

音声スペクトルのローカルピークの音声認識への寄与に関する研究

著者	古賀 秀昭
号	1112
発行年	1989
URL	http://hdl.handle.net/10097/12061

氏 名	古 賀 秀 昭
授 与 学 位	工 学 博 士
学位授与年月日	平成 2 年 2 月 14 日
学位授与の根拠法規	学位規則第 5 条第 2 項
最 終 学 歴	昭 和 40 年 3 月 東京電機大学工学部電気通信工学科卒業
学 位 論 文 題 目	音声スペクトルのローカルピークの音声認識への 寄与に関する研究
論 文 審 査 委 員	東北大学教授 城戸 健一 東北大学教授 木村 正行 東北大学教授 曾根 敏夫 東北大学助教授 牧野 正三

論 文 内 容 要 旨

第 1 章 序 論

人間同士の意思疎通の手段として音声は、文字に比べれば非常に負担が少なく、容易なものである。文字以外の押しボタンやキーボードによる意思疎通も音声に比べれば手間と時間のかかり方は大きく、また、それらの操作には、そのための熟練または知識を必要とする。音声による情報入力の利点を列举すると次のようになる。

- 1) 情報の入力速度が速い。文字を書くよりも音声の方が 2 倍以上速い。
- 2) 作業の操作性を増す。手足、眼、耳などを同時に使って別の作業をしながら並列的に情報の入力ができる。
- 3) 電話機などが入力端末に使えるので、経済的であり、また、遠隔地から入力できる。

多くの機械、たとえば、ワードプロセッサへの入力、銀行などでの現金の出し入れ、新幹線の切符の購入、ロボットによる各種の操作、身体障害者の自動車運転、通訳電話などが音声で可能になれば随分と便利になる。このように音声認識の研究は人類の福祉の向上に貢献することができる。

音声認識の研究は 1950～1960 年代に始められた。以来、デジタル処理技術の進展によって線形予測法、PARCOR 分析法などで音声分析の精度は向上したが、音素認識の精度は不十分であった。その後は認識単位は単語となり、動的計画法 (DP) によって単語音声確認精度は向上した。しかし、単語数を多くすると記憶容量や処理量が多くなって実用的でない。そこで、最近では音素を認識単位とする方法に変わりつつある。

人間は音声信号に少々雑音があっても、また、悪い周波数特性の伝達系を通して母音はかなりよく認識できる。このことから、母音の認識にとっては雑音にマスクされにくいスペクトルのピークの周波数すなわちローカルピークが重要と考えられる。従来、ローカルピークの抽出には帯域通過フィルタ (BPF) が使われているが、BPF ではその特性から周波数分解能やピーク検出性能に問題があるので、本研究では FFT による分析法を用いることにした。また、母音の認識にとってローカルピークが重要であることを聴取実験と機械認識とで比較検討した。

第 2 章 ローカルピークの抽出と母音認識

音声資料として単母音と単語中母音を用いることにした。単母音のデータ収集の目的で音響データ収録システムを作製した。このシステムでは AD 変換器のサンプリング周波数は 30kHz 程度まで可能であり、24kHz サンプリングでは 5.46 秒間のデータを収集できる。また、このシステムは FFT の機能とデータをホストコンピュータに転送する機能も持っている。

ローカルピークの抽出を BPF 法と FFT 法とで行い、両者を比較した。

単母音による open 実験では、FFT 法による認識結果は BPF 法による認識結果より認識率が 0.7 % 向上し、単語中母音では同じく open 実験で男声の場合に最大 4.2 %、女声の場合に最大 12.5 % 向上した。これは、BPF 法に比較して FFT 法が不要のローカルピークを抽出することが少なく、また、ローカルピークの欠落が少ないためと考えられる。

第 3 章 スペクトルへの A 特性付加の効果

音声信号から抽出したローカルピークには、認識にとって不要なローカルピークが低周波に存在することがあり、認識に悪影響を及ぼす。これに対し、音声信号のスペクトルに騒音計の A 特性を付加すると不要のピークが抑制され、認識率の向上が図れると考えられるので、その効果を検討した。単母音では A 特性を付加すると open 実験の場合、認識率が 2.7 % 向上し、単語中母音では同じく open 実験で男声の場合に 2.6 %、女声の場合に 2.5 % 向上した。

音声収録に際して A 特性を持つ回路に音声信号を通すと聞きやすくなるということがある。そこで、低周波成分の大きい自動車車内騒音の重畳した音声信号にこの A 特性を通すと騒音が減少し、音声の認識率が向上することが期待できるので、それを検討した。その結果、100km/h で走行中の自動車の場合、C 特性では SN 比は -10.8dB であるが、A 特性を付加すると、SN 比が 14.6dB、A 特性をもう一度付加すると SN 比は 19.1dB となって、SN 比が大きく改善された。さらに、車内騒音の重畳した単母音について認識実験 (標準パターンは騒音のない単母音を作成、close 実験) を行った。その結果、100km/h の速度では、A 特性では C 特性に比べて認識率が 31% 改善され、A 特性 2 回付加では 38% 改善され、騒音が無いときの認識率と同程度になり、A 特性の付加は非常に有効であった。

第 4 章 認識実験結果と聴取実験結果との比較

ローカルピークの周波数は人間の母音認識の手がかりとして重要と考えられる。そこで、単語中

の母音 500 個について男女性10名による聴取実験を行うとともに、同じ音声資料を用いてローカルピークによる機械認識と、一般に行われている LPC ケプストラム係数による機械認識実験を行い、これら 3 者を比較検討した。検討は 4 種類の方法で行った。

- ①聴取実験結果の混同表と機械認識結果の混同表の相関関係による比較では、ローカルピークによる認識の方が LPC ケプストラム係数による認識より聴取実験に近い。
- ②機械認識実験で正認識したものでかつ聴取実験でも 10 人中 8 人以上が正答した数を比較したところ、女声ではローカルピークによる認識が LPC ケプストラム係数による認識より聴取実験に近い。
- ③機械認識実験で誤認識したものでかつ聴取実験でも 10 人中 8 人以上が誤答した数の比較では、両者に差は無い。
- ④聴取実験結果と機械認識の際の事後確率との相関関係を見ると、女声についてローカルピークによる認識の方が LPC ケプストラム係数による認識より聴取実験に近い。

以上の結果より、全体的にはローカルピークによる認識が LPC ケプストラム係数による認識より、聴取実験に近いという結論が得られた。

第 5 章 単語中母音の認識率の改善

ローカルピークは人間の母音認識にとって重要な手がかりとなっているとは言えるが、ローカルピークだけによる認識では、それを音声認識システムに組み込むほどの認識率にはならない。そこでまず、ローカルピークだけでどの位まで認識できる可能性があるかという最大可能認識率を求めた。その結果、男声は 97.9%、女声は 94.5% となり、相当高率の認識ができる可能性があることが分かった。

そこで、認識率を向上させる方法として、大きく 2 種類を検討した。一つはローカルピークだけを用いて、認識アルゴリズムの検討で認識率を向上させること、二つはローカルピークに補助情報を付加することである。

1. 認識アルゴリズムの検討

①標準パターン作成の検討

標準パターン作成用のデータで、まず、close 実験を行い、誤認識したデータを標準パターンの中から差し引いて新たな標準パターンとする。これは、誤認識されたデータは調音結合や雑音の影響を受けているので別の母音とみなすべきであると考えたことによる。この結果、認識率が open 実験では男声の場合 1.5%、女声の場合 0.8% 向上した。

②ピッチ周波数による性差の利用

男女声を混合して認識すると、ホルマント周波数の違いによって、男声、女声に分けた場合より認識率が低下する。そこで、あらかじめ標準パターンは男声用と女声用の 2 種類を作成しておき、入力音声信号のピッチ周波数によって、標準パターンを切り替えて認識を行った。その結果、認識率が open 実験で 3.1% 向上した。

③ 3 フレームの認識結果の多数決の利用

従来は音声データ 3 フレームのうち、標準パターンを 3 フレームで作成し、認識は中央フレームに

対して行ったが、ここでは3フレームとも認識して、その多数決によって認識を行った。その結果、認識率が open 実験では、男声の場合は1.8%、女声の場合は1.3%向上した。

これら3方法は独立であるので、併用することによって、さらに認識率の改善が期待できる。

2. 補助情報の利用

①スペクトルのパワーの二乗平均値の利用

ローカルピークだけで認識した後、その音声データのパワーの二乗平均値を各チャンネルごとに求め、あらかじめ設定しておいた、スレッシュホールドと比較して、認識結果を修正する方法を開発した。この結果、認識率が男声の open 実験で1.8%向上した。

②ローカルピークの回りのスペクトルの利用

ローカルピークはパワースペクトルの最小二乗近似直線より上のピークのみを“1”にして、他は“0”にする方法であるが、ローカルピークの広がりにも音韻性があると考えて、最小二乗近似直線より上のスペクトルは全部“1”にして、最小二乗近似直線より下は全部“0”にする方法で認識した。この結果、認識率が open 実験で男声の場合に3.5%、女声の場合に2.3%向上した。

③スペクトルの傾斜の利用

ローカルピークとスペクトルの傾斜を併用した。傾斜はスペクトルの全帯域、低域、高域の3種類とした。この結果、認識率の向上が open 実験で男声の場合に8.6%、女声の場合に7.3%と著しかった。

④スペクトルの傾斜の比の利用

ローカルピークは伝達系の周波数特性の変化に強い特徴を持つが、スペクトルの傾斜はそれには弱い。これに比べて、傾斜の比（低域/全帯域、高域/全帯域）は強いと考えられるので、これをローカルピークと併用して、伝達周波数特性を変化させて認識実験を行った。ローカルピークだけの場合は周波数特性の変化の影響をわずかしか受けませんが、傾斜との併用ではその影響が著しい。しかし、傾斜の比との併用では、周波数特性の変化の影響はほとんど無く、傾斜の比が周波数特性の変化に強い特徴量であることが明らかになった。

この4種類の方法のなかでは、伝達系の周波数特性の変化を考慮すればローカルピークとスペクトルの傾斜の比の併用が最も有効であった。

第6章 結 論

1. 本論文では音声認識の特徴量として音声スペクトルのローカルピークの周波数（ローカルピーク）を用いている。ローカルピークの抽出のための周波数分析に、従来はBPFが使用されているが、ここではFFTを用いて、両者を比較した。単母音と単語中母音について認識実験を行い比較したところ、FFT法がBPF法より優れていることが明らかになった。

2. 音声のスペクトルに騒音計のA特性を付加して認識実験を行ったところ、認識率が向上した。また、自動車車内騒音にA特性を付加すると騒音が低減し、車内騒音下单母音のSN比が著しく改善された。さらに、車内騒音下单母音にA特性を付加すると認識率は騒音の無いときと同じ程度に改善された。

3. 単語中母音について聴取実験と機械認識（特徴量はローカルピークと LPC ケプストラム係数）を行い，3 者を比較検討した結果，ローカルピークによる認識が聴取実験に近いことが分かった。

4. ローカルピークだけでは認識率が十分でないので，認識率の向上を図った。認識アルゴリズムの検討ではピッチ周波数の利用が最も効果が大きく，補助情報の利用では，伝達系の周波数特性の変化を考慮に入れば，スペクトルの傾斜の比が最も有効であった。

今後さらに検討すべき問題としては時間軸上の情報の利用と言語学的情報の利用が残っている。

論文審査の要旨

情報化社会の進展に伴い、人間にとって最も容易な情報入出力手段である音声を、人間-機械間に利用することが望まれている。これを受けて音声認識の研究が活発に行われているが、まだ多くの問題が残されている。特に音声認識に用いられている特徴量には人間の聴覚特性を考慮したものが少なく、音声認識の高精度化のためには、それを考慮した特徴量の研究が必要である。著者は、聴覚的特徴量であるスペクトルのローカルピークに着目し、その音声認識への寄与についての研究を行った。本論文はその成果をまとめたもので、全編6章より成る。

第1章は序論である。

第2章では、スペクトルのローカルピークの抽出方法について検討し、新しく提案した高速フーリエ変換を利用した抽出法が、従来のバンドパスフィルタを利用した方法より優れていることを認識実験により示している。

第3章では、小さな音に対する聴覚のラウドネス特性に基づく騒音計のA特性をスペクトルに付加した上でローカルピークを抽出することにより、母音認識率が改善されることを述べている。特に自動車内騒音中での母音の認識では、A特性付加の効果は大きく、A特性を2回付加することによって、無騒音時の認識率と同程度の認識率が得られている。これは、実用上重要な成果である。

第4章では、聴取実験による人間の誤り傾向を、スペクトルのローカルピークによる認識実験の誤り傾向、および他の特徴量による認識実験の誤り傾向と比較検討している。4種類の方法で検討した結果、スペクトルのローカルピークによる認識が、一般に行われているLPCケプストラム係数による認識よりも、聴取実験結果に近いという結論を得ている。これは非常に興味ある結果である。

第5章では、単語中母音の認識率改善のために、3種類の認識アルゴリズムと4種類の補助情報の利用について検討している。その結果、伝達系の周波数特性の変化が起り得る環境条件では、スペクトルの傾斜の比をローカルピークと併用するのが最も有効であることを示している。

第6章は結論である。

以上要するに本論文は、聴覚的特徴パラメータであるスペクトルのローカルピークの母音認識への寄与について研究し、人間の聴覚による認識や他の方法による自動認識と比較検討することにより、種々の新しい知見を得たもので、情報工学並びに音響工学の発展に寄与するところが少なくない。

よって、本論文は工学博士の学位論文として合格と認める。