか。そもそも，物理的に異なる二つの音ストリーム（例えば，父・母の「おはよう」）を何故「同一である」と感覚するのだろうか？「聞こえた音を音韻（仮名）表象に変換し，音韻列としての同一性を認知する」との仮説も可能であるが，発達心理学はこれを否定する。何故なら，分節音及び音韻意識は「後天的に学習されるもの」だからである。本研究は，上記問いを数学及び物理の問題として捉え，「音色の相対音感」という新概念を提案することで解く。提案する枠組みは，一つの帰結として「孤立音を音韻として同定する能力は音声言語運用の必要条件ではない」という命題を主張するが，欧米圏に数多く存在する発達性ディスレクシアが該当する症状を呈している。
キーワード　音声模倣，話者不変量，音色差異，相対音感，発達性ディスレクシア

Is the ability of identifying a given [a] sound as phoneme /a/ necessary for spoken language competence?

N. MINEMATSU†, T. NISHIMURA††, K. SAKURABA†††, S. ASAKAWA†, and D. SAITO†

†,††The University of Tokyo, †††Kiyose-shi Welfare Center for the Handicapped
E-mail:{mine,asakawa,dsk_saito}@gavo.t.u-tokyo.ac.jp, nt-tazuko@ams.odn.ne.jp, sakuraba@mtd.biglobe.ne.jp

Abstract　Developmental psychology tells that infants acquire language through the vocal imitation but no infants try to imitate the voices of their parents. It is known that myna birds imitate the voices and sounds of their keepers. Why don't infants imitate the voices and sounds? Since the timbral characteristics of sounds are completely controlled by the shape of the sound generator, the voice imitation requires the same shape of the vocal tube that the parents have. Considering this reason, it is impossible for infants to imitate the voices of their parents. Then, why don't they *try* to imitate them and why do they perceive the identity between the two different sound streams, e.g, mother's "Good morning" and father's "Good morning"? Some readers may reply that infants decode the input streams into two sequences of phonemes and perceive the identity between the two phonemic sequences. Developmental psychology, however, denies this proposal because it claims that the segments and the phonemic awareness are learned later than the vocal imitation. In this work, taking the above question as one of the questions in mathematics and physics, it is answered by introducing a new concept of *relative timbre*. The proposed framework claims that the ability of identifying a given linguistic sound as phoneme is not required for spoken language competence. As far as the authors know, the cases are easily found in developmental dyslexics.
Key words　vocal imitation, speaker invariance, timbral difference, relative sense of sounds, developmental dyslexia

1. 幼児，九官鳥，そして，音声合成システム

子供の言語発達を考えた場合，幼児の聞く声の大半は母親，父親の声である。自らが話せるようになると，その子の聞く声の半分は（大人になっても）自らの声である（speech chain）。このように，人が接する言語音は，音響的（話者的）には非常に偏った音ばかりである。幼児の言語獲得は「音声模倣」という言葉で表現されるが [1]，この時，最も聞き慣れた両親の声（音）を模倣しようとする幼児はいない。一方，九官鳥の「音声模倣」に目を向けると，彼等は声（音）を真似ることが分かる。車，ドア，動物の声，様々な音を真似る [2]。優秀な九官鳥は聞けば飼い主が分かる [2] が，どんなに優秀な幼児を聞いても，親を当てることはできない。九官鳥は「音」を学習し，その「音」を長い鳴管を使って生成する。そして，恐らく「音」のモデルを内部的に構築し，以前聞いた「音」に反応する。

音声合成システムを考えてみる。波形編集合成に代わり，近

年では，HMM による音響モデルを用いた HMM 合成が注目されている [3]。学習話者（通常は一名）による数百〜数千文の音声試料を与えると,「音」と音素（異音）の対応を学習する。そして，学習試料に無い異音列をテキストで与えた場合でも，その異音列に相当する「音」を生成するようになる。しかしこの場合，得られるのは学習話者の声である。音声合成システムは「音」を学習し,「音」のモデルを構築し，与えられた異音 ID 列に沿って「音」ストリームを生成する。以上を考えれば,「音」ストリームが学習者の声と似てくるのは，至極当然である。学習者の声とどれだけ似ているのか，が，評価指標にもなる。そのため市販する音声合成器は，著作権が放棄された音声試料を使わざるを得ない。話者が容易に特定できる音声合成器を発声者の許可なく販売すれば，確実に訴えられることになる。

幼児が両親の声を真似ないのは，訴えられることを避けるためだと主張する人は皆無であろう。親と子の声道形状の相違を考えれば，両親の声を模倣することは物理的に不可能である。幼児，九官鳥，音声合成システム，と並べた場合，音声合成システムが九官鳥シミュレータであることがよく分かる。当然，話者変換技術を用いれば音声合成システムは他人の声を出すようになる。しかし「幼児は親の声を模倣後，親に隠れて話者変換の技を学ぶ」と主張する人も皆無だろう。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。

父親がある語を教える。子供が「音声」を模倣する。母親が別の機会に同一の語を教える。そして子供がまた「音声」を模倣する。この時，父親に対してより太い「声」で反応し，母親に対してより細い「声」で反応することは無い。そもそも何故，物理的には異なる音ストリームを同一であると感覚できるのだろうか？発達心理学は「彼等は音韻意識が未発達であるため音声を音韻（モーラ）列として認知することが困難である」と主張する [4](注1)。これに従えば,「音韻列としての同一性」を前提とした議論は不適切である。そして彼等は「与えられた音韻列に対して，各音韻を音に変換する」技が使えない状態で，両親と会話を楽しむ。幼児と音声合成は完全に異なることが分かる。

幼児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するのであれば，音そのもの（つまり声）に対する音響モデリング技術，即ち，音の生成モデル（generative model）は甚だ不適切である。結局，母親の「おはよう」，父親の「**おはよう**」，幼児の「オハヨウ」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。

2. 「音」の何をモデル化すべきなのか？

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 [1]）を獲得し，その後，個々の分節音を獲得する」と主張する [5]。筆者らの一部は，この話者不変と思しき「語ゲシュタルト」の音響的定義を，発達心理学，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」の定義が無いまま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。

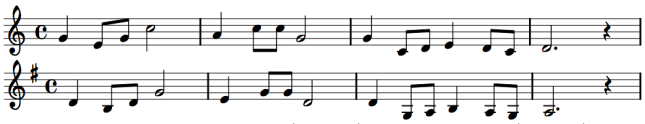
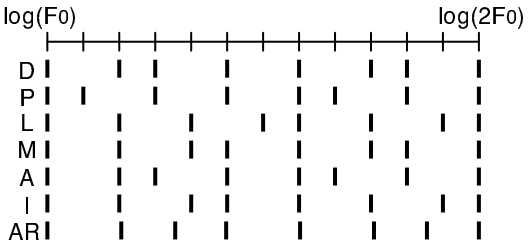


図 1　とあるメロディー（ハ長調）とその移調版（ト長調）



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

そもそも，二つの音の同一性を感覚するのに，その二音の物理的同一性が必要なのだろうか？人間は他の霊長類と異なり，全く異なる物理特性を有する二音を（ある環境下では）「同一である」と感覚する能力を持っている [7]。相対音感である。

2.1　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（カラオケに行ってキーを上げ下げしても）「ソーミソドー，と聞こえてきます」と彼らは主張する(注2)。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は意味はない。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

(注1)：そもそも彼らは「しりとり」を行なうことが困難である [4]。

(注2)：声を出さずに「ソーミソドー」と心の中でつぶやいた時と全く同一と思われる感覚・記憶が，無意識的に再生される，と言う主張である。

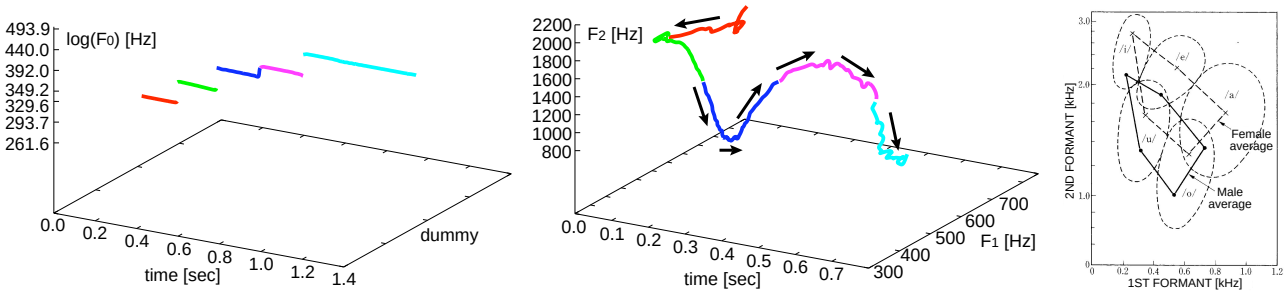


図 3　F_0 の動的変化としての CDEFG と音色の動的変化としての/aiueo/，及び，日本語母音図

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」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，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。シンボルを並べてメロディーが構成されるのではない。

2.2　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，下記に示す様に，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/aiueo/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/aiueo/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 [10]。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

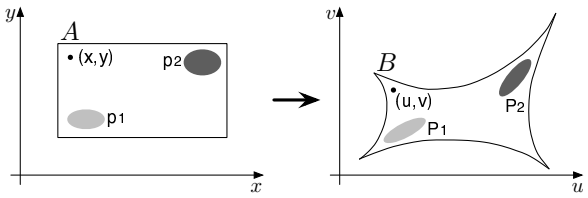


図 4　一対一対応関係を有する二つの空間 A と B

（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 [11]，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関数に対して，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，立証されることになる。

3.　非言語的音響変動不変の音声の構造的表象

3.1　2 つの空間における頑健な不変量

図 4 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 と B の間に一対一の対応があれば，空間 A の分布 p_i は空間 B の分布へと写像され，それを P_i とする。この時，次の等式が常に成立する [12]。

$$\oint\!\!\oint_A \sqrt{p_1(x,y)p_2(x,y)}dxdy \equiv \oint\!\!\oint_B \sqrt{P_1(u,v)P_2(u,v)}dudv$$

上式は，量子化学の世界では「重なり積分」と呼ばれる量であり(注3)，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，ヘリンジャ距離でも成立する一般的性質である（頑健な不変性）。

3.2　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

(注3)：この場合，分布は電子雲を指す。任意の二電子雲間の「重なり積分」を全て集めたのが「重なり行列」となる [13]。分子軌道法などで使われる。

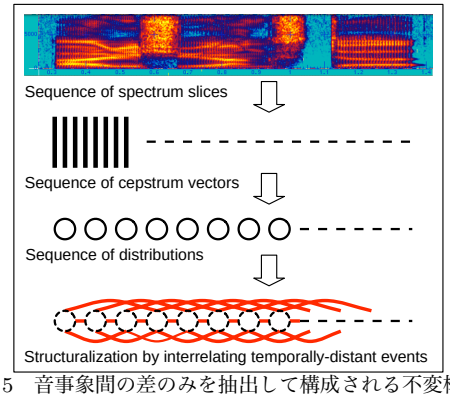


図 5　音事象間の差のみを抽出して構成される不変構造

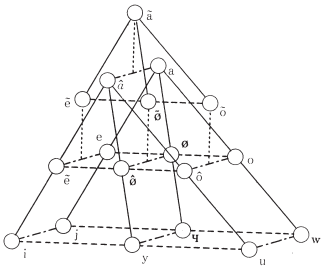


図 6　ヤコブソンによるフランス語の母音・準母音構造 [14]

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 [14]。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。

4.　音的実体を全く使用しない構造的音声処理

4.1　連続母音系列発声をタスクとした音声認識

図 5 に示した，音声の音的実体を一切捨象した物理表象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 [15]。

図 7 にその枠組みを示す。入力音声を構造化し，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**図 8** に示す様に，片方の構造を回転及び平行移動して両構造を合わせた上で照合する。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）は構造の回転として，音響機器特性の差異（伝達関数の掛け算）は構造の平行移動として解釈される(注4)。回転＆平行移動後の音響スコアは二つの距離行列を用いて計算されるが，これはタ

(注4)：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 [16]，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。

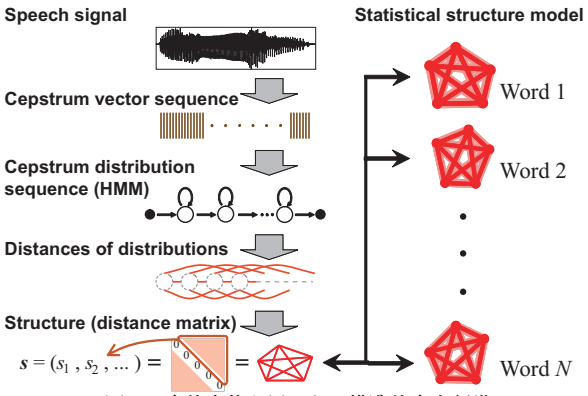


図 7　音的実体を用いない構造的音声認識

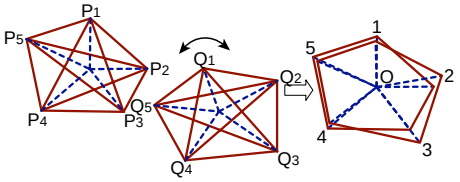


図 8　回転及び平行移動を通して行なう音響照合

表 1　音的実体を用いない構造的音声認識結果 [%]

	HMM(4,130)	HMM(260)	提案手法 (8)
単語単位	97.4	82.1	96.6
母音単位	98.8	90.4	98.6

ンパク質の構造解析などで用いられている手法と同一である。

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 1** に示す。学習話者 260 名, 4,130 名の不特定話者 HMM+CMN 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) とほぼ同等の性能を示している。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約 99%が非常に少ない学習話者数で同定できて「しまう」事実は，甚だ驚嘆に値する。声に含まれる言語情報は，音的実体ではなく，音色の動きとして符号化されている，と解釈すべきであろう。

4.2　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，重たいために声帯振動周期が長くなるためである。このような純粋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，移調前後で曲の同一性認知が有意に遅れ [17]，困難になる場合もある。

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。このような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，この両者の隔たりに強い不自然さ（恣意性）を感覚している。

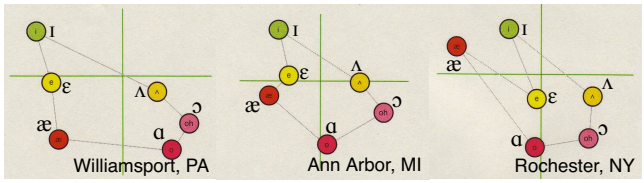


図 9　米国方言における母音構造（但し一部の母音）の差異 [18]

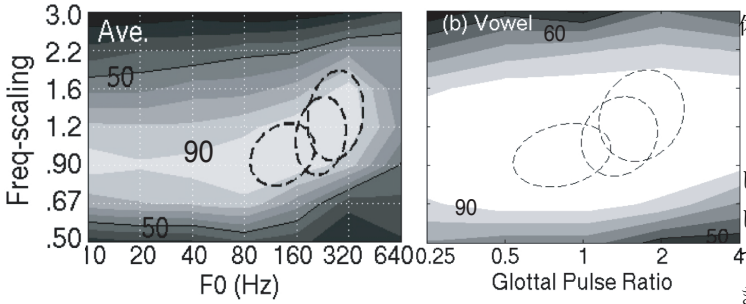


図 10　孤立母音及び連続音声中の母音同定 [19]〜[21]

4.3　音なのか音群の体系なのか？

第 2.1 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 [18]。例えば **図 9** に米国における母音構造の地方差を示す。声道長の正規化を行った後に，F_1– F_2平面上にプロットされたデータである。

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，この双子の発音は（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(注5)。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/は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。

4.4　幼児が学ぶ「もの」に根ざす音声合成系とは？

九官鳥は提示された「声」から何を学び，何を模倣するのか？一方幼児は，提示された「声」から何を学び，何を模倣するのか？両者の違いは何なのか？これを考えた場合，音コピーマシンを目指す音声合成技術は，些か不可思議な技術体系と言わざるを得ない。「要素分解＋再合成」の枠組みの上で音コピーマシンを追求する分析合成系は，符号化など，その存在価値は理解できる。しかし，九官鳥シミュレータとして位置づけられる音声合成技術を模索したとして，果たしてそれが，音声言語の出力を担う技術になり得るのだろうか？幼児が学ぶ,「声」の中に符号化されているコンテンツ（即ち「音声」）に根ざした音声合成系を考えた場合，次のような枠組みを検討すべきである。それは，生成対象とするコンテンツを「声」として出力しようとする個体の身体的特徴（声道長など）が与えられて初めて,「声」が生まれる枠組みである。身体情報が無ければ「声」が定義不能な枠組みである。この場合，第一義的に必要なのは「声」モデルではなく,「音声」モデルである。これに対して身体が与えられて初めて,「声」が生まれる枠組みである。

(注5)：但し，発話スタイルに相違が生じることは考えられる。

5.　母音は音名なのか，階名なのか？

前節で，音色の相対性に着眼した考察を行なった。音高に対しては相対音感は広く認知されているものの，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。この絶対音感を拠り所として，例えば音声認識の場合は，数万人の音声から統計モデルを構築して「音→音韻」変換を模索し，音声合成の場合は，入力された音韻列に対して「音韻→音」変換を模索してきた。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。

図 3 に示す母音図から分かるように，日本語の場合，話者による違いを考えても，母音間の重なりはそれほど大きく無い。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った場合，音の連続ストリームの中にある母音はどうなのだろうか？孤立母音の場合は困難であるにも拘らず，連続ストリーム中であれば容易である場合，これこそ，音色に関する階名認知として考えることができる。

先行研究にその答えを見ることができる [19]〜[21]。**図 10** 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(注6)。縦軸の値 y に対して，$170/y$[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，$160x$[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[cm] の小人となると，160[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，とたんに同定率が上昇する。65[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。孤立音の同定はできないが，連続ストリームに対しては，個々の音事象を同定できる。これは，階名認知そ

(注6)：厳密には，親密度データベース [22] における最低親密度単語群である。よって，音素配列的には正しい日本語である。無意味語と記したのは，上記 DB の開発者が「未知語と考えて差し支えない」と言及しているからである。

のものである。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？更に問うてみたい。音ストリームを音韻列として表記・認知できたとして，果た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（ラーラ音感者）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」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労する。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 [23]。米国では現在，教科書は音声 CD 添付が義務付けられている [24]。視覚障害を含め，読めない子供が数多く存在するからである。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？幼児にとって必要なのは，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，それがどう視覚化されるのか，は楽しい朝食を囲む際に何ら必要ない。

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。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 [25]。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 [26]。当然音声言語は無い。

6. ま と め

第 2.1 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。

本稿をここまで読まれた読者に対して，最後に一言問うてみたい。「“Happy Birthday” の歌を一番歌い易い音高で歌って下さい。」と依頼され，歌ったとする。そして「何故，貴方の歌の平均ピッチは 100[Hz] なのですか？そして，何故，初めの “ハ” の母音部分の平均ピッチは 90[Hz] なのですか？」と聞かれた時に何と答えるだろうか？音声科学の知識のある者なら「私の声の絶対的な高さは，私の声帯の長さ，重さ，固さが決めている事項ですから，私が制御しているのではなく，親からの遺伝情報（身体）が決定していると言えます。私が制御しているのは，音高変化の動的パターンだけですよ。」と答えるかもしれない。次に「じゃあ，初めの “ハ” の母音部分の第一フォルマント周波数は何故，700[Hz] なのですか？」と聞かれたらどうだろう？「私の “ア” のフォルマント周波数（音色の絶対的特性）は，私の声道の長さが決めている事項ですから，私が制御しているのではなく，やはり，親からの遺伝情報（身体）が決定していると言えます。私が制御しているのは，?????だけですよ。」と答えたとして，“?????” には何を入れるべきだろうか？筆者らはここに「音色変化の動的パターン」という言葉を入れて考えている。何故ならば，調音器官の運動は音色の動的変化を意味するからであり，更に，音色の動的変化パターンは，話者が制御できない身体性に不変な形で表象することが可能だからである。「**おはよう**」も「おはよう」も「オハヨウ」も，皆，同じ音響パターンとして観測することが可能だからである(注7)。筆者らはこの音響パターンこそ，幼児が模倣する「音声」，即ち，発達心理学の言う「語ゲシュタルト」であると考えている。

文　献

[1] 早川, 月刊言語, 35, 9, pp.62–67 (2006)
[2] 宮本, 音を作る・音を見る, 森北出版 (1995)
[3] http://hts.sp.nitech.ac.jp/
[4] 原, コミュニケーション障害学, 20, 2, pp.98–102 (2003)
[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－「移動ド」教育システムに学ぶ, 春秋社 (2005)
[10] 古井，ディジタル音声処理，東海大学出版会 (1985)
[11] 高橋他，信学技報，SP-2006-162，pp.13–18 (2007)
[12] 峯松他，春音講論，1-P-12，pp.147–148 (2007)
[13] 武次他，早わかり分子起動法，裳華房 (2003)
[14] R. Jakobson *et al.*, Notes on the French phonemic pattern, Hunter, N.Y. (1949)
[15] S. Asakawa *et al.*, “Automatic recognition of connected vowels only using speaker-invariant representation of speech dynamics,” Proc. InterSpeech (2007, accepted)
[16] M. Pitz, *et al.*, IEEE Trans. Speech and Audio Processing, 13, 5, pp.930–944 (2005)
[17] 宮崎，日本音響学会誌，vol.60，no.11，pp.682–688 (2004)
[18] W. Labov *et al.*, Atlas of North American English, Mouton and Gruyter (2005)
[19] D. Smith *et al*, J. Acoust. Soc. Am., 117(1), pp.305–318 (2005)
[20] 青木他，秋音講論，2-P-6，pp.373–374 (2004)
[21] 林他，春音講論，2-Q-27，pp.473–474 (2007)
[22] 天野他，日本語の語彙特性，三省堂 (2000)
[23] 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)
[24] 河村，“DAISY を活用したディスレクシアの方々への支援”，日本障害者リハビリテーション協会セミナー「ディスレクシアの支援・デンマークでの活動から」より (2006)
[25] 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)
[26] 深見，ひろしくんの本，vol.5，中川書店 (2006)

(注7)：但し，この音響パターンは（恐らく）4 次元以上の空間を要求するため，人間の意識活動において，これらを視覚的に認知することは困難であろう。