

音声研究 ----私の反省と提言----

Review of my Speech Research and Proposal of a New Goal.

中島隆之†
Takayuki Nakajima

1. まえがき

昭和 39 年に国研(旧電子技術総合研究所)に入所、国家プロジェクト「パターン情報処理」で音声認識WGを担当するなど 20 余年間音声研究に従事した。その後、産業界に移り、シャープ(株)技術本部の研究所ではVRの研究や国家プロジェクトRWC Pのマルチモーダル・インターフェース等の研究を推進、また、(株)パトリスにおいては特許文書の自動処理の研究など一貫して研究生生活を過ごした。

本文では、国研における音声科学、音声情報処理技術の推進、そして産業界における商品化の過程に触れた者として、各過程を振り返るとともに今後の音声研究への期待を述べたい。

なお、本文は文献[1]を基礎に各テーマに至った動期や感想を主にして書き直したものである。

2. 調音次元の音声処理

私は昭和 44 年頃音声研究をはじめた。音声波形の分析により音韻情報を取り出すことを目的に、それまでにあった代表的分析法およびカルマンフィルタを音声分析に適用する研究など独自にプログラムして実際の音声に適用した。利用可能主記憶装置が数 10 キロ語程度の電子計算機を賃借して約 2 年間「紀伊半島に小型の台風 7 号が発生し」という 2.7 秒間の音声だけを対象にしてスペクトルパターンを XY プロッターで書き出し、厚紙にコピーして切り出し、時間方向に張り合わせて立体模型をつくった。分析法やパラメータを変えた幾つもの模型を作りあらゆる角度から観察した。昭和 45 年パターン情報処理大型プロジェクト(PIPS)が発足した。パターン情報部音声認識研究室という名の研究室が新設されて、そこで音声認識を目的とする研究を開始した。

当時話題になり始めていた、板倉氏らの線形予測分析法さらに PARCOR 分析法を追試して、これが有望な方法として目に留まった。中でも PARCOR 係数が不均一径音響管の反射係数に形式的にせよ対応しているという知見は、音声声道という調音次元のパラメータで表現できる可能

性を示すのではないかと直感させるものがあり興奮を覚えた。筆者のオリジナルな音声研究はこれを基点としているが、その直感が生じた要因には、筆者が以前に属していた応用部音響研究室の「アナログ声道モデルによる音声合成の研究」で声道に馴染んでいたことがあったと思う。

2.1 適応逆フィルタ法による声道形推定法[2]

連続音声においては、調音結合が強く表われ音節などの短い単位の標準パターンによって認識することを困難にしている。調音結合のしくみを明らかにするためには発声器官の効きの特徴に立ち戻った考察が有効として、東大、他では発話時の X 線映画の解析によっても研究が進められているが、音声から声道形を推定する逆問題は興味ある問題であった。

私は、「PARCOR 係数が、不均一径音響管の反射係数に形式的に対応している」ことがこの問題解決に使えるかを実験的に確かめることにした。

当初は、実音声の母音部波形を細工して、これを偏自己相関分析して音響管形状にもどしてみたが、見通しが得られなかった。そこで基本に立ち返り、声道のモデル化によるシミュレーション実験をして見ることにした。すなわち、種々の音響管形状を与え、これをインパルス波形のピッチ列で駆動して擬似音声波形を合成し、これを偏自己相関分析して音響管形状にもどし合成時と推定時の形状を比較してみた。何と両者は、ほとんど完全に一致するではないか。この時の驚きと感動は、生涯忘れられないものである。

この実験により、私は声道形推定の問題は音声波形から声道インパルス応答波形を分離する問題に帰着されると確信したのである。

そこで、声道の伝達特性の特徴を実験的に調べた結果、声道を長さの等しい不均一断面の管の接続で近似したとき、隣り合うセクションの面積が数倍程度で比較的滑らかな母音型声道の周波数伝達特性の概形は、ほとんど平坦で傾きは存在しないと考えて良いことが明らかになった。これは実音声の母音のパワ・スペクトル概形の傾きや、概形の途中の曲がりは声帯波と放射の特性のもので

あることを示している。この「放射特性を含めた声帯波特性」は、①低次の共振系であること、②声道のようなQの高い共振は存在しない、③その特性は、人によって異なり、また時間的にも音韻の違いによっても刻々変化する、等の特性から成ると考えた。この条件を満足させるため、未知パラメータを持つ、拘束付きの2次及び3次系の自己回帰モデルを組合せたモデルを創造した。

実音声波によってこの声帯波モデルの未知パラメータを推定し、その逆フィルタリングによって声道のインパルス応答を分離し、得られたPARCOR係数が唇側からの反射係数に対応するものとして、声道形に変換すると、夢に見たX線写真やファントの論文で見た声道形が現れた。

これまで低次のホルマント周波数とそのバンド幅から声道を推定した例はあるが、実音声の分析で直接精度良く声道形を求めた例は無かった。

この結果を昭和48年2月東北大学で開催された日本音響学会・音声研究会で発表して注目された。翌年、ストックホルム音声セミナーで発表[3]し、SCRLの脇田氏によるものを含め日本からの二件の論文が話題となり声道形推定に関する特別セッションが急遽開かれ活発な意見が交わされた。議論は、ホルマントと声道形の関係が不定であることに関連して逆問題の解が単一に定まるかに集まった。筆者の見解は先に実験で示したように、声道のインパルス応答と声道形は一对一の関係があるので、解のユニーク性の問題ではなく声道インパルス応答分離の精度の問題と考えている。ちなみに脇田の方法では声帯波と放射特性を固定特性で与えており、その問題点は把握済みであった。

2.2 声道形をベースとするその後の研究

(1)声道形の局所的特長による母音記述

母音の特徴はパワー・スペクトル、さらにはホルマントで表されるとというのが常識であったが、連続音声ではホルマントの安定かつ連続性の保障はないなどの課題を感じていた。

私は、声道形の推定に成功したこともあり調音領域のパターンの特徴に期待し、母音と幾つかの子音を不均一音響管で表現して長さ方向の局所的特徴の組合せを手動学習した閾値論理の実験を行った。具体的には、母音を前舌母音と後舌母音に分け、調音位置、調音の強さ、声門位置、唇の開き、に対応する調音特徴を調音次元の特徴としてこれらを推定された声道形から得た。

9名の話者による各4首の俳句中の(無声化を除く)すべての母音を使い、閾値論理網を形成した。最初、話者は男性1名とし、このデータにより基本的な構造を形成する。以後、話者を1名ず

つ増加し、識別誤り、及びrejectが生じた場合には、さらに新しい調音特徴の発見と、より詳細な記述により開催論理網の規模を拡張する。話者を9名まで拡げて、各音韻にそれぞれ複数のカテゴリーを設定した識別機構を作成した。これによって、9名の同じ資料を識別した結果は、各母音の「わたり」を除く母音の中央の連続した数フレームについて98~100%の識別率が得られた。

(2)調音モデルと調音パラメータ抽出

声道形の推定は私のオリジナルの研究のため、その正当性(精度と信頼性)を示すことにエネルギーを費やすことになった。

一つの方法としてこれを調音器官の横断面に写像してその動きをX線による動きを重ねて観察することをした。さらに調音モデル化による調音パラメータを抽出も試みた。

調音モデルは、下顎と舌を一体として1円筒で表す1シリンダーモデルにより日本語五母音で観測されていた舌の動きと矛盾しないことを確認し、また、舌全体を前舌部と後舌部に分けた2シリンダーモデルにより、子音/d/と/g/から母音へのわたりの前舌の動きも矛盾なく得られることを示すことが出来た。このような方法で間接的ながら声道形推定の精度を示した。しかし、これらのパラメータ推定は非線形問題となり実用的とは言えない。

そこで、筆者は、X線映画で観測した顎と舌の動きを分析し、これに基づいて声道断面積関数を顎、舌に対応する関数の線形結合で表現する線形調音モデルを考案した。このモデルは1シリンダーモデルより精度が高く、線形代数方程式を解くことで顎、舌、唇の開きの3種の調音パラメータの抽出が可能となった。

(3)音韻フィルタ群による音韻抽出

音韻フィルタ群による音韻抽出の考えは、連続音声について区分化を直接行わず、各音韻を並列的处理によって抽出するものである。すなわち、各フィルタは、それぞれ受け持ちの音韻を抽出するように形成される。各音韻フィルタは、パワー・スペクトル傾斜の概形などのパラメータから定める「候補検出」の部分と、候補として検出された部分について、適応形音声分析システムで得られるすべてのパラメータを必要に応じて照合し、前後の音韻、継続時間、調音パラメータの動きを詳しく検討して一つの音韻とわたりとを決定する「判定部」から成る。

音韻フィルタ群形成のための資料として、前述の4首の俳句を使い、話者を1名から10名まで順次増加し、記述の不備を補い各音韻フィルタを形成した。

このように形成した音韻フィルタの出力は、縦軸の音韻 /i/, /e/, /a/, /o/, /u/, /ya/, /yu/, /yo/, /wa/, /wo/, /s/, /h/, /p/, /t/, /k/, /m/, /n/, /g/, /b/, /d/, /g/, /r/ として、各音韻の出現が、横軸の時間方向に長さの異なる矩形で示される。

実験の結果、当時全ての音韻が区別出来たわけではないが、これらの音の特徴をあらゆる手段で見付け出して検証しようとした枠組みを与えたものと言える。すなわち、一つの分析法から全音韻を抽出できるとするのではなく、各音韻について個別に、音声分析法まで遡って最適な抽出法は何かと考えることであった。

3. 子音の特徴を求めて

声道形推定の余勢をかい、短区間分析、定常雑音除去、さらには音韻フィルタ群を組み込んだ適応型音声分析装置を作成した。しかし音韻フィルタ群は論理規模の拡大をせざるを得ず一人で出来る範囲を超えるものであった。

この主要原因に子音の本質的特長が明確でないことがあった。これまでの音声分析は周波数特性の極を抽出するものであり、これまでの分析法では重要な周波数特性を見落としているのではないかという懸念が生じた。

この問題は、実音声进行分析する方法だけでは解決しない。音源位置と鼻腔を持つ声道モデルの伝達特性を計算できる合成手段により、音響的現象を声道レベルのパラメータで記述し、その特徴をノモグラムとして体系的にまとめることにした。

3.1 声道モデルのノモグラム[4]

周波数特性を有する反射係数と伝搬定数及び音源位置を声道パラメータとした声道伝達特性計算のデジタル計算機向きの巡回計算法を導き、その具体的な手順を示した。これにより、音響的及び調音的条件が声道伝達特性に及ぼす効果を合成実験的に見積ることができる。

本ノモグラムの教えるところ、多くの子音の物理次元での本質的特長は、スペクトルの零にあると言わざるを得ない。たとえば音源が声道の途中にある音の場合でも声道伝達特性の極の数と位置には全く変化はなく、私が本質的特長と考える音源位置に関する情報は、先の極特性を保存した状態で、音源から奥の部分の声道形の伝達特性が零として加わることが明らかになった。

このノモグラムにより、従来の全極形の音声分析法の限界を痛感し、子音の本質的特長の分析に精密な極零分析の必要性を実感した。そこで、国研での音声研究が可能な最後の機会を以下概説す

る精密なPSE（パワ・スペクトル包絡）分析を新たに開発して子音の本質的特長の発見に賭けることにした。

3.2 ピッチ対同期形PSE分析[5]

子音の本質的特長には極零分析が不可欠との見解に至った。ピッチ対同期形PSE分析法はこの考えを厳密に実現した音声分析の一つである。すなわち、有声音では、ピッチ同期で切り出したピッチ対（2ピッチ分波形）に高分解能フーリエ変換を施して高精度パワ・スペクトルを得る。このパワ・スペクトルをピッチ周波数間隔でサンプリングし、PSEを推定する。なお、このPSE特性は、直線周波数軸の対数パワ・スペクトルをピッチ周波数間隔で標本化し、これに余弦級数をあてはめたものである。

この分析法により、これまで無視されてきた零が明瞭に抽出できることは確認したが、先のノモグラムから予測される子音の特徴が得られるが完璧といえる段階ではなく時間切れとなったのは残念である。

しかし、副次的に面白い現象を見出すことになった。これはピッチ構造を持つ音声波形の分析にのみ現れる現象と考えられるので、その法則性を以下の非定常波形のスペクトル・モデルとして纏めることにした。

3.3 非定常音声波形のスペクトル・モデルと零位相インパルス応重畳形音声波形合成法[6]

ピッチ同期形PSE分析法を実音声に適用し、高精度（高サンプル数）フーリエ変換で得た調波構造を持つパワ・スペクトルをピッチ周波数間隔で標本化してみると、そのサンプル点がパワ・スペクトルの調波構造のピークと一致せず、ピークの左右にずれる周波数帯域があることを見つけた。この現象は、従来の音声分析では、1フレームの切り出し波形に256点程度のFFTを適用するのに対し、本方法では、常識外れの4096点のFFTを適用して周波数分解能を上げることによって発見されたのである。

これは2ピッチ分の波形の前後の波形の形状に違いがある場合に起こる現象である。

そこで、ピッチ構造を持つ非定常態の波形についてフーリエ変換による調波構造を持つパワ・スペクトルとそのスペクトル包絡との関係をピッチ周波数と関連して調べることにした。このため明確に異なる2つの非対象形三角波のピッチ時間差を持つ対にPSE分析を適用して検討した。その結果をまとめたのが非定常態音声波形のスペクトル・モデルである。

このモデルにおいて新たにFMスペクトルなるものを提案した。FM（周波数変調）スペクトルとは、横軸を周波数とし、その軸上の等間隔標準化周波数において、調波構造のピークが各標準化周波数からずれる（周波数）値を縦軸に表したものである。無論、定常態音声波形の分析では、FMスペクトルは生じない。

ちなみに明確に異なる2つの非対象形三角波のピッチ時間差を持つ対のPSEと、この2つの三角波の前後を入れ替えた場合のPSEは、当然全くおなじであるが、FMスペクトルの極性は+と-とが逆転するのである。この現象は、たった1フレームの定常態分析の結果に非定常態情報が織込まれていることを示すものであり、音声ピッチ構造を持つことによって生じる現象と考えると、極めて興味深い。

聴覚が定常態分析をしている証拠はないが、音声は、幸いピッチ波形のため、仮に定常態分析でもスペクトル変化の方向を示す情報が捉えられるのである。まして聴覚がより優れた非定常態分析をするならさらに正確に子音情報を識別できると考えられる。

一方、非定常音声波形のパワ・スペクトルモデルを得たことにより以前よりずっと抱いていた音声合成に係わる問題点、すなわち、「音声合成における原因不明の異音の発生」の原因解明へのヒントが頭をよぎったのである。

従来の音声分析合成系では、パラメータの量子化を行わない場合でも異音の発生など音質の劣化を経験する。特に高ピッチの女性の美声の合成は困難であった。これまでの代表的な音声合成では、フレーム間隔と呼ぶ等間隔時点にパワ・スペクトル包絡を表すパラメータを与えてフィルタを構成し、これに変化するピッチ波形を音源として入力して音声波形を得ている。通常音声では、パワ・スペクトル包絡パラメータと音源ピッチの両方が同時に変化しているわけである。上記モデルの教えるところ、この場合FMスペクトルが発生し、しかも合成音声では、それが不自然な形になっているのではないかと考えた。

ちなみに、パワ・スペクトル包絡パラメータと音源ピッチのうち、片方だけ変化させた音声合成では異音は全く発生しないことを確認した。

こうして生まれたのが、零位相インパルス応答の重ね合わせによる音声合成法である。この方法ではパワ・スペクトル包絡特性（PSE）を対称形インパルス応答波形にするためにフーリエ変換を利用した。零位相波形をピッチ周期で重ねあわせることにすれば、時間基準位置が波形の中央で

明確なために正確なピッチ周期が保障される。

この方法によれば、従来の任意の位相特性を音源インパルス列で駆動したときに発生する出力波形の位相のバラツキが抑えられ、時間領域で対象波の対のためFMスペクトルも生じない筈と考えた。

この目論み通り、この方法により極めて高い女性の声でも異音を発生させることがなく安心して音声合成ができるようになったのである。

この合成音の客観的で定量的な評価は今後の課題であるが、合成音は原音と対で示され、主観的ながら「この合成音には、従来の方法でしばしば発生していた耳障りな異音がなく、澄んでいて、しかも自然音声のもつボリューム感があり良い声である」という評価を得た。

この方法は、その後、異音を生じさせないで超高ピッチまでを安心して合成する方法として実用化されているという情報も頂いた。本合成法の開発は、過去の音声合成で抱いていた異音の発生の課題をずっと心に留めていたところに、FMスペクトルの存在を見つけたことが契機となったと考えている。

また、本PSE分析・合成系は子音の本質的特徴を抽出する他、人間の音声知覚のしくみの解明の道具となれば幸いと思っている。

4. 二十一世紀の音声研究への提言

4.1 音声技術の最終目標は社会インフラ

2002年度米国大統領予算教書の補助資料[7]では、連邦政府諸機関が実施する現在の情報技術研究開発活動と解決すべき今後の課題に焦点を当てている。ここでは**情報技術の提供する社会基盤（インフラストラクチャー）**なしには国家の経済競争力や重要目的の達成はありえないとしている。

音声技術関連の項目を引用（同文献,pp.21-22）すると、チャレンジ#7「人間の能力と万人の進歩に対する支援」にあつて、情報技術は、「すべての人が自分の個人的能力や技能を高めるための手助けができる潜在的可能性をもつ」と位置付け、

第1に、人とコンピュータとの相互作用の研究は、たとえばコンピューティング・デバイスが人間の記憶・注意持続時間、感覚認識（視覚、聴覚、触覚等）および理解力等を増大させることにより、人間の能力に寄与しうる範囲を拡大することが目的である。

エンドユーザーに重点を置いた研究は、ユーザーの能力や操作性を拡張するという観点から、インターフェース、検索エンジン、通信技術等のITコンポーネントを再創出し、翻訳や、音声で操作することが目標である。

例えば、機械コンピュータ対話、パターン認識、自動筆記等の言語技術の研究は、人間とコンピュータの間の手を使用しない通信、外国語の資料の機械による翻訳、文章のコンピュータによる分析、会議における討論の音声認識や、話された言葉の筆記等を目指している。

これらの能力は、軍事や国家安全保障の場で極めて重要な用途を持つのみでなく、他の領域でも極めて広範囲にわたる用途をもっている。この研究は、年齢あるいは身体的能力の如何を問わず、すべての個人が豊かで自立した生活を送ることができるように補助する技術、また、そのためのツールとデバイスの開発を支援する。

長期的研究のニーズとしては、「言語間ならびに話された言葉と書かれた言葉の間の翻訳、音声言語照会システムを含む言語工学技術」のほか、

チャレンジ# 8 知識世界の管理と実現の項目 (同文献,pp.23-24) には、

アーカイブ構築に関連して「現在の検索ツールは、内容の記述からは音声情報または映像情報を探し出すことは出来ない」という記述があるように、音声認識、自然言語、画像認識を要素とするプロジェクト例として、デジタルライブラリ計画がある。人間を主体とする自然で知的な関係が必要とされている。

大統領のレベルの国家戦略として、これら情報技術は人間の能力を拡張する潜在技術として大きな期待が賭けられていることは間違いなく、また、**社会インフラ**として位置付けられていることを認識すべきであろう。

4.2 音声研究の進め方

音声研究は魅力あるテーマであり、多くの成功と不成功の歴史をもつがそれらの経験が必ずしも統合された人類の財産となっていない。音声研究の成果への評価も必ずしも定まっていない。音声合成は成功したが、音声認識の利用は限定的である。30年前に素朴に期待された音声認識のレベルは実現されたともいえるが、そのレベルは出来てみれば努力の割に成果が今ひとつで、何か虚しい。

音声認識は、究極のものが出来ない限り、途中成果は受け入れられないという難しさがあると思われる。途中成果が受け入れられるためには、限定された応用の世界を創出することが必要とされ、事実この方向で多くの努力がなされた。

過去の研究を否定する必要はなく、その埋もれた中からも有効なものを発掘する態度も必要であろう。研究活動が、オリジナルを競うあまり、他人の研究結果を十分に吟味しないまま、前の研究

の隙間を埋めるだけの研究に陥っていたことも懸念される。メーカーで商品化プロセスに触れた者として、基礎研究でも、研究を正しく評価してより高いものにしていく意識改革と仕組みの構築が必要に思われる。

音声認識は、目標自体「神の領分」に踏み込むものとして真剣に議論された時期もあり、また、脳の研究の重要性を強調する考えもある。しかし脳の情報処理に全てを帰着させるのは無責任であり、いま「**人類の共有財産として積み上げをすべき**」目標を考えたい。

4.3 何を為すべきかの私案

それは、ユニバーサルな**音韻 (音素) 抽出機能**の実現ではないか。本来、音韻論はユニバーサルの考えによるものであり、これを押し進めると、音韻を抽出する機能を持つシステムは、**世界に一つ完成すれば良いはず**である。連続音声中で全ての音素が正しく発声されるわけではないが、変形、消失、付加にも規則性は存在する。あくまでも発声された範囲での音素抽出が出来ればよい。

日本語の音韻は、全体の一部であり、このような回り道は必要ないと言われそうであるが、連続音声中で意識していない音は、別の言語のそれに対応していることも容易に想像される。これは筆者が、実データによる音韻体系化の試みから学んだことである。不特定話者の汎用音声認識にはこのシステムが必要である。

誤解を恐れず言えば、万国語共通の音韻認識システムは、これを人類の財産として「**(概) 国際音声記号抽出システム**」原器 (デジュール標準) があればよい。

過去、この部分は、従来各国・個々の研究者が独自性を競い合って多大の費用と労力をつぎ込んできたが相互利用をするという考えは無かった。外国における、外国語の音韻に関する結果は、日本語では通用しない、などと漠然と考えている。しかし世の中は、長い歴史をもつ音声研究の成果の利用に期待している。

インフラとは、電気やガスのように基本的なコンポーネントが利用可能な形で提供される事であろう。こう考えると、21世紀にこれまでの方法論が通用するであろうか。共通の社会インフラの部分と、各国語に対応して成果を競争する部分を分ける必要があろう。

実現方法の詳細を提起する知恵は持たないがリナックス OS の成功は参考にすべきである。また応用システムのための応用インターフェース (API) を定めた音声分析法を含めた音韻別のコンペ方式もあり得る。この場合、音声データペー

スとテストコレクションを用意することが重要である。

今後は、国際レベルで、社会インフラとして何を提供できるかを考え、焦点を絞った目標を立てて推進することが望まれる。この場合、いたずらに結果を求めるのではなく、サイエンスの立場で自由に評価して知恵と努力が蓄積できる仕組みが考えられるべきであろう。

かつての国家プロジェクトでは、電子計算機が高価であったこともあり政治家をはじめ専門家以外の誰にも分かるキャッチフレーズで目標を掲げて大きな予算を確保し、研究成果はデモンストレーションで示すことが行われた。研究者は、宣伝の大きな目標に不安を感じながらも当面の研究費の確保を優先したような気がする。

いま、研究環境は当時とは全く異なり、安価に良い環境が構成できる。同じ志を持つ有志が国内外と結束して知恵を出し合い、成果をうまく結集することができれば、音声認識も新しい世界を開くことができると確信する者である。

5. あとがき

本論文を書くよりかなり前の時期に論文タイトルの提出を求められた。今回の企画タイトルが、失敗からの云々、ということであったので、タイトルに私の反省という副題をつけた。しかし、いざ論文を書く段になると、懺悔を書かなければいけないのかな、という暗い気持ちに陥った。正直そのような気持ちもないので、本文の前半は、私の音声研究を振り返り、各研究テーマが変遷していった経緯、あるいは何故そのテーマを考えついたのか、また、各テーマ間の関連や研究中の驚きや感動なども素直に書かせて頂くことにした。

この中から、自らの研究結果に対する内省に基づいてテーマの変遷があったこと、またその副産物として過去から抱いていた例えば音声合成に関する課題を解決するヒントが得られることがあること、などを読み取って頂ければ幸いである。

また、企画の目的に、「技術者・研究者に一步前に入るための勇気を、若い人に伝えたい、云々」ということであったので、後半は、まず、米大統領の一般教書の補足資料の内容を紹介した。現在、わが国では、残念ながらこれに並ぶものが見当たらない。この資料では、情報技術を社会インフラと捉えており共感を覚える。音声研究者はこのような具体的な高い目標をもつべきであろう。

音声研究の成果を社会インフラとして人類共通の財産に位置付けること、そして、選択と集中の

時代にあって音声技術の万国共通な具体的なシステム作りあげを提案した次第である。ここには、本文 2.2(3)節で紹介した音韻フィルタ群のイメージがあり、もし世界の音声研究者が、各自抽出を担当する音韻をきめて協力し合えば道が拓かれるのでは、等と無責任ながら夢を書かせて頂いた。このような視点でたまたま拝読した産総研(旧電総研)音声聴覚情報研究グループの2001年度年報の中に、実に「言語に共通な音声符号系(IPAに基づく)を用いた音声検索」[8]があり問題意識を共有できる。

過去に音声研究に熱中し、現在でも関心を持つ者として、若い研究者が過去に捉われない大胆な発想で音声研究を進め、大いに発展することを願うものである。

文献

- [1] 中島, "[フェロー記念講演] 21世紀の音声研究への期待と提案," 信学技報, SP2002-46, pp.25--30, Jun. 2002.
- [2] 中島, 他, "デコンボリューションによる声道形の推定と適応型音声分析システム," 音学誌, vol.34, no.3, pp.157-166, Mar.1978
- [3] T. Nakajima, et al., "Estimation of Vocal Tract Area Functions by Adaptive Inverse Filtering Methods and Identification of Articulatory Model," Proc. Speech Communication Seminar Stockholm, Aug.1974
- [4] 大村, 中島, "調音と音声パワースペクトラムのノモグラムの検討," 音学会誌, vol.46, no. 6 pp.468478, June1990
- [5] 中島, 鈴木, "パワースペクトル包絡(PSE)音声分析・合成系," 音学会誌, vol.44, no.11, pp.824-832, Nov.1988
- [6] 中島, 鈴木, "非定常波形のスペクトル・モデルに基づくピッチ対同期形PSE分析法," 音学会誌, vol.44, no.12, pp.900-908, Dec.1988
- [7] 国家科学技術会議, "ネットワーク及び情報技術研究開発," 2002年度大統領予算教書補足資料, 情報技術研究開発に関する省庁間作業部会報告書(AITEC訳), pp.21-24, Jul.2001
- [8] 田中, 他, "汎用音声符号系への符号化と音声処理システムの構築," 音学会, 研究発表会資料, no.2-5-14, Mar.2002