

# 英語音声ファイルから発音が近い和文の自動生成 - 空耳自動生成システム -

榎間 祐太 堀 幸一郎 水谷 雄樹 濱川 礼  
中京大学 情報科学部 情報科学科

## 1. はじめに

本論文では、英語音声ファイルから日本語の空耳を自動的に生成するシステムについて述べる。空耳とは、「外国語の音を何らかの意味を持った日本語に聞こえるように工夫する言葉遊びの一種」である。

現在テレビ番組やネット上で空耳事例が多くあるが、面白いものはそれほど多くはない。それは、視聴者から投稿された空耳を紹介し評価するテレビ番組「タモリ倶楽部」で過去 17 年間での最高評価が全体の 2.5% しかない例から鑑みることができる[1]。そこで、我々はユーザが面白い空耳を思い付くことを支援することを目的とし、空耳自動生成システムの開発を行った。

## 2. 本システムの特徴

本システムは、英語の音声ファイルを入力することで、日本語の空耳をテキストで出力する。英語を使用するのは、日本の義務教育で学習され、日本人が日常で触れる機会が多いからである。

空耳には「音の位置がずれる」、「変形していく単語」、「音階の錯覚」、「雑音の七変化」、「背後や頭上を通る音」などがある[2]。我々はその中で「音の位置がずれる」、「音階の錯覚」を元に「発音が小さく聞き取りづらい部分を考慮しない」という特徴が、「変形していく単語」を元に「文字の区切りを都合よく解釈する」という特徴があると考えた。本システムではこの 2 つの特徴を考慮して、日本人が思い付く空耳に近い空耳を生成するために、それぞれ「音声編集機能」と「漢字変換機能」を実装した。また、生成された空耳の信憑性を検証するための機能として「音声比較機能」を実装した。

## 3. システムの全体構成

本システムは、「音声編集・認識部」、「漢字変換部」、「音声比較部」の 3 つの機能から構成されている[図 1]。

Automatic generation of Japanese which pronunciation is near from English voice file

-Mishearing automatic operation generation system-

Yuta Sakakima, Koichiro Hori, Yuki Mizutani

and Rei Hamakawa

Chukyo University Department of information science

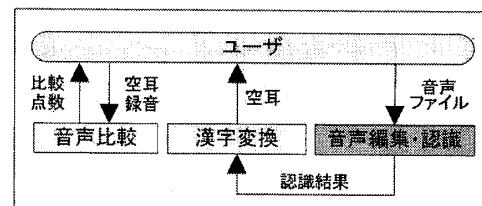


図 1 システム構成図

## 4. 音声編集・認識部

2 章でも述べた「発音が小さく聞き取りづらい部分を考慮しない」という特徴を表現するために、小さな発音が一定区間以上連續する箇所を音声データからあらかじめ削っておく。そして、小さな発音を削った音声データを、日本語音声認識エンジンによって音声認識を行うことで空耳を生成している。

また、音声を削る度合い(カット率)を変化させることで認識結果が変化する。どの認識結果が最も良いかは聞き手によって変化するため、英語音声の認識で良い出力結果が得られることが多かった 3 種類のカット率を設定し、カット A, B, C とした。この 3 つの結果を空耳として出力する。

[表 1]にカット率の具体的な数値を示す。音声波形の中で最大の振幅を max とし、max の a% 以下の値が b/16000s 続いた区間の振幅を 0 とし、音声をカットする。

表 1 音声のカット率ごとの具体的な数値

カット率	a (%)	b (1/16000s)
カット A	2	100
カット B	6	200
カット C	10	400

[図 2]に音声編集を行った際の認識結果を示す。

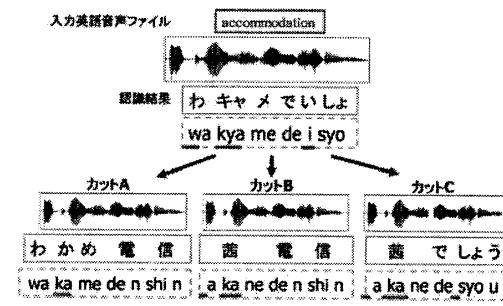


図 2 音声編集による認識結果の変化

## 5. 漢字変換部

2章で述べた、「文字の区切りを都合よく解釈する」という特徴を表現するために、形態素解析ツール MeCab を利用し、音声認識部の結果に対しその文章で考えられる全ての品詞分解の結果を出力する。その結果を日本語の文章に近づけるために品詞ごとに漢字変換を行う。IME API による漢字変換候補リスト取得機能と IME の学習機能を利用することで、ユーザがよく目にする文字の変換候補を取得している。この候補の組み合わせにより、一つの文章から同音異義の文章を複数生成している。[表 2]に漢字変換の例を示す。

表 2 漢字変換例

例文	意味 1	意味 2	意味 3
たいとる	タイトル	鰐怒る	炊いとる
あくまで	空くまで	悪魔で	あ熊で

## 6. 音声比較部

空耳出力結果をユーザが読んで録音する。入力した音声と録音した音声を、ケプストラム分析[3]を用いて音の特徴量を抽出しグラフ化と点数化を行う。この機能により、音声分析の観点から出力された空耳を比較し、聴覚だけでなく視覚でも出力された空耳が正しいと判断できる。ケプストラム分析は FFT ケプストラム分析を利用する。FFT はフレーム長 4096 点とし、分析範囲をフレーム長の  $1/2$  ずつシフトさせながら分析を行う。

続いて、点数の求め方を説明する。まず、1 フレーム長 4096 点を 16 で割り、1 区間 256 点の 16 区間に分割する。

各区間の類似度を以下の式で求める。

$$\cos\left(\frac{2}{3}S(n)\right) \quad 0 \leq S(n) \leq \frac{3}{4}\pi$$

$S(n)$  は類似度を求めるための値で、区間  $n$  の 256 点ずつの平均の差分値の絶対値を取ったものとする。また、3 分の 2 の乗算は、特徴量の値を 268 フレーム分プロット調査した結果、絶対値 2 以下のものが 64.2% となっていることから、差分値が 0 の時 1 を取り、差分値の増加と共に減少し、2 付近で 0 になる曲線を  $\cos$  で近似するために行っている。そして、全体の類似度の平均を取り、百分率を求め点数とする。[図 3]に類似度のグラフを示す。

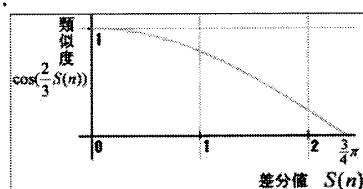


図 3 類似度のグラフ

## 7. ユーザインターフェース

[図 4]に本システムのユーザインターフェース、および空耳の実際の例「automobile」の出力結果を示す。

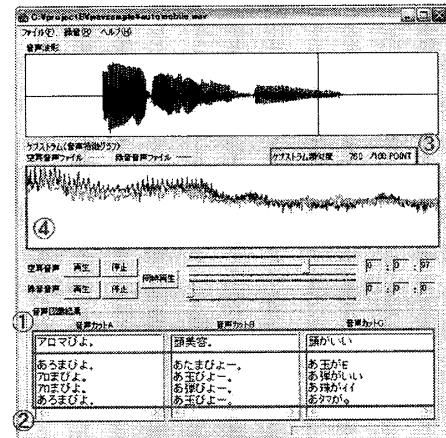


図 4 ユーザインターフェース

- ①音声編集・認識部で生成された空耳
- ②①を漢字変換した結果の表示
- ③FFT ケプストラム分析によるグラフ表示
- ④結果を録音することによって点数表示

## 8. まとめと今後の展望

ユーザ 18 名を対象に本システムを使用した評価を 5 段階評価で行ったところ、「面白い空耳を生成することができたか」という項目で平均 4.2 点という高評価を得ることが出来た。また、あるユーザが「accommodation」という英語の空耳、「わかめ電信」と「茜でしょう」を見て、「わかめでしょう」という新たな空耳を思い付いたことから、ユーザが面白い空耳を思い付くことを支援するという目的を達成することが出来た。

今後の改善点として、長文の認識精度が悪いことがあげられる。これは、音声認識エンジン自体が長文の認識を苦手としていることが原因であり、長文を入力した場合単語ごとに区切って認識させるなどの工夫が必要である。

今後の展望として、出力された空耳を利用して、ネイティブの外国語発音を覚えやすくする学習支援ツールへの利用などが考えられる。

## 参考文献

- [1] 空耳アワーアップデート  
<http://www7a.biglobe.ne.jp/~soramimiupdate/>
- [2] 空耳 -聴覚の錯覚-  
<http://www.brl.ntt.co.jp/IllusionForum/basics/auditory/index.html>
- [3] 小林隆夫：“音声のケプストラム分析、メルケプストラム分析”，電子情報通信学会技術研究報告, Vol.98, No.261 pp. 33-40 (1998-09)