A Bi-lingual Mandarin-To-Taiwanese Text-to-Speech System

Min-Siong Liang, Ke-Chun Chuang, Rhuei-Cheng Yang*, Yuang-Chin Chiang†, Ren-Yuan Lyu**

Dept. of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan
*Dept. of Computer Science and Information Engineering, Chang Gung University,Taiwan
†Inst. of Statistics, National Tsing Hua University, Hsin-chu, Taiwan
siong@msp.csie.cgu.edu.tw, rylyu@mail.cgu.edu.tw Tel: 886-3-2118800 ext 5709

Abstract

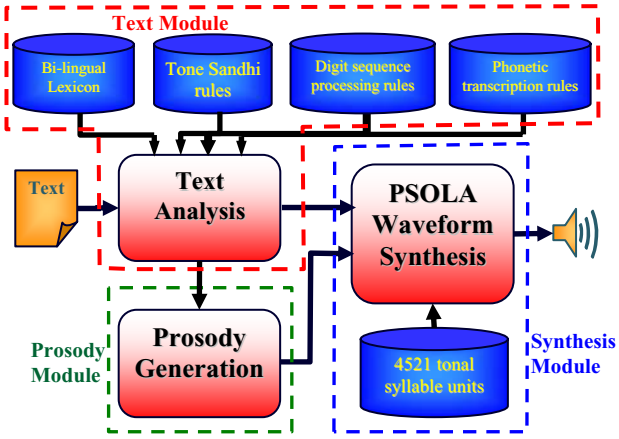
In this paper, we describe a bi-lingual Text-to-speech (TTS) translational system, which converts Chinese Text into Taiwanese speech, by using bilingual lexicon information. This TTS system is organized as three functional modules, which contain a text analysis module, a prosody module, and a waveform synthesis module. We conducted an experiment to evaluate the text analysis and tone-sandhi modules. About 90% labeling accuracy and 65% tone-sandhi accuracy can be achieved. With this proposed Taiwanese TTS component, a talking electronic lexicon system can be built to help those who want to learn these 2 languages.

1. Introduction

There exist three languages in Taiwan, including Mandarin, Taiwanese and Hakka. Over 90% people use Mandarin, but Mandarin in Taiwan is quite different from Beijing Mandarin. This situation is quite similar to that between British English and American English. On the other hand, there are about 70% people whose mother-tongue is Taiwanese. Taiwanese originated from Min-nan, which is widely used in South China, Taiwan, Hai-nan and overseas. Unfortunately, due to lack of elementary education for this language, most people can not read or write Taiwanese even though they speak and listen to it very well. Therefore, there is not much linguistic resource for Taiwanese. Besides Mandarin and Taiwanese, there are 10% people who speak Hakka. Like Taiwanese, Hakka linguistic resource is very few. Most people in Taiwan are at least bi-lingual, while some of them are even trilingual. In recent years, Taiwan government started to pay much more attention to mother-tongue education especially in Taiwanese and Hakka. Since Taiwanese text material is few, it is a good idea for learning Taiwanese by inputting Mandarin text into a TTS system and output Taiwanese speech. Although there are many TTS systems in Mandarin [1][2][3], but there are few TTS systems TTS systems constructed for Taiwanese [4][5].

It is very difficult to translate Mandarin text into Taiwanese text. The major reason is that Taiwanese has not an official written form. It exist three main written forms of Taiwanese nowadays including full-Chinese (FC) characters (e.g. he is very happy today, /伊今仔日心情真好/), full-Roman (FR) script (e.g. he is very happy today, /i2-gin2-a1-lit7-sim2-zing2-zim2-her4/) and Mixed Chinese –Roman (MCR) (i.e. articles include Chinese character and Roman script, e.g. he is very happy today, /伊今a4日心情zim2-her4/). On the other hand, there are at least two pronunciations in Taiwanese words. Those pronunciations can be categorized into literate pronunciation and oral pronunciation. Literate pronunciations are usually used in reading classical literature and there exists some correspondence to Mandarin. Instead, oral pronunciations are usually used in daily lives and some different dialects exist such as Taipei, Tai-nan, I-Lan and Lu-gang. Besides written form and multiple pronunciation issues, there is not a suitable phonetic symbol sets for multilingual text analysis.

The TTS system proposed in this paper 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>. We will describe the process of the tone-sandhi rules for text analysis module, since Taiwanese is a tonal language. In waveform synthesis module, the system adopts TD-PSOLA to modify the waveform by adjusting the prosody parameters of selected units so that the synthesized speech sounds more natural.



<Fig. 1> The TTS system architecture is composed of 3 major functional modules, namely a text analysis module, a prosody module, and a waveform synthesis module.

2. Text analysis module

2.1. The Formosa Phonetic Alphabet (ForPA)

The most widely known phonetic symbol sets used to transcribe Mandarin Chinese are the Mandarin Phonetic Alphabet (MPA, also called Zhu-in-fu-hao) and Pinyin (Han-yu-pin-yin), which have been officially used in Taiwan and China, respectively, for many years. However, both systems are inadequate for application to the other members of the Chinese language family, like Taiwanese (Min-nan) and Hakka. Among the phonetic systems useful for Taiwanese and Hakka, there are Church Romanized Writing (CR, also call Peh-oe-ji, 白話字) for Taiwanese and the Taiwan Language Phonetic Alphabet (TLPA) for Taiwanese and Hakka. Because the same phonemes are represented using different symbols in Pinyin, CR and TLPA, it is confusing to learn these phonetic systems simultaneously. For example, the syllable "pa (ㄅㄚ)"

in TLPA and "pa (ㄆㄚ)" in Pinyin may be confused with each other because the phoneme /p/ is pronounced differently in the two systems. Therefore, it is necessary to design a more suitable phoneme set, the newly proposed ForPA, for multilingual speech data labeling [6].

2.2. Word Segmentation and Mandarin-Taiwanese translation (Sentence-to- morpheme)

In spite of that the Han-Roman orthography is the most common style for writing Taiwanese in the contemporary Taiwan, all of three types of written texts (i.e. FC, FR and MCR scripts) can be analyzed by our text analysis module. Since there is no natural boundary between two successive words, we must segment text into word sequence first. In fact, each of 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. In our Mandarin-Taiwanese bi-lingual lexicon, each Chinese character string in the lexicon contains one pronunciation in Mandarin and at least two pronunciations in Taiwanese with corresponding transcription in ForPA. Each word in Taiwanese contains one literature (classic) pronunciation and at least one oral pronunciation. The statistics of Mandarin-Taiwanese lexicon is shown in <Table.1>. We use the bilingual pronunciation dictionary as the knowledge source to do word segmentation algorithm based on the sequentially maximal-length matching, which segment Mandarin sentence into maximal-length words combination. The following paragraph will describe word segmentation algorithm:

Step1: If S is the input sentence，transform S into character string C={C_1 C_2 C_3 ...}, where the number of Chinese characters in string C is N, C_i is the minimum combination element and the maximum number of the Chinese character of one word in lexicon is n.

Step2: Search the pattern with {C_1 C_2 C_3 ... C_n} in the lexicon

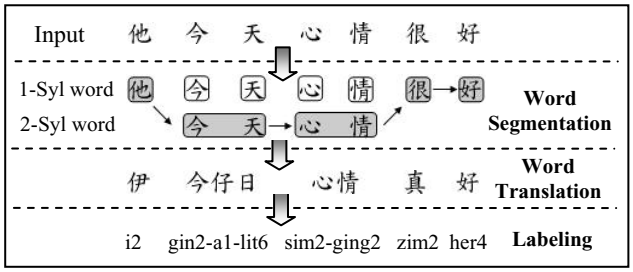
Step3: If the pattern exists, the pattern {C_1 C_2 C_3 ... C_n} is denoted as a word. Repeat Step2 with a new pattern {C_{n+1} C_{n+2} C_{n+3} ... C_{n+n}}.

Step4: If the pattern does not exist, repeat Step2 with a new pattern {C_1 C_2 C_3 ... C_{n-1}}.

Finally, we straightly translate Mandarin into Taiwanese word-by-word, and transcribe Taiwanese words phonetically in ForPA (see in Section 2.3). An example of the segmentation, translation and transcription progress is shown as <Fig 2>, where the input Mandarin sentence is "他今天心情很好" (he is very happy today).

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



<Fig 2> The text analysis progress, which combine word segmentation, translation and labeling (see section 2.3), where the input Mandarin sentence is "他今天心情很好", "1-Syl Word" is denoted as one-syllable word and so on.

2.3. Labeling (morpheme-to-phoneme) and Normalization of the digit sequences

For each segmented word, there exists not only one Taiwanese pronunciation for it. To deal with the multiple-pronunciation problem, two stage strategies are used. The first stage: Choose the oral pronunciation first to do labeling. If the oral pronunciation does not exist, the literate pronunciation will be considered. The second stage: We build a searching network with pronunciation frequencies as node information and pronunciation transitional frequencies as arc information for each sentence. The best pronunciation is then conducted by Viterbi search.

Another important issue for text analysis is the normalization of the digit sequences. However, for digits, there are also 2 manners (literate and oral) 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 as shown in <Table. 2>. In addition, if a digit sequence does not represent a quantity, it is pronounced with the classic pronunciation.

Position	Pronunciation
Ten Million	Read 0-9 as oral pronunciation
Million	Read 0-9 as oral pronunciation
Hundred Thousand	No sound for 1, 2 as literate pron. and others as oral pron.
Ten Thousand	If the digit is 0 in hundred thousand position, read 0-9 as oral pron. If the digit is not 0 in hundred thousand position, 1,2 read as literate pron. and others as oral pron.
Thousand	Read 0-9 as oral pronunciation
Hundred	Read 0-9 as oral pronunciation
Ten	No sound for 1, 2 as literate pron. and others as oral pron.
Unit digit	If the digit is 0 in hundred thousand position, read 0-9 as oral pron. If the digit is not 0 in hundred thousand position, 1,2 read as literate pron. and others as oral pron.

<Table.2>: The rules for normalization of the digit sequences where pron. denote "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 consonants,

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. The sandhi rules for each lexical tone are shown as <Fig. 3>, which is called the "tone sandhi sailboat", where tone "1" must change to tone "2", tone "2" must change to tone "3", tone "3" must change to tone "4", tone "4" must change to tone "1", tone "5" may change to tone "2" or tone "3" for two different major sub-dialects, and tone "4" may change to tone "9" for "HaiKau" sub-dialect.

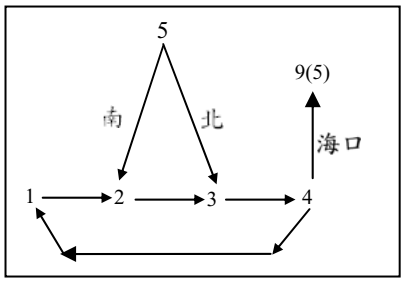
In addition, the <Fig. 4> shows the tone-sandhi rule in stop tones "6"and "7", where tone "6" and "7" must change into "8" and "6" in three kinds of stop-syllables with tail symbols "-p,-t,-k", instead the tone "6" and "7" must change into "3" and "4" in the kind of stop-syllables with tail symbol "-h".

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>.

ForPA	dong1	dong2	dong3	dong4	dong5	dong9
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

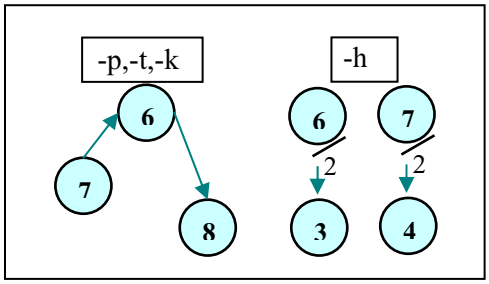
<Table.3>: ForPA: Formosa Phonetic Alphabet, Ch: an example Chinese Character, F0: the fundamental frequency contour, RF: relative frequency level, H: High; M: Middle; L: Low, R:Rising; F: Falling; S: Stop



<Fig. 3> "Tone-sandhi Boat": The Taiwanese tone sandhi rules, where tone "5" may change to tone "2" or tone "3" for two different major sub-dialects, and tone "4" may change to tone "9" for "HaiKau" sub-dialect.

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



<Fig.4> The tone-sandhi rule in stop tones "6" and "7", where tone "6" and "7" must change into "8" and "6" in three kinds of stop-syllables with tail symbols "-p,-t,-k", instead the tone "6" and "7" must change into "3" and "4" in the kind of stop-syllables with tail symbol "-h".

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 abundant news was collected 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. We choose a set to cover all distinct Chinese characters and minimize the number of sentences from 169,040 sentences. Due to time constraints, we choose preceding 200 sentences for manual transcription by two Taiwanese linguistics experts. The 200 sentences cover 41% of all distinct Chinese character, which occur in candidate articles. In Comparison with 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 shown as <Table. 5>.

From the <Table 5>, we find the system can segment and translate most articles accurately into Taiwanese word and

reach over 97% accuracy. If we do not consider tone-sandhi, the system can translate 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.

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

<Table 5> The statistics of performance in parts of word segment and transfer, labeling and tone-sandhi.

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 [7].

There are many variety of synthesis method. We adopt the most popular method TD-PSOLA to modify the prosodic feature of selected units. Synthesis components are used to not only raise or lower pitch but also enlarge or shrink duration. After the analysis of tonal syllables, we can gather duration and short pause information in each syllable. By the information mentioned above, 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.

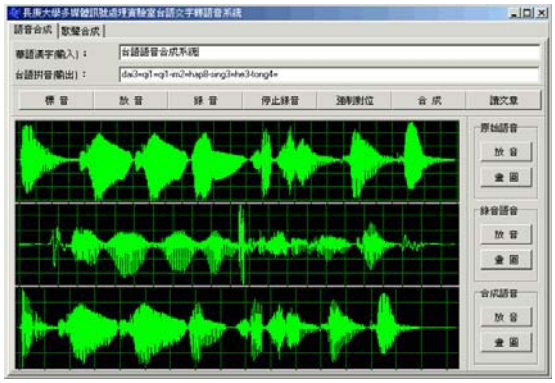
Waveform synthesis module

6. Applications with Taiwanese TTS system

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 and corresponding Taiwanese speech. It is a good and friendly tool to support those who want to learn Taiwanese. On the other hand, as mentioned about Section 2.1, the ForPA is a more suitable phonetic alphabet set for Taiwanese. In order to spread ForPA, it is necessary to construct a Taiwanese interactive phonetic alphabet learning tool. When we type various kinds of existing Taiwanese syllables in ForPA, the tool will pronounce simultaneously. In addition, the TTS system also can sing in Taiwanese with music of song.

7. Conclusion

As shown in <Fig. 5>, we have successfully constructed a mandarin-to-Taiwanese TTS system. Hence, most Mandarin articles can be translated into Taiwanese and automatically converted to a speech signal in Taiwanese. This is great helpful for those who want to learn mother-tongue language in Taiwan. However, there are still a lot to do. In the future, we should improve tone-sandhi for more accurate speech synthesis.



<Fig. 5> The interface of Taiwanese TTS system

8. References

[1] Chu, M., H. Peng, Y. Zhao, Z. Niu, E. Chang, "Microsoft Mulan - a bilingual TTS system", *ICASSP '03*, Volume: 1, pp. I-264 - I-267, 6-10 April 2003.

[2] Chou, F. C., C. Y. Tseng, L. S. Lee, "A set of corpus-based text-to-speech synthesis technologies for Mandarin Chinese", *IEEE Transactions on Speech and Audio Processing,* Volume: 10, Issue: 7, pp. 481 – 494, Oct. 2002.

[3] Chen, Sin-Horng, Shaw-Hwa Hwang and Yih-Ru Wang, "An RNN-based prosodic information synthesizer for Mandarin text-to-speech", IEEE Transactions on Speech and Audio Processing, pp. 226 – 239, Volume 6, Issue 3, May 1998.

[4] Lin, Chuan-Jie and Hsin-Hsi Chen, "A Mandarin to Taiwanese Min Nan Machine Translation System with Speech Synthesis of Taiwanese Min Nan." International Journal of Computational Linguistics and Chinese Language Processing, pp. 59-84, 4(1), February 1999.

[5] Lyu, R. Y., Z. H. Fu, Y. C Chiang, H. M. Liu, "A Taiwanese (Min-nan) Text-to-Speech (TTS) System Based on Automatically Generated Synthetic Units", *ICSLP2000*, Oct. 2000.

[6] Liang, M. S., R. Y. Lyu, Y. C. Chiang, "An Efficient Algorithm to Select Phonetically Balanced Scripts for Constructing A speech Corpus", *IEEE-NLPKE 2003*, October 26-29, 2003, Beijing, China.

[7] Chou, F. C., C. Y. Tseng, "Corpus-based Mandarin Speech Synthesis with Contextual Syllabic Units Based on Phonetic Properties", *ICASSP98,* pp. 893 s- 896 vol.2, 12-15 May 1998.