

特集論文

字幕付きテレビニュース放送を素材とした語学学習教材作成システム

A Construction System for CALL Materials from TV News with Captions

小林 聡
Satoshi Kobayashi

豊橋技術科学大学 情報処理センター
Computer Center, Toyohashi University of Technology
skoba@cc.tut.ac.jp

田中 敬志
Takashi Tanaka

豊橋技術科学大学大学院 情報工学専攻
Dept. of Information and Computer Sciences, Toyohashi University of Technology
ttakashi@slp.ics.tut.ac.jp

森 一将
Kazumasa Mori

(同上)
kmori@slp.ics.tut.ac.jp

中川 聖一
Seiichi Nakagawa

豊橋技術科学大学 情報工学系
Dept. of Information and Computer Sciences, Toyohashi University of Technology
nakagawa@slp.ics.tut.ac.jp

keywords:

CALL, language learning material, Japanese learning, English learning, listening, TV-news, caption, speech processing

Summary

Many language learning materials have been published. In language learning, although repetition training is obviously necessary, it is difficult to maintain the learner's interest/motivation using existing learning materials, because those materials are limited in their scope and contents. In addition, we doubt whether the speech sounds used in most materials are natural in various situations. Nowadays, some TV news programs (CNN, ABC, PBS, NHK, etc.) have closed/open captions corresponding to the announcer's speech. We have developed a system that makes Computer Assisted Language Learning (CALL) materials for both English learning by Japanese and Japanese learning by foreign students from such captioned newscasts. This system computes the synchronization between captions and speech by using HMMs and a forced alignment algorithm. Materials made by the system have following functions: full/partial text caption display, repetition listening, consulting an electronic dictionary, display of the user's/announcer's sound waveform and pitch contour, and automatic construction of a dictation test. Materials have following advantages: materials present polite and natural speech, various and timely topics. Furthermore, the materials have the following possibility: automatic creation of listening/understanding tests, and storage/retrieval of the many materials. In this paper, firstly, we present the organization of the system. Then, we describe results of questionnaires on trial use of the materials. As the result, we got enough accuracy on the synchronization between captions and speech. Speaking totally, we encouraged to research this system.

1. ま え が き

これまでに、多くは英語を対象としてリスニングに関する語学学習教材が開発され、またリスニング学習に関する研究もなされている [竹蓋, 山田 97, 中川 97, 峯松 00, アンケート, 東條 95]。しかし、それらによって提供されている話題やリスニングの対象は (例えば、数個のスキットや、英語の /l/ と /r/ の聞き分けなどに) 限定されている。加えて、既存の教材に含まれる発話は、多くの

場合は教材として収録されたものであり、ネイティブの自然な発話とは異なっていると考えられる。

音声の自然性については、ラジオやテレビ等のメディアを通しての外国語を語学学習に利用する試みが以前から行なわれている [東條 95, 東條 00]。その結果、現在、語学学習教材として CBS や CNN などのニュース放送を利用した CAI ソフトウェア [SIM 97, PE 00] や、教材のビデオ、テープなど、様々なものが市販されている [アンケート]。

例えば[SIM 97]は、CBSのニュース放送を素材とし、SIM方式(同時通訳方式)でのリスニングの練習を取り入れ、英語の語順で聞き取ることを可能とするよう開発されている。これにより日本人の「返り読み」の癖を矯正するとしている。また、リズム法則を取り入れた発音の練習も行なえ、不十分ではあるが発声のリズム的な評価も行なえる。[PE 00]は紙面による教材(技術英語)と、それに対応するニュース(7種類、各2~4分)の動画と音声収録されたCD-ROMからなり、これを用いてリスニング学習を行なう。紙面では英文に対応する日本語訳、キーワードとなる単語の訳が示されている。また、それぞれのニュースには字幕が付けられており、ニュース映像の画面に重ねて表示される。

このような教材は繰返し学習には適しているが、やはりいずれも提供される話題は少数に限定されている。

そこで多様な話題を提供することを目的として、各々の教師がニュース放送を録画し、それを教材とすることも多い。しかし、BS放送などを通して、英語字幕付きの英語によるニュース放送(PBS等)などを用いてそれを教材化することはできるが、CAIソフトウェアにあるような様々な機能を持たせることは、学習者個人や語学教師には到底困難である。

語学学習には繰返し学習が必要ではあるが[Darwing 89]、既存の教材では話題が限定されている。そのため、学習者が興味を持ち続けられるか否かという問題がある。それに対し、より広い話題や時期に即した話題に触れられるならば、学習者が外国語学習に興味を持ち続けることが可能になると考えられる。しかし、適当な話題のニュース放送などを用いて、学習者個人や語学教師が語学学習教材を作成するには限界があり、また、幅広い話題の内容を教材として利用するためには手間がかかるので、教材を半自動的に手軽に作成できるツールが必要となる。

本研究においては、字幕付きのテレビニュース放送をもとに、自然な発話による豊富な話題の提供、時期に即した話題の提供、教材の繰返し視聴、単語の語義表示、聞き取りテストの作成、ネイティブ発話と学習者発話の比較などの機能を持つ、市販のソフトウェアの機能を強化した語学学習教材(CALL:Computer Assisted Language Learning)を、半自動的に作成するシステムについて述べる。

以下、本システムの概要を2節に、教材化処理について3節に、字幕と音声の同期について4節に、5節では教材およびプレイヤーの機能を紹介する。6節では、作成した教材を試用してもらった際の被験者評価結果についてまとめる。最後に7節において今後の課題や展望について述べる。



図1 日本語の放映時画面例

2. 本システムの概要と特徴

今回述べる教材化システムには英語版と日本語版があるが、いずれも本質的には同様の機能を持っている。そこで、ここでは日本語版を中心に述べる。

現在、NHKの「ニュース7」では、音声認識技術を用いた字幕放送が、文字放送を通して行なわれている[安藤 01]。これにより毎日、映像と音声、その発話内容の字幕が得られる。本研究で構築しているシステムは、この映像と音声、字幕から、語学のリスニング教材を作成しようというものである。

ただし、アナウンサーの音声の字幕化には、音声認識および認識結果の確認・修正作業が行なわれているため、放送時の字幕の表示には実際の音声に対して数秒から十数秒の遅れがある。そのため、図1に示すように、字幕はニュース放映時の画面や音声とはずれが生ずる。図1の例であれば、映像は「通信白書」など通信関係の映像になっているものの、字幕は、その前の話題である気象についての情報となっている。

また、NHKによる字幕付きニュース放送の場合、字幕は映像や音声と異なり、文字放送を通して放送されている。そのため、単にニュース放送をビデオに録画するのみでは字幕は収録されないため、字幕を教材として使うことはできない。

これらの問題により、放送時の映像と字幕をそのまま教材とすることは難しい。そのため本研究では、2枚のTVチューナボードをPCに組み込み、映像と音声およ

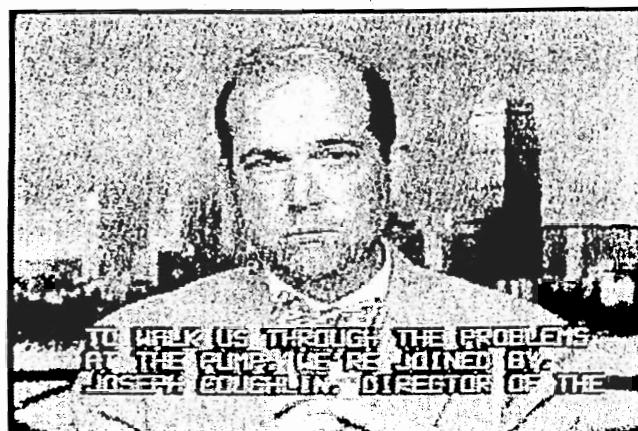


図 2 英語の放映映画画面例

び字幕を同時にコンピュータ内に収録する。また、映像と音声および字幕を取得した後、字幕の遅れを修正するため、音声と字幕との同期を自動的に取るようにした。

また、英語版に関しては、BS 放送の PBS ニュースを利用している (図 2)。これは字幕が映像として取得できる。この字幕を現在は人手で書き写しコンピュータにタイプ入力している。また、PBS などではアナウンサーの書き起こしテキストが Web[PBS] から参照することができ、このテキストを取り込んで教材化することも可能となっている。

前述の市販 CAI ソフトウェアや先行研究には、発話の自然性の問題や話題が限定されているという問題が見られた。それらに対して、本システムによって作成される教材は、教材の素材としてニュース放送を利用し、またコンピュータを用いているため、以下のような利点がある。

- (1) 自然に近い発話。
- (2) 時期に即した、また豊富な話題。
- (3) 教材作成の省力化が可能。
- (4) 多数の教材の蓄積が可能。
- (5) 特定部分の繰り返し視聴が容易。
- (6) 単語単位での辞書引きが可能。
- (7) 穴埋め形式の聞き取りテストなど、学習問題の自動作成が可能。
- (8) 学習者の音声を収録し、アナウンサー音声との波形や基本周波数、発声タイミングの比較が可能。
- (9) 字幕の提示において、発声箇所に対応して表示色を変えていくことにより、発声位置の確認が可能。
- (10) 字幕のフルテキスト表示や要約表示、表示無しなどの切り換えが可能。

特に上記 (2) により、学習者は学習への興味を持ち続けられると考えられる。

また、テキストとして字幕が保存されているため、多数の教材が蓄積された場合には、特定の話題についてニュースや学習者の母語による関連ニュースの検索が容易にな

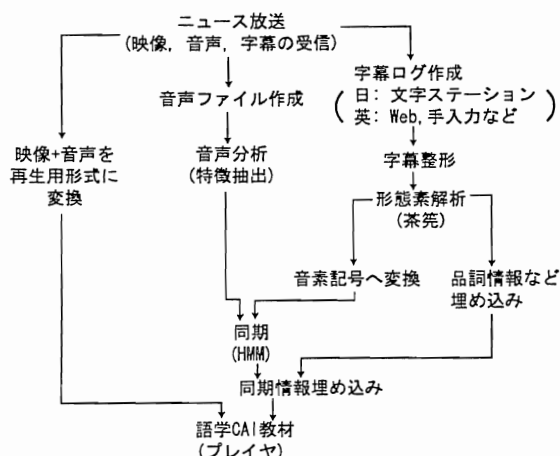


図 3 処理の流れ (日本語ニュース)

ることも考えられる。さらに、上記 (3), (7) で述べたように、コンピュータを利用した自動的な教材作成を目指しているため、個々の学習者が独自にニュース放送を収録し、教材として用いることも可能になる。このように、本論文で述べる教材化システムおよび教材プレイヤーは、移植性や汎用性を高めるために、Java を用いてプログラム開発を行なっている。教材に附属するツールや辞書についても、利便性を高めるためにできるだけ著作権フリーなものを用いている。

3. 教材化作業の手順

先に述べたように、本システムの日本語版は、NHK による字幕付きニュース放送 (「ニュース 7」) を下に、自動的な教材化を目指している。本日本語版システムの構成と処理の流れを図 3 および以下に示す。英語版システムも本質的に同じである*1。

- (1) 図 3 に見られる「音声、映像、字幕の受信」には、2 枚の TV チューナーボードを使用している。1 枚は、映像+音声の収録に使用し、MPEG1 形式で記録する。もう 1 枚は文字放送の受信が可能な TV チューナーボードであり ((株)IO データ機器の GV-BCTV3/PCI、および字幕表示のための附属ソフトウェア「文字ステーション」)、字幕放送を収録している。英語版の場合は映像から書き写すか Web から取得する。
- (2) 現在のところ (1) で得た字幕放送のログを、無用なスペースを除くなど整形すると共に、ログの文字化けの修正と確認を行なう。
- (3) 字幕の形態素解析を茶筌を用いて行なう [茶筌]。また、形態素解析の結果の確認・修正を行なう。英語

*1 日本語版と英語版の両者の主な違いは、字幕の取得方法、同期用の形態素解析 (英語版では品詞 tagger と発音辞書を使用)、音響モデル (日本語版では音節モデルを、英語版では音素モデルを使用)、辞書 (和英/英和) などである。

```

<CAI>
<VIDEO src="d:\\ttakashi\\yuki.mov" />
<SOUND src="d:\\ttakashi\\yuki.raw" />
<JIMAKU>
<ONE_JIMAKU start="0" end="1531">
  <WORD POS="接続詞" NORMAL="ところで" PRONOUNCE="ところで" START="0" END="820">ところで</WORD>
  <WORD POS="名詞-固有名詞-組織" NORMAL="雪印乳業"
    PRONOUNCE="ゆきじるしにゅうぎょう" START="820" END="1320">雪印乳業</WORD>
  <WORD POS="助詞-係助詞" NORMAL="は" PRONOUNCE="は" START="1320" END="1531">は</WORD>
  <WORD POS="記号-読点" NORMAL="," PRONOUNCE="," START="1531" END="1531">,</WORD>
</ONE_JIMAKU>
<ONE_JIMAKU start="1531" end="4972">
  <WORD POS="名詞-副詞可能" NORMAL="きょう" PRONOUNCE="きょう" START="1531" END="1704">
    きょう</WORD>
  <WORD POS="助詞-格助詞-一般" NORMAL="から" PRONOUNCE="から" START="1704" END="1948">から
</WORD>
  <WORD POS="名詞-一般" NORMAL="牛乳" PRONOUNCE="ぎゅうにゅう" START="1948" END="2217">牛乳
</WORD>
  <WORD POS="助詞-並立助詞" NORMAL="と" PRONOUNCE="と" START="2217" END="2428">と</WORD>
  :
  :

```

図 4 日本語版の字幕例
発話内容:「ところで雪印乳業は、きょうから牛乳と.....」

版では Brill's Tagger を使用して品詞解析を行なう [Brill].

- (4) 手順 (2) の形態素解析の結果として得られる読みを利用し、音節系列を作成する。英語版では、発音辞書を利用して音素系列を得る。
- (5) 手順 (3) と同様に、形態素解析の結果として得られる個々の形態素の品詞情報などを生成する。英語版では、品詞タグを用いて品詞を得る。
- (6) 録画から得られる音声 (12kHz サンプリング) を分析し特徴パラメータ系列に変換する。
- (7) 音声認識技術を利用して [SPOJUS, 甲斐 97], 手順 (3) で得た字幕の音節系列と音声 (音素) の特徴パラメータ時系列との照合により、字幕と音声との同期情報を作成する。
- (8) 手順 (6) の結果から、句読点単位および単語単位での開始時刻を得て、句読点単位での戻り・先送りを可能とするリンク情報を作成する。
- (9) 手順 (4) で得た品詞情報など、および手順 (7) で得た同期情報および戻り・先送りに関するリンク情報を、字幕の文字列に書き加える。これらの情報を含んだ字幕例を図 4 に挙げる。
- (10) 最終的に、再生用に形式の変換を行なった映像+音声と上記の情報を含んだ字幕により、Java と QuickTime を利用したプレイヤーを用いて再生する語学学習教材となる。得られた語学学習教材の画面は、図 5, 図 6, 図 7 のようになる。

上記手順 (1)~(10) 中、手順 (2) を除き、個々の手順においては自動化がなされている。手順 (2) では、字幕取得プログラムの制約により字幕が正確に収録されていない場合 (字幕放送の制御コードの問題で、数字が文字化けする) や、字幕に現われていない間投詞、言い直しがあるため、取得された字幕ログに問題がないかの確認および修正を人手で行なっている。その後、4 節で述べる同期計算プログラムに音声データと手順 (2) で整形した字幕を入力すると、図 4 に示すような教材用のデータが出力されるようになっている。またその他に、音声データを MPEG1 形式のビデオデータから分離する、ビデオフォーマットを MPEG1 形式から QuickTime ビデオ形式に変換する、などの細かい作業については現時点では手作業で行なわなければならない。本システムを用いて教材を作成する作業 (手順 (1) は除く) にかかる時間は、コンピュータの習熟度合いにもよるが、数分のニュース放送のデータであれば、フォーマット変換や計算時間を含めて全体で 30 分~1 時間程度、このうち教材作成者の作業時間は 10~40 分程度である。

このようにして得られた語学学習教材は、Java で記述され QuickTime の機能を利用するプレイヤーによって実行される。この教材とプレイヤーにより、以下のような機能が実現されている。

- ニュース画面および音声と同期した字幕の提示。
- タイムスライドによる任意箇所での再生。
- 字幕単位での戻り、先送り。

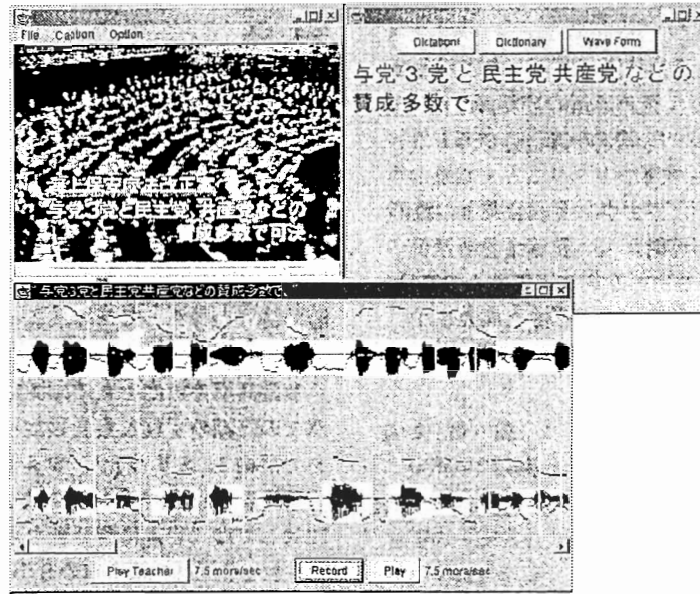


図 5 日本語版の教材画面例: 音声と字幕の同期および音声波形とピッチ表示画面 (上:アナウンサー音声, 下:学習者音声)



図 6 日本語版の教材画面例: 辞書引き画面

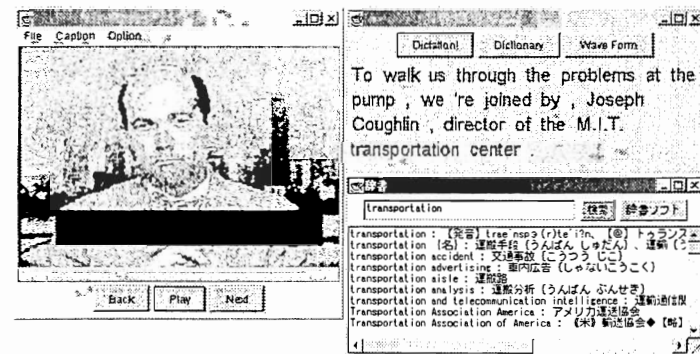


図 7 英語版の教材画面例: 辞書引き画面

- 字幕表示の切り替え (ひらがなのみでの表示, 特定の品詞のみの表示など).
- 字幕の提示において, 発声箇所に対応して表示色を変えていくことにより, 発声位置の確認が可能.
- アナウンサー音声の波形・ F_0 (基本周波数)表示.
- 学習者音声の波形・ F_0 表示.
- 学習者の音声を収録し, アナウンサー音声との波形や基本周波数の比較が可能.
- アナウンサーと学習者音声の発声タイミングの表示.
- 穴埋め形式の聞き取り問題の自動作成と採点.
- 形態素単位での辞書引き [英辞郎].
- 再生速度の変更.

4. 音声と字幕の同期

表 1 音声試料の分析条件

サンプリング周波数	12 kHz
ハミング窓幅	21.33 m 秒 (256 ポイント)
フレーム周期	8 m 秒 (96 ポイント)
LPC 分析次数	14 次
特徴パラメータ	LPC メルケプストラム (4 フレーム・セグメント, 10 次元 * 4 フレームを圧縮 → 20 次元) + Δ CEP(10 次元) + $\Delta\Delta$ CEP(10 次元) + Δ POW (1 次元) + $\Delta\Delta$ POW (1 次元)
(日本語)	
特徴パラメータ (英語)	LPC メルケプストラム (10 次元) + Δ CEP(10 次元)
音響モデル (日本語)	音節モデル, 5 状態 4 出力分布, 離散継続時間付き出力分布型 HMM 4 混合ガウス分布, 全共分散行列 155 名, 2 万文 (ASJ+JNAS) で学習
音響モデル (英語)	音素モデル, 5 状態 4 出力分布, 離散継続時間付き出力分布型 HMM 4 混合ガウス分布, 全共分散行列 326 名, 3260 文 (TIMIT) で学習
モデル数	114(日本語), 39(英語)

現在, NHK による字幕放送は, 音声から数秒から十数秒の遅れを伴っている. PBS の場合, 発話と字幕はほぼ同時に提示されている. しかし, いずれの場合も, 単語単位での同期情報は放送からは得られない. 本システムでは, 単語単位の発声タイミングを示すことも機能の 1 つとしているため, 単に放送されたタイミングでの字幕の提示だけではなく, 音声と字幕の同期を新たに取る必要がある. 丸山らも TAO の「視聴覚障害者向け放送ソフト制作技術研究開発プロジェクト」の一環として音声と字幕の同期化の研究を行なっている [丸山 98]. これはワードペアのスポッティング法に基づく手法であり, 沸き出し誤りが生じる手法である.

本システムの音声と字幕の同期を取る際の音声の分析条件等を表 1 に示す. 字幕のかな文字列 (あるいは音素記号列) からそれに対応する音節 (音素) 単位 HMM の連結を構成し, 入力音声とマッチングする.

現状では, 30 秒から 1 分半程度ずつの 1 まとまりのニュースの音声区間を切り出している. それに対応する字幕に対して, 日本語では形態素解析ソフト ChaSen [茶筌] のかな表示を用いて音節系列を作成し, 英語では CMU 発音辞書 [CMU 辞書] から音素系列を作成する. そして切り出した音声区間と, 字幕から得られる音節 (音素) 系列に対応する隠れマルコフモデル (HMM) を連結した音響モデルを One Pass Viterbi アルゴリズム [中川 88] を用いて最適な時系列同士のマッチングを行ない, 時間の同期情報を得る. ただし, この際, 無音区間をパワーの閾値により除く処理を行なっている.

この同期の精度を正確に測定するため, 各名詞の終端点での同期のずれを集計したところ, 日本語の音声データで平均 14 ミリ秒 (サンプル数 39), 英語の音声データで平均 92 ミリ秒 (サンプル数 25) となった. 英語の音声に対しては, 音声に背景音などの比較的大きなノイズが含まれるため (英語音声の S/N は 16dB~20dB, 日本語音声の S/N は 30dB~33dB^{*2}), また音響モデルの精度が不十分 (音響モデル作成のためのトレーニングデータ量が少ない) であることが, 日本語音声の精度よりも悪くなる原因となっている. しかし 6.2 節で述べる字幕表示の被験者実験結果では, 日本語, 英語ともに同期精度で高い評価を得ており, 字幕表示においては違和感のない精度が達成されているものと考えられる.

しかし, いずれの場合もアナウンサーの発話内に含まれるポーズ, 言い淀みや間投詞の問題がある. 字幕にはアナウンサーの言い淀み等は表示されない. そのため, 字幕と音声との同期を取る際に, 発話内に言い淀み等が含まれることによって音声との照合結果の尤度が低くなり, 字幕に対して最適な音声区間が得られない場合がある. ポーズについては, ポーズを除いた音声試料に対して音声照合を行なうことで対処し, 言い淀み等については音声認識システムの未知語処理技術を用いて, 任意の音節列 (日本語版) や音素列 (英語版) のマッチングで吸収するようにしている [甲斐 99].

また, 読み付与誤りやアナウンサーの言い直しによって, 発話内容と音節 (音素) 系列が対応しない場合があるが, ほとんどの場合音節 (音素) 系列の前後関係により違いが吸収される. 例えば, 発話内容が「さいこうよんひやくにじゅうえん (最高 420 円)」であり, 付与された読みが「さいこうよんにぜろえん」であった場合, 前後関係の「さいこう~えん」があることによって「よんにぜろ」の HMM 系列に「よんひやくにじゅう」の音声が無理矢理吸収されることになる. 実際に, 日本語の音声データにおいて発話内容と読み付与が対応していない部分の同期精度を調べたところ, 15 単語中 12 単語 (80%) において読み付与の正しい他の部分と同程度の同期精度

*2 音声自動認識では, S/N が 20dB より悪くなると急激に認識率が低下する. 研究段階でよく用いられているクリーン音声の S/N は 40~50dB 程度である.

が得られていることがわかった。英語の音声データにおいて、単語辞書に登録されていない未知語に対しては発音記号列がわからないため、任意の音素に対して大雑把な照合が取れるガベージモデル(全音素から学習したモデル)を用いて未知語処理を行っている。実際に、未知語を含むニュース放送を用いてガベージモデルを未知語の個所に挿入した場合、違和感の少ない精度で字幕が表示されることを確認している。

もう 1 つの問題として(特に日本語版において)、スタジオ外のレポートからの発話の場合には字幕が存在しない問題がある。そのため、字幕と同期する音声領域の決定が困難になる。この場合、字幕ログに各字幕の記録時刻を付け、その時刻を参考にアライメントを取ることによって解決可能であると考えている。

5. 機能の詳細

本節では、作成された教材およびプレイヤーの機能を述べる。

5.1 映像+音声の再生

映像・音声は、字幕に含まれる句読点単位で再生される。このため、特定部分の繰り返し視聴が容易になっている。また、連続しての再生も可能である。加えて、同単位での先送り、戻りも可能である。

若干の再生速度の変更も、サンプリング・レートの制御により可能になっている。これにより、発話速度が速いために聞き取り難い箇所もゆっくりと再生させ、聞き取ることが可能となる。今後は、発話中のポーズ長の制御による話速制御を実現する予定である[中村 94]。

5.2 字幕

字幕の提示は、漢字かな混じりでの提示(標準用)と、かなのみでの提示(初心者用)、特定の品詞のみの提示(中級者用)、および提示なし(上級者用)の 4 種類の提示方法が選択可能である。今後、学習者のレベルに合わせてより詳細な字幕提示方法の制御を実現する予定である(学習者にとって難しい漢字のみかなで、残りは漢字で表示する、要約の提示など)。

5.3 発音学習支援

このプレイヤーは、字幕に対応するアナウンサー音声の波形と基本周波数(F_0 , 声の高さに対応する物理量)を表示することが可能である。また、学習者の音声を録音し、同様に音声波形と F_0 を表示できる。これにより、発話速度と声の高さ、大きさについて、アナウンサーの音声と比較が可能である(図 5)。両者を比較することで、アクセントやイントネーションの例示ができ、学習の参考になると考えられる。

また、音声と字幕の同期をとるアルゴリズムを用いて音声波形と発声単語の対応付け結果も表示できる。図 5 下の音声波形とともに表示されている縦棒は単語境界を表し、上段の教師音声(アナウンサー音声)と下段の学習者音声の単語境界を線をつなぎ、教師音声と学習者音声の発話タイミングの比較が容易にできるようになっている。また、発話された時間長と字幕から教師と学習者の発話速度を計算し、日本語では mora/sec, 英語では word/sec で表示することができる。

今後、学習者音声の評価を行なう機能を追加する予定である[峯松 00, 中村 01, 三輪 01]。

5.4 辞書引き

字幕中の語句の辞書を引くことも可能である。この際、表示されている字幕の一部を選択した後、プレイヤー画面右下の「辞書」ボタンを押すことで、目的とする語句の英訳や和訳が表示される(図 6, 図 7)。また、マウスポインタを目的の単語上に乗せることにより、簡略な意味表示も可能である(図 6 右上の「日本」)。

現在、日本語版の日英辞書および英語版の英日辞書にはともに「英辞郎」[英辞郎]を用いている。また日本語版では、中国語での意味表示にも対応している。意味表示には、適当な日中辞書がなかったため、Unicode によって記述され、日本語による検索が可能な中日辞書(版權フリーなもの)を利用している。この結果、検索文字列は日本語列で与えるが、検索結果の画面では中日辞書の形式で表示される。辞書の形式については、今後日中辞書の形式のものを作成し、置き換える予定である。

5.5 ディクテーション問題作成

字幕中の特定の品詞に対する聞き取りテスト作成および採点機能を有している。この機能により、字幕中には特定の品詞の箇所が空欄として表示され、その空欄に対応する回答を入力する。入力後、正解とともに評価結果が表示される。

このような機能は、先に形態素解析や品詞のタグ付けを行なっているため、容易に実現できる。例えば、学習者が苦手な「数量表現」部分の聞き取りテストを容易に作成できる。

今後、ディクテーション問題や内容理解問題の作成など教材作成に教師が介入できるようにしていく予定である。

6. システムの機能評価

本節では、本システムによって作成された日本語版および英語版の教材を被験者に試用してもらった際に行なったアンケート調査の結果について述べる。被験者の一覧を表 2 に示す。留学生の日本語学習者は日常会話は勿論、日本語の講義も理解できる程度である。一方、日本人学生の英語学習者は、日常会話もおぼつかない程度のもの

である。アンケートは基本的に5段階評価(1:悪い~5:良い)で行ない、一部自由記述項目もある。教材の試用は30分程度ずつ、2回に分けて行なった。2回目の試用の後、アンケートに回答してもらった。

表2 被験者

日本語版	日本語教師 2名 (日本人)
	留学生 6名 (うち3名は中国人, 3名は その他東南アジアからの留学生)
英語版	英語教師 2名 (日本人)
	学生 5名 (日本人)

6.1 同期精度

字幕と音声の同期精度については、英語版では全被験者の評価値の平均が5.0、日本語版では4.75と、いずれも十分な精度が得られているという評価を得た。

6.2 字幕表示

発声に対応して字幕の色を変えていく表示については、聞き取りにも学習にも役立つとの評価が得られた(英語版と日本語版それぞれ「聞き取りの役に立つ」が4.8, 4.5, 「学習の役に立つ」が4.6, 4.8)。

字幕の全文表示・名詞のみ表示などの切り替え機能は学習の役に立つか否かについては、英語版と日本語版で評価が分かれ、英語版では被験者全員の評価の平均が3.6であるのに対し、日本語版では4.1と高い評価を得た。

字幕の全文かな表示(日本語版のみ)が必要かという設問に対し、学生は2.8、教師は5.0と、その評価に大きな差が出た。これは、日本語能力が高い被験者であったためと考えられる。なお、学生2名と教師1名から「漢字には振仮名が必要」という意見が出ており、振仮名表示機能の追加によりこの評価は変化すると考えられる。

6.3 辞書引き

辞書引きの語彙数については英語版で2.3、日本語版で2.9と評価が低く、辞書の語彙数について不満があった。しかし現在英和辞書(英語版)で28万語彙、和英辞書(日本語版、中国語対応以外)で54万語彙、中日辞書(日本語版、中国語対応)で3万語彙と、語彙数は必ずしも少ないわけではない。そのため、ここで得られた評価は、単語の活用などによる語形変化による問題、あるいは使用辞書の想定ドメインと、ニュース放送というドメインでは使用される語彙に違いがあるためと考えられ、改善を要する。

また辞書の内容が役に立ったかどうかについては、日本語版で4.3、英語版では3.3と比較的良好な評価を得た。

辞書引き機能が学習の役に立つかという設問については、英語教師の評価が3.5と低い、他は4.6以上の評価を受け、学習に役に立つと評価された。

これらの結果から、辞書の語彙という面から見ると不備はあるものの、オンライン辞書の有効性が示された。

なお、中国人用日本語版教材の辞書には中日辞書を使用しているが、学習の役に立つが(中国人留学生のみで、3.3)、使い易くはなく(同2.3)、日中辞書が必要(同5.0)という結果であった。

6.4 ディクテーション問題

ディクテーション問題は、学習の役に立つとの評価が得られた(英語版4.3、日本語版4.2)。今回のアンケートでは、穴埋め箇所が多すぎる、何回でも聞き直せる方がよい、正解をより明示的に示す方がよい、単語だけでなく句も対象とする方がよいなどの意見が得られた。今後寄せられたこれらの意見を下に改善を行なう。

6.5 波形と基本周波数の表示

アナウンサー音声の波形表示は学習の役に立つか否かについては、教師と学生とで評価が分かれ、英語版学生で3.4、英語版教師で2.5、日本語版学生で3.3、日本語版教師で2.0となった。

学習者自身の音声波形の提示については、多くは3.0以上の評価を与えているが、英語版教師は1.5と低い評価を与えている。

基本周波数についても同様で、アナウンサー音声、学習者音声のいずれに対しても日本語版では教師は4.5と高く評価しているが、英語版教師は2.0と評価している。

波形や基本周波数の表示によるアナウンサー音声と学習者音声の比較が役に立つか否かについても同様に日本語版教師の評価は3.0と3.5と比較的高めの評価を行なっているのに対し、英語版教師の評価はいずれも2.5と低いものであった。

本評価実験では、試用者に音声波形と基本周波数曲線を直接提示していたが、学習者にはそれらの音声学的な意味が分からないのではないかという意見があり、これが評価に現われたと考えられる。

今後、音声波形や基本周波数曲線の提示については、それらを直接提示するのではなく、学習者に分かり易い形式を用いることを検討していきたい。

しかし、全体としてみれば、学習の役に立つと評価されていた(英語版学生4.8、英語版教師3.0、日本語版学生4.0、日本語版教師5.0)。

6.6 他の学習法との比較

他の学習方法(テレビおよびラジオの語学講座、市販ソフトなど)と比べてみた場合、本システムにより作成した教材による学習方法の方が高い評価を受けた(英語版学生4.0、英語版教師3.0、日本語版学生4.2、日本語版教師5.0)。これは、本システムにより作成される教材には辞書引き、再生箇所の選択などインタラクティブ性があることによる結果と思われる。

6・7 授業や独習への導入

授業には、日本語版は使えるが(日本語版教師 5.0)、英語版はそれほどでもない(英語版教師 3.5)という評価となった。これは、本システムによって作成される教材は、上級者向けであると評価されている(日本語版教師 4.75、英語版教師 5.0)ことと符合する。理由としては、やはり素材がニュースであるため、話速、使用される語彙が上級者向けであることが挙げられた。今後、中級者向けのコースウェアの開発を進めていく必要がある。

しかし独習用教材としては高く評価された(日本語版教師 4.0、日本語版学生 3.7、英語版教師 5.0、英語版学生 3.6)。

6・8 今後必要な機能

今後必要な機能として、学習者音声の発音やイントネーション、リズムの評価機能については、英語版使用者と日本語版使用者とで評価が分かれた(英語版 3.7、日本語版 4.5)。日本語版使用者は必要とし、英語版使用者はさほど必要とは思っていないことが分かる。これは、学習者の語学力レベルの差と考えられる。

さらに、辞書の他に関連する語彙や例文などの表示が必要とされた(全平均 4.3)。これはコンピュータ利用によって関連ニュース検索により実現が可能だと考えている。

7. む す び

本論文では、字幕付きテレビニュース放送から、語学学習教材を半自動的に作成するシステムについて述べた。本システムは、収録したテレビニュース放送とその字幕を下に、語学のリスニング(とスピーキング)のための教材を半自動的に作成する。また、単にニュースを再生するだけでなく、コンピュータを利用することの利点として、字幕単位での戻り・先送り、辞書引き、聞き取り問題の自動作成と採点、アナウンサーと学習者の音声波形と基本周波数の表示なども可能にしている。

ただし、スタジオ外からのレポートの際に見られるような字幕と音声との不一致(一般に字幕化されていない)の処理が問題点として残っている。

今後、上記の問題点を解決するとともに、以下のような機能の実現を考えている。

- (1) 手動作業部分の自動化。
- (2) 音声信号の理解しやすい波形等での提示や、学習者の発音の自動評価。
- (3) 英語による副音声との同期(日本語版)。
- (4) 学習者の母語によるインタフェースや辞書、意味表示。例えば、日中辞書の導入(日本語版)。
- (5) 学習者のレベルに応じたコースウェアの設計・実現。例えば、教材の字幕を学習者のレベルに合わせ、漢字をかんで表示するなどの機能の提供(日本語版)。

(6) 教材化作業(特に問題作成)への教師の介入を可能にする、オーサリングシステムや多様な形式の問題作成への対応。

(7) World Wide Web 上におけるニュース映像や音声、さらにその発音に近い書き起こしテキスト[テレビ朝日、TBS、CNN]を用いた教材化への対応や関連ニュースの検索等。

また、今回は教材作成システムの機能とその評価について述べたが、今後、授業等への導入による評価を行っていく予定である。

謝辞

本研究は、文部科学省 科学研究費補助金 特定領域研究(A)「メディア教育利用」の「音声言語処理技術と学習者モデルを用いた語学学習システムの研究」(課題番号 12040104)の研究助成を受けて行なわれている。本研究では、NHKにより放送されている字幕付きニュース放送を使用している。この使用を快諾していただいたNHK放送技術研究所の関係諸氏に深く感謝する。

また、中日辞書、日中辞書についてご協力いただいた、研究生の李麗芳さんに感謝いたします。

◇ 参 考 文 献 ◇

- [Brill] “品詞解析プログラム Brill's Tagger”, <http://www.cs.jhu.edu/brill/>.
- [CMU 辞書] “The CMU Pronouncing Dictionary”, <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.
- [CNN] “CNN ニュース”, <http://www.cnn.com/video/>.
- [Darwing 89] Darwing, Tracey M.: “Information type and its relation to non-native speaker comprehension”, *Language Learning*, 39, pp. 157-172 (1989).
- [PBS] “The Online News Hour”, <http://www.pbs.org/newshour/>.
- [PE 00] “PE English for Professional Engineers”, *English Express 別冊*, vol.6, 朝日出版(2000).
- [SIM 97] “SIM スーパーエルマー CBS コース”, 株式会社ソース, 東京 SIM 外国研究所(1997).
- [SPOJUS] “SPOJUS-SYNO 日本語連続音声認識システム”, <http://www.slp.tutics.tut.ac.jp/SPOJUS/>.
- [TBS] “TBS ニュース”, <http://www.tbs.co.jp/>.
- [アンケート] 豊橋技術科学大学中川研究室: “英会話 CAI ソフトアンケート結果”, <http://www.slp.tutics.tut.ac.jp/CALLsoft/>.
- [安藤 01] 安藤彰男, 今井亨, 小林彰夫, 本間真一, 後藤淳, 清山信正, 三島剛, 小早川健, 佐藤庄衛, 尾上和穂, 世木寛之, 今井篤, 松井淳, 中村章, 田中英輝, 都木徹, 宮坂栄一, 磯野春雄: “音声認識を利用した放送用ニュース字幕製作システム” *電子情報通信学会論文誌*, Vol. J84-D-II, No.6, pp.877-887 (2001).
- [英辞郎] “英辞郎”, <http://www.alc.co.jp/ejjiro/index.html>.
- [甲斐 97] 甲斐充彦, 中川聖一: “冗長語・言い直し等を含む発話のための未知語処理を用いた音声認識システムの比較評価”, *信学論*, Vol. J80-D-II, No.10, pp.2615-2625 (1997).
- [甲斐 99] 甲斐充彦, 廣瀬良文, 中川聖一: “単語 N-gram 言語モデルを用いた音声認識システムにおける未知語・冗長語の処理”, *情報処理学会論文誌*, Vol.40, No.4, pp.1385-1394 (1999).
- [竹蓋] “こうすれば聞こえるヒアリング”, 竹蓋幸生監修, アルク.
- [茶筌] “日本語形態素解析システム 茶筌”, <http://cl.aist-nara.ac.jp/lab/nlt/chasen/>.
- [テレビ朝日] “ANN ニュース”, <http://www.tv-asahi.co.jp/ann/news/web/index.html>.

- [東條 95] 東條加寿子:“衛星放送のニュースを利用した教材:英語リスニング教材としての妥当性の考察”,九州女子大学紀要,第31巻,pp.97-110(1995).
- [東條 00] 東條加寿子:“英語教育とメディアの接点-メディアが伝える英語情報の教材化に関する考察-”,九州女子大学紀要,第36巻3号,pp.15-29(2000).
- [中川 88] 中川聖一:“確率モデルによる音声認識”,電子情報通信学会,(1988).
- [中川 97] 中川 聖一, Alan A. Reyes, 鈴木 英之, 谷口 泰広:“音声認識技術を利用した英会話 CAI システム”,情報処理学会論文誌,第38巻第8号,pp.1649-1658(1997).
- [中村 94] 中村章,清山信正,池沢龍,都木徹,宮坂栄一:“リアルタイム話速変換型受聴システム”,日本音響学会誌,Vol. 50, No. 7, pp.509-520(1994).
- [中村 01] 中村直生,峯松信明,中川聖一:“HMM を用いた英単語単発声からの発音上の癖推定”,電子情報通信学会技術研究報告,SP2000-127,pp.41-48(2001)(2000).
- [丸山 98] 丸山一郎,阿部芳春,江原暉将,白井克彦:“ワード列ベアモデルによる字幕送出タイミング検出の検討”,電子情報通信学会音声研究会,SP98-34(1998).
- [峯松 00] 峯松信明,藤澤友紀子,中川聖一:“英単語発音上の癖の自動推定・視覚化とそれに基づく発音能力の韻律的評定”,電子情報通信学会論文誌 D-II, vol. J83-DII No.2, pp.486-499(2000).
- [三輪 01] 三輪多恵子,中川聖一:“日本人英語とネイティブ英語の韻律特徴の比較”,日本音響学会講演論文集,1-2-12,pp.229-230(2001)(2000).
- [山田 97] 山田玲子, W. Strange, 久保田理恵子:“日本語話者に対する米語母音聴取訓練”,日本音響学会講演論文集,1-3-9,pp.379-380(1997).

[担当委員:美濃導彦]

2001年11月12日 受理

著者紹介

小林 聡(正会員)



1991年豊橋技術科学大学情報工学課程卒業。1994年同大大学院工学研究科情報工学専攻修了。1999年静岡大学博士後期課程電子応用工学専攻満期退学。1999年豊橋技術科学大学知識情報工学助手。2000年同大情報処理センター助手。博工。コンピュータの教育利用の研究に従事。

田中 敬志



1978年生。2001年豊橋技術科学大学情報工学課程卒業。現在同大学院修士課程情報工学専攻在学中。音声言語処理に関する研究に従事

森 一将



1978年生。2000年豊橋技術科学大学情報工学課程卒業。現在同大学院修士課程情報工学専攻在学中。音声言語処理に関する研究に従事

中川 聖一(正会員)



1948年生。1976年京都大学大学院博士課程修了。同年京都大学・情報工学科・助手。1980年豊橋技術科学大学・情報工学系講師。1990年同大学教授。1985-1986年カーネギーメロン大学客員研究員。音声情報処理,自然言語処理,人工知能の研究に従事。工学博士。1977年電子通信学会論文賞,1988年度IETE最優秀論文賞,2001年電子情報通信学会論文賞受賞。著書「確率モデルによる音声認識」(電子情報通信学会編),「音声・聴覚と神経回路網モデル」(共著,オーム社),「情報理論の基礎と応用」(近代科学社),「パターン情報処理」(丸善)など。