The Role of Sound Change in the Speech Recognition System:

A Phonetic Analysis of the Final Nasal Shift in Mandarin

楊孝慈　James H. Yang
國立雲林科技大學應用外語系
Department of Applied Foreign Languages
National Yunlin University of Science and Technology
yanght@yuntech.edu.tw

摘要

過去十幾年來，電腦語言學家一直嘗試要設計一個能辨識標準語的語音系統，並能兼顧個別使用者的語音差異。目前，一些語音辨識系統已能針對使用者的年齡和性別，處理不同的音頻問題。然而，語音變化的問題還沒納入考慮。因此本文提議將語音變化的現象，整合到語音辨識系統的設計。本文以中文的字尾鼻音變化爲例，分析它的語音特質，並討論它對語音辨識系統的影響，最後提出一個能提升辨識字尾鼻音的解決方案。

Abstract

Over the past decade, computational linguists have been striving to design a speech recognition system that is able to identify standard speeches and to accommodate sound variables caused by different individual accents. Furthermore, some speech recognition programs have been able to learn and identify distinctive sound frequencies due to the user's age and gender. Nevertheless, regular sound alterations that occur in different varieties of a language have never been seriously considered in the design of the speech recognition system. Accordingly, this study proposes to incorporate the socio-phonological information about regular sound modifications to enhance the performance of Automatic Speech Recognition. To illustrate this point, this study investigates and analyzes the acoustic variation of the syllable-final nasal shift from the velar to the dental, which has been discovered to be one of the distinctive sound features that makes the variety of Mandarin spoken in Taiwan differ from that spoken in China. Following the phonetic analysis, this study discusses the effect of the nasal merger on the development of phonology-dependent speech technologies. It concludes by proposing a preliminary resolution to the identification of syllable-final nasals for the design of Automatic Speech Recognition.

關鍵詞：中文語音學，語音變化，鼻音，語音辨識系統
Key words: Mandarin phonetics, sound change, nasal, speech recognition system

1. Introduction

My motivation to explore the nasal merger of Mandarin spoken in Taiwan originates from an incident in my life. My brother's first baby was born on August 31st, 1999. He gave his son a name called *Geng-ren* /kəŋ.ʐən/[1] (耕仁, meaning "to cultivate benevolence"). Interestingly, I found, as a native speaker of Mandarin in Taiwan, that I would easily mispronounce his name as *Gen-ren* [kən.ʐən] (跟人, meaning "to follow people"), rather than its standard pronunciation. Since the name was subject to mispronunciation and misunderstanding, I suggested to my brother that he should change the name; hence, he later selected another name *Jia-he* (家和, meaning "harmony in the family"). Because of this interesting incident, I came to realize that many native speakers of Mandarin in Taiwan seem to merge the syllable-final velar nasal /ŋ/ with the dental nasal /n/, hence neutralizing such minimal pairs as *geng* /kəŋ/ (耕, "to cultivate") and *gen* /kən/ (跟, "to follow"). To explore this possible sound change, I later conducted a speech production experiment, which is discussed in the subsequent section.

2. Speech Production Experiment

To investigate the possible nasal merger observed above, I addressed three research questions:

1) Is the syllable-final nasal modification a free variation or a conditioned alteration?
2) Does it occur in Mandarin spoken in Taiwan, China, or both?
3) Is it an ongoing or complete sound change?

To address these questions, I invited 30 native Mandarin speakers to participate in the speech production experiment, including 11 males and 19 females, who were students of the University of Hawaii at Manoa. Fifteen of them were from Taiwan, and another fifteen were from China. They were all young adults with the average age of 27, the eldest subject being 36 years old and the youngest one being 21.

For this experiment, I designed a questionnaire to understand the subjects' basic sociolinguistic backgrounds. In addition, I also created sixty easy and interesting riddles to elicit spontaneous speech data. The answers to the riddles included the test words needed for this study--that is, words which end with three types of rhymes: –ing, -eng, and –ang. Three samples of the riddles are displayed below:

---

[1] The character *geng* (耕) is *phonemically* transcribed as [gəŋ] in *Guó-tái Shuāngyǔ Cídiǎn* [Mandarin-Taiwanese Dictionary], edited by Xing-chu Yang (1992).

Table 1. some examples of the riddles used for speech data collection

Riddle	Answer	Pinyin
晚上會發光的昆蟲，是什麼蟲? (What kind of insect lights in the evening?)	螢火蟲 (Firefly)	Yíng-huǒ-chóng
在海邊指引船出入港的，是什麼塔? (What kind of building stands on the coast and guides boats in and out of a harbor?)	燈塔 (Lighthouse)	Dēng-tǎ
全壘打是什麼球類運動的用語? ("Home run" is a term of what sports?)	棒球 (Baseball)	Bàng-qíu

The test words were randomly mixed with 10 irrelevant words in order to avoid the subject's awareness of what words were being examined. I interviewed each participant at a time and tape-recorded each interview. The subject was first asked to answer the questionnaire which included questions concerning his or her basic sociolinguistic background. Next, the interview proceeded with a riddle game, which was carried out in a relaxed atmosphere to collect data from the subject's virtually spontaneous utterances. The informant was told to give the answer to every riddle as soon as possible. If having no idea about the answer, the informant would be given clues to say the test word. Furthermore, if the response was perceived to be not loud enough for sound analyses, the informant would also be asked to say the answer once more and aloud. The recorded data were later analyzed on the computer using Praat, a sound-editing program. The findings are presented in the ensuing section.

3. Findings

Regarding the first research question as to whether the final nasal shift from the velar to the dental (/ŋ/>/n/) occurs without syllabic constraints or appears only in certain environments, the results show that the nasal merger is not a free variation, but a conditioned sound change, which can be formulated by the following phonological rule:

(1) Nasal Fronting:
/ŋ/→[n]/{i, ə}_____.

For example, the word 經 *jing* (/tɕiŋ/, "pass") is regularly pronounced by the Taiwanese respondents as 金 *jin* (/tɕin/, "gold") according to Rule 1. Furthermore, this change in the rhyme also causes lexical neutralization. For instance, the word *jing-yu* (鯨魚, whale) is recurrently pronounced by the Taiwanese respondents as *jin-yu* (金魚, goldfish). Consequently, this nasal merger leads to lexical neutralization, creating homophones sharing the nasal endings with different meanings.

Notably, the nasal merger in Mandarin spoken in Taiwan is not only conditioned by the preceding vowel (Rule 1) but is also blocked by the bilabial onset, which is regularized by the following rule:

(2) Vowel Labialization:
/əŋ/→[oŋ]/[labial] _____

To my knowledge, such discovery has not been specifically analyzed in any previous studies. For instance, the sound *meng* (/məŋ/) is pronounced as *mong* (/moŋ/), rather than *men* (/mən/) according to Rule 2, the vowel labialization rule. Obviously, this sound modification displays articulatory assimilation because the vowel is labialized due to the influence of the initial bilabial consonant /m/. This sound change, however, might not merely occur because of sound assimilation, but it might also exist to constrain the creation of homophones; for example, if the word *meng* (/məŋ/, 夢, dream) is changed according to the nasal fronting rule, it would become neutralized with another word *men* (/mən/, 們, depressed), thus resulting in homophones. By contrast, if the word *meng* is changed according to the vowel labialization rule, it would be pronounced as *mong* (/moŋ/), which does not appear in Mandarin vocabulary. Therefore, the vowel labialization rule does not simply occur for ease of articulation but may also fill the vocabulary gap while avoiding creating homophones.

In addition, the results demonstrate that the syllable-final nasal shift described above occurs mainly in Mandarin spoken in Taiwan, instead of China. Specifically, when the preceding vowel is /i/, the final nasal merger occurs 96 percent of the time in the native speakers of Mandarin from Taiwan (MT). By contrast, the final nasal alteration takes place only 38 percent of the time in the native speakers of Mandarin from China (MC). Moreover, when the preceding vowel is /ə/, the nasal shift occurs 95 percent of the time in MT, while only 3 percent of the time in MC. Nevertheless, the nasal merger never occurs when the preceding vowel is /a/. The occurrence percentage of the nasal merger in Mandarin is displayed as follows:

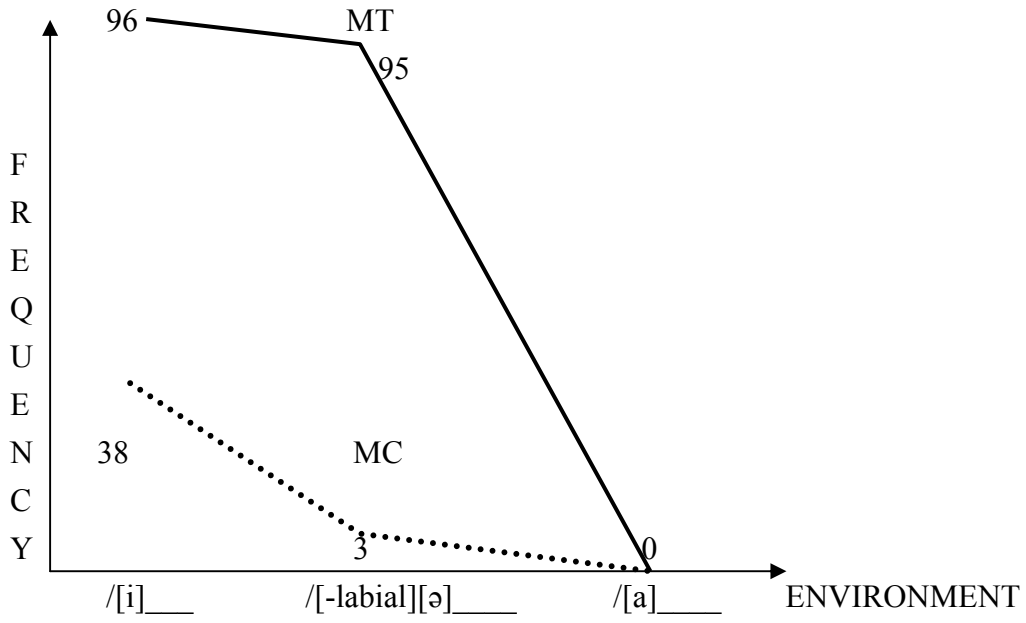

Figure 1 Occurrence percentage of the syllable-final nasal merger in three environments[2]

As Figure 1 displays, the syllable-final velar nasal /ŋ/ in MT merges nearly completely with the dental nasal /n/ when preceded by /i/ or /ə/. By comparison, MC in general does not undergo the nasal shift.

Taken together, all of the Taiwanese respondents underwent the nasal fronting (Rule 1). More than 95 percent of the time they displayed the nasal shift when the final nasal was preceded by the vowel either /i/ or /ə/. By comparison, Tse's 1992 survey suggests that 73% of his Taiwanese informants could not distinguish the syllable-final nasal minimal pairs. Accordingly, the final velar nasal merger with the dental has evolved into a nearly complete status in MT.

However, the final nasal modification described above only appears sporadically in MC. Specifically, Rule 1 occurs only 20 percent of the time in the Chinese responses. Comparatively speaking, Rule 1 occurs more than 95 percent of the time in the Taiwanese responses, making MT differ significantly from MC ($P<0.05$). While some of the Chinese informants from southern China also consistently undergo the nasal shift , the findings do not allow this study to conclude that the final nasal modification occurs regularly in southern China. First, although three of the speakers from southern China recurrently displayed the nasal merger, a couple of them were able to pronounce the test words according to the standard pronunciations without changing the final velar nasal into the dental one. Additionally, the number of the speakers from southern China was very few; only five informants participated in this study. Accordingly, the nasal shift from the velar to the dental seems to present an ongoing phonological process of confusion and interchange in

[2] The percentage of the nasal merger is obtained by excluding the pronunciations that follow Rule 2.

southern China. To confirm whether the nasal merger is common in southern China, an empirical and quantitative study is needed in the future.

In summary, the results demonstrate that the nasal shift (Rule 1) is nearly complete in MT, leaving very few lexical "residues," which are usually high frequency words, such as *sheng* (生, life) and *qing* (清, clear), instead of *shen* (身, body) and *qin* (親, kin), respectively. Theoretically speaking, this nasal shift is consistent with Zee's prediction of the nasal shift from –ng to –n in Chinese dialects (1985) and is opposed to M. Chen's theory of unidirectionality of the nasal shift from –n to –ng (1972, 1973, 1975). However, the syllable-final velar nasal remains unchanged when preceded by the non-palatal low vowel /a/.

4. Phonetic Analyses

The perceived transcriptions of the nasal shift in question are also supported by acoustic analyses. First, the spectrogram of the word *jing-yu* (鯨魚, whale) pronounced by the Taiwanese Mandarin speaker shows that the phonological change is a syllable-final nasal shift, rather than a nasal deletion with the preceding vowel nasalized, as shown below:

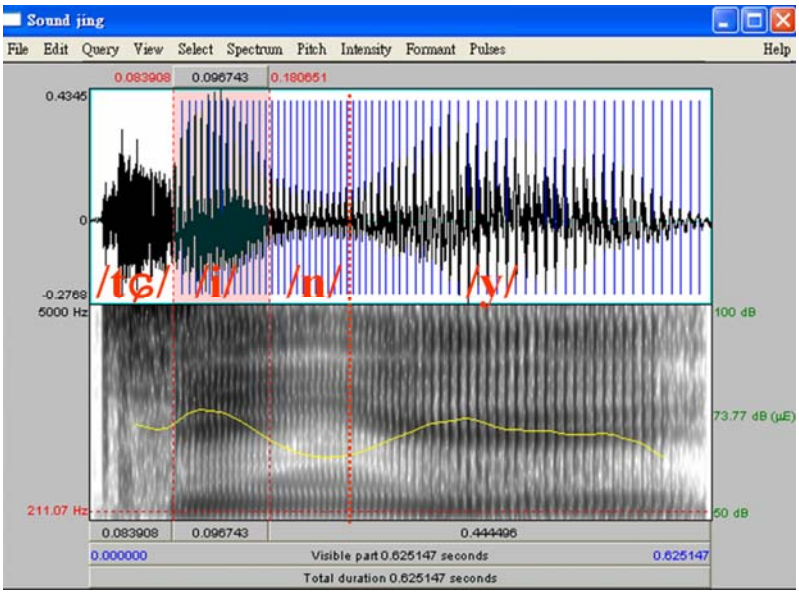

Figure 2. Spectrogram of the test word *jing-yu* (鯨魚, whale) pronounced by the Taiwanese informant

The darker parts of the sound reflect the spectrograms of the vowels, while the whiter parts signal those of the consonants (Ladefoged, 2003). In addition, the sound analysis via Pratt (a computer program for sound analyses) demonstrates that the syllable-final nasal still remains as shown in the middle whiter part of Figure 2.

The speculation that the sound alteration in question is a nasal deletion with the preceding vowel nasalized can also be cleared by comparing the spectrograms of the two

sounds *jī-yu* and *ji-yu*, as shown below.

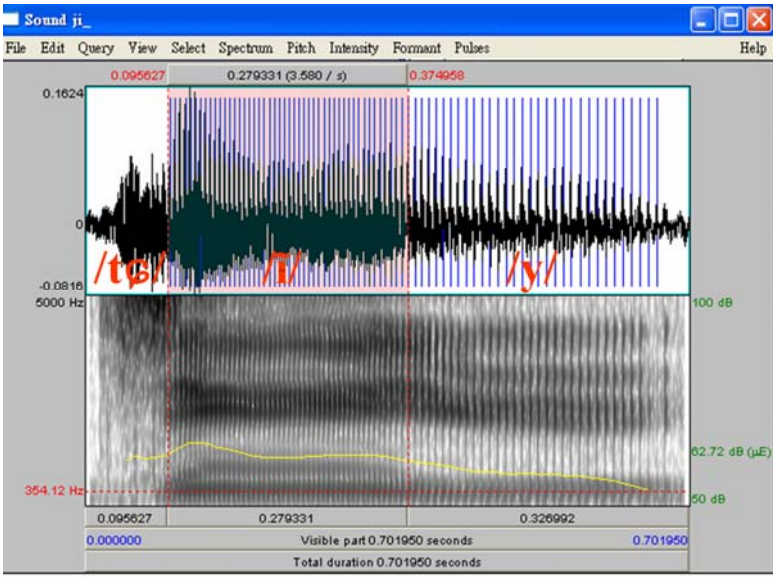

Figure 3. Spectrogram of the sound *jī-yu*

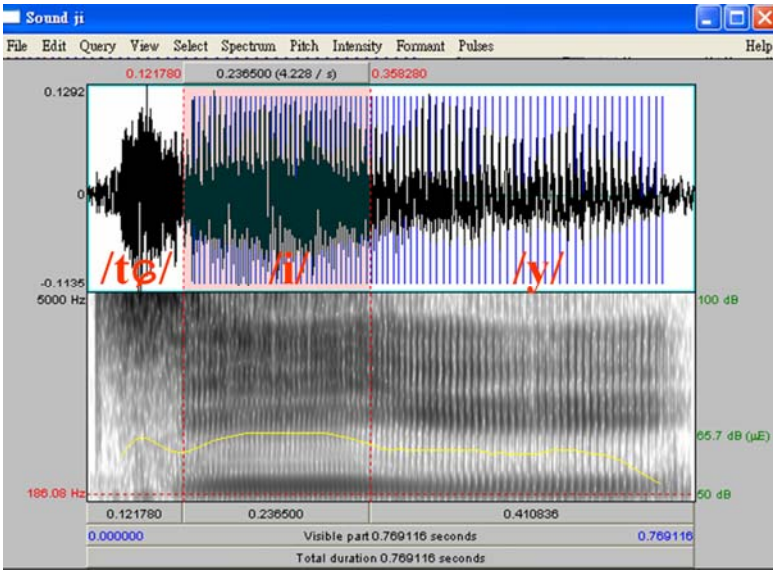

Figure 4. Spectrogram of the sound *ji-yu*

At a glance, none of Figures 3 and 4 look similar to Figure 2. Figure 3 (the spectrogram of the sound *jī-yu*) does not display any whiter part between the vowels /i/ and /y/, but only parallel lines equally black present before the vowel /y/. By contrast, Figure 2 exhibits a whiter part between the two vowels /i/ and /y/. Therefore, the nasal does not disappear but remains.

Furthermore, Figure 4 (the spectrogram of the sound *ji-yu*) also looks distinct from Figure 2. Because the spectrogram of the word *jing-yu* (鯨魚, whale) pronounced by the

Taiwanese informant neither look similar to the spectrogram of the sound *jī-yu*, nor does it look similar to that of *ji-yu*, the sound modification in question should not be the nasal deletion.

Nonetheless, the spectrogram of the word *jin-yu* (金魚, goldfish) pronounced by the same Taiwanese informant looks similar to Figure 2, although their durations and intensities are slightly different, as presented below:

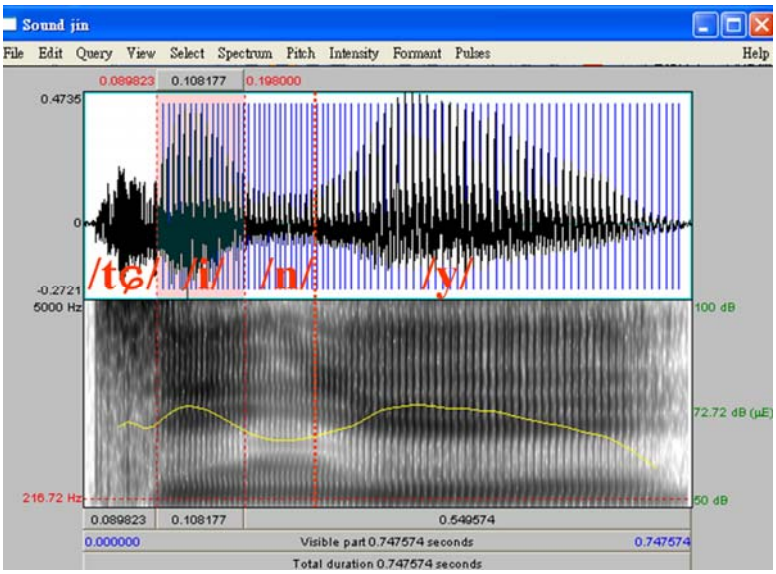

Figure 5. Spectrogram of the test word *jin-yu* (金魚, goldfish) pronounced by the same Taiwanese informant

The spectrogram similarity between Figures 2 and 5 indicates that the Taiwanese Mandarin speaker does not distinguish the minimal pairs *jing-yu* (鯨魚, whale) and *jin-yu* (金魚, goldfish), but pronounces them the same, changing the final velar nasal into the dental. Furthermore, this acoustic analysis is also supported by the discovery from my interview with the Taiwanese informant after the riddle game.

By comparison, the spectrogram of the word *jing-yu* (鯨魚, whale) pronounced by the Chinese informant is distinct from the spectrogram of the same word pronounced by the Taiwanese informant, as shown below:

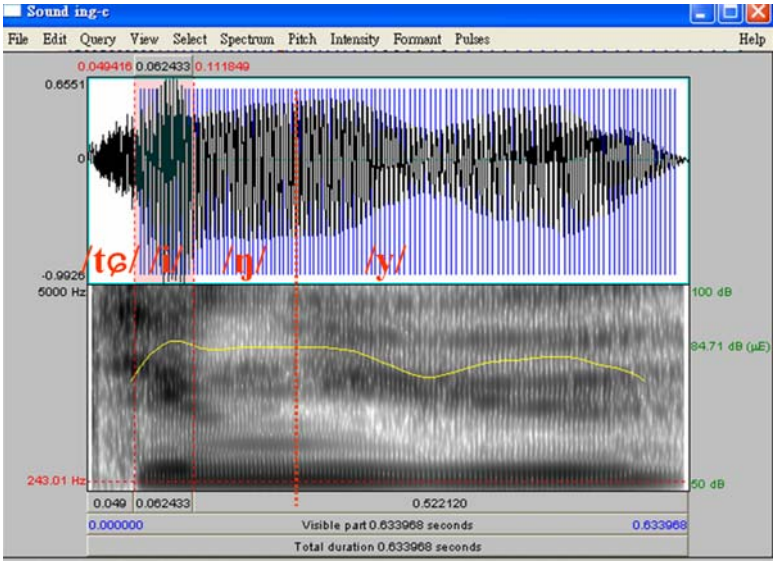

Figure 6. Spectrogram of the test word *jing-yu* (鯨魚, whale) pronounced by the Chinese informant

Apparently, Figure 2 does not look similar to Figure 6; accordingly, it is safe to infer that the final velar nasal does not retain in MT, but must shift into another, either the bilabial or the dental.

Finally, the comparison between Figure 2 and the spectrogram of the bilabial nasal demonstrates that the nasal does not shift to the bilabial because the two spectrograms are distinct from each other, as shown in Figure 7:

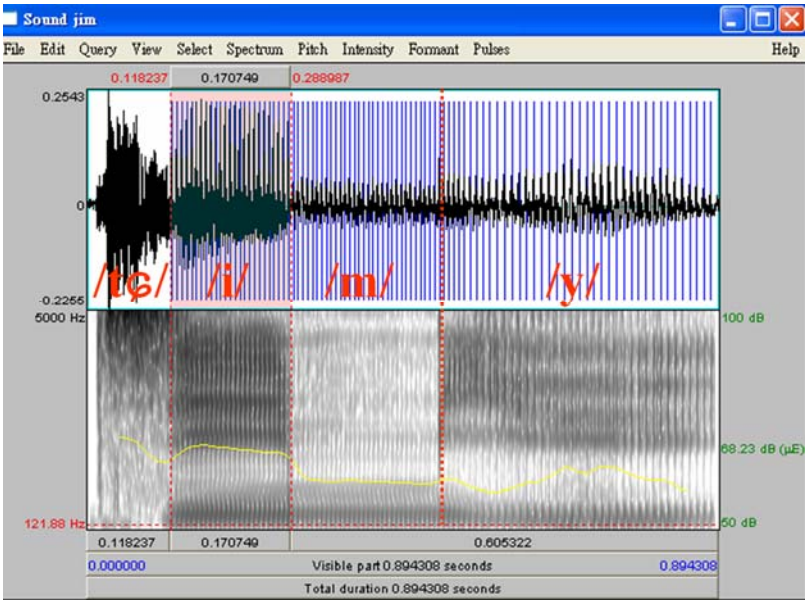

Figure 7. Spectrogram of the sound jim-yu

The spectrogram comparison between Figures 2 and 7 eventually assures us that the syllable-final velar nasal does not switch to the bilabial but the dental.

To summarize, the spectrograms presented above indicate that the syllable-final nasal changes from the velar to the dental when the preceding vowel is /i/ or /ə/, instead of vanishing with the preceding vowel nasalized.

5. Implications for Automatic Speech Recognition

As phonetically presented above, the syllable-final velar nasal shift to the dental regularly occurs in the variety of Mandarin spoken in Taiwan. While this nasal merger in phonologically independent words might not influence computer intelligibility, it might interfere with the computer word-identification process of the syllable-final nasal minimal pairs. For instance, because of the nasal shift, Taiwanese Mandarin speakers tend to say the word *gao-xing* as *gao-xin*, changing the velar nasal into the dental one. This sound alteration might not mislead the speech recognition system to misidentify *gao-xing* (高興) as *gao-xin* (高信) because no such word as *gao-xin* (高信) exists in Mandarin. Nonetheless, this nasal shift has resulted in considerable homophones, neutralizing many minimal pairs, such as *jing-yu* (鯨魚, whale) and *jin-yu* (金魚, goldfish), and *sheng-gao* (升高) and *shen-gao* (身高). Consequently, when a Taiwanese Mandarin speaker says *jing-yu* (鯨魚, whale) as *jin-yu* (金魚, goldfish), the system might misidentify the intended word as *jin-yu* (金魚, goldfish) rather than *jing-yu* (鯨魚, whale), hence retrieving the wrong word. To resolve word-identification problems like this, a possible approach might be to train the user to produce the standard pronunciations of the minimal pairs differing only in nasal endings. Specifically, the user might be trained to articulate the minimal pairs as the standard pronunciations so that the recognition system is able to retrieve the right words.

Such sound training, however, is neither easy nor realistic. While the speech recognition program may provide the standard pronunciations of the syllable-final nasal minimal pairs for the user to imitate and articulate, there is evidence that most people tend to mishear new sounds and mispronounce them according to their habitual articulation (Ohala, 1992, 2001). Furthermore, some recent research has demonstrated that Taiwanese Mandarin speakers tend to preserve their distinctive and unique sound features as their Taiwanese identities, instead of following the standard pronunciations associated with China (Hsu, 2005). Accordingly, a viable resolution might program the automatic speech recognition to distinguish syllable-final nasal minimal pairs, as represented in Table 2.

Table 2. Example of the minimal pairs that differ in the syllable-final nasals following the vowels /i/ or /ə/

-ing/-in	-eng/-en
qing-xin/qin-xin 輕信/親信	cheng-jio/chen-jio 成就/陳舊
ying-qi/yin-qi 英氣/陰氣	sheng-gao/shen-gao 升高/身高
jing-ying/jin-yin 經營/金銀	seng-qing/shen-qin 聲請/申請
xing-xiang/xin-xiang 星象/新象	zheng-zhi/zhen-zhi 整治/診治
jing-yu/jin-yu 鯨魚/金魚	zheng-feng/zheng-fong 政風/陣風

Table 2, however, is not a complete list. Further research is needed to include as many minimal pairs of the syllable-final nasals as possible in the corpus of the speech recognition system.

Most crucially, research into the context where minimal pairs differing in nasal endings is needed to enhance the performance of word identification. For example, when the user says the word *ying-qi* (英氣) as *yin-qi* (陰氣), the speech recognition system might retrieve both of the words for the user to choose. Nonetheless, if we would like to advance automatic speech recognition, we need to investigate the discourse where each of the words occurs. For instance, the word *ying-qi* (英氣) often collocate with such words as *huan-fa* (煥發) and *lin-ran* (凜然), while the word *yin-qi* (陰氣) does not. Furthermore, the word *ying-qi* (英氣) often occurs when the discourse includes words like *zhen-pai* (正派), *hao-shuang* (豪爽), and *guang-ming* (光明). Therefore, future research into the discourse where syllable-final nasal minimal pairs occur will facilitate the speech recognition system to retrieve the intended words.

To conclude, this study has analyzed the phonetic attributes of the syllable-final nasal shift from the velar to the dental that occurs nearly completely in Mandarin spoken in Taiwan. To accommodate this nasal shift, a computational linguist needs to include as many minimal pairs ending with the nasal rhymes as possible in the corpus of the speech recognition system. In addition, to improve the efficiency of the word identification, this study has also proposed to investigate the discourse where each of the minimal pairs appears. In short, future research into the syllable-final nasal minimal pairs and the context of their usages is needed to enhance the identification accuracy of the speech recognition system.

References

[1] J. K. Tse, "Production and perception of syllable final [n] and [ŋ] in Mandarin Chinese: An experimental study.," *Studies in English Literature,* pp. 143-56, 1992.

[2] E. Zee, "Sound change in syllable-final nasal consonants in Chinese," *Journal of Chinese Linguistics,* vol. 13, pp. 291-330, 1985.

[3] M. Chen, "Nasals and nasalization in Chinese: Exploration in phonological universals," University of California at Berkeley, 1972.

[4] M. Chen, "Cross-dialectal comparison: A case study and some theoretical considerations," *Journal of Chinese Linguistics,* vol. 11, pp. 38-63, 1973.

[5] M. Chen, "An areal study of nasalization in Chinese," *Journal of Chinese Linguistics,* vol. 3, pp. 16-59, 1975.

[6] P. Ladefoged, *Phonetic data analysis: an introduction to instrumental phonetic fieldwork*. Oxford: Blackwells, 2003.

[7] J. J. Ohala, "What's cognitive, what's not, in sound change," *Lingua e Stile,* vol. 27, pp. 321-362, 1992.

[8] J. J. Ohala, "An Account of Sound Change," University of California, Santa Barbara: The LSA Institute, 2001.

[9] H.-J. Hsu, "The leveling of neutral tone in Taiwan Mandarin and its new identity," in *2005 international conference of applied linguistics* Chia-yi, Taiwan, 2005.