

報告番号

※ 第 950 号

主論文の要旨

題名 統計的手法による音声分析合成系
に関する研究

氏名 板倉文忠

音声を遠隔地に伝送し、あるいは長期間記録に留めておくためには、従来、「音響波形保存の原理」が採用されてきた。すなわち、発生された音声による音圧の変化をできるだけ忠実に伝送媒体あるいは記録媒体の変化に対応させるといふ原理である。この原理に基づく電話や蓄音器は、発明後ほぼ1世紀におよぶ技術的改善により、高性能かつ経済化が達成され、今日の情報化社会に貢献している。

今日の電話伝送系においては、1回線の音声を伝送するためには、アナログ伝送のばあい周波数帯域として300~3400Hz、信号対雑音比として26dB以上、デジタル伝送のばあいは、毎秒56000ビットの伝送容量をもつ伝送路が使用されている。この伝送容量は、発声による言語情報発生速度や聴覚の情報受入れ能力に比較して数倍ないし数十倍大きい。言いかえれば、今日の電話伝送路を通過する信号は、伝送しなければならない情報に比べて、極めて冗長度の高い信号であることがわ

かる。

このような音声信号のもつ冗長性に着目して、音声もより狭い周波数帯域あるいはより低い情報伝送容量で伝送・記録するための一連の技術を開発することが音声分析合成系の研究目標である。音声分析合成系においては、はじめに、音声を分析し、その中から音声の伝送・記録に不可欠な情報を担っているいくつかの特徴パラメータを抽出する。(分析過程) これらの特徴パラメータは、原音声に比べて低い情報容量の符号に変換され、伝送・記録される。(符号化過程) つぎに、伝送・記録された信号は、再び特徴パラメータに復号化され、更に音声に復元される。(合成過程)

このような過程を含む音声分析合成系の研究は、1939年ベル電話研究所のW・H・Dudleyによってスペクトルチャンネルボコーダが発表されて以来、数多くの研究が報告されているが、それらの多くは「音声のスペクトル情報、(すなわち、周期性の有無、周期性の有るばあい

の基本周期および周波数スペクトル分布)が保存されていれば、原音声の情報内容は正しく再現される」という仮説に立つ、いわば、「スペクトル保存の原理」に基づいている。「スペクトル保存の原理」に基づく音声分析合成系が、従来の「音響波形保存の原理」に基づく音声伝送・記録方式に比べて、極めて高い情報圧縮能力をもつことが指摘されながらも、明瞭性、自然性などの伝送品質の点で実用に耐えるものが実現されず、広く実用に供されるに至っていない現状である。

このように、従来の音声分析合成系の研究が十分な成果をおさめられなか、た最大の理由は、音声信号自体が複雑な時間変化を示す不規則な変動を含むため、これまでの決定論的な解析方法が必ずしも有効でないことによると考えられる。たとえば、声道駆動音源信号には、無声音のばあいも勿論、有声音のばあかも基本周期、音源波形に種々なゆらぎが含まれている。このような不規則変動要因によ

る分析精度の低下を最小限に保つためには、変動の存在を積極的に考慮に入れた統計的手法を導入する必要があると考えられる。

本論文の目的は、音声分析合成系におけるスペクトル情報の分析・符号化・合成過程に関して、統計的手法により統一的な理論的な検討をおこない、新しい音声分析合成系を提案し、実験的にその有効性を確認することにある。

本論文は、つぎに示す7つの章から構成されている。

第1章は、序論であり、音声分析合成系の目的、研究動向の概論と、音声分析合成系の基礎となる音声発生機構および聴覚の特性に関する概説をおこなう。つぎに、音声分析合成系の研究における問題点の所在を示し、本研究の方針を明らかにする。

第2章は、時間離散的定常確率過程であり、第3章以下の考察を進めていくうえで数理的取扱いの基礎となる時間離散的定常確率過程

のスペクトル解析について述べる。

第3章は、音声スペクトル密度の最適識別の理論である。はじめに、音声発生過程の物理的な特性および聴覚のスペクトル弁別特性を考慮にいて、音声信号を、零点をもたない有理スペクトル密度の確率過程にモデル化する。つぎに、このモデルに基づいて、短時間の音声信号の尤度関数を求め、スペクトル密度で特徴づけられる音声信号の最適識別法の理論を展開する。その結果、音声スペクトル密度の最適識別法は、入力信号の短時間自己相関係数の遅延時間の小さい一部を、スペクトルパラメータで重みづけした値によって判定する一種の線形判別関数であることが知られた。この理論から得られた結果は、パターンマッチングボコーダのような音声分析合成方式の基礎をあたえるばかりでなく、音声の機械認識に通じた新しい分析の方向を示すものである。なお、ここで誘導された入力信号の尤度関数は、第4章で述べる音声スペク

トル包絡パラメータの抽出法の出発点を与える。

第4章は、スペクトル包絡の抽出である。

音声スペクトル包絡パラメータの抽出法として、最尤スペクトル推定法と呼ばれる手法を提案する。この方法は、スペクトル密度に含まれる未知母数を、最大尤度基準により決定する方法である。この方法は、従来もっとも精密な音声分析法とされている「合成による分析 (Analysis-by-synthesis) 法」と同様に、入力信号のスペクトル包絡に、モデルとして想定したスペクトルをある適合誤差尺度に関してもっともよく適合させる方法である。最尤スペクトル推定法の特徴は、用いる適合誤差尺度が、入力信号のスペクトル包絡の山部分には鋭敏であるが、谷部分には鈍感であるという性質をもつため、スペクトル包絡パラメータが人間の聴覚のスペクトル弁別特性と整合した形で抽出される点にある。また、最尤スペクトル推定法によれば、スペクトル包絡パラメータが、入力信号の短時間自己相関係数

の遅延時間の少ない一部を係数とする連立一次方程式の解として一意に決定されるという著しい利点をもつ。さらに、この方法は、ホルマント周波数およびバンド幅の決定にも応用でき、その有効性が実験的に確認された。

第5章は、音声基本周期の抽出法である。

音声分析において、スペクトル包絡の抽出と同様に重要であり、かつ困難とされている問題として、音声基本周期の抽出がある。ほぼ同じ波形の概周期的な繰返しである有声音から、その基本周期を抽出することは、容易に見えるが、従来相関法などがかなり高度な手法を用いても、十分な精度が得られなかった。

その原因は、基本周期による波形の繰返しの外に、ホルマントによる見かけ上の周期が無視できないためである。この原因による基本周期の抽出誤差を減ずるために、先に述べた最尤スペクトル推定法によって抽出したスペクトル包絡パラメータを用いて、入力信号のホルマント構造による自己相関関数の見かけ

上の周期性を除去し、真の基本周期を正しく抽出する方法として、変形相関法を提案し、その有効性が実験的に示される。また、従来もつとも正確な基本周期抽出法として知られている「ケプストラム法」と比較して、同等以上の性能を有し、しかも、処理が単純であることが示される。

第6章は、最尤スペクトル推定法による音声分析合成方式である。第4章、第5章に示した分析過程により抽出された音声のスペクトル包絡パラメータおよび音源パラメータから、音声を復元するための合成過程について検討する。合成過程は、基本周期、振幅および周期性尺度から駆動音源信号を復元する部分と、音源信号からスペクトル包絡パラメータを用いて原音声のスペクトル包絡を復元する部分よりなる。駆動音源信号の復元に関しては、パルス音源と雑音音源の切替え方法について考察する。スペクトル包絡の復元に関しては、スペクトル包絡パラメータの量子化に

ともなう音声合成用ディジタルフィルタの安定性の問題について考察する。その結果、スペクトル包絡パラメータを量子化しなければ、音声合成用ディジタルフィルタは常に安定なこと、ならぬに、量子化による不安定化を緩和するための1つの方法としてスペクトル等化による方法を示し、この方法により量子化ビット数を約2/3に減じ得ることが示される。

第7章は、最尤スペクトル推定法による音声分析合成方式の通話品質である。この章では、本方式の情報伝送容量を決定する種々のシステムパラメータが通話品質におよぼす影響について、対比較試験、および明瞭度試験により定量的な検討を加える。その結果、本方式が、商用のバルス符号変調方式の1/10の情報伝送容量において、明瞭度等価減衰量は約3dB、最適レベルにおける単音明瞭度は約90%であり、従来の「スペクトル保存の原理」に基づく音声分析合成系に比較して優れた伝送品質を有する事が確認された。

これらの研究で得られた音声分析合成方式の原理は、単に音声の高効率伝送記録を実現する際の基本技術となるばかりでなく、近年その需要が高まっている音声によって電子計算機から人間へ応答する装置(音声応答装置)の応答可能な語彙数を経済的に増大するための有効な方法を提供するものであり、現在実用化のための検討が進められている。また、この研究の経過で得られた音声分析および合成の手法は、音声科学ならびに実用的な音声情報処理技術の進歩に役立つものと信ずる。