立石賞特別賞の受賞記念講演概要

デジタル時代の音声符号化・合成・認識に関する音声分析根幹技術の発明

名古屋大学 名誉教授 板　倉　文　忠

1. 音声情報処理

1.1 歴史：電話と蓄音機の発明

電話は 1876 年にグラハムベルが音声を電気信号に変換しその信号を相手に伝えて再び音声に戻す，所謂，電気音響変換器として発明されました。一方，蓄音機は 1877 年にトーマス・エジソンによって発明され，音の空気振動を蝋管の溝の凹凸変化に刻み込んで，それを再び元の波形に戻すという，音をアナログ的に記録・再生する装置で，エジソンは，この蓄音機の発明が人生で最も興奮した時であったと史記に書いています。

1.2 人の発声のモデル

さて，ここで発声と言うことを少し考えてみます。人間は脳で考えたことを口で発声する訳ですが，これをモデル化すると図１の様なことになっています。私が主として研究してまいりましたのは，その発声の部分でありまして，脳がどういうふうに動いているか，その構造がどうなっているかということについて，当時はまだ直接観測したり，データを取ったりすることが難しい時代でした。しかし，発声については図１のように，肺から送られてきた空気流が声帯という器官で空気振動になり，それが口腔を介して口から音声として放射され相手に伝わる訳です。声帯は音の基本的振動を作るところで音源と呼ばれています。それが声道とよばれる舌や口腔，唇の動きによって，言葉特有の変換を受け，音声となって放射される訳です。これが音声発生のメカニズムでして，脳に比べると非常に単純なシステムでございます。従いまして，この部分は，かなり数学的な扱いもできるのではないかと考えたわけであります。

実は，このような仕組みを具現化した先行研究に Voder というものがあります。これは，電気回路によって声帯の振動のような波形を電気的に作り，それをボーカルトラクト（声道と呼ばれる音響管）に相当するフィルターにかけますと人間の声に近いものが作れます。そのボーカルトラクトの特

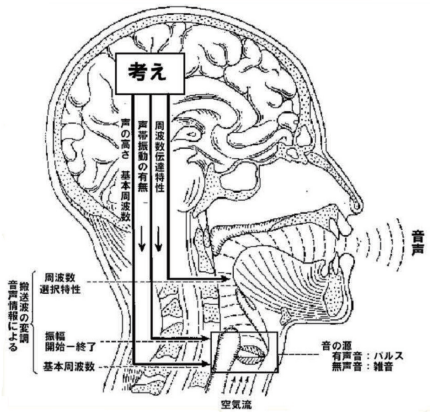

図１　人の発声のモデル

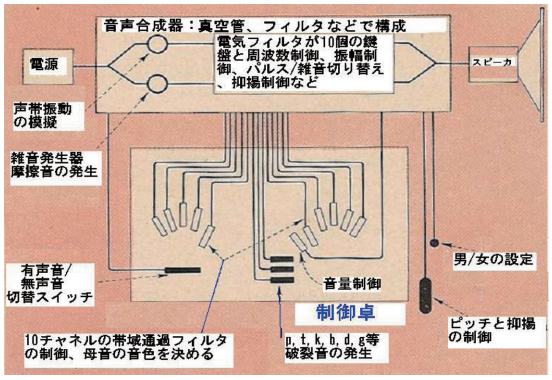

図２　音声合成電子オルガン Voder の原理

性をオルガンの鍵ように 10 本の指で周波数をコントロールすることによって人間の声を合成できるはずだということをベル研究所の Homer Dudley が 1939 年（今から 80 年前）にメモを残しております。

1. 3　音声帯域圧縮方式 Vocoder の発明

この原理を実用的に近づけた一つの研究が音声帯域圧縮方式 Vocoder というものです。これは，1928 年太平洋横断電信ケーブルが敷設された時，その電信ケーブルを介して音声を直接送ることは出来ないかという議論がされました。ただ，当時の電信ケーブルで送れる最大周波数は 100 Hz 程度で，音声波形をそのまま送ろうとすると 3000 Hz 程度の周波数帯域幅が必要でした。そこで音声信号の帯域幅を 100 Hz くらいに圧縮しようと考えられた Vocoder というものがあります。これは，先程の Dudley が 1928 年に提案していました。

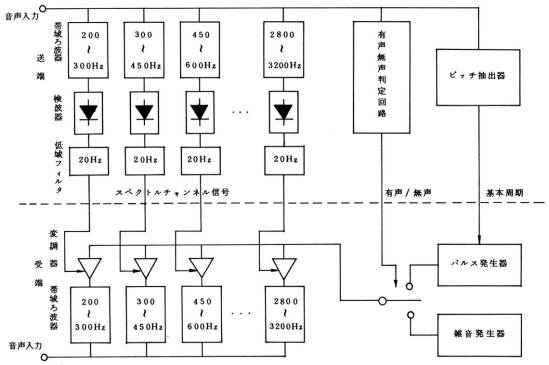

図 3　Dudley. H のチャネルボコーダ

図 3 がボコーダの原理図です。簡単に言えば，人間の声を周波数分析する濾波器とそれを合成する濾波器を送信側と受信側に置き，それぞれの周波数成分の強さを整流して取り出し，それを濾波器へ振幅情報として入力し音声を合成するというものです。要するに，アナログ的なフィルター技術を使った試みでありますが，実際の電信ケーブルで音声を伝送することには使われなかったようです。

2.　音声分析合成系：

このように人間の声を分析して，それをパラメータ（数値）として取り出し低ビットレートで符号化して伝送し，再び元の音声を合成する音声分析合成系の技術が必要になった訳です。この音声合成分析系が上手く機能すれば，人間の声の重要な部分をきちんと分析できたという一つの証拠になるわけです。その意味で，私は音声処理の基本技術と考え，それをライフワークとして研究してきました。その後，その中核となる PARCOR，LSP という数学理論を考え付いて，今日ではそれが，色々な所に利用されるようになってきたという次第でございます。

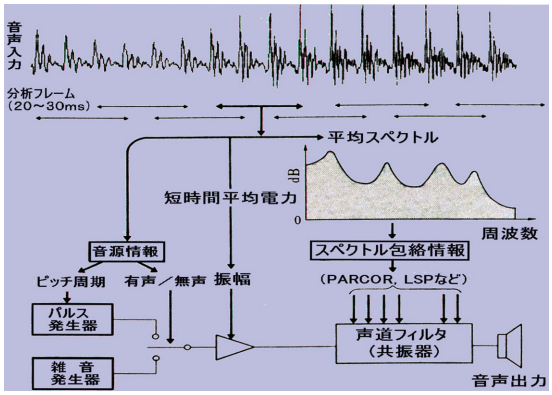

図 4　低ビットレート音声符号化の原理

音声合成分析系を図 4 で説明します。最上段が音声のアナログ波形で，これを 20～30 ms の区間（分析フレーム）に細かく分割し，各区間の信号の強さや声の高さ等，平均的な周波数成分を抽出し，これを以て，先程の Vocoder と同じように，音声を合成・復元するというものです。これは，音声波形をそのまま伝送する代りに，人間の声の平均的なスペクトルと音源の情報を伝送するという意味で音声波形そのものを伝送する従来の電話とは全く異なる仕組みになっている訳です。

2. 1　はじめの挑戦と失敗

この研究を進めるにあたって，まず自分の声をサウンドスペクトログラムで分析した所，私の声が少ししわがれていて教科書に載っているようなきれいな声紋パターンが得られませんでした。そこで私は，人間の声と言うものは非常

に複雑な変動を持つ信号ではないかと考え，まず，それを確率過程と見做してモデル化した方が良いのではないかと考えました。ということでその数学モデルに基づいて確率過程の統計的なパラメータと称する母数を抽出し，それによって認識すればもっと良い音声認識システムができるのではないかと考え研究をスタートしました。

当初私は，それをアナログ的なフィルターで実現しようとしたのですが，上手くいかないことが分かり，当時研究室に導入されたミニコンFACOM270-20 を使ってデジタル信号処理を意識した数理分析からスタートしました。その結果，音声の最適識別に必要な統計量は，初めのｐ個の自己相関関数で抽出できることを明らかにし，これを使って音声の分析をスタートさせたわけであります。その研究をまとめた成果は，電信電話公社電気通信研究所の成果報告として 1966 年に出版されました。(図 5)

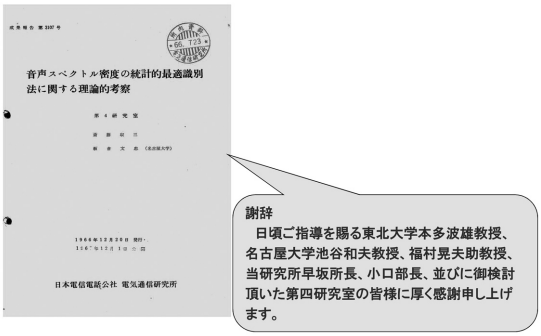

図 5　理論のまとめ

2.2　音声分析合成方式に挑戦

その頃，研究室の指導者であった齋藤収三先生から，「実は音源の周波数を抽出するためにピッチ抽出という問題が一番難しい，それを何とかしないといけない」と言われました。そこで，このピッチ抽出という問題に取り掛かり，変形相関法という新しい理論を編み出し，加えて先程の理論と統合することによって最尤スペクトル推定法による音声分析という方式を 1967 年に提案し，実験的にも確認しました。その合成音は予想以上に自然で明瞭性も高いことから，その結果を翌年，東京で開かれた第 6 回 ICA（国際音響学会議）で発表する機会に恵まれました。偶々同じセッションでベル研究所の Atal, Schroeder から音声の適応線形予測符号化に関するが発表があり，私の研究と共通する点が多いということで線形予測音声符号化の先駆的な研究として認められ大変光栄でした。その原理図が図 6，図 7 です。

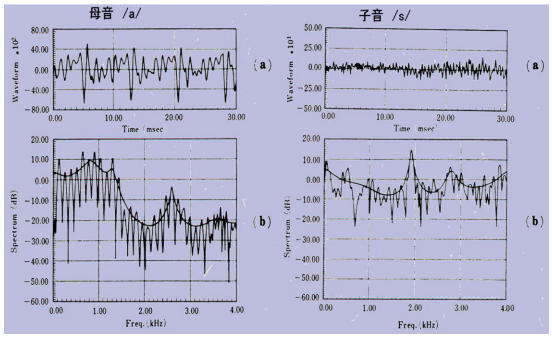

図 6　ML スペクトル推定の例

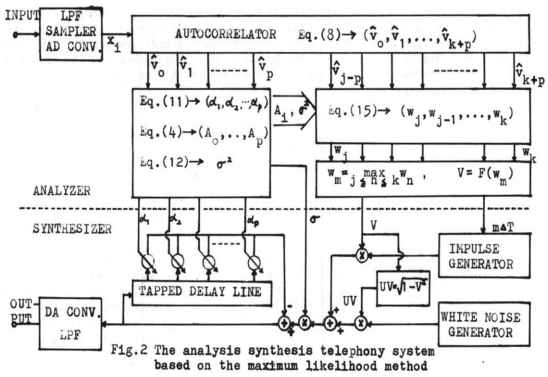

図 7　ML スペクトル推定法のブロック図

2.3　最尤推定法から PARCOR 方式への発展

こうした研究をさらに発展させ PARCOR 型音声分析合成方式を編み出し（図 8，図 9），1969 年第 7 回 ICA にて発表しました。当日は丁度アポロ 11 号の月面着陸の実況中継をやっていたため，聴講者が少なく殆ど反響がなかったのは残念でした。

その後，1970 年に音声合成器を実際にハードウエアで試作することになり，まずは音声を合成する部分を試作したのですが，図 10 のように非常に大きな装置になりました。しかし，こうした努力のお蔭で，NTT で実用化しよう

PARCOR係数の絶対値が全て1未満であれば，分析フィルタの逆フィルタである合成用格子型フィルタは安定である事を理論的に示す事ができた．

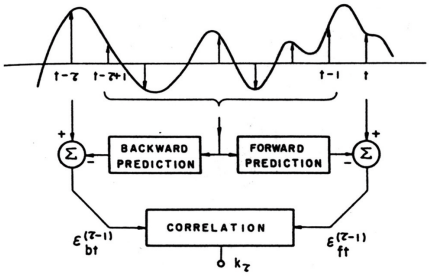

Fig. 1. Definition of the partial autocorrelation coefficients.

図 8　PARCOR 係数の定義

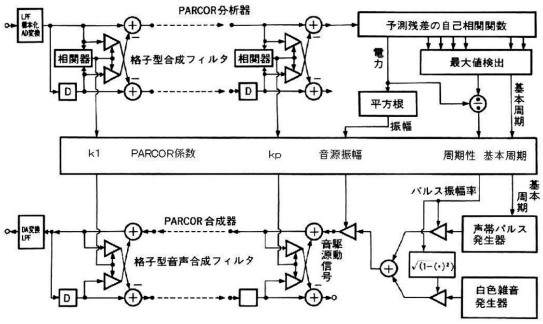

図 9　格子形ディジタルフィルタ PARCOR 音声分析合成系

との機運が高まり，ANSER という音声自動応答装置が開発されました。

そうこうする内に，1976 年に TI が Speak & Spell という音声合成を使った商品（おもちゃ）を発表した訳ですが，乾電池で動く非常に小さな装置で音声合成ができるようになりました。

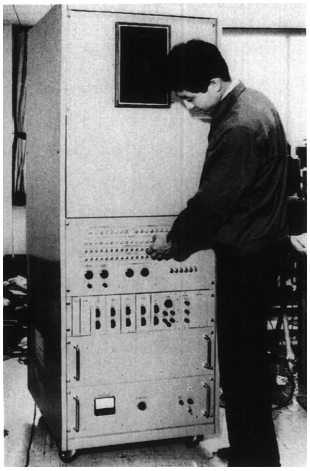

図 10　PARCOR 音声合成器（1970 年試作，NTT 提供）

2.4　最尤スペクトル推定法の音声認識への応用

その後，音声認識にも線形予測符号化（LPC）という方式が使われ始め，図 11 は，私

ベル研での滞在の2年目には，Rosenberg氏と協力して，音声認識，話者認識，音声応答よりなる3-modeシステムを作成し，航空座席予約システムとして実験公開し，好評を博した．これらの研究開発が端緒になり，しばらく停滞していたベル研究所の音声認識の研究が息を吹き返し盛んになった

（残念ながら,それも2002年までのことのようである.：最近Lucent Bell研は音声処理グループをほとんど解散した．またATT Labs.においても音声グループは著しく縮小した。）

図 11　3-mode（音声認識，話者認識，音声応答）システム

が 33 歳から 35 歳頃にベル研究所で行ったデモンストレーションの写真です。当時ベル研究所では，音声認識研究がストップしていましたが，この研究がきっかけになって，再び活発な研究が行われるようになりました。

2.5　LSP 方式の誕生

LSP 方式も私がベル研究所にいた時に，そのきっかけを見つけ，それが現在では，世界の携帯電話の音声分析部として広く使われています。LSP は従来のものに比べて，① パラメータ量子化誤差の影響が少ない　② パラメータを補間した時のスペクトル再現精度が高いという長所を持っていまして，他のパラメータに比べて極めて優れているということで現在では世界中で使われています。その貢献に対して，2017 年 8 月 ISCA（International Speech Communication Association）から ISCA メダルが授与されました。

図 12 が LSP 音声合成フィルタの回路図に相

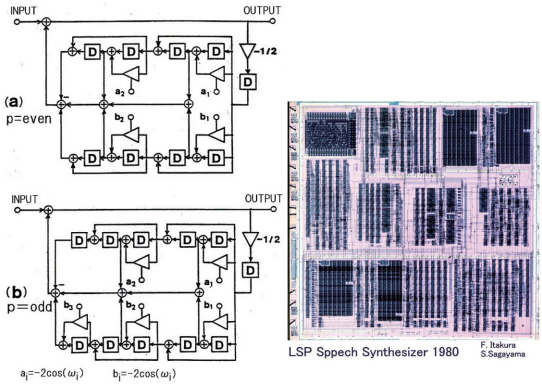

図 12　LSP 音声合成フィルタの構造

当するものです。これをLSIチップにしたのが横の写真です。このチップの設計にあたっては，（現）明治大学の嵯峨山茂樹先生の大きな貢献をいただきました。

3. 総括：単純な最適化原理の活用
― 阿呆の一つ覚え ―

私は，こうした研究をやってきましたが，基本的には，出来るだけ単純なアルゴリズムで音声を分析しよう，簡単に言いますとガウスが考えた最小二乗法をベースにいろいろな問題を解決してきました。

このように，音声の研究と言いましても，音声そのものの勉強だけでなく，関連する数学の勉強をし，そこで古くから使われていた考えを音声と言う具体的な研究に応用して研究を進めてきた訳であります。そういう意味で，研究をする時には，あまり他の人がやっていることにとらわれることなく，自分でその現象を最も本質的に表現しているものは何であるかということを考えて進めることが重要ではないかと考えております。

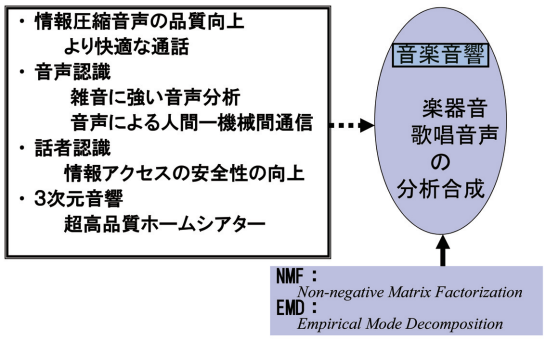

図13　現在の研究と今後の展望

4. むすび

音声情報処理の研究は，発声生理，知覚，音声学などの基礎的研究と関連しながら，最近のマイクロエレクトロニクスやソフトウエア技術をベースに，過去50年の間に長足の進歩を遂げてまいりました。音声情報処理の目標は，立石財団の趣意でもありますように，人類の夢である人間とシステムの自然な対話（人間相互間と同様な）を実現することであろうと考えています。ここで紹介した音声分析合成技術は，「千里の道の一里塚」にすぎませんが，数理的な基礎が強固であることから安心して応用でき，実用性の高いものであると考えています。

今後は，音声信号を単に言語情報の伝達メディアとしてだけでなく，音楽など感性メディアの研究にも展開していくことが望ましいと考えております。（図13）

謝辞・文献

ここに記したことは，筆者の過去45年の音声処理に関する研究開発経験の初期に行われたものです。この間，電電公社通研基礎研究部第4研究室に在任中の斎藤収三室長はじめ，橋本新一郎，橋本清，脇田寿，小池恒彦（故），山本啓，筧一彦，好田正紀，佐藤大和，古井貞熙，鹿野清宏，北脇信彦，中津良平，村上憲也，東倉洋一（故），嵯峨山茂樹，小林勉，箱田和雄，河原英紀，誉田雅彰，匂坂芳典，長渕裕実，管村昇，林伸二，相川清明，守谷健弘，伊藤憲三，杉山雅英氏など，厳しい指導者・先輩のご指導と優秀な同僚のご協力により達成できたものであり，ここに厚く御礼申し上げる次第です。最後に，名城大学在任中，音声・音響研究室の中心となって活躍いただいた畏友坂野秀樹准教授，並びにこの度の立石賞受賞にあたり，お世話になった財団関係各位に併せて深甚なる感謝を申し上げます。

論文目録抄録

1） 板倉，福村，斎藤，“音声の最適識別法に関する一考察”，信学全大（1966.11）
2） 板倉，斎藤，“偏自己相関関数による音声分析合成系”，音講論集，（1969.10）
3） 板倉，斎藤，西川，小池，“PARCOR形音声応答装置”，音講論集，（1970.5）
4） 板倉，“線形予測係数の線スペクトル表現”，音声研資S75-34（1975）
5） 嵯峨山，板倉，“複合正弦波モデルによる音声分析合成系”，音声研資S79-6（1979）
6） 管村，板倉「線スペクトル対（LSP）音声分析合成方式による音声情報圧縮」信学論（A）Vol. J64-A, No. 8,（1981）
7） 板倉“スペクトル符号化にもとづく音声分析合成”，音響学会誌37, 5（1981）