A Taiwanese Text-to-Speech System with Applications to Language Learning

Min-Siong Liang[1], Rhuei-Cheng Yang[2], Yuang-Chin Chiang[3], Dau-Cheng Lyu[1], Ren-Yuan Lyu[2]
[1.] *Dept. of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan*
[2.] *Dept. of Computer Science and Information Engineering, Chang Gung University,Taiwan*
[3.] *Inst. of Statistics, National Tsing Hua University, Hsin-chu, Taiwan*
E-mail: {siong,gang}@msp.csie.cgu.edu.tw, rylyu@mail.cgu.edu.tw Tel: 886-3-2118800ext5967

Abstract

The paper describes a Taiwanese Text-to-speech (TTS) system for Taiwanese language learning by using Taiwanese / Mandarin bilingual lexicon information. The TTS system is organized as three functional modules, which contain a text analysis module, a prosody module, and waveform synthesis modules. And then we set an experiment to evaluate the text analysis and tone-sandhi. A 89% labeling and 65% tone-sandhi accuracy rate can be achieved. With adopting proposed Taiwanese TTS component, talking electronic lexicon system, Taiwanese interactive spelling Learning tool and Taiwanese TTS system can be built to help those who want to learn Taiwanese.

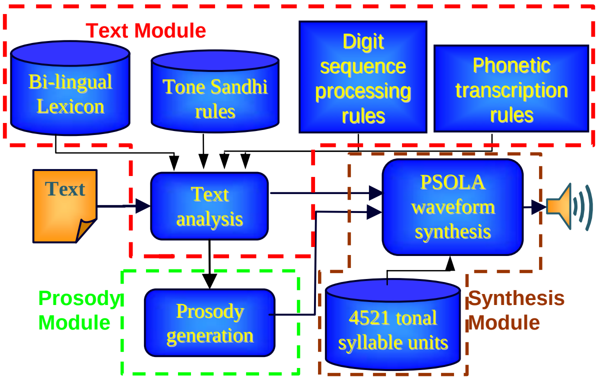
1. Introduction

Speech-based technologies interface is a trend toward future e-Learning. [1] In Taiwan, Taiwanese is one of three major languages (Mandarin, Taiwanese and Hakka) and is widely used as the native tongue of more than 75% population in Taiwan. Unfortunately, due to lack of elementary education for this language, most people can not read or write Taiwanese although they speak and listen to it every day. In recent years, Taiwan government start to pay much more attention to mother-tongue education, and made more effort and budget for it. But learning Taiwanese has at least two problems: one is that the Taiwanese articles or teaching materials are few in comparison with Mandarin, the other is that most Taiwanese texts consist of Chinese characters and English characters, which most people do not know how to read. Therefore, it is a better way for learning Taiwanese that input Mandarin text and output Taiwanese speech by a TTS system.

In this paper, we attempt to construct a Taiwanese TTS system, which should be able to tranform any modern Mandarin or Taiwanese articles into Taiwanese speech for reading out.

Since Taiwanese is a tonal language, some special processes about the tone-sandhi rules should be considered. [3] Besides, the system also adopts TD-PSOLA to modify the waveform by adjusting the prosody parameters of selected units so that the synthesis speech sounds more natural. This TTS system is composed of 3 major functional modules, namely a text analysis module, a prosody module, and a waveform synthesis module. The system architecture is shown as in <fig.1>.

This paper is organized to describe all 3 major modules in detail and the evaluation of text analysis and tone-sandhi in the following sections, and finally a discussion, application and conclusion are given.



<fig 1> The TTS system flow chart

2. Text analysis module

In spite of the fact that we have many experiences on dealing with Taiwanese text, it is still difficult to transcribe Mandarin text into Taiwanese text [2][3][4][5][6]. The major reason is that Taiwanese has not been assigned as an official language historically and the written form is not consistent at all. However, due to the construction of bilingual lexicon, this work becomes easier. In the following paragraphs, we will describe text analysis in detail.

2.1. The Formosa Phonetic Alphabet (ForPA)

The Mandarin Phonetic Alphabet (MPA, also called Zhu-in-fu-hao) and Pinyin (Han-yu-pin-yin) are the most widely known phonetic symbol sets to transcribe Mandarin Chinese. They have been officially used in Taiwan and Mainland China respectively for a long time. However, both two systems are designed only for Mandarin. It's necessary to design a more suitable phoneme set to begin with multilingual speech data collection and labeling. An example of ForPA is listed in <Table 1> [7].

2.2. Word Segmentation and Mandarin-Taiwanese transcription (Sentence-to-word)

Since there is no natural boundary between two successive words, we must segment text into word sequence first. We use Mandarin-Taiwanese bi-lingual lexicons for text analysis. Each item in the lexicons contains a Chinese character string, which is transcribed into Mandarin with Formosa Phonetic Alphabet (ForPA). There is at least a Taiwanese word corresponding to a Mandarin word [7]. Every word in Taiwanese has at least two pronunciations, containing literature (classic) and oral pronunciations. The statistics of Mandarin-Taiwanese lexicon is shown as <table.2>.

We use the bilingual pronunciation dictionary as the knowledge source and then apply a word segmentation algorithm based on the sequentially maximal-length matching in the lexicon.

2.3. Labeling (morpheme-to-phoneme)

For each segmented word, there may exist not only one pronunciation. To deal with the multiple-pronunciation problem, two strategies are adopted. One is the oral pronunciation has priority for transcription. Another is that build a network with pronunciation frequencies as node information and pronunciation transitional frequencies as arc information has been constructed for each sentence. The best pronunciation is then conducted by Viterbi search.

2.4. Normalization of the digit sequences

Another important issue for text analysis is the normalization of the digit sequences. In fact, each of almost Taiwanese single-syllabic words has 2 distinct manners of pronunciation: one for classic literature like poems, and the other for oral expression in daily lives. However, for digits, these 2 manners of pronunciation exist in daily lives. The manner of pronunciation depends on the position of the digit in a sequence, which can be summarized in rules. In addition, if a digit sequence does not represent a quantity, it is pronounced digit by digit as the classic pronunciation.

3. Prosody analysis module

Like Mandarin, Taiwanese is a tonal language. Traditionally speaking, it has seven lexical tones, two of which are carried in syllables ended with stop vowels, such as /ak/ and /ah/ (called entering-tone traditionally) and the other five are carried in those without stop-vowels (called non-entering tone traditionally). Let's define the number 1 to 7 to encode the 7 Taiwanese tones as follows: "1" High-Level（like東），"2" Mid- Level （like洞），"3" Low-Falling （like棟），"4" High-Falling （like黨），"5" Mid-Rising（like同），"6" High-Stop（like獨），"7" Mid-Stop（like督）.An example of these 7 tones with one corresponding Chinese character for each tone is shown in <table.3>. Some phonetic/acoustic characteristics, including contour of fundamental frequency (*F0*), the description of relative frequency level (*RF*), and the proposed tone-to-digit (*TD*) mapping are also shown. In this table, one can also find 2 additional tones, namely "8" Low-Stop and "9" High-Rising, which are necessary for tone-sandhi issue discussed in next paragraph.

The tone sandhi issue is relatively complex in Taiwanese. Every Taiwanese syllable has 2 kinds of tones called the lexical-tone and the sandhi-tone depending on the position it appears in a word or a sentence. One of the most frequently referred sandhi rules says that , for most cases, if a syllable appears at the end of a sentence, or at the end of a word, then it is pronounced as its lexical tone, otherwise, it is pronounced as its sandhi tone[2]. The sandhi rules for each lexical tone is as follows:

(1) tone "1" will change to tone "2";
(2) tone "2" will change to tone "3";
(3) tone "3" will change to tone "4";
(4) tone "4" will change back to tone "1";
(5) tone "5" may change to tone "2" or tone "3" for two different major sub-dialects;
(6) tone "6" will change to tone "8";
(7) tone "7" will change to tone "6".

The above is summarized in <fig.2>, which is called the "tone sandhi sailboat".

Other finer aspect like triple adjective, where the first character of 3 duplicative adjectives will carry a very different tone other than the traditional 7 lexical tones mentioned previously. We map such a "High-Rising" tone to digit "9", and call it tone "9". The tone sandhi rules for triple adjectives are summarized in <table.4>.

4. The evaluation of text analysis and prosody module

After the progress of text analysis and prosody, an experiment is set to evaluate the performance. The main target of the experiment examines accuracy rate of automatic transcription, which produced text analysis and prosody modules, in comparison with manual transcription. The evaluation can be organized to three stages mentioned below:

Stage 1: collect abundant news from internet. The choice of the news has no bias on special categories as possible. Sentences longer than 20 Chinese characters are removed. In the end, the total news contains 7,573 articles, 169,040 sentences and statistics are shown in <table 5>.

Stage 2: choose a set to cover all distinct Chinese characters and minimize the number of sentences from 169,040 sentences. Due to time constraints, we just choose preceding 200 sentences for manual transcription by two Taiwanese linguistics experts. As shown in fig 3, the 200 sentences cover 41% of all distinct Chinese character, which occur in candidate articles.

Stage 3: compare the automatic transcription with manual transcription. There are three kinds of results, which are word segmentation evaluation, labeling evaluation and tone-sandhi evaluation. The results of these evaluations are presented as <table 6>.

From the <table 6>, we find the system can segment and transcribe most article accurately into Taiwanese word and reach over 97% accuracy. If we do not consider tone-sandhi, the system can transcribe article into correct pronunciation close to the 88% rate and the most errors happen in names and out of vocabulary. Because the Taiwanese has uniform tone-sandhi rules, it is acceptable that the accuracy rate of tone-sandhi is lower.

5. Waveform synthesis module

Before we explain operation of synthesis module in the system, it is necessary to denote what INITIAL/FINAL is. An INITIAL/FINAL format can describe the composition of Taiwanese syllable. INITIAL is the initial consonant and FINAL is the vowel (or diphthong) part with an optional medial or a nasal ending [10].

There are many variety of synthesis method. We adopt the most popular method TD-PSOLA to modify the prosodic feature of selected units [8].One of the preliminary task to mark pitch period for tonal syllables. In order to finish pitch mark, we apply an algorithm, which find a pitch period first and then label all local maximums within pitch period in voiced part. [9] Those local maximums are taken for pitch mark. With pitch mark, all tonal syllables can be segmented as .a succession of synthesis components. As shown in <fig 4>, Synthesis components are used to not only raise or lower pitch but also enlarge or shrink duration.

Furthermore, after the analysis of tonal syllables, we can gather duration and short pause information in each syllable. By the information, the synthesis speech will be accomplished in below cases:

Case 1: if the syllable consist of unvoiced consonant (p-, t-, g-, k-, z-, s-, c-, h-), the system just modify duration of the unvoiced INITIAL, and modify duration and pitch of FINAL.

Case 2: the system will modify duration and pitch both on INITIAL and FINAL if there do not exist unvoiced INITIAL.

Case 3: replace the short pause with a zero-value section.

6. Applications to Taiwanese Language Learning

By adopting proposed Taiwanese TTS system, a Taiwanese talking electronic lexicon can be built. We can input Taiwanese or Mandarin words, and then the output is a list of Taiwanese words associated with the input words. The interface of talking electronic lexicon system is shown as <fig 5>. By filling out any Chinese word in top left blank space, the bottom left memo will list other candidate words in bottom right area. In top right area, we can press the lexical-tone or sandhi-tone button to play the pronunciation of the word. Therefore, the electronic talking lexicon system is a good and friendly tool to support those who want to learn Taiwanese pronunciation or write Taiwanese articles. In addition, the extension of the Taiwanese talking electronic lexicon, we can play any Chinese documents in Taiwanese to aid those who just understand Taiwanese. The system interface is shown as <fig 3>.

On the other hand, as mentioned about Section 2.1, the ForPA is a more suitable phonetic alphabet set for Taiwanese. Therefore, in order to spread ForPA, it is necessary to construct a Taiwanese interactive phonetic alphabet learning tool, which consists of Taiwanese TTS component. When we type various kinds of existing Taiwanese syllables in ForPA, the tool will pronounce simultaneously. By the tool we can learn a new language phonetic system more quickly. The <fig 6> shows the interface of interactive phonetic system learning tool.

7. Conclusion

As shown in <fig.4>, we have successfully constructed a Taiwanese TTS system from bi-lingual for contextual learning. Hence, the most Mandarin article can be transcribed into Taiwanese and automatic generation of a speech signal in Taiwanese. This is great helpful for those who want to learn mother-tongue language in Taiwan as shown <fig 5><fig 6>.

However, there are still a lot to do. In the future, we should improve tone-sandhi for more accurate speech synthesis. In the other hand, it is imperative to use signal processing techniques to smooth the waveform to reduce discontinuity in our future TTS system.

8. Reference

[1] Walsh, P., J. Meade., "Speech enabled e-learning for adult literacy tutoring", The 3rd IEEE International Conference on Advanced Learning Technologies, 9-11 July 2003, Page(s): 17 -21.

[2] Ren-yuan Lyu, Zhen-hong Fu, Yuang-chin Chiang, Hui-mei Liu, "A Taiwanese (Min-nan) Text-to-Speech (TTS) System Based on Automatically Generated Synthetic Units", *ICSLP2000*, Oct. 2000

[4] Ren-yuan Lyu, Chi-yu Chen, Yuang-chin Chiang, Min-shung Liang, "A Bi-lingual Mandarin/Taiwanese(Min-nan), Large Vocabulary, Continuous Speech Recognition System Based on the Tong-yong Phonetic Alphabet (TYPA)", *ICSLP2000*, Oct. 2000, Beijing, China

[5] Yuang-chin Chiang, Zhi-siang Yang, Ren-yuan Lyu, "TAIWANESE CORPUS COLLECTION VIA CONTINUOUS SPEECH RECOGNITION TOOL", *ICSLP2000*, Oct. 2000, Beijing, China

[6] Dau-cheng Lyu, Min-siong Liang, Yuang-chin Chiang, Chun-nan Hsu, Ren-yuan Lyu,"Large Vocabulary Taiwanese (Min-nan) Speech Recognition Using Tone Feature and Statistical Pronunciation Modeling" , *Proceedings of 8th European Conference on Speech Communication and Technology (EuroSpeech 2003*), Sep 1-4, 2003, Geneva, Switzerlan

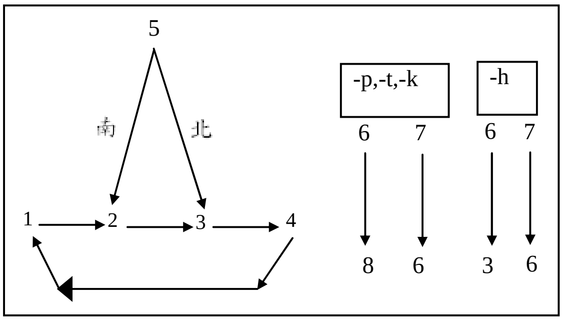
[7] Min-siong Liang, Ren-yuan Lyu, Yuang-chin Chiang "An Efficient Algorithm to Select Phonetically Balanced Scripts for Constructing A speech Corpus", *Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE-NLPKE 2003)*, October 26-29, 2003, Beijing, China

[8] Donovan, R. E. and P. C. Woodland, "A hidden markov-model-based trainable speech synthesizer", Comp. Speech & Lang., 1999.

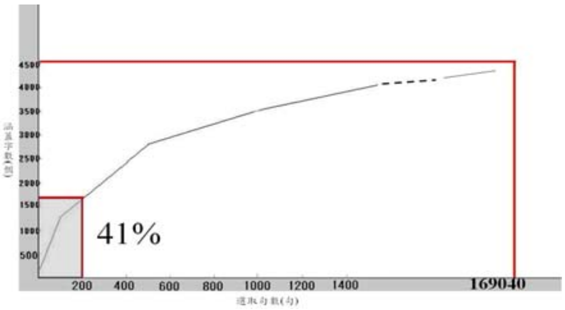
[9]Yuang-chin Chiang, Ren-zyun Chen, Ming-jie Tian, Ren-yuan Lyu, "DIMSU: A Speech Database with Pitch Marks", *Proceedings of Oriental COCOSDA 2003*, Oct 1-2, 2003, Sentosa, Singapore

[10] Fu-chiang Chou, Chiu-yu Tseng, "Corpus-based Mandarin Speech Synthesis with Contextual Syllabic Units Based on Phonetic Properties", ICASSP98

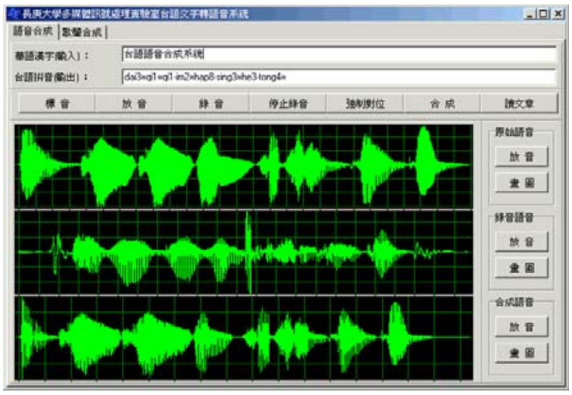
Tables and Figures



<fig.2> The Taiwanese tone sandhi rules



<fig.3> The coverage rate of 200 sentences to the total sentences with respect to distinct Chinese character

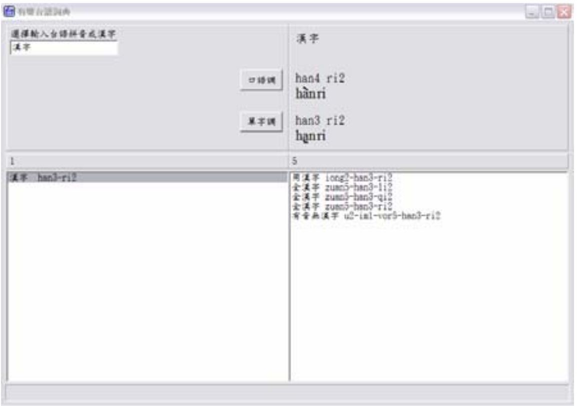


長庚大學多媒體訊號處理實驗室台語文字轉語音系統

華語漢字(輸入)： 台語語音合成系統

台語拼音(輸出)： dai3=gi1=gi1-im2=hap8-sing3=he3-tong4=

<fig 4> The interface of Taiwanese TTS system



選擇輸入台語拼音或漢字

漢字

漢字

口語調 han4 ri2
hànri

單字調 han3 ri2
hanri

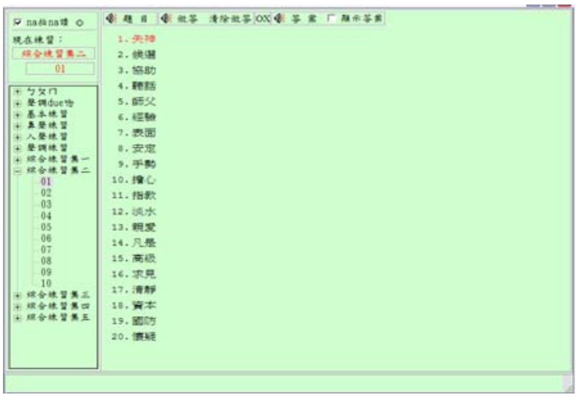
1

漢字 han3-ri2

5

用漢字 iong2-han3-ri2
全漢字 zuan5-han3-li2
全漢字 zuan5-han3-qi2
全漢字 zuan5-han3-ri2
有音無漢字 u2-im1-vor5-han3-ri2

<Fig 5>The interface of Taiwanese electronic talking lexicon



現在練習：
綜合練習集二
01

- ㄅㄆㄇ
- 變調due怖
- 基本練習
- 鼻聲練習
- 入聲練習
- 變調練習
- 綜合練習集一
- 綜合練習集二
  - 01
  - 02
  - 03
  - 04
  - 05
  - 06
  - 07
  - 08
  - 09
  - 10
- 綜合練習集三
- 綜合練習集四
- 綜合練習集五

題目 做答 清除做答 OX 答案 □ 顯示答案

1. 先神
2. [illegible]
3. 協助
4. 聽話
5. 師父
6. 經驗
7. 表面
8. 安定
9. 手勢
10. 擔心
11. 指數
12. 淡水
13. 親愛
14. 凡是
15. 高級
16. 求見
17. 清靜
18. 資本
19. 國防
20. 懷疑

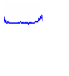
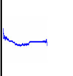
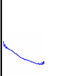
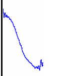
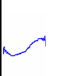
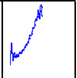
<Fig 6>The interface of Taiwanese interactive spelling Learning tool

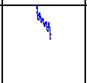
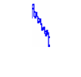
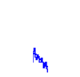
ForPA	Syllable (字)	IPA	Pinyin	MPA
i	i(一)	i	y;i	ㄧ
u	u(吳)	u	w;u	ㄨ
yu	yuan(原)	y	yu;ü	ㄩ
ii	zii(資)	ɨ	i	
-nn	ann(餡TH)	ã		
-p	ap(壓TH)	-p		

<Table 1>: The partial example of the phone set for languages in Taiwan, decoded in four different phonetic alphabet including ForPA, IPA, MPA, and Pinyin. An example of syllable and Chinese character （字） are also shown in the second column

	LP-Taiwanese	OP-Taiwanese	Total
1-Syl	2319	8040	10359
2-Syl	21337	49222	70559
3-Syl	7163	11367	18530
4-Syl	55	15525	15580
5-Syl	1	711	712
6-Syl	0	497	497
7-Syl	0	478	478
8-Syl	0	195	195
9-Syl	0	3	3
10-Syl	0	20	20
Total	30875	86060	116935

<Table 2>: The number of pronunciation of bi-lingual Lexicons, including literature pronunciation (LP) and oral pronunciation (OP) in Taiwanese (Syl: syllable)

ForPA	dong 1	dong 2	dong 3	dong 4	dong 5	dong 9
Ch	東	洞	棟	黨	同	
F0						
RF	HL	ML	LF	HF	MR	HR
TD	1	2	3	4	5	9

ForPA	dok6	dok7	dok8
Ch	獨	督	
F0			
RF	HS	MS	LS
TD	6	7	8

<Table 3> ForPA: Formosa Phonetic Alphabet, Ch: an example Chinese Character, F0: the fundamental frequency contour, RF: relative frequency level, H: High; M: Middle; L: Low, R:Rising; F: Falling; S: Stop

lexicical	1	2	3	4	5	6	7
sandhi-tone	9	9	4	1	9	9	6

<Table.4> The tone sandhi rules for triple adjectives

Duration	2002/8/19~2003/5/27
# of news	7573
# of total sentences	169,040
# of distinct CC	4513

<Table 5> The statistics of candidate news, where Duration means the period of those news, CC denote Chinese character and # denote number.

	Expert1	Expert2
Word Seg & Transfer	97.80%	98.76%
Labeling	89.96%	88.27%
Tone-sandhi	65.43%	62.43%

<Table 6> The statistics of performance in parts of word segment and transfer, labeling and tone-sandhi.