社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2008－SLP－74（6）
2008 / 12 / 9

音声言語運用が要求する認知的能力と音声言語工学が構築した計算論的能力

峯松 信明†

† 東京大学大学院工学系研究科，〒113-8656 東京都文京区本郷 7-3-1
E-mail: mine@gavo.t.u-tokyo.ac.jp

あらまし 波形素片やスペクトル素片をテンプレートとして保有し，音響照合・音声生成を行なう方法論から，HMM や GMM に代表される数理統計的な音響モデリング技術の台頭によって，音声認識・合成技術の性能・柔軟性は著しく向上した。しかし音声言語工学が構築して来た計算論的能力と，音声言語運用に関する人間の認知的能力には大きな差異があることは否めない。音声認識では，多様な話者性に対処するために数千・万の話者を使って音響モデルを構築するが（かつ，適応技術を用いたモデル補正が必要となる），幼児の音声言語獲得過程を考えると，聴取する声の多くは母親，父親，自身の声であり，非常に限られた話者性の音声である。音声合成に目を向ければ，合成器が生成するのは学習話者の声である。しかし父親の太い声を模倣する幼児はいない。親の声の物真似を通して音声言語を獲得した事例は存在しない。音声言語工学は音声言語に基づくマン・マシンインターフェイスの構築を目的としているが，上記した差異は技術的未熟さから来るのだろうか？それとも，技術を構築する研究者の人間理解の未熟さから来るのだろうか？本稿では，音声言語の獲得に難を示す重度自閉症者や，音声言語を獲得しない霊長類などの情報処理能力を参照しつつ，音声言語運用が要求する認知的能力と音声言語工学が構築した計算論的能力について考察する。
キーワード 音声言語運用，音声言語工学，多様性と不変性，相対音感，自閉症，霊長類，進化，構造音韻論

Cognitive competence required for spoken language performance and computational competence realized by spoken language engineering

Nobuaki MINEMATSU†

† Graduate School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan
E-mail: mine@gavo.t.u-tokyo.ac.jp

Abstract The performance and flexibility of speech recognition and synthesis technologies have been remarkably enhanced by introducing statistical modeling of speech sounds, before which, waveform-based or spectrum-based templates were stored and used for acoustic matching and speech generation. However, a large gap cannot be denied yet between the computational competence realized by spoken language engineering and the cognitive and language competence of humans. Acoustic models used for speech recognition are often built after collecting utterances of thousands of speakers. Human infants, however, acquire spoken language by hearing a remarkably speaker-biased speech corpus; mother, father, and themselves. In speech synthesis, a synthesizer generates speech sounds of the speaker used to train the synthesizer. No infant, however, impersonates its parents to acquire spoken language. Although the aim of spoken language engineering is to realize the man-machine interface based on spoken language, a large gap still exists between humans and machines. The author wonders whether this gap is due to the technical immatureness or due to researchers' immatureness of understanding humans. Considering the information processing of severely damaged autistics, for whom spoken language is very difficult to acquire, and that of the other primates than humans, who cannot acquire spoken language, this paper discusses the cognitive competence required for spoken language performance and the computational competence realized by spoken language engineering.
Key words spoken language performance, spoken language engineering, variability and invariance, relative sense of tone, autistics, primates, evolution, and structural phonology

1. 刺激の物理的多様性とその認知的不変性

環境からの刺激を個体が受容し，環境に対して応答を返す。この受容・応答のループが個体と環境との間のインタラクションを生む。しかし，同一の刺激はしばしば異なった様態をもって受容される。ある犬を見る。その犬を異なる角度から見る。当然網膜像は異なるが，我々は同一性を容易に認知する。朝日の下の花と，夕焼け空の同一の花は，色みは異なるが，同一の花として認知する。男性歌手によるハミングと，女性歌手による同一メロディーのハミングは基本周波数は異なるが，同一メ

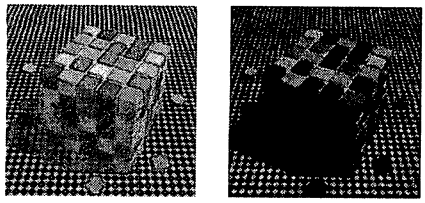


図1　異なる色眼鏡を通して見た同一のルービックキューブ

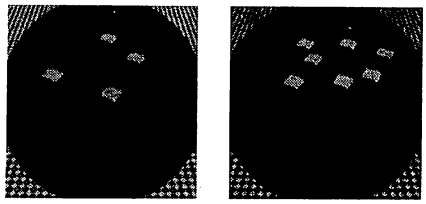


図2　対象となる部位以外を隠した絶対的な色知覚

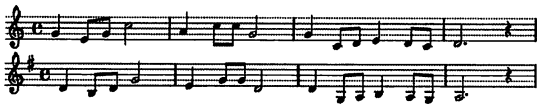


図3　ハ長調（上）とト長調（下）の同一メロディー

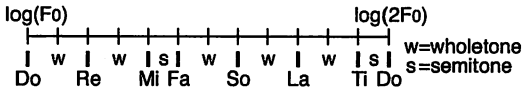


図4　長調におけるオクターブ内の音配置

ロディーとして認識する。男性の「おはよう」と女性の「おはよう」は，その音色が大きく異なるが，我々は言語的同一性を容易に発見する。これら刺激の物理的多様性は，何れも「静的バイアスによる刺激変形と，変形に不変な認知様式」と解釈できる。近年の心理学研究によれば，この同一性認知は，類似した情報処理が関与していることが示唆されている [1]〜[3]。

図1は，同一のルービックキューブを黄眼鏡，青眼鏡で覗いた場合の「見え」を表現している。両者において対応する部位は，絶対的には異なる波長を有するが，我々は両者に同一の色ラベルを振り，また，両キューブの同一性を認知する。また，左キューブ上面には4つの青部位を，右キューブ上面には7つの黄部位を認知するが，これらを単独で観察すれば，同一の色であることが分かる（**図2**参照）。即ち，絶対的に異なる色を「同じ」と判断し，絶対的に同一の色を「違う」と判断する(注1)。

同様の認知は，音高においても観測される。**図3**に示す二つの音系列は，同一メロディーのハ長調（上）とト長調（下）であるが（女性のハミングと男性の同一メロディーのハミングに対応），両者の同一性認知は，通常容易である。更に，聴取者が言語化可能な相対音感者であれば，両メロディーを同一のドレミ列（ソミソドラドドソ）で書き起こす。ここで，ハ長調の最初の音と，ト長調の最初の音は絶対的には異なるにも拘らず，彼らは同じ音（ソ）と判断する。更に，ハ長調の最初の音と，ト長調の4番目の音は絶対的には同一であるにも拘らず，彼らは異なる音（ソとド）と主張する。絶対的に異なる音高を「同じ」と判断し，絶対的に同一の音高を「違う」と判断する。

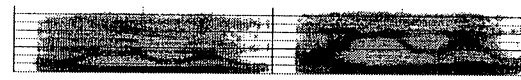


図5　長身（左）と短身（右）の話者による/あいうえお/の発声

センサー（受容器）の出力する物理量（波長，基本周期など）のみに基づいて人間の認知を説明しようとすれば，上記現象は説明困難となる。心理学研究によれば，これら認知の不変性（恒常性）は，刺激群のコントラストを用いた情報処理に基づくと考えられている [1]〜[3]。各刺激の絶対的物理量は容易に変形するが，対象刺激と周辺の刺激群との関係（コントラスト）は不変である。**図4**に長調のオクターブ内音配置を示す。「全全半全全全半」という音配置は調不変であり，メロディー中の2音（時間的に離れていてもよい）が三全音の音高差を持つ場合，それらは「ファとシ」或は「シとファ」のいずれかとなる [4]。このような不変的関係性を制約条件として，相対音感者はメロディーをドレミで書き起こす。彼らのドレミは階名である(注2)。

さて，静的バイアスに不変な認知様式は，進化的にどこまで遡れるのだろうか？色の不変的認知は蝶や蜂などの昆虫にも見られ [5]，進化的には非常に古い。朝日の下の花と夕焼け空の花の同一性が認知困難であれば，蜂蜜には辿り着けないのかもしれない(注3)。さて，音高に対する相対音感はどうだろうか？進化人類学研究によれば，異なる調の同一メロディーに対する同一性認知は，人間以外の霊長類には極めて難しいことが示されている [6]。即ち，音高の相対音感は進化的に極めて新しい。

2.　音声の物理的多様性とその認知的不変性

色，メロディー同様に，音声も静的バイアスによって多様に変化する。**図3**を女性と男性のハミングだとすれば，このバイアスは声帯の長さ・重さの性差に由来する。一方，声道の長さ・形状の性差に起因するバイアスは，音声の音色（スペクトル包絡）を変形する（**図5**参照）。この音色バイアスに対する人間の不変的認知は，どのような情報処理を基盤としているのだろうか？色の不変的認知に対して「蝶は数千の色眼鏡の試着を通して各色の統計モデルを構築する」と主張する仮説を筆者は聞いたことがない。そもそも，各色にラベル（カテゴリ）を振るという作業すら，花やキューブの同一性認知には必要無い。しかし音声言語工学では，**図5**に示す音ストリームを音シンボル列を通して眺め，各音シンボルの音響量を数千の喉形状を通して観測し，その絶対量を統計的にモデル化する方法論が業界標準となっている。生態学的及び進化論的に考えた場合，この方法論は極めて不自然である。言語化可能な相対音感者は，調不変にメロディーを書き起こせるが，孤立音の同定は困難である。コントラストが無いからである。音声の場合，孤立音の音素同定は容易であり，これは音色の絶対音感に相当する能力である。音声言語工学は，この絶対音感（音→シンボル変換）能力の計算機上での実装を，音声言語能力の計算機上での実装に対する必要条件として扱っているが，これは正しいのだろうか？

幼児の言語獲得過程を考える。幼児の聞く声の大半は母親，

(注1)：冊子は白黒印刷となるので，是非，下記で確認することを勧める。
http://www.lottolab.org/illusiondemos/Demo%2012.html

(注2)：絶対音感者にとってはドレミは音名であり，ドレミ列は調に依存する。

(注3)：興味深いことに，蟻にはこの能力は無いようである。

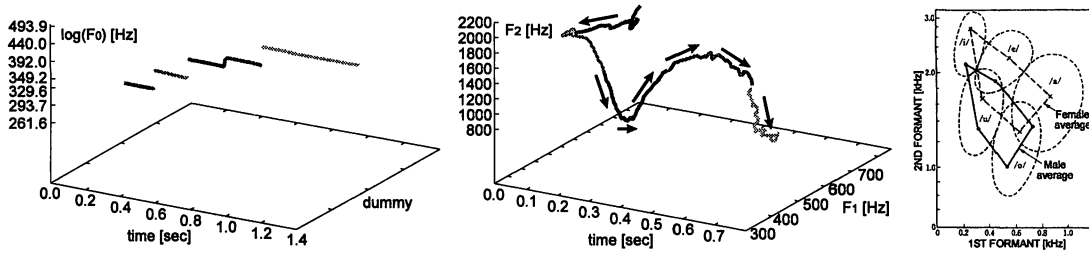
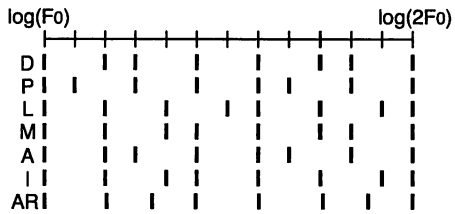


図 6　F_0 の動的変化としての CDEFG と音色の動的変化としての/aiueo/，及び，日本語母音図



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

図 7　6 種類の古典的教会音階とアラビア音階

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。構音障害の場合を除いて，人が聴取する言語音の話者性は極めて偏りが大きい。さて，幼児の言語獲得は「音声模倣・学習」という言葉で表現される [7]。親の発声を積極的に模倣する行為が観測される。この行為は動物学的には非常に稀な行為であり，人間以外の霊長類では観測されない [8]。人間以外では鳥，クジラ，イルカに見られる。さて，幼児の音声模倣は，親の声の音真似をする訳では無い。しかし，九官鳥は音を真似る。車，ドア，犬，猫など，音を真似る。人の声も音の一部でしかない。優秀な九官鳥は聞けば飼い主が分かる [9] が，どんなに優秀な幼児を聞いても親は当てられない。このように，動物の音声模倣は基本的に音の模倣となるが [10]，人間の幼児は，個体サイズを越えた，奇妙な音声模倣をする。彼らは親の声の何を真似ているのだろうか？

「親の声をシンボル（音韻）列に落として，その後，個々の音韻を自らの口で生成する」という説明は甚だ不適切である。彼らは音韻意識が未熟であり，「しり取り」も困難な状況にある [11](注4)。発達心理学で「幼児は単語全体の語形・音形（語ゲシュタルト）を獲得し，その後，個々の分節音を獲得する」と主張する [7], [12], [13]。語ゲシュタルトに話者の情報が含まれていれば，父親の声を音真似することになる。つまり，この語ゲシュタルトは話者不変の全体的な音パターンである必要がある。色や音高に関する不変的認知は，各刺激の物理量の絶対的認知ではなく，刺激群間のコントラストに基づく全体的認知が基本となる。生態学的，進化論的に考えれば，音声の不変的認知も同様の枠組みで検討することが自然であると考える。

筆者は，このような着眼点に基づき，音色の相対音感（音色コントラストに基づく話者性に不変な音声の全体的表象）を提唱している [14], [15]。ここでは，音高の相対音感と音色の相対

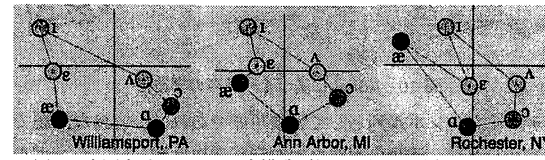


図 8　米国方言における母音構造（但し一部の母音）の差異 [16]

音感の等価性について更に幾つかの事実を示したい（**図 6** は両者を特徴量空間のトラジェクトリとして表現している）。音高の相対音感は「声帯の長さ・重さの個体差という静的バイアスを越えた音高パターン認知」を可能にしている。音色の相対音感は「声道の長さ・形状の個体差という静的バイアスを越えた音色パターン認知」を実装するために提案している。音高におけるコントラストであるが，長調以外の不変的関係性を**図 7** に示す。いくら移調しても満たされる音高コントラストの制約条件である。一方，日本語の五母音の配置を**図 6** に示す。男女間で同様の配置をとる。これを多次元の移調と捉える（音色は多次元，音高は 1 次元）。また，**図 7** にある多様な配置を音色で考えれば，それは欧米の方言に対応する。**図 8** に米語方言の幾つかを示す。欧米では，方言によって母音配置が変形する。

言語化可能な音高の相対音感者は，孤立音のドレミ同定が困難である。音声の場合これは当てはまらないが，例えば，巨人や小人の孤立母音を提示すると，母音同定が困難になることが示されている（フォルマント周波数は声道長に依存するため，**図 6** の母音図の領域外に位置する）。しかし，無意味語であっても連続モーラ列を提示されると，母音の同定が可能となることも示されている [17]〜[19]。これは，孤立音の同定は出来ないが，メロディーはドレミで書き起こせる相対音感者と類似している。結局，巨人の孤立母音と小人の孤立母音では両者の同一性が認知困難となった場合でも，連続ストリームの中ではその同一性が認知できることになる。幼児の場合，音韻カテゴリーが未獲得であれば，孤立音同定は原理的に出来ないが，音声模倣はできる。成人であっても同様に，孤立音同定が出来なくても，音声模倣は可能である。言い換えれば，孤立音同定や音韻カテゴリーの獲得は音声模倣の必要条件では無い [11], [20]。

このように考えると，ある言語障害の存在理由が物理的に説明可能となる。言語化困難な相対音感者(注5)は，次の問いに答えることが困難である。「次に示すメロディーの三番目の音を覚えて下さい。その後，別のメロディーを提示します。同じ音が出て来たら手を挙げて下さい。」階名であれ，音名であれ，メ

(注4)：例えば中国語には母音が 30 種類以上ある。中国語を母語とする幼児の音声模倣は，これら母音カテゴリを全て習得した後に始まる訳では無い。

(注5)：**図 3** の同一性は容易に認知可能であるが，これらをドレミで表記できない。「ラーララーラ」としか表現できない音感の持ち主である。

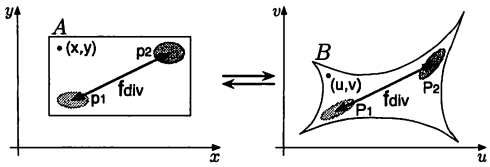


図 9　連続かつ可逆な変形に対して不変な f-divergence

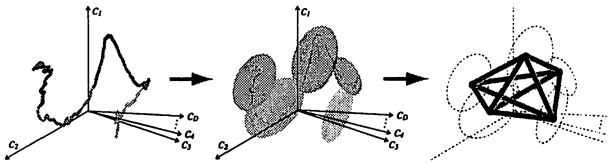


図 10　f-div. に基づく一発声の構造化

ロディーをシンボル列として表象可能であれば，容易な課題である。しかし，言語化困難な相対音感者には難しい。シンボル列としての表象が困難だからである。同様の現象を音声で考える。「次に示す発声の三番目の音を覚えて下さい。その後，別の人の別の発声を提示します。同じ音が出て来たら手を挙げて下さい。」音声をシンボル列へと変換できない方々（成人）がいるのだろうか？彼らは，音声言語運用は問題無いが，文字言語の使用は困難となるはずである。**図 6** の母音図では母音間の重なりは大きくないが，カテゴリの種類が増えれば，自ずと重なりは増す [21]。当然，音声の相対的特性に頼った音声認知がより強くなることが予想される。母音カテゴリ数の多い英語圏では，(言語化出来ない音高の相対音感者が存在する様に)，文字言語の使用に困難を示す方々が，当たり前のように存在することを示唆する。事実，欧米には「読み書き」のみに特異的に困難を示す読字障害（失読症，dyslexia）者が当たり前のように存在する [13]。筆者はこの障害の存在を知らずに，音声の物理特性と話者の多様性のみから，この障害の存在を予言していた。

3. 音声の構造的表象とその実験的検証

話者の違いは音声をどのように変形させるのか?先行研究で議論された多くの話者不変量は全て，周波数の線形変換（$\hat{f}=\alpha f$）を仮定し，その上で音の実体に対する変換不変量を検討している [22]〜[25]。音声合成における声質（話者）変換では，通常，より複雑な写像関数が用いられる。この事実を考慮すれば，上記の枠組みで話者の多様性を十分表現することは困難である。筆者の提唱する音声の構造的表象は，如何なる連続かつ可逆の変換に対しても不変となるコントラスト量に基づいている。

全ての事象は特徴量空間の点ではなく，分布として記述される。**図 9** に示すように，次式で定義される f-divergence は，連続かつ可逆な全変換に対して不変量となる（十分性）。更に，$\oint M(p_1(x),p_2(x))dx$ の形で 2 分布に関する量を定義した場合，それを変換不変にすれば，必ず f-div. になる（必要性）[26]。

$$f_{div}(p_1(x,y),p_2(x,y)) = \oint p_2(x,y)g\left(\frac{p_1(x,y)}{p_2(x,y}\right)dxdy$$

この f-div. を用いて一発声を構造化する様子を**図 10** に示す。トラジェクトリーを一旦有限個の分布列に変換し，全ての分布間距離を f-div. で計測する。最終的に距離行列として表象されるが，距離行列は一つの幾何学構造を規定するため，本表象を

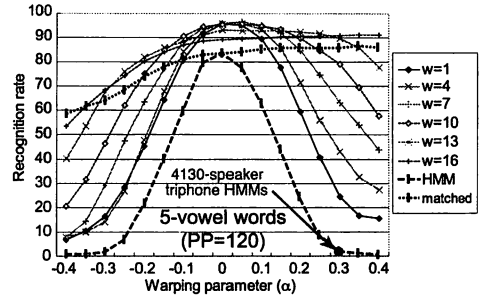


図 11　5 母音単語からなる語彙セットに対する認識率

音声の構造的表象と呼ぶ。音高の相対音感同様，コントラストのみを抽出している。なお，構造表象に基づく音響的照合は，話者適応・正規化を施した後の音響スコアを，話者適応・正規化を明示的に施すことなく算出することが可能であり，話者性に対する極めて高いロバスト性が示されている [27]。

詳細な実験条件などは参考文献に譲るが，高いロバスト性に関して実験結果を一つ紹介しておく。提案している構造表象は，極めて強い不変性を持つため，異なる単語が同一であると評価されることがある（強すぎる不変性問題）。また，N 個の事象に対して ${}_NC_2$ 個のコントラストが得られるため，パラメータ数が $O(N^2)$ で増える（高すぎる次元数問題）。これらを次元分割や LDA を併用することで解決し，日本語 5 母音を並び替えて構成される 120 単語の孤立単語認識を行なった。結果を**図 11** に示す。学習音声は 8 名の成人話者，評価音声は別の 8 人の成人話者の音声である。なお評価話者については，周波数ウォーピングにより巨人化（α<0），小人化（α>0）した音声も用いており，人工的に不一致条件を実現した。**HMM** が通常の単語 HMM を使用した場合であり，不一致条件では認識率が容易に下落する。**matched** は，各 α の値毎に単語 HMM を構築し，不一致条件が無い状態での認識率である（理論的な最高性能）。それ以外は構造表象による認識性能である。何れの場合も，成人話者の音声を用いて構造単語モデルを構築している。w は次元分割のパラメータ（ブロックサイズ）である。成人話者の音声のみから構造単語モデルを構築しているにも拘らず，w=16 の時には，個々の条件で構築し直した HMM と同等あるいはそれ以上の性能を示している。不一致時の認識率の下落は数千人の話者を用いた音素音響モデルでも避けることは出来ず，図には 4,130 人の話者から構築したトライフォンモデルを使用した時の性能についても一部示している。なお [27] には，音素バランス単語セットに対する構造表象の性能も掲載している。

4. 音声言語運用が要求する認知的能力と音声言語工学が構築した計算論的能力

絶対音感が極めて強くなると，人間であっても，移調前後でメロディーの同一性認知が遅れるようになる [28]。「異なる音は異なる」として認知される。オーケストラの音合わせで使われる基準音は A4 であり，通常 440[Hz] であるが，ホールや季節によってこれは上下する。極端な絶対音感者は，この変更が耐えられない。440[Hz] からずれたら，それは基準音ではない。

刺激の絶対的特性にのみ基づく情報処理システムを構築すれ

ば，絶対音感者の様にそのロバスト性は極めて低くならざるを得ない。ある先天的障害を持つと，次のような挙動を示すことがある。ある犬を見る。その犬を異なる角度から見る。そこには同一性は無く，二匹の犬として認知される [29]。**図 1** を見ただけで，左キューブ上面の青部位と右キューブ上面の黄部位が同一色であることを見破る [30]。移調前後でメロディーの同一性認知が困難となる強い絶対音感を持つ [29]。彼らは自閉症者であり，しばしば極めて優れた記憶力を持ち，刺激の詳細な局所的・具体的情報をそのまま記憶する [31], [32]。逆に言えば，刺激群の中に埋め込まれた抽象的なパターンの抽出に困難を示す [29]。自閉症者の知覚世界は健常者のそれとは大きく異なり，言語活動が可能な自閉症者（アスペルガー症候群や高機能自閉症）の多くは，自らを異星人と語る [33]〜[35]。昆虫でも可能な情報処理が阻害されていれば，異星人と名乗るのも無理は無い。

自閉症者であり動物学者である T. Grandin 博士は，局所的・具体的な情報を詳細に記憶する彼らの（そして自らの）情報処理戦略は，動物のそれと類似していることを指摘している [32]。事実，霊長類を自閉性の高い個体として利用し，重度自閉症者の支援を研究している例もある [36]。また，刺激をありのまま受け止めて記憶し，効率的な情報の取捨選択に困難を示す彼らの情報処理は，人工知能の世界で広く知られる「フレーム問題」と関連づけて議論されている [37]。健常者は「不用な情報を，不用であると判断することなく捨て去る能力」を通常獲得しているが，これを持たないロボットは，環境の変化に逐一反応し，無限の変化の可能性を考えざるを得なくなる。これがフレーム問題であり，自閉症者はこの問題を抱えつつ生活している。当事者自身の言葉を借りれば「自閉は情報の便秘」となる [38]。

重度自閉症者にとって音声言語は，通常，その利用が極めて困難なツールである。文字言語が第一言語となる場合もある。但し，フォントが変わると読めなくなることもある。当然，母親の声は理解可能だが，父親の声は認識困難となる例もある [35]。電話越しの母親の声も難しい。九官鳥のように，父親の声をそのまま真似る自閉症児 [39] や，七色の声を持つと呼ばれる声優・中村メイコの声をそっくり音真似する自閉症児 [40] が報告されている。しかし，音声言語運用は彼らには困難である。

音声言語情報処理の性能を飛躍的に向上させた，音声の数理統計的モデリングは，1) マイクから取得した物理量を詳細に表象する技術，2) 不可避の変形を被った同一カテゴリの複数の音資料を統計的にモデル化する技術，3) 新たな変動に対してモデルを適応・変形する技術，等によって構築されている。観測量を O とし，O の生成が要因 A, B, C に依存するとするならば，

$$P(A,B,C|O) \propto P(O|A,B,C)P(A,B,C) =$$
$$P(a_0,b_0,c_0)\prod_{i=1}^{T} P(o_i|a_i,b_i,c_i)P(a_i,b_i,c_i|a_{i-1},b_{i-1},c_{i-1})$$

となり，$P(O|X)P(X)$ を最大化する X を求める問題へと帰着される。スペクトル包絡を O とすれば，それを変形させる要因は無数にあり，問題は爆発する。一方，要因群を言語的要因 A と（時不変な）非言語的要因 B, C に分類し，後者に非依存な観測量 O' を定義できれば，問題は $P(O'|A)P(A)$ の最大化と簡素化される。このような方策を採らなければ，O の多様性を十分カバーできる学習データが得られない場合，認識性能は容易に劣化する。そのため，適応的に $P(O|X)$ を補正することになるが，幾ら学習データを集めても新たな要因は容易に生成され，その都度補正をかける。筆者は，現在の業界標準となっているこの戦略と，不用な情報を捨てられず，その結果フレーム問題に悩み続ける自閉症者の情報処理とに類似性を感じている。

そもそも，音高の絶対量に基づいて判断を行なう絶対音感者は，ある音高カテゴリの定義を変えることは出来ない。出来ないのが絶対音感だからである。音の絶対量に基づいて事象 x と事象 y の類似性を検証する方法論を採りながら，ある音カテゴリの定義（音響モデル）を，都合により，あれこれ変える方法論は，論理的な矛盾を抱えていると考察することもできる。

音声言語工学が目指すのは「人間が行なう情報処理の実装」である必要は無い，との意見は古くから指摘されている。しかし，「生態系が長い進化の過程を経て獲得してきた静的バイアスに頑健な情報処理」とも異なると思われる方法論を築き上げ，自らを異星人と呼ぶ自閉症者とよく類似した情報処理系を構築しているのも事実である。ASR という名称の A は何を意味するのだろうか？automatic, autistic あるいは alien だろうか？認知ロボティクスの世界では 90 年代に，フレーム問題で悩むロボットと自閉症児の類似性が指摘され [41]，今日に至っている。音声言語工学も，この問題に真剣に取り組むべき時が来ているように考える。生態系が「コントラスト抽出に基づく全体的なパターン形成を通して静的バイアスの対処法を獲得して来た」のであるならば，本来構築すべきは，音声ストリームに埋め込まれた不変パターン・構造を抽出する技術の構築であろう。近年ヨーロッパにて，人間の音声コミュニケーション能力を獲得できる人工エージェントのプロジェクトが発足している [42]。J. Hawkins 博士が提唱する memory-prediction theory を主要な枠組みとして採択しているが，彼は著書の中で次の様に述べている [43]。“I believe a similar abstraction of form is occurring throughout the cortex. Memories are stored in a form that captures the essence of relationships, not the details of the moment. The cortex takes the detailed, highly specific input and converts it to an invariant form. Memory storage and recall occur at the level of invariant forms.”

5. 音声の構造的表象の言語学的妥当性

本稿をまとめる前に，提案している音声の構造的表象に対する言語学的考察をしておきたい。実は構造的表象は，極めて古典的な言語論に再度焦点を当てているに過ぎない。例えば，音素の定義を意味論から切り離して言語学に求めた場合，二種類の定義が導かれる [44]。1) A phoneme is a class of phonetically-similar sounds and ..., 2) A phoneme is one element in the sound system of a language having a characteristic set of interrelations with each of the other elements in that system. 第一の定義が音の絶対量に着眼した統計モデルを築き上げたことは自明である。第二の定義は，音素は他の音素との関係によって初めて定義される，としている。この定義を押し進

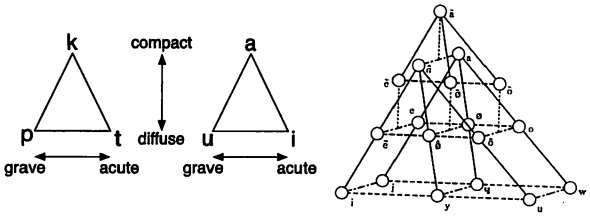


図 12　ヤコブソンの構造音韻論

めると，The phonemes cannot be defined acoustically and they are a set of abstractions [44]. となる。さて，どちらが第一義だろうか？近代言語学の祖ソシュールまで遡るならば，相対的定義が第一義であることが分かる [45]。Language is a system of only conceptual differences and phonic differences. What defines a linguistic element, conceptual or phonic, is the relation in which it stands to the other elements in the linguistic system. The important thing in the word is not the sound alone but the phonic differences that make it possible to distinguish this word from the others. ソシュールの言葉に啓蒙され，ヤコブソンによって「ある音素と音素はどう異なるのか」ついて検討され，弁別素性が提案された。**図 12** に示す母音・子音三角形はその一例である [46]。注意すべきは，元来弁別素性は音的差異の表現手段であって，音素の表現手段ではない(注6)。やがて音韻構造はより複雑なものとなる（**図 12** 右 [47]）。このような構造が性別・年齢を問わず存在すると考える。音声の構造的表象は構造音韻論の物理的実装である。

6. ま と め

音声言語運用が要求する認知的能力と音声言語工学が構築した計算論的能力という観点から，従来の方法論と，筆者が提唱している音声の構造論とを概観した。筆者は音声の実体に基づく処理を否定するものではない。人間とて長い進化プロセスの産物でしかない。刺激の実体に対する情報処理が基盤としてあり，その上に，実体の変形に不変なコントラストを使ったプロトコルが走っていると考えている。既に従来の実体論と提案する構造論との融合について検討を開始している。更には，異メディアに対しても同様の枠組みの応用を検討している。

文　献

[1] R. B. Lotto *et al.*, "An empirical explanation of color contrast," *Proc. the National Academy of Science USA*, 97, 12834–12839, 2000

[2] R. B. Lotto *et al.*, "The effects of color on brightness," *Nature neuroscience*, 2, 11, 1010–1014, 1999

[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－「移動ド」教育システムに学ぶ, 春秋社, 2005

[5] A. D. Briscoe *et al.*, "The evolution of color vision in insects," *Annual review of entomology*, 46, 471–510, 2001

[6] M. D. Hauser *et al.*, "The evolution of the music faculty: a comparative perspective," *Nature neurosciences*, 6, 663–668, 2003

[7] 早川, 月刊言語, 35, 9, pp.62–67, 2006

[8] W. Gruhn, "The audio-vocal system in sound perception and learning of language and music," *In: Proc. Int. Conf. on language and music as cognitive systems*, 2006

[9] 宮本, 音を作る・音を見る, 森北出版, 1995

[10] 岡ノ谷, 春音講論, 1-7-15, 1555–1556, 2008（個人的な質疑応答を含む）

[11] 原, コミュニケーション障害学, 20, 2, pp.98–102, 2003

[12] 加藤, コミュニケーション障害学, 20, 2, pp.84–85, 2003

[13] 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

[14] N. Minematsu, "Mathematical evidence of the acoustic universal structure in speech," *Proc. ICASSP*, 889–892, 2005

[15] 峯松他, 音声言語情報処理研究会, 2007-SLP-67-14, 75–80, 2007

[16] W. Labov *et al.*, *Atlas of North American English*, Mouton and Gruyter, 2005

[17] D. Smith *et al*, "The processing and perception of size information in speech sounds," *J. Acoust. Soc. Am.*, 117, 1, 305–318, 2005

[18] 青木他, 秋音講論, 2-P-6, 373–374, 2004

[19] 林他, 春音講論, 2-Q-27, 473–474, 2007

[20] 峯松, "「あ」という声を聞いて母音「あ」と同定する能力は音声言語運用に必要か？～音声認識研究からの一つの提言～", 日本語学 4 月号, 187–197, 明治書院, 2008

[21] J. Hillenbrand *et al.*, "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* 97, 5, 3099–3111, 1995

[22] S. Umesh *et al.*, "Scale transform in speech analysis," *IEEE Trans. Speech and Audio Processing*, 7, 1, 40–45, 1999

[23] T. Irino *et al.*, "Segregating information about the size and shape of the vocal tract using a time-domain auditory model: the stabilised wavelet-Mellin transform," *Speech Communication*, 36, 181–203, 2002

[24] A. Mertins *et al.* "Vocal trace length invariant features for automatic speech recognition," *Proc. ASRU*, 308–312, 2005

[25] 益子, 秋音講論, 3-7-2, 105–106, 2005

[26] 喬他, 信学技報, SP2008-51, 49–54, 2008

[27] 朝川他, 秋音講論, 2-P-3, 113–116, 2008

[28] 宮崎, 日本音響学会誌, 60, 11, 682–688, 2004

[29] U. Frith, 自閉症の謎を解き明かす, 東京書籍, 2005

[30] D. Ropar *et al.*, "Do individuals with autism and Asperger's syndrome utilize prior knowledge when pairing stimuli?" *Developmental Science*, 4, 4, 433–441, 2001

[31] T. Grandin, 我, 自閉症に生まれて, 学研, 1994

[32] T. Grandin, 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

[33] O. Sakcs, 火星の人類学者, 早川書房, 2001

[34] 泉, 地球生まれの異星人, 花風社, 2003

[35] 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, 2005

[36] 北澤, 自閉症治療に挑む心理学と神経科学, 自閉症スペクトラム研究, 社会技術研究開発事業「脳科学と社会」研究開発領域, 領域架橋型シンポジウム, 2008

[37] 藤居, 自閉症～「からだ」と「せかい」をつなぐ新しい理解と療教～, 新曜社, 2007

[38] ニキ, スルーできない脳～自閉は情報の便秘です～, 生活書院, 2008

[39] R. Martin, 自閉症児イアンの物語～脳と言葉と心の世界, 草思社, 2001

[40] 深見, ひろしくんの本 (V), 中川書店, 2006

[41] 小嶋, "ロボットに「心の理論」は教えられるか？", 発達心理学シンポジウム資料, 1998

[42] ACRONS (Acquisition of Communication and Recognition Skills) http://lands.let.ru.nl/acorns

[43] J. Hawkins *et al.*, *On intelligence*, Henry Holt, 2004

[44] H. A. Gleason, *An introduction to descriptive linguistics*, New York: Holt, Rinehart & Winston, 1961

[45] F. D. Saussure, 一般言語学講義, 岩波書店, 1940

[46] R. Jakobson *et al.*, *Preliminaries to speech analysis*, MIT Press, Cambridge, MA, 1952

[47] R. Jakobson *et al.*, *Notes on the French phonemic pattern*, Hunter, N.Y. 1949

(注6)：ヤコブソン自身，最終的には「音素を素性の束」として素性を音素定義に用いたが，筆者はこれを，「ヤコブソンの勇み足」と考えている。