

聴覚情報補償のための文字・記号等視覚情報挿入形式体系化の研究

電子情報学科情報工学専攻 高橋 秀知

要旨：聴覚障害者のための情報補償としての文字・記号等の視覚情報を使用する方法は、近年発達したマイクロエレクトロニクスの驚異的發展の基に、著しく高速・多機能な文字・記号挿入装置が、廉価に入手・使用可能な段階に成りつつある。従来からの経験的な文字・記号挿入装置のノーファウを生かした、対象映像の聴覚情報を視覚情報として、効果的に文字・記号を挿入する「対映像文字・記号挿入形式の理論化・体系化研究」を行う必要性を痛感する。この観点に立った一考察を論ずる。

キーワード：聴覚情報補償、視覚情報挿入、挿入形式、理論化、体系化

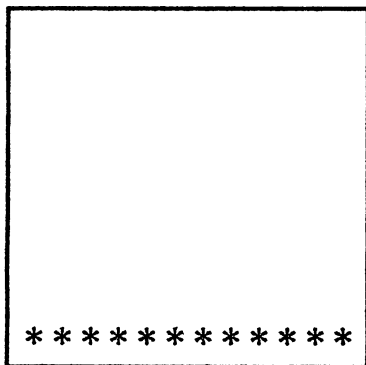
1. 前書き

聴覚障害者とのコミュニケーションまた、音声情報等の補完のため文字等の映像への挿入及び提示の一手段として、映像に対する文字等視覚情報の効果的な挿入形式は、文字等挿入装置の高速化・多様化と挿入技術の進歩を背景として、対象映像への効果的な挿入形式を再検討し、体系化することの必要性を痛感する。この観点に立って、効果的な挿入形式によって得られた「文字等の挿入映像を教育の場で最大限に利用し、効率的かつ効果的な教育を目途とした研究」を考えている一人として、「これまでの経験的な知識を背景とした挿入形式を再検討し、理論的体系化を確立すると同時に、これらを活用した教育方法の確立」を計りたい。本論文は、当該映像の内容・表現に対応した効果的な文字等の挿入形式とそ

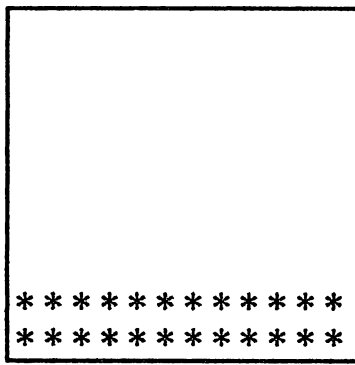
れによる理解度及び視感覚の疲労度との関連を理論的・体系的に研究する第一歩として、「その効果を予測し、その体系化の手掛かりを研究」するものである。

2. これまでの文字等挿入形式の概略

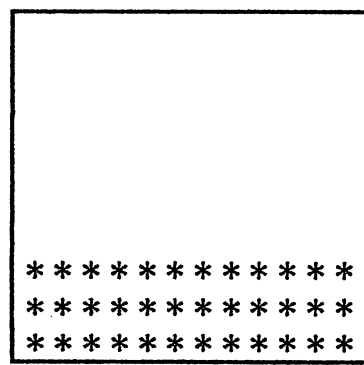
現在日進月歩の勢いで発展しつつあるマイクロエレクトロニクスの技術的向上により、以前には考えられなかった高速・多機能型文字等挿入装置が、比較的廉価に開発される状況と成っている。そこで、これら高速・多機能型文字等挿入装置を利用した「効果的な対映像文字等挿入形式」の理論的体系化への第一歩とする。これまでは、高速・多機能型文字等挿入装置が容易に利用出来ない環境にあったため、「経験的な文字挿入形式」が利用されて来たと思う。



第1図



第2図



第3図

例えば、第1図・第2図のように、画面下部に1行または2行を、16文字～21文字／行程度で文字を挿入する形式が多く採用されていたようである。これに対して、実際の映像は、

- 2.1. 講演会等の映像（特徴：大部分は音声情報）
 - 2.2. 実験・実況等の映像（特徴：音声及び状況情報）
 - 2.3. 劇映画・ドラマ等の映像（特徴：音声情報発信元の明示及び状況情報）
 - 2.4. その他
- これら多種類映像に対して、文字情報の挿入位置及び形式の選択枝は、
- 2.5. 1行の文字数
 - 2.6. 1画面の行数
 - 2.7. 縦学・横書き描写
 - 2.8. 文字の挿入位置・形式
 - 2.9. その他

があり、更に、文字の挿入方式にも、

- 2.10. 既製映像に対する文字挿入
- 2.11. 映像を作成しながらの文字挿入

がある。このような状況下に対して、前述のような画一的な文字挿入形式で、これまでは既製映像に対処してきた感がある。そこで、次章に於いて、各場面の特徴的考察を試み今後の進展を計る。

3. 各場面の特徴的考察

先ず始めに、前章の2.10)と2.11)について言及して置きたい。

3.1. 既製映像に対する文字挿入

通常の文字等の挿入方式で、既製対象映像に文字等を挿入する場がこれに該当する。この場合の特徴は、

- 3.1.1. 何回でも希望箇所に、希望する形式で、文字等の挿入が可能である。
- 3.1.2. 従って、当該映像に対する聴覚情報を可能な限り正確かつ確実に挿入できる。
- 3.1.3. この挿入方式を、仮に「バッチ処理方式」と命名して置く。

これに対して、

3.2. 映像を作成しながらの文字挿入

の方式は、

- 3.2.1. 現在作成中の映像に、「映像作成の進行と同時に文字等の挿入」を要求される。
- 3.2.2. 従って、失敗・誤入力、原則として許容されない。
- 3.2.3. この挿入方式を、仮に「リアルタイム処理方式」と命名して置く。

ここで、バッチ処理方式は、先刻ご承知と思われるの

で、リアルタイム処理方式について若干記述する。

3.3. リアルタイム処理方式の困難的特徴

前述したように、この処理方式は、「映像を作成しながら文字等の挿入を同時進行させる」ところに困難的特徴が存在する。以下、順次その困難性を記述する。その主たる原因は、「現在進行中の製作映像より流れてくる聴覚情報を的確に捕らえ、それを視覚情報に変換提示・挿入する」ことにある。この方式は、通常次の行程を経て行われる。

3.3.1. 音声情報の文字化

音声情報は順次淀みなく流れてくる。この情報を「視覚情報としての文字等情報に変換」する必要がある。音声情報を文字等情報に変換する方式として、現在考えられている方式には、音声認識装置による方式とキーボードよりの入力方式の2方式が一般に考えられている。

3.3.1.1. 音声認識方式

この方式は、特定話者方式と不特定話者方式とに大別される。特定話者方式は、文字通り特定人の音声のみを認識する方式で、その他の人々の音声を認識することは出来ない。一方、不特定話者方式は、不特定多数の人々の音声を認識する方式である。現段階に於いては、不特定話者方式の方は、未だ実用化の段階に至っていないようである。これに対して、特定話者方式は、特定・限定的に実用化されている。

3.3.1.2. キーボード入力方式

音声情報をキーボード・オペレータが聴取し、その情報をキーボードより入力する方式である。この方式の問題点は、オペレータの入力速度にある。我々の通常発声速度は、大略300字／分程度のものであるが、これに追従するためのキータッチ速度は、通常のタッチ入力速度の限界にある。従って、この速度で長時間にわたる入力作業は不可能である。そこで、マルチタッチ・マルチ入力方式が開発された。この方式は、一度に複数のキーをタッチ入力し、複数文字を発生する方式である。この方式の開発により、350字／分の文字入力が可能となった。この速度は、通常のニュース放送の発声速度程度である。このように入力に関しては、ある程度追従可能な方式が開発されたが、次の難関として、「かな・漢字変換」がある。

3.3.2. かな・漢字変換について

日本文は「かな」と「漢字」の混合形態を取っている。周知のように、漢字には「同音異義語」が存在する。「かな・漢字変換」の際に、多数存在する同音異義語の中から、一度で「適切な漢字を選択する」ことは至難の業である。更に、選択された「かな・漢字混合文」を

「表出された音声情報と同速度」の高速で合成・挿入するための「超高速文字等挿入装置」を必要とする。

3.3.3. 高速文字等挿入装置について

幸いにも、この超高速文字等挿入装置は開発されたため、現在の困難点は、「かな・漢字変換」の一点に集約されつつある。

したがって、これからの議論は、「バッチまたはリアルタイム処理方式で、自由に文字等を希望する箇所に、希望する形式で、対応映像に視覚情報が挿入できること」を前提としている。

4. 映像と字幕

次に典型的な映像・字幕の関係について考察を行う。

講演者の音声情報が、重視される画像に於いては、講演者そのものの画像より、音声情報を文字化し、字幕として提示することが必要である。この時、

4.1. 音声の文字化を可能な限り忠実に表現提示する方法

4.2. 対象者の知識に合わせて要約提示する方法の2方法が考えられる。

4.1. については、大量の字幕を、

4.1.1. 何時・何処に、

4.1.2. どの様な形式で、

4.1.3. どの様な速度で、

挿入するかが、効果的な挿入形式としての検討項目となる。

第4図に於いて「いとう」は、画面による使い分けの差であると考えられるが、「いとう」の併用は、対話等

画面上で区別をする必要な映像の場合に有効である。一方、「あとえ」との関係は、それとやや異なる。「あ」が動的イメージを与えるのに対して、「え」は、より静的イメージを与える。従って、被挿入映像の状況によって、使い分ける必要がある。

4.1. について

「縦書きの場合」第5図参照

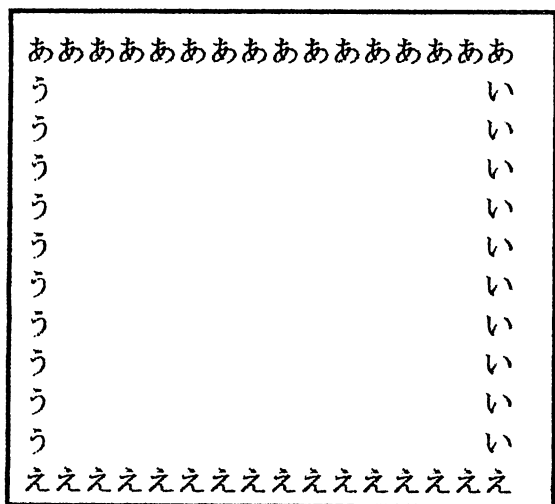
この場合は、日本語本来の形態となるため、横書き形式の場合より、より豊かな情緒性が提供されよう。特に、最近のワープロの様に、行書体の文字ホントを持っている場合には有効性が予測される。

「横書きの場合」第6・7図参照

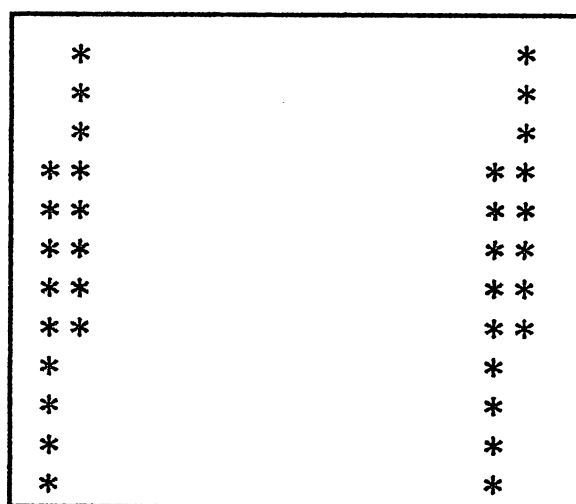
図のように、通常12～21字/行の間で使用されている。特殊な場合を除いて、3行は、多すぎるようである。我々が、文字を読むとき、一つの視野に入っている文字群の漢字を拾い読みをする傾向がある。従って、「ひらがな・カタカナ」だけの文章は読みにくい。7図の場合は、6図の場合の視線が左から右へ、2往復するのに対して、1往復での読み取りが可能となる。但し、いずれの場合にも、左端最下部の文字には、注意力がやや劣るようである。

4.2. について

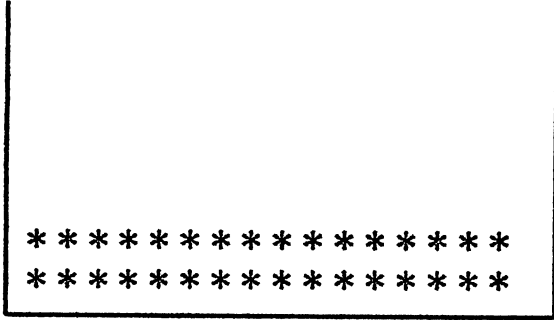
相当量の文字が挿入されるため、提示文字を全部読み取れるか否かは、読み手の能力次第となる。そこで、これを救済する方法として、ビデオテープを読み手の能力に合った速度で再生する方法も考えられる。但し、この場合、映像がノーマル状態でないため、実験状況等のように、画面の動きを重視する場合には利用できない。要



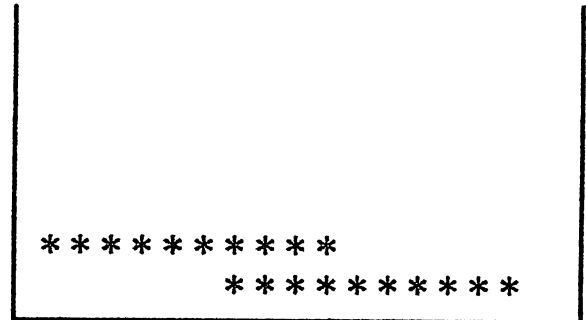
第4図



第5図



第6図



第7図

約の場合には、文字量が少ないため、取扱いは簡単となるが、要約の定義が大変に難しい。要約する人の学識・経験に左右されがちになるからである。要約は、一種の創作と云える。一方、適切な要約は、講演内容をより正確にかつよりの確に提示することが可能となる。

5. 実験等画像内容も重視する文字提示様式

実験の様子や絵画の鑑賞のように、原画像も提供情報として大切な場合には、

5.1. 挿入文字量を減らす方法

5.2. 文字と画像とを切り放す方法

等が考えられる。

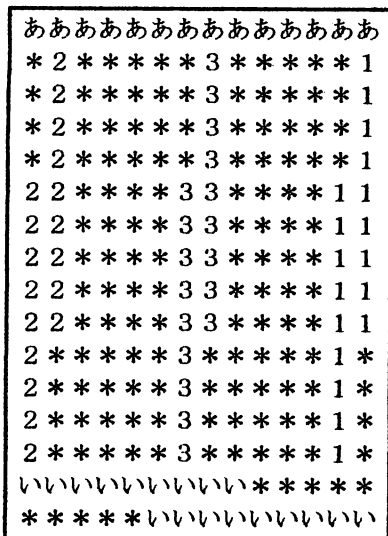
5.1. について 第8図参照

映像の注視に影響しないような位置と量の文字挿入が必要である。従って、基本的には下段に2行としても、映像の様子によっては、上段または縦書きの左右及び中央と様々な形式を、臨機応変に採用する必要がある。

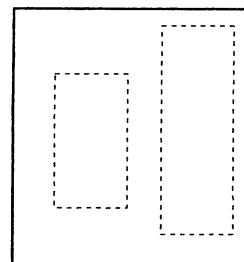
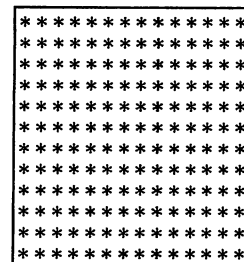
5.2. について 第9・10図参照

説明文字群及び映像とを分離して、説明文をより良く理解すると同時に、映像も注視可能の状況にする。この場合、第9図と第10図の2種類が考えられる。

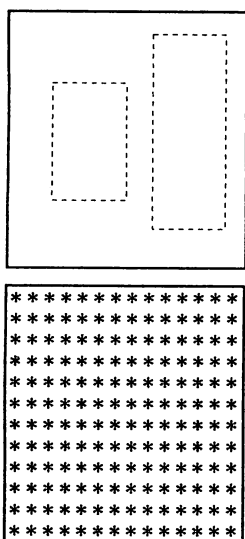
第9図について



第8図



第9図



第10図

説明文によってこれから出てくる映像の予備知識を提供し、その後、実際または実状を、映像によって提供する形式である。これに反して、第10図は、第10図について

始めに、映像を提示するのであるから、知識・概念の確認または、問題の提示等には、より適していると考えられる。

以上のような困難点は、科学技術の進歩と共に暫時解決されて行くことを願っている。この背景の基に、どのような文字・記号の挿入が、「より効果的であるか」を研究し、理論化し体系化することが大切である。

6. 体系化に関連するパラメータ

文字等の効果的挿入に関連するパラメータとして、概念的に考えると、

- 6.1. 挿入文字数・記号数
- 6.2. 挿入文字体系・記号体系
- 6.3. 挿入文字・記号の色彩化の度合い等装飾度合
- 6.4. 挿入形式（縦書き形式／横書き形式）
- 6.5. 挿入方式（静的「パッチ的」方式／動的「実時間的」方式）
- 6.6. その他

が考えられる。以下、体系的考察を行うためのデータ・パラメータについて研究する。

映像内聴覚情報の視覚情報への変換部位は、大略的に、次の2つに分類される。

- 6.7. 台詞と言われている言葉の部分
- 6.8. 音楽等の背景描写の部分

それらには、次の4個のパラメータが考えられる。

6.8.1. 疲労感に対するパラメータ

長時間の聴覚代行としての視覚の使用による疲労度の時間変化

6.8.2. 内容理解に対するパラメータ

視覚情報に変換された音声情報量と表示内容による理解度

6.8.3. 背景映像の認識に対するパラメータ

挿入視覚情報量に対する被挿入映像の情報認識度

6.8.4. 提示時間に対するパラメータ

変換視覚情報の提示時間と教育効果

がある。従って、体系化には、概括的に見ても、これら $5 \times 4 \times 2 = 40$ 個のパラメータから成る「多変数関数」を扱うことになる。

7. 体系化に関する私案

一方、我々が日常聴覚情報を受容している場合、特別の場合を除いて、一般的には、聴覚的にも、視覚的にも疲労を感じてはいない。これと同じ様に、聴覚障害者の視覚情報の受容に対しても、疲労感のない、しかも可能な限り背景映像の情報をも認識出来る形式が望ましい。一般的に横書き形式が考えられているが、次の場合には、縦書き形式の方が効果的と考える。

7.1. 背景描写の効果をも出す場合、特に、草書の様な字体が有効であると予測する。

7.2. また、左右に分けて文字挿入することにより、固有の意味を持たせる場合には有効的と考える。

8. 静的表示方式（字幕方式）と動的表示方式（スクロール方式）

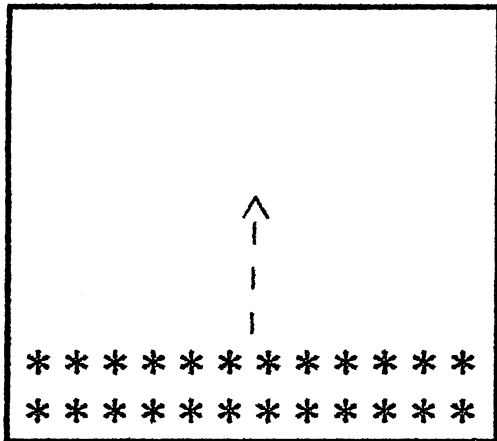
これは難しい問題であるが、一般的に云って、

8.1. 静的表示の場合には、

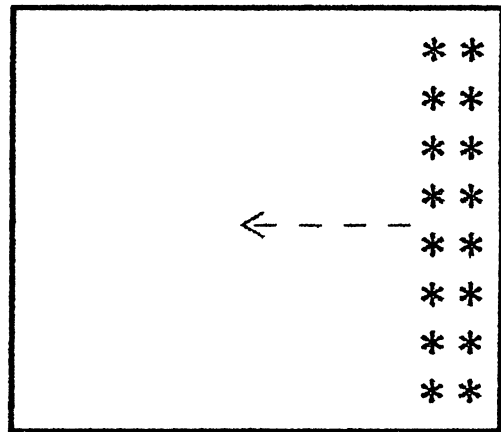
或る意味の、一種のパターン認識が可能のようで、短時間での情報受容によって、各々の文字を読まなくとも、内容の大略が掴めるのではないか？特に、第1図の場合には、その感が深い。第2図では情報量が少ないし、第3図は視覚のジャンプが2回あるため、長時間に亘る連続的な情報取得に対しては、疲労度の増大が予想される。更に、2行目の始めの字は、位置的に他と比べて、注意力が少ないとの報告も有るようで、その辺の考慮も必要と考える。

8.2. 動的表示の場合には、第11・12図参照

背景映像への注意が失われがちである。これは、次々に出てくる文字を読まねばならず、その読み取りに気を取られるためと予測する。我々が、「ナナメヨミ」をする場合いには、全部の字を追わないで、漢字に対する、



第11図



第12図

一種のパターン認識を行っている。従って、静的表示の場合でも、カナ文字だけの場合と漢字の混在率の高い場合とでは、文字数に対する認識率は、異なるものと思われる。この観点に立つと、広報の様に、より注意深く読んでもらうことを必要とする場合には、動的表示の方が良いとも考えられる。この場合には、動く速度を十分に考慮しないと、理解不足や間延び等の現象が生じ易い。

以上の考察から、第13図の様な空間曲線が、体系的考察の第一歩として考えらる。これは、体系化に先だつての経緯を背景にした研究の一例である。

9. 結語

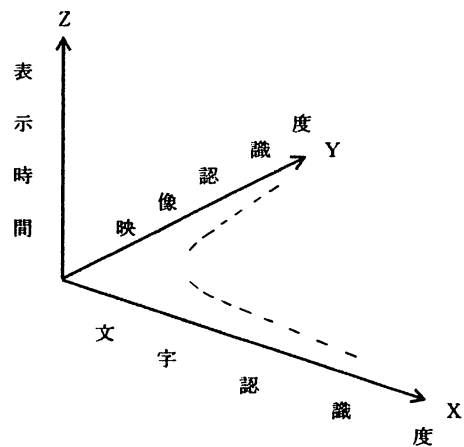
これまでの考察は、字幕挿入一般について行ったものであるが、これらの対象となる映像にも次の4種類が考えられる。

9.1. 絵画・実験の説明の様に、背景映像そのものの観察が重要で、説明は補助手段であり、学術的に正確性を要するもの

9.2. 学術講演の様に、映像より挿入視覚情報が重要で、しかも正確性を要するもの

9.3. ドラマの様に、映像と台詞・説明が等価に重要であるもの

9.4. アニメーションの説明の様に、背景映像そのものからの得られる情報が重要で、説明は、補助手段であり、正確性より、フィーリング的説明が必要であるもの等、それぞれについて、条件の異なるパラメータを必要とするため、大変に複雑と成るが、簡単なものから、だんだんと体系化が出来ればと考えている。



第13図

「参考文献」

1) 西川 俊・高橋 秀知・小林 正幸・石原 保志・柴田 邦博

「聴覚障害者のためのリアルタイム字幕表示システム」
電子情報通信学会論文誌 D-II Vol.J78-D-II No.11
PP.1589-1597(1995)

2) 小林 正幸・西川 俊・石原 保志・高橋 秀知
「聴覚障害者のためのキーボードの連弾入力方式によるリアルタイム字幕提示システム」

映像情報メディア学会誌 Vol.51 No.6 pp.886-895 (1997)

Study of Systematic Captioning of Words and Signs for the Assistance of Voice and Music Information

Information Science and Electronics

Hidechika TAKAHASHI

ABSTRACT

The methods of using words and signs as visual information for the assistance for the hearing impaired has been made easy by the high speed multi-use captioning device according to the microelectronics and then it has become very cheap recently.

And then I have been studying the theoretical and systematic effective captioning style for the video picture by using an experienced captioning technique recently.

keywords: assistance of the hearing impaired / captioning of voice information / captioning style / theoretical and systematics