編譯論叢
第二卷 第一期（2009 年 3 月），101-150

專業口譯需要字正腔圓嗎？

張鳳蘭

常有人說專業口譯發音咬字要像播音員一樣字正腔圓，但專門探討口譯員發音咬字的實證研究卻不多見。本文旨在透過實證調查探究口譯員的發音咬字對聽眾的影響。問卷調查的對象是彰化師範大學三個班級的學生共 121 位。本研究採取假貌相比測試法 (matched guise technique)，由 4 名女性專業口譯分別錄製了 3 個發音版本的錄音檔，包括標準版、自然版和口音版。口譯員兩位來自中國大陸，兩位來自臺灣；其中一位臺灣口譯員也受過專業廣播訓練並主持了多年的電台廣播工作。學生們每聽完 1 個錄音檔，便在李克特 7 等尺度量表 (7-point Likert Scale) 上就口譯的專業程度和聽眾的喜好程度評分。問卷調查在中國大陸東北大學重複進行以探討大陸聽眾的標準和期望是否與臺灣相似或甚至更高。參與調查的爲瀋陽東北大學 3 個班級的學生共 89 位。兩岸問卷調查的結果顯示，在譯文內容無虞的情況下，口譯員發音愈標準聽眾覺得愈專業，專業程度和聽眾的喜好程度呈現高度的正比關係。像播音員一樣字正腔圓有加分作用，但並非必要。

關鍵字：口譯、口譯員、發音、口音、聽眾觀感、聽眾喜好、假貌相比測試法、專業程度、語言態度

收件：2008 年 10 月 30 日；修改：2009 年 1 月 10 日；接受：2009 年 1 月 12 日

張鳳蘭，國立彰化師範大學英語學系暨翻譯研究所助理教授，E-mail: chang_ann@hotmail.com。

Compilation and Translation Review
Vol. 2, No. 1 (March 2009), 101-150

Do Interpreters Need to Sound Like Broadcasters?

Feng-lan Chang

It is widely believed that good enunciation is a basic quality of an interpreter, but few empirical studies have focused on this aspect. The current research investigates how interpreter enunciation affects audience perception. One hundred and twenty-one students from three classes at the National Changhua University of Education served as raters. They listened to four sets of matched guises of speeches recorded by four professional interpreters, two from China and two from Taiwan. One of the Taiwanese interpreters was also a broadcaster. Each interpreter recorded a reading passage in three pronunciation versions: Standard Mandarin, a Natural guise and an Accented guise. Students rated perceived professionalism and their preference on a 7-point Likert scale. The study was replicated in China with 89 students from three classes at Northeastern University as raters. Results of both studies indicated when content of rendition is of no concern, more standard Mandarin during interpreting is perceived as more professional and favorable. However broadcaster-level enunciation is "icing on the cake," but not essential for interpreters.

Keywords: interpreting, interpreters, enunciation, accent, audiences' perception, audiences' preference, matched guise, professionalism, language attitude

Received: October 30, 2008; Revised: January 10, 2009; Accepted: January 12, 2009

Feng-lan Chang, Assistant Professor, Department of English & Graduate Institute of Translation and Interpretation Studies, National Changhua University of Education, E-mail: chang_ann@hotmail.com

Background and Significance

Most experts, scholars, and professional interpreters agree good enunciation is essential for interpreters (Chou & Chen, 1995; Zhang, 1999; Mei, 2000; Liu, 2001; Liu, 2005; Zhong, 2001; Pochhacker, 2004). Kuo (2004) emphasizes exquisite enunciation and intonation (字正腔圓) enhance the professionalism and charm of an interpreter, making one stand out; otherwise, one would be buried among other interpreters. Interpreters should be cautious that mild misarticulation might lead to the discomfort of the listeners, whereas serious ones may actually result in confusion or even misinterpretation.

While it is widely believed good enunciation is a basic quality of an interpreter, empirical studies devoted to this issue are scarce. Previous research on language attitudes has largely studied subjects' reactions to two languages or dialects, where these two choices are available to the speaker in the speech community being surveyed. The interpreting scenario is applicable by definition. Kalmar, Zhong, and Xiao (1987) examined language attitudes in the Guangzhou area. Results were consistent with many studies of areas where a "high" variety (here, Mandarin) coexisted with a "low" variety (in this case, Cantonese). Speakers of Cantonese recognized the social advantages associated with fluent use of Mandarin, but remained loyal to their local dialect. This study also indicated female judges gave substantially higher ratings to Mandarin than did male judges.

The similarity-attraction process has been shown to be central to the process of people evaluating other people in social psychology and linguistic studies (Thatcher, 2004). The more similarity is perceived between the judge and the person being judged, such as background, social class, and language,

the more positively that speaker is evaluated (Coupland, 1980). Hence, the convergence strategies, where the speaker tries to sound more like the listener, supported the data of Berg's (1986) investigation of language choice among shop assistants at a variety of department stores in Taiwan.

While the convergence theory offers an explanation for why listeners tend to prefer those who sound like themselves, other studies have suggested that similarity is not always an asset. Giles (1971) demonstrated that the same local dialect could elicit favorable evaluations on qualities relating to personal attractiveness and integrity but negative evaluations on competence. In other words, a more standard accent is often perceived as more competent, but not necessarily more attractive, suggesting another underlying mechanism is operating besides the similarity-attraction dynamics.

Yet another factor to be taken into consideration is the perceived group identity. Listeners' evaluations are often influenced by the perception a speaker belongs to or is from outside the speech community of the judge (Tajfel, 1974; Thatcher, 1995).

To address effectively the aforementioned dimensions on language attitudes, the most widely used method is the matched guise technique (Thatcher, 2004). The rationale is that it is impossible to compare reliably between the speaker representing "exquisite enunciation" and one that is "less exquisite" because many other factors could differ between the two speakers such as timbre, pitch, and speech rate, etc. One may end up comparing between apples and oranges. The matched guise design attempts to control extraneous factors by using a single speaker to perform in multiple variants or "guises", as an actor might change costume and make-up, appearing in the same play in a second role (Thatcher, 2004). In the matched-guise design, the voice of a speaker is held constant, allowing only the independent variables (in this case,

for example: broadcaster-like enunciation vs. accented pronunciation) to vary.

Lambert et al. (1960) pioneered the matched guise technique. In this branch of research, a number of studies on native speakers' evaluation of native varietal speakers are relevant to this project. Arthur, Farrar, and Bradford (1974) examined attitude toward a variety of English. Forty-eight Anglo-American UCLA students rated four pairs of matched guise voices of standard English and Chicano English (a non-standard ethnic dialect). Dialect differences consistently affected rating, with standard English guises receiving more positive evaluation. This finding was supported by the data of Sebastian and Ryan (1985). Their 120 subjects rated two pairs of matched guises representing Spanish-accented English and Standard English. Speakers with the heaviest accent were the least positively evaluated. Similar results were also found in Thatcher's 1995 matched-guise study in which 210 native speakers of Mandarin in Taiwan evaluated Mandarin use by native and non-native Mandarin speakers. A significant preference was shown for use of Beijing Mandarin when the speaker was believed by the raters to be a native speaker of Mandarin.

Thatcher (2004) explored preference of audience for Chinese-B and Chinese-A interpreters working into Chinese using a matched guise format. Seventy-one university students from Taiwan listened to recordings of nine interpreters' voices reading a passage and rated the characteristics of the speakers using a seven-point Likert scale. Four of the samples on the stimulus tape were actually dual performances by two speakers, who performed once each in Taiwan Mandarin and Beijing Mandarin guises. The author advised that Chinese-B interpreters adopt as standard a Beijing Mandarin accent as possible.

In the language attitude literature, Thatcher's work (2004) is one of the few studies examining audiences' perception of interpreters in Taiwan. Nevertheless, the "interpreters" in her study are not real interpreters, but

skilled speakers of Mandarin as a second language and one Taiwan native. Furthermore, the study did not ask raters to judge directly and specifically the speaker's professionalism as an interpreter, which is considered a central quality of a successful interpreter. In addition, the author also pointed out that her study did not investigate how rater characteristics may have affected their judgment.

Existing research has indicated that females exhibited better listening skills than males in first or second language learning (Larsen-Freeman & Long, 1991 quoted in Oxford & Herman, 1995). Eisenstein (1982) also demonstrated females performed significantly better than males on a dialect discrimination task and females were able to recognize dialects of greater or lesser prestige. Yu (2005) indicated females have greater linguistic awareness than males. Ng (1992) studied end user's reactions to the performance of student interpreters. Gender differences were observed in the study. Males were found to focus on the interpreters' lexical choice and overall fluency, while females tended to attach more significance to the correctness of grammar. Different majors in college have been shown to demonstrate different dominant intelligence. More language-related majors such as Chinese and law majors exhibit better linguistic intelligence than science majors (Yu, 2005; Hsu, 2006). Prior training or exposure to interpreting might also affect audience expectations. Kurz (1993) conducted three surveys on interpreter-related qualities in conferences involving medical personnel, engineers and diplomats, respectively. It was found that the average ratings by these three different user groups were consistently lower than those by the conference interpreters in Buhler's 1986 survey of AIIC (International Association of Conference Interpreters) members. In addition, users' expectation profiles were also found to differ according to their professional backgrounds.

Since 1989, quite a few researches have embarked on the exploration of user-expectations of interpreting services in the western world (Kurz, 1989, 1993, 1994, 1996; Gile, 1990; Ng, 1992; Marrone, 1993; Vuorikoski, 1993, 1998; Kopczynski, 1994; Mack and Cattaruzza, 1995; Moser, 1995, 1996; Collados Ais, 1998; Andres, 2000). Several studies have shown among the important quality criteria of interpreting services, content (e.g. faithfulness to the original) is relatively more important than form (e.g. native accent or pleasant voice) (Buhler, 1986; Marrone, 1993; Kurz, 1993; Kopczynski, 1994; Moser, 1995, 1996). Funded by AIIC, Moser (1995 & 1996) directed the most elaborate survey on user expectations. Using a questionnaire with both open-ended and specific questions, 94 AIIC interpreters conducted a total of 201 interviews on speakers or audiences at 84 different meetings. The meetings were categorized into four groups: large technical meetings, small technical meetings, large general meetings and small general meetings. A marked preference for faithfulness to meaning was found. A lively animated voice was considered important but accent was rated fairly unimportant. On the contrary, Collados Ais (1998) found interpretations with melodious delivery and mistakes were generally rated better than interpretations with a monotonous delivery and total sense consistency.

Ru (1995) conducted one of the few user-expectation surveys on the Mandarin Chinese market. Twenty Taiwanese conference interpreters and 166 conference attendees were asked to rank the relative significance of seven quality parameters of interpretation: pronunciation (發音), fluency of delivery (流暢度), logical cohesion of rendition (一致性), appropriate speed (速度適中), faithfulness to the original (忠實度), use of correct terminology (專業語彙), and pleasant voice (悅耳聲音). The 166 attendees were from five conferences categorized into three types: business, engineering, and religious meetings. It is

interesting that pronunciation received fairly high ranking from the audience: second place by the audience of religious meetings and fourth place by the entire audience. The Taiwanese audiences attached greater significance to pronunciation than do westerners to native accent.

Ru pointed out "pronunciation" in her study was more related to enunciation and intonation. In addition, her targeted subjects were native speakers of Chinese. She speculated the difference in the rating of pronunciation vs. native accent between Chinese and western audiences might be related to one's mother tongue. This argument merits further exploration.

Among the important quality criteria of interpreting, content is shown in many studies to be more important than form. In the condition when content is not a concern and being held constant, is more standard enunciation considered more professional? Furthermore, if more standard is perceived as more professional, does an audience necessarily prefer such enunciation? Is it possible that Taiwanese audiences might prefer more natural pronunciation than carefully articulated enunciation because the former is how most people speak in Taiwan and therefore is considered more personable? On the contrary, would a typical mainland Chinese standard enunciation be perceived by audiences in Taiwan as more distant and therefore less attractive? By the same token, would "standard" Mandarin produced by a Taiwanese interpreter be "accepted" by audiences in China as professional or attractive? Are the underlying expectations and standards similar on the two sides of the Taiwan Strait? With increasing contacts and exchanges across the Taiwan Strait, aspiring Taiwanese interpreters wishing to pursue a career on both sides of the Strait might find this information helpful. Such information can help interpreters enhance their professionalism and quality of interpreting. It can also assist teachers of interpreting to focus more precisely on the training of their students.

Purposes and Research Questions

In this study, the enunciation of four interpreters was manipulated in order to investigate empirically how it affects audiences' perceptions. On the premise that content of rendition is of no concern, this study is designed to answer three specific questions:

1. Is more standard Mandarin Chinese in interpreting perceived as more professional?
2. Do audiences prefer more professional sounding enunciation?
3. How do rater characteristics such as gender, major of study, and prior exposure to interpreting, affect judgment of the interpreters' enunciation?

Predicted Hypotheses

1. More standard Mandarin Chinese during interpreting is perceived as more professional.
2. Perceived professionalism is positively correlated with rater's preference.
3. Rater characteristics, such as major, gender and prior exposure to interpreting affect his/her evaluation of the interpreter. Females are predicted to be more demanding of enunciation than males. English majors, especially those who have taken interpreting courses, are expected to be more peculiar about enunciation than non-English majors.

Methodology

Recruitment of Stimulus Readers

Speakers for the stimulus audio files are four professional interpreters, two from mainland China and two from Taiwan. All speak standard Mandarin and at least one dialect (or are capable of faking one). Only females were recruited in order to eliminate the speaker's gender effect.

Voice One ("Taiwan A").

Voice One is a 50-year-old Taiwanese female from Taipei. She received her interpreting training in Taipei. Upon the completion of her training, she worked as an in-house simultaneous interpreter for a well-known TV station in England for two years. After her return to Taiwan, she worked for a broadcasting company in Taipei as a program host and, on the side, as a free lance interpreter for more than 10 years. Thanks to her background and training in the media, her enunciation is representative of a broadcaster. She also speaks Taiwanese and Cantonese.

Voice Two ("Taiwan B").

Voice Two is a 39-year-old Taiwanese female from central Taiwan (Nantou). She holds a Master's degree in interpreting from the US. Voice Two has been working as a free lance interpreter and translator in Taiwan for one year. She speaks Mandarin and fakes the Taiwanese accent well.

Voice Three ("China A").

Voice Three is a 39-year-old Mainland Chinese female from Beijing.

She left Beijing for Hong Kong at the age of 13. After finishing her college education in Hong Kong, she obtained her Master's degree in interpreting from the US. Voice Three has worked as a free lance conference interpreter since 1993 and has done numerous assignments for the US government. She spent the first eight years of her life in Hainan Island and nearly ten years in Hong Kong. Therefore, besides Mandarin, she also speaks Cantonese and the Hainan dialect.

Voice Four ("China B").

Voice Four is a 26-year-old female from Qingdao China. She also holds a master degree in interpreting from the US. She has worked as an in-house interpreter in Shanghai for approximately one year. Besides Mandarin, she also speaks the Qingdao dialect.

Stimulus Digital Audio Files

Each of the four interpreters recorded three versions of a 166-word reading passage (see Appendix A) with a digital recorder, representing Standard Mandarin, Natural Mandarin, and Accented Mandarin, respectively. In the Standard Mandarin guise, interpreters did their best to produce exquisite enunciation and intonation. In the Natural Mandarin guise, interpreters were told to substitute several retroflex consonants / ㄓ , ㄔ , and ㄕ / purposely with their non-retroflex counterparts / ㄗ , ㄘ , and ㄙ / respectively. Only these three retroflex consonants were manipulated and there was no intonation problems involved. The total number of substituted / ㄓ , ㄔ , or ㄕ / among the four interpreters ranged from three to nine in the entire 166-word passage. Matched-guise design aims to make within-subject comparisons in order to

eliminate between-subject extraneous factors. Therefore, the number of mis-articulated consonants was not kept constant across interpreters since the goal of this project is to compare the three guises of the same interpreter. It was as expected that the interpreters produced different total numbers of substitute / ㄓ , ㄔ , or ㄕ /. Two interpreters ("China B" and "Taiwan A") had very few, only three and four substitutions, respectively, while the other two ("Taiwan B" and "China A") had more, seven and nine respectively. Detailed distribution of the substitutions is listed in Appendix B. If any interesting pattern is observed in the relationship between the total number of substitutions and audience perceptions, such a topic can be specifically investigated in further studies by employing the same interpreter to manipulate the retroflex consonants. To determine an audience's tolerance threshold of inconsistency in retroflex consonants is beyond the scope of this study.

In fact, in the planning stage of the project, "Taiwan A" was asked to fake all the desired guises instead of recording different interpreters; however, there were only so many guises one could fake successfully to convince listeners. She was not able to fake the mainland accent. Consequently, other mainland Chinese interpreters were used instead.

In the Accented Mandarin guise, interpreters read with an obvious dialectal accent. "China A" recorded a moderate Cantonese accent by inserting several typical Cantonese words such as [gai] for「介」and [si] for「司」. "China B" spoke a Quingdao dialect and so the tones differed distinctively from Mandarin Chinese. "Taiwan A" produced a heavy Taiwanese accent by mispronouncing all the retroflex consonants and replacing [f] for [h] in words such as「非」and 「方」. "Taiwan B" also performed the same substitutions but her Taiwanese accent was not as exaggerated. The speed of recording was controlled at about 166 words/47 seconds (+/- 2 seconds). The reading task

was chosen instead of samples from live interpreting because the latter would make it difficult to control variations in vocabulary, grammar, information content, and speech rate in order to maintain sufficiently realistic differences between samples (Thatcher, 2004).

Instrument: Survey

Based on Thatcher's 2004 study, the 7-point Likert scale was adapted to answer the three proposed questions. There are two items, professionalism and rater's preference, for each of the 12 recordings (see sample in Appendix D). Each recording is about 47 seconds in length; consequently, the entire survey takes approximately 15 minutes to complete.

The reliability of the survey was established by five graduate students from the Interpreting Program of National Changhua University of Education (NCUE) serving as raters. Among the five students, three were second-year graduate students and one was a third-year graduate student. The other one was a second-year graduate student in the Translation Program but she had taken several interpreting courses with the interpreting majors and therefore can also be considered a sophisticated listener. The inter-rater reliability index was adopted to assess the degree of agreement among the 5 raters on the 12 guises. The reliability index is 0.75.

Subjects

The study was first conducted in Taiwan and then replicated in China for a comparison. In the Taiwan study, three classes of students (N = 121) from NCUE in central Taiwan served as raters. One was the freshmen class from the Physics Department (n = 36) representing science majors who were primarily males. The other two were English majors, with the sophomore (i.e.,

English-N, *n* = 39) and the senior (i.e., English-Y, *n* = 46) classes differing in that the latter had taken a one-year interpreting course. The senior class did not receive any accent-related training in other courses. Most English majors were females (see Table 1). University students were used as raters because it was difficult to recruit actual conference attendees to participate in the survey outside a conference environment.

Table 1. Gender Breakdown of the Taiwanese Raters

Class	Gender		Total
	Male	Female	
Physics	26	10	36
English-N	3	36	39
English-Y	9	37	46
Total	38	83	121

In the China study, three classes of students (*N*=91) from Northeastern University in Shenyang China served as subjects. There were two subjects with incomplete data, so the subsequent analyses were based on responses from 89 students. The freshmen class from the Automation Department (*n* = 38) represented science majors who were mostly males. The other two are English majors, with the freshmen class (*n* = 31) and the junior class (*n* = 20) differed in that the junior class had taken a one-year interpreting course. As with their Taiwanese counterparts, the juniors did not receive any accent-related training in other courses. Most English majors are females (see Table 2). The majority of the students were from North China.

Table 2. Gender Breakdown of the mainland Chinese Raters

Class	Gender		Total
	Male	Female	
Automation	25	15	38*
English freshmen	5	26	31
English juniors	5	15	20
Total	35	54	89

*Missing data in the Automation class.

Procedures

The survey was administered to the Taiwanese students (i.e., raters) by the author, their regular classroom English teacher, during a regular class period. In China, the raters' regular classroom English teacher conducted the survey after class. For both studies, all three classes did the survey in the same language lab to ensure uniformity in the administration of the evaluation.

Instructions to the raters were pre-recorded into an MP3 file in Mandarin Chinese by the author (see Appendix C) and played back to them. After the instructions were played, raters were presented with the 12 recorded guises of speeches and asked to rate two parameters, perceived professionalism of the interpreter and the degree they like the interpreter's performance (i.e., rater's preference), on the 7-point Likert scale and then identify the origin of the speaker. The 12 guises were presented in a random order for each class to avoid the order effect. A different random ordering was used for each of the six classes.

A repeated measures analysis of variance (ANOVA) was conducted to test the three proposed hypotheses. There are two between-subject factors: *Class* (3 levels) and *Gender*; two within-subject factors: *Interpreter* (4 levels) and *Version* of guises (3 levels). The α level is set at 0.05 level.

Results and Discussion

For each research question, the data from the Taiwan survey will be presented first, followed by those of the China study. The mean scores of perceived professionalism and rater's preference for each interpreter and guise in the Taiwan study are presented in Table 3.

Table 3. The Mean Scores of Perceived Professionalism, Rater's Preference, and Percentage of Correct Guesses of Speaker Origin among Interpreters and Guises in the Taiwan Study

	China A			China B			Taiwan A			Taiwan B		
	Standard	Natural	Accented	Standard	Natural	Accented	Standard	Natural	Accented	Standard	Natural	Accented
Perceived Professionalism (overall)	5.40	3.42	3.06	5.64	5.39	3.84	5.92	4.36	2.42	5.81	4.02	2.95
Rater's Preference (overall)	4.68	2.93	2.64	5.07	4.86	3.06	5.52	3.83	2.60	5.57	3.78	2.69
% of Correct Guesses of Speaker Origin	94.21	55.37	73.55	90.08	38.84	99.17	81.82	85.95	93.39	94.21	86.78	89.26

Q1: Is more standard Mandarin Chinese in interpreting perceived as more professional?

Taiwan Data.

Data in Table 3 indicated more standard articulation received higher scores for all four interpreters. Across all interpreters, the Standard guises were scored the highest, followed by the Natural guises with the Accented guises scoring the lowest. Please also refer to Figure 1 for a graphic visualization.

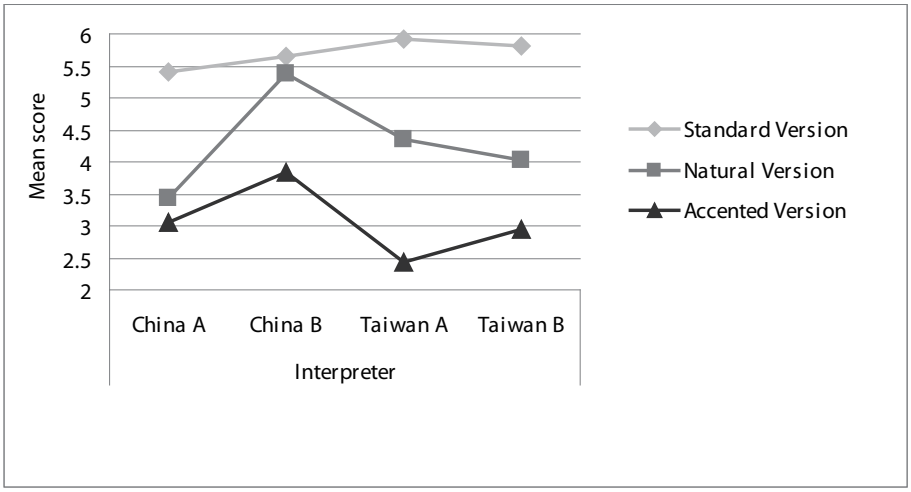

Figure 1. The Mean Scores of Perceived Professionalism by Interpreters and Guises

The repeated measures ANOVA result is summarized in Table 4 (see Appendix E). The *Version* effect is statistically significant [F (2, 230) = 268.77, p < .0001]. The result of the Tukey's post-hoc comparison procedure showed the overall mean scores for the three versions of guises were significantly different from one another, with the average for the Standard guises (5.69) being the highest, followed by the Natural guises (4.30) and the Accented guises scoring the lowest (3.07).

The matched-guise design attempts to control extraneous factors by having the same speaker perform in different variants or guises; hence, it is more meaningful to compare the three guises of the same interpreter. Further analysis was performed with simple main effects and Tukey's post-hoc comparison procedures. The results are presented in Tables 5-1 through 5-4 (see Appendix E).

The result of Tukey's comparison showed that in professional scores,

each of the mean score of four interpreter's three guises was significantly different from one another except for "China B's" Standard vs. Natural guises. This means that across the interpreters, the listeners perceived the three guises as distinctively different with the more standard guises receiving the higher professional scores except for "China B". It is because "China B's" Natural guise had only three misarticulated retroflex consonants. Such subtle difference resulted in a slightly lower mean score in professionalism (Standard 5.64 vs. Natural 5.39), but the difference was not statistically significant. These results support predicted hypothesis one that more standard Mandarin during interpreting is perceived by the college students as more professional.

For the Standard guises, as shown in Table 3 and Figure 1, the ranking of mean scores from high to low is : "Taiwan A": 5.92 > "Taiwan B": 5.81 > "China B": 5.64 > "China A": 5.40. The Taiwanese interpreters were rated higher in professionalism than the mainland Chinese ones. It is interesting to note that "Taiwan A" the broadcaster, received the highest scores. However, with further analysis, simple main effect was performed to investigate the differences among the four interpreters' scores for the Standard guises. The resulting F value [= 5.96, df= (3, 1035), $p < .001$] indicated that the four interpreters' mean professionalism scores for the Standard version were somehow statistically significantly different. Tukey's post-hoc comparison procedure was performed. The results are summarized in Table 6-1(see Appendix E).

The outcome from this study is consistent with similarity-attraction theory (Coupland, 1980; Berg, 1986; Thatcher, 2004) in that Taiwanese listeners found the two Taiwanese interpreters to be more professional than the two mainland Chinese interpreters. As shown in Table 3, the percentage of correct guesses of speaker origin for the four interpreters were high ("China A": 94.21%; "China B": 90.08%; "Taiwan A": 81.82% and "Taiwan B": 94.21%) meaning the listen-

ers were aware of the speaker's origin. The ranking here supports the theory of group identity (Tajfel, 1974; Thatcher, 1995) that when the speaker is perceived as from inside the group of the listener, she is more positively evaluated.

For the Natural guises, the ranking of the mean scores from high to low is: "China B": 5.39 > "Taiwan A": 4.36 > "Taiwan B": 4.02 > "China A": 3.42 (see Table 3 and Figure 1). Fascinatingly, the number of mispronounced retroflex consonants in the order of the above ranking was three for "China B", four for "Taiwan A", seven for "Taiwan B" and nine for "China A" respectively. The results showed the guises with fewer misarticulated retroflex consonants were rated more positively in professionalism.

As with the Standard guises, the result of simple main effect [$F= 47.31$, $df= (3, 1035)$, $p < .0001$] was significant. Results of Tukey's procedure are presented in Table 6-2 (Appendix E). Each pairwise comparison was statistically significantly different from each other, meaning these four Natural guises sounded different to the 121 raters. Nonetheless, as explained in the stimulus digital audio files section, these Natural guises were produced by four different interpreters and, consequently, there was a confounding factor of interpreter. The focus of this matched-guise study is to make within-interpreter comparisons. Having the same interpreter misarticulate different numbers of consonants to control this factor will be done in future studies.

As for the Accented guises, it has been stressed that making between-interpreter comparisons is not the purpose of this study, especially since the type and severity of accents varied among the interpreters. Consequently, no between-interpreter comparisons on the Accented guises will be pursued in this project.

Taken together, the data of this Taiwan survey support the widely accepted belief that good enunciation is essential for interpreters because

the more standard Mandarin Chinese is perceived as more professional. This finding is consistent with the results of Thatcher's 2004 study in that when interpreters were perceived as native speakers of Mandarin, the Beijing guise rather than the Taiwanese Mandarin guise seemed likely to garner more positive ratings on both overall better impression and competence variables. The finding is also in line with the conclusions of several previous studies indicating the more standard accent is more positively evaluated whereas the heavier dialectal accents received more negative ratings (Arthur et al., 1974; Sebastian & Ryan, 1985; Thatcher, 1995).

China Data.

The mean scores of perceived professionalism and rater's preference from 89 students for each interpreter and guise are presented in Table 7. The ANOVA results are summarized in Table 8 (Appendix F).

Table 7. The Mean Scores of Perceived Professionalism, Rater's Preference, and Percentage of Correct Guesses of Speak Origin Among Interpreters and Guises in the China Study

	Interpreters											
	China A			China B			Taiwan A			Taiwan B		
	Standard	Natural	Accented	Standard	Natural	Accented	Standard	Natural	Accented	Standard	Natural	Accented
Perceived Professionalism (overall)	4.73	3.34	2.66	5.99	5.70	2.10	5.61	4.30	3.66	5.55	3.87	2.93
Raters' Preference (overall)	4.45	3.07	2.40	5.82	5.53	2.42	5.20	3.79	3.43	5.24	3.90	2.94
% of Correct Guesses of Speaker Origin	71.91	60.67	37.08	73.03	95.51	97.75	49.44	78.65	86.52	39.33	85.33	97.75

As shown in Table 8, the *Version* effect is statistically significant [F (2,166) = 310.51, p < .0001]. The results of Tukey's procedure indicated the average for the Standard version was the highest (= 5.47), followed by the Natural version (= 4.30), with the Accented version scoring the lowest (= 2.84). Furthermore, a significant *Interpreter × Version* interaction [F (6,498) = 36.81, p < .0001] was found. Consistent with the Taiwan survey, across all interpreters, the Standard guises received the highest ratings, and the Accented guises scored the lowest. For a clear visualization refer to Figure 2.

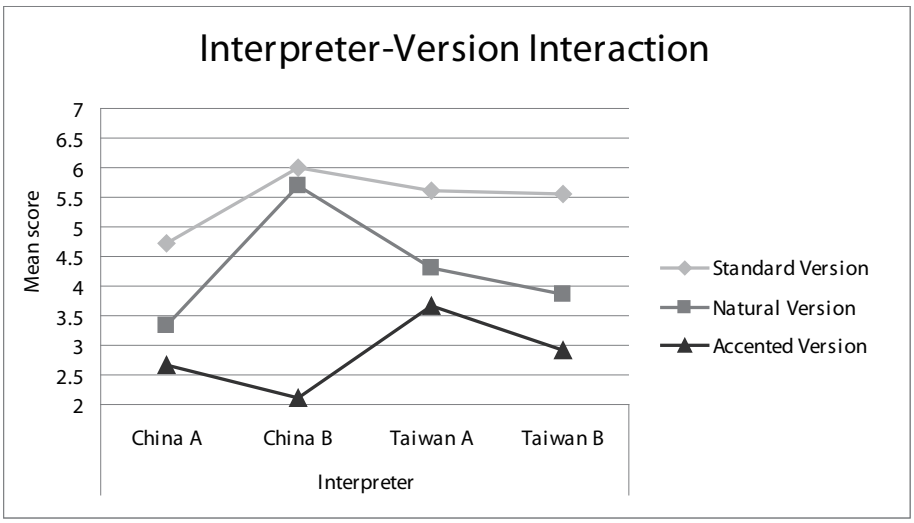

Figure 2. Visual Presentation of the Mean Scores of Perceived Professionalism Among Interpreters and Guises

Tables 9-1 through 9-4 (Appendix F) summarize the results of the simple main effects and Tukey's procedure. Results indicated all comparisons were significantly different except for the comparison between "China B's" Standard and Natural guises. This outcome mirrored the result of the Taiwan study. "China B's" Natural guise only missed three retroflex consonants.

Compared with her Standard guise, this subtle difference was still perceived by the mainland Chinese students, as the Natural guise received a slightly lower professional rating (5.7 vs. 5.99). However, the difference was not statistically significant, meaning that when only two or three retroflex consonants were mispronounced, the overall impact was negligible.

For the Standard guises, as shown in Table 7 and Figure 2, the ranking of mean professionalism scores from high to low is : "China B", 5.99 > "Taiwan A", 5.61 > "Taiwan B", 5.55 > "China A", 4.73. As a significant simple main effect [$F= 20.02$, $df = (3, 747)$, $p < .0001$] was obtained, the result of Tukey's procedure is summarized in Table 10-1. The only non-significant pairwise comparison among the four Standard guises is "Taiwan A" vs. "Taiwan B". The result is different from that of the Taiwan study. For the mainland Chinese students, "China B" sounded significantly more professional than the two Taiwanese interpreters.

As expected, the interpreter receiving the highest professionalism score from the mainland Chinese students was from China ("China B"). However, surprisingly, the two Taiwanese interpreters were evaluated as more professional than "China A". They also received higher preference ratings (see Table 7). A possible explanation is revealed by the percentage of correct guesses of interpreter origin as shown in Table 7. The percentages were 49.44% and 39.33% for "Taiwan A" and "B" respectively, which were much lower than Taiwanese students' 81.82% and 94.21% correctness. The data indicated more than half of the mainland Chinese students thought the two Taiwanese interpreters were from China in origin. As a result, their receiving of higher professionalism ratings can be explained by the similarity-attraction theory. It is interesting to note that mainland Chinese students tended to think interpreters with more standard enunciation were all from China.

With regard to the Natural guises, the ranking of the mean scores from high to low is the same as that of the Standard guises: "China B", 5.70 > "Taiwan A", 4.30 > "Taiwan B", 3.87 > "China A", 3.34 (Table 7 & Figure 2). The results of simple main effects and Tukey's post-hoc comparison procedure are presented in Table 10-2.

The outcome is identical to the finding of the Taiwan study. The guises with fewer misarticulated retroflex consonants scored higher in professionalism. A caveat to note here is the identical pattern in the ranking found in both studies. It is just an interesting observation of the results instead of valid analysis due to the experimental design. Detailed explanation has been given previously in the stimulus digital audio files and in the discussion of the Taiwan data.

Overall, the findings of the mainland study also support the belief that good enunciation is essential for interpreters because the more standard Mandarin Chinese during interpreting was perceived as being more professional.

Q2: Does audience prefer more professional sounding enunciation?

Taiwan Data.

Based on data from 121 students, Pearson correlation coefficient between perceived professionalism and rater's preference was 0.77, indicating a fairly high positive correlation. The correlations between these two variables across the 12 recordings were also high (0.51 - 0.78, see Table 11) except for Taiwan A's Accented guise ($r = 0.20$). A close examination of the data revealed for both the professionalism and preference scores, the listeners' responses were fairly spread out along the 7-point scale, resulting in a low correlation. This means that with Taiwanese A's heavy accent, the listeners' responses were

more multi-polar. Some may have found it unprofessional but cute, amusing or entertaining while others might have found it totally unacceptable, leading to a low correlation between the professionalism and preference scores. To explore specific reasons, the questionnaire can elicit an explanation in the future studies.

Since the result of the survey supports the common view that more standard Mandarin is perceived as more professional, the following discussion will focus on the Standard guises. As shown in Table 3, Taiwanese raters preferred Taiwanese interpreters (Rater's preference scores: "Taiwan B": 5.56> "Taiwan A": 5.5 > "China B": 5.07 >"China A": 4.69). The similarity-attraction theory also applied in this case. Interestingly, "China A" represented a typical mainland Chinese officer's articulation. As predicted, Taiwanese listeners found it less appealing. One may again explain it with the group identity theory or similarity-attraction theory.

Table 11. Pearson Correlation Coefficients between Professionalism and Preference

Interpreter	Version	Correlation coefficient
China A	Standard	0.58
	Natural	0.69
	Accented	0.59
China B	Standard	0.71
	Natural	0.78
	Accented	0.51
Taiwan A	Standard	0.75
	Natural	0.59
	Accented	0.20
Taiwan B	Standard	0.71
	Natural	0.65
	Accented	0.54

Intriguingly, "Taiwan A", "Taiwan B" and "China B" were not significantly different from one another in professional scores. Nonetheless, when rater's preference scores were examined, the resulting F value [= 12.43, df= (3, 1026), $p < .0001$] showed the four interpreters' mean preference scores for the Standard version were also somehow statistically significantly different. The results of Tukey's post-hoc comparison procedure are summarized in Table 12.

Table 12. Results of Tukey's Comparison Procedure on Mean Preference Scores of Standard Version

	Taiwan A	China B	China A
Taiwan B (5.57)	0.05	0.50*	0.89*
Taiwan A (5.52) broadcaster		0.45*	0.84*
China B (5.07)			0.39*
China A (4.68)			

Note. Minimum Significant Difference = 0.35.

* significant at 0.033 level

China Data.

The overall Pearson correlation coefficient between the two ratings, perceived professionalism and rater's preference, was 0.85, indicating a highly positive correlation. The correlations between these two aforementioned ratings across the 12 recordings were mostly high, ranging from 0.65 to 0.94 (see Table 13). The only two exceptions were the Accented guise of "China B" (0.38) and the Natural guise of "Taiwan A" (0.45). Examination of frequency plots revealed that, for these two guises, the ratings of professionalism and preference were relatively more spread out along the 7-point scale, meaning the raters' responses were more varied on these two guises. "China B's" Accented guise is

pure Qingdao dialect. It is possible that the mainland Chinese's reaction to this heavy accent was similar to how the Taiwanese students reacted to the heavy Taiwanese accent. Some might have found it unprofessional but amusing while others might have found it totally unacceptable. Again, to determine specific reasons, one can elicit explanations in the questionnaire in future studies. As for "Taiwan A's" Natural guise, it only missed four retroflex consonants in the 166 words. Some of these mainland Chinese northerners might have perceived it to be acceptable. Had the students not been from Beijing (there was only one Beijing native), the scope of acceptability for "standard Mandarin" might have been greater.

Since the current findings also support the hypothesis that more standard Mandarin is perceived as more professional, the following discussion on rater's preference focused on the Standard guises only. As shown in Table 7, mainland Chinese raters preferred Taiwanese interpreters over "China A", one of their own ("China B", 5.82> "Taiwan B", 5.24 > "Taiwan A", 5.20 > "China A", 4.45). This ranking is similar to that of the professional ratings, only with the two Taiwanese interpreters switching their order. As mentioned previously, a possible explanation is that more than half of the raters took the Taiwanese interpreters for mainland Chinese. Therefore, the similarity-attraction theory again can be applied. Interestingly, "China A" represented a typical mainland Chinese officer's formal and serious articulation. As predicted, Taiwanese raters found it less appealing. However, do the current data reflect mainland Chinese are also tired of its formality and find it distant?

Table 13. Pearson Correlation Coefficients Between Professionalism and Rater's Preference

Interpreter	Version	Correlation coefficient
China A	Standard	0.84
	Natural	0.77
	Accented	0.75
China B	Standard	0.84
	Natural	0.77
	Accented	0.38
Taiwan A	Standard	0.78
	Natural	0.45
	Accented	0.94
Taiwan B	Standard	0.65
	Natural	0.67
	Accented	0.80

Q3: How do raters' characteristics such as gender, major of study, and prior exposure to interpreting, affect their judgment of the interpreters' enunciation?

Because the correlation between perceived professionalism and rater's preference scores is high, the following discussion will focus on the professional scores unless otherwise specified.

Taiwan Data.

Results of repeated measures ANOVA indicated statistically significant three-way interaction effects including *Interpreter×Version×Class* and *Interpreter ×Version×Gender* (see Appendix E).

Figure 3 illustrated the *Interpreter×Version* interaction effects of the three classes. Among them, the overall patterns of the physics majors and the English sophomore class were more similar. With the Standard version, the English senior class differed from the other two classes in that its trend line declined on "Taiwan B". In the senior class, the Taiwanese broadcaster, "Taiwan A", scored higher in professionalism than "Taiwan B" (6.07 vs. 5.35). However, the trend line of the other two classes went upward instead. Another difference was observed in "China B's" Natural guise with three mispronounced consonants. For both the sophomore and the physics classes, her Natural guise almost overlapped with her Standard guise. For the physics majors, her natural version even scored slightly higher than her Standard version (5.58 vs. 5.56). Nonetheless, for the senior class, the two guises were distinctively different. For both the Standard and Natural versions, the professional ratings given by the physics class were more positive than the other two classes across interpreters. Overall, the English senior class appeared more differentiating in the interpreters' pronunciation. They seemed to tune in to the subtle differences more than the other two classes, suggesting prior training in interpreting made a difference. This finding is consistent with the predicted hypothesis and prior research, i.e., people with different professional backgrounds or prior exposure to interpreting present different expectation profiles (Buhler, 1986; Kurz, 1993, 2002; Yu, 2005).

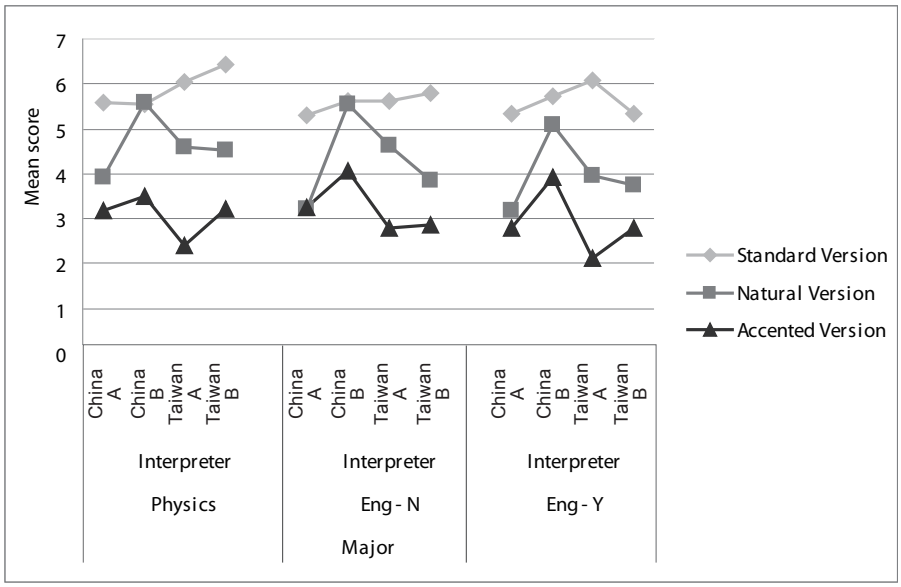

Figure 3. The *Interpreter×Version* Interaction of Three Classes in Taiwan

Figure 4 illustrated the three-way *Interpreter×Version×Gender* interaction effect. Please refer to Table 1 for the breakdown of raters' gender. The main difference between male and female raters was on the two Taiwanese interpreters. For both the Standard and Natural versions, the trends of male raters went upward while those of the females went downward. As mentioned previously, "Taiwan A" is the broadcaster and her Natural guise missed four retroflex consonants while "Taiwan B" missed seven. The female listeners' ratings were as predicted but the males' were the opposite. Furthermore, in 10 out of the 12 conditions (4 interpreters × 3 versions), males gave higher scores than females.

Taken together, as predicted, females tended to be more differentiating in the pronunciation of the interpreters. This finding is in line with prior research indicating females have better listening skills (Eisenstein, 1982; Larsen-Freeman

& Long, 1991).

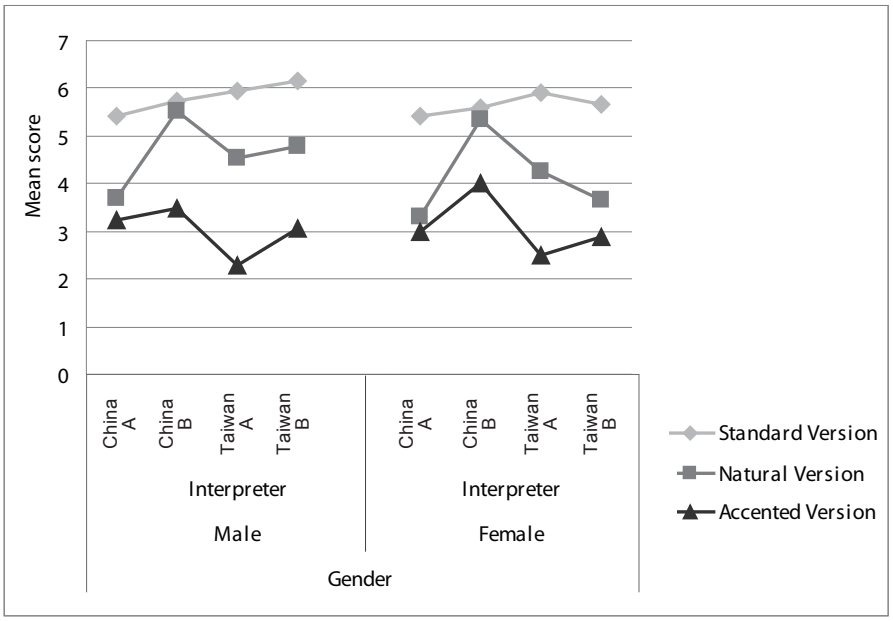

Figure 4. The *Interpreter×Version* Interaction of Males vs. Females in Taiwan

China Data.

As shown in Table 8 (Appendix F), the statistically significant three-way *Interpreter × Version × Class* interaction effect is interesting. Figure 5 illustrated the mean scores of perceived professionalism among interpreters and guises for the three classes. At first glance of the three figures, the mean professionalism scores from the Automation class appeared obviously different from those from the two English classes. It was completely unexpected that for the Automation class, "Taiwan A's" Accented guise, a strong and even amusing accent, was perceived as more professional than her Standard and Natural versions.

Her Accented guise also received a fairly high preference rating (average = 5.47) from the Automation students. This unexpected outcome might suggest that Automation students were less differentiating and listened with less attention to the interpreters' pronunciations as hypothesized. However, another possibility could be that "Taiwan A's" Accented guise was exaggerated slightly and the Automation class might have not responded seriously which could have affected the validity of the outcome. Hence, a more natural accented guise is recommended for future studies.

The patterns of the two English-major classes were similar. The freshmen class appeared to perceive "Taiwan A" as more professional than "Taiwan B" in both the Standard and Natural guises. "Taiwan A" is a typical broadcaster whose Standard guise flowed gracefully. Her Natural guise also missed three fewer retroflex consonants than "Taiwan B" (4 vs. 7 misarticulations). Another difference between the two English classes was observed in "China A's" Natural vs. Accented guises. The English junior class unexpectedly perceived "China A's" Accented guises as more professional than the Natural version. Overall, the mean pattern of the freshmen class is more in line as hypothesized. Contrary to Taiwan raters, the English class who had prior training in interpreting was not attuned to the subtlety as much as the freshmen class. Possible reasons could be related to characteristics of the junior class, such as the length of training, the instructor, or the students not being auditorily sensitive.

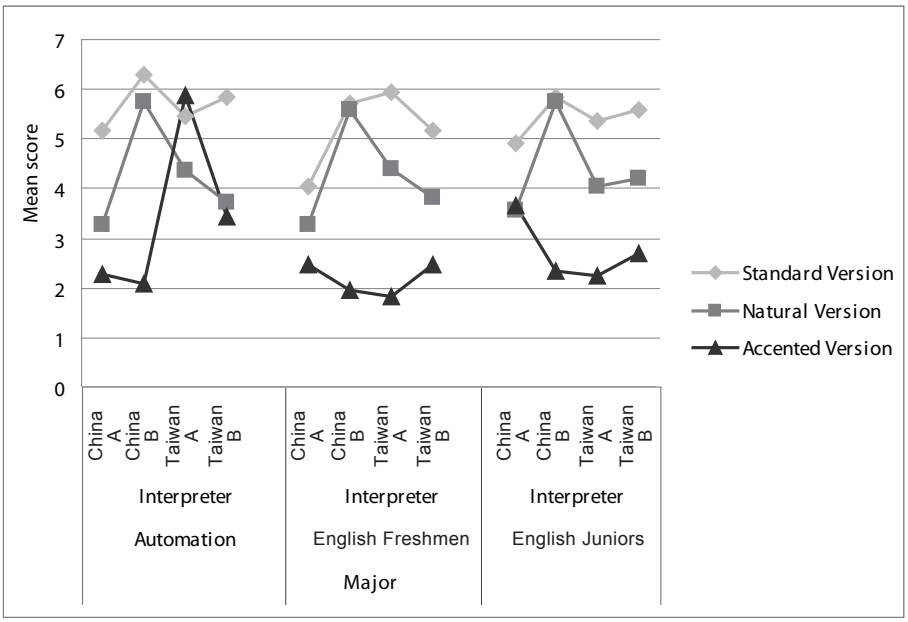

Figure 5. The Interpreter×Version Interaction of Three Classes in China

According to Table 8, the *Gender* effect was significant [F (1, 83) = 6.4, p = .0133]. Results support the hypothesis that females are more demanding of an interpreter's enunciation. The gender difference echoes the finding of the Taiwan study and previous research (Eisenstein, 1982; Ng, 1992).

Figure 6 demonstrates the mean scores of perceived professionalism among guises and classes for the two genders. In most of the conditions (7 out of 9), the mean professionalism scores from males were higher than those from females. This finding, once again, supports the hypothesis. It is also consistent with the finding in the Taiwan survey.

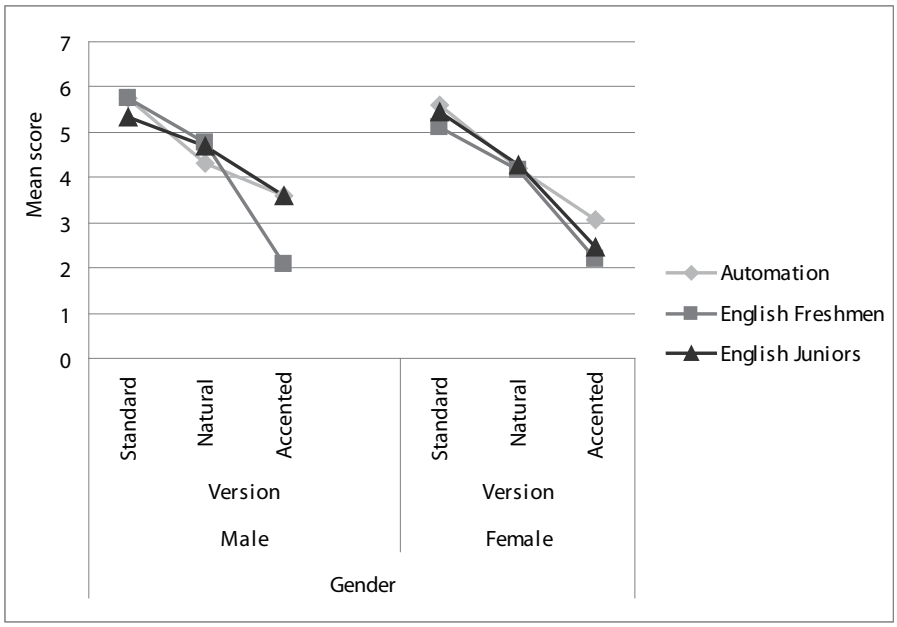

Figure 6. The *Interpreter×Version* Interaction of Males vs. Females in China

Conclusions

In order to verify empirically the common claim that interpreters ought to have exquisite enunciation and intonation (字正腔圓), audience perceptions of professional interpreters were examined in this research project. Three classes from NCUE were surveyed in response to four interpreters' three matched guises. The survey was replicated in China where three classes from Northeastern University in Shenyang served as raters. On the premise that content of rendition is of no concern and held constant, the results were analyzed to answer the following three research questions:

1. Is more standard Mandarin Chinese in interpreting perceived as more professional?

2. Do audiences prefer more professional sounding enunciation?
3. How do rater characteristics such as gender, major of study, and prior exposure to interpreting affect judgment of the interpreters' enunciation?

Firstly, based on the data from 121 Taiwanese and 89 mainland Chinese subjects, more standard Mandarin was perceived as more professional across interpreters. Across interpreters, the Standard guise was rated as the most professional, and the Accented guise the least. In both studies, the top interpreter with the highest professional ratings was from the same country as that of the audience. Furthermore, the raters from both sides of the Taiwan Strait evaluated the Standard guise of "China A" as the least professional and favorable, possibly due to "bureaucrat-like" formality, which might have sounded distant.

Secondly, a high correlation was found between raters' preference and perceived professionalism (0.77 for Taiwan study and 0.85 for China study respectively), meaning audiences also prefer more standard sounding enunciation. Preference scores from the 121 Taiwanese college students for the two Taiwanese interpreters were not statistically significantly different. "Taiwan A", the broadcaster, received the highest professional scores but "Taiwan B" was not significantly different from her in either professional or rater's preference scores. The 89 Chinese students even preferred "China B" and "Taiwan B" over "Taiwan A" (Table 7). The outcome of this study suggests exquisite enunciation is evaluated positively, but a professional interpreter does not necessarily need to sound like a broadcaster in order to be perceived as professional or to win audience approval.

Finally, rater characteristics were found to affect their judgment of the interpreters' pronunciation and performance. In both locations, males tended

to be less demanding in judging enunciation as exhibited in their giving higher professional scores than females.

Differences between the Taiwan and China Surveys

Three interesting differences were observed. Firstly, the ranking of the four interpreters is different. As expected, Taiwanese students rated the Standard guise of a local Taiwanese interpreter-cum-broadcaster ("Taiwan A") as the most professional. However, the mainland Chinese listeners gave a mainland Chinese interpreter ("China B") the highest professional ratings and the mean rating was significantly higher than the second place "Taiwan A". Therefore, to northern Chinese mainlanders, a mainland Chinese accent still sounded more professional than a Taiwanese interpreter-cum-broadcaster.

Secondly, regarding accents, the mainland Chinese listeners from northern China were less able to distinguish the origin of a speaker. The lower overall percentage of correct guesses in speaker origin (China 72.75% vs. Taiwan 81.89%) reflected this outcome. When each of four interpreters' three guises was more closely examined statistically, Cochran-Mantel-Haenszel Statistic = 30.94 ($df=1$, $p < .0001$) was significant. This result indicated the relationship between the origin of the raters and the correctness of identifying interpreters' origin existed in some "Interpreter-Version" conditions. The results are listed in Table 14. The condition with p-value less than 0.0042 (= 0.05/12) can be considered significant.

Table 14. Breakdown of Percentage of Correct Guesses in Interpreters' Origin

Interpreter	Version	% of correct		Chi-square	p
		China	Taiwan		
China A	Standard	71.91	94.17	19.52	<.0001
	Natural	60.67	55.00	0.67	.412
	Accented	37.08	73.33	27.55	<.0001
China B	Standard	73.03	90.00	10.32	.0013
	Natural	96.59	38.33	73.90	.0001
	Accented	97.75	99.17	0.72	.3955
Taiwan A	Standard	50.00	93.33	50.84	<.0001
	Natural	79.55	86.67	1.88	.17
	Accented	86.52	81.67	0.88	.3475
Taiwan B	Standard	39.77	93.33	70.27	<.0001
	Natural	85.39	87.39	0.18	.6756
	Accented	97.75	90.00	4.91	.0266

Statistically significant conditions included: China A's Standard and Accented guises and the two Taiwanese interpreters' Standard guises. "China A's" Accented guise had a moderate Cantonese accent. More than 60% of the mainland Chinese students made an incorrect judgment. Ninety-four percent of Taiwanese students could identify "China A's" Standard guise was recorded by a Chinese mainlander. Anecdotally, during the survey, most Taiwanese students laughed when presented with "China A's" Standard version and indicated the speaker must be from China. Surprisingly, only approximately 72% of mainland Chinese students made correct guesses. This contrast is somewhat hard to explain. About 50-60% of mainland Chinese students guessed the two Taiwanese

Standard guises were produced by Chinese mainlanders. Chinese mainlanders tended to think that more standard guises were produced by mainland Chinese interpreters. A possible explanation is that China is a large country with numerous dialects as well as accents. It is therefore more difficult for the Chinese mainlanders to identify correctly a speaker's origin. According to the author's personal experience from traveling in China, when the mainland Chinese hear poorly articulated Mandarin or Mandarin with an accent unfamiliar to them like Taiwanese, they often think the speaker is from one of the remote provinces such as Canton or Fujian, where people speak with strong southern accents. A Taiwanese accent is just as foreign to northerners as is Cantonese or any other southern accent. On the contrary, there are only limited variations of accents in Taiwan, either Taiwanese Mandarin or non-Taiwanese Mandarin. Therefore, discrimination becomes relatively easier. This can be seen in "Taiwan B's" Accented guise. The percentage of correct guesses for her Standard and Accented versions were both high (93% and 90%), likely because the accents can be easily identified as Taiwanese due to their familiarity. In her Accented guise, she substituted [f] with [h] which is typical of Taiwanese Mandarin. This feature is so salient, which makes guessing easier than her Natural version (87%).

Thirdly, prior training in interpreting exhibited different effects in these two studies. In the Taiwan study, the English majors with prior training in interpreting appeared to be the most discriminating among the three classes on the listening tasks. However, in the China study, the English class with prior training in interpreting was not more discriminating than the English freshmen.

In conclusion, results of both studies illustrated more standard Mandarin is perceived as more professional and favorable. Broadcaster-level exquisite enunciation and intonation is "icing on the cake," but not essential for professional interpreters.

Limitations and Future Plan of Study

The stimulus audio recordings are not real samples from live conferences and the raters are not actual conference attendees. University students are often considered representing the age group that in their future careers may have the opportunities and occasions to use interpreting services and attend international conferences. However, they still do not fully represent all conference attendees. In addition, demography of the raters is another limitation. Finally, adopting two different dialects in the China study could possibly introduce another uncontrolled variable if one were to make between-interpreter comparisons.

For further studies, it will be informative to employ the same interpreter to manipulate the retroflex consonants to determine audience's tolerance threshold of inconsistency in retroflex consonants, because in both studies three substitutions out of 166 words were perceived as negligible but four were not. It will also be interesting to explore audiences' perception in locations, such as Hong Kong and the U.S., where Mandarin Chinese is not the first language. In addition, a larger sample with more balanced group sizes should be taken into consideration. For instance, students from business or law schools might be good candidates, as the male-female ratio would be more balanced. Finally, the questionnaire can be designed to elicit the rater's explanation of judgment for potentially controversial guises to investigate the rationale directly.

Bibliography

Andres, D. (2000). *Konsekutivdolmetschen und Notizen. Empirische Untersuchung mentaler Prozesse bei Anfangern in der Dolmetscherausbildung und professionellen Dolmetschern (as cited in Kurz, 2001)*. Unpublished dissertation, University of Vienna, Vienna, Austria.

Arthur, B., Farrar, D., & Bradford, G. (1974). Evaluation reactions of college students to dialect differences in the English of Mexican-Americans. *Language and Speech, 17,* 55-270.

Berg, M. van den. (1986). Language planning and language use in Taiwan: Social identity, language accommodation and language choice behavior. *International Journal of the Sociology of Language, 59,* 97-115.

Buhler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua, 5*, 231-235.

Collados Ais, A. (1998). *La evaluación de la calidad en la interpretación simultanea. La importancia de la communicación no verbal (as cited in Kurz, 2001).* Granada: Editorial Comares.

Coupland, N. (1980). Style-shifting in a Cardiff work setting. *Language in Society, 9,* 1-12.

Eisenstein, M. (1982). A study of social variation in adult second language acquisition. *Language Learning, 32,* 367-391.

Giles, H. (1971). Patterns of evaluation in reaction to RP, South Welsh and Somerset accents speech. *British Journal of Social and Clinical Psychology, 10,* 280-281.

Gile, D. (1990). L'évaluation de la qualité du travail par les délégués: Une étude de cas (as cited in Kurz, 2001). *The Interpreters' Newsletter, 3*, 66-71.

Hsu, T. (2006). *An innovative measurement: Multiple-intelligences inventory for English language learners.* Unpublished master's thesis, Providence University, Taichung, Taiwan.

Kalmar, I., Zhong, Y., & Xiao, H. (1987). Language attitudes in Guangzhou, China. *Language and Society, 16,* 499-508.

Kopczynski, A. (1994). Quality in conference interpreting: Some pragmatic problems. In M. Snell-Hornby, F. Pochhacker, & K. Kaindl (Eds.), *Translation studies. An interdiscipline* (pp. 189-198). Amsterdam: John Benjamins.

Kurz, I. (1989). Conference interpreting—User expectations. *Coming of Age. Proceedings of the 30th Annual Conference of the American Translators Association,* 143-148.

Kurz, I. (1993). Conference interpretation: Expectations of different user groups. In

F. Pochhacker, & M. Shlesinger (Eds.). *The interpreting studies reader* (2002) (pp. 313-324). New York: Routledge.

Kurz, I. (1994). What do different user groups expect from a conference interpreter? *The Jerome Quarterly, 9*, 3-7.

Kurz, I. (1996). *Simultandolmetschen als Gegenstand der interdisziplinären Forschung (as cited in Kurz, 2001).* Vienna: WUV Universitätsverlag.

Kurz, I. (2001). Conference interpreting: Quality in the ears of the user. *Meta, 46,* 394-409.

Lambert, W., Hodgson, R., Gardner, R., & Fillenbaum, S. (1960). Evaluation reactions to spoken languages. *Journal of Abnormal and Social Psychology, 60,* 44-51.

Larsen-Freeman, D. & Long, M. (1991). *An introduction to second language acquisition research.* London & New York: Longman Group Limited.

Mack, G., & Cattaruzza, L. (1995). User surveys in simultaneous interpretation: A means of learning about quality and /or raising some reasonable doubts. In Tommola, J. (Eds.). In *Topics in interpreting research* (pp. 51-68). Turku: University of Turku.

Marrone, S. (1993). Quality: A shared objective. *The Interpreters' Newsletter, 5*, 35-41.

Moser, P. (1995). *Simultanes Konferenzdolmetschen. Anforderungen und Erwartungen der Benutzer (as cited in Kurz, 2001).* Endbericht im Auftrag von AIIC, Vienna, SRZ Stadt und Regionalforschung GmbH.

Moser, P. (1996). Expectations of users of conference interpretation. *Interpreting, 1* (2), 145-178.

Ng, B. (1992). End users' subjective reaction to the performance of student interpreters. *The Interpreters' Newsletter, 7*, 43-55.

Oxford, R., & Herman, M. (1995). Adults' language learning strategies in an intensive foreign language program in the United States. *System, 23,* 359-386.

Pochhacker, F. (2004). *Introducing interpreting studies.* New York: Routledge.

Sebastian, R., & Ryan, E. (1985). Speech cues and social evaluation: Markers of ethnicity, social class and age. In Giles, H. & St. Clair, R. (Eds.), *Recent advances in language communication and social psychology* (pp. 113-143). London: Lawrence Erlbaum, Ltd.

Tajfel, J. (1974). Social identity and intergroup behavior. *Social Sciences Information, 13,* 65-93.

Thatcher, T. (1995). *Native speaker evaluation of guise choice in Mandarin Chinese.* Unpublished doctoral dissertation, The University of Texas at Austin, Austin, Texas, U.S. A.

Thatcher, T. (2004). *Some observations on preferred Mandarin guise for native and non-native*

interpreters working into Mandarin Chinese. Unpublished master's thesis, Fujen University, Taipei, Taiwan.

Vuorikoski, A. (1993). Simultaneous interpretation—User experience and expectation. In Picken, C. (Ed.), *Translation—the vital link.* Proceedings of the XIIIth World Congress of FIT (vol 1, pp. 317-327). London: Institute of Translation and Interpreting.

Vuorikoski, A. (1998). User responses to simultaneous interpreting. Unity in Diversity? In Bowker, L., Cronin, M., Kenny, D., & Pearson, J. (Eds.), *Current trends in translation studies* (pp. 184-197). Manchester: St. Jerome Publishing.

Zhong, S. (2001). *An introduction to English-Chinese interpretation.* Taiwan: Lighthouse Publishing.

汝明麗（1995）。**從使用者觀點探討口譯品質與口譯員角色**。輔仁大學翻譯學研究所碩士論文，未出版，臺北縣。

周兆祥、陳育沾（1995）。**口譯的理論與實踐**。 臺北：臺灣商務。

林郁如、雷天放、陳菁（廈門大學口譯教研小組）（2006）。**口譯教程**。上海：上海外語教育。

張維爲（1999）。**英漢同聲傳譯**。北京：中國對外翻譯。

梅德明（2000）。**英語高級口譯資格證書考試高級口譯教程**。上海：上海外語教育。

郭岱宗（2004）。優美、快速的中英同步翻譯教學。載於柴明顥（主編），**口譯的專業化道路：國際經驗和中國實踐**（頁280-296）。上海：上海外語教育。

游俊閔（2005）。**大學生多元智慧之研究**。大葉大學教育專業發展研究所碩士論文，未出版，彰化縣。

劉和平（2001）。**口譯技巧：思維科學與口譯推理教學法**。北京：中國對外翻譯。

劉和平（2005）。**口譯理論與教學**。北京：中國對外翻譯。

Appendix A: Text of Recording and Source Text

譯文：工藝品公司介紹

各位來賓、各位朋友：

大家好！非常榮幸能通過這次會議提供的平台與大家相聚，並向大家介紹一下我們的公司。我們是一家專業設計、生產和銷售工藝品的公司，從公司創立至今，我們已經走過了整整 15 年的歷程。15 年前，我們擁有員工 50 名，今天，擁有員工 1,700 名；15 年前， 我們租用廠房 4,500 平方米，今天自建廠房 24,000 平方米，這些數字，都清晰地記錄了我們創造的輝煌。

Source Text: Introduction to an Arts and Crafts Company

Distinguished guests, dear friends:

I feel honored to have this opportunity to introduce our company to you through the platform provided by this conference. Funded 15 years ago, we are a company specializing in the design, manufacture and sales of handicrafts. Fifteen years ago our staff numbered 50; now the figure is 1,700. Fifteen years ago we rented a facility covering a floor space of 4,500 square meters; now our own facility covers a floor space of 24,000 square kilometers. These figures speak loudly and clearly of our success.

Appendix B: Distribution of Retroflex Substitutions

Interpreter	Distribution
China A: 9 Substitutions (shown as underlined)	工藝品公司介***紹*** 各位來賓、各位朋友： 大家好！非***常***榮幸能通過這次會議提供的平台與大家相聚，並向大家介***紹***一下我們的公司。我們***是***一家專業***設***計、生產和銷售工藝品的公司，從公司創立***至***今， 我們已經走過了***整整***15 年的歷程。15 年前，我們擁有員工 50 名，今天，擁有員工 1,700 名； 15 年前，我們租用廠房 4,500 平方米，今天自建廠房 24,000 平方米，***這***些數字，都清晰地記錄了我們創造的輝煌。
China B: 3 Substitutions	工藝品公司介紹 各位來賓、各位朋友： 大家好！非常榮幸能通過這次會議提供的平台與大家相聚，並向大家介紹一下我們的公司。我們***是***一家專業設計、生產和銷售工藝品的公司，從公司創立***至***今，我們已經走過了整整 15 年的歷程。15 年前，我們擁有員工 50 名，今天，擁有員工 1,700 名；15 年前，我們租用***廠***房 4,500 平方米，今天自建廠房 24,000 平方米，這些數字，都清晰地記錄了我們創造的輝煌。
Taiwan A: 4 Substitutions	工藝品公司介紹 各位來賓、各位朋友： 大家好！非常榮幸能通過這次會議提供的平台與大家相聚，並向大家介***紹***一下我們的公司。我們是一家專業設計、生產和銷售工藝品的公司，從公司創立***至***今，我們已經走過了整整 ***1***5 年的歷程。15 年前，我們擁有員工 50 名，今天，擁有員工 1,700 名；15 年前，我們租用廠房 4,500 平方米，今天自建廠房 24,000 平方米，這些數字，都清***楚***地記錄了我們創造的輝煌。
Taiwan B: 7 Substitutions	工藝品公司介紹 各位來賓、各位朋友： 大家好！非常榮幸能通過***這***次會議提供的平台與大家相聚，並向大家介紹一下我們的公司。 我們***是***一家專業***設***計、生產和銷售工藝品的公司，從公司創立***至***今，我們已經走過了整整 15 年的歷程。***1***5 年前，我們擁有員工 50 名，今天，擁有員工 1,700 名；15 年前，我們租用廠房 4,500 平方米，今天自建***廠***房 24,000 平方米，這些***數***字，都清晰地記錄了我們創造的輝煌。

Appendix C: Instructions

以下你會聽到幾位專業口譯員的錄音，有些口譯來自中國大陸， 有些來自臺灣。每聽完一位口譯員的錄音之後，請針對他的專業水平以及你個人對該口譯的喜歡程度評分，然後再判斷他的口音是大陸口音還是臺灣口音。謝謝。

Appendix D: Sample Rating Form

主修 _____，年級 _____

學號 _____

性別 _____，是否學過口譯 ____ 是 ____ 否

語言背景：_____ 國臺語雙語， _____ 只會說國語

請針對下列口譯員的專業水準及你個人對該口譯的喜歡程度評分

Speaker 1

毫不專業	→	→	→	→	→	非常專業
1	2	3	4	5	6	7

毫不喜歡	→	→	→	→	→	非常喜歡
1	2	3	4	5	6	7

你認爲這是大陸口音_____ 臺灣口音 _____ ？

Appendix E: Taiwan Data

Table 4. Repeated-measures ANOVA Summary Table in the Taiwan Study

Source	df	SS	MS	F
Class	2	16.80	8.40	2.11
Gender	1	0.89	0.89	0.22
Class × Gender	2	0.60	0.30	0.08
Error(I)	115	458.73	3.99	
Interpreter	3	85.42	28.47	33.20*
Interpreter × Class	6	10.51	1.75	2.04
Interpreter × Gender	3	4.52	1.51	1.76
Interpreter × Class × Gender	6	2.50	0.42	0.49
Error(II)	345	295.93	0.86	
Version	2	758.53	379.27	268.77*
Version × Class	4	10.37	2.59	1.84
Version × Gender	2	5.92	2.96	2.1
Version × Class × Gender	4	2.66	0.66	0.47
Error(III)	230	324.56	1.41	
Interpreter × Version	6	84.69	14.11	20.16*
Interpreter × Version × Class	12	17.96	1.49	2.14*
Interpreter × Version × Gender	6	9.93	1.65	2.36*
Interpreter × Version × Class × Gender	12	11.12	0.93	1.32
Error(IV)	690	483.16	0.70	
Total	1451	2584.80		

* $p < .05$

Table 5-1. Results of Tukey's Comparison Procedure on the Mean Professionalism Score of Three Guises of Interpreter "China A"

	Natural	Accented
Standard (5.40)	1.99*	2.35*
Natural (3.42)		0.36*
Accented (3.06)		

Note. F (2, 920) = 83.44, p < .0001.
* significant at 0.025 level

Table 5-2. Results of Tukey's Comparison Procedure on the Mean Professionalism Score of Three Guises of Interpreter "China B"

	Natural	Accented
Standard (5.64)	0.25	1.80*
Natural (5.39)		1.55*
Accented (3.84)		

Note. F (2, 920) = 73.47, p < .0001.
* significant at 0.025 level

Table 5-3. Results of Tukey's Comparison Procedure on the Mean Professionalism Score of Three Guises of Interpreter "Taiwan A"

	Natural	Accented
Standard (5.92)	1.56*	3.50*
Natural (4.36)		1.94*
Accented (2.42)		

Note. F (2, 920) = 196.65, p < .0001.
* significant at 0.025 level

Table 5-4. Results of Tukey's Comparison Procedure on the Mean Professionalism Score of Three Guises of Interpreter "Taiwan B"

	Natural	Accented
Standard (5.81)	1.79*	2.86*
Natural (4.02)		1.07*
Accented (2.95)		

Note. F (2, 920) = 125.55, p < .0001.
* significant at 0.025 level

Table 6-1. Pair wise Comparison of the Standard Version among Different Interpreters

	Taiwan B	China B	China A
Taiwan A (5.92)	0.11	0.28	0.51*
Taiwan B (5.81)		0.17	0.40*
China B (5.64)			0.23
China A (5.40)			

Note. Minimum Significant Difference = 0.30.
* significant at 0.033 level

Table 6-2. Pair wise Comparison of the Natural Version among Different Interpreters

	Taiwan A	Taiwan B	China A
China B (5.39) missing 3 consonants	1.03*	1.37*	1.97*
Taiwan A (4.36) missing 4 consonants		0.34*	0.94*
Taiwan B (4.02) missing 6 consonants			0.60*
China A (3.42) missing 9 consonants			

Note. Minimum Significant Difference = 0.30.
* significant at 0.033 level

Appendix F: China Data

Table 8. Repeated-measures ANOVA Summary Table in the China Study

Source	df	SS	MS	F
Class	2	22.83	11.41	2.95
Gender	1	24.79	24.79	6.40*
Class × Gender	2	1.47	0.73	0.19
Error(I)	83	321.51	3.87	
Interpreter	3	96.22	32.07	33.71*
Interpreter × Class	6	64.55	10.76	11.31*
Interpreter × Gender	3	1.22	0.41	0.43
Interpreter × Class × Gender	6	6.99	1.17	1.22
Error(II)	249	236.94	0.95	
Version	2	829.88	414.94	310.51*
Version × Class	4	47.03	11.76	8.80*
Version × Gender	2	2.62	1.31	0.98
Version × Class × Gender	4	18.49	4.62	3.46*
Error(III)	166	221.83	1.34	
Interpreter × Version	6	158.14	26.36	36.81*
Interpreter × Version × Class	12	188.55	15.71	21.94*
Interpreter × Version × Gender	6	7.51	1.25	
Interpreter × Version × Class × Gender	12	14.77	1.23	
Error(IV)	498	356.57	0.72	
Total	1067	2621.91		

* $P < .05$

Table 9-1. Results of Tukey's Comparison Procedure on the Mean Professionalism Rating of "China A's" Three Guises

	Natural	Accented
Standard (4.73)	1.39*	2.07*
Natural (3.34)		0.68*
Accented (2.66)		

Note. F (2, 664) = 63.38, p < .0001. Minimum Significant Difference = 0.37.
* significant at 0.025 level

Table 9-2. Results of Tukey's Comparison Procedure on the Mean Professionalism Rating of "China B's" Three Guises

	Natural	Accented
Standard (5.99)	0.29	3.89*
Natural (5.70)		3.60*
Accented (2.10)		

Note. F (2, 664) = 289.25, p < .0001. Minimum Significant Difference = 0.37.
* significant at 0.025 level

Table 9-3. Results of Tukey's Comparison Procedure on the Mean Professionalism Rating of "Taiwan A's" Three Guises

	Natural	Accented
Standard (5.61)	1.31*	1.95*
Natural (4.30)		0.64*
Accented (3.66)		

Note. F (2, 664) = 85.69, p < .0001. Minimum Significant Difference = 0.37.
* significant at 0.025 level

Table 9-4. Results of Tukey's Comparison Procedure on the Mean Professionalism Rating of "Taiwan B's" Three Guises

	Natural	Accented
Standard (5.55)	1.68*	2.62*
Natural (3.87)		0.94*
Accented (2.93)		

Note. F (2, 664) = 128.81, p < .0001. Minimum Significant Difference = 0.37.
* significant at 0.025 level

Table 10-1. Pair wise Comparison of the Standard Version Among Interpreters

	Taiwan A	Taiwan B	China A
China B (5.99)	0.38*	0.44*	1.26*
Taiwan A (5.61)		0.06	0.88*
Taiwan B (5.55)			0.82*
China A (4.73)			

Note. Minimum Significant Difference = 0.36.
* significant at 0.033 level

Table 10-2. Pair wise Comparison of the Natural Version Among Interpreters

	Taiwan A	Taiwan B	China A
China B (5.70)	1.40*	1.83*	2.36*
Taiwan A (4.30)		0.43*	0.96*
Taiwan B (3.87)			0.53*
China A (3.34)			

Note. F (3, 747) = 72.49, p < .0001. Minimum Significant Difference = 0.36.
* significant at 0.033 level