

論文 / 著書情報
Article / Book Information

論題(和文)	音声および音響
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	第3部 最近25年間における電子情報通信技術の発展, , , pp. 232-237
Citation(English)	, , , pp. 232-237
発行日 / Pub. date	1991,
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1991 Institute of Electronics, Information and Communication Engineers.

A 基礎・境界

1. 音声および音響

古井 貞 熙 (日本電信電話株式会社)

1.1 音 声

1.1.1 音声分析法

最近の25年間に、音声情報処理技術は目覚ましい進歩を遂げてきたが、そのきっかけとなる最も大きなエポックをもたらしたのは、1968年に東京で開かれた第6回国際音響学会議で発表された、電電公社(現NTT)電気通信研究所の板倉、斎藤による最尤法を用いた音声分析合成法と、米国ベル研究所のAtalらによる予測符号化法の二つの論文である。この二つの論文は、いずれも音声波形を線形予測して情報圧縮するもので、確率過程論など数理的手法を背景としており、その後、線形予測分析法(linear predictive coding, LPC)と総称されて、一つの大きな学問体系をなすに至った。

その後、線形予測分析法の枠組みの中で、情報圧縮率、合成フィルタの安定性、パラメータの補間特性などに優れた、パーコール(PARCOR)係数、LSP(線スペクトル対)パラメータなどが提案され、これらを用いた高い品質の音声分析合成系が実現された。線形予測分析法の優れた特徴に、スペクトル包絡が比較的少数のパラメータで能率的に表現できるだけでなく、これらのパラメータで構成した逆フィルタに音声を通すことにより、声道特性に対応したスペクトル包絡特性を自動的に除去できる点がある。その残りの波形(残差波形)の自己相関関数(部分自己相関関数)を計算することにより、ピッチ(声帯の振動周期)が正確に抽出できる。このことは、音源(声帯)情報と声道情報を別々に抽出して、それらの情報から音声を再合成する音声分析合成系の性能向上に大きな貢献を果たした。

ケプストラム(cepstrum)は、対数スペクトルの逆フーリエ変換として定義され、地震波の解析を目的として、1960年代の前半に考案された方法であるが、音声の音源情報と声道情報のように、畳み込み

の関係にある情報を自動的に分離できる性質があるため、初めはピッチ抽出の方法として音声に適用された。その後、スペクトル包絡を表現する方法としても優れた性質をもつことがわかり、その上、線形予測係数から直接的に計算できるいわゆるLPCケプストラムは、線形予測分析法とケプストラム分析法の両者の長所を持ち、比較的少ない計算量で求めることができるため、特に音声認識のための音声分析の中心的方法として、今日まで広く用いられるに至っている。

人間による音声生成のメカニズムのモデル化に関しても、種々の研究が行われているが、中でも特筆すべきものは、石坂、Flanaganによる声帯振動の2質量モデル(two-mass model)である。これは、声帯を声道に近い側と、気管に近い側の二つの部分に分け、その間をバネで結合したモデルである。このモデルに各種の生理的条件や実測値を代入し、電気的な等価回路を導いて、声道との結合を含む計算機シミュレーションにより、声帯振動が起こる(発振)条件、開口断面積の時間変化、通過体積速度の時間変化、声帯の振動周波数などを求めた。その結果、このモデルは、実際の人間の声帯の振動の仕方をよく表現していることが確認された。

音声の特徴の能率的表現法(情報圧縮法)として、極めて大きなインパクトを与えている方法に、ベクトル量子化法がある。これは、波形やスペクトル包絡パラメータの複数の値の組(ベクトル)をまとめて、一つの符号で表現する方法である。これにより極めて高い情報圧縮率が得られるため、音声符号化のみならず音声認識にも広く用いられており、スペクトル空間のノンパラメトリックな表現法として、音声認識における話者適応化法などの新しい技術を生み出している。ベクトル量子化をさらに進めたマトリックス(セグメント)量子化についても研究が行われている。

1.1.2 音声符号化法

音声符号化の方法には、上で述べた音声分析合成系（ボコーダ）の他に、波形符号化方式がある。前者が音声の生成モデルに基づいて、音声をパラメータに変換し符号化する方法であるのに対し、後者は波形をできるだけ忠実に再現しようとするものである。波形符号化方式のうち、最近の25年間に提案された主な方式に、音声信号の統計的性質や動特性を利用した、適応予測符号化（APC）、帯域分割符号化（SBC）、適応変換符号化（ATC）、適応ビット割当 APC（APC-AB）などがあり、分析合成系と波形符号化方式の組み合わせとして、マルチパルス符号化（MPC）、符号駆動線形予測符号化（CELP）などが提案されている。特に CELP は、4～8 kbps 程度の情報量において、比較的高い符号化音声を実現できるため、デジタル自動車電話、携帯電話などの符号化方式の基本として、広く用いられつつある。

CELP 方式では、音声波から短時間毎に線形予測分析によってスペクトル包絡を推定し、この推定結果に基づく合成フィルタに通したのちの出力が、実際の音声波形にできるだけ近づくように、最適な音源波形をベクトル符号帳から選択する。符号帳にはあらかじめ、種々のランダム雑音を蓄えておく。ランダム雑音の代わりに種々のパルス列を用いるのが MPC である。CELP の一種である VSELP (vector sum excited linear prediction) 方式は、音源ベクトルを少数の基底ベクトルの線形結合で表現し、さらに多段のベクトル量子化法を用いることにより、伝送符号誤りに強く、符号帳探索が高速で、しかも符号帳の記憶容量が少ないという特徴を実現している。VSELP 方式は、日米のデジタル自動車電話における標準の音声符号化方式に採用されている。

これらの電話音声を対象とした数 kbps 程度の符号化の他に、数 100 kbps 程度の高品質オーディオの符号化や、数 100 bps 程度の極低ビット符号化の研究も進められている。音声符号化の重要な共通技術の一つに、ノイズシェイピングの提案がある。これは、音声の周波数スペクトルよりも 15 dB 程度低いいき値以下の雑音は知覚されない聴覚的マスキング現象を利用して、各周波数の雑音のスペクトルが、いき値以下となるように、雑音のスペクトルを調整

する方法である。これにより、同じ S/N 比でも聴覚的な雑音を低減することができる。

1.1.3 音声合成法

音声合成の基本的方法は、単語や定型文の音声波形をそのまま蓄えておいて、必要に応じて接続、再生する録音編集方式、それを分析合成法によってパラメータの形で蓄えるパラメータ編集方式、任意の言葉が生成できる規則合成方式の3種類に大別できる。このうちパラメータ編集方式は、線形予測分析法などによる音声分析合成技術の進展により、音声品質の大幅な向上が達成された。これにより、録音編集方式と合わせて、音声合成が駅の案内放送などかなり広い範囲で実際に用いられるようになってきた。

しかしこれらの方法では、生成できる言葉の種類が限られてしまうので、多様なサービスへの適用をめざして、規則合成方式の研究が盛んに行われるようになった。特に最近では、かな漢字まじりで書かれた通常の任意の文章を、人間が手を加えることなく音声に変換できる、文・音声変換（テキスト音声合成）技術が進んできている。その一連の処理の第1段階である文章解析に関しては、大量の辞書情報や、種々の文法および経験的規則が整備され、それに基づいて音素や韻律を決める音韻および韻律規則の高度化が進んでいる。音声を生成する段階に関しても、いわゆる調音結合を考慮して前後の音素の影響を反映した音素単位などを用いることにより、合成音声品質が大きく進歩した。

1.1.4 音声認識法

音声認識の形態は、区切って発声された単語を認識する単語音声認識と、単語を連続して発声した音声を認識する連続音声認識に大別できる。後者はさらに、連続数字音声のように、単語の連なり方に特に制約がなく、音としての性質のみを用いて認識する連続単語音声認識と、通常の文章のように文法や意味に関する規則に従って単語が連なっているものを認識する文音声認識（音声理解）に分類できる。単語音声認識に限定しても、周囲の雑音や音声の個人差への対処など難しい問題が存在し、実用システムや装置はまだ必ずしも広く用いられるところまでいっていないが、この25年間に研究対象は単語から

文章へ大きく展開しており、多数の新しい技術が開発されている。

音声認識の最も基本的な部分は、音声スペクトルの時系列を抽出しパラメータ表現する段階であり、種々の研究が行われてきたが、ケプストラムと、その動的特徴であるデルタケプストラムを用いる方法が提案され、これによって認識性能の大きな改善が実現された。次の段階は、標準パターンと入力音声のパラメータ系列の比較であるが、迫江、千葉らによって提案された、動的計画法に基づくDPマッチング法は、時間軸の非線形伸縮を許して最適なマッチングが能率的に実行できる方法として、その後の音声認識の研究に極めて大きなエポックとなった。DPマッチング法は、長い間音声認識の基本的方法として用いられてきたが、最近ではそれに代わって、隠れマルコフモデル(HMM)が、中心的方法として用いられるようになってきた。HMMは、音声の時間軸の伸縮だけでなく、スペクトルの変動も含む種々の変動を統計的なマルコフモデルによって表現する方法であり、種々の優れた特徴を有している。比較的最近、ニューラルネットの利用に関しても種々の研究が広く行われているが、まだ従来の方法を大きく上回るには至っていない。

連続単語音声の認識法に関しては、DPマッチング法を単語内と単語間の両方に適用する2段DP法や、その改良であるLB(level building)法などが種々提案され、連続発声された数字系列の認識などに用いられている。通常の記事音声の認識に関しては、階層モデル、黑板モデル、ネットワークモデルなどの種々の構成モデル、縦型、横型などの種々の木探索法、文脈自由文法を中心とする種々の構文および意味の表現法が研究され、1000語程度の語彙を対象とし、タスクを限定した連続音声認識システムの研究が行われるようになった。最近では、音響処理部とそれに続く言語処理部を合わせて、通信理論に基づく確率モデルとして定式化する方法が広く用いられている。この際、音響処理部はHMMによってモデル化し、言語処理部には単語列や文字列の生起確率を用いた統計的言語モデルを用いることが多い。大語彙の連続音声認識においては、音響の基本単位として音素を用いる方法の研究が盛んに行われ

るようになった。

音声認識における重要な課題の一つとして、話者による音声の違いへの対処法がある。その一つの方法は、話者が変わるたびにすべての語彙を学習しなおす方法(特定話者型)であるが、これでは語彙数が大きくなったときに対処できないので、話者による変動を初めから取り込んだモデルを用いる方法(不特定話者型)や、新しい話者の短い音声を用いて、システムをその話者の音声に自動的に適応させる方法(話者適応型)の研究が活発に行われるようになり、ベクトル量子化法などを用いた種々のアルゴリズムの提案や比較が行われている。

1.1.5 話者認識法

音声に含まれる個人性情報を用いて、誰の声であるかを自動的に判定する話者認識の方法は、判定に用いる言葉をあらかじめ決めておく発声内容依存型と、任意の言葉を用いることができる発声内容独立型の方法に分けることができる。前者に関しては、ケプストラムとデルタケプストラムの時系列のDPマッチング、あるいはHMMを用いる方法によって、高い精度が実現されるようになったが、後者に関しては、種々のアルゴリズムの提案段階にある。その中心的方法は、各話者特有のパラメータ分布を、ベクトル量子化の符号帳、あるいは複数のガウス分布を用いた確率モデルで表現する方法で、これにより大幅な性能向上が達成されつつある。

1.2 応用音響

音声を中心とする音による通信技術は、デジタル技術、コンピュータなどの急速な進歩により、大きな変貌を遂げた。その結果、応用音響の技術分野は、電気音響学、デジタルオーディオ技術、不規則振動・信号処理・制御、聴覚・環境心理、ヒューマン・マシンインタフェースなど多くの分野に密接な関連を持つに至った。以下、電気音響変換技術を核とする電話機の進展、音響・音場情報処理などについて述べる。

1.2.1 電話音響技術

1960年代から1990年頃までの間の電話機の歴史は、通信機から家電製品への脱皮を大きな流れとする激動の歴史であった。電話音響技術は、その中で

要求される技術を実現してきたが、その変化を導いた重大な要素は電話機への半導体回路の積極的な導入であった。

1960年代には、米国以外の国では電気通信システムを政府機関が一元的に管理運営しており、我が国でも電話音響技術は電電公社主導のもとに進歩してきた。1964年より導入された600形電話機は、その体制で戦後復興のため培われてきた技術の集積であり、これによって日本の電話機は世界水準にならぶ音響技術を獲得した。特に音響機器を定量的に設計する技術は世界のトップレベルとなり、日本製の電話用音響機器の米国などへの大量輸出の原動力となった。また1970年に発表された600L高損失加入者用電話機は、おそらく世界最高の感度をもつ電話機であった。しかし諸外国が電話通話品質の尺度としてRE（通話当量）を採用する中で、我が国はAEN（明瞭度等価減衰量）を捨てなかったため、600形電話機は諸外国の電話機に比べ、通話品質設計がやや特殊なものとなった。

1973年の石油ショック以来情勢は激変した。1978年に発表された601形電話機は何よりも物価上昇に対応するための改良形であったが、同時にAEN、RE両者の通話品質規格を満足する設計を実現した。それ以後の社会情勢は単一機種 of 電話機の量産を許さなくなったため、この電話機が我が国の旧型電話機、すなわち炭素送話器を用い、増幅器を用いない電話機の技術の集大成となった。1968年の米国でのカータホン裁定は、実質的に通信回線に接続する端末機器を自由化するものであったが、電電公社仕様の電話機以外の使用を認めなかった我が国でもこの方向への圧力が高まり、ついに1985年、電電公社の民営化と同時に、実質的に全ての電話機が自由化された。従来主観測定に頼ってきた電話機の通話品質評価を客観測定化する研究が結実して、AURALシステムが実用化されたのが前年の1984年であり、間一髪で間に合ったといえる。我が国の電話機の設計思想はここで大幅に転換し、半導体回路を積極的に取り入れた電子化電話機に脱皮した。

我が国における電子化電話機の嚆矢は1970年の大阪万博で電電公社が試用したコードレス電話機であろう。送・受話器には同一の小形電磁変換器が用

いられた。この技術はその後通常の電話機に適用され、1973年発表の700P電話機となった。ただし、電子回路は全IC化されていたが、一部に混成ICが用いられるなど高コストの問題を残していた。1980年前後より外国で電話機回路のモノリシックIC化が盛んになり、特に米国では開発途上国製の安価な製品が市場に現れた。中には技術が未熟で、電子化電話機の評判を落とした例もあった。自由化1年前の1984年に電電公社が発表した、800P電子化電話機「ハウディ」シリーズは、1チップICによる本格的な電子化電話機であり、送・受話器とリングには小形軽量の圧電変換器が用いられ、機構部品も小形化された。この電話機が自由化以後の電話音響技術の源流となった。

しかし、家電製品となった電話機に市場が要求したのはデザイン、機能の多様化であった。制度の改正によりコードレス電話機が爆発的に売られるようになり、音声情報圧縮技術と半導体メモリを活用した留守番電話機とともに電話機の使い方が多様化された。また多くの家庭用電話機がファッション商品化し、一家に数台の時代となった。最近の我が国の電話機の生産量は、電電公社の最盛期の2倍に達している。

現在の電話音響技術の課題の一つは高品質化であり、1988年のCCITTのADPCMによる高品質音声伝送方式の勧告と、同年のINSネット64のサービスインは、ステレオセット並の音の出る電話機への道を開いた。いま一つの課題は無線回線の有効利用への対応であり、安価で高能率のCODECの実現は、電話音響技術の重要なテーマといえる。また、机上のマイクロホンとスピーカで通話するハンズフリー電話機は古くから実用化されていたが、まだ問題が多く普及に至っていない。この技術の改善には音場制御技術の進歩に期待するところが大きい。

1.2.2 音響情報処理

1965年のCooleyとTukeyによる高速フーリエ変換（FFT）の提案、1968年の野田・南雲らによる適応処理アルゴリズムの提案、1972年のNHKと日本コロムビアによるPCM録音機の開発などにより、デジタル音響情報処理の時代が到来した。その後、多数のオーディオメカによる、コンパクト

ディスクなどデジタルオーディオメディアの研究開発、FFT 分析をはじめとするデジタル信号処理技術、精密計測及び音場可視化技術、マイクロホンアレーによる指向性制御技術、音響エコーキャンセラ技術、騒音源と逆位相で音を放射させる付加音源を新たに設置して、室内騒音を低下させるアクティブ騒音制御など、音響情報処理の研究分野は急速に拡大進展した。

雑音の中から所望の信号のみを抽出する技術は、Wiener フィルタをはじめ、これまで多くの研究がなされてきた。通常の室内での音声・音響情報を対象とした音響情報処理では、複数のマイクロホン（マイクロホンアレー）を用いた多点受音処理に多くの関心が寄せられており、複数のマイクロホンで受音した信号をデジタルフィルタで処理することにより、広い周波数帯域にわたって、音場の条件に適応した所望の指向性を実現する可能性が見いだされている。

適応処理アルゴリズムの応用は、これまで非現実的であると思われていた、数千以上の共振周波数が密集し複雑に変動する室内音場のエコーや騒音を、制御の対象として取り込むに至った。音響エコーキャンセラは、拡声通話系内に回り込む数百 ms から 1 s に及ぶ室内エコーを適応予測処理によって消去することができる。室内での通話者の大きな移動などに追従する高速適応処理手法の開発が今後望まれている。

適応信号処理技術の発展は、臨場感通信の実現に不可欠な音場再生技術に大きな変革を起こしつつある。1960 年代後半から 1970 年代前半における、ベル研究所とゲッチング大学の研究は、大きなコンサートホールで聴取者が聞き取る音を、コンピュータシミュレーションによって正確に再現する可能性を実証した。この研究は、当時は両耳効果の研究や音響心理実験など特殊な目的に限られていた。しかし、両耳効果に対する関心の高まりや、近年の信号処理技術の進展などにより、これからの高品質通信、仮想現実感への応用が期待されるため、トランスオーラルステレオとも呼ばれ、多くの研究が進められている。

音響理論におけるコンピュータの利用は、拡声通

話系のエコー経路や明瞭度など室内の複雑な音波の伝搬特性の予測・分析、あるいはコンサートホールの設計などの手段として、不可欠なものとなっている。積分方程式による残響理論、代数学や関数論的な手法による音響伝達特性の解析・制御、過渡的な現象を解析するウェーブレット変換などには、従来手法の枠組を越えた新たな理論の展開が見られる。また、音響情報を用いた設備診断技術、両耳受聴・聴覚モデル、音響環境心理と大脳生理学、コンピュータ・オーディオメディア処理など、多くの新しい学問・技術の芽が誕生している。

1.3 超 音 波

超音波の技術分野は広く、非線形音響、水中音響、超音波トランスデューサ、超音波物性、超音波計測、医用超音波、強力超音波、表面波部品、超音波機能部品など、多岐にわたっている。特に超音波機能部品、非線形音響は、最近 10 年間で新しく発展している部門である。また、超音波計測における超音波顕微鏡や、医用超音波関連では、超音波診断機器の世界市場の 1/3 のシェアを日本が占めるに至っている。以下、いくつかの例について述べる。

1.3.1 超音波機能デバイス

超音波振動の動力的応用の一つとして考えられる超音波モータ、共振子の一つの応用とも考えられる超音波ジャイロスコープの研究などが発展して、超音波機能部品という技術部門が確立されるに至った。超音波モータは、超音波振動エネルギーを回転運動や直線運動エネルギーへ変換・利用する部品であり、その研究開発は、1982～1985 年の指田、熊田らの研究を契機に活発に行われている。低速度・高トルク、変速器のない直接駆動の実現、良好な制御性などが一般的な特徴で、カメラの自動焦点機構、ロールスクリーン式カーテンの駆動などにすでに応用されている。特性の安定性・耐久性の向上、超音波モータの特徴を生かした応用分野の開拓（出力 mW 以下のモータなど）が進み、日本で生まれたこの技術が広く世界で受け入れられることが期待される。

1.3.2 弾性表面波部品

弾性体の表面に沿って伝搬する波（弾性表面波）

は、1885年にRayleigh卿によって理論的に予見されたものである。工学的には、1964年に柴山・佐藤が表面波の発生・検出に成功したのに続き、電極配置が周期的構成をもつすだれ状電極の考案により、弾性表面波フィルタの研究が急速に進展した。各一對の素電極から出た波が干渉しあって、ある波長の波は強め合いあるものは弱め合う結果、周波数特性は帯域通過形となる。東北大学、NTT研究所、東芝など多数の機関の研究により、1980年代には、テレビジョン受像機の中間周波増幅回路用フィルタなど民生機器用表面波部品では、日本が世界市場の1/2以上のシェアを占有するに至っている。広帯域、高速度なアナログ信号の処理が表面波部品の特徴であり、デジタル素子との共存、発展が期待される。

1.3.3 非線形音響

Westervelt (1963) によるパラメトリックアレー音源の提案以来、非線形音響現象に関する関心が高まり、最近では音響ソリトンへの興味などを契機として、新たな非線形音響の分野が発展しつつある。パラメトリックアレー音源は、一つの送話器より周波数の異なる二つの超音波を同時に同方向に放射し、それらの音波の非線形相互作用によって生じる差周波数成分（差音）を利用する音源である。差音を低周波に設定することにより、伝搬損失が少なく、かつ鋭い指向性をもつ音源を実現できるのが特徴である。そのため、最近のパラメトリック音源による海底探査では、海底下およそ25 mにある10数mの土砂層の厚みまで観測することも可能になった。また、差音を可聴周波数帯域に設定して、特定

の領域の中にいる人だけに音声や音楽を伝える、オーディオスポットライトへの応用も考えられている。

非線形波動の伝搬には、奇妙な現象が見いだされている。非線形波形歪と伝搬速度の周波数依存性（分散性）による波形変化とのつりあいの下に現れる、安定な孤立波をソリトンという。一般に伝搬速度の分散性は波形のくずれを伴うため、超音波を利用した計測では大きな問題となっている。ソリトンの形成は、分散性媒質内においても波形の変化の少ない伝搬モードが得られる可能性を示すものであり、今後の研究が期待される。

上記の応用音響および超音波の部分に関して、NTT研究所東山博士に多大な協力を頂いた。その内、電話機技術については富士通研究所大賀博士に御討論頂き、東京工業大学上羽教授には、超音波に関する文献の紹介を頂いた。厚く感謝の意を表する。

参 考 文 献

- (1) 古井貞熙：“デジタル音声処理”，東海大学出版会（1985）
- (2) Furui S. and Sondhi M.M. (Ed.): “Advances in Speech Signal Processing”, Marcel Dekker, New York (1992)
- (3) 中川聖一：“確率モデルによる音声認識”，電子情報通信学会（1988）
- (4) 城戸健一、曾根敏夫、柴山乾夫、山口公典、中鉢憲賢：“基礎音響工学”，日本音響学会編，音響工学講座1，コロナ社（1990）
- (5) 古井貞熙：“音響・音声工学”，近代科学社（1992）