

[依頼講演] 構造音韻論の物理実装に基づく学習者表象と発音学習支援

峯松 信明[†]

[†] 東京大学大学院情報理工学系研究科

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

あらまし 「二人として同じ学習者はいない。」「地球上には約 20 億の英語学習者が存在する。」この二つの事実は、地球上に約 20 億種類の英語が存在することを意味する。学習者を記述する文字の創成を考えた場合、この 20 億の対象物を描ききる枠組みが必要となるが、学習者が生成した文の誤りや、生成した音声の誤りを記録するだけでは不十分である。人間の言語活動を「脳内言語システム」の出力と捉え、そのシステム動作を決定する各種パラメータ群を定義することで統一的な記述方法を考案し、パラメータ値の差異でもって 20 億の学習者を描く方法論の確立が必要である。言い換えれば、言語学という学問を人間一人一人の言語活動を記述する学問として再構築する必要がある。本研究では言語学の祖と言われるソシュールの主張を数学的・物理的に解釈することで構造音韻論を物理実装し、その上で、20 億の学習者を描ききる手法を提案すると共に、可能となりつつある発音学習支援環境について報告する。

キーワード ソシュール言語学, 構造音韻論, 構造不変の定理, 音響的普遍構造, 発音学習

[Invited talk] Physical implementation of structural phonology and its use to represent individual language learners

Nobuaki MINEMATSU[†]

[†] Graduate School of Information Science and Technology, University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0031 Japan

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

Abstract No two students are the same. There are 2 billion students of English on this planet. These two facts indicate that there are 2 billion different kinds of Englishes on this planet. A good method to describe every student has to adequately deal with the 2 billion different students. But it is clear that detection of errors in sentences or speech generated by students is not sufficient by itself. The true method is supposed to regard linguistic activities of a human comprehensively as output of his/her linguistic system in brain. The system is represented by the method as a set of parameters and differences of parameter values are expected to reflect those between two students. In other words, linguistics has to be reconstructed so that it can describe not only languages but also individual humans. In this work, Sausurre's claim on language is interpreted mathematically and structural phonology is reconstructed on physics. Using physically-implemented structural phonology, some proposals are made for pronunciation training.

Key words Sausurre, structural phonology, theorem of the invariant structure, acoustic universal structure, pronunciation training

1. はじめに

How are you? Fine, thank you. And you? 数十年にも渡って繰り返されてきた会話である。筆者はこの会話が、現在の外国語学習の実態を見事に表現しているように感じている。彼らは本当に Fine なのだろうか? How are you? を「あなたはどのような状況にありますか?」と解釈した場合、正しい応答は I don't know how I am. You don't know either, do you! であろう。「二人として同じ学習者はいない。」「地球上には 20 億の英語学習者がいる。」対象言語を英語に限定した場合でも、20 億種類の異なる英語が存在することになる。この 20 億の対象

物を描ききる枠組みが存在して初めて How are you? と問いかける権利が教師に与えられるべきでは無いだろうか? 記述を可能とする文字がなければ、あるのはカオスのみである。

言語学、そしてその下に位置する音韻論、音声学。言語学や音韻論は「言語」を記述するために生まれた科学であり、音声学は「言語音」をその調音特性、音響特性に基づいて記述するために生まれた科学である。「人間」を単位としてその言語活動を描く科学はどこにあるのだろうか? 本年になって、言語障害学と言語学とが対話を始めた [1]。脳障害により言語活動の一部が阻害された方々を一人一人描くために、チョムスキーの普遍文法がどこまで寄与できるのか、その模索が始まった。人間

があつて初めて言葉は存在する。学習者のみならず、母語話者として二人として同じ言語を運用する人はいないと考えれば、本来言語学は人間を単位として生まれるべきではなかったのか？

人間を単位として言語学を再構築する場合、各人間が一つの言葉を保有するとしても、その数は60億を越える。これらを全て描く方法論は可能だろうか？言語学者（即ち人間）の作業には落せない。可能とするためには、計算機上で行なうしか道は無い。計算機上で行なうためには、言語学を数学的に、物理学的に捉え直す必要がある。本研究では筆者が提唱する「構造不変の定理」という数学的事実を用いてソシュールの言語哲学を数学的、情報理論的に再解釈することから議論を始める。

本研究の対象は発音学習である。この場合、調音音声学や音響音声学など既に物理学上の音声科学が存在する。しかし、これらが提供する表象は性別、年齢、話者性、音響機器特性、伝送特性など種々の非言語的な要因が混入し、不可避免的に、雑音混じりの表象とならざるを得ない。また、学習者の発音の誤りを逐一検出したとして、How are you?という問いに答えるための「全学習者に対する統一的な記述方法」が得られた訳ではない。本来行なうべきは、人間の言語活動を「脳内言語システム」の出力として捉え、そのシステム動作を決定する各種パラメータ群を定義することで統一的な記述方法を考案し、パラメータ値の差異でもって学習者間の違いを論じる方法論の創成である。学習者の出力したサンプル群を観測するだけの方法論では、学習者群の分類すら十分には行なえない。その裏に潜む歪んだシステムを描くことが必要である。発音学習を対象にした場合、これは音韻論の物理的実装を意味することになり、本研究では、構造音韻論を「構造不変の定理」を用いて物理実装する。

2. 構造不変の定理

ある数学的定理（事実）について述べる。本定理は、複数の物理現象から構成される幾何学的構造が、ある数学的変換に関して不変であることを保証する。なお、ユークリッド空間に存在する n 点に対して、全ての二点間距離（即ち距離行列）を求めると、その n 点で張られる構造は一意に規定される。

意味のある記述が分布としてのみ可能な物理現象を考える。複数の分布に対して（分布群）、全ての二分布間距離（距離行列）を求める。二分布間距離として、バタチャリヤ距離やクルバック・ライブラ距離を用いた場合、各分布に対して単一の任意一次変換（アフィン変換）を施しても、二分布間距離は不変である。任意の三分布間距離が三角不等式を満たす場合、この距離行列によって張られるユークリッド空間の構造は一意に規定され、それは一次変換によって一切不変である。

分布間距離尺度としてバタチャリヤ距離を考える。

$$BD(p_i(x), p_j(x)) = -\ln \int_{-\infty}^{\infty} \sqrt{p_i(x)p_j(x)} dx \quad (1)$$

ここで $q = \int_{-\infty}^{\infty} \sqrt{p_i(x)p_j(x)} dx$ とすると

$$0.0 \leq q \leq 1.0 \quad (2)$$

となり、これをある事象が起きる確率として解釈すれば^(注1)、 BD はその事象の生起に対する自己情報量として解釈される。二つの分布が単一ガウス分布に従う場合、以下が成立する。

$$BD = \frac{1}{8} \mu_{ij}^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} \mu_{ij} + \frac{1}{2} \ln \frac{|(\Sigma_i + \Sigma_j)/2|}{|\Sigma_i|^{\frac{1}{2}} |\Sigma_j|^{\frac{1}{2}}} \quad (3)$$

μ_i は i の平均ベクトル、 μ_{ij} は $\mu_i - \mu_j$ を、 Σ_i は i の分散共分散行列である。上式に対して $x' = Ax + b$ と一次変換を施しても距離は不変であり、また、各分布が混合ガウス分布に従った場合でも不変性は保存される。第一項は平均ベクトル差異に基づく距離であり、分散共分散行列の平均行列で両分布状況を代表させた場合のマハラノビス距離である。第二項は平均ベクトルに無関係に定まる量であり、両分布の分布形状差異に起因する距離である。さて、事象 k が生起する確率を q_k とした場合、その q_k に対して $q_k = \int_{-\infty}^{\infty} \sqrt{p_i(x)p_j(x)} dx$ を満たす分布 i, j を規定すれば、事象 k に対する自己情報量は分布 i, j 間のバタチャリヤ距離として幾何学的に解釈することが可能であり、全事象より得られる情報源の（平均）情報量はこのバタチャリヤ距離の期待値となる。距離の大小を分布間の関係の大小として解釈すれば、シャノンの情報理論が定義する全ての情報は、ある物理現象間の関係の大小として解釈することも可能であろう。

なお、バタチャリヤ距離で構成される距離行列において、必ずしも任意の三分布間距離が三角不等式を満たす訳ではない。この場合、ユークリッド空間において構造を張ることは不可能となるが、 n 分布間の情報論的關係は一切不変である。

広く知られているように、点群で張られる構造に対して一次変換（アフィン変換）を施すと、それは構造を歪める。アフィン変換は画像処理の世界では、構造を歪める変換として広く使われている。物理現象を点ではなく分布として捉えることで、即ち、現象の存在を非局所的に捉えることで、かつ、事象間の距離を情報論的に解釈することで、構造の不変性は数学的に保証される。一次変換を $x' = Ax + b$ とした時に、行列 A の乗算、ベクトル b の加算に何らかの明確な物理的意味を考えるとができる場合、張られる構造はこれら物理的操作に対して一切不変であり、ある種の普遍構造を呈することとなる。

3. 不可避免的に混入する非言語的特徴のモデル化

「構造不変の定理」を用いたソシュール言語哲学の数学的解釈を述べる前に、音声コミュニケーションチャンネルに不可避免的に混入する非言語的特徴を、工学的にモデル化する。ここでは議論を簡単にするために、最も単純化したモデルを考える。

音声認識の世界では、音声コミュニケーションに付随する雑音・歪みは主に三種類に分類される。加算性雑音、乗算性歪み、そして線形変換性歪みである。テレビ、ラジオなどの背景雑音が加算性雑音の典型例であるが、この種の雑音は考慮しない。それらは物理的な抹消が可能であり（スイッチを消す、別の部

(注1) : q の定義より、次元的には確率となる。また、対象とする事象であるが「二つのサンプル群 i, j に対して両サンプル群が同一の母集団からの任意抽出であると観測者が認識する」という事象もその例として考えられる。

屋に移動するなど), 不可避な雑音ではないからである。

収録系, 伝送系, 再生系の音響機器による差異が生じる歪みが乗算性歪みの典型例である。スペクトル系列の時間平均パターンが話者特性を近似できることから分かるように, 話者性の一部も乗算性歪みとなる。この種の歪みは意味論的にゼロ点が存在せず, 抹消することは不可能である。音声事象をケプストラムベクトル c で表現した場合, 乗算性歪みはベクトル b の加算として表現され, その結果, $c' = c + b$ となる。

話者が異なれば声道形状は異なる。聴取者が異なれば聴覚特性は異なる。聴覚特性の差異は聴取後の聴覚イメージを聴取者間で異なるものにする。なお, メル尺度やバーク尺度は平均特性に過ぎない。これらは線形変換性歪みの典型例であり, 不可避な歪みである。声道形状の差異, 聴覚特性の差異も対数スペクトルに対する周波数ウォーピングとして工学的に実装される。[2] によれば, 単調増加かつ連続であれば, 如何なる周波数ウォーピングもケプストラムの次元では, 行列 A を掛ける演算に数学的に変換される。即ち, $c' = Ac$ である。

音声コミュニケーションには様々な雑音・歪みが混入するが, 不可避的なもののみに限定した場合, それは乗算性及び線形変換性の歪みとなる。異なる歪み源は異なる A 或いは b を呈するが (A_i と b_i), これらの歪みが統合されて得られる (統合) 歪みも, 最終的には $c' = Ac + b$ として記述される。即ち不可避的な歪みはアフィン変換としてモデル化される。

4. 構造音韻論の物理実装と学習者表象

4.1 ソシユール言語哲学と構造音韻論

言語学の祖, 構造主義の祖と言われるソシユールによれば「言語が含むのは, 言語体系に先だって存在する観念でも音でもなくて, ただ, この体系から生じる概念的差異と音的差異だけである。」[3] この言語哲学に啓蒙され, Jakobson によって弁別素性が提案され [4], それに基づいて構造音韻論は確立された。ソシユールの言語哲学に基づいて考えた場合, 弁別素性は本来, 異なる音素間の差を表現するための手段である。音素を弁別素性の束として考え, 各音素を (他音素との関係を直接的に議論せずに) 定義する方法論は, ソシユール言語哲学から乖離するものである [5]。音素間の差・関係のみに着眼した例として, Jakobson によるフランス語音素の幾何学的構造 [6], 及び Halle によるロシア語音素の樹型図 [7] を上げることができる。図 1 に Jakobson による音素群の幾何学的構造を示す。

4.2 構造不変の定理に基づく構造音韻論の物理実装

音韻論では, 音声に不可避的に混入する非言語的な特徴は, 研究者の頭の中で一切捨象される。捨象の後, 彼らは何を見たのか? 図 1 に示す, 音素群を相対的に構造として捉える方法論を純粹に幾何学的に考えれば, 各音素の絶対的な (位置) 情報が捨象される。音響空間内に音素群を配置し, その構造を考える。この時, 捨象される情報が非言語情報となれば, 不可避的に混入する非言語的な特徴を保持しない音声の構造的・物理的表象が定義される。さて「構造不変の定理」より, 一次変換に対して不変な構造を定義することが可能であることが示されている。また, 不可避的に混入する非言語的な特徴は (最も簡略

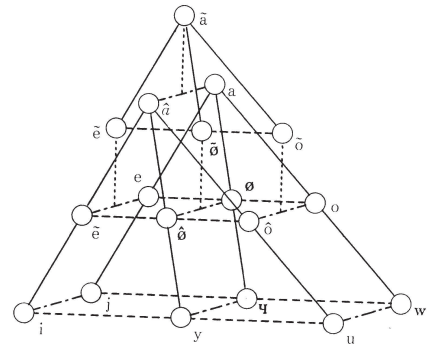


図 1 Jakobson によるフランス語音素の幾何学的構造

Fig. 1 Jakobson's geometrical structure of some French phonemes

なモデル化を行えば) ケプストラムベクトルの一次変換となる。問題は音素を分布として表現することの妥当性である。

自然音声から抽出された 1 ピッチ波形を計算機上で繰り返して接続すると (即ち完全なる周期波形とした場合), それはブザーのような音となる。即ち, (周期性の意味で) 完全なる音声はブザー音となる。人間はこのような完全なる音声は生成できない。人間は不完全なる音生成器であり, その意味において音声は不可避的に揺れており, 意味のある音声の物理表象は音響空間内の点ではなく, 分布として行なうべきである。

以上の考察より, 音素を音響空間内の分布として捉え, 分布群となった音素群に対して「構造不変の定理」を適用して得られる構造が, 構造音韻論の物理実装となり, その構造は非言語的な特徴に対して一切不変となることが示された。本研究ではこの構造を音響の普遍構造と呼ぶ。点群によって構成される構造に対するアフィン変換は, 幾何学的には, 拡大・縮小, せん断, 回転, 平行移動などに分類される。これらの中で二点間距離が不変な変換は回転と平行移動である。即ち A を掛ける演算は構造の回転として, b を足す演算は構造の平行移動として解釈される。人間の成長による音声の音響的变化が声道長の延びのみによって生じるならば, その変化は音韻構造の回転として解釈される。15 年ほどかけてゆっくり回る回転である。なお, 音素を混合ガウス分布として近似した場合でも, その二音素 (分布) 間距離は一次変換によって不変である。

以下, 本音声表象で学習者発音を個人を単位として網羅的に記述することによって何が可能となるのか, という点に絞って現在までに得られている結果を概観する。個々の結果の詳細は参考文献を参照して戴きたい [8]~[11]。

5. 20 億人の学習者の記述を目指して

5.1 1 学習者の記述

音素群の情報論的構造化によって, 音声物理に不可避的に混入する非言語的な特徴は, それを表現する次元を失い, 残された構造的表象は純粹に物理的 (音響的) かつ言語的表象と言える。学習者音声よりこの音響的・言語的構造を抽出した場合, そこには何らかの構造的歪みが観測されることが容易に推測される。[13] では乳児の言語獲得を, 一つ一つの音の獲得ではなく, 音群の中に張られる関係・構造の獲得であるとモデル化している。即ち, 粗い構造から詳細な構造へと音韻構造意識が変

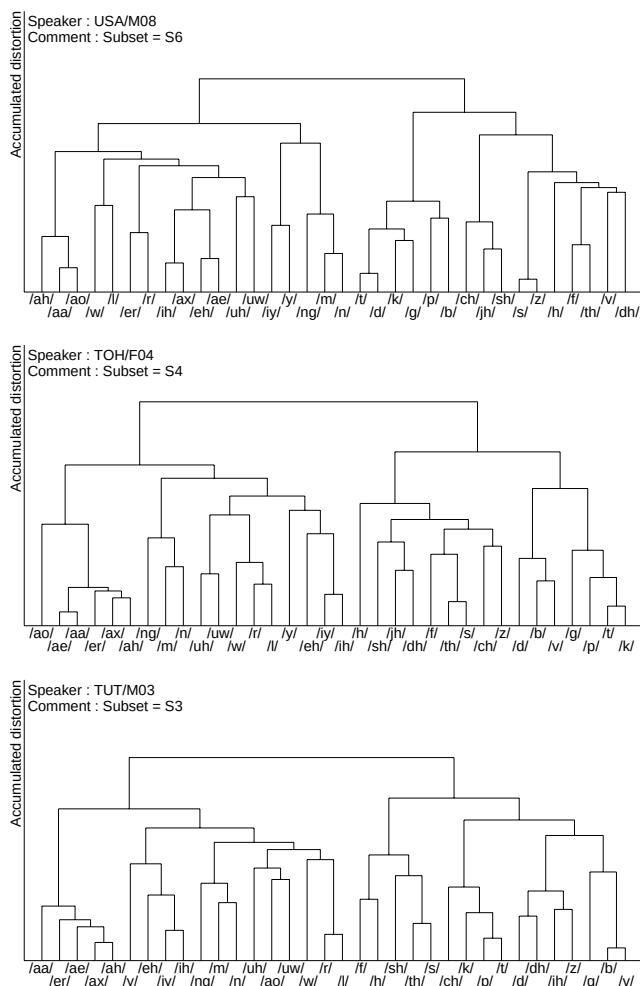


図2 米語話者（上）と日本人（中・下）に対する音素樹型図
Fig. 2 Phoneme tree diagrams produced by an American and two Japanese students

化することこそ、言語意識の獲得であるとしている。音韻構造意識という観点から外国語学習を考えた場合、それは歪んだ構造から正常な構造への遷移としてモデル化され、本研究で提案する音響的普遍構造は、この歪んだ構造として学習者一人一人を描く方法論を提供することが可能である。

ERJ (English Read by Japanese) データベース [12] 中の各話者に対して、特定話者音響モデル (HMM) を構築し、ボトムアップクラスタリングにより、各話者の音素樹型図を作成した。幾何学的構造は距離行列と等価であるが、樹型図は距離行列 (即ち構造) を二次元平面上に視覚化したものである。図2に米語話者と日本人による音素樹型図の例を示す。ERJ データベースによれば、二人の日本人学習者の音素生成に関する発音習熟度 (英語教師による評価) は 2.25 (中) と 2.20 (下) であった。日本人樹型図には /r/ & /l/, /s/ & /th/, /z/ & /dh/, /f/ & /h/, /iy/ & /ih/, /v/ & /b/ などの子音混同、/ah/, /ax/, /er/, /ae/ などの母音縮退、更には schwa (/ax/) の不適切な発声など、日本人英語に特徴的な癖が明確に表現されている。英語音声学によれば、母音は弱勢となると schwa に近づくと言われるが、schwa が調音的にも音響的にも母音中心に位置していることを考えると、これは母音構造のサイズが小さくなることと同値で

ある。既に強勢・弱勢母音を用いた分析により、構造のサイズが調音努力におよそ対応することを実験的に確認している [11]。なお、樹型図は距離行列を視覚化する一手法に過ぎず、多次元尺度法など、より良いインターフェイスを模索する必要がある。

5.2 2 学習者間の距離尺度

2 学習者を比較する場合、距離行列の各要素を比較することも可能である (局所的比較)。ここでは、2 つの構造間の大局的差異をスカラー量で表現することを考える。A の乗算は回転であり、b の加算はシフトである。一つの構造を回転及びシフトさせ、他方の構造と対応する点 (分布) 同士を近づける形で重ね合わせる。回転及びシフト後の対応する点 (分布) 間距離の総和の最小値を「構造間差異」として定義する。得られた距離行列を点間距離の集合として考え、また、距離行列における全ての 3 点間距離が三角不等式を満たしている場合を考える。この場合純粋にユークリッド空間における二つの構造の重ね合わせ時の構造間差異となる。既にこの差異を、距離行列のみから解析的に求めることに成功しているが、ここでは誌面の都合上割愛させて戴く。なお、距離行列をベクトルと見なして計算される 2 行列間のユークリッド距離が、A と b に対する適応処理 (即ち回転とシフト) を施した後の、対応する音素間距離の平均値に近似的に比例することを、実験的に導いている [9]。

5.3 全学習者の分類

2 学習者間距離がスカラー量で表現されると、任意の学習者集団に対し、学習者距離行列が定義でき、学習者全体をボトムアップクラスタリングすることが可能となる。例えば英語学習者は地球上に約 20 億人存在すると言われるが、彼らを個人の単位で記述し、全学習者を分類することは、技術的には可能である。これが実現されれば、地球上にどのような英語がどのような状況で分布しているのか、その地図を描くことが、学習者個人を単位として行なえることになる。既に [9] において、ERJ の 202 人の話者に対するボトムアップクラスタリングを試みているが、音素間距離 (音韻群構造) は、読み上げ文セットや読み上げ方にも依存してくるため^(注2)、意味のある学習者樹型図を得るためには、細かく収録条件を制御することが必要である。ERJ における文セット 6 に対して行なった学習者の分類の様子を図3に示す。リーフノードの数値は各学生の発音スコア (5 点満点、米国人教師による評価) である。約 30 人の日本人を 3 つに分類した時に、一人の学生で 1 クラスタが形成されているが、この話者は 202 名中の唯一のバイリンガル話者である。

6. 音韻構造歪みに基づく発音評価

6.1 教師・学生間における音韻構造歪みの定量化

2 者間の大局的な構造的比較を行ない、その差異をスカラー量で表現できることを示した。これは、学習者と教師とを、その音韻構造歪みのみに基づいて比較できることを意味する。上

(注2)：例えば、音素の音響的実態である単音の音響的特性は、前後の音素によって変化するため、音響モデルとして得られる分布も音素環境に依存したものとなる。また、発話速度によっても構造は変化する。例えば発話速度を落した場合、各音素に対応するスペクトルがより定常となるため、分布が急峻となり、構造は全体的に大きくなる傾向がある。

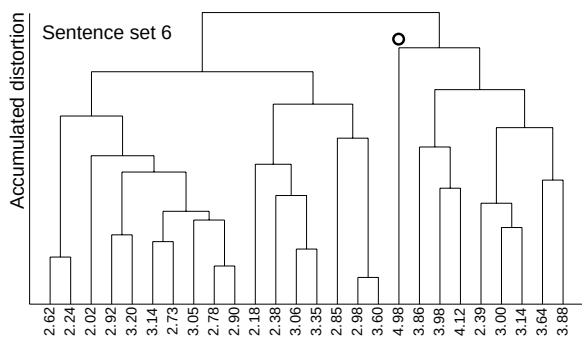


図3 文セット6の話者に対する分類

Fig. 3 Bottom-up clustering of the students of set-6

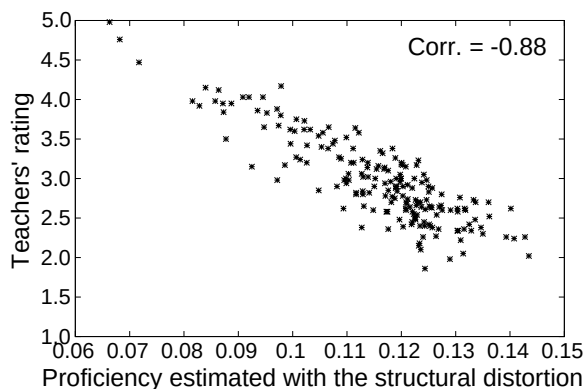


図4 音韻構造歪みに基づく自動発音評定

Fig. 4 Automatic rating based on phonological distortions

述した2話者間距離を用いてERJの日本人学生を自動評定した結果が図4である。横軸が構造間差異による自動評定点、縦軸が教師による評定点である。非常に良好な関係が得られている。各話者は8つの文セットの何れかを読み上げており、文セットの違いに起因する構造歪みが埋め込まれた形で、各学習者の構造的表象が求められている。そのため、教師との差異（行列間差異）を求める際に相関値が最大となるよう、距離行列に対してマスキングを施している[9]。

6.2 発音評定における究極の安定性を求めて

筆者は英語劇の舞台役者としての経験を持ち、口周り・腹周りの筋肉武装から英語発音に取り組んだ一人である。しかし純粋な日本人である以上、日本語訛りの英語を発声することも容易である。そこで、以下に示す興味深い実験を行なった。

筆者による英語音声（通常の英語A、故意に訛った英語B）、及び米語話者二名による英語音声（男性M、女性F）、合計4種類の英語音声を用意した。F/A/Bの様子を表1に示す。MとFは話者が異なるだけで収録条件は同一である。次にB、M、Fから音響モデル（HMM）を作成した。Aをこれら3種類の発音と比較した場合、AはBよりもM、Fに音響的に近いと判定されるのだろうか？英語教育学的にAはBよりもFに近いと判定されるべきであるとするならば、この条件が音声工学的には最も困難な条件である。表1に示されているように、AとFは習熟度以外の、非言語的な特徴は全て異なっている。一方、AとBは習熟度という言語的な特徴以外は、全て揃えてある。即

表1 実験で使った三種類の英語発音

Table 1 Three kinds of pronunciations used in the experiment

話者	USA/F12(F)	筆者 (A)	筆者 (B)
性別	女	男	男
年齢	??	36	36
マイク	Sennheiser	特価品	特価品
録音室	防音室	リビング	リビング
AD	SONY DAT	PowerBookG4	PowerBookG4
習熟度	perfect	good	Japanized

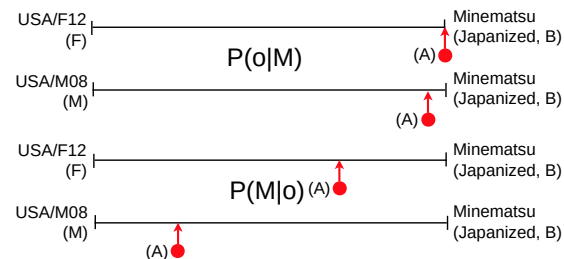


図5 $P(o|M)$ 及び $P(M|o)$ に基づく自動発音評定

Fig. 5 Proficiency rating with $P(o|M)$ and $P(M|o)$

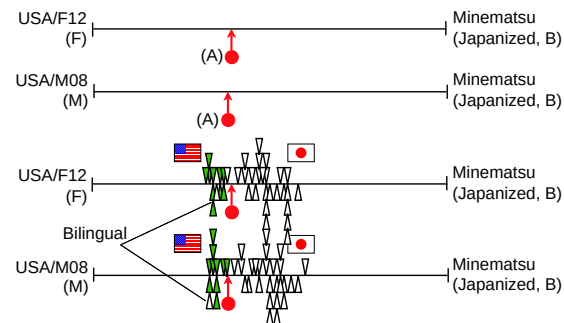


図6 音韻構造歪みに基づく自動発音評定

Fig. 6 Proficiency rating with structural distortion score

ち、発音比較を行なう技術が、音声に不可避免的に混入する非言語的な特徴に左右されると、AをBよりFに近いと判定することは甚だ困難となる。類似性を判定するための尺度としては、1) 音声認識で広く使われる音響スコア $P(o|M)$ を使用、2) 入力話者と音響モデルとの相性を正規化する目的で広く使われる音響スコア $P(M|o)$ を使用、3) 本研究で提案する構造歪みに基づく音響スコアを使用、の3種類の場合を検討した。

結果を図5、図6に示す。 $P(o|M)$ の場合、Aは限りなくBに近いと判定される。全ての音響的特徴を用いて判定していることを考えれば、当然の結果である。 $P(M|o)$ 使用時は、MとFとで評定が異なっている。相性を正規化する技術を用いているにも拘わらず、異なる教師が異なる評定を下している結果となった。図らずも、CALLシステムが時として「不安定性である」と敬遠される様子が露呈された。構造歪みに着目したスコアでは、安定した判定が行なわれている。図6には、同一文セット（セット6）を読み上げた米語話者、他の日本人の位置も示している。バイリンガル話者以外の全ての話者が、筆者（A）以上に筆者（B）に「音響的に」近いと判定されている。これは、スペクトル比較を原理とする方法論では不可能である。

7. 言語構造歪みに基づく発音評定

7.1 英語特有の語彙構造との相性に基づく発音評定

日本人英語では音素混同が頻発する。音素混同は異なる単語の音響的類似性を高め、音響空間内の語彙密度の増加を促す。その結果、単語混同を誘発し、不明瞭な英語となる。単語を音素連結で表現した場合、単語間距離は音素系列間距離となり、これは学習者表象で用いていた音素間距離行列のみを用いて計算できる。即ち、語彙中の全単語に対して単語間距離を求めれば、図1で示した幾何学的構造は、単語をノードとして構築可能である。そしてこの単語群構造は母語話者と学習者で構造的差異を持つこととなる。当然のことながら、提案する学習者表象（距離行列）及び英語語彙だけを用いて語彙密度も計算可能である。発音教育の目標を native-sounding な発音ではなく intelligible な発音とする声が高まってきているが、intelligible な発音を物理的に定義することが困難であることも事実である。「語彙密度が低い発音がより intelligible である」と考えればその定義は可能となる。図7に推定された語彙密度と教師による発音スコアとの相関の様子を示す。本手法は、学習者と教師との間の音響的な比較を一切行わず、学習者の発音が、対象とする言語が持つ語彙構造にとってどの程度「相性が良いのか」という観点から評定した結果であると解釈できる。

7.2 矯正すべき音素順序の推定

提案する学習者表象（距離行列）は、非言語情報を一切含まないため、学習者・教師間で「関係の部分置換」が意味を持つ。即ち、行列の一部を置換することで語彙密度の大幅な低減が可能である場合、その関係の是正が、学習者にとってまず初めに矯正すべき対象であると解釈できる。「音素 p と他の全ての音素との距離群」を「音素 p の部分関係」と定義し、どの部分関係置換を優先すべきかについて、学習者 RYU/F06 を対象に検討した。結果を図8に示す。音素の矯正順序である。日本人が頻繁に混同する音素が優先度高く抽出されている。本手法を用いれば、原理的には、20億の学習者に対して20億通りの教示が可能となる。また部分関係置換後の音素樹型図を、発音学習直後の学習者状況の予測図として提示することも可能である。

8. ま と め

外国語学習及び言語能力回復を目的とした場合の言語学のあるべき姿を考察し、言語学を、人間を単位としてその言語活動を記述する学問として再構築する必要性を述べた。その再構築は、言語学を数学・物理学の上に構築し、計算機上のアルゴリズムの形で表現することを要求する。そこで、ソシュール言語学の数学的・物理学的・情報論的解釈を行なった。筆者の解釈によれば、筆者が提唱する「構造不変の定理」を音声物理に適用したものが「音的差異」であり、また、全「音的差異」を集め、幾何学的構造に落したものが構造音韻論の物理実装である。物理実装された構造音韻論に基づいて学習者表象を提案し、地球上の全ての英語の生息の様子を地図として描くことが可能であることを述べた。更に、安定した発音評定及び、20億通りの音素矯正順位など、発音学習への応用についても報告した。

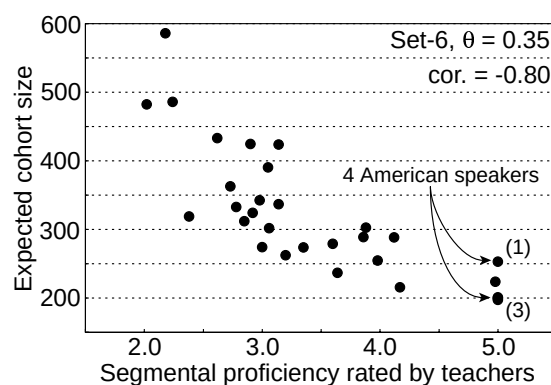


図7 語彙密度に基づく発音の自動評定

Fig.7 Result of pronunciation proficiency estimation based on lexical density

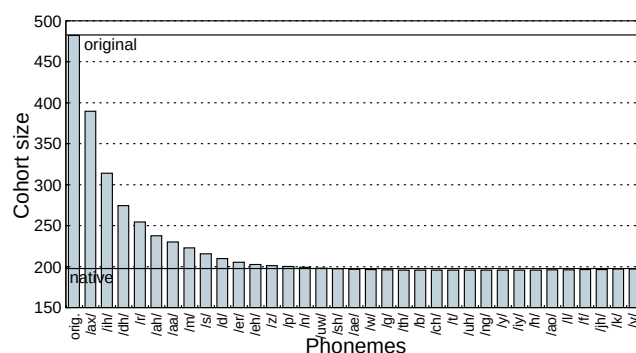


図8 関係の部分置換に基づく語彙密度の減少

Fig.8 Lexical density reduction by sequential replacements of phonemic relations

文 献

- [1] 上智大学国際シンポジウム基礎資料「言語学と言語障害学との接点」(2004)
- [2] M. Pitz *et al.*, "Vocal tract normalization as linear transformation of MFCC," Proc. EUROSPEECH, pp.1445-1448 (2003)
- [3] ソシュール, 小林英夫訳, "一般言語学講義", 岩波書店 (2003)
- [4] R. Jakobson *et al.*, Preliminaries to speech analysis: the distinctive features and their correlates, MIT Press, Cambridge (1952)
- [5] 例えば, 2004年5月に開催された日本フランス語学会シンポジウム「ソシュール研究の現在」にて加賀野井秀一(講談社選書メチエ「知の教科書・ソシュール」の著書)の弁
- [6] ローマン・ヤコブソン, 服部四郎編, "構造的音韻論", 岩波書店 (1996)
- [7] M. Halle, The sound patterns of Russian, The Hague: Mouton (1959)
- [8] 峯松, "音声に内在する音響的普遍構造とそれに基づく語学学習者モデリング", 信学技報, SP2003-179, pp.25-30 (2004)
- [9] 峯松, "音声の音響的普遍構造の歪みに着目した外国語発音の自動評定", 信学技報, SP2003-180, pp.31-16 (2004)
- [10] 峯松, "音響的普遍構造と言語的普遍構造の整合性に基づく発音明瞭度の測定", 信学技報, SP2003-181, pp.37-42 (2004)
- [11] 朝川他, "音響的普遍構造のサイズと単語境界における音響的分離度に着目した米語音声の音響分析", 信学技報, SP2004-28, pp.53-58 (2004)
- [12] 峯松他, "英語 CALL 構築を目的とした日本人及び米国人による読み上げ英語音声データベースの構築", 日本教育工学会論文誌, vol.27, no.3, pp.259-272 (2004)
- [13] S. E. Blache, The acquisition of distinctive features, University Park Press, Baltimore (1978)