研究概要報告書

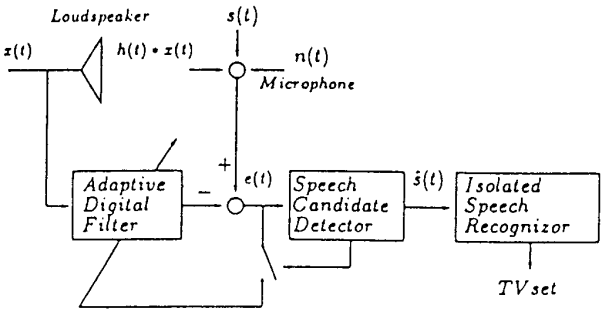
資料－９

研究題目	マスキングモデルを利用した騒音中における音声認識システムの開発	報告書作成者	宇佐川　毅
研究従事者	宇佐川　毅		
研究目的	マン・マシン・インターフェスにおいて音声の持つ優位性としては、補助器具無しでの遠隔操作が可能である点、特殊な訓練が不用である点などがあげられ、現在では離散単語の音声認識のための基本技術はおおむね完成している。にも関わらず広く普及しない原因の第一に、従来技術における耐騒音性がいまだ不十分であることが指摘される。音声認識における騒音の問題は、各方面からの研究が進められているが、音声認識装置のフロントエンドとして提案されてるものの多くは有色性雑音や非定常雑音に対して十分な性能を発揮できていないのが現状である。申請者は、従来から主たる騒音源がある程度特定できる状況を想定し、騒音を適応デジタルフィルタによって除去するとともに、マスキング現象を模した聴覚モデルを用いた音声によるリモートコントロールシステムを提案している。このシステムでは、特定話者での単語認識において、白色雑音のみならず、音声などの有色性がつよく非定常な妨害音が存在する場合にも、耐騒音性を２５ｄＢ以上改善できることを明らかにしてきた。また、マスキングモデルを構成する際には、心理物理実験に基づくデータが必要となるが、詳細なモデルを構成する際に必要となる網羅的な実験はなく、おのずとモデルの精度には限界がある。このため、心理物理データに依存せず、生理学的な知見に基づく基底膜上の数値モデルを用いたモデル（ディジタル蝸牛モデル）を用いた音声パラメータの抽出手法を比較検討する。本研究は騒音の影響を受けにくい音声パラメータの抽出手法を開発することにより、実環境下で利用可能な音声認識システムを開発することを目的とし、実環境に近い騒音環境下での音声認識実験を行うことにより、提案の手法の有効性を検討する。提案する音声認識システムは、種々の種類の騒音への対応が可能であり、音響機器や医療機器などの音声によるリモートコントロール・工業用ロボットの音声による制御など、従来騒音が問題となっていた環境下での音声認識装置の応用が可能となる。		

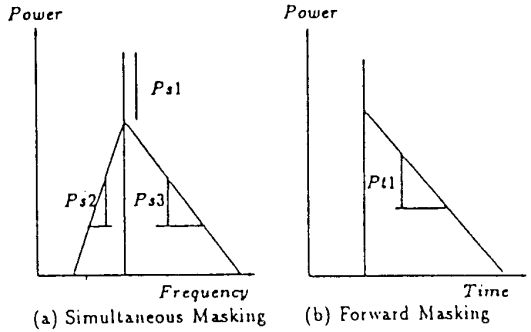
様式－９

研究概要報告書

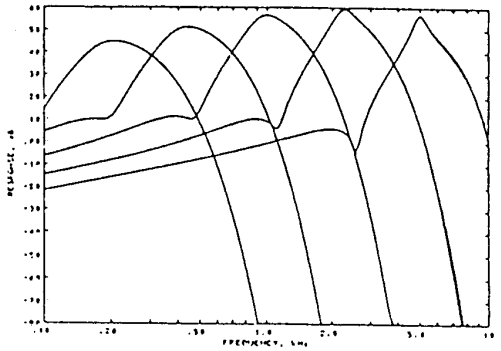
研究内容	実環境下で動作する音声によるリモートコントロールシステムを実現するには、周囲騒音の影響を無視することはできない。騒音中の音声認識システムを構築するには、信号処理技術に基づく騒音の抑制を音声に抽出という方法も検討されているが、音声認識段階、特に音声の特徴パラメータの抽出段階での騒音対策が有効であると考えられる。このような立場から、本研究では騒音の影響を受けにくい音声パラメータの抽出手法として、マスキングモデルを用いた手法とディジタル蝸牛モデルを用いた手法を比較検討し、より実用的な音声認識システムの開発を目指した。従来検討してきたマスキングモデルを用いた場合にも、テレビの音声リモコンを想定した実験では、雑音として白色雑音のみならず楽器音や音声であっても、認識対象である音声信号のレベルに対して、テレビの音声信号（放送音）のレベルが１５ｄＢ程度まで高くとも、即ちＳＮＲが－１５ｄＢ以下でも、十分な認識率が得られていた。ただ、このマスキングモデルを一般化し改良するためには、広範な心理物理実験による実験結果が必要であると考えられる。しから現実には、そのような心理実験を短期間に終えることは極めて困難である。このため、マスキング現象が、主として基底膜上で生じているものと捉え、J.M.Kates（１９９１）提案する生理学的知見に基づいたディジタル蝸牛フィルタを用いたモデルを構成し、これによりマスキング現象を模擬した。これら二つのモデル、即ちマスキングモデルおよびディジタル蝸牛モデルを、音声認識システムにおける雑音抑制機能として利用した場合の性能について比較検討した。その結果、マスキングモデルにより得られた耐騒音性と、ほぼ同程度の性能がディジタル蝸牛モデルによっても得られることが明らかとなった。即ち、テレビの音声リモートコントロールシステムを構成した場合、ＳＮＲ＝－１５ｄＢまで特定話者での単語認識率で９０％以上の性能が得られている。ここで、具体的な装置をＤＳＰなどを用いて構成することを想定した時、マスキングモデルはＦＦＴを用い入力信号を周波数領域に変換した後、マスキングによるスペクトルの変形を周波数毎に順次行っていたため、処理の並列化や分割が比較的困難と考えられる。一方、ディジタル蝸牛フィルターは、時間領域でのフィルタ処理であり、基底膜上の出力信号は各フィルタの出力として捉えられる。このため、基本となる処理の大半は並列化することもできと考えられる。



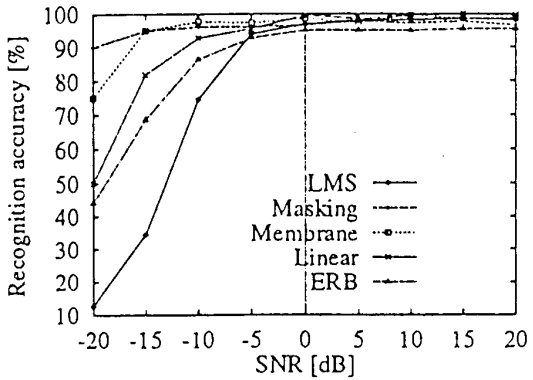
システムテムの全体構成（マスキングモデルまたは、ディジタル蝸牛モデルは、単語認識部に含まれる）



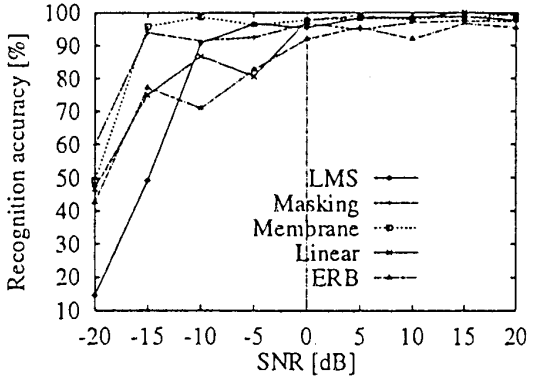
マスキングモデル



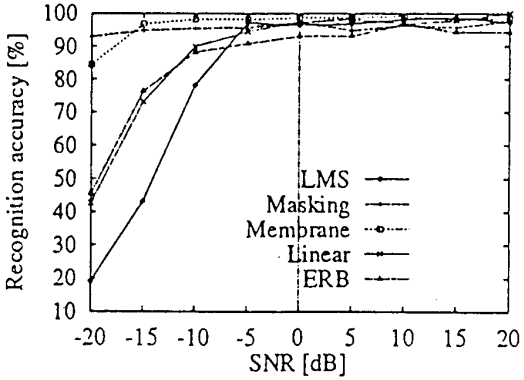
ディジタル蝸牛フィルター



音声認識結果（白色雑音の場合）



音声認識結果（男性音声が雑音の場合）



音声認識結果（オルガンが雑音の場合）

（注：　フローチャート図、ブロック図、構成図、写真、データ表、グラフ等　研究内容の補足説明にご使用ください。）

様式－１０