Exploring Sibilant Merge Patterns for Speaker Profiling in Taiwan

Yu-Leng Lin[1,*], Bruce Xiao Wang[2]

1,* Institute of Linguistics, National Chung Cheng University, Taiwan
2. Department of English and Communication, The Hong Kong Polytechnic University
* Lngyllin@ccu.edu.tw

Abstract

This study investigates the evolving linguistic landscape of Taiwan by analyzing the phonetic distinction between alveolar and retroflex sibilants in spontaneous Taiwanese Mandarin speech among speakers from northern and southern Taiwan. Analyzing 2,256 tokens, the study found that place of articulation was the only significant factor influencing center of gravity values, with gender, region, and Min proficiency showing no significant effects. Contrary to earlier studies, this research reveals that contemporary speakers, particularly the younger generation, consistently maintain distinct phonetic contrasts. The findings have implications for forensic phonetics, emphasizing the importance of regional and generational variations in spontaneous speech for speaker identification.

1 Introduction

Taiwan Mandarin (TM) exhibits a unique phonological feature: significant variability in the articulation of alveolar and retroflex sibilants. In standard Mandarin, the phonemic contrasts between alveolars /s ts ts^h/ and retroflexes /ʂ tʂ $tʂ^h$/[1] play a crucial role in phonological identity. This distinction, however, is highly speaker-dependent, showing significant variation across individuals and regions. The merging and/or merger-in-process between alveolar and retroflexes in TM have long been reported and argued in previous studies (Ing, 1984; Kubler,

1985; Lee-Kim & Yun-Chieh, 2022; Steffen Chung, 2006).

1.1 Taiwan Southern Min influence on sibilant merging

Early work attributed this merging trend largely to Taiwan Southern Min (TSM) influence. Studies by Ing, 1984 and Kubler, 1985 claimed that this sibilant merging might arise from language contact with Taiwan Southern Min (TSM) which lacks retroflex, where speakers with higher TSM proficiency are more likely to have merged alveolar and retroflex fricatives than those who have lower TSM proficiency.

However, later studies rejected this claim (Chuang et al., 2019; Lee-Kim & Yun-Chieh, 2022) where they found that TSM proficiency does not necessarily lead to higher degree of sibilant merging. Nevertheless, previous studies consistently reported substantial between-speaker variation in the realization of retroflex, ranging from palato-alveolar to dental/apical.

The current study investigates the sibilant merging in TM, focusing on variations across different regions (i.e., north/south of Taiwan) and genders at both group and individual levels. Unlike other commonly adopted methods, which often involve reading tasks in controlled lab settings, this study used a map-based Q&A approach to encourage spontaneous speech, allowing sibilants of interest to appear naturally in participant responses. Through this approach, the current study aims to contribute to forensic applications within Taiwan, providing insights into how individual and group levels in sibilant

---

[1] We acknowledge that the retroflexes in Mandarin may not align precisely with the traditional articulatory definition, where in the back of the tongue tip contacts postalveolar region. Nevertheless, for consistency with prior studies, we employ these terminologies herein.

merging may serve as distinguishing features in phonetic analysis.

A particular focus is on the hypothesis proposed by early studies (Ing, 1984; Kubler, 1985), which suggests that speakers with higher proficiency in TSM are more likely to merge alveolar and retroflex sibilants than those with lower TSM proficiency. This hypothesis highlighted the potential influence of bilingualism and cross-linguistic interaction on phonological processes in TM. However, subsequent research has challenged this view. Chuang, Sun, Fon, and Baayen (2019) and Lee-Kim and Chou (2022) found that TSM proficiency does not necessarily correlate with a higher degree of sibilant merging, suggesting that the factors influencing this phonetic variation are more complex and may involve other sociolinguistic or individual differences. Despite the differing interpretations, these studies consistently report substantial between-speaker variation in the realization of retroflexes, with realizations ranging from retroflex or post-alveolar/palato-alveolar to dental or apical articulations.

1.2   Implications for forensic phonetics and speaker profiling

Understanding these phonetic nuances is not only theoretically significant but also has applied value in forensic phonetics. Forensic phonetics is a specialized field that focuses on the analysis of speech and voice patterns to aid in legal investigations for various purposes (Jessen, 2007, 2008; Rose, 2002). Forensic phonetics provides valuable tools for such analyses, particularly through speaker profiling. Speaker profiling (Schilling & Marsters, 2015) involves estimating a speaker's gender, regional accent, socioeconomic status, educational background, and other characteristics based on their speech patterns, particularly when only a recording of the offender is available. The goal is to assist law enforcement agencies in narrowing down the origin of the questioned speaker. There are broadly two approaches to speaker profiling. The first is the human-centered phonetic approach, which includes acoustic and auditory analysis (Cambier-Langeveld, 2010). The second is the automated method, which uses an automatic system (Brown, Franco-Pedroso, & González-Rodríguez, 2021). Among these, the phonetic approach is the most widely used. In the context of sibilant merging in TM, the variability among TM speakers presents both challenges and opportunities for forensic experts. On one hand, the lack of consistent phonemic contrasts complicates the task of speaker profiling, as traditional phonetic markers may not be reliably present. On the other hand, the unique patterns of sibilant merging within an individual's speech could serve as valuable idiosyncratic features for speaker profiling, especially when traditional phonetic markers are ambiguous or absent.

The current study investigates the phenomenon of sibilant merging in TM, focusing on variations across different regions and gender at both group and individual levels. By analyzing these variables, the study aims to uncover patterns that could be potentially used for speaker profiling in relation to region and gender factors in Taiwan. The findings are expected to contribute to the field of forensic phonetics by providing insights into how regional and gender-based differences in sibilant merging might inform speaker profiling processes. Ultimately, this research underscores the importance of understanding phonetic variation in TM not only as a linguistic phenomenon but also as a critical tool in forensic casework, where the ability to accurately profile speakers can have profound implications for justice and legal outcomes.

The paper is organized as follows. Section 2 outlines the methodology. Section 3 presents results and discussion. Section 4 concludes with an overall discussion and considers the implications for future research.

2   Method

2.1   Participants

Twenty participants, aged 18 to 35, were recruited from National Chung Cheng University in Taiwan. They were evenly divided by gender and region, with 5 males and 5 females from both northern and southern Taiwan. Northern Taiwan was defined as Taipei, Taoyuan, Kinlong, Hsinchu, or Miaoli, while southern Taiwan included Chiayi, Tainan, Kaohsiung, or Pingtung. All participants had lived in their respective regions until the age of 18. The average age of the participants was 23.71 years (SD = 3.36).

All twenty participants reported exposure to Min between infancy and nine years of age. In addition, they all spoke English as a second language, having begun learning it between the ages of 2 and 12. To assess the influence of TSM, participants rated their listening and speaking proficiency in TSM using a seven-point Likert scale, where 1 indicated that participants barely knew the language and 7 represented native-level proficiency. TM and English language proficiency self-rated by participants are shown in Table 1.

2.2 Mandarin production task

To elicit the most natural language data, the experiment was designed in a Q&A format. Participants were shown a printed map of the Taipei MRT and were asked questions such as, "How many stops are there between station A and station B?" and "How do you travel from point A to point B?" Because the names of MRT stops did not include [ts] and [tʂʰ], additional questions were asked to elicit target words containing these sibilants. This approach encouraged participants to focus on finding the correct answers, leading them to pay less attention to their speech. The target words elicited for the production experiment are listed in Table 2. The consonants consisted of voiceless coronal fricatives and affricates, specifically alveolars /s ts tsʰ/ and retroflexes /ʂ tʂ tʂʰ/. These consonants were placed in the syllable-initial position of the first syllable, followed by the nuclei [i], [a], [o], [u], and [ɨ], with or without coda consonants.

2.3 Procedure

Before the formal experiment began, the experimenter briefly introduced the participants to the different MRT lines and the cardinal directions (north, south, east, west) on the map, and then proceeded with the questions. Participants were instructed to give detailed answers without looking at the written questions and to use the full names of MRT lines and stations, avoiding abbreviations such as "red line." A total of 59 questions were asked, but only those containing alveolars and retroflexes were analyzed. The number of keywords with alveolars and retroflexes was balanced, with five keywords featuring alveolars in the initial syllable and five featuring retroflexes. Each keyword was elicited at least ten times. The Q&A session lasted approximately 25 to 40 minutes, and each participant received monetary compensation for their time.

2.4 Acoustic and statistical analysis

1ˢᵗ syllable consonant	PoA	IPA	English gloss
[ts]	alveolar	[**ts**uei] [tɛɪn]	"nearest"
[tsʰ]	alveolar	[tɕʰɨ] [tʂɑŋ]	"Qizhang MRT station"
[s]	alveolar	[san] [tʂʊŋ]	"Sanchong MRT station"
[tʂ]	retroflex	[tʂɨ] [ʂan]	"Zhishan MRT station"
[tʂʰ]	retroflex	[tʂʰɨ] [fan]	"eat"
[ʂ]	retroflex	[ʂɨ] [pʰaɪ]	"Shipai MRT station"

Table 2: Sample elicited words.

This study analyzed the Center of Gravity (CoG) of specific keywords. The data, collected from a TM production task, were recorded at a 44.1 kHz sampling rate and preprocessed by filtering out frequencies below 500 Hz (using a pass Hann band filter with a 500 Hz lower limit and 22050 Hz upper limit in Praat (Boersma & Weenink, 2024; version 6.4.16). The annotated sound files marked the onset of frication at the first

		F	M
N	TM listening	7.00 (±0.00)	7.00 (±0.00)
	TM speaking	6.80 (±0.45)	7.00 (±0.00)
	TSM listening	3.60 (±1.52)	4.60 (±2.19)
	TSM speaking	3.20 (±1.10)	4.20 (±2.39)
	English listening	4.60 (±1.14)	5.20 (±0.84)
	English speaking	4.20 (±1.30)	4.80 (±0.84)
S	TM listening	6.80 (±0.45)	7.00 (±0.00)
	TM speaking	6.80 (±0.45)	6.80 (±0.45)
	TSM listening	5.20 (±0.84)	4.80 (±1.10)
	TSM speaking	3.80 (±2.28)	4.20 (±1.92)
	English listening	4.00 (±0.71)	4.60 (±1.67)
	English speaking	3.00 (±1.22)	4.20 (±1.92)

Table 1: Mean language proficiency ratings and standard deviation (in brackets) on a seven-point scale (1= barely know, 7 = native-level).

appearance of white noise above 1000 Hz in the spectrogram and the end at the beginning of the following vowel (Li, 2008). For spectral analysis, a 23.2 ms Hamming window was applied to the midpoint of the frication. Measurement values were extracted using a custom-written Praat script, with two measurements taken: one spectral moment of the frication noise and the CoG.

A total of 3440 tokens were collected from participants' responses to 59 questions. However, only the keywords containing retroflex ([ʂ, tʂ, tʂʰ]) and dental ([s, ts, tsʰ]) sibilants were analyzed. In total, 2256 tokens from twenty participants were
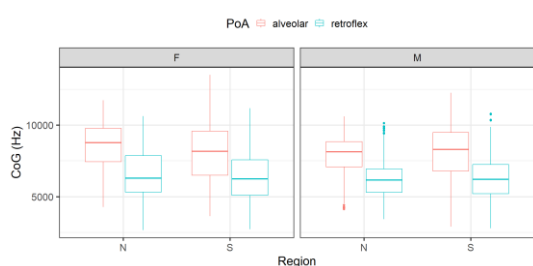
Figure 1. Interaction between PoA, region and gender (N: North, S: South).

included in the analysis. A linear mixed-effects model was used, implemented in the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R version 4.4.1 (R Core Team, 2024).
The independent variables in the study included place of articulation (PoA: alveolars vs. retroflexes), region (Southern Taiwan vs. Northern Taiwan), gender (male vs. female), and Min proficiency (measured on a Likert scale from 1 to 7). A linear mixed-effects model was applied to the data, with PoA, region, and gender as fixed effects, and word and speaker as random effects. Further, we conducted *Pearson's* correlation test to investigate if speakers with higher TSM proficiency would have higher degree of alveolar-retroflex merging, namely, a lower difference in the CoG values.

3   Results and discussion

The linear mixed-effects model analysis revealed that none of the fixed effects, except for PoA, had a significant impact on the center of gravity (CoG) values. Specifically, the results indicate that alveolars and retroflexes remained distinct and separable based on their CoG values

($p < .01$). However, there were no significant differences in CoG values between male and female participants or between speakers from northern and southern Taiwan. See Figure 1.

These results contrast with earlier findings (Ing, 1984; Kubler, 1985), potentially due to differences in the participant groups. The current participants belong to a younger generation characterized by increased mobility, facilitated by improved inter-city transportation in Taiwan and the widespread promotion of Mandarin. In earlier studies, participants typically acquired TSM as their first language and only began learning TM in elementary school. In contrast, the present generation receives a more diversified language education. TSM is just one of several domestic languages offered in elementary schools, alongside Hakka, indigenous languages, Mindong, and Taiwan Sign Language. However, these languages are taught for only one hour per week, and not all students choose to study TSM. As a result, participants in the current study had less exposure to TSM compared to earlier cohorts. The hypothesis that higher proficiency in TSM correlates with a greater likelihood of merging alveolars and retroflexes assumes that TSM is the dominant language, with TM playing a secondary role. In such cases, greater TSM proficiency would be expected to reduce the acoustic differences—such as in the CoG values—between alveolars and retroflexes in TM. However, the participants in the present study experience a reversed linguistic environment: TM is now the dominant language, while TSM is less prominent. This linguistic shift likely contributes to the maintenance of the contrast between alveolars and retroflexes in TM. It is plausible that these participants, who are proficient in TM, TSM, and English, are better equipped to preserve the distinction between these sounds rather than merging them. This interpretation is further supported by the data presented in Figure 2. The figure demonstrates that participants with higher Min proficiency were more likely to maintain a clear distinction between alveolars and retroflexes, although the correlation is low (*Pearson's r* = .08).
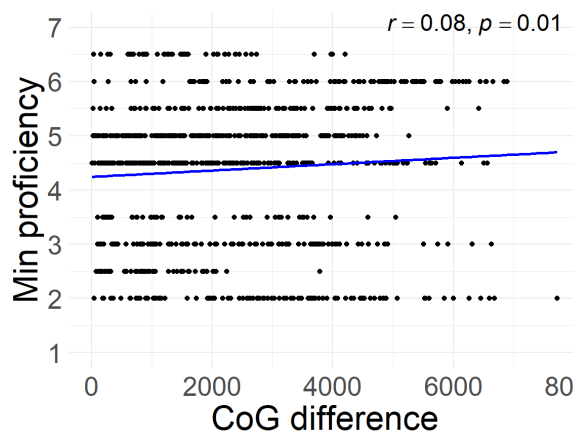
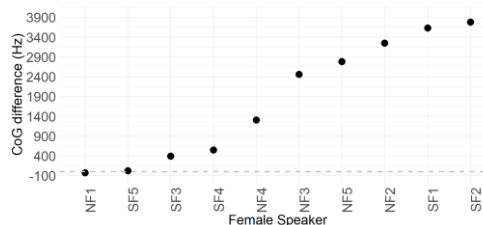Figure 2. Correlation between Min proficiency and CoG difference.

Figure 3. CoG difference of PoA produced by females.
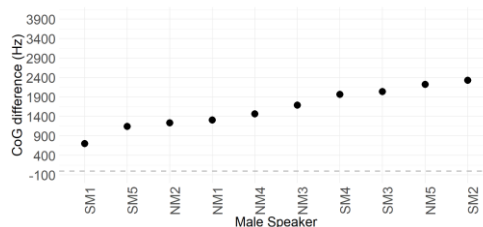
Figure 4. CoG difference of PoA produced by males.

Examining individual speakers reveals findings that support the argument by Lee-Kim and Chou (2022) that an ongoing sound change is occurring in the younger generation, where the variation between alveolar and retroflex sibilants spans a continuum—from complete merger to distinct contrasts. Although there was no significant main effect of gender, Figures 3 and 4 suggest that females exhibited greater variation in their sibilant production compared to males. Specifically, the range in CoG values between alveolars and retroflexes was -100–3900 Hz for females, while for males, the range was narrower, at 700–2400 Hz. This suggests that females generally maintained a more pronounced separation between the two contrasts, with a total range difference of 4000 Hz for females, compared to 1700 Hz for males. This finding aligns with Labov's sociolinguistic claim that women tend to favor and maintain more prestigious linguistic forms, particularly those that preserve phonological contrasts (Labov, 2001).
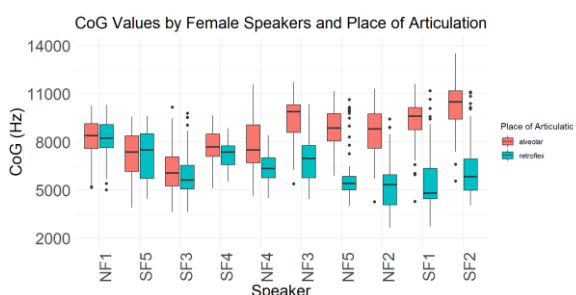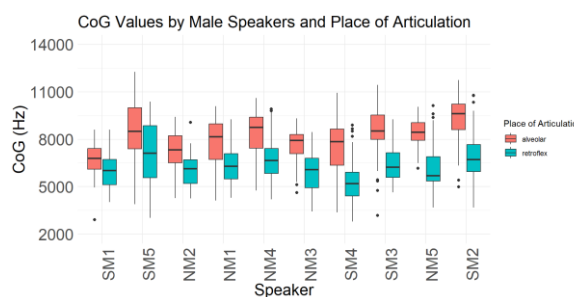

Contrary to earlier studies (Ing, 1984; Kubler, 1985), southerners in this study appeared to maintain a more pronounced separation between alveolars and retroflexes, with a total range difference of 4000 Hz, while northerners exhibited a smaller range difference of 3400 Hz (see Figures 5 and 6). This may be linked to our earlier argument that individuals proficient in TM, TSM, and English are better equipped to preserve distinctions between these sounds rather than merging them. Further supporting this interpretation, Table 1 shows that southerners demonstrated higher Min proficiency, with TSM listening scores of 5.00 and TSM speaking scores of 4.00, whereas northerners scored 4.10 in listening and 3.70 in speaking[2].

---

[2] According to Lee-Kim and Chou (2022), scores higher than 5 are considered fluent, while those below 3 are considered weak. In the current study, southerners' TSM listening scores fall into the fluent category.
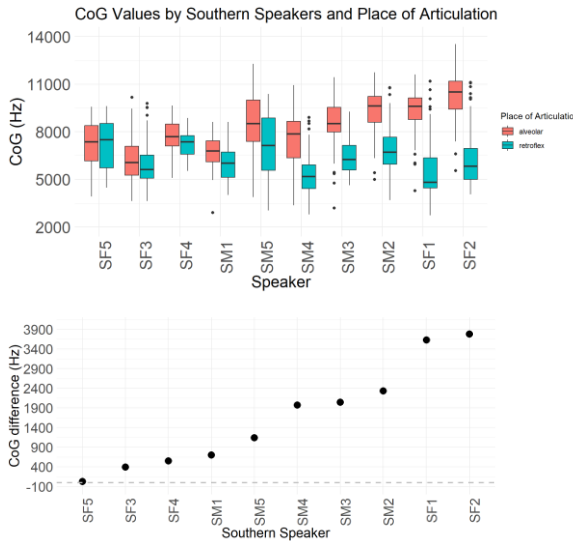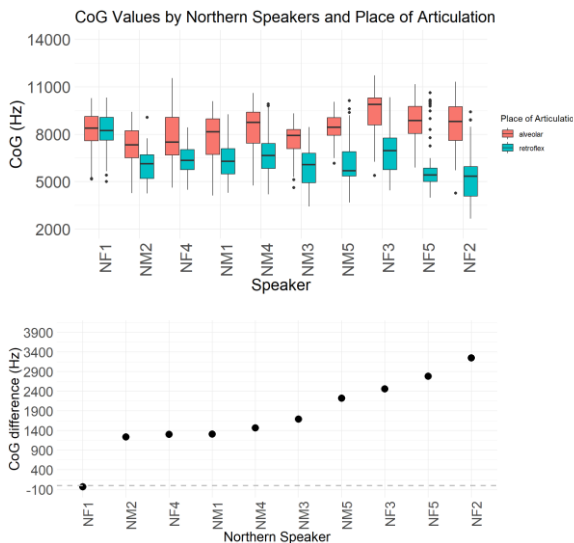
Figure 5. CoG difference of PoA produced by southerners**.**

Figure 6. CoG difference of PoA produced by northerners**.**

4   Implications and discussion

This study sheds light on the evolving linguistic landscape of Taiwan by examining the phonetic distinction between alveolar and retroflex sibilants in TM among speakers from northern and southern Taiwan. Through analyzing 2256 tokens containing retroflex and dental sibilants, the study reveals that PoA was the only significant factor influencing CoG values, while gender, region, and Min proficiency did not produce significant effects. The participants, regardless of background, consistently maintained distinct phonetic contrasts between alveolars and retroflexes.

The results reflect a broader phenomenon of language change. As Lee-Kim and Chou (2022) suggest, ongoing sound changes in younger generations indicate that variation between alveolar and retroflex sibilants now spans a continuum—from complete merger to clear contrasts. This study's findings align with this idea, highlighting that even in a shifting linguistic environment, contrasts are maintained, though the degree of separation can vary.

Gender also plays a nuanced role in this evolving linguistic landscape. Although there was no significant main effect of gender, female participants exhibited greater variation in their sibilant production. The wider range in CoG values observed among females suggests that they are more likely to preserve contrasts than males. This observation echoes Labov's (2001) claim that women prefer and maintain more prestigious linguistic forms, reinforcing the idea that sociolinguistic factors intersect with phonetic variation.

Regional differences in phonetic behavior further emphasize the dynamic nature of language change. In contrast to earlier studies, which suggested greater phonetic merging in southern Taiwan, this study found that southern speakers maintained a more pronounced separation between alveolars and retroflexes. This may be linked to higher language proficiency among southern participants, who demonstrated greater proficiency in TSM and TM. These findings point to the complex and adaptive nature of language use across different generations, regions, and linguistic experiences in Taiwan.

In terms of speaker profiling for forensic applications, the merging patterns of alveolar and retroflex sibilants might not serve as a reliable indicator for distinguishing regional accents. Traditionally, southern speakers are more likely to exhibit a merged accent compared to northern speakers. However, this was not observed in our analysis. This discrepancy is likely due to the shifting linguistic landscape of sibilant merging in Taiwan, influenced by changes in language policy and increased mobility. While phonemic contrasts might not always align with traditional expectations, the unique patterns of sibilant articulation could still provide valuable, individualized markers for forensic experts. As linguistic behavior evolves in Taiwan, understanding these nuances becomes crucial for

accurate speaker profiling. This research not only contributes to the linguistic field but also underscores the practical applications of phonetic analysis in legal contexts, where subtle speech patterns may aid in delivering justice.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Boersma, P., & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program].

Brown, G., Franco-Pedroso, J., & González-Rodríguez, J. (2021). A segmentally informed solution to automatic accent classification and its advantages to forensic applications. *International Journal of Speech, Language and the Law, 28*(2), 201-232.

Cambier-Langeveld, T. (2010). The role of linguists and native speakers in language analysis for the determination of speaker origin. *International Journal of Speech, Language & the Law, 17*(1).

Chuang, Y.-Y., Sun, C.-C., Fon, J., & Baayen, R. H. (2019). Geographical variation of the merging between dental and retroflex sibilants in Taiwan Mandarin.

Ing, R. (1984). Issues on the pronunciations of Mandarin. *The World of Chinese Language, 35*, 6-16.

Jessen, M. (2007). Speaker classification in forensic phonetics and acoustics. *Speaker classification I: fundamentals, features, and methods*, 180-204.

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass, 2*(4), 671-711.

Kubler, C. C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics, 27*(2), 156-176.

Labov, W. (2001). Principles of linguistic change, vol. 2: Social factors Oxford. *UK: Blackwell*.

Lee-Kim, S.-I., & Chou, Y.-C. (2022). Unmerging the sibilant merger among speakers of Taiwan Mandarin. *Laboratory Phonology, 13*(1).

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese.* The Ohio State University,

Rose, P. (2002). *Forensic speaker identification*: cRc Press.

Schilling, N., & Marsters, A. (2015). Unmasking identity: Speaker profiling for forensic linguistic purposes. *Annual Review of Applied Linguistics, 35*, 195-214.