

音色の変復調を通して考える音声コミュニケーション

峯松 信明[†]

[†] 東京大学大学院情報理工学系研究科, 〒 113-0033 東京都文京区本郷 7-3-1

E-mail: mine@gavo.t.u-tokyo.ac.jp

あらまし 音声における「音響的多様性と知覚的不変性」は、古くから音声科学・工学の分野における研究対象となってきたが、未だに解決に至っていない問題の一つである。近年筆者は、数学的に不変性が保証された「関係的不変性」を用いてこの問題に対する一つの解を導いている。即ち、如何なる連続・可逆な空間写像に対しても不変となる f -divergence を用いて、音声の動き成分を抽出し、これのみを用いて発声を表象する方式を提案した。本稿では、この表象方式に対して電気通信及び進化人類学の観点から一つの解釈を加える。音声生成はしばしば、話者がもつ（その話者特有の）音色（スペクトル）を、構音器官を動かすことで変調させるプロセスとして解釈される（音色変調）。この場合、音声認識、即ち、音声からその言語的内容（メッセージ）を抽出するプロセスは、音色復調として考えることになる。音声コミュニケーションを音色の変・復調プロセスとして考えた場合、ヒトとサル（サル）の構音器官に関する運動能力の差異、音情報に対する認知能力の差異と非常に整合性のある議論が展開できるようになる。更に本稿では、言語的内容の伝搬を担う音声の音響的特徴は、その多くが超分節的特徴（韻律的特徴）であることの主張も行う。

キーワード 変・復調, 音色, 音声コミュニケーション, 人類学, 自閉症, 超分節的特徴

Speech communication modeled as timbre modulation and demodulation

Nobuaki MINEMATSU[†]

[†] Graduate School of Information Science and Technology, The University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

E-mail: mine@gavo.t.u-tokyo.ac.jp

Abstract Perceptual invariance against a large amount of acoustic variability in speech has been a long-discussed question in speech science and engineering and it is still an open question. Recently, we proposed a candidate answer for it based on mathematically-guaranteed relational invariance. Here, completely transform-invariant features, f -divergences, are extracted from speech dynamics of an input utterance and they are used to represent that utterance. In this paper, this representation is interpreted from a viewpoint of telecommunications and evolutionary anthropology. Speech production is often regarded as a process of modulating the baseline timbre of a speaker's voices by manipulating the vocal organs, i.e. spectrum modulation. Then, extraction of the linguistic content from an utterance can be viewed as a process of spectrum demodulation. This modulation-demodulation model of speech communication has a good link to known morphological and cognitive differences between humans and apes. The model also claims that a linguistic content is transmitted mainly by supra-segmental (prosodic) features.

Key words (de)modulation, timbre, speech communication, anthropology, autism, supra-segmental features

1. はじめに

音声には、話者・環境に起因する、多様な音響歪みが不可避免的に混入する。しかし、人の音声コミュニケーションにおいては、これら（静的な）非言語的要因は支障にならないことが一般的である。例えば身長 250cm ほどの世界一の巨人と 70cm ほどの世界一の小人（成人）とが、世界一の声色差をものともせず会話するシーンが、テレビで放映されている（図 1 参照）。

自動音声認識において、これら話者・環境の違いに頑健なシステムを構築する場合、数千・万人の声を集めて数理統計的な

手法により不特定話者音響モデルを構築したり、話者・環境が変わる度に音響モデルを（その話者・環境に合わせて）適応・修正して対処することが一般的である。これらは、現在の音声認識の枠組みが、音声が進ぶ言語的情報の同一性を、音響的の同一性を通して検証する枠組みとなっていることに由来する。

幼児の言語発達を考えた場合、彼らが聞く声の多くは両親の声であり、また自らが話し始めると、聞く声の約半分は自らの声となる。即ち人は、非常に話者性の偏りの大きい音声の聴取を通して、頑健な音声認識能力を獲得する。話者バランスの取れた大規模コーパスを要求する訳では無い。更に、言語の獲得



図1 世界一の巨人と世界一の小人 (大人)

は音声模倣を基本とするが、彼らの模倣は音響的模倣（即ち、声帯模写）ではない。父親の太い声と自らの可愛い声の声色の違いを無視して、父親の「おはよう」と自らの「おはよう」との言語的同一性を感覚する。つまり、言語的に同一であることは音響的同一性を必要としない。それでは、両発声の何が同一なのだろうか？言い換えれば、幼児は、父親の太い声の何を真似ているのだろうか？太い「おはよう」を音韻列（平仮名列）として解釈し、個々の音韻を自らの小さな口で音に変え、可愛い「おはよう」となる、という説明は不適切である。彼らの音韻意識はまだ未発達だからである[1]。研究者は音韻列を通して音声を表象しがちだが[2]、幼児の場合これとは大きく異なる。

「おはよう」と「おはよう」の物理的共通項について、発達心理学では例えば「語全体の語形・音形」[3]「語の全体ゲシュタルト」[4]「related spectrum pattern」[5]などの用語で指し示されることがある。しかし、この話者差を超えた発声の共通項に対する物理的定義は発達心理学では議論されていない。筆者らは先行研究において、話者の違いを声空間の空間写像として捉え、話者を越えた音声特徴量を写像不変量として解釈し、語ゲシュタルトの物理実装を試みた[6],[7]。刺激の多様な変化に対する認知の不変性は心理学的には「知覚の恒常性」と呼ばれるが、[6],[7]では色知覚やメロディー知覚の恒常性に倣う形で音色（声）の恒常性を考察した。そして、二音間の音色差を変換不変に計測する手法を提案し、提案手法に基づいて発声を表象し、超頑健な孤立単語認識システムを実装した[8]。

本稿では、提案された新しい音声表象論・技術に対して、変・復調に基づく電気通信技術の観点から新しい解釈を加える。音声生成を音色変調、音声認識を音色復調として音声コミュニケーションをモデル化すると、進化人類学研究で示されている、サルとヒトとの構音器官の運動能力差異、更には、音情報に対する認知能力差異に関して、整合性の高い議論が可能となる。

2. 知覚の恒常性と関係的不変性

2.1 色とメロディーにおける知覚の恒常性

図2は、同一のルービックキューブを黄・青眼鏡で覗いた「色み」を表現している。両者において対応する部位は、異なる波長を網膜に届けるが、我々は両者に同一の色ラベルを振り、両キューブの「色み」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つの黄部位を認めるが、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3参照）。即ち、異なる色を「同じ」と判断し、同一

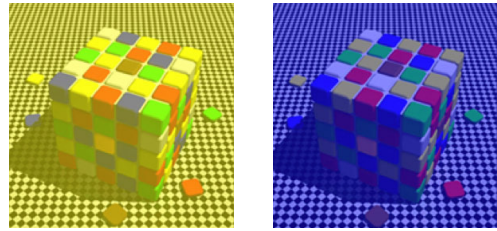


図2 異なる色眼鏡を通して見た同一のルービックキューブ [10]

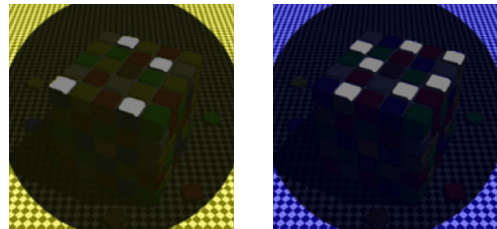


図3 対象となる部位以外を隠した絶対的な色知覚 [10]



図4 ハ長調（上）とト長調（下）の同一メロディー

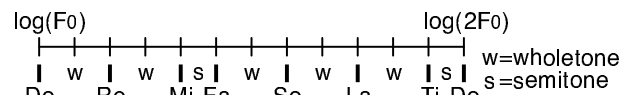


図5 長調におけるオクターブ内の音高配置

の色を「違う」と判断する^(注1)。我々の認知が、個々の要素刺激の物理特性のみでは説明困難であることを示す好例である。

図4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、メロディーの階名書き起こしが可能な相対音間者は、両者を「ソミソド・ラドドソ」と書き起こす。ここで、ハ長調の最初の音と、ト長調の最初の音の基本周波数は異なるが、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番目の音の基本周波数は同一であるが、彼らは異なる音（ソとド）と主張する。異なる音高を「同じ」と判断し、同一の音高を「違う」と判断する。更に色知覚と同様、孤立音のみを提示すれば階名書き起こしは不可能となる。これも、我々の認知が、個々の要素刺激の物理特性のみでは説明困難であることを示す好例の一つである。

心理学研究によれば、これら知覚の不変性は、刺激群の関係性・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[9]~[11]。各刺激の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性・コントラストは不変である。図5に長調のオクターブ内音高配置を示す。「全全半全全全半」という音高配置は調不変であり、メロディー中の二音（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[12]。このような不変の関係性を制約条件として、相対音感者はメロディーを階名で書き起こす。よって孤立音では、階名は知覚できない。

(注1)：冊子が白黒印刷であれば、是非、下記で確認することを勧める。

<http://www.lottolab.org/illusiondemos/Demo%2012.html>

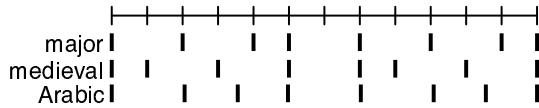


図 6 音高配置の通時的・共時的変形

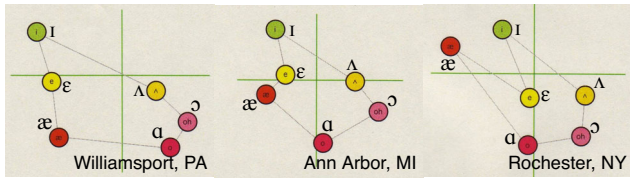


図 7 米語方言における母音配置の多様性 [13]

2.2 関係性・コントラストの通時的・共時的変形

図 5 の音高配置は、音高配置の一例に過ぎず、時代・地域などに依存して異なる音配置が使われている。図 6 に長調、中世の教会音階（の一つ）、アラビア音階を示す。長調の曲を後者二つで演奏すれば、相対音感者であっても（これらの音階に不慣れであれば）階名書き起こしは困難となる。特にアラビア音階での演奏は「調律がずれた楽器」の演奏のように聞こえるが^(注2)、アラブ人にとっては馴染みのあるメロディーとなる。通常、人はその地域・時代で使われる音配置を、音楽経験を通して獲得する。個々の音高の基本周波数を記憶する訳では無い。

同様のことは音声でも成立する。方言である。日本語の場合、方言差はアクセントやイントネーションの変形を指すことが多いが、外国語の場合、韻律的な変形と同様に母音の音色が変形することを意味する。図 7 に米語方言の幾つかを示す。声道長正規化後の幾つかの単母音を第一、二フォルマント周波数平面上にプロットしている [13]。当然、各地方で生まれ育った場合、この母音配置を言語経験を通して獲得することになる。

3. 写像不変量に基づく構造的表象とその応用

3.1 完全なる写像不変量とそれに基づく音声表象

声の多様性は、収録機器や伝送機器の音響特性の差異や、話者の年齢・性別・体格の差異に起因し、音声の音色（スペクトル包絡）を多様に変形する。[6], [7] では、音声の知覚恒常性、言い換えれば、音声の話者不変表象を、前節の知覚の恒常性に倣い、話者不変なコントラスト量のみを用いて導出した。

ある特徴量空間を考える。全ての事象は分布として記述されたとする。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となる（十分性、図 8）。更に 2 事象に関する量を $\oint M(p_1(x), p_2(x)) dx$ で定義した場合、それを変換不変とするのは f -div. のみである（必要性）[14]。

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

この f -div. を用いて一発声を構造化する様子を図 9 に示す。特徴空間における発声軌跡を有限個の分布列に変換し、全ての

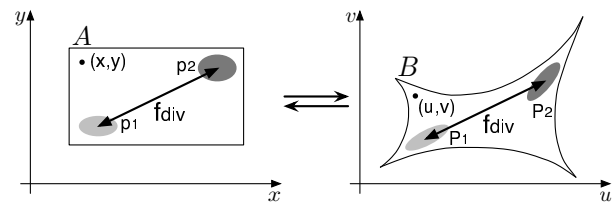


図 8 連続かつ可逆な変形に対して不変な f -divergence

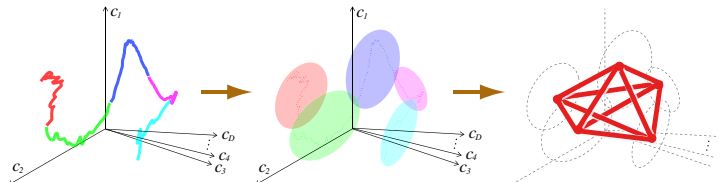


図 9 f -divergence に基づく一発声の構造化

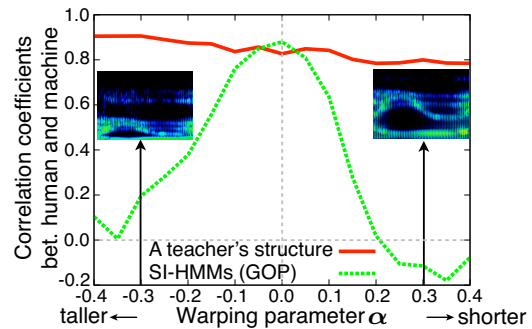


図 10 構造表象と GOP を使った発音習熟度自動評価

分布間距離を f -div. で計測する。最終的に完全写像不変な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（形態）を規定するため、本表象を構造的表象と呼んでいる。

3.2 構造的表象の音声工学応用

構造表象に基づく音響的照合は、話者適応を施した後の音響スコアを、話者適応を明示的に施すことなく算出可能である。そのため、例えば孤立単語認識システムにおいて、話者性に対する極めて高いロバスト性が示されている [8]。なお、全空間写像に不変であれば、異なる単語を同一と判定することが起きてしまう（単語の違いも話者の違い音色の違いである）。話者差を表現する写像群にのみ不変となるよう、不変性の制御を、部分空間における構造表象を用いることで可能にしている [8]

[15] では、外国語発音評価をタスクとして本表象を応用している。話者変換技術を通して、巨人化・小人化した学習者を含め、多様な話者性を持つ学習者群を用意した。そして、各学習者毎に、複数の発声を用いて音素 HMM を構築し、音素距離行列を求めた（変換不変な発音表象）。一人の教師音声から得られた音素距離行列と各学習者行列との比較を通して発音習熟度を推定した。母語話者教師による手動発音習熟度と、自動推定スコアとの相関を求めたところ、一人の教師行列しか用いていないにも拘わらず、高い相関値が得られた（図 10 参照）。

従来法として母語話者 HMM による事後確率を求める GOP (Goodness of Pronunciation) 法を用いたところ、学習者身長の変化に伴い相関値は下落し、やがて 0.0 となった。この下落は、学習者身長に応じて母語話者 HMM を適応することで技術的には対処できる。しかしこの場合、発音の評価ではなく、声

(注2)：「子犬のワルツ」をアラビア音階で演奏した MIDI を下記にアップロードしている。一度聴取することを強く勧める。

<http://www.gavo.t.u-tokyo.ac.jp/~mine/material/western.wav>

<http://www.gavo.t.u-tokyo.ac.jp/~mine/material/arabic.wav>

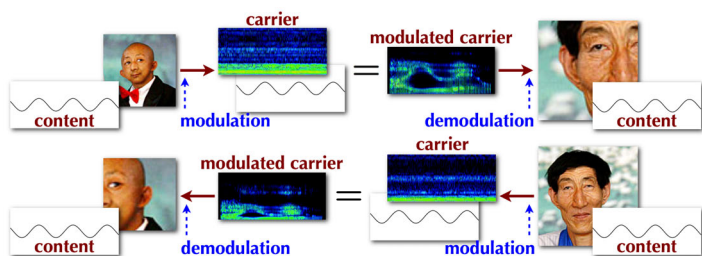


図 13 伝えるべき情報のあるスペクトル変調とスペクトル復調

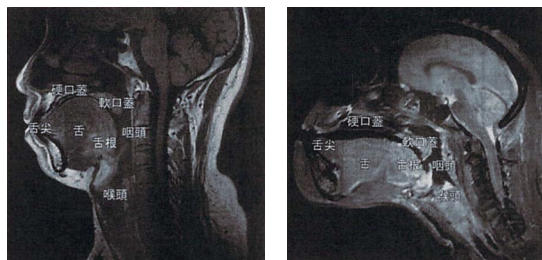


図 14 人間とチンパンジーの構音器官の形態的差異 [27]

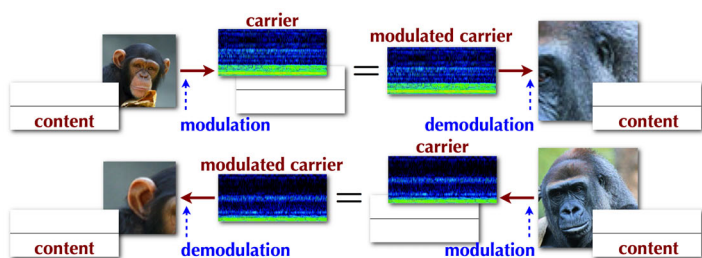


図 15 伝えるべき情報の無いスペクトル変調とスペクトル復調

メッセージ（情報）を取り出すことになる。これが復調である（図 12 は周波数変復調の例）。この観点から音声コミュニケーションを考えれば、音声生成は構音器官を使ったスペクトル変調であり、音声認識は搬送波を除去してメッセージを抽出するスペクトル復調となる。音声コミュニケーションの変復調モデルから図 9 を考えれば、これはスペクトル復調の一方式となる。図 13 は提案する変復調モデルを、概念的に示している。

5.2 進化人類学・心理学的観点から考える変復調モデル

スペクトル変調は、肺からの呼気流を口腔へ送り込むこと、そして、舌運動による口腔形状の変形が必要条件であるが、この二つのプロセスはヒト以外の霊長類では困難であることが示されている。図 14 にヒトとチンパンジーの構音器官の様子を示す。チンパンジーの場合、食道と気道は立体交差しており、基本的には、鼻が気道に通じ、口が食道に通じる。そのため、肺からの呼気流の多くは鼻に抜けてしまうようである [27]。更には、彼らの舌の形状変化の自由度がヒトと比べて極端に低いことも、解剖学的に検証されている [28]。これらを考えると、チンパンジーが柔軟なスペクトル変調器を有することは困難であると言える。図 15 にその様子を概念的に示している。肺からの呼気流の一部は口腔から放射されるが、口腔形状を柔軟に変形させることが困難であるため、変調が起きず、情報を載せることが困難である様子を図は示している。しかし、個体差による口腔形状差異は、当然、ベースラインスペクトルを変形させる。チンパンジーは声による個体同定を行うことができる

が [29]、これは個体間の口腔形状の差異が反映されたスペクトルの差異に基づいた情報処理であると考えられる。

ヒトは何故、スペクトル変調器を有することができたのか？これに対しては、直立歩行による喉頭の下落や、前頭葉の肥大による調音運動能力の増大が要因として挙げられている [27]。

スペクトル変調を通してヒトとサルの違いを考察したが、スペクトル復調に関してはどうかであろうか？主旋律のみから成るメロディー（周波数変調）に対して、ベースラインピッチ（搬送波）を変えることは移調に相当する。ヒトの場合、異なる搬送波を用いたとしても、移調前後のメロディーの同一性を容易に感覚できる。しかし [30]～[32] が示しているように、サルの場合、ベースラインピッチ差異を無視して（搬送波を除去して）、移調前後のメロディー（メッセージ・情報）の同一性を感覚することは難しい。つまり、周波数復調することが難しい。

第 1. 節に述べたように、幼児は体格を越えた（ベースラインスペクトルの差異を超えた）音声模倣を行う。他個体の発声を積極的に模倣する行為は、霊長類ではヒトだけに観測される [33], [34]。それ以外の動物種では、鳥、イルカ、クジラにおいて音声模倣行為が観測されている [34]。しかし、彼らの音声模倣は基本的に音響的模倣、即ち、声帯模写である [34] (注 7)。音声コミュニケーションの変復調モデルから考えた場合、動物の音声模倣行為は、メッセージ（情報）を複製するのではなく、ただ単に、変調された搬送波の複製を試みているに過ぎない。

進化的・心理学的観点から、音声コミュニケーションの変復調モデルを考察した。動物の場合、音ストリームの中に潜む個体差を越えた音響パターンの抽出は困難であり、音そのものから直接的に取得可能な情報のみが処理対象となっている、と解釈できる。現行の音声認識の音響モデリング技術も、音声スペクトル（変調搬送波）そのものをモデル化対象としている。筆者らが提案する音声の構造的表象は、搬送波変調によって送付される情報・メッセージのみを対象としたモデル化を試みており（即ち復調）、進化人類学・進化心理学的に考えれば、技術を進化させることに相当する、と考えている。

5.3 発達障害学的観点から考える変復調モデル

ヒトの場合、構音障害でない限り、スペクトル変調器を所有することになる。しかしこの場合でも、スペクトル復調器に不具合があれば、音声コミュニケーションは困難となることが予想される。メッセージ（情報）の抽出が困難になるからである。

この予想に該当する事例としては、重度自閉症者が挙げられる。音声模倣が声帯模写的になる様子が報告されている。七色の声を持つと呼ばれる声優の中村メイ子の多様な声をそっくり真似る例 [36]、相手の声をそっくり模倣した例 [37]～[40]、声だけに限らず、車のエンジン音、プリンタ音、ドア音など、様々な音響音を声帯模写する例 [39], [40] は、自閉症関連図書には頻出する記述である。刺激音をそのまま記憶し、複製を試みている訳だが、重度自閉症者の場合、音声コミュニケーションの獲得が困難となるのは周知の事実である。中には、母親の音声は

(注 7)：九官鳥やオウムの模倣も同様、音響的模倣である [35]。彼らは車、ドア、犬、猫、そして人の声を真似る。人の声も音でしかない。

理解できるが、母親以外の音声への対応が難しい例もある[41]。しかし電話越しであれば、母親の声も難しくなるようである。

自閉症（アスペルガー症候群）者として世界で初めて書籍を出版した[38] グランディンは動物学の教授であるが、彼女は、自閉症者と動物の情報処理における類似性を指摘している[42]。いずれも、入力刺激の詳細をそのまま記憶・保持する様子を報告している^(注8)。入力情報を無意識的に取捨選択できず（変復調の例で言えば、搬送波を取り除くことができずに）、汎化能力に乏しく、情報過多の渦に巻き込まれる様子は多くの自閉症関連図書に散見する記述である[40],[43]～[45]^(注9)。なお、重度自閉症者のモデルとしてサルを用いた応用行動分析研究例もある[47]。自閉症者の多くは絶対音感保有者である[48]。

6. まとめと考察

筆者は先行研究において、音声科学・心理学の分野で言われる「音声の知覚恒常性」、発達心理学の分野で言われる「語ゲシュタルト」を物理的に説明することを目的として音声の構造的表象を提案し、各種のタスク（音声認識、発音評価、方言分類）において、その技術的有効性を示してきた。本稿では、この構造的表象を、進化人類学、進化心理学、発達心理学、発達障害学、更に、電気通信の観点から考察し、音声コミュニケーションの音色の変復調モデルとしての解釈を述べた。

このモデルに基づいてサルからヒトへの進化を概観してみる。本来、警戒音やクーコールなどのより原始的な伝達、及び、個体を同定するために使われていた声、直立歩行、前頭葉の肥大に起因する形で可能となった柔軟な舌・顎運動によって、音色（スペクトル）が様々に時間変化するようになった（変調）。と同時に、その変調搬送波から、音声生成者の個体差を越えて、メッセージ（情報）を抽出するための復調方式も認知的に獲得するに至った。先天的な構音障害者は変調モジュールに不具合を有することになるが、この場合でも音声認知は健常者と同様に行われる。その一方で、先天的に復調モジュールに不具合があると考察される重度自閉症者は、音声コミュニケーションに困難を抱えることになる。進化人類学では言語起源を論じる時に、構音器官の進化が論じられることが多いが[49],[50]、上記の事実より、変調側（生成側）と同様に、復調側（認知側）への着眼も非常に重要であると、筆者は考えている。

文 献

- [1] 原, コミュニケーション障害学, 20, 2, 98–102, 2003
- [2] 小林他, 言語処理学会第 15 会年次大会発表論文集, P2-16, pp.534–537
- [3] 加藤, コミュニケーション障害学, 20, 2, 84–85, 2003
- [4] 早川, 月刊言語, 35, 9, 62–67, 2006
- [5] P. Lieberman, *Child Phonology vol.1*, Academic Press, 1980
- [6] 峯松他, 信学技報, SP2005-12, 1–8, 2005
- [7] 峯松, 信学技報, SP2008-84, 31–36, 2008
- [8] N. Minematsu *et al.*, *Proc. Speech and Computer*, 35–40,

(注8)：自閉症者はしばしば（両親の発声ではなく）テレビやラジオのコマーシャルを真似るが、グランディンによれば「両親の発声は常に変動するが、コマーシャルは常に同一の音響刺激を提供してくれるから。」とのことである。

(注9)：ある当事者は、自閉症とは「情報の便秘」である、と述べている[40]。同様に、自閉症を（人工知能の世界で言う）「フレーム問題」が解けない症状として説明する書籍もある[45][46]。

- 2009
- [9] R.B. Lotto *et al.*, *Proc. the National Academy of Science USA*, 97, 12834–12839, 2000
- [10] R.B. Lotto *et al.*, *Nature neuroscience*, 2, 11, 1010–1014, 1999
- [11] 谷口, 音は心の中で音楽になる, 北大路書房, 2003
- [12] 東川, 読譜力ー「移動ド」教育システムに学ぶ, 春秋社, 2005
- [13] W. Labov *et al.*, “Atlas of North American English,” *Mouton de Gruyter*, 2005
- [14] Y. Qiao *et al.*, “A study on invariance of f -divergence and its application to speech recognition,” *IEEE Transactions on Signal Processing*, 58, 2010 (to appear).
- [15] M. Suzuki *et al.*, *Proc. Int. Workshop on Automatic Speech Recognition and Understanding*, 574–579, 2009
- [16] J. Morais *et al.*, *Cognition*, 7, 323–331, 1979
- [17] C. Read *et al.*, *Cognition*, 24, 31–44, 1986
- [18] R. Port, *New Ideas in Psychology*, 25, 143–170, 2007
- [19] 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
- [20] D. Saito *et al.*, *Proc. ICASSP*, 4485–4488, 2008
- [21] S. Greenberg *et al.*, *Proc. ICASSP*, 1647–1650, 1997
- [22] H. Hermansky *et al.*, *IEEE Trans. SAP*, 2, 4, pp.578–589, 1994
- [23] “Special Session: Auditory-inspired spectro-temporal features,” *Proc. INTERSPEECH*, 2008 (for example).
- [24] “Special Session: Novel modulation decompositions of signals: theory and applications,” *Proc. ICASSP*, 2010 (for example).
- [25] S. K. Scott, *Proc. INTERSPEECH*, 10–13, Keynote speech, 2007.
- [26] <http://en.wikipedia.org/wiki/Modulation>
- [27] 葉山, ヒトの誕生～二つの運動革命が生んだ奇跡の生物種～, PHP 新書, 1999
- [28] H. Takemoto, *American Journal of Primatology*, 70, 966–975, 2008
- [29] S. Kojima, *A search for the origins of human speech – auditory and vocal functions of the chimpanzee*, Trans Pacific Press, 2003
- [30] M.R. D’Amato, *Music Perception*, 5, 453–480, 1988
- [31] A.A. Write *et al.*, *J. Exp. Psychol. Gen.* 129, 291–307, 2000
- [32] M.D. Hauser *et al.*, *Nature neurosciences*, 6, 663–668, 2003
- [33] W. Gruhn, *In: Proc. Int. Conf. on language and music as cognitive systems*, 2006
- [34] 岡ノ谷, 春季音講論, 1-7-15, 1555–1556, 2008 (質疑応答含む)
- [35] 宮本, 音を作る・音を見る, 森北出版, 1995.
- [36] 深見, ひろしくんの本 (V), 中川書店, 2006
- [37] 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
- [38] T. Grandin, 我, 自閉症に生まれて, 学研, 1994
- [39] L.H. Willey, アスペルガー的人生, 東京書籍, 2002
- [40] ニキリンコ, スルーできない脳～自閉は情報の便秘です～, 生活書院, 2008
- [41] 東田他, この地球にすんでいる僕の仲間たちへ, エスコアル, 2005
- [42] 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
- [43] 綾屋他, 発達障害当事者研究, 医学書院, 2008
- [44] 泉, 僕の妻はエイリアン, 新潮社, 2005
- [45] 藤井他, 自閉症, 新曜社, 2007
- [46] 渡部, 鉄腕アトムと晋平君, ミネルヴァ書房, 1998
- [47] 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
- [48] U. Frith, 自閉症の謎を解き明かす, 東京書籍, 1991
- [49] W. T. Fitch, *Trends in Cognitive Sciences*, 4, 7, 258–267, 2000
- [50] W. T. Fitch *et al.*, *Proc. the Royal Society*, B, 268, 1477, 1669–1675, 2001