

音声分析合成の基礎技術とその音声符号化への応用

板倉 文忠

名城大学 〒468-8502 名古屋市天白区塩釜口 1-501

E-mail: itakuraf@ccmfs.meijo-u.ac.jp

あらまし Dudley (1936 年) に端を発する音声分析合成系 (Vocoder 技術) は 1966 年に確率過程論をベースにした新しい形に近代化した。その代表例がここに述べる最尤 (ML) 方式, PARCOR および LSP 音声分析合成方式 (以下 ML 方式, PARCOR 方式, LSP 方式と略す) などである。その原理は, 音声信号から音声の周波数スペクトルのパラメタを統計的な手法により分析抽出し, それらを低ビットレートで符号化伝送し, 再びもとの音声を合成する技術であり, 現在, 音声帯域圧縮伝送, 音声合成, あるいは音声認識の前処理などの音声処理の一つの基礎技術になっている。これらの方式は, それぞれ, 1966 年, 1969 年, 1975 年に日本電信電話公社電気通信研究所 (通研) で発明された。この資料は, 音声分析合成系の発展の経緯を著者個人の経験を中心に纏めたものである。

キーワード 音声分析合成, Vocoder, Dudley, Sigsaly, ML, PARCOR, LSP, LPC

Fundamentals of Speech Analysis and Synthesis and its Applications to Speech Coding

Fumitada ITAKURA

Faculty of Science and Technology, Meijo University, Shiogamaguchi, Tempaku-ku, Nagoya, 468-8502
Japan

E-mail: itakuraf@ccmfs.meijo-u.ac.jp

Abstract ML, PARCOR and LSP speech analysis-synthesis methods are low bit rate speech coding methods which serve as a fundamental method for speech processing, such as speech band-width compression methods and feature extraction for speech synthesis/recognition. ML, PARCOR and LSP methods were invented in 1966, 1969 and 1975, respectively, at the Electrical Communication Laboratory of NTT. In this article, a brief history of the PARCOR and LSP will be described.

Keyword : Speech Analysis and Synthesis, Vocoder, Dudley, Sigsaly, ML, PARCOR, LSP, LPC

1. 音声分析合成系 Vocoder 技術の歴史

人は誰でも仲間と互いに意思を疎通させたいという欲望を持っている。意思の疎通にもっとも簡便な手段は音声言語によるものと言ってもよいだろう。ただし音声言語は発話者の近くにいる人にはすぐに伝わるが, 数十メートル離れると聞こえなくなってしまう。また時間が経てばすぐに消えてしまう。音声言語のこの欠点を補うために, 19 世紀の末に Graham Bell の電話機 (1876), Thomas Edison の蓄音機 (1877) が発明された。電話機が発明された当初は, その技術の重要性よりも, むしろ目の届かないほど遠くの人

の声が聞こえるという神秘性が人々の興味を惹つけたようだ。また, Edison も, 蓄音機は「多くの発明の中で最も感動した発明だった」と述懐している。この 2 大発明が相次いでなされたのは決して偶然ではなく, 1877 年には, Edison も電話の発明改良に注力しており, 音声による空気の振動を記録することができれば, その振動記録から声を再生し, あるいは電信技術により遠方に伝え声として再生できると確信し, ついに蓄音機 Speaking Phonograph を発明したと言われている。

20 世紀にはいると真空管, 電気音響技術, 高周波技術などの進歩とともに電話技術は, 急速に発達したが, その基本原理は Bell や Edison が注目

したアナログ的な音声波形をできるだけ忠実に伝送したり蓄積する方法であった。音声信号は、主として約 4000Hzまでの周波数成分を含んでいる。従って、明瞭な音声を送信するためには4000Hzの周波数帯域が必要になる。ところが 20世紀初頭ごろまでに敷設されていた電信用大陸間横断ケーブルの伝送可能な周波数帯域はせいぜい 100Hz程度に限られていた。電話の発明後約 50 年経った 1920 年代に、何とかして旧来の電信ではなく、なまの音声で大陸間を結びたいという要望が生まれた。

この要望に応えるべく 1928 年米国のベル電話研究所の Homer Dudley は、人間の発声のメカニズムと人の聴覚の特性をうまく利用して、音声を 4000Hz より狭い帯域で伝送する装置ボコーダ (Vocoder, Voice Coder) を発案した。Dudley の最初の着想は、次のようなものであった：すなわち、音声は人の発声器官により生成されたものであるから、受話器の代わりに発声器官を模擬する音声合成器を用意し、送話側では、人の音声を分析して声の特徴を抽出し、これを受信側に送ってやれば、音声合成器によって元の声に近い音声再生できるはずだ。

図 1 に示すように、人が話すとき、まず考えを頭に浮かべ、それを声にするために声の高さ、有声音・無声音の区別（声帯振動の有無）、音素による音色の特徴を、調音神経をとおして発声器官に伝える。声帯では有声音か無声音かによりパルスあるいは雑音を音源として発生する。この音源信号が喉頭・咽頭・舌・口蓋・唇・鼻腔・鼻孔などの調音器官の音響的共鳴作用により、音素を特徴づける周波数選択特性が形成され、明瞭な音声生成されるというアイデアであった。音声波形そのものは、時間とともにかなり速い振動成分を含み、約4000Hzの周波数帯域を占めるが、それぞれの調音器官の動きは一般に緩やかで、せいぜ

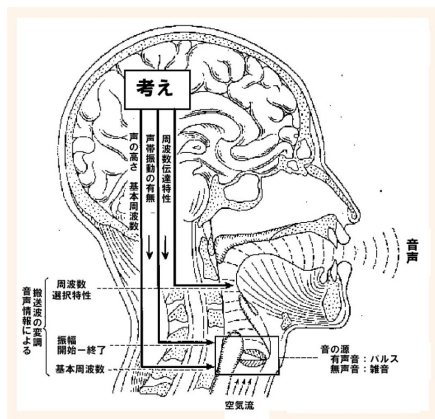


図 1. 音声の生成メカニズム(Dudley 考案:1928)

い20Hzくらいまでの周波数帯域を占めると考えられる。この考えをDudleyは、「音声のキャリア特性」と呼んだ。すなわち音声は、音声をキャリア (エネルギーの運び手) として、これを音声言語情報 (振幅、有聲/無聲の区別、基本周期、周波数選択特性、) により変調したものであるというモデルである。実際、Dudleyは、音声信号から調音器官の動きを抽出し、それらから音声を合成することを試みたが、この処理は大変困難 (現在の技術でも難しい) あった。1936年になって、調音器官の動きの直接抽出する代わりに、音声の周波数スペクトルを抽出して伝送し、受信側で音声を合成する、いわゆる、チャンネルボコーダを発明したのである。1939年にDudleyは、このチャンネルボコーダにより、音声の帯域を200~300Hzに圧縮できることを実演して示すことができた。

図2は、チャンネルボコーダの原理を示す。送信側では音声信号を多数の帯域通過フィルタと整流平滑回路によって周波数分析し、周波数スペクトルを抽出する。また信号の周期性の有無および周期がある場合は基本周期(ピッチ)も抽出する。これらの周波数スペクトルと周期性に関する情報を伝送し、受信側では周期性の有無に応じてパルス発生器あるいは雑音発生器を制御して音源信号を作成し、これを周波数スペクトルに応じて変調し、分析側と同様な帯域通過フィルタに通した後、総合して音声を再生する。

70年も前にDudleyが考案したこのアイデアは、現在も受継がれ、音声符号化、音声合成、音声認識などの研究開発の出発点として生かされている。当時の電気回路技術では、チャンネルボコーダは大変大掛かりな装置であり、また受信側で再生できる音声の品質が不十分であったため、一般の大陸間商用電話に実用されることはなか

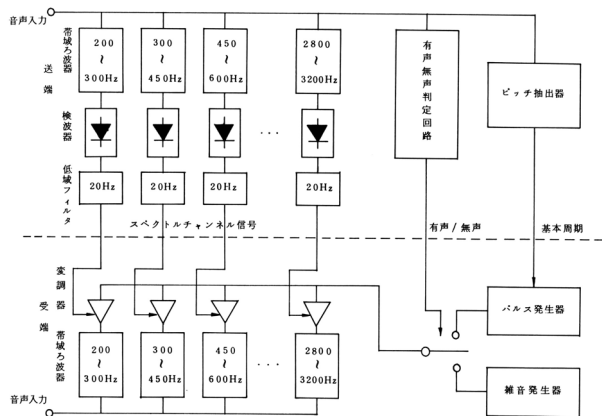


図 2. Homer Dudley のチャンネルボコーダ(1936)



図 3. 大陸間無線秘話電話システム Sigsaly

ったが、第二次世界大戦中1943年4月、ロンドンのチャーチル首相とワシントンのルーズベルト大統領の秘密会談用の大陸間無線秘話電話システムSigsalyとして実用された。この装置の重さは約55トン、消費電力は30kWにのぼり、とても事務室に入るようなものではなかった。(図3)Sigsalyは、ワシントン-ロンドンについて、北アフリカ、パリ、ハワイ、グアム、シドニー、マニラに、また終戦後にベルリン、フランクフルト、東京など全部で12ヶ所に設置され、1946年まで3000回の極秘会談に使われたという記録が残っている。

2. 音声分合成系の近代化

Dudley に端を発する Vocoder 技術は 1966 年に確率過程論をベースにした新しい形に近代化した。その代表がこれから述べる PARCOR および LSP 音声分析合成方式（以下 PARCOR 方式、LSP 方式と略す）などである。音声信号から音声の周波数スペクトルのパラメタを統計的な手法により分析抽出し、それを低ビットレートで符号化伝送し、再びもとの音声を合成する技術であり、現在、音声帯域圧縮伝送、音声合成、あるいは音声認識の前処理などの音声処理の一つの基礎技術になっている。(付録図面参照)

3. 最尤スペクトル推定法

筆者は、1965年3月に名大大学院修士課程を修了し、同年4月に博士課程に進学したが、その直後指導教授の宇田川先生が急逝されたため、専門を音声・音響分野に変更することになった。修士課程で統計的パターン認識や時系列解析の勉強をしてきたため、音声認識を統計的パターン認識の立場から追及してみようと考えた。その当時、筆者は音声分野について最小限の知識しか

持ち合わせていなかったが、福村晃夫先生のご指導で、Fant 先生の "Acoustic Theory of Speech Production" の特別輪講（受講者は私1名）のお陰で、一応の予備知識を獲得することができた。

1966年4月に、NTTの奨学金を得て通研実習生として音声の研究を始めた。その当時、通研に基礎研究部が新たに設置され、斎藤収三第四研究室のもとに、音声・聴覚・画像・映像に関する新進気鋭の研究者が集まり、活気に満ちていた。

通研に来て、まず自分の声をサウンドスペクトログラムで分析してみたが、私の声あまり良くないせいか教科書に載っているようなきれいなフォルマント構造を示すスペクトルパターンが表れない。そのときに音声は想像以上に複雑な信号で多くのランダムさを含んでいるに違いないと考えた。そこで、確率過程の時系列解析の手法を使って音声分析を試みる事にした。初めは、音声波形を時間的に連続な p 次の全極型定常確率過程としてモデル化し、そのパラメタの違いを最適判別する問題を考察した。この問題はモデルの母集団確率過程の十分統計量を求める事に帰着されるが、連続波形モデルでは、十分統計量が波形の $1 \sim p$ ($=10$) 階までの微係数の電力で与えられる事が解り、早速、実際に電子回路を製作して母音のスペクトルを分類する実験を始めた。当然の事ながら、波形の高階（10階！）の微係数は、わずかな付加雑音によっても大きく乱され、このアイディアは見事に失敗に終わった。

ちょうどそのころ、通研基礎研究部4研に音声研究用にAD/DA変換器を装備した計算機(FACOM 270-20)が導入される事になっていたので、デジタル信号処理を意識して、時間的に離散的な p 次の全極型定常時系列モデルにもとづく検討を行い、昭和41年度電気通信学会全国大会において初めて発表した⁽¹⁾。それによると、モデルの最適判別に必要な十分統計量は、初めの p 個の自己相関係数値である。従って、分析のための計算量も少なく、連続確率過程の場合に問題になっていた高階の微係数は入ってこないの、応用上も好都合であった。この研究の副産物として、全極型モデルの最尤推定法と音声スペクトルの合成法が発見され、ML音声分析合成方式と言う概念が生まれた。これは、後に線形予測符号化(LPC)と呼ばれる事になったものと数学的に等価である。この理論によれば、音声合成に必要なパラメタが、入力波形の自己相関を係数とする p 元連立方程式（正規方程式）の求解により推定できる。またフォルマント周波数の推定は、こうし

て求めたパラメタを係数とする p 次代数方程式の求根に帰着される。

ここまで解ってくると、音声分析合成方式に挑戦したくなり、斎藤収三室長に相談したところ、「分析合成方式の重要な問題点は音声基本周期などの音源情報の抽出であるから、まずその問題を解決しなければならない」という忠告を受けた。偶然にも、最尤法ではスペクトル包絡を逆フィルタの形で抽出しているのを、それを使えば基本周期情報が抽出できる事に気がつき、早速シミュレーションをしてみた結果、従来最も有力とされていたケプストラム法に比べて遜色のない方法である事が判り、これを変形相関法と名付けた。最尤音声分析合成方式と変形相関法を組み合わせる事によって、ML方式が一応完成したのが1967年夏である。その分析合成音は予想以上に自然でありまた明瞭性もあった。これらの結果をまとめて、1968年8月東京で開かれた第6回国際音響学会議(6thICA)で発表する機会を得た。図らずも、同会議の同じセッションで、ベル研究所のSchröder, Atal両氏による音声の適応線形予測符号化(APC)が発表され、これらが現在の線形予測音声符号化の先駆的な論文と言えよう。

4. ML方式からPARCOR方式への発展

ML方式を使って、低ビットレート符号化の詳細な実験を進めている内に、 α パラメタ(LPC係数)を10ビット以下に量子化すると、時々著しい

異音が発生する問題に遭遇した。その原因は、パラメタ量子化により合成用帰還形(IIR)フィルタが発振するためである。これを抑えるために、いろいろな対策を打ってみたが原理的な解決法とは言えなかった。1969年3月、東京神田の古書街で一冊の数学書 Szegő 著 Orthogonal Polynomials(1956)に出合った。ばらばらと立ち読みしてみたところ懸案の問題を解決する糸口が見つかりそうな気がしたので買い求め、読み進む中に音声合成フィルタ伝達関数の分母の多項式が直交多項式により展開出来る事に気が付いた。また、直交多項式の漸化式によって、 p 元の連立線形方程式がいわゆる Levinson の再帰法によって解ける事も明らかになった。さらに直交多項式の漸化式をデジタルフィルタの中に埋め込む事によって、前向き予測誤差と、後ろ向き予測誤差を格子型フィルタによって構成できる事、および2つの予測誤差の相関係数が、音声波形の偏相関係数(PARCOR; Partial Correlation)に相当し、それらの絶対値が全て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。(図4) 1969年7月21日に開催された日本音響学会の音声研究会でA4版20ページの詳細な資料を作成して発表したのであるが、当日が月ロケット・アポロ11号の乗組員が月面に降りた日であり、研究会参加者はわずか数名であり反響もほとんどなかった。その後、PARCOR方式は、

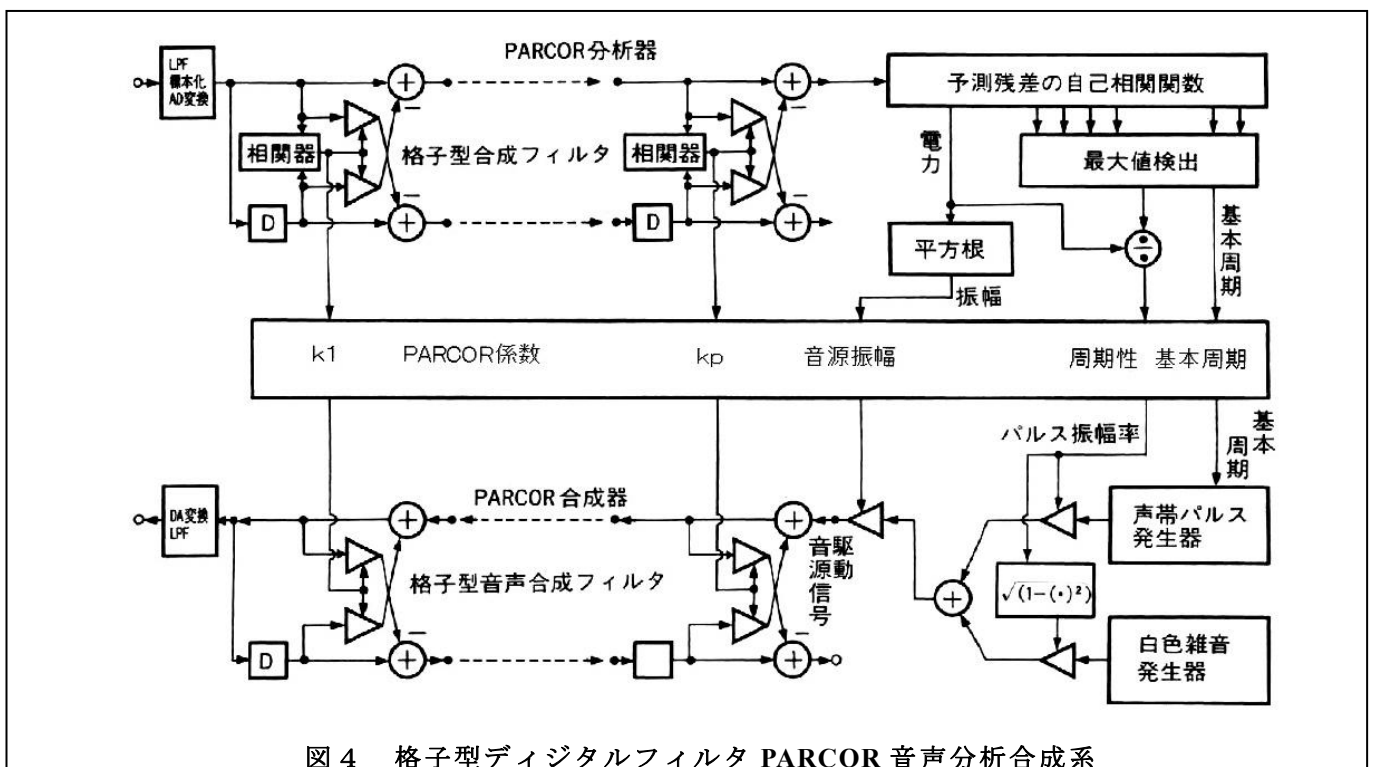


図4 格子型デジタルフィルタ PARCOR 音声分析合成系

第7回国際音響学会議（1971年，Budapest）に投稿した格子型音声分析合成フィルタの理論と、音声通信と処理に関する会議（1972年，ボストン，ICASSPの前身）のPARCOR係数の最適量子化の論文によってようやく国際的にも認知された。

PARCOR方式が誕生した1969年当時は、LSI技術も未熟だったので、そのハードウェアによる実現は簡単ではなかった。例えば、1970年に試作したPARCOR音声合成器は、汎用プリント基板約400枚からなり、5V・50Aの直流電源を消費する巨大な代物であった。しかし、この装置の実演に促され、通研の基礎研究部と宅内研究部の協力により開発され、PARCOR方式は、パラメタ編集方式の音声応答装置として、1971年に実用化され³⁾、その後、各種の電話によるデータサービスに供された。

1976年には、米国TI社からSpeak&Spellという商品名の形で、PARCOR合成の機能をもつ音声合成LSIが民生用に作られ、広く実用化されるようになった。また、米国国防省は、1976年LPC10-eと呼ばれるFederal Standard 1015を制定し、軍用秘話通信方式を実用化した。この方式の原理はPARCOR方式そのものであり、音声を2400bpsで符号化する方式である。この方式を実行するために必要な計算速度は約7Mipsであり、比較的少ないが、実用的な騒音環境での音声品質は、MOS値で2.3程度であり、一般の商用通信用の音声符号化方式としては、不十分であったと言わざるを得ない。

その後、1996年になって、音源情報として混合駆動LPC（MELP）が新しい軍用通信の標準として採択された。この改良によって、計算量はLPC-10eの6倍くらいに増加するが、2400bpsにおけるMOS値は、3.2に向上した。

5. 最尤スペクトル推定法による音声認識

PARCOR方式の基礎的検討が一段落した後、1973年8月から2年間、ベル研究所音響研究室の客員研究員として過ごすことになった。当時のベル研は、世界の電子情報通信のメッカとも呼ばれていた。音声分野においても、Fletcher, Dudleyほか多数の聴覚・音声・音響分野のパイオニアがうまれた事は周知の通りである。当時の音響研究室には研究室長のFlanagan博士の周りには、Rabiner, Schafer, Sondhi, Atal, Jayant, Coker, Umeda, Hall, Ishizaka, Noll, Rosenberg, Sessler, Berkley, Allen, Samber, Levinson, Chrochiere 諸氏のような現在のデジタル信号処理、音声・音響

処理の開拓者がきら星のごとくそろっていて、通研にもましてきわめて活気のある研究室であった。

Flanagan氏から与えられたテーマは、”Digital Processing of Speech”ということで、ほとんど何の拘束がないようなテーマであった。ベル研での滞在期間は限られていたので、最尤スペクトル推定法で求めたスペクトル距離尺度とDTW（Dynamic Time Warping）を統合した単語音声認識システムを試作した。（図5）ベル研での滞年の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。

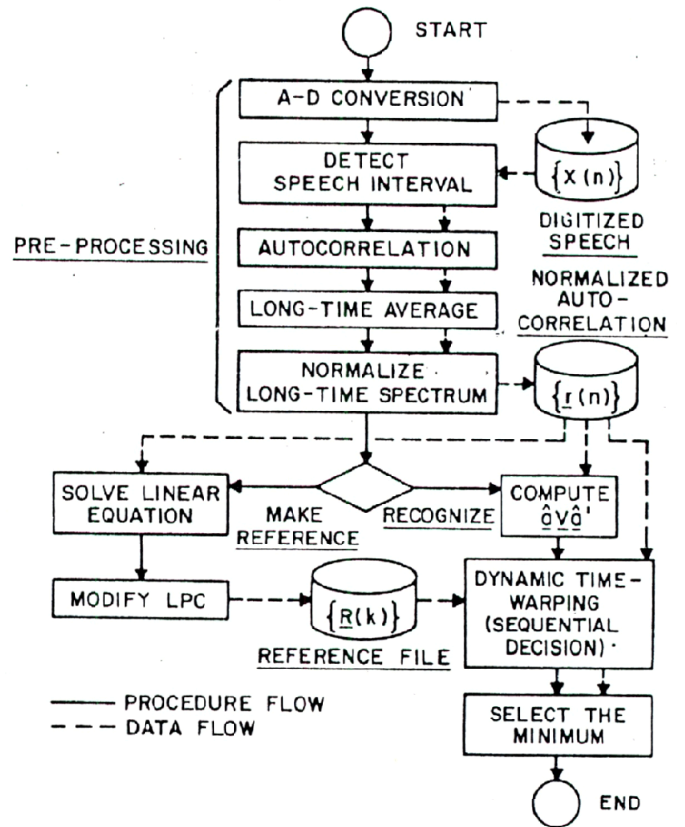


図5. 最小予測残差原理の音声認識への応用

入力音声信号の自己相関関数を求め、長時間スペクトルに関する平均化と正規化を行う。標準パターンは変形予測係数の形で蓄積し、未知信号の正規化自己相関関数との内積をとって、距離尺度を計算する。時間軸の非線形伸縮はDPマッチングにより行う。

(IEEE Trans.ASSP-23, No.1, 1975)

6. LSP 方式の誕生

ベル研滞在中の 1974 年の暮れ頃、図書室に行って数学書を漁っていたところ、Toeplitz 行列のスペクトル分解に関するものを見つけた。この理論を PARCOR 合成フィルタと結びつけると、全極型のフィルタの線スペクトル表現が可能であることがわかった。

PARCOR 係数は、声道断面積関数を p 個の等長音響管でモデル化した時の断面積の不連続点における波動の反射係数であると見なされる。唇端は体積流に対して完全開放のため反射係数 = -1 と見なし、声門端は、開放時には反射係数 = 1、閉鎖時 -1 とモデル化する。このように音響管の両端の反射係数を ± 1 にすると、波動エネルギーが閉じこめられ、声道共振系は無損失系になる。従って、音響管の伝達関数は線スペクトル状になる。

唇端からかぞえて $i=1, 2, \dots, p$ 番目の反射係数を k_i とする。 m 次の前向き予測演算子、後向き予測演算子を $A_m(D), B_m(D)$ 、 $D=$ 遅延、とすると、次の漸化式が成立する：

$$\left. \begin{aligned} A_m(D) &= A_{m-1}(D) - k_m B_m(D), \quad A_0(D) = 1 \\ B_m(D) &= D[B_m(D) - k_m A_{m-1}(D)], \quad B_0(D) = D \end{aligned} \right\} (1)$$

声門端反射係数 $k_{p+1} = \pm 1$ の時、全極型伝達関数の線スペクトル周波数は、次式の ω に関する方程式の根として与えられる。

$$\left. \begin{aligned} P(z^{-1}) &\equiv A_{p+1}(z^{-1})|_{k_{p+1}=1} = 0 \\ Q(z^{-1}) &\equiv A_{p+1}(z^{-1})|_{k_{p+1}=-1} = 0 \end{aligned} \right\} (2)$$

$$\text{ここで, } P(z^{-1}) = 0, \quad Q(z^{-1}) = 0 \quad (3)$$

の根が単位円周上にあることに注意すると、

$P(z^{-1}), Q(z^{-1})$ は、次のように因数分解できる：

$$\left. \begin{aligned} P(z^{-1}) &= (1 - z^{-1}) \prod_{i=2,4,\dots,p} (1 - 2\cos\omega_i z^{-1} + z^{-2}) \\ Q(z^{-1}) &= (1 + z^{-1}) \prod_{i=1,3,\dots,p-1} (1 - 2\cos\omega_i z^{-1} + z^{-2}) \end{aligned} \right\} (4)$$

ここで p は偶数とした。このとき、LSP (線スペクトル対周波数) $\{\omega_i\}$ は、次のように奇数番目

と偶数番目のものが、必ず交互に並ぶように順序づけることができる。すなわち

$$\{0 < \omega_1 < \omega_2 < \omega_3 < \omega_4 < \dots < \omega_{p-1} < \omega_p\}$$

この条件が満たされる時、またそのときに限り、

$$1/A_p(z^{-1}) = 2/\left[(P(z^{-1}) + Q(z^{-1})) \right] \quad (5)$$

で表される音声合成フィルタは安定であることが示されている。下の図 6 は、式(4), (5)により求まる LSP 方式の音声合成フィルタのブロック線図である。この原理に基づいて、音声分析合成系を構成し、その性能を把握する努力が、1976 年から 1981 年にかけて精力的に行われ、その成果に基づいて、LSP 音声合成用 LSI の開発も行われた。

LSP は、PARCOR 係数に比べて、パラメタ量子化誤差の影響が少なく、また、パラメタを時間的に補間したときのスペクトル再現精度が高いという長所を持つ。このため、LSP パラメタは、現在も多くデジタル携帯電話の音声符号化方式に利用されている。

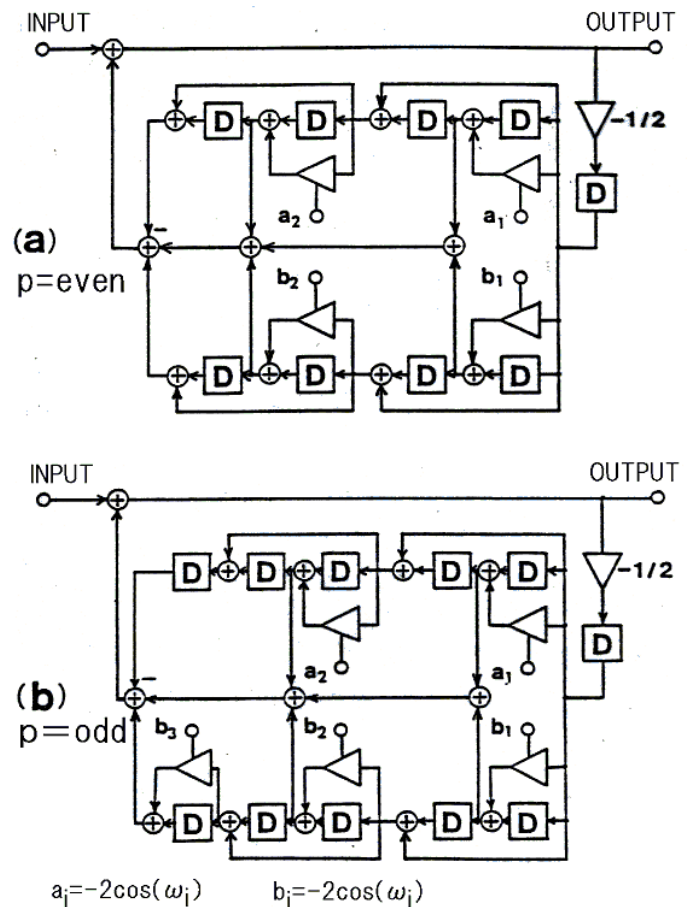


図 6 LSP 音声合成フィルタの構造

7. ハイブリッド音声符号化 CELP

以上のべた、音声分析合成系は、音声信号伝送情報量を 2400bits/s 程度まで圧縮することができるが、その背景には、音声が入間の発声器官により生成された信号であり、そのスペクトル包絡は全極形の伝達関数で近似され、更に音源信号は有声音の場合準周期的パルス列、無声音の場合はランダム雑音でモデル化できるという暗黙の仮定がなされている。雑音や歪を含む現実の電話音声においても、スペクトル包絡に関する仮定は、さほど深刻なものではないが、音源信号に関しては、上記の仮定は重大で、しばしば音声分析合成は破綻する。言い換えれば、音源信号を単純なパルス列でモデル化することは困難であり、より現実の信号に則した音源信号のモデル化が必要になる。

そこで、音声分析合成系の単純パルス駆動に代わって、残差駆動、マルチパルス駆動、あるいは符号駆動の線形予測符号化が提案された。ここでは、世界中の携帯電話で、音声を 4000 乃至 8000bits/s に符号化し、実用に供しうる音声品質をもつ符号駆動線形予測符号化(CELP)について、紹介する。

図 7 に CELP エンコーダ・デコーダの符号化手順を示す。最初に入力音声信号の線形予測分析を行い、線形予測係数を求める。これを LSP パラメータに変換し、最も近いものを LSP コードブックの中から選ぶ。さらに後のコードブック探索のため

に、選んだ LSP パラメータを再び線形予測係数に変換しておく。次にピッチ分析、固定符号帳、利得符号帳の探索を順に行い、量子化された予測係数で合成した音声が入力信号に近くなるようなインデックスを求めます。

デコード側では、駆動信号は適応符号帳と固定符号帳の信号にそれぞれ利得を掛けて加算する事により得る。駆動信号を合成フィルタに通すと音声再生される。適応符号帳の実体は過去に生成した駆動信号そのものである。音声信号は周期性が強いので過去に生成した駆動信号を保存しておく、ピッチ周期情報に基づいて再利用する。固定符号帳は予め決められた複数の波形データの集まりであり、そのインデックスが符号データになります。特に ACELP の場合は固定符号帳として振幅が一定のパルス(振幅情報が無い)を使用するため、パルスの位置情報だけで駆動信号が表現できる。そのため符号帳のためのメモリが必要なくなる。またエンコーダにおいては代数的な特徴を利用して、少ない演算処理で符号帳の探索を行う事ができる。利得符号帳には適応符号帳と固定符号帳の利得情報が両方とも含まれている。2種類の利得情報を組にしたものが複数用意されており、そのインデックスが符号データになる。

通常、合成フィルタには線形予測フィルタが用いられる。合成フィルタの情報として線形予測係数と相互変換が可能で、かつ符号化に適した LSP パ

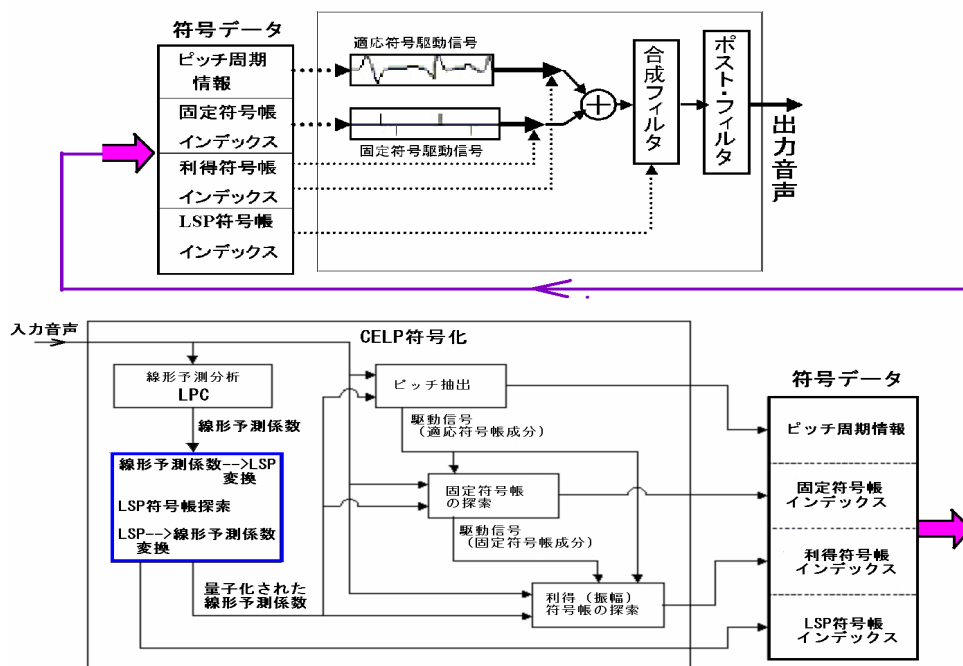


図 6 符号駆動線形予測符号化 (CELP) の構成

ラメタが使用される。たくさんの LSP パラメタが符号帳に用意されており、そのインデックスが符号データになる。

CELP は、音源情報およびスペクトル情報に対して、徹底的にベクトル符号化の考えを適用して情報圧縮を実現し、波形情報も、ある程度保存することにより、入力信号に含まれる雑音や歪への耐性を向上している。

8. むすび

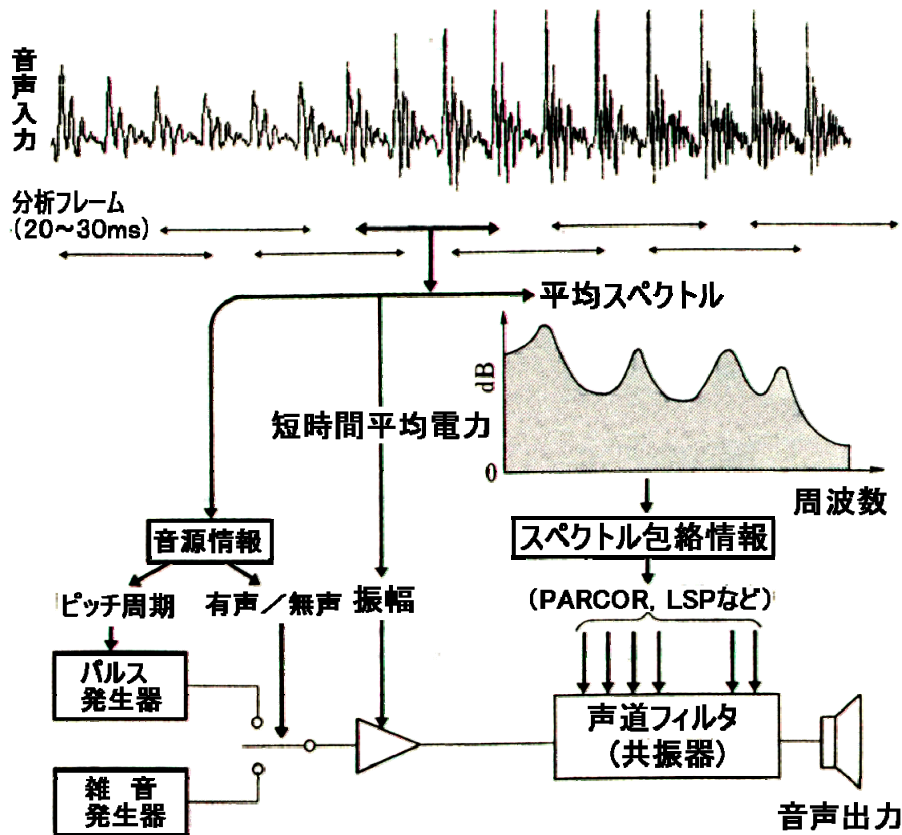
音声情報処理の研究は、音声生理、知覚、音声学などの基礎的研究と関連しながら、最近のマイクロエレクトロニクスとソフトウェア技術をベースにして、過去40年の間に長足の進歩を遂げてきた。

音声情報処理の目標は、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。ここで紹介した音声分析合成技術は、「千里の道の一里塚」にすぎないが、数理的な基礎が強固であるため安心して応用でき、携帯電話や IP 電話への適用に見られるように実用性も高いものであった。

一方、現在、精力的に研究開発が進められている音声対話システムは、巨大なソフトウェアシステムである。それ故、その構成要素の中に数理的に脆弱なものが多々見受けられる。こうした弱点を取り除き、理論的にも強化してこそ、信頼できる人間・機械間の対話が実現されると考える。

文献：(ごく初期のものに限定)

- 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(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, “複合正弦はモデルによる音声分析合成系”, 音声研資 S79-6(1979)
- 6)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)



付録図面：低ビットレート音声符号化の原理