Speaker Identity of Cantonese-English Bilingual Speakers

Robert Bo Xu, Donghui Zuo, Peggy Mok

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong

xuborobert@gmail.com, donghui@cuhk.edu.hk, peggymok@cuhk.edu.hk

Abstract

Previous studies have shown that language background plays an important role in speaker identification. However, very few studies have focused on voices of bilingual speakers recognized by bilingual listeners. 79 Cantonese-English bilingual speakers participated in a voice line-up identification experiment, in same-language (the languages of target and line-up) and cross-language conditions. The results show that listeners had a higher accuracy in identifying voices in the same language as the target. They were also more confident in their judgment of a more familiar language. It is argued that accent is important in the identification of bilingual speakers. It was also found that there was no correlation between listeners' confidence and their accuracy of identification.

Index Terms: speaker identification, bilingual voice, voice line-ups, forensic phonetics, memory for voice

1. Introduction

Remembering and recognizing the voice of a particular person is an important cognitive ability and social skill. Speaker identification is also a useful, sometimes critical, tool to assist conviction or acquittal in court cases of various sorts, when an earwitness is available. However, the accuracy of speaker identification is influenced by many factors discussed in the literature, such as speaker familiarity, voice disguise, witness preparation and line-up construction.

So far only a few studies have investigated the influences of language background on speaker recognition. Most of them focused on the judgment by monolingual speakers; even fewer dealt with bilingual speaker identification. In reality, the witness to a crime may be required to identify someone who does not speak his most familiar language. Moreover, language contact and bilingualism is very common in many parts of the world. It is possible that both the witness and the perpetrator may be bilingual speakers. Therefore, it is worth investigating bilingual speakers' ability in voice recognition, and how bilingualism affects speaker identification. This research aims to fill the gap in this understudied area.

1.1. Influence of language background on monolingual speaker identity

Only a few studies have examined the relationship between language background and speaker recognition with contradictory results. Inspired by the other-race effect in face recognition, which stated that people has an advantage in recognizing faces in their own race, Goldstein et al. [1] tested whether this effect could be generalized to voice recognition. They came to the conclusion that there was no difference between voice recognition of foreign voices and native voices, i.e., there is no other-race effect in voice recognition.

However, other studies on monolingual speakers showed that language background does affect voice identification. Thompson [2] showed that English monolingual speakers identified English voices significantly better than they did with Spanish voices. Goggin et al. [3] also argued that language familiarity played an important role in speaker identification. They showed that English and German monolingual speakers were better in identifying voices in their native languages. In addition, further studies by Köster and Schiller [4] and Wester [5] showed that typological difference of the target language from monolingual speakers' native language did not impact their identification. In other word, as long as the voice is in a language other than the listeners' native language, what language it is does not affect the outcome of the identification.

The native language advantage found in the above studies can be extended to familiar dialects and accents. Sjostrom et al. [6] found that voices in a more familiar accent could be identified more accurately than those in a less familiar accent. Nevertheless, Thompson [2] found that accented speech does not necessarily impact speaker recognition in comparison to unaccented speech. He showed that the accuracy of monolingual English speakers in identifying an English voice with strong Spanish accent was not significantly different from either an English voice or a Spanish voice.

While Schiller and Köster's study [7] provided support for some previous studies [2][3], they also raised a logical question: if language familiarity benefited voice recognition, is linguistic information an important factor in the process? The question was answered by Schiller et al. [8]. In their study, they replaced all syllables in a natural German passage with 'ma' to minimize the linguistic cues to the target language. As a result, German and English monolingual speakers did not display significant difference in recognizing the target voice. This result implied that linguistic information played an important part in speaker identification.

Besides identification accuracy, the confidence of the witnesses has been discussed in the literature. Jurors' perception of earwitness evidence could be influenced by the confidence of the witnesses in their judgment. However, studies have shown that confidence-identification correlation was either insignificant or relatively low [9][10]. Studies concerning with the influence of language background also showed that the speakers had a greater confidence towards a familiar language in the recognition tasks, even though the correlation between accuracy and confidence is not reliable [2][3].

1.2. Identity of bilingual speakers

Bilingualism put an interesting perspective to the discussion on the influence of language background to speaker identity. However, only a few studies have investigated this topic. Goggin et al. [3] found that language condition does not have

a significant effect on identification accuracy by English-Spanish bilingual listeners. This result showed that fluent bilingual speakers had approximately equal performance in both languages.

Other studies focused on identification of bilingual speakers (in the form of cross-language pairs) by monolingual listeners. Kim [11] shows that monolingual English speakers do not differ from bilingual Korean speakers in identifying a Korean voice from a line-up of Korean-accented English voices. Previous studies [5][12] employed a discrimination design to show that monolingual speakers performed best when they were asked to match voices in their native language, worse in an unknown language, and worst when they coped with cross-language pairs (native vs. unknown). Winters et al. [12] claimed that listeners' judgment depended on whether the speakers used the same language. Wester [5] argued that this cross-language disadvantage is caused by unbalanced source of linguistic and indexical information due to the language switch.

1.3. The present study

This study focused on speaker identification by Cantonese-English bilingual speakers in Hong Kong. No previous studies have incorporated bilingual speakers as both the target voice to be identified and the listeners to identify the voice. In addition, although Wester [5] has tested the different language pairs in his study, he used a discrimination test, not voice line-ups that are usually employed in forensic practice. Previous studies [5][12] demonstrated a cross-language disadvantage for monolingual speakers through native-unknown language pairs. As bilingual speakers are familiar with both languages, it is unknown whether this disadvantage is applicable to bilingual speakers. In order to acquire a complete picture of bilingual speaker identity, the present study has a novel approach in studying speaker identification: having bilingual listeners identifying voices of bilingual speakers from a voice line-up in two language conditions: within the same language (Cantonese and English) or across the two languages. Finally, this study investigates how confident bilingual speakers are when they are making the judgment in different language conditions, and how reliable their confidence is.

2. Method

2.1. Voices and Materials

10 female native speakers of Cantonese who spoke fluent English with a Hong Kong accent were invited to the recording session, which took place in a sound-treated recording booth, with a sampling rate of 22050 Hz. All of them were students at the Chinese University of Hong Kong. They were required to give a long answer to each of the 4 questions (2 in Cantonese and 2 in English). For the Cantonese questions, the speakers were asked to describe: a) a travel experience, b) a dining experience. For the English questions, the subjects were asked to talk about: a) their study at the university, b) their study in secondary school. The recordings were screened by the authors and 3 voices with distinctive characteristics (e.g., harshness and creakiness) were excluded. The remaining 7 female voices, which all had modal-to-breathy phonation, comparable speech rates and similar styles, were used in the identification experiment. One subject's speech samples on dining experience (in Cantonese) and university study (in English) were shortened to two one-minute samples. These samples were used as the "target voice", for familiarization during the exposure in the experiment. Voices of two other speakers were auditorily more similar than others to the target voice, as judged by the authors (Foil A and Foil B). All speakers' (including the target subject) speech samples on travel experience (in Cantonese) and secondary study (in English) were cut into 30-second samples. To minimize the effect of context on identification, the orders of the sentences and phrases were randomized within each narrative, and the randomized versions were used in the voice line-ups. Long hesitations and words with prominent individual characteristics were removed.

2.2. Subjects

79 Cantonese-English bilingual listeners participated in this experiment. They were students at CUHK, speaking Hong Kong Cantonese as their native language and having learned English from an early age. They use English predominantly in academic settings. None of the subjects reported any history of hearing or speech loss or impairment. All listeners received class credit for their participation.

2.3. Procedure

The 79 listeners were randomly assigned to one of the four language conditions: Cantonese target and Cantonese line-up (CC), English target and English line-up (EE), Cantonese target and English line-up (CE), and English target and Cantonese line-up (EC).

The experiment took place in a quiet room. It consisted of three stages: 1) The listeners listened to the target voice via a headphone without being told what they would need to do later (exposure); 2) after exposure to the target voice, they took a five-minute mandatory break by filling in a detailed questionnaire about their language background; 3) after the break, the participants were informed about the task. Further instructions were given and the line-up was played on the computer. The subjects listened to all seven 30-second speech samples completely one after another. At the end of the line-up, they were allowed to repeat any of the samples as many times as they like. Finally, they were asked to decide whether the voice they were exposed to appeared in the parade, and if so, which one. They were also asked to rate the confidence of their judgments on a 9-point scale and give some reasons of their judgment.

3. Results

The identification accuracy and confidence rating were calculated within each condition. The general pattern is shown in Figure 1. The accuracy of judgment is better for the same-language conditions (EE better than CC) than the cross-language conditions (EC better than CE). The confidence rated by the listeners appears in another order. Listeners felt most confident in the CC condition and least confident in the EE condition, with CE and EC in the middle.

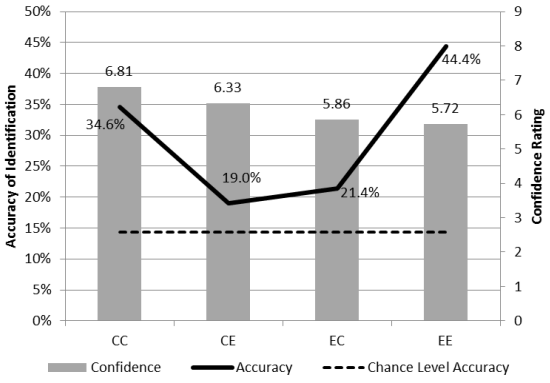

Figure 1: *Accuracy and confidence of all subjects across four conditions.*

3.1. Identification accuracy

Table 1 shows the identification accuracy across the four conditions. The participants in the same-language conditions (CC, EE) outperformed participants in the cross-language conditions (CE, EC). The chance level of the choice is 1/7=14.3%. A chi-square test shows that the performances in the same-language conditions are significantly higher than chance level (the CC condition: $\chi^2(1, N=26)=8.756$, $p=0.0031$; The EE condition: $\chi^2(1, N=18)=13.347$, $p=0.0003$). In contrast, a chi-square test shows that the performances in the different-language conditions are not significantly higher than chance level (the CE group: $\chi^2(1, N=21)=0.386$, $p=0.386$; the EC group: $\chi^2(1, N=14)=0.817$, $p=0.3660$). This indicates that a line-up in the same language as the target voice has positive impact on speaker identification. This result agrees with a previous study [6] that cross-language condition is more difficult than same-language condition.

Table 1. *Identification accuracy across conditions.*

Condition	CC	CE	EC	EE	Total
Correct	9 (34.6%)	4 (19.0%)	3 (21.4%)	8 (44.4%)	24 (30.4%)
Incorrect	17 (65.4%)	17 (81.0%)	11 (78.6%)	10 (55.6%)	54 (69.6%)
Total	26	21	14	18	79

The result has also shown than between the same-language conditions, listeners performed better in the EE condition than the CC condition, probably because the Cantonese accent in the English samples has disclose some extra information of the speaker identity to the listeners. Between the cross-language conditions, listeners were more accurate in the EC condition than the CE condition.

Symmetric measures show that there is an association between language condition and accuracy in the overall data (Cramer's V=0.217). However, there is little association between language condition and accuracy within the same-language conditions (CC vs. EE, phi=0,099) and within the cross-language conditions (CE vs. EC, phi=0.029). The association across all conditions suggests that bilingual listeners may react to the difference between same-language condition and cross-language condition. However, as little association is found within same-language or cross-language conditions, it indicates that as long as the languages of the target and the line-up are the same, which language (English or Cantonese) is spoken does not make a difference. Similarly, as long as the languages of the target and the line-up are different, which language is the target and which forms the line-up also do not matter.

Figure 2 shows the percentage of choices in every language condition. Two foils (A and B) are auditorily very similar to the target voice. They attracted many false alarms. The fact that Foil A and Foil B together account for more errors than all the other foils shows that many listeners were capable of remembering some features of the target voice, even though their memory is not accurate enough for them to identify the target.

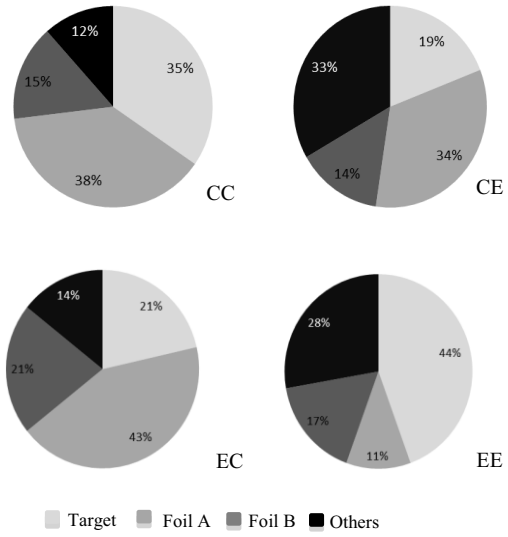

Figure 2: Percentage of choices in each condition.

3.2. Confidence of judgment

As shown in Figure 1, the mean of confidence rated by the listeners decreased in this order: CC (6.8), CE (6.3), EC (5.9), EE (5.7). As native Cantonese speakers, the subjects felt more confident in identifying a voice in Cantonese than in English. This pattern echoes previous findings that listeners appeared to be more confident in a more familiar language. The result here implies that the confidence of judgment is highly sensitive to language background. Even though speakers are bilingual and fairly proficient in English, they were still more confident to judge in Cantonese than in English. Person's correlation shows that there is a strong correlation between condition and confidence rating ($r=0.299$, $p<0.05$). However, no significant difference was found across the means of confidence in all conditions after a Bonferroni correction.

Pearson's correlation shows that there is no correlation between confidence and accuracy. This result, together with previous findings [12][13], calls for strong caution when confidence of identification by the witness is to be taken into consideration in real forensic cases. It can be unreliable.

4. Discussion

This study focuses on the patterns of speaker identification by bilingual listeners. Bilingual listeners identified voices more accurately when the target voice and the voice line-ups are in the same language. However, even though the listeners are more familiar with Cantonese (their native language), their performance in the CC condition did not surpass their

performance in the EE condition. This result agrees with [3], indicating that bilingual speakers in this study have equal abilities in identifying the languages they have mastered.

Between the same-language conditions, the speakers performed slightly better in the EE condition than the CC condition. As shown in [6], a familiar accent offered useful information for speaker identification. It can be assumed that listeners in the present study could have retrieved some extra cues to the speaker identity from the Cantonese accents in the English voices. However, since no association between the language condition and accuracy within the two conditions has been found, further studies are needed to examine the role of accents for bilingual speaker identity.

The result in this study also shows that the cross-language disadvantage on monolingual speakers found in previous studies [5] can be extended to bilingual speakers. Wester [5] argued that this was owing to the mismatch between indexical information and linguistic information. According to him, both sorts of information could be drawn from the native language, but only indexical information (not linguistic information) could be retrieved from a foreign language; as a result, this mismatch would burden the cognitive load and lower the identification accuracy. However, this argument cannot explain our results, as bilingual listeners have the access to both sorts of information from both languages, yet the identification pattern of bilingual speakers is very similar to that of monolingual speakers in [6]. We would argue that it is the misplacement of both linguistic information and indexical information that has caused this cross-language disadvantage. On the one hand, as [9] has shown, linguistic information plays an important part in voice recognition. When the line-up switched to the other language, much of the linguistic information in the target language was lost. On the other hand, it is possible that some indexical information is inherently related to specific languages, and was thus lost in the cross-language pairs.

Although the accuracy of speaker identification in this study is higher than chance level, the false alarm is also very high. Many speakers chose someone with a similar voice (Figure 2), while the highest accuracy rate is only 44% among all conditions. Therefore, earwitness evidence, even though sometimes very useful, needs to be evaluated with much caution.

Last but not least, the confidence rating data have confirmed previous findings that speakers were more confident in recognizing a familiar language. In this study, even though speakers were familiar with both languages, they came from a Cantonese community and were dominant in Cantonese. Thus they had the highest confidence in their Cantonese-Cantonese judgment, followed by the identification of a Cantonese target voice with an English line-up, and worst in their English-English judgment. Previous studies did not provide clear data on the correlation between confidence level and identification accuracy. This study has shown that there is no correlation between confidence and accuracy. As a consequence, the confidence level earwitnesses have should have no bearing on the evidence they provide through voice line-up identification.

5. Conclusions

This study fills a gap in the literature on the identification of voices from bilingual speakers by bilingual listeners, using a voice line-up paradigm. The result shows that some effects previously found on monolingual listeners, such as the cross-language disadvantage, higher confidence for a familiar language, are also applicable to bilingual listeners. It was illustrated that linguistic information and accents play important roles in bilingual speaker identity. We argued that the loss of linguistic information and some indexical information because of the language switch is the reason that causes the cross-language disadvantage. It was also found that there is a high false alarm in the identification; and the confidence of judgment was not correlated to the correctness of the judgment. Therefore, earwitness evidence should be taken with great caution in legal practice.

6. References

[1] Goldstein, A. G., Knight, P., Bailis, K. and Conver, J., "Recognition memory for accented and unaccented voices", *Bulletin of the Psychonomic Society*, 17(5), 217-220, 1983.

[2] Thompson, C. P. "A language effect in voice identification", *Applied Cognitive Psychology*, *1*, 121-131, 1987.

[3] Goggin, J. P., Thompson, C. P., Strube, G., and Simental, L. R. "The role of language familiarity in voice identification. *Memory & cognition"*, *19*(5), 448-458, 1991.

[4] Koster, O., & Schillert, N. O., "Different influences of the native language of a listener on speaker recognition", *Forensic Linguistics*, *4*(1), 1350-1771, 1997

[5] Wester, M., "Talker discrimination across languages", *Speech Communication*, *54*(6), 781-790, 2012.

[6] Sjostrom, M., Eriksson, E. J., Zetterholm, E., & Sullivan, K. P. H., "A bidialectal experiment on voice identification", *Linguistics*, 2008.

[7] Schiller, N. O., and Köster, O., "Evaluation of a foreign speaker in forensic phonetics: A report", *Forensic Linguistics*, *3*(1), 176-185, 1996.

[8] Schiller, N. O., Koster, O., & Duckworth, M., "The effect of removing linguistic information upon identifying speakers of a foreign language", *Forensic Linguistics*, *4*(1), 1350-1771, 1997.

[9] Procter, E. E., "The effect of the distributed learning on the identification of disguised voices", Unpublished MA Thesis, University of Guelph, Guelph, ON, Canada, 2002.

[10] Saslove, H. and Yarmey, H. D., "Long-term auditory memory: speaker identification", *Journal of Applied Psychology*, 65, 111-116, 1980.

[11] Kim, H.-hyun, "Cross-linguistic voice recognition between Korean and English0", *Proceedings of Association for Forensic Phonetics and Acoustics*, 1990-1991, 2007.

[12] Winters, S. J., Levi, S. V., and Pisoni, D. B., "Identification and discrimination of bilingual talkers across languages", *The Journal of the Acoustical Society of America*, 123(6), 4524-38, 2008.