

音声合成と画像情報を組合わせた英会話学習支援システム

日下大輔[†] 堀 幸雄[†] 今井慈郎[†]
香川大学[†]

1. はじめに

近年、社会の国際化によって日常的なコミュニケーションやビジネスで英語能力を必要とする場面が多くなっている。英語能力は個人の所得に対しても関係しているとの報告もあり[1]、英語の重要性は高まっている。

本稿では、国際化社会に学生や社会人が対応するための、音声合成技術を活用した英会話学習支援システムを提案する。本システムが提供する機能は英語能力を向上させる練習方法[2]であるリスニングとディクテーション練習機能である。

2. 関連研究

本システムは、リスニングとディクテーション練習機能を提供する。先行研究として中川氏の研究[3]が挙げられる。中川氏は字幕と副音声付きテレビニュース放送を利用して、ディクテーション教材を半自動的に作成できる教材作成システムを提供している。中川氏のシステムでは音声が付いている動画を教材とする必要があった。一方、本研究で開発するシステムは Web に存在するテキストを合成音声で読み上げ、リスニング教材とすることで、教材へと加工できる資源の幅を増やすことができる。そのため、Web に存在するニュースのテキストを利用することが可能となる。

また中川氏は論文の中で既存の市販語学学習ソフトは視覚的フィードバックが少ない問題を述べている。よって、本研究では音声の意味に対応した画像を Web より収集して、ユーザに示すことを考えている。

画像を提供することで英語の学習を支援するシステムの例として imaizumi 氏の開発した Web アプリケーションの「英単語シアター」[4]がある。そのほかの例として、ユーザが入力したテキストに対して画像を表示する赤松大典氏が開発した「さしえショー」[5]がある。状況を Flash のアニメーションとしてユーザに提供し、状況と共に英語を学ぶ「LittleFox」[6]という英語学習サイトがある。本研究で開発するシステムは、「LittleFox」のように会話が行われている状況を画像からユーザに印象付けて、ユーザが状況にあった英会話を行えるように練習できる英会話学習支援システムを開発する。

3. システムの詳細

本システムの全体構成を図1に示す。本システムは3つのモジュールから成る。Web から英語ニュースサイトに掲載されているニュース文章を収集するための「Web 情報収集モジュール」、Web から収集したニュース文章の音声化と文章に適合する画像を探索する「メディア作成モジュール」、ユーザに提供するリスニング・ディクテーション教材を作成する「ディクテーション教材作成モジュール」である。

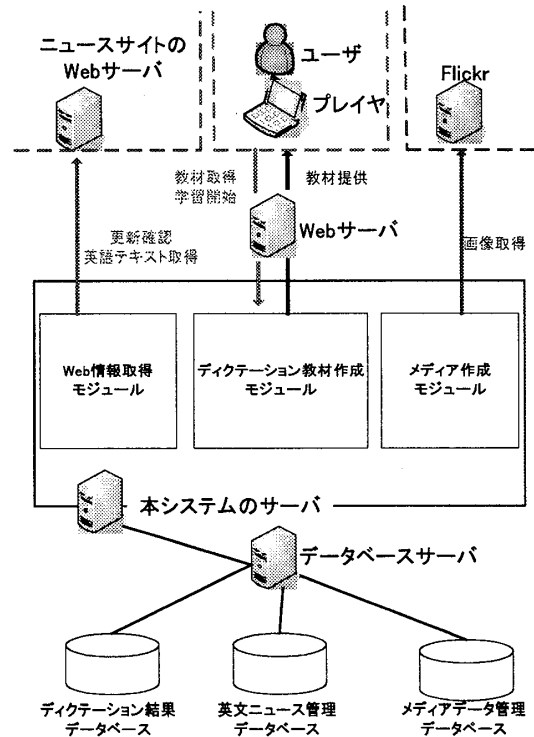


図1. システムの全体図

3. 1. Web 情報収集モジュール

Web からのテキスト情報収集の流れを図2に示す。必要なサイトのみの情報を収集するため、収集したいサイトの URL をサーバ上にテキスト形式でまとめておく。

Web から収集したニュース文章情報はテキストファイル形式でサーバに保存する。Web 上のニュース文章をテキスト形式にするために、テキストブラウザの w3m を利用する。w3m を使用することで、Web に掲載された html ファイルに存在する html タグを取り除いたテキスト情報を収集できる。保存するテキストファイルは、一文ごとに改行されて分けられている状態で保存する。

A study on the English Conversation Support System Using artificial-voice technology and images

[†] Daisuke Kusaka · Kagawa University

[†] Yukio Hori · Kagawa University

[†] Yoshiro Imai · Kagawa University

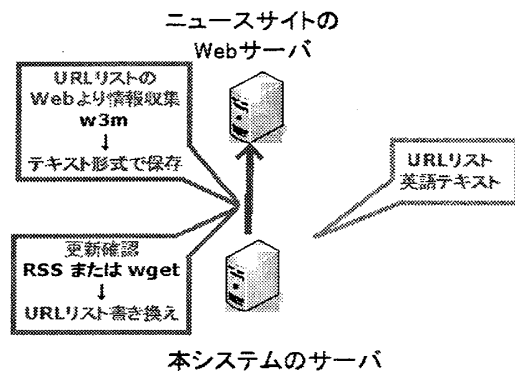


図 2. Web 文章の取得

3. 2. メディア作成モジュール

このモジュールは Web 情報収集モジュールで保存したテキストファイルを元に音声ファイル作成と画像収集を行う。

保存したテキストファイルは、SAPI5 を利用して音声合成を行い、テキストの内容を読み上げた音声ファイルを作成する。

保存したテキストの文章は単語ごとに空白文字で区切られている。ユーザへリスニング・ディクテーション中に出力する画像は単語毎の意味を連想させる画像とする。単語を検索条件とし、画像を共有するコミュニティサイト「Flickr」に FlickrAPI 利用の画像検索を行い画像を取得する。

音声ファイルの作成と画像取得はユーザのシステムへの待ち時間を短縮するため、Web 情報を取得した時点で、ファイルをサーバに保存する。

3. 3. ディクテーション教材作成モジュール

ユーザに教材を利用してもらうためのインタフェースは JavaScript による GUI インタフェースを使用する。ユーザはインタフェースの最初の画面で、次の (1) から (3) を選択できる。

- (1) 自分の興味のある英語記事のカテゴリの選択。
- (2) リスニングモードで起動するかディクテーションモードで起動するか選択。
- (3) 問題文の順序が元のテキストファイルの通りか、またはランダムかを選択。

ディクテーションの問題では、文章中の単語をランダムに隠す処理を行う。

ディクテーション時、ユーザに提供する情報は、ディクテーション用の穴あきテキスト、音声ファイル、および画像である。ユーザはテキストの穴を埋めることでディクテーションを行う。

本システムでは、音声や画像を同期させる必要がある。また、システムが読み上げている部分を、ユーザにわかりやすく表示する必要がある。そこで、教育効果を高める目的で、英語ニュースの音声ファイルの再生にあわせて、再生されている単語にハイライトを付与し、現在の再生位置を明示する。このハイライト効果を高めるため、辞書ファイルに存在した約 5 万個の単語の音声ファイル作成とその音声の再生時間の調査を行った。その結

果、ハイライトの進捗を音声再生と同期させることが可能となった。以下に調査結果に基づく、単語読上げ時間と音声再生時間を示す。

(a) 辞書の単語の音声ファイル化

単語毎に SAPI5 モジュールを使って音声ファイルを作成した。1 つの単語につき ID-単語の文字数. wav 形式でサーバに保存した。

(b) 単語の平均読み上げ時間の調査

(a) で作成した音声ファイルの再生時間を調べることで、システムに読み上げられる単語の文字数 x 毎の平均再生時間を算出できる二次関数を求めた。音声ファイルを文字数別に平均再生時間を算出し、 $f(x)=ax^2+bx+c$ の a, b, c の値を求めると表 1 の結果となった。

表 1. 平均再生時間を算出できる係数の値

係数	係数の値	誤差
a	2.89×10^{-4}	$\pm 1.56 \times 10^{-4}$ (54.06%)
b	4.67×10^{-2}	$\pm 6.02 \times 10^{-3}$ (12.87%)
c	7.88×10^{-1}	$\pm 4.97 \times 10^{-2}$ (6.304%)

4. まとめ

本稿では、英語ニュース Web 等の文章を音声合成で読み上げて動的なリスニング・ディクテーション教材をユーザに提供し、英会話・英語ニュース内容に関係する画像をユーザに提供するシステムを提案した。

Power Law of Practice[7]等でシステムを利用したユーザと利用していないユーザでディクテーション問題に解答するのに掛かった時間や、正解率を比較することでシステムの評価実験を行いたい。

参考文献

- [1] 松繁 寿和
英語能力と所得の関係に関する実証分析(IV-1 部会 経済と教育(2)), 日本教育社会学会大会発表要旨収録, No.52(20000915) pp. 191-192
- [2] 杉浦正利, 竹内彰子, 馬場今日子
リスニング能力養成のための自律学習: ディクテーションの効果, Studies in language and culture Vol.23, No.2(20020000) pp. 105-121
- [3] 中川聖一
音声言語処理技術を用いたテレビ放送やビデオ映像からの語学学習教材の半自動作成システム,
The 18th Annual Conference of the Japanese Society for Artificial Intelligence, 2004
- [4] imaizumi
英単語シアター, <http://imagine-lab.net/english/>, 最終更新日: 2007 年 6 月 19 日
- [5] 赤松大典
さしえショー, <http://sashie.jp/sashieshow/>, 最終更新日: 2007 年 9 月 21 日
- [6] (株)ザッパラス, <http://www.zappallas.com/>, LittleFox
- [7] J. Anderson: Cognitive skills and their acquisition, L. Erlbaum Associates, 1981.