

スペクトラムパターンを利用した 聴覚障害者コミュニケーション支援の研究

酒井 啓† 太田 昌孝† 前野年紀†

†日本 IBM サービス (株)

†東京工業大学 総合情報処理センター

聴覚障害者のうち視覚障害を持たないろう者は、他者とのコミュニケーションを手話、筆談、口話(読唇)などの視覚的手段に頼っているが、音声による通常の会話に比べてさまざまな制約がある。

そこで、音声入力データを解析して得られるスペクトラムパターンや他の特徴量をろう者に視覚的に提供するシステム VAD(Visual Aid for the Deaf) を試作した。

各音素(子音、母音)はほぼ視覚的に弁別可能である。今後はシステムの改良を図っていくが、利用者の訓練次第では相手の発話内容を実時間で認識できるようになることが期待できる。

Communication Support System for the Deaf using Sound Spectrum Pattern

Akira Sakai†, Masataka Ohta†, Toshinori Maeno†

†IBM Japan Service Business Company Ltd.

Shimo-Tsuruma 1623-14, Yamato-shi, Kanagawa 242, Japan

†Computer Center, Tokyo Institute of Technology

Oh-okayama 2-12-1, Meguro-ku, Tokyo 152, Japan

The deaf, who isn't sight impaired, uses various visual communication methods (sign languages, writing, lip reading etc.). But, compared to usual verbal communication, conventional visual communication is restricted in various ways.

We developed VAD (Visual Aid for the Deaf) which, in real-time, converts characteristics of voice such as spectrum pattern into visible form.

With VAD, one can visible identify most of the phonemes. We will continue to improve the system, hoping that, with some training, the deaf can visibly recognize speech.

1 はじめに

現在、聴覚障害者のうち視覚障害を持たないろう者(以下、ろう者と表現する)は、手話、筆談、口話(読唇)、FAX、パソコン通信などの視覚的手段により、他者とのコミュニケーションを行なっている。しかし、音声による通常の会話に比べ、これらの手段は次のような問題点を抱えている。

- 手話：相手側も手話を習得している必要がある。
- 筆談：会話の速度が遅い。
- 口話(読唇)：ろう者側が相手の話し方に慣れる時間が必要である。また、完全に理解できるわけではない(聞き漏らしがある)。
- FAX：相手側にも FAX が必要である。実時間性に欠ける。
- パソコン通信：相手側にもパソコン、モデム等が必要である。

本研究の目的は、音声の周波数分布を解析し、時間を追って図にしたスペクトラムパターンやその特徴量をろう者側に直接提供することにより、こうした視覚的コミュニケーションを補完、援助しようというものである。

スペクトラムパターンを分析し、発話内容や話者の特定を行なうことは、これまでも音声入力インタフェースの研究などで行なわれている。特に、実時間でのスペクトラムパターンの解析は、最近、ワークステーションの力を借りて、誰でも容易に行なえるようになってきた[1]。

例えば、44.1KHz でサンプリングされた音声データに対して512点の高速フーリエ変換を行ない、スペクトラムパターンを実時間で生成しようとする場合、実効で2MFLOPSの浮

動小数点演算能力が必要となる。しかし、CPUの速度が年々2倍で向上している昨今、最新のワークステーションやPCのCPUの能力は、それを行なうに十分なものになってきている。

今回紹介するVAD(Visual Aid for the Deaf)も、現有の音声入出力機能を持つ一般的なワークステーションSUN SS10上に試作されたものである。

音声認識の分野では、スペクトラムパターンの特徴は各種とらえられており、その特徴量の抽出自体は難しくない[2][4]。ただ、この特徴量をコンピュータで弁別してパターン認識する場合には、話者による違いや、人間には自明な文脈に関する知識の欠如が認識率を下げる原因となっているのである。

そこで、本研究の基本方針として、パターン認識技術に頼らず、特徴量自体を直接人間の視覚に提供し、話者や文脈にかかわる判断は人間に任せることにした。将来的には、この機能を実装した携帯用の電子機器を持たせることにより、ろう者にも通常の音声会話の能力を持たせることを目指す。

2 聴覚障害者支援システムの現状

デスクトップ型パーソナルコンピュータによるろう者用の訓練装置は、国内外ですでに商品として実用化されており、例えば、次のようなものがある。

- Speech Viewer(IBM)
- スピーチトレーナー(富士通)

しかし、それらは、音声の存在や違い(音の大小や高低)を認識させることや、後処理としてスペクトラムパターンを表示することによる発話の訓練にとどまっている。

また、Voice-Aid[3]のような、プッシュボタンと音声合成を組み合わせた会話補助システ

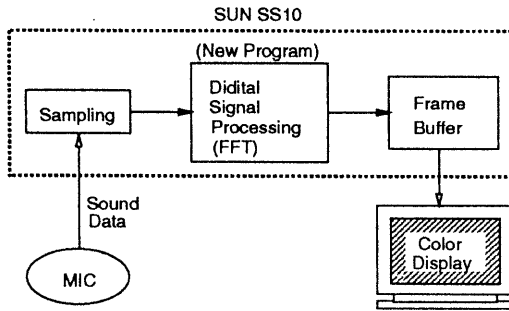


図 1: VAD の概要

ムも研究されている。これはプッシュボタン入力(トーン)の組み合わせで入力されたメッセージを、音声合成によって相手に送るというものであるが、ろう者が使用する場合は、メッセージを送ることはできても相手のメッセージを受け取る手段がない。

現在、明瞭に発話できるろう者も多くなっているが、家族や親しい知人のように慣れている人が相手にできれば、話を通じないという人もまだ多い。そうした(発話の不明瞭な)人の場合、この会話補助システムと今回の試作システムを組み合わせることで、電話を利用した双方向コミュニケーションが可能になると思われる。

3 VAD(第1版)

3.1 VAD の概要

VAD システムは、マイクから入力した音声を高速フーリエ変換(FFT)し、その結果として生成されるスペクトラムパターンを、カラーディスプレイ上に実時間で表示するものである。システムの概要を図1に示す。

音声入力のエンコードタイプは各種用意されているが、今回は16bit PCMを用いた。この場合、サンプリングレートとして8KHz,16KHz,11.025KHz,22.05KHz,44.1KHz 等が可能であ

る。サンプリング時のパラメータとしては、他にゲイン、チャンネル数、入力精度等を指定できる。

3.2 スペクトラムパターンの表示

スペクトラムパターンは、音声スペクトルの時間的な変化を、濃淡図形によって目に見える形で表示したものであり、周波数成分の強度がその度合いに応じて濃淡で表示される。これを生成・表示するプログラムを作成した。

3.3 評価

作成したプログラムでテストしたところ、母音の音素(a,i,u,e,o)の弁別、および子音のある程度の弁別は可能であった(写真1)。しかし、

- 母音に比べて子音の表示時間が短くわかりにくい。
- 子音に同じように見える物があって区別しにくい。
- 入力音声が大きすぎると表示が見にくくなる。

等の問題があった。

一方、このシステムでは自分や他人の声を直接見ることができるため、Speech Viewer やスピーチトレーナーと同様、聴覚障害者の発話訓練用としても有用なものに思われた。

例えば、筆者(酒井)はろう者であり、発音も不明瞭であるが、「い」と「え」の発音がほとんど同じになってしまうことが表示パターンをみただけで把握できた。このシステムを使って、正常な発音の表示パターンを見ながら、自分の発音をそれに近づけていくように訓練すれば、発音の矯正に十分役立っただろう。

3.4 口話との比較

VADは、入力音声の周波数分布を解析して得たスペクトラムパターンをろう者が直接見て、相手の発話内容を理解しようとするものであり、「人間がパターン認識を行なう」という点で、ろう者のコミュニケーション手段の一つである口話(読唇)と類似していると考えられる。そこで、口話の特徴について述べる。

口のパターンを読むというのは直感力に左右されると思われるが、実際、ろう者にも口話に対する得手、不得手はあり、誰でも口話ができるわけではない。口話のできるろう者はおおむね、ろう学校で早期に口話教育を受け、日常的に(学校や職場で)健聴者の中でコミュニケーションを行なう機会が多い。つまり、必要に迫られて日常的に使っているのである。

そして、この口話には、以下のような特徴がある。

1. 話者によって話し方(口のパターン)が異なるため、慣れるまで時間がかかるが、慣れたら認識率は向上する。口の動きのはっきりしている人は分かりやすい。
2. 異なる単語で口のパターンが同じに見えるものがある。通常、前後の文脈から判断するが、言い直してもらるか、筆談による補助が必要な場合もある。
3. 見たことのない(口の動きの)パターンは分からないことが多い。専門用語、業界用語など。これはろう者側の語彙力に左右される。
4. 残存聴力のある人の場合、口が見えないと発話内容が理解できない人でも、補聴器による補助は上記の特徴に対して有効であり、認識率を大きく向上させる。
5. 口話の習得にかかる時間は、個人差もあるが、大人の場合は数カ月かかる。個人

の適性(直観力、推理力、想像力)がものをいう。

そして、VADで表示されるスペクトラムパターンを読んで発話内容を理解しようとする場合も、口話との類似性から、

- 相手側のスペクトラムパターンに慣れるにつれて、認識率が向上する。
- スペクトラムパターンの特徴をもっとはっきりわかるように表示できれば、認識率が向上する。
- 訓練(慣れるまで)には、ろう者側の適性もあるが、かなりの時間(数カ月?)がかかる。
- 残存聴力のある人の場合、補聴器の補助が有効である。

等の特徴が生じるものと考えられる。

4 VAD(第2版)-表示の工夫

4.1 表示の工夫

次に、第1版の問題点を解決するため、プログラムに以下のような改良を加えた。

1. FFTのとき、入力音声データを重複してFFTを行なうようにして、子音の表示時間を十分に得る(写真2)。
2. 表示データに対して周波数補正を行ない、高音部を強調するようにして、子音が明確に表示されるようにする。例えば、 $1+f+4f^2$ といった関数をかける(写真3)。
3. 表示データに対して対数変換を行なうようにして、入力音声の大き過ぎる場合でも弁別可能な表示を得るようにする(写真4)。

4.2 評価

この改良されたプログラムでテストしたところ、問題点はほぼ解消し、子音もかなり明確に弁別できるようになった。

5 VAD(第3版)-他の特徴量の併用

5.1 ケプストラムの併用

ケプストラムは、音声波形の短時間振幅スペクトルの対数の逆フーリエ変換として定義され、スペクトル包絡と微細構造を近似的に分離して抽出できる特徴がある[4]。微細構造は母音に見られる比較的細かい周期のパターン、スペクトル包絡は周波数による変化のゆるやかなパターンである。

ケプストラムは母音と子音の弁別に有効であり、その弁別結果を用いて

子音は緑色、母音は赤色

というように表示色に反映させることによって、表示されるスペクトラムパターンをよりわかりやすいものにした(写真5, 写真6)。

また、ワークステーションやディスクのファン等から発生する音が意外に大きいため、そうしたバックグラウンドノイズの影響を除去するように改良した。

5.2 評価

子音と母音が明確に区別できるようになったことで、子音と母音のまとまりである音節を認識しやすくなった。また、バックグラウンドノイズの影響を除去したことで、表示されるスペクトラムパターンがクリアになった。

6 おわりに

このシステムによって音素の弁別が視覚的に可能になり、表示もある程度満足のいくものになったので、何人かのろう者にシステムを試用してもらうことを考えている。そのことも含めて、今後は、試作システムでの実験を重ね、ろう者の認識しやすいスペクトラムパターンの表示方法を模索していきたい。

将来的に、ろう者が実時間で会話を認識できるようになる可能性は大きいと思われるが、学校での授業、職場での会議、電話での会話、TVやラジオなど、日常的に使用することで慣れていく必要がある。そのために、ワンボードコンピュータを利用した、携帯のためのなるべく小さなスペクトラムパターンの解析・表示装置を開発していきたい。

なお、VADの(SUN SS10で動く)プログラムは公開しているので、興味のある方は筆者まで問い合わせられたい。連絡先は以下のとおりである。

E-mail: vad@necom830.cc.titech.ac.jp

参考文献

- [1] 橋本, 永田, 竹林: "ワークステーションにおける音声認識インタフェースの検討", 情処学会 HI 研究会, HI-46-3(1992)
- [2] 千葉, 上澤, 三宅: "計算機上での音声データ表現の一検討", 情処学会第46回全国大会, 2-245(1993)
- [3] 宮里: "プッシュボタン-仮名文字変換方式とその応用", 情処学会 HI 研究会, HI-27-4(1989)
- [4] 古井: "デジタル音声処理", デジタルテクノロジーシリーズ6, 東海大学出版会

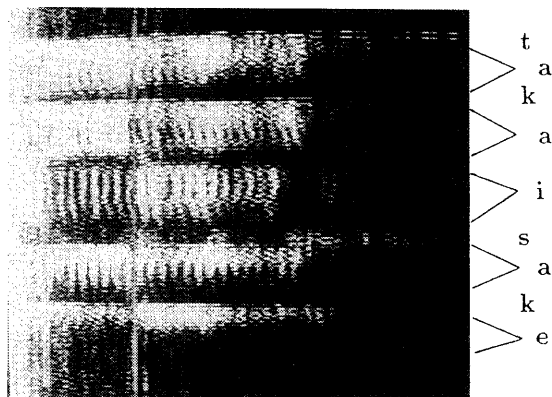


写真1: 第1版

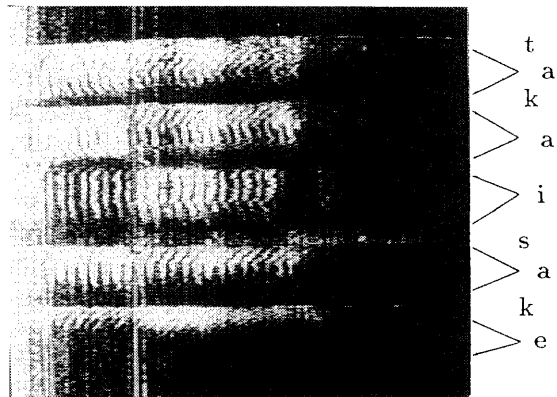


写真2: 第2版-重複してFFT

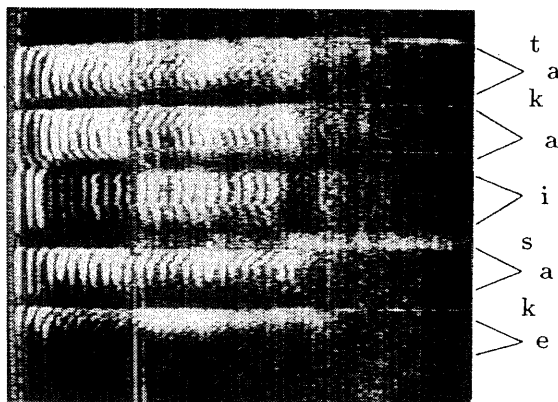


写真3: 第2版-さらに周波数補正

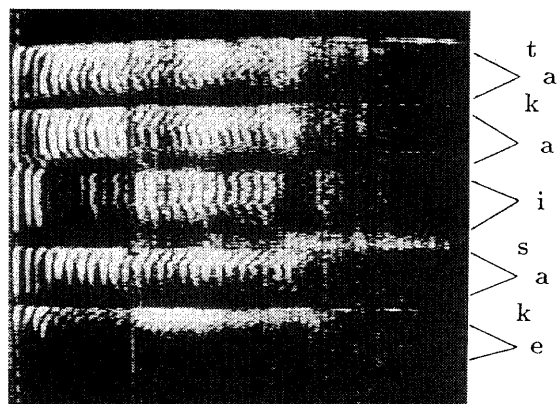


写真4: 第2版-さらに対数変換

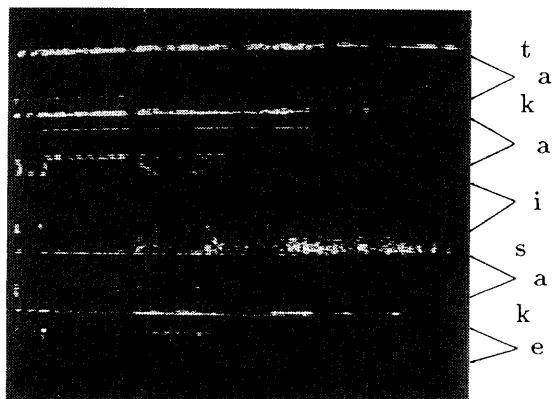


写真5: 第3版-緑色(子音)成分

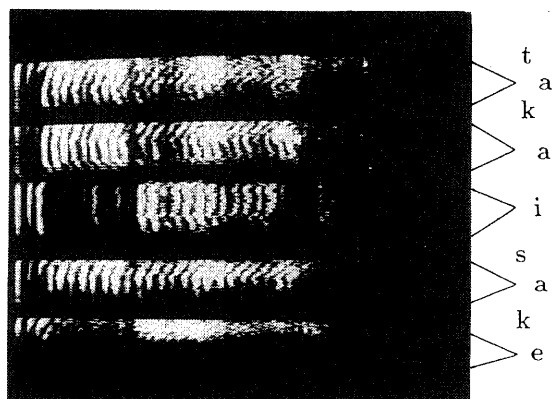


写真6: 第3版-赤色(母音)成分

スペクトラムパターンの表示 (「タカイサケ」)