行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※※

※ 多面向、多語者台語語料的收集及 ※
※ 不特定語者連續語音辨識（2/2） ※

※※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：■個別型計畫　□整合型計畫

計畫編號：NSC 89－2118－M－007－004 023

執行期間：89年8月1日至90年7月31日

計畫主持人：江永進

共同主持人：呂仁園

計畫參與人員：

本成果報告包括以下應繳交之附件：

- □赴國外出差或研習心得報告一份
- □赴大陸地區出差或研習心得報告一份
- □出席國際學術會議心得報告及發表之論文各一份
- □國際合作研究計畫國外研究報告書一份

執行單位：國立清華大學統計學研究所

中　華　民　國　90　年　10　月　31　日

行政院國家科學委員會專題研究計畫成果報告

多面向、多語者台語語料的收集及不特定語者連續語音辨識(2/2)

計畫編號：NSC89-2118-M-007-004
執行期限：88 年8 月1 日至90 年7 月31 日
計畫主持人：江永進 清華大學統計學研究所

一・摘要

這篇報告總結自 88 年 8 月 1 日至 90 年 7 月 31 日本校（清華大學）執行國科會補助專題計畫「多面向、多語者台語語料的收集及不特定語者連續語音辨識」研究 e 成果。語料 e 收集 gah 整理是一樁費時、費力 e 大工程，而且整理 e 人所付出 e 心力，尤其是對邊緣化 e 語言如台語來講，會曉書寫、閱讀 e 人口稀少，有能力替台文注音 e 人才更加是難 cue。利用"會曉聽"，阮第一年研製利用 TTS (Text-To-Speech) e 編輯軟體 tcpedit.exe 收集有注音斷詞消息 e 台文語料；藉用"會曉讀"，阮第二年探討使用語音辨識技術 e 注音辦法，藉著記錄讀台語文章 e 語音，利用連續語音辨識加上強制辨識網路，自動辨識轉錄出語料 e 注音消息，可以當時得著二種語料，一種是有注音資料 e 文字語料、另外一種是語音。用音節來算，用強制辨識網路之下 e 音節辨識率是 96.05%。阮同時也針對錯誤定所 e 辦法探討。本報告主要是針對第二年 e 成果報告。

關鍵詞

台語大詞彙語音辨認、語音處理、隱馬可夫模型、錯誤定所

ABSTRACT

Corpora, in their different forms for different purposes, have been the bases for modern natural language processing technology. Taiwanese, as other language members in the Sino-Tibet family, has been marginalized due to many reasons. One of the consequences of this marginalization is that no standard written script exists, and thus collecting corpus for these languages has been extremely difficult. By (almost) arbitrarily selecting the *hanlor* written script (mixture of *hanzi* and roman characters), we are still facing the problem that only few people are capable of phonetically transcribing a given Taiwanese text. On the other hand, reading a Taiwanese text is easier due to the existence of many commonly used *hanzi*. By recording a person's reading of Taiwanese text, we use a continuous speech recognizer for Taiwanese to automatically transcribe the text, and end up with two kinds of corpora, one in text, one in speech. The accuracy of the automatic phonetic transcription is about 96.05% in syllable count. For marginalized languages, this automatic transcription can be very useful for corpus collection if proper error spotting scheme is implemented.

Keywords

Speech Recognition, Taiwanese Large-Vocabulary Speech Recognition, Hidden Markov Model, Error Spotting

二・計畫緣由

現在自然語言 e 相關處理技術，統計式 e 語言模型是關鍵 e 辦法。馬可夫語言模型 (Markov language model)，是將語言發生 e 過程當做 Markov 過程，對自然語言處理 e 成功

扮演關鍵 e 地位；由馬可夫模型所衍生出來 e 隱藏式馬可夫模型(Hidden markov model)甚至是語音辨識成功 e 數學基礎。但是統計式模型內底 e 參數需要非常大量 e 語料來訓練、估計 Markov 模型 e 轉移機率，只有豐富 e 語料才會凍訓練出來穩定 e 模型參數。Di 華語、英語方面，電子語料豐富而且已經有人整理相當數量 e 語料庫，但是台語因為種種因素困難，語料少，電子語料 iau 卡少，經過整理 e 語料閣進一步稀有，理由不外是過去 e 政治壓制，文字形式 e 無一致，台語教育不足等等。這使得台語語音辨識 e 進展受著真大 e 限制。

台語是75%住在台灣的人民日常生活 e 語言。根據1996年世界語言年鑑估計，全世界講閩南語 e 人口至少八億八千萬和4千9百萬人，以這個人口數來講，位居是全世界第21名 e 語言。在過去十多年以來，台灣本地語音辨認 e 研究完全集中di 華語（國語，Mandarin）e 研究，因為台灣本質上是一個移民社會，社會上除了華語，閣有台語、客語、甚至原住民語等多語並行，所以阮試著ui多面向、多語者台語語料 e 收集ga不特定語者連續語音辨識，期望將來能有結合客語語音辨認 e 研究。

語料 e 收集 gah 整理是一椿費時、費力 e 大工程，而且整理 e 人所付出 e 心力閣無一定受著適當 e 重視(語料庫本身 e 努力無容易發表論文)；di 語料整理人力方面，華語 dor 有稀有 e 感歎，欠少教育 e 台語，當然只有閣卡嚴重；對本計劃所設定 e 目標，注音語料庫來講，上困難 e 是聲調 e 註記(單字調/口語調(本調/變調))，這是因為台語 e 聲調本來 dor 比華語複雜，閣因為變調 e 關係，難以 di 短期來訓練出熟手。

雖然注音人才難 cue，但是有二個事實可以幫助。第一是“會曉聽” e 能力，第二是“會曉讀” e 能力。

會曉聽 e 能力，畢竟 iau 閣真濟人具備，如果用 TTS 唸出注音，那麼聽起來怪怪 e 所在，可能 dor 是注音 m 對 e 所在。在本計劃 e 第一年中，阮撰寫 TTS (Text-To-Speech)幫助 e 注音工具程式 tcpedit.exe，並且用來對張春凰 e “青春 e 路途”散文集全部注音。tcpedit 是一個自動斷詞、自動注音、改正注音、改正聲調 e 工具，因為有 TTS e 幫助，普通會曉台語 e 人 dor 可以操作，實質上解決人才稀少 e 問題。

會曉讀台文 e 人比起聽有台語 e 人雖然加足少，但是畢竟 iau 是有，而且讀文章需要 e 氣力遠小於手工拍字 e 注音。所以，在第二年 e 計劃中，除了繼續對張春凰第二本散文集“愛在土地發酵”用 tcpedit 注音以外，阮也繼續探討使用連續語音辨識 e 技術，希望會凍幫助加速語料 e 收集。zit 部份現在只是技術 e 探討，阮 iau 無滿意到大量用來收集語料 e 程度。描述如下。

假設有一篇文章，咱希望得著斷詞、注音、以及口語調 e 消息，那麼如果有人將文章唸過，錄音起來(每句錄音做一個檔案)，然後利用特製 e 語音辨識 e 技術，辨識出來，差不多 dor 是注音好 lo。dizia，“差不多”至少有三個意思，使得阮停留在探討 e 階段：第一是辨識並無 100%正確(音節正確率大約是 96%)，如何合理指出可能 m 對 e 所在，是一個問題；第二是阮 e 辨識工作並無包括聲調消息(這是現在大部分華語語音辨識 e 通例)，如何加上聲調辨識，需要進一步研究；第三是斷詞消息並無一定包括在辨識結果當中。

zit 份報告主要是針對語音辨識 e 語料收集法，分做以下幾個部份：首

先紹介實驗材料，然後有關阮所使用e語音辨認e一些細節，包括前級訊號處理層次、語音單位e選取等等，第五、第六節分別是實驗結果和錯誤定所。這項研究已經寫成 2 篇碩士論文[劉 00][劉 01]，3 篇會議論文[Lyu00.1] [Lyu00.2] [Chiang00.1]以及一篇期刊論文[Lyu99.1]於國內及國際學術會議發表。其中[Chiang00.1]zit 篇會議論文亦當作本報告e附件。

三・實驗材料

實驗e材料統計如下：

	Syllable count	Speaker A	Speaker B
Article 1	1497	16.98 min	17.34 min
Article 2	1499	17.34 min	18.02 min
Article 3	2324	26.04 min	26.77 min

Table 1. Experimental Data. Each speaker recorded each article for three times

三篇文章分別是 di 散文集"青春e路途"擷取而來[1][2]，並且藉由兩個語者將來錄完。這兩位語者並且錄製了一些平衡詞語音去訓練語者相關辨識器。

原始文字語料是由 tcpedit 完成注音e，其中值得報告e是變調。

台語e變調規則，應該包括二個部份：什麼時陣愛變調，愛變什麼調。但是過去e著作研究，主要只是愛變什麼調e記錄而已，阮非常趣味(亦非常需要)想 veh 知影，什麼時陣愛變調。

單獨e台語詞發音，一般來講，除了最後一個音節以外，攏需要變調，所以 di 歸句e情形，阮定簡單變調是：除了句尾音節以外，其他攏變調。雖然 zit 個規則無考慮輕聲變調之類e變調，但是，簡單變調e效果如何呢？

變調e部分就拿校正好e變調 ham TTS 系統標e簡單變調做比較，阮分做兩大部分來表示，第一部份是舒喉五調e變調，第二部分是束調(入聲)e變調。Di 舒喉五調e變調方面，伊e正確率達到 97.9%，但是 di 束調e變調這方面伊e正確率只有 69.7%而已。

Di 變調校正e過程中，發現/h/結尾e音節伊e變調明顯卡亂，所以阮進一步將束調e部分閣分做/h/結尾e音節 gah /p/、/t/、/k/結尾e音節兩部分，結果/p/、/t/、/k/結尾e音節變調e正確率 98.4%，/h/結尾e音節變調e正確率只有 17.3%，所以事實真明顯，束調部分變調e正確率會 zia 低，完全是受著 /h/結尾e音節伊e正確率所影響。

整體來講，如果 mai 將/h/結尾e音節一併計算，簡單變調e正確率達到 97.9%，就算將/h/結尾e音節一併計算，簡單變調e正確率 ma 閣有 93.6%。

四・系統所用e演算法

為了消除人類嘴唇e輻射現象，先做預強調(pre-emphasis)放大，係數為 0.975。音框的權重視窗(window)用漢明視窗(Hamming Window)，其長度為 16 毫秒(ms)，每 8 毫秒移動一次，所以有 50%的重疊(overlap)。再取 12 維度(Dimension)的梅爾刻度e倒頻係數，與一維度e能量，組成基本e 13 維度梅爾刻度倒頻係數。再以這 13 維度做基礎，取其一階微分，總共 26 維度e梅爾刻度倒頻係數。

Di這項研究中我們採用聲母/韻母模型以及音素模型，附帶考慮了音節內及跨音節e連音耦合現象，運用連續型隱馬可夫模型(CHMM)為基本e語音模型。由於漢語e單音節性，一般認為音節內e連音耦合現象嚴重過跨音節的連音耦合現象，因此，過去e研究者有時ganna考慮音節內e連音耦合[Lyu95]。di本研究中，阮進一步

考慮跨音節 e 連音耦合，由於模型參數總數一下子膨脹得太快，故模型間的訓練語音資料共用就成了重要的課題。這也在研究中考慮著。

除了聲學模型以外，語音辨識閣需要辨識網路。一般連續語音 e 辨識問題需要複雜 e 網路，但是在本研究中，阮已經知影文字 e 正確答案，所以可以針對每一句錄音，用各“字”e 破音字讀音，強制製作出加足小 e 強制辨識網路，這不但增加辨識 e 效率，也增加正確率。

五・實驗 e 結果

使用強制型 e 辨識網路，將錄好 e 語音資料丟給這個辨識器去辨識。而所得 e 結果如下：

	Speaker A	Speaker B	Average
Syllable correct rate (%)	96.06	96.04	96.05
Sentence correct rate (%)	71.59	69.74	70.67

計算正確率 e 標準答案是由第一年計劃中得著 e 語料當做標準，比較得著 e。音節正確率可以達到96%是主要是因為強制網路 e 結果，而且並無計算聲調 e 正確率。

六・錯誤定所

為著得著高品質 e 語料，不論是什麼理由 e 錯誤，咱攏需要去cue出來，實施人工糾正；指出可能錯誤所在，叫做錯誤定所(error spotting)。阮有試驗二種錯誤定所 e 方法。

A、藉用概似度 e 錯誤檢查

以每個音節做單位來講，比較概似度 e 分數，猜測得分卡低 e 可能是錯誤，並且發出警告。

為著將所有 e 錯誤都包括入來，阮不得不將警告尺度調低。但是這就使得一半左右 e 音節必須要做檢查。也因為如此，假警報率會接近 50%。

B、藉由複製語音辨識 e 錯誤檢查

如果將文章重複 e 錄製兩三遍，我們就可以比較辨識出來 e 翻譯，在音標上有什麼不同。藉由我們比較六組文章翻譯 e 結果，我們發現 iau 閣有 1.27% e 錯誤，而這些錯誤都發生在翻譯器會給定相同 e 結果。

七、計畫成果自評

在本計畫中，我們已經完成以下的工作目標：

(1) 進一步了解台語 e 變調。
(2) 擴充了台語發音辭典。
(3) 擴充台語語音資料庫。
(4) 音素模型、右相關音素聲學模型、左右相關音素模型及跨音節左右相關音素模型。
(5) 使用錯誤 e 發現來減低人工校正的時間。
(6) 藉由相似度 e 錯誤檢查、藉由複製語音辨識 e 錯誤檢查。
(7) 本計畫成果已寫成**2**篇碩士論文外，亦寫成**3**篇會議論文及**1**篇期刊論文分別發表於國內外學術會議及期刊上。詳見**[Lyu99.1] [Lyu00.1] [Lyu00.2] [Chiang00.1][Lio00][Lio01]**，其中**[Chiang00.1]**並作為本文附件。

八、參考文獻

[1][Chiang94] 江永進, "台音式輸入法 version4.1", 臺灣新竹清華大學統計所, 1994

[2]CunHong Diunn張春凰. *The Way of*

Youth. DaiLe Publishing, Taipei, Taiwan. (In Taiwanese.) 1994.

[3][Lyu99.1] Ren-yuan Lyu, Yuang-chin Chiang, Wen-ping Hsieh, Ren-zhou Fang, 'A Large-Vocabulary Speech Recognition System for Taiwanese (Min-nan)', To appear in May, 2000, *Journal of the Chinese Institute of Electrical Engineering*

[4][Lyu00.1] 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)', *International Conference on Spoken Language Processing 2000 Conference*, Jun. 2000

[5][Lyu00.2] 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', *International Conference on Spoken Language Processing*, Jun. 2000.

[6][Chiang00.1] Yuang-chin Chiang , Zhi-siang Yang, Ren-yuan Lyu, 'Taiwanese Corpus Collection via Continuous Speech Recognition Tool', *International Symposium on Chinese Spoken Language Processing*, Jun. 2000.

[7][Lio00]劉惠玫，「用TTS輔助台語語料之處理」，清華大學碩士論文，89年1月

[8][Lyu00]劉映吾，「苗栗腔客家話特定語者之大詞彙辨識」，清華大學碩士論文，90年6月

附件：相關著作（台語語音語音資料庫）
appear at Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP2000), Oct.16~Oct.21, 2000, Beijing, China

TAIWANESE CORPUS COLLECTION VIA CONTINUOUS SPEECH RECOGNITION TOOL

Yuang-chin Chiang[2], Zhi-siang Yang[1], Ren-yuan Lyu[1]
[1.] Dept. of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan
[2.] Inst. of Statistics, National Tsing Hua University, Hsin-chu, Taiwan
Email: rylyu@mail.cgu.edu.tw, rylyu@ms1.hinet.net; Tel: 886-3-3283016ext5677

ABSTRACT

Corpora, in their different forms for different purposes, have been the bases for modern natural language processing technology. Taiwanese (MinNan), as other language members in the Sino-Tibet family, has been marginalized due to many reasons. One of the consequences of this marginalization is that no standard written script exists, and thus collecting corpus for these languages has been extremely difficult. By (almost) arbitrarily selecting the *hanlor* written script (mixture of *hanzi* and roman characters), we are still facing the problem that only few people are capable of phonetically transcribing a given Taiwanese text. On the other hand, reading a Taiwanese text is easier due to the existence of many commonly used *hanzi*. By recording a person's reading of Taiwanese text, we use a continuous speech recognizer for Taiwanese to automatically transcribe the text, and end up with two kinds of corpora, one in text, one in speech. The accuracy of the automatic phonetic transcription is about 96.05% in syllable count. For marginalized languages, this automatic transcription can be very useful for corpus collection if proper error spotting scheme is implemented.

1. INTRODUCTION.

Corpus has been the basis for modern natural language processing techniques. For different purposes, we need different kind of corpus. For marginalized languages such Taiwanese and Hakka, corpus collection can be formidable. The reasons include: no widely accepted written script, only few self-educated people can write in their native languages, and even less people is capable of the necessary background knowledge for a seemingly simple task such as phonetically transcribing a Taiwanese text.

For Taiwanese, there are at least three kinds of written scripts: all in *hanzi* (Chinese character), all in roman characters, and *hanlor* (mixture of *hanzi* and roman characters). Text written in *hanlor* has the advantages that it is easier to create, easier to read (both writer and reader do not have to learn more Taiwanese *hanzis*), and easier to process electronically (no user-defined characters to worry).

Given a Taiwanese *hanlor* text in electronic form, we have corpus in its raw form. Various language processing tool such as automatic segmentation and automatic phonetic transcription can then be readily applied if e-dictionary exists for Taiwanese. Using a 60K-dictioanary of Daiim input method [1], we are able to transcribe the phones of text with 85% accuracy in base-syllable count. The reasons for this far from perfect performance are that text-to-phone conversion technique is not perfect, that Taiwanese has a vast amount of *hanzis* with more than one pronunciation, and that the dictionary does not cover all the words and characters in the text. Manual correction seems inevitable.

But the lengthy manual correct proves to be costly, if not impossible. Due to the lack of Taiwanese education, it is difficult to find a person who is efficiently capable of this correction, especially for tones.

However, thanks to the easiness of *hanlor* written script, an educated native speaker can read Taiwanese text without much difficulty. By recording the reading of a speaker, we are able to use our speaker-dependent Taiwanese continuous recognizer to transcribe speech data. We thus have two kinds of corpus under our disposal: a text corpus with phonetic transcription, and a speech corpus for further study for recognition. Note that only speaker dependent speech recognizer is needed to serve our purpose of phonetic text corpus collection.

This paper is organized as follows. Section 2 describes our scheme of transcription using continuous speech recognizer. Section 3 discusses some error spotting schemes. Last section is for conclusion.

2. AUTOMATIC PHONETIC TRANSCRIPTION OF TEXT VIA SPEECH AND TEXT DATA.

In this section, we will describe our phonetic transcriber using known text and continuous speech recognizer. The performance of the transcriber will be reported.

A Taiwanese *hanlor* text consists of *hanzis* and words in roman characters with the latter representing the sounds. A Taiwanese *hanzi* can have several pronunciations, and even words in roman characters could have pronunciation variation, since Taiwanese is a tonal language and words in roman characters usually is in base-syllable form, that is, syllables without tone mark.

Suppose that we want phonetically transcribed text. An automatic phonetic transcriber could work as follows. Given a Taiwanese *hanlor* text, a speaker's reading of the text is recorded, and a search network of this text is constructed for use with continuous speech recognizer by connecting the possible pronunciations of each word or character. See Fig.1 in appendix for an example. The speech and the network are used for recognition, and then the phonetic transcription reported. This restrictive network can boost the correct rate of the recognition. We then end up with a corpus with phonetic transcription, which in turn can be used to improve the performance of a speech recognizer for more general purposes.

The setup of our recognizer is rather standard. See, for example, [2]. The speech is recorded at 16k sampling rate using a noise-canceling microphone in a relatively quite environment. Speech frames of 16 msec with frame shift 8 msec are used for short time analysis. Each frame is pre-emphasized, Hamming window applied, and mel-ceptrum and delta-ceptrum of dimension 12 each are computed. Together with energy and delta-energy, the feature vector is of 26 components. For the continuous hidden Markov model part, we use both inter-syllabic and intra-syllabic right-context-dependent phones as basic units. Each unit is modeled by a three-states and three-gaussian-mixtures, and trained as a speaker dependent model by a set of speech data on a Taiwanese phonetically balanced corpus. Note that the recognition units are those of base syllable, and thus recognition result cannot have information about tones. We use HTK toolkit [3] for the training and recognition process.

As a comparison, our previously reported Taiwanese large vocabulary recognizer achieve 93.2% syllable correct rate under the same setup [4][5]. And a continuous speech recognizer performs only 55.6% syllable correctness [6] due to the lack of ample corpus to train the language model in the recognizer.

The experiment is performed on the following materials. Three articles from the prose book "The Way of Youth" [7] are read by two speakers. The two speakers also record some speech for our phonetically balanced corpus [4][5] to train the speaker dependent recognizers. The experimental materials are summarized in Table 1.

	Syllable count	Speaker A	Speaker B
Article 1	1497	16.98 min	17.34 min
Article 2	1499	17.34 min	18.02 min
Article 3	2324	26.04 min	26.77 min

Table 1. Experimental Data. Each speaker recorded each article for three times

Those speech data are then recognized using the restrictive network built from the pronunciations variations. Table 2 shows the findings.

	Speaker A	Speaker B	Average
Syllable correct rate (%)	96.06	96.04	96.05
Sentence correct rate (%)	71.59	69.74	70.67

Table 2. Recognition correctness

Note that even with a much smaller network, it is not perfect in the phonetic transcription. A further error analysis shows that most of the errors come from the confusing set such as velar and pre-alveolar consonants. Other minor errors include accent differences, the reading of speaker being incorrect, the pronunciation dictionary not fully covering the correct pronunciation, and combination of two syllables into one due to strong co-articulation.

3. ERROR SPOTTING.

Whatever the reason the errors come from, we still need to find some way to relieve the error correction process from having to go through the whole text. Error spotting seems a reasonable direction to pursuit. If the transcriber can issue a warning on the syllable that is possibly incorrect based on certain criteria, and the warning percentage is reasonable, and the warnings include all that are errors, then manual inspection effort can be reduced significantly. Two methods of error spotting are considered in our study.

A. Error Spotting by Likelihood.

By inspecting the average log-likelihood scores (over all speech frames) generated by recognizer for each syllable, we might guess that one with low average log-likelihood might be an error and issue a warning. A simple criterion is to decide a threshold for each syllable, and issues a warning if a given recognized syllable has lower average log-likelihood.

However, to include all the errors in the recognition results, we have to set the thresholds so low that almost half of the syllables need to be inspected, that is, the false alarm rate can reach 50% to include all the true alarms.

Maybe this alarming criterion is too simple minded, but we did not pursuit further.

B. Error Spotting by duplicate speech recognition.

With two or more sets of speech data on the same text, we can compare their phonetic transcription, and issue a warning if there is a difference. By comparing the six sets of transcription results using the above speech data, we find that there are still 1.27% of the syllables that are recognition error but the transcriber gives the same result, that is, our recognizer consistently makes same mistakes not issuing warning. This result is somewhat disappointed.

4. CONCLUSIONS.

Given the non-perfect performance of the state of the art speech recognizer based on the hidden Markov model technique, it is expected that our phonetic transcriber will not be perfect. An additional reasonable error spotting mechanism will serve our corpus collection purpose. Although the two error spotting methods in this study are less than ideal, we hope that finer setup in the transcriber can improve the error spotting performance: higher speech sampling rate for better consonants discrimination, better alarm criteria based on the likelihood scheme. For marginalized languages such as Taiwanese, this automatic phonetic transcription scheme can be crucial for corpus collection.

Tone transcription is also an important part in the corpus, especially for Taiwanese. In addition to be a tonal language, Taiwanese is also rich in tone sandhi. Basically tone sandhi in Taiwanese consists of two problems: when and where-to. The question when tone sandhi happens is still not well understood, and a corpus with tone information will help. It is an issue needs to be studied.

Another technique useful for phonetic transcription is using the text-to-speech system. By generating a speech based on the phonetic transcription that might contain errors, and play the speech to ears of a native Taiwanese speaker, one can easily judge if possible errors exist according its naturalness. Note that the speech generation process can be as simple as concatenating monosyllabic speech files. This is a study currently undergoing.

REFERENCE

[1] YuangChin Chiang. *Daiim Input Method for Windows 95/98/NT*. Version 4.1. National TsingHua University, Hsinchu, Taiwan. 1994.

[2] L. Rabiner and B.H. Jung. *Fundamental of Speech Recognition*. Prentice-Hall International. 1993.

[3] Entropic Research Laboratory, Inc. HTK: Hidden Markov Model Toolkit. 1993.

[4] Ren-yuan Lyu, YuangChin Chiang, Ren-jou Fang, Wen-ping Hsieh, "A Large-Vocabulary Taiwanese (Min-nan) Speech Recognition System Based on Inter-syllabic Initial-Final Modeling and Lexicon-Tree Search", *ROCLING XI Conference*,

p.139~p.149, Hsinchu. Aug. 1998.

[5] Ren-yuan Lyu, YuangChin Chiang, Wen-ping Hsieh, Ren-zhou Fang, Zhi-xiang Yang, Zong-yi Lin. "A Large-Vocabulary Taiwanese (Min-nan) Multi-syllabic Word Recognition System Based upon Right-Context-Dependent Phones with State Clustering by Acoustic Decision Tree," *International Conference on Spoken Language Processing,* Sydney, Australia. Nov. 1998.

[6] Zhi-xiang Yang. *An Initial Study On Continuous Taiwanese Speech Recognition.* Master Thesis, Dept. of Electrical Engineering, Chang Gung University, TauYuan, Taiwan. 1999

[7] CunHong Diunn. *The Way of Youth.* DaiLe Publishing, Taipei, Taiwan. (In Taiwanese.) 1994.

APPENDIX

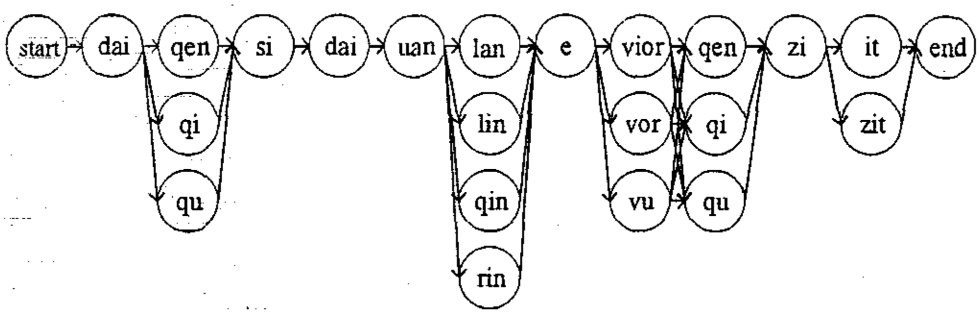

Fig.1. Example network built from a sentence. (Taiwanese is one of the mother tongues of Taiwanese. 台語是台灣人 e 母語之一。)