

## 視覚障害者のための電子図書館 その6 — Text-To-Speech 合成音声による電子録音図書 —

筑波技術短期大学視覚部一般教育等<sup>1)</sup> 同鍼灸学科<sup>2)</sup>

村上佳久<sup>1)</sup> 上田正一<sup>2)</sup>

**要旨**：電子録音図書には大きく分けて二種類ある。1つは、ボランティアなどが実際に声に出して録音した、音読み録音図書を電子化したもの。もう一つは、冊子体の本をOCRなどで電子テキスト化し、合成音声で処理し(読ませて)電子録音図書化したものである。ここでは、即応性に優れた合成音声による電子録音図書について、作成方法や利用方法などについて検討する。

**キーワード**：電子録音図書、合成音声、Text-To-Speech、OCR

### 1. はじめに

Text To Speech (Text-To-Speech) (略号 TTS)：「音声合成」とは、一般に、テキスト文字列から音声を生成する機能のことで、Text-To-Speech は、それぞれの言語でテキストを読み上げるシステムを言う。

日本語の場合は、日本語 Text-To-Speech Engine を用いて処理を行うが、日本語の中に含まれる英語については、Text-To-Speech のシステムによって異なるが、対応する英語の Text-To-Speech Engine が処理する場合もある。

したがって、テキスト文が存在すれば、文章読み上げ機能として簡単に利用できる技術である。視覚障害者が利用する場合は、コンピュータを利用する前提として、画面読み合成音声ソフトウェアなどで、画面の操作を合成音声の補助を受けながら行うので、既に、この機能については馴染みが深いと言える。

この合成音声を利用する場合には、合成音声エンジンの違いにより、読み方などが異なるので留意が必要である。また、専門用語の読み方の辞書などを用意しないと、鍼灸や理学療法などの医学系に利用される用語の読みに大きな間違いが頻発するので注意しなければならない。

また、合成音声で読ませるテキスト文を作成するためには、出版社から提供されるテキスト文を利用するか、OCR 技術により、冊子体の本をテキスト化するしかない。

そこで、本研究では、OCR 技術によるテキスト文作成と、合成音声による電子録音図書の2点について、その作成方法や技術的問題などを検討する。

### 2. OCR によるテキスト文作成

OCR (Optical Character Recognition (光学的文字認識) または、Optical Character Reader (光学式文字読取装

置))を利用すると、全盲などの視覚障害者が読むことのできない本の活字をテキスト化し、合成音声で読ませることにより、本の内容を知ることができる。従来は、ボランティアなどに対面で朗読してもらうことでしか本の内容を知ることができなかったが、コンピュータ技術の発達により、独力で操作可能となり、対面朗読なしに本の内容を把握することが可能となった。

この OCR 機能と合成音声機能を組み合わせたものは、盲人用読書機と呼ばれ、製品化されているが、英語の場合は比較的実用的であるが、日本語の場合は、OCR 機能の部分に問題があり、実用的とは言えない部分が多い。その原因としては、日本語特有の問題が挙げられる。

#### 2. 1 OCR 機能

OCR 機能では、大きく分けて次の2つのソフトウェアエンジンからなる。

画像データから文字を認識するエンジン

文章のレイアウトを認識するエンジン

文字認識のエンジンは、スキャナで光学的に読み取られた画像データを文字コードに変換するための OCR の基本的な部分である。2002年12月現在では、OCR 製品を販売しているメーカーでも技術的に大差はなく、自社原稿では、認識率98~99%程度のもので一般的である。

しかし、問題はレイアウト認識エンジンである。英語に比べて日本語には利用する文字数が非常に多い。また、英語に比べて、文字の利用方向が多彩である。例えば、新聞である。現在の新聞各紙は、右上から左下に向かって、多段の段組で、縦書きで展開している。また、見出しは、大きな活字で、縦書きと横書きの混合である。更に見出し中に縦書きと横書きが混在する

場合もある。

これに対して、最近の一般の文章は、ワープロの影響もあって、左上から右下に向かって横書きで展開される。また、手書きの文章では、右上から左下に向かって縦書きが一般的である。また、文章中に写真や絵がある場合はこれを認識してその部分は文字認識を行わないようにすることも必要である。

## 2. 2 OCR 認識

このように見ていくと、最近の日本語表示の仕方の問題があるように感じるが、複雑な表示方法のため、レイアウト認識には大きな差が生じる。

OCR の一般的な流れを記述すると、

- 1) スキャナから画像データを読み取る
- 2) 画像データを整形する (原稿の向きを修正する)
- 3) 文章のレイアウトを認識し、画像の文字データ部分だけを切り出す
- 4) 画像の文字データ部分を文字認識しテキスト化する
- 5) テキスト化された文章を機械的に自動校正し、文章の間違いを訂正する
- 6) テキスト書式をワープロなどのアプリケーションにあわせて、出力する

高額で、高機能な OCR はこれらの機能全てを有しているが、廉価版などではいくつかの機能が省略される。そこで、いくつかの OCR ソフトで、実際にどの程度の性能を有するかテストを行って検討した。

テスト方法は、3種類である。

- 1) A 4 用紙に1000文字の原稿  
(上下左右20mm 空け、明朝体・ゴシック体)
- 2) 新聞切り出し (A 4)
- 3) テクノレポート (図版入り)

1 は OCR 文字認識エンジンの性能のチェックで、2・3 はレイアウト認識エンジンのチェックである。なお、今回のテストでは、認識後の後処理による文字校正機能は利用しないようにしている。また、視覚障害者の利用を考慮して、画面読み合成音声ソフトで利用できる OCR ソフトウェアを選択した。視覚障害者向けの OCR ソフトでは、今回のテストのメーカーと同じ認識エンジンを有するので認識率はほぼ同等だと思われる。

### 2. 2. 1 テスト1の結果のまとめ

メーカー	明朝体	ゴシック体
S社	99%	99%
R社	99%	99%
M社	99%	99%
N社	99%	99%
A社	98%	99%

(ワープロ A 4 原稿、1000文字、領域指定後認識)

統計処理上、危険率5%でメーカー間の有意差はない。

### 2. 2. 2 テスト2の結果のまとめ

メーカー	レイアウト認識	文字認識
S社	17%	78%
R社	18%	81%
M社	15%	82%
N社	21%	83%
A社	19%	79%

(新聞原稿、自動認識、レイアウトの順で判断)

新聞原稿のレイアウト認識では、認識率の悪さが目立つ結果となった。

1つには、縦書きと横書きが混在したレイアウトであること。

本文原稿は、右上から左下方向で、縦書きで書かれているが、見出しは左から右へ横書きや縦書きの混在であることが主たる要因と思われる。

レイアウト認識後の文字認識では、文章の切り出しが悪いためか80%程度の認識率に留まっている。

### 2. 2. 3 テスト3の結果のまとめ

メーカー	レイアウト認識
S社	80%
R社	85%
M社	75%
N社	80%
A社	75%

(テクノレポート、自動認識、レイアウトの順で判断)

タイトル表示と要約、図と表が混在する2段組原稿では、2段組の原稿の順番と表題などの順番を間違えるケースが多かった。論文原稿なので比較的レイアウト認識は良好といえる。

レイアウト認識では、特に新聞原稿の認識が悪いが、一般的な本や論文、また、国家試験などの問題では、テクノレポート同様の結果が得られるものと考えられ

る。しかし、文字認識の精度に比べて、レイアウト認識の精度が悪いので、そのままでは、合成音声向けの原稿としては利用できないので、人の手で校正が必要である。

### 2. 3 図書のデジタル化

一般的に図書のデジタル化は、以下のような作業手順で行われる。

#### 1) 図書のコピー

冊子体の本を切断せずに見開きでA4かA3にコピーし、中央部分の曲がった部分を光学補正して、全ページをコピーする。ページ送りは自動的に行われ、コピーされる。

冊子体の本を切断できる場合（本が破棄処分できる場合）は、この工程は省かれる。

#### 2) スキャニング

コピーされた図書は、ADF（自動原稿送り装置）付きのスキャナで連続して画像データ化される。場合により両面原稿の自動読み取りも行われる。

#### 3) 画像データのデータベース化

スキャナで読み込まれた画像データは、サーバに蓄積されデータベース化される。この時点で画像データのみをPDF化し、電子図書データとして流用する機会が多い。

#### 4) OCR化

画像データをOCRで連続認識しテキストデータに変換する。日本語補正処理により機械的に校正を行い、人的作業による校正時間を出来るだけ少なくする。多くの大学では、人的校正の時間を行うだけの余裕がないために、このレベルで作業を止め、テキストデータとしている場合が見られる。

#### 5) 校正

人的作業により、原本と照合しながら校正を行う。最も時間がかかる部分である。

#### 6) データベース化

校正（2校以上）されたテキストデータは、データベース化され電子図書として運用される。著作権処理も行われる。

現在、多くの大学などで電子図書館が構築され運用されているが、図書データのデジタル化にはOCR技術は欠かせないものである。しかし、OCRの認識性能から図書のテキスト化は、出来るだけ行わず、画像データによる運用で留めている場合を多く見かける。それは、校正に係る人と時間の問題があまりにも大きいためである。

しかし、視覚部図書館の場合は、テキストデータから、拡大図書や電子点字図書、電子録音図書が作成されるため、テキストデータを作成せざるを得ない。そこで、OCRに関わる問題点を見直し、作業の効率化を検討する。

### 2. 4 レイアウト認識

OCRには前述のようにレイアウト認識に問題がある。自動レイアウト認識後に自動文字認識をさせると、必然的に認識率は低下する。その状態で人的校正を行っても時間がかかり効率が悪い。そこで、レイアウト認識を手動で行い、その後に文字認識を行わせることで、認識率の変化を調べた。

#### 2. 4. 1 2) 原稿、手動レイアウト認識後文字認識

メーカ	文字認識（手動）	文字認識（自動認識）
S社	92%	78%
R社	94%	81%
M社	91%	82%
N社	94%	83%
A社	90%	79%

認識率が飛躍的に向上したが、ワープロ原稿よりも認識率が悪い。これは、縦書きと横書きが混在したレイアウトであることが主たる原因と思われる。

#### 2. 4. 2 3) 原稿、手動レイアウト認識後文字認識

メーカ	文字認識（手動）	文字認識（自動認識）
S社	97%	80%
R社	98%	85%
M社	98%	75%
N社	98%	80%
A社	97%	75%

テクレポ原稿では、手動でレイアウト認識を行うと文字認識は飛躍的に向上した。ワープロ原稿には若干及ばないものの高い文字認識精度を誇っている。さらに、多くのOCRソフトでは、文字認識に自信がないときは幾つかの候補を出して選択するように表示している。

2. 4. 2の場合、これらの疑問候補から正しい文字を選択すれば、認識率はほぼ全てのOCRソフトで99%となり、校正にかかる労力を低減することが出来る。

これらの結果から、手動でレイアウト認識を行い、その後に人的な校正を行えば最も効率的なことが分かる。

## 2. 5 電子図書作成のシステム化

そこで、以下のような作業手順を実験的にシステム化して運用してみた。

### 1) 図書のコピー

自動修正機能付きの見開き本コピー機は非常に高価なので、通常のコピー機でコピーするか、本を切断して利用する。

### 2) スキャニング

両面読み取り機能付きの ADF を利用して、連続的にスキャニングを行い、サーバに名前を付けて保存する。

### 3) OCR その1 (手動レイアウト認識)

サーバの画像データを OCR ソフトに読み込み、手動でレイアウト認識を行う、その後、自動文字認識を行い、OCR の編集形式で認識結果のデータをサーバに保存する。

### 4) OCR その2 (初校)

サーバに保存された OCR 編集形式のデータを読み込み、認識時の候補から文字を選択し、初めの校正(初校)を行い、テキストデータ化し、サーバに保存する。

### 5) 校正

ワープロやエディタ上で、テキストデータを読み込み、校正を行う。(1校)

本来は、その後別の人により、2校を行うのであるが、時間的問題や人的問題からこの時点で、校正を終了する。なお、この時点での認識率は、99.7%程度であるから、A4原稿の1頁あたり3文字程度の誤字が含まれることになる。

このレイアウト認識を手動で行う手順と、前述の自動レイアウト認識を行う手順とで、人的な労力を比較すると、

#### 手動レイアウト認識

マウス操作により手動でレイアウト認識する作業  
認識結果より文字候補を選択する作業(初校)

校正する作業(1校)

#### 自動レイアウト認識

校正する作業

一見すると、自動レイアウト認識の方が作業量は少なく感じるが、認識率の結果より校正する文字数が非常に多いため、手動レイアウト認識の方が係る時間が少ないことが示唆された。

以上の結果から、OCR でテキスト原稿を作成するためには、自動レイアウト認識を手動に切り替えることによって、認識率の向上を見た。

## 3. 合成音声化

電子テキスト化された本を重度視覚障害者である全盲が利用するためには、合成音声による読み上げを利用する。この場合の合成音声の読み上げは、画面読み合成音声ソフトのような OS を視覚障害補償するソフトウェアである必要はない。テキスト文を読み込んで、合成音声出力できるソフトウェアで有ればよい。画面読み合成音声ソフトも2つのエンジンから成り立つが、必要なのは、テキスト文を合成音声で出力できる、合成音声エンジンである。

### 3. 1 問題点

ところで、この合成音声エンジンについて、視覚部では、考慮すべき次のような問題が存在する。

- 1) 外字
- 2) 発音辞書
- 3) アクセント

1つは、外字である。視覚部で、教材として利用する場合、特に鍼灸学科などでは、JIS 文字コード外の所謂「外字」と呼ばれる存在を無視できない。何故なら、教科書や参考書、また国家試験でも外字は利用されているためである。特に東洋医学関係に外字が多く必要不可欠である。そこで、合成音声でも外字をサポートしている必要がある。

2つに、合成音声が読む、辞書の問題である。例えば、「内側」は、一般的には「うちがわ」と発音されるが、解剖学用語では、「ないそく」と発音する。したがって、鍼灸や理学療法の解剖学関係の用語では、「ないそく」と発音する必要がある。

しかし、「内側」を「ないそく」と発音させると、一般的な用例の場合でも「ないそく」と発音してしまう危険性がある。よって、用語に合わせた専門の用語辞書が必要となる。

3つに、音声のアクセントと音質の問題である。人の声に比べて、合成音声の声質は無機的で、長時間の聞き流しに苦痛を感じやすいと言われるが、電子録音図書では、10時間以上の録音物も少なくないので、聞きやすい合成音声が求められる。

現在、合成音声ソフトウェアは、十数社から販売されているが、コールセンター向けなどの組込用応答を除き、一般的に入手可能な物は5社程度である。視覚障害者向けの利用で定評があるのは、画面読み合成音声に利用されている、IBM 社製合成音声エンジンと Ricoh 製合成音声エンジンである。

これらは、合成音声エンジン単体としては、外字に

対応し、ユーザ辞書により発音辞書を登録できる。アクセントもある程度の変更が可能である。

### 3. 2 運用実験

そこで、実際に鍼灸関連の教科書・参考書を利用して、実運用テストを行った。

対象は、鍼灸学科3年生、5名で、強度弱視や手動弁、指数弁である。

テキスト：経絡経穴概論、東洋医学概論、解剖学、生理学（生理学を除いて外字が必要）

発音用語辞書：各図書専用の辞書を作成

アクセント：各々の学生の好みに合わせる

合成音声ソフト:Pro-Talker97(IBM)、雄弁家 V2(Ricoh)

利用メディア：MD (80min)

なお、合成音声出力は、Pro-Talker97は、ソフト側でwav ファイル出力が可能なので、直接 wav 出力し、NetMD 経由で MD に書き込んだ。雄弁家 V2も、ファイル出力が可能であるが、今回は、合成音声出力を別のパソコンで取り込み、そのデータを NetMD を利用して MD に出力した。MD は、LPMD（長時間再生2倍・4倍モード）もあるが、等速で利用した。

また、ボランティアがカセットテープに録音した同一のデータを MD に出力して比較対照とした。

5名の学生の使用後の感想は次のようなものであった。5段階評価で、1が悪い、3が普通、5が良いで評価した。

#### 3. 2. 1 合成音声と人声の比較

被験者	合成音声	人声
あ	3	4
い	3	5
う	4	4
え	2	5
お	3	4
平均	3. 0	4. 4

結果は予想通り、合成音声の方が悪いが、合成音声をお我慢できないとした学生は無く、補助手段として合成音声は有効である事が示唆された。

ボランティアの読まない図書などで、即応性が求められる場合は、5名とも合成音声をサポートとした。また、IBM 製と Ricoh 製での比較を行ったところ、

5段階評価で、1が悪い、5が良いで評価した。

#### 3. 2. 2 合成音声メーカーの比較

被験者	IBM 製	Ricoh 製
あ	3	4
い	4	3
う	4	4
え	3	3
お	4	4
平均	3. 6	3. 6

こちらのメーカーの比較は、個人によって異なり、合成音声のメーカー別による有意差はない。

以上の結果から、本来は、人声による録音図書のほうが望ましいが、即応性や利便性から、OCR 出力テキストデータによる合成音声 (Text to Speech) 録音図書も電子図書として利用すべきメディアである事が示唆された。

実際に電子図書閲覧室などで、OCR 機能を利用してテキストデータを作成し、合成音声で図書を読んでいる鍼灸学科の3年生の意見では、

- ・100%ではなくとも即応性や簡易に調べたい場合は、わかる程度でよい
  - ・本来は、種類別の音声登録辞書を用意して欲しい
  - ・OCR のレイアウト機能を何とかして欲しい
- などの意見が寄せられたが、技術的問題を理解しているようである。

#### 4. 電子図書館への道

電子図書館の構築に当たり、拡大文字作成用のテキストデータを作成するために、点字変換するためのテキストデータ作成に OCR 機能は必要不可欠であるが、作業全体の流れから OCR 機能を精査されることは無かった。従来、OCR 機能の自動認識率だけが問題とされ、どのようにすれば作業の TCO 削減がはかれるかと言った問題に取り組んで来なかったのは、残念な事である。

今回の結果から、OCR の認識率の比較を検討しても無意味である事が示唆され、レイアウト認識による認識率の低下が最も問題である事が判明した。自動レイアウト認識よりも手動によるレイアウト認識は、全体の認識率を向上させるため、レイアウト認識を手動で行うことによる、人的作業量の増加は、結果として校正の作業量を著しく軽減するため、手動で行う事が TCO 削減につながる事が判明した。

実際に作業量を計算すると、レイアウト認識を自動で行うと、校正にかかる時間は、A4原稿当たり、約

30分で、レイアウト認識を手動で行った場合は、10分程度である。レイアウト認識にかかる時間は、5分程度なので、大幅に作業時間を軽減できる。

作成されたテキストデータを元にして、様々な電子図書の展開がある。

原本

- ・録音図書作成（朗読）
  - 録音図書（カセット、MD）
  - 電子録音図書（mp3、NetMD）
  - DAISY
  
- ・電子テキストデータ作成（OCR）
  - 電子図書（拡大文字）
  - 電子点字図書（点訳）
  - 電子録音図書(Text to Speech)

## 5. おわりに

OCR によるテキストデータ作成により、Text to Speech での電子録音図書の展開が非常に容易となった。当然の如く、合成音声を利用するための、読み辞書や外字への対応や出力メディアの選択などについて、システム化を行う必要があるが、視覚障害者の電子図書館構築に当たりデジタル化や様々なデータの相互変換を容易に行い、利用者の要望に即した配信を行うためには、基本となる原本データのデジタル化が最も重要な要因である。

多くの大学の電子図書館では、図書館自身で図書のデジタル化を行っており、著作権処理も図書館が主体で行っている。また、電子図書の配信も図書館の重要な仕事の一部である。

残念な事に、視覚部では、電子データ作成に対する視覚部全体の意思統一が行われておらず、様々な問題を抱えている。ネットワークやサーバ、外字や様々な辞書、視覚障害補償など部門毎に独立した事象で行っている。

技術的な問題を無視して、「ちょっとやれば出来る」とか「Windows になってからは技術的に簡単だ」等と言う人は、未だに少なからず存在するが、電子図書館は、データの共有化を強力に押し進めるため、外字や様々な辞書、ネットワークや視覚障害補償ソフトウェアなども含めて共有せざるをえない状況を創出する。

最新の技術を利用するためには、技術に対する正しい評価が必要不可欠である。最新の IT 技術が視覚障害者の学習支援を積極的に推進する事を願ってやまない。

## 参考文献

Biermann, Grate Ideas in Computer Science, 1990, MIT Press

## Text-To-Speech Digital recording books by voice synthesis

MURAKAMI Yoshihisa<sup>1)</sup> UEDA Shoichi<sup>2)</sup>

<sup>1)</sup> Department of General Education, Division for the Visually Impaired, Tsukuba College of Technology

<sup>2)</sup> Department of Acupuncture/ Moxibustion, Tsukuba College of Technology

**Abstract :** There are two kinds of digital recording books. In one, a volunteer reads a book, the content is recorded, and the data is digitalized. In the other a book with a digital text is made with an OCR (Optical character recognition). And the text data is digitalized by a voice synthesis. Here, the making method and use, etc. about a digital recording books by an excellent voice synthesis are examined.

**Key Words :** Digital recording book, Voice synthesis, Text-To-Speech, and OCR