TR-IT-0277

音声合成システム CHATR における 日本語話者による英語の音声合成

田村則和, 藤澤謙, ニック キャンベル

1998 9 18.

ABSTRACT
多言語音声合成システム CHATR で日本語話者による英語の音声合成を行った。まず英語話者により英語音声を合成し、そのケプストラム情報をターゲットとして日本語話者で英語の音声合成を行った。

Contents

Chapter 1

始めに

現在エイ・ティ・アール音声翻訳通信研究所では、音声合成研究の一環として多言語音声合成システム CHATR の研究を行なっている。

このシステムは、大量の音声データベースの中から最も適切な音声波形を抽出し接続するものであり、信号処理を行わないために自然に近い音声が得られる。

今まで音声合成をする際、日本語文章内に英語が混在するとき、英語の部分だけを英語話者で出力していた。これでは話者の声が異なる部分が入るため、音質的にも話者の特徴的にも違和感が持たれた。

そこでこの問題を解決するため、日本語話者で英語を出力する方法が提案されている [1]。本稿ではまず、英語話者による英語音声を合成し、その音声波形のケプストラム情報を得る。それをターゲットとして日本語話者の音声データベースより音素を選択し、CHATR でつなぎ合わせて英語音声を合成した。

Chapter 2

日本語話者による英語の音声合成の概要

2.1 解説

日本語話者での英語出力の方法は2通り考えられる。

方法1: カタカナ英語

方法2: ネイティブに近い英語

前者は発話にふさわしい音素を決定するため、各英語の音素ごとに対応させたマッピングテーブルを使用して、日本語の音素表記にする。例えば、"interesting" という単語をまず英語の音素表記にし (ih・n・t・r・ax・s・t・ih・ng)、マッピングテーブルを使用して日本語の音素表記にする (i・n・t・r・u・s・t・i・N)。

ところが日本語の母音空間と英語の母音空間には差があるため、日本語の音素表記では英語の音素表記をすべてカバーしきれない (Figure 2.1 2.2)。例えば "cup" と "cap" の母音のように日本語では区別されてない音が同じ音素 /a/ になってしまい、英語初心者がしゃべるカタカナ英語の発話になってしまう。

後者は適当なターゲット情報を用意し、それにより近い音素を日本語の音声データベースから選択する。前者と違って日本語の音素 /a/ の中でも区別され、ネイティブに近い英語発話となることが期待される。

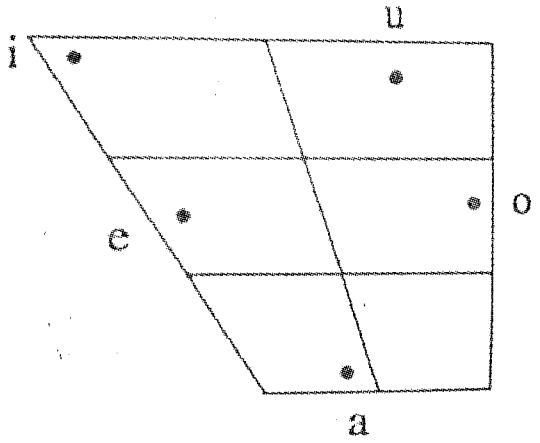

Figure 2.1: 日本語の母音空間

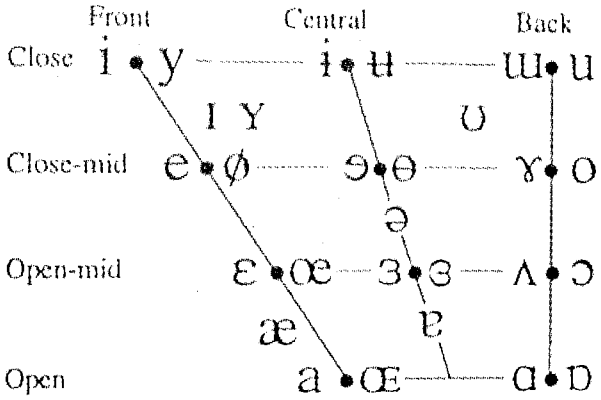

Figure 2.2: 英語を含む母音空間

2.2 アルゴリズム

以下の 2 段階で音声合成を行う。CHATR で行っている詳細は Appendix A.2 を参照。

1. 英語 DB による英語の音声合成

   英語文章を入力して英語合成音声を CHATR で作る。そこで選ばれた音素に対応するフレームごとのケプストラムへのインデックスをターゲット情報として保存しておく。(Figure 2.3)

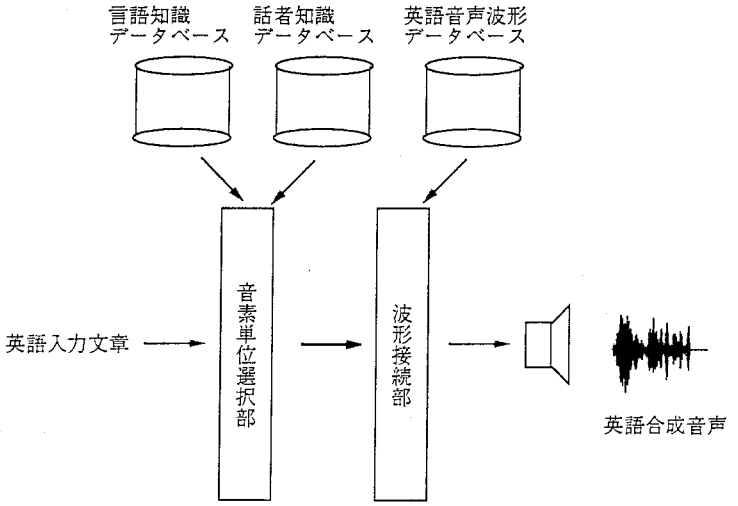

Figure 2.3: 英語 DB による英語の音声合成

2. 日本語 DB による英語の音声合成

conv_romaji で英語の音素表記を日本語の音素表記に変換した音素列を CHATR に入力する。

1 で得たケプストラムをターゲットとして、音素単位候補とのケプストラム距離を計算する。

ケプストラム距離の最小なものを音素として選択し音声合成を行う。(Figure 2.4)

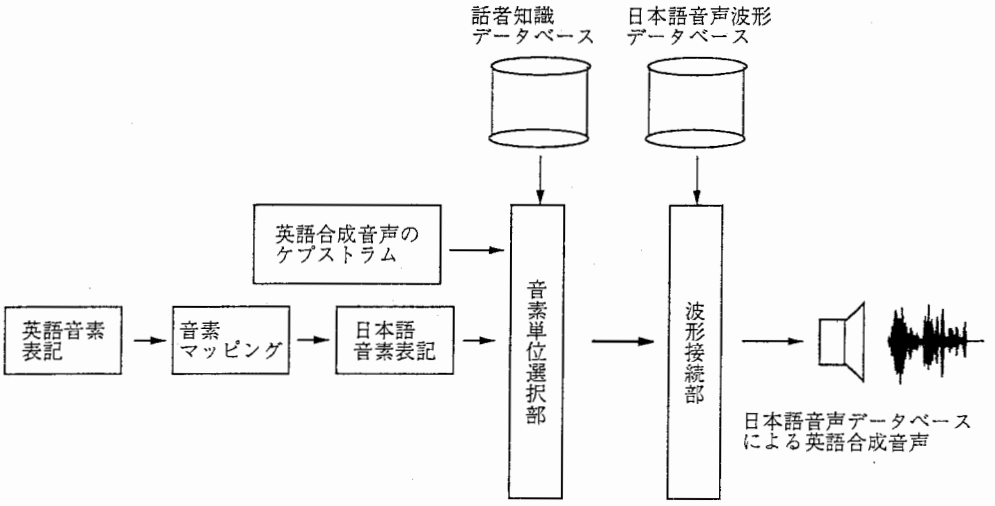

Figure 2.4: 日本語 DB による英語の音声合成

2.3 音素ラベル変換

conv_romaji を用いて英語の音素表記を日本語の音素表記に変換する。conv_romaji の詳細は Appendix A.1 を参照。

2.3.1 マッピングテーブル

英語音素列を日本語音素列に変換するマッピングテーブルを以下に示す。

英語	日本語	英語	日本語	英語	日本語
ax	u	axr	a	aa	aa
ao	a	ah	a	ay	ai
aw	au	ae	a	ea	ea
ia	ia	ua	ua	el	l
en	N	er	a	eh	e
ey	ei	iy	ii	ih	i
uh	u	uw	uu	em	m
oh	o	ow	oo	oy	oi
y	y	r	r	l	r
m	m	n	n	ng	N
nx	N	jh	j	ch	chi
zh	z	sh	sh	th	s
dh	z	p	p	b	b
d	d	dx	d	t	t
k	k	g	g	f	f
v	b	z	z	s	s
hh	h	w	w	sil	#
brth	@	brth			

2.4 ケプストラム距離

適切なターゲット情報により近い音素を選択するため、ターゲットのケプストラムベクトルと音素単位候補のケプストラムベクトルとのケプストラム距離を比較対象とする。

2.4.1 ケプストラム距離の計算

ターゲット波形1音素のフレーム数を M_t、フレーム t のケプストラムを

$$c_{1,t}(0), c_{1,t}(1), c_{1,t}(2), \ldots, c_{1,t}(M_t)$$

合成波形1音素のフレーム数を M_u、フレーム t のケプストラムを

$$c_{2,t}(0), c_{2,t}(1), c_{2,t}(2), \ldots, c_{2,t}(M_u)$$

とした時、フレームごとの自乗ケプストラム距離は

$$d(t) = \sum_{k=0}^{M_t} (c_{1,t}(k) - c_{2,t}(j))^2$$

となる。ここで j は $k * \frac{M_u}{M_t}$ の整数値をとる。全体のケプストラム距離は

$$d = \frac{1}{M_t} \sum_{t=0}^{M_t} d(t)$$

となる。

2.4.2 計算結果

英語音素列を日本語音素列に変換するマッピングテーブルにおいて、その変換が適切かどうかをみるため、1つの例をあげる。

ターゲット波形音素として、英語話者の音声データベース中にある音素 /ax/ を選択した。合成波形音素として、日本語話者の音声データベース中にある音素 /u/ と /m/ を選択した。この時、日本語話者の音声データベース中に出てくる全ての音素を対象としてターゲット波形音素とのケプストラム距離を計算し、その中の最小なものを以下に示す。

- 例 1

  ターゲット波形音素　：ax
  合成波形音素　　：u

  ケプストラム距離　　：216.818

- 例 2

  ターゲット波形音素　：ax
  合成波形音素　　：m

  ケプストラム距離　　：624.296

Chapter 3

評価

3.1 方法

日本語 DB による英語の合成音声がいかに英語 DB による英語の合成音声に近いかを 5 段階で評価した。音声試料、話者、被験者は以下のとおりである。

音声試料　ランダムに選択した英単語を以下の方法で合成
1. conv_romaji で変換した音素表記で合成
2. ケプストラムをターゲットとして合成

話者　日本語男性話者 (MHT)

被験者　1 人 (田村)

3.2 結果

実際の音声試料は Appendix B を参照。

日本語の音素列　: conv_romaji で英語の音素列を日本語の音素列に変換したもの
conv_romoji　: conv_romaji で変換した音素表記で合成した評価
cepstrum　: ケプストラムをターゲットとして合成した評価

英単語	日本語の音素列	conv_romoji	cepstrum
interesting	intrustiN	2	3
forever	furebu	1	1
trouble	trabr	2	2
difficult	difikurt	2	2
rump	ramp	3	4
lamp	ramp	3	4
expect	ikspekt	2	2
shopping	shopiN	3	4
deep hole	diip hoor	3	4
your family	yo famuri	2	3
平均		2.3	2.9

極めて良い (5) – 極めて悪い (1)

Chapter 4

まとめと今後の課題

4.1 まとめ

ケプストラムをターゲットとして音声合成を行った結果、conv_romaji で変換した音素列の合成音よりは英語に近く聞こえた。特に、conv_romaji では同じ音素列となる rump と lamp ではケプストラムをターゲットとしたほうは、違いがよくわかった。

ただし、conv_romaji を基本として音声合成を行っているため、conv_romaji の変換であまりにも英語とかけはなれているとケプストラムをターゲットとしても意味がない。

4.2 今後の課題

conv_romaji でうまく変換されない英単語をどうやって出力するかが問題となる。

- conv_romaji 内のマッピングテーブルを改正する。その後にケプストラムをターゲットとして音声合成を行う。
- 英語 1 音素に対し日本語の音素複数個を対応させる。そして、CHATR 内で選ばれたそれぞれの音素単位候補とのケプストラム距離を計算し、その中で最小なものを音素として選択する。例えば、/ax/ に対して /u/, /a/, /i/ の 3 音素を対応させる。

Bibliography

[1] ニック キャンベル : “外国語の音声合成について”, 日本音響学会講演論文集, Sep. 1998

[2] 佐藤俊則 , “今日からあなたも chatr ユーザー” ,
http://www.itl.atr.co.jp/local_info/department/dept2/dvi/jchatr.dvi,
April 8th 1996

Chapter 5

謝辞

本研究の遂行にあたり、多くの御指導、御助言を頂いたニック キャンベル室長、直接指導者の藤澤謙研究員他第二研究室の皆様、ATR 音声翻訳通信研究所の皆様に感謝します。

1998 年 9 月 18 日
田村 則和

Appendix A

使用したプログラムおよびファイル

A.1 conv_romaji

変更点

以前までは、chatr-0.91 のライブラリをリンクする方法をとっていたが、chatr-0.94 のライブラリをリンクできるように、jlts_lookup() 関数を mlts_lookup() に変更した。

また default で入力した文章をローマ字変換して出力していたが、日本語の音素列に変換して出力するようにした。

更に、不必要なデバッグ情報は表示しないようにした。

変更前の conv_romaji は以下のパスにある。

```
/home/as53/simomura/SunOS/Chatr/xphwang/conv_romaji/conv_romajiSunOS
```

変更後の conv_romaji は以下のパスにある。

```
/home/as60/xntamura/conv_romaji/conv_romaji
```

プログラムの解説

```
conv_romaji : [-debug] [-nokey] [-kakasi]
                  [-f <input filename> ] [-E][-J][-K][-G]
```

- 機能
  入力されたファイル中の文を CHATR で合成するリスプコマンドに変換して出力する。(ファイルの中の文は、英語、日本語、韓国語、ドイツ語のいずれか。)

  -debug : デバッグ情報を出力。

  -nokey : 言語を表すキー (例: $/E/$) の出力を抑える。

-E ：英語の音素表記に変換。

-J ：日本語の音素表記に変換。

-K ：韓国語の音素表記に変換。

-G ：ドイツ語の音素表記に変換。
default は、日本語の音素表記に変換。

-kakasi :?

- 使用例：英語から英語

```
csh > more test.txt
ATR Interpreting
csh > conv_romaji -f test.txt -E
(Say (Synth (Utterance Phoneme(/E/ 3  ihnter1prihtihng ))))
```

- 使用例：英語から日本語

```
csh > conv_romaji -f test.txt t4.txt
(Say (Synth (Utterance Phoneme(/E/ 3  inta'pritiN ))))
```

- 使用例：英語 + 日本語から日本語

```
csh > more test2.txt
hello こんにちは
csh > conv_romaji -f test2.txt
(Say (Synth (Utterance Phoneme(/E/ 3  huroo' /J/ 3 koNnichiwa 5
/E/ 3  ))))
```

A.2 CHATR

変更点

英語話者で英文章の合成音声を作る際、リスプコマンド

```
chatr> (speaker_nes)
nes
chatr> (SayText "合成したい英文章を入力 ")
```

を入力する。そこで選ばれた音素に対応する波形へのインデックスを保存できるようにした。リスプコマンド

```
chatr> (Save UnitLabels2 "-")
Save file ? ok -> 1, no -> 0
```

を入力すると、保存するかどうか聞いてくるので、ok の 1 を入力する。保存ファイルは以下のパスにある。

```
/home/as60/xntamura/chatr-0.94_2/src/cep_target_data/wavdata
```

また、英文章を保存するため、リスプコマンド

```
chatr> (Save XWords "-")
Save string and conv_romaji ? ok -> 1 , no -> 0
```

を入力すると、保存するかどうか聞いてくるので、ok の 1 を入力する。保存ファイルは以下のパスにある。

```
/home/as60/xntamura/chatr-0.94_2/src/cep_target_data/stringdata
```

この保存ファイルを使い、conv_romaji で日本語の音素列をファイルに保存する実行ファイル、日本語の音素列を保存したファイルは以下のパスにある。

```
/home/as60/xntamura/chatr-0.94_2/src/cep_target_data/play_conv_romaji
/home/as60/xntamura/chatr-0.94_2/src/cep_target_data/phonemedata
```

次に日本語話者に切り替える。

```
chatr> (speaker_MHT)
MHT
```

保存された波形のケプストラムをターゲットとし、音声合成を行うため

```
chatr> (load "/home/as60/xntamura/chatr-0.94_2/src/cep_target_data/nus_params")
chatr> (load "/home/as60/xntamura/chatr-0.94_2/src/cep_target_data/phonemedata")
```

を入力する。

変更した CHATR は以下のパスにある。

```
/home/as60/xntamura/chatr-0.94_2/src/main/chatr
```

Appendix B

結果

各英単語の音声ファイルは以下のパスにある。フォーマットは AIFF で保存してある。

英語話者 nes による英語合成音声

`/home/as60/xntamura/test_data/English/{1,2, ... ,14}_E-test.wav`

conv_romaji で英語の音素列を日本語の音素列に変換した、日本語話者 MHT の英語合成音声

`/home/as60/xntamura/test_data/Conv_romaji/{1,2, ... ,14}_Conv-test.wav`

ケプストラムをターゲットとした、日本語話者 MHT の英語合成音声

`/home/as60/xntamura/test_data/Cepstrum/{1,2, ... ,14}_Cep-test.wav`

1 から 14 までの数字に対応した英単語を以下に示す。

1. interesting
2. forever
3. trouble
4. dificult
5. appointment
6. expect
7. shopping
8. deep hole
9. your family
10. I am a student.

11. ramp
12. lamp
13. rump
14. lump

サンプルは以下の URL を参照。

http://www.itl.atr.co.jp/%7Efujisawa/DEMO/CepTarget/index.html