JASA

ARTICLE

The effects of ethnic bias and face on identification, accentedness judgements and intelligibility of Cantonese accents in Hong Kong

Grace Wenling Cao,[1] Him Cheung,[2] and Peggy Mok[3,a]

[1]*School of Languages, Cultures and Linguistics, University College Dublin, Dublin, Ireland*

[2]*School of Psychology, Speech and Hearing, University of Canterbury, Christchurch, New Zealand*

[3]*Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong Special Administrative Region of China*

ABSTRACT:

Social information such as a talker's ethnicity, gender, and age are found to affect accent perception and attitudes. While existing research primarily focuses on English-dominant communities, this study aims to fill the gap by examining the impacts of ethnic bias and face on three Cantonese accents in Hong Kong. Nine groups of 20 Hong Kong Cantonese listeners were exposed to three Cantonese accents (i.e., Hong Kong local Cantonese, Mandarin-accented, and English-accented Cantonese) in three conditions of visual cues (i.e., a silhouette, a South Asian face and a White face). For accent identification, seeing a South Asian face in a mismatch condition led to more errors compared to seeing a White face in the same condition. For intelligibility, an enhancement of intelligibility was found when the face and accent were misaligned (e.g., an English accent matched with a South Asian face), supporting the general adaptation mechanism instead of the expectation mechanism. We argue that listeners might perceive South Asian and White faces as the same broad social category "foreigners/outgroup members," resulting in a similar enhancement effect in the aligned and misaligned conditions. A dual-activation mechanism is proposed to account for the complementary effect of phonological and visual cues on accent perception. © 2025 Acoustical Society of America.

https://doi.org/10.1121/10.0035783

I. INTRODUCTION

Hong Kong is a diverse and multilingual metropolis, with the majority ethnic group in Hong Kong being Chinese (91.6%, HKSAR Census and Statistics Department, 2021). Among the ethnic minorities, Southeast Asians accounted for 4.6%, South Asians for 1.4%, Whites for 0.8%, and other Asians for 0.5%. Among the different minority groups, their social statuses are not equal. Ethnic white people have historically enjoyed privileged status due to Hong Kong's colonial history, while South Asian minorities often face marginalization in society. From the 1840s to 1997, Hong Kong was under British colonial rule, with English-speaking white people in power. During this period, white people benefited from legal and employment privileges, asserting a higher social status than the local Chinese (Carroll, 2007).

South Asian minorities first settled in Hong Kong during the Opium War of 1839–1942, primarily for British military service, labour, trade, and law enforcement purposes (Erni and Leung, 2014a). For instance, Indian troops and merchants started to settle in Hong Kong during the British invasion in the 1840s (Kwok and Kirti, 2003), and Pakistanis began to come to Hong Kong in the 1960s after

the independence of their country (Erni and Leung, 2014a). Due to the language barriers, cultural differences, and the segregation policy of British colonial rule, the interactions between South Asians and the local Chinese were limited during the colonial time. After the handover in 1997, the segregation between Hong Kong Chinese and South Asian ethnic minorities continued. Ku *et al.* (2006) found that 69% of the 200 South Asian minorities in their survey were either fair or poor in speaking and understanding Cantonese or did not know how to speak or understand it at all. Over 75% of the South Asian respondents were engaged in elementary occupations or worked as service workers, shop sales, or clerks. When they were asked whether Hong Kong people treat people of different race equally, over 62% of them disagreed or strongly disagreed. Chan and Wong (2005) conducted another survey investigating Hong Kong people's attitudes towards ethnic minorities. They found that over 60% of Hong Kong Chinese respondents agreed that Hong Kong people have negative perceptions towards South Asian/Southeast Asian ethnic minorities, and 41% of the respondents reported having no acquaintances who are South Asian minorities. All these reports show that South Asian minorities are often neglected and discriminated against in their education, employment, and health care (Erni and Leung, 2014b).

a)Email: peggymok@cuhk.edu.hk

Given the different social dynamics between the three race groups (i.e., Hong Kong Chinese, White, and South Asian minorities), Hong Kong people have various attitudes towards different English accents and Cantonese accents used by these groups. For instance, people in Hong Kong perceive inner circle English varieties such as American English, British English, or Australian English as more prestigious and attractive than Indian English and Filipino English (Chan, 2016; Zhang, 2009). Studies on Hong Kong people's attitudes towards Cantonese accents tend to focus on Mandarin-accented Cantonese (Gu, 2011; Shum *et al.*, 2023), also known as Tamjai accent (Yip and Catedral, 2021). For example, Gu (2011) found that mainland Chinese immigrant students who do not speak Cantonese or speak non-native accents of Cantonese were teased by their classmates and felt excluded from the local community. To our knowledge, there are no studies on Hong Kong people's attitudes towards English-accented Cantonese or Cantonese spoken by South Asian minorities. Thanks to a few White actors who were active in Hong Kong public media, such as Mr. Gregory Charles Rivers (河國榮) from Australia and Mr. Brian Thomas Burrell (布偉傑) from the US, Hong Kong people are not unfamiliar with English-accented Cantonese. In comparison, there is no clear prototype or standard representation of South Asian Cantonese in Hong Kong. In daily interactions, Hong Kong Chinese tend to communicate with South Asians in English. Given that studies on people's attitudes towards different Cantonese accents are scarce, this study aims to fill this gap and explore Hong Kong people's awareness and attitudes towards Cantonese accents.

A. The impact of face on accent perception

Previous studies have found that social cues, including a talker's ethnicity, gender, sexuality and age can affect listeners' perception of speech (Babel and Mellesmoen, 2019; Babel and Russell, 2015; Drager, 2011; Hay *et al.*, 2006; Johnson *et al.*, 1999; McGowan, 2015). The hybrid models of exemplar-based theories, such as those proposed by Johnson (2006) and Pierrehumbert (2001, 2002) are commonly used to explain the impact of social cues. According to the exemplar-based framework, social cues are stored alongside linguistic cues in the exemplars. When thousands of exemplars accumulate, abstract social categories such as "Asian/South American" or "Chinese/Brazilian" are formed and subsequently affect perception through top-down tunning. Three related mechanisms have been proposed to explain the different effects of visual cues on accent perception. The first mechanism suggests that exposing listeners to visual cues that are associated with social bias could inhibit the perception of non-native accented speech (Kang and Rubin, 2009; Kutlu *et al.*, 2022a; Rubin, 1992; Yi *et al.*, 2013). Both Rubin (1992) and Kang and Rubin (2009) tested American undergraduates' accent ratings and comprehension of native English, which was presented alongside either an Asian face or a White face. The results of both studies revealed that when the audio was paired with an Asian face, listeners perceived the speaker as having a stronger accent

and performed less effectively in comprehension tests compared to when the audio was paired with a White face.

The second mechanism argues that it is not the bias but people's expectations that affect accent perception. The expectation mechanism suggests that listeners' expectations of the accent could be manipulated by the visual cues presented; if their expectations match with the accent they hear, a facilitation effect should be observed. One example of this mechanism is McGowan (2015). The study tested American English listeners' transcription of Mandarin-accented English, in incongruent (e.g., a White face) and congruent (e.g., an Asian face) conditions. Results suggested that the congruent conditions outperformed the incongruent conditions for both experienced and inexperienced listeners. McGowan (2015) explained that the expectation arose from seeing an Asian face matched with hearing Mandarin-accented English, which in turn facilitated the intelligibility of accented speech.

Closely related to the expectation mechanism, the third mechanism proposes that listeners' intelligibility of different accents could vary depending on whether they apply targeted adaptation or general adaptation to accent perception (McLaughlin and Van Engen, 2023; Melguy and Johnson, 2021). While the expectation mechanism tends to focus on the (in)congruency between expectation and accents, the third mechanism addresses whether the same facial cues could give rise to different expectations based on the adaptation strategies listeners apply. For instance, Melguy and Johnson (2021) examined the visual facilitation effect when listeners' social expectations (e.g., seeing a European, East Asian, or South Asian face) aligned with the Mandarin accent either on a general level or on a targeted level. They predicted that if listeners employed a targeted adaptation strategy, they would perform better when the Mandarin accent was matched with the East Asian face compared to the cases when the Mandarin accent was matched with the European or South Asian face. If listeners used a general adaptation strategy, they were predicted to have similar performance when the Mandarin accent was matched with the East Asian prime or the South Asian face because both faces would activate a general category of social cues associated with Asian speakers. In contrast to their predictions, they found no significant effect of visual cues on intelligibility.

Although no significant effect of the face was found, Melguy and Johnson (2021) brought in an important discussion: people's social categorisation of the face (i.e., ethnicity of the speaker) might affect the adaptation strategies in accent perception. Depending on experience, people can categorise faces at varying levels of specificity. For instance, Indian people can be categorised into the social group of "Indians" at a specific level; they can also be broadly categorised as "Asian" at a broader level. Another perspective of categorising face is based on ingroup or outgroup membership. For instance, studies of face categorisation suggest that people perform better in recognising ingroup members compared to outgroup members (Bernstein *et al.*, 2007; Lakshmi *et al.*, 2021; Meissner and Brigham, 2001). Ray

J. Acoust. Soc. Am. **157** (3), March 2025

Cao *et al.* 1619

et al. (2010) also found that people strictly applied an ingroup-outgroup distinction in face recognition. Once an outgroup feature was detected, a member would be excluded from ingroup membership even though she/he might have presented other ingroup features, leading to a lower recognition rate.

B. The impact of attitudes and other social factors on accent perception

Although several key studies (Kang and Rubin, 2009; McGowan, 2015; McLaughlin and Van Engen, 2023; Melguy and Johnson, 2021; Rubin, 1992) have contributed to the development of the three mechanisms, they tend to interpret the influence of face on accent perception as the impact of the visual cues. Kutlu et al. (2022a) argued that it was not the additional visual cues that mediate the accent perception but the social implications associated with a particular race. More specifically, Kutlu et al. (2022a) investigated American college students' accent perception of South Asian English, American English, and British English matching with a White (ingroup) or a South Asian face (outgroup). They found that Indian English matched with White faces was transcribed more accurately than Indian English matched with South Asian faces. Such an effect was also observed in ratings of accentedness: American English was rated as more accented when it was matched with South Asian faces than White faces. The emphasis on the impact of race and its associated social information (ingroup vs outgroup) would urge more consideration of social factors when developing the mechanisms for accent perception.

Previous studies have identified several social factors that can influence accent perception, including the experience of listeners (Babel and Russell, 2015; McGowan, 2015; McLaughlin and Van Engen, 2023), the social network and locale of the listeners (Kutlu et al., 2022a,b) and language attitudes (Babel and Russell, 2015; Kang and Rubin, 2009; Yi et al., 2013). Babel and Russell (2015) investigated implicit and explicit measures of ethnic bias towards Asian Canadians in relation to accent perception and ratings. Interestingly, their findings indicated that neither type of attitude predicted listeners' transcription and accent ratings. Another study by Yi et al. (2013) discovered that listeners who developed stronger associations between ethnic White faces and Americans and between Asian faces and foreign individuals were more likely to exhibit enhanced transcription abilities for native English sentences than nonnative sentences when presented with audio-visual cues. Kang and Rubin (2009) measured listeners' attitudes using a matched-guise design (Zahn and Hopper, 1985), where participants were led to believe that they were listening to audio produced by a native Chinese speaker. The study revealed that when primed with this belief, listeners rated native English recordings as more accented and less intelligible.

Another research gap is that previous studies mainly focused on English-dominant and White-dominant societies, such as the US and Canada. The non-native accents in these societies are often associated with attitudes like unattractive (Babel and Russell, 2015; Kang and Rubin, 2009; Yi et al., 2013). The homogeneity of societal backgrounds in previous studies makes it less applicable to other non-English dominant multilingual societies. As reviewed in Sec. I A, Hong Kong has diverse racial groups with different social dynamics, making it a perfect locale to expand accent perception research to a non-White and non-English speaking community.

C. The present study

This study explores the impact of different visual cues (i.e., a silhouette, a South Asian face and a White face) on Hong Kong Chinese's perception of three Cantonese accents: local Cantonese, Mandarin-accented Cantonese and English-accented Cantonese.

The silhouette condition was chosen for two reasons. First, it serves as a baseline to compare with the other two conditions where different racial cues are presented. This allows us to explore a broader mechanism of how voice and/or face affect accent perception. Second, it serves as a condition to test the bias mechanism (see Table I). The choice of a White face and a South Asian face was based on the historical and sociolinguistic contexts of Hong Kong. As reviewed in Sec. I A, White people in Hong Kong have enjoyed privileges since the colonial period. Previous studies also confirm that inner circle English, such as American English and British English, is favoured by Hong Kong people (Cao, 2018, 2024; Chan, 2016; Zhang, 2009). When Cantonese accents are matched with a White face, we predict that positive attitudes would be elicited among Hong Kong Chinese participants. In contrast, South Asian minorities in Hong Kong have been marginalised in their education, health care, and career development. Previous research also found that 60% of Hong Kong Chinese respondents agreed Hong Kong people have negative perceptions towards South Asian/Southeast Asian ethnic minorities (Chan and Wong, 2005). Therefore, we predict that Cantonese accents paired with a South Asian face would elicit relatively more negative

TABLE I. Predictions based on three mechanisms for Research Question 2.

Visual cues	Local Cantonese (CAN)			Mandarin accent (MAN)			English accent (ENG)		
	Silhouette	South Asian	White	Silhouette	South Asian	White	Silhouette	South Asian	White
Bias mechanism	no bias > bias								
Expectation mechanism								mismatch <	match
Adaptation mechanism					general <				match
		general =	general					general =	general

attitudes compared to the condition where Cantonese accents are paired with a White face.

The selection of the three Cantonese accents is based on the degrees of familiarity within the local Hong Kong community. The local Cantonese accent was selected to represent the standard variety of Hong Kong Cantonese, as well as the most familiar accent for Hong Kong people. The other two accents were Mandarin-accented Cantonese and English-accented Cantonese. The Mandarin accent is more familiar to Hong Kong people compared to English-accented Cantonese. The presence of this Cantonese accent has become increasingly common in Hong Kong due to a significant influx of immigrants from Mainland China. According to the Hong Kong Population Census, from 1997 to 2023, approximately 1.18×10^6 mainland Chinese immigrants settled in Hong Kong through the One-way Permit Scheme (HKSAR Census and Statistics Department, 2023). This means that, on average, one out of every seven residents in Hong Kong would be a "new immigrant" from Mainland China. As Mandarin is the lingua franca in Mainland China, many mainland Chinese immigrants would likely speak Cantonese with a Mandarin accent (Yip and Catedral, 2021). In contrast, ethnic White residents only account for 0.8% of the Hong Kong population and only 7.1% of them can speak Cantonese (HKSAR Census and Statistics Department, 2021).

The reason for not choosing South Asian Cantonese is twofold. First, among English-accented Cantonese, Mandarin-accented Cantonese, and South Asian Cantonese, South Asian Cantonese is most unfamiliar to Hong Kong people. According to the HKSAR Census and Statistics Department (2021), only 8.5% of South Asians in Hong Kong are able to speak Cantonese. In daily interactions, Hong Kong Chinese tend to communicate with South Asians in English, as 27% of South Asians in Hong Kong reported being able to speak English, which is significantly higher than the 8.5% who can speak Cantonese. Moreover, the diverse groups within the South Asian community in Hong Kong, including Indian, Pakistani, and Nepalese suggest that it would be difficult to select a specific South Asian Cantonese for the study. Based on these practical reasons, South Asian Cantonese was not selected. Although people who speak Cantonese with an English accent in Hong Kong come from different English-speaking countries, the variation of their accents is smaller compared to South Asians whose first language is Hindi, Panjabi, or Nepalese. Thus, English-accented Cantonese was selected.

This study asks two research questions:

- Research Question 1: Does face affect people's accentedness ratings and attitudes towards the three Cantonese accents?
  Building upon previous findings and considering the historical and sociolinguistic contexts of Hong Kong, we predict that Hong Kong people's accentedness ratings and attitudes may become more negative when seeing a South Asian face compared to seeing a White face.

- Research Question 2: Does face influence the intelligibility of the three Cantonese accents? If yes, which mechanism(s) can best account for the findings?

Our predictions for Research Question 2 are summarised in Table I. If the bias mechanism is correct, we would expect that in the local Cantonese condition, seeing a South Asian face would inhibit intelligibility compared to seeing a White face and a silhouette.

Following the expectation mechanism, we would expect that in the English-accented condition, listeners in the match condition (i.e., seeing a White face) would outperform the mismatch condition (i.e., seeing a South Asian face).

For the adaptation mechanism, we would expect that people's adaptation strategies are influenced by their social categorisation of the faces. If they categorise the face at a specific level (e.g., categorising a White face as "English people"), a targeted category would be activated when a White face is paired with the English accent. If they categorise the face at a broader level (e.g., categorising a White face as "non-native/foreign speakers/outgroup members"), a general category might be activated when a White face is paired with the Mandarin accent. In this case, we would expect listeners to have better intelligibility in the targeted condition (a White face + ENG accent) than in the general condition (a White face + MAN accent).

Another possibility is that Hong Kong Chinese listeners might categorise both a White face and a South Asian face as a broader social group "non-native speakers/foreigner/ outgroup members," and a general category would be activated. In this case, we would expect the same facilitation effect whether listeners see a South Asian face or a White face. If this effect is robust, we should observe this in both Mandarin and English accents.

II. METHOD

A. Participants

For the study, 180 participants (90 females and 90 males) were recruited from a university in Hong Kong through a university Mass Mail system. They were all native speakers of Hong Kong Cantonese, and their ages ranged from 19 to 39 years old (mean = 22 years old). Participants were equally divided into nine groups of 20 people and each group was assigned to one of the nine conditions (three accent conditions × three face conditions). Equal numbers of female and male participants were assigned to each group.

Regarding participants' language background, all participants were proficient in both English and Mandarin. Participants spoke Hong Kong English (Cao, 2018), an emerging new variety of English with Cantonese influence based on British English. Their English proficiencies ranged from 3 to 5** in the Hong Kong Diploma of Secondary Education Examination (HKDSE), which are equal to 5.5 to 7.5 in the International English Language Testing System (IELTS) framework. For their Mandarin proficiency, their self-report proficiency on a scale of 1–5 was used to assess

J. Acoust. Soc. Am. **157** (3), March 2025

Cao *et al.* 1621

their Mandarin level, as not many participants had taken the standardised test of Mandarin proficiency. The average self-report proficiency in Mandarin was 4.2 (listening), 3.7 (speaking), 4.2 (reading) and 4.0 (writing), and their self-report proficiency in English was 4.0 (listening), 3.7 (speaking), 4.2 (reading) and 3.8 (writing).

B. Stimuli

1. Audio

Forty-five Cantonese sentences (see the supplementary materials in Cao, 2025; Sec. A) were created for this project. Each sentence contained seven to eight Chinese characters and was placed in unpredictable contexts, such as "明仔嘅阿爸係逃犯" (English translation: Father of Ming is a fugitive).

Three male speakers including a native Cantonese speaker, a Mandarin-accented Cantonese speaker, and an English-accented Cantonese speaker were recruited to read the sentences in their natural accent and at a normal speed. Their readings were recorded in Praat at the sampling rate of 44.1 kHz in a quiet room. The native Cantonese speaker was born and grew up in Hong Kong and spoke Cantonese with a local accent. The Mandarin-accented Cantonese speaker was a native speaker of Mandarin Chinese who had been working in Hong Kong for three years by the time the recordings were made. The English-accented Cantonese speaker was a Canadian White person who had been living and working in Hong Kong for over ten years.

2. Face

Five different images of male South Asian faces and male White faces were each selected from the Internet. Five faces were selected because we wanted to elicit the general category of each type of face instead of a particular face. To reduce the impact of judgemental bias resulting from people's socioeconomic status, the figures depicted in the images were all dressed in suits or wearing shirts. A post-experiment test was run on each set of faces, with a group of 20 Hong Kong people who did not participate in the experiment rating the South Asian faces and another group of 20 rating the White faces. They were asked to rate each face on its attractiveness, friendliness, and approachability from 1 to 10. There were no significant differences in attractiveness (South Asian$_{mean}$ = 5.24, White$_{mean}$ = 6.25), friendliness (South Asian$_{mean}$ = 6.0, White$_{mean}$ = 6.3), and approachability (South Asian$_{mean}$ = 5.9, White$_{mean}$ = 6.0) in linear mixed effect models (Cao, 2025; Sec. B).

C. Procedure

Participants were given a Qualtrics link (Qualtrics, 2024) and asked to complete all the tests in a quiet environment using headphones. The experiment consisted of an accent identification task, a transcription task, and a survey of language background and attitudes, which took about 20 min to complete.

In the transcription task, participants were instructed to transcribe all the sentences using Chinese characters. They

were instructed to use spelling for segments they could not identify. The combinations of sentences and faces remained consistent across all conditions, while the order of the 45 Cantonese sentences was randomized. Prior to the transcription, participants completed a trial transcription to familiarize themselves with the task. Once the experiment started, participants were presented with a face positioned at the centre of the screen, accompanied by an audio file located below the image. By clicking the play button, participants could listen to the audio and enter their transcription in the provided blank space. There was no time limit imposed for the transcription task, and each audio is about 2–3 s long, played only once. There was a two-minute break in the middle of the transcription task.

Following the transcription task, an accent identification task was introduced. Participants were required to identify the specific Cantonese accent they had just heard from the following seven options: a Western accent (such as American or British), a South Asian accent (such as India or Nepal), a Southeast Asian accent[1] (such as Pilipino or Indonesia), a Japanese/South Korean accent, a Mainland Chinese accent, a local Cantonese accent, and an "others" category. Participants had to select the most appropriate accent category that they believed corresponded to the accent they had heard during the transcription task. Participants were also asked to rate the degree of accentedness of the Cantonese accent they had heard from 1 (no accent at all) to 6 (a very strong accent).

Regarding the attitude survey, participants judged the accent they just heard on 19 attitudinal traits (see Table II) from 1 (strongly disagree) to 4 (strongly agree). They also

TABLE II. Factor loadings of the rotated component matrix of 19 differential semantic traits. Note: Traits of the same factor are emboldened.

Traits	Factor 1 (Status)	Factor 2 (Friendliness)	Factor 3 (Competence)	Factor 4 (Energy)
Humble	**0.850**	−0.041	−0.035	−0.068
Polite	**0.843**	−0.144	0.112	0.018
Logical	**0.779**	0.071	−0.041	−0.152
Reliable	**0.751**	0.139	−0.074	−0.113
Well-educated	**0.685**	0.185	−0.079	0.052
Confident	**0.615**	−0.028	0.059	0.283
Cool	**0.569**	−0.117	0.006	0.476
Friendly	0.233	**0.790**	−0.088	0.064
Popular	−0.012	**0.762**	0.184	0.063
Considerate	0.095	**0.755**	0.097	−0.189
Willing to help	−0.012	**0.547**	0.390	−0.048
Humorous	−0.073	**0.518**	0.135	0.171
Wealthy	−0.065	−0.027	**0.821**	0.050
Competent	0.193	0.279	**0.686**	−0.052
Intelligible	−0.032	0.524	**0.561**	−0.098
Strong leadership	−0.190	0.371	**0.511**	0.148
Open	0.037	0.106	−0.071	**0.754**
Active	−0.206	−0.062	0.103	**0.716**
Energetic	0.483	0.116	0.003	**0.562**
Eigenvalue (rotation)	*4.198*	*2.958*	*1.989*	*1.860*
Total %variance explained		*58%*		

completed a language background questionnaire about their personal information such as age, sex, and whether they had contact with Mainland Chinese, Whites, or South Asians in their daily life.

D. Data preparation

In the transcription task, each sentence was meticulously assessed character by character by two research assistants. A point would be awarded only if the correct character was used. As Cantonese is mainly spoken with variations in the writing system, in cases where there was no consensus on the correct character, for example, between 無 and 冇 (without), or 老竇 and 老豆 (father), both answers were considered correct. There were 45 sentences in total, and each sentence contained between seven and eight characters. The transcription accuracy was represented by two columns: one column denoted the number of correctly transcribed characters, while the other indicated the number of incorrect characters. These two columns were used as the grouped binomial dependent variable for logistic mixed effect regressions in subsequent analyses.

Regarding the attitude survey, the reverse traits were recoded first and then the ratings of the 19 traits were imported to R for principal component analysis (PCA) using the psych package (Revelle, 2024). The suitability of the data for factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) test. The KMO value was found to be 0.8, suggesting that the data is good for factor analysis. A varimax rotation including orthogonal factors was used. Four factors with an eigenvalue larger than 1 were gained (i.e., *Status*, *Friendliness*, *Competence*, and *Energy*, see Table II) and a scree plot was plotted (Cao, 2025, Sec. C) using the factoextra package (Kassambara and Mundt, 2020). In total, they accounted for a combined total of 58% of variances observed in the data. Factor scores of these four factors were used for subsequent analyses.

III. RESULTS

A. Accent identification and accentedness judgement

The accuracy of accent identification was calculated by dividing the number of participants who correctly identified the accent by the total number of participants ($N = 20$). Table III shows the accuracy for each condition. As listeners selected an answer from seven options, the percentage of selecting it by chance was 14.3%.

For the local Cantonese condition, the exposure to White faces did not change participants' judgement on accent identification but seeing South Asian faces reduced

the accuracy. Compared to other accent conditions, the effects of the face on the local Cantonese accent were relatively weaker. This might be due to the high familiarity of local Cantonese accents by the participants. As all the participants were native speakers of Cantonese who were born and grew up in Hong Kong, the acoustic cues of local Cantonese accents were robust enough for them to complete the identification task.

For the Mandarin accent condition, exposure to a South Asian face or White face reduced the accuracy compared to the condition of a silhouette. Face successfully deceives the participants into identifying a wrong accent. A similar effect of the face was found in the English accent condition. When participants saw a White face that matched the accent, the accuracy increased; when they saw a South Asian face that did not match the accent, the accuracy decreased.

It is worth pointing out that South Asian faces tend to have a stronger impact on accent identification than White faces when the accents and the faces are mismatched. For instance, in the condition of local Cantonese, the accuracy of accent identification increased from 85% (i.e., the control condition-silhouette) to 90% when they saw a White face. However, the accuracy decreased to 65% when South Asian faces were presented. Similarly, when participants heard Mandarin-accented Cantonese, the mismatch between the faces and the accent led to a decrease in accuracy, but a larger decrease was observed in the South Asian faces (from 90% to 15%) than in the White faces (from 90% to 55%).

Table IV shows the results of accentedness judgement. A linear regression was run using the accentedness judgement as the dependent variable, and face, accent, and the interaction between face and accent were used as the fixed effects. The conditional R^2 is 0.71. No significant effect of face was found. Regarding the effect of accent, a significant difference in ratings on accentedness was found between local Cantonese ($\text{mean}_{\text{Cantonese}} = 4.8$) and the other two accents [$\text{mean}_{\text{Mandarin}} = 2.2$, standard deviation Mandarin ($\text{SD}_{\text{Mandarin}}) = 0.6$, $p < 0.001$; $\text{mean}_{\text{English}} = 2.2$, $\text{SD}_{\text{Mandarin}} = 0.6$, $p < 0.001$]. No significant difference was found between the Mandarin and English accents ($p = 0.7$).

B. The impact of face on language attitudes

Ratings of 19 attitudinal traits were reduced to four factors in PCA. After PCA, each participant had four factor scores for each accent, representing their attitudes towards the accent in different dimensions. For each accent, a linear regression was run using factor scores as the dependent

TABLE III. Accuracy of accent identification.

	Local Cantonese	Mandarin accent	English accent
White face	90%	55%	80%
South Asian face	65%	15%	35%
Silhouette	85%	90%	55%

TABLE IV. The mean and SD (in bracket) of accentedness judgement. Note: A lower score indicates a stronger accent that was rated based on a 6-point scale.

	Local Cantonese	Mandarin accent	English accent
White face	4.7 (1.1)	2.4 (0.6)	2.5 (1)
South Asian face	4.7 (0.9)	2.2 (0.5)	2.1 (0.6)
Silhouette	5 (0.7)	2 (0.7)	2.2 (0.6)

J. Acoust. Soc. Am. **157** (3), March 2025

Cao *et al.*   1623

TABLE V. A summary of all the significant comparisons and their p values in the linear regression models for language attitudes. "ns" represents no significance.

		Local Cantonese	Mandarin accent	English accent
Factor 1 (Attractiveness)	Silhouette vs South Asian	0.013	<0.001	0.008
	Silhouette vs White	ns	<0.001	<0.001
	South Asian vs White	ns	0.014	<0.001
Factor 2 (Friendliness)	Silhouette vs South Asian	ns	<0.001	ns
	Silhouette vs White	ns	<0.001	00.025
	South Asian vs White	ns	<0.001	ns
Factor 3 (Competence)	Silhouette vs South Asian	0.053	<0.001	ns
	Silhouette vs White	ns	<0.001	<0.001
	South Asian vs White	0.02	ns	<0.001
Factor 4 (Energetics)	Silhouette vs South Asian	<0.001	<0.001	<0.001
	Silhouette vs White	<0.001	<0.001	<0.001
	South Asian vs White	<0.001	ns	ns

variable and face (silhouette, South Asian, White) as the independent variable. Results of linear regressions suggest several significant comparisons, as shown in Table V and Fig. 1.

When comparing the ratings of South Asian and White faces across the three accents, the results reveal notable differences in perceptions among Hong Kong Chinese participants. When listening to the local Cantonese accent,

participants found that the presence of either a South Asian face or a White face did not significantly affect their ratings of the speaker's status and friendliness. However, a South Asian face was associated with perceptions of greater competence and energy compared to a White face.

In the Mandarin accent condition, participants rated speakers with a South Asian face as less friendly but more prestigious than those with a White face. Regarding the
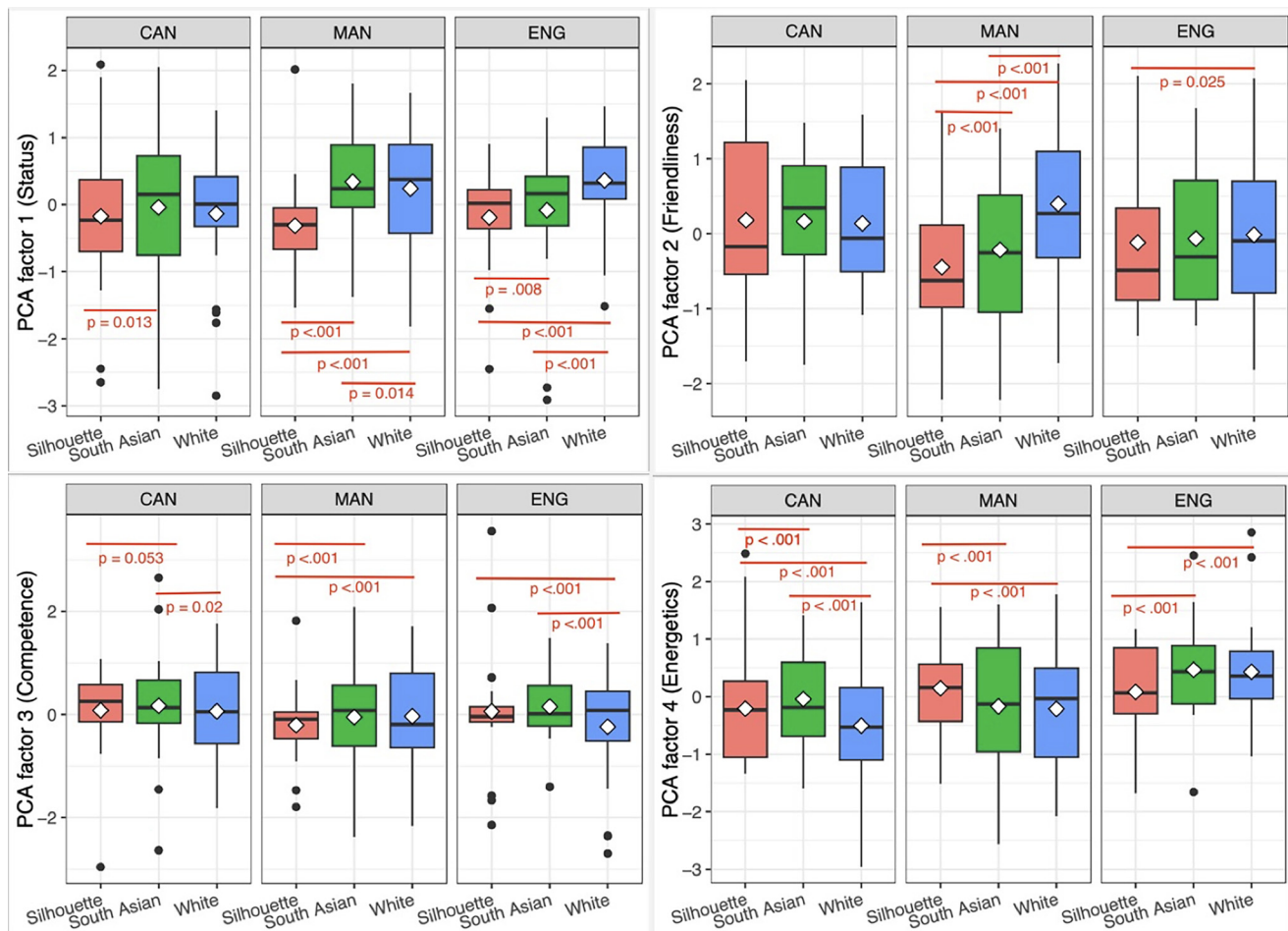

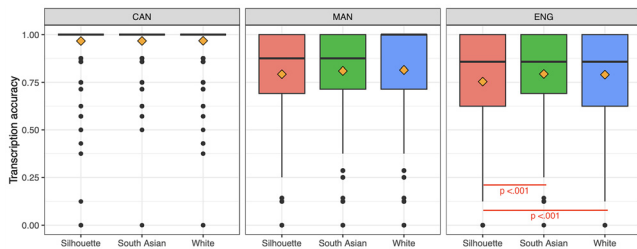FIG. 1. Factor scores of four factors by accents and face conditions.

FIG. 2. Transcription accuracy in boxplots with means (the diamond in the bar).

English accent, an English speaker with a South Asian face was perceived as less prestigious yet more competent than one paired with a White face.

While we predicted that seeing South Asian faces would elicit more negative attitudes than seeing White faces in accent perception, some findings support this prediction. For instance, when English-accented Cantonese was paired with a South Asian face, the accent would be rated as less prestigious than when the English accent was paired with a White face. Similarly, when Mandarin-accented Cantonese was paired with South Asian faces, the accent would be rated less friendly than pairing with White faces.

C. Transcription task

1. The full dataset

The full dataset was modelled using *glmer()* function from *lme4* package in R (version 4.2.3). The two columns containing the numbers of correct characters and incorrect characters were used as a grouped binomial dependent variable. Fixed effects contained the face (silhouette, South Asian, White), accent (local Cantonese, Mandarin, English), and the interaction between face and accent. Random effects included intercepts for groups, face images and sentences, and random slopes for the effect of accent by sentences. Figure 2 plots the transcription accuracy of all sentences in three accent conditions.

Results in Table VI suggest a significant effect for **accent**. The transcription accuracy of English-accented Cantonese ($p < 0.001$) and Mandarin-accented Cantonese ($p < 0.001$) was significantly lower than the local Cantonese

TABLE VI. Model estimates for log-likelihood of correct character transcription for the full dataset containing three accents. [*] represents $p < 0.05$.

	Est.	SE	z-value	p
(intercept)	3.858	0.176	21.86	**<0.001**
Face (South Asian vs Silhouette)	0.005	0.096	0.049	0.96
Face (White vs Silhouette)	0.033	0.096	0.345	0.73
Accent (MAN vs CAN)	−2.125	0.199	−10.67	**<0.001**[*]
Accent (ENG vs CAN)	−2.391	0.202	−11.86	**<0.001**[*]
Face (South Asian): Accent (MAN)	0.127	0.107	1.186	0.24
Face (White): Accent (MAN)	0.136	0.107	1.264	0.21
Face (South Asian): Accent (ENG)	0.270	0.105	2.558	**0.0105**[*]
Face (White): Accent (ENG)	0.215	0.106	2.033	**0.0421**[*]
Marginal/Conditional R^2			0.36/0.85	

accent. When the reference of accent changed from Cantonese to English in the model, there was no significant difference in the transcription accuracy between the English and Mandarin accents [Est = 0.27; standard error (SE) = 0.21, z-value = 1.276, $p = 0.2$], suggesting that the transcription difficulty was similar between the two accents.

2. The face effect

To answer Research Question 2, the full dataset was split into three subsets by accents. A different generalised logistic mixed effect model was run separately on each subset of accent. The model used the grouped binomial variable as the dependent variable, the face (reference = silhouette) as the fixed effect, and groups, face images, and sentences as random intercepts. Results of the three models suggest that seeing different faces significantly changed participants' intelligibility of English-accented Cantonese (Table VII) but not Mandarin-accented Cantonese and local Cantonese (Cao, 2025; Sec. D). An item analysis of face images was conducted separately for White and South Asian faces. Although the *post hoc* analysis suggested no difference in the attractiveness, friendliness and approachability ratings among these faces, the item analyses revealed that one South Asian face and one White face consistently elicited lower transcription accuracy in three accents (Cao, 2025; Sec. E). Given that both South Asian faces and White faces had one face eliciting lower transcription accuracy, the difference in images should not affect the face effect at the group level.

In the local Cantonese condition, the bias model predicted that intelligibility would be reduced in the South Asian face condition compared to the silhouette and the White face condition due to the bias against South Asians in the Hong Kong community. The non-significant effect of face on the local Cantonese accent did not support this hypothesis. However, as results in Sec. III B showed that participants in the present study did not fully associate South Asian faces with a lower social status, the results should be interpreted with caution.

In the English accent condition, the expectation model predicted that the intelligibility would be higher when face and accent were matched (i.e., ENG + White face) than when the face and accent were mismatched (i.e., ENG + South Asian face). Results of the English accent model suggest a non-significant difference between the South

TABLE VII. Model estimates for log likelihood of correct character transcription for the English accent subset. [*] represents $p < 0.05$. The reference level of the face changed to South Asian for the comparison between White and South Asian.

Accent = ENG	Est.	SE	z-value	p
(intercept)	1.47	0.19	7.85	**<0.001**[*]
Face (South Asian vs Silhouette)	0.27	0.04	6.17	**<0.001**[*]
Face (White vs Silhouette)	0.25	0.04	5.61	**<0.001**[*]
Face (White vs South Asian)	−0.03	0.05	−0.57	0.569
Marginal/Conditional R^2			0.008/0.78	

J. Acoust. Soc. Am. **157** (3), March 2025

Cao *et al.*   1625

Asian face and White face condition. In other words, we did not find supporting evidence for the expectation mechanism.

Under the adaptation mechanism, if there is a general adaptation, we predicted that the intelligibility of English and Mandarin accents would be improved in the South Asian face and White face conditions because listeners might perceive both types of faces as a more general category such as "non-native speakers/foreigners/outgroup members." Results partially confirm this hypothesis: seeing a South Asian face ($p < 0.001$) and a White face ($p < 0.001$) both significantly improved the transcription accuracy compared to seeing a silhouette in the English accent condition. However, the significant effect of face was not found in the Mandarin accent condition.

We also predicted that if there is a targeted adaptation, higher intelligibility should be observed when the White face was paired with the English accent compared to the case when the White face was paired with the Mandarin accent. This was because we assumed that a targeted adaptation would be activated when the White face paired with the English accent, facilitating the accent processing. Cross-accent comparison is possible because there was no significant difference in the overall transcription accuracy between the English and the Mandarin accents, as indicated in Sec. III C 1. A *post hoc* test of the model for the full dataset in Sec. III C 1 did not support this hypothesis: no significant difference was found in this comparison (White.MAN − White.ENG = 0.19, SE = 0.21, $z = 0.891$ $p = 0.373$).

To summarise, no supporting evidence was found for the bias mechanism, the expectation mechanism, and the targeted priming of the adaptation mechanism. Part of the results supported the general priming adaptation mechanism.

3. The effect of language attitudes

To answer Research Question 2, the influence of language attitudes on intelligibility was tested. In Sec. III B, four factors of language attitudes were gained from the PCA, representing four dimensions of attitudes: Status, Friendliness, Competence, and Energy. Using the model of each accent (in Sec. III C 2) as the base, each attitude factor was interacted with the face and tested using the function analysis of variance (ANOVA) for model comparisons. Table VIII summarises the model comparison results (see Cao, 2025; Sec. F for the full results). Adding Factor 1 (Status) and Factor 3 (Competence) significantly improved

TABLE VIII. Summary of the model comparison results. A significant *p*-value indicates a significant improvement in model fit by adding the attitude factor. "ns" represents no significance.

	CAN	MAN	ENG
Face × Factor 1 (Status)	**<0.001**	**<0.001**	**0.004**
Face × Factor 2 (Friendliness)	ns	**<0.001**	ns
Face × Factor 3 (Competence)	**0.016**	**<0.001**	**0.035**
Face × Factor 4 (Energetics)	ns	ns	ns

the model fit for CAN, MAN, and ENG accents. Adding Factor 2 (Friendliness) also improved the model fit for MAN. In other words, participants' attitudes towards faces affected their intelligibility of accented speech.

Figure 3 shows the correlations between attitudes towards face and intelligibility for the three accents. For the interaction of Factor 1 (Status) and South Asian faces, the more prestigious the South Asian faces were rated, the more intelligible the Mandarin accent was perceived. However, when participants heard English-accented Cantonese, a negative correlation was found between South Asian faces' perceived prestigiousness and intelligibility. Regarding the interaction of Factor 1 (status) and White faces, no strong correlation between attitudes and intelligibility was found when participants heard Mandarin-accented or English-accented Cantonese.

It is worth mentioning that the correlation patterns of South Asian faces were very similar to those of White faces in Factors 2, 3, and 4. Only in Factor 1 did the ethnicity of faces show significant differences, as shown in Table IX. This suggests that how people perceive the status of the two ethnicities can have different impacts on their intelligibility of Cantonese accents.

IV. DISCUSSION

A. The impact of face and attitudes on accent perception

The first research question explored the impact of different ethnic faces on accent identification, accentedness ratings, and attitudes. First, seeing different faces significantly affected the accuracy of accent identification, but South Asian faces tend to show a stronger influence than White faces. For instance, the accuracy of accent identification for Mandarin-accented Cantonese dropped by 75% when participants saw a South Asian face, but the accuracy only dropped by 35% when they saw a White face. The accentedness ratings also showed a similar trend. The Mandarin and English accents were rated as more accented when they were presented with a South Asian face compared to a White face. This trend is in line with the findings of Kutlu *et al.* (2022a). Kutlu *et al.* (2022a) found that American English was perceived as more accented when it was paired with a South Asian face compared to a White face. Kang and Rubin (2009) also found that people rated American English as more accented when they were primed to believe that the speaker was Chinese. The different degrees of influence of South Asian faces and White faces on accent identification and accentedness ratings are likely due to the ethnic bias that the Hong Kong Chinese participants have towards White people and South Asian minorities. For instance, Ku *et al.* (2006) conducted a survey with 200 South Asian minorities in Hong Kong and 61.5% of them disagreed or strongly disagreed that Hong Kong people treat people of different races/nationalities equally. Moreover, 67.5% of them reported being racially discriminated in Hong Kong. Among the four main ethnic minorities in Hong Kong, a
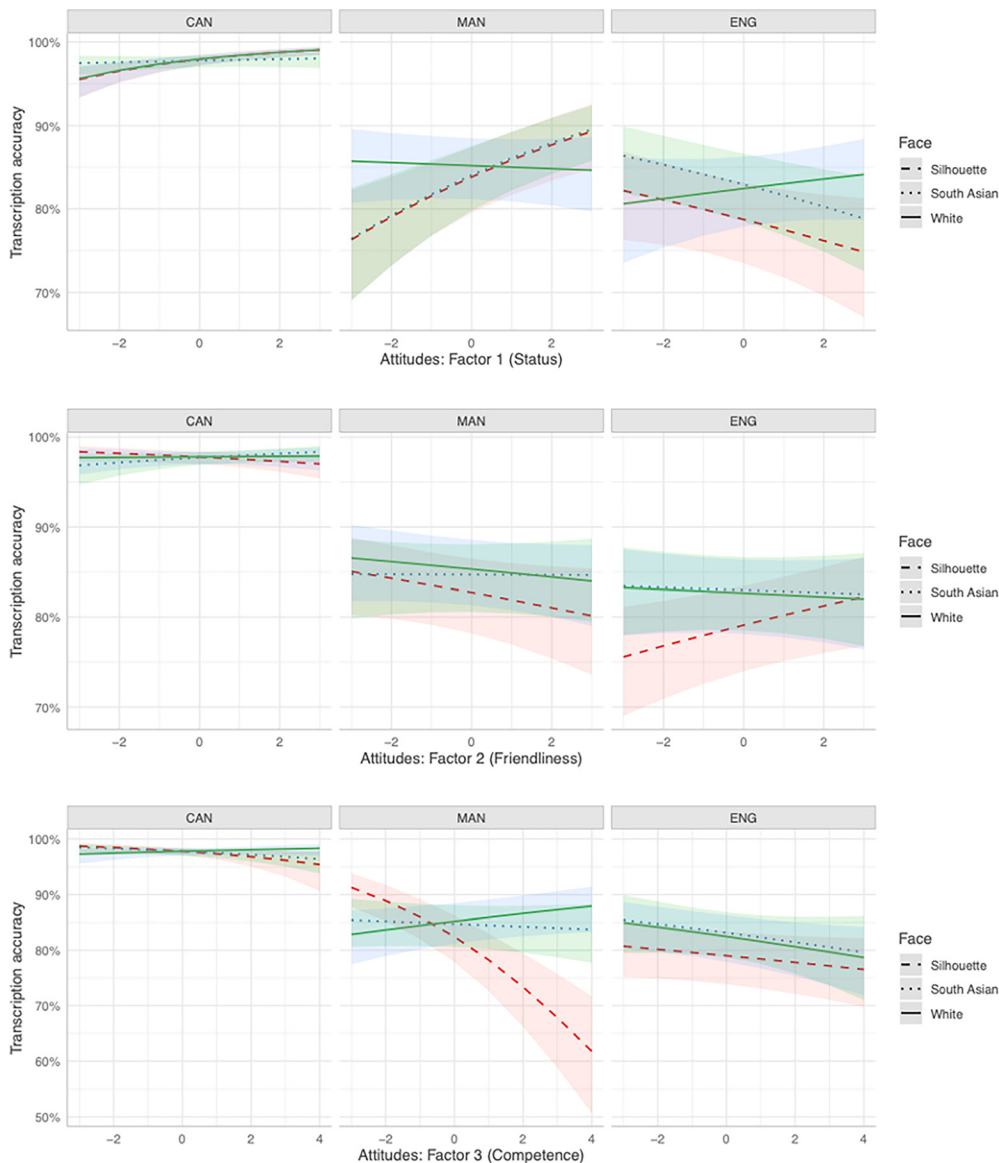
FIG. 3. The predicted transcription accuracy for the interaction of face and Factor 1 (the top panel), face and Factor 2 (the middle panel), and the interaction of face and Factor 3 (the bottom panel) for three accents.

TABLE IX. Summary of *p* values in pair-wise comparisons for the interactions between face and attitude factors. A significant *p*-value indicates a significant improvement in model fit by adding the interaction. "ns" represents no significance.

	Comparisons	CAN	MAN	ENG
Factor 1 (Status)	Silhouette vs South Asian	0.005	ns	ns
	Silhouette vs White	ns	<0.001	0.046
	White vs South Asian	0.006	<0.001	0.013
Factor 2 (Friendliness)	Silhouette vs South Asian	0.024	ns	ns
	Silhouette vs White	ns	ns	0.05
	White vs South Asian	ns	ns	ns
Factor 3 (Competence)	Silhouette vs South Asian	ns	<0.001	ns
	Silhouette vs White	0.02	<0.001	ns
	White vs South Asian	0.026	ns	ns
Factor 4 (Energetics)	Silhouette vs South Asian	ns	ns	ns
	Silhouette vs White	ns	ns	ns
	White vs South Asian	ns	ns	ns

higher percentage of Indian (74%) and Pakistani (84%) ethnic minorities reported experiencing racial discrimination compared to Filipinos (44%). It is possible that Hong Kong Chinese participants' negative stereotypes towards South Asian people influenced how they identified Cantonese accents and how they judged accentedness.

In terms of the impact of face on attitudes, a mixed result is found. Some findings support our hypothesis, that is, more negative attitudes would be elicited when participants see a South Asian face compared to a White face. For instance, we found that when English-accented Cantonese was paired with a South Asian face, it was rated as less prestigious than when the same accent was paired with a White face. Similarly, when Mandarin-accented Cantonese was associated with South Asian faces, it was perceived as less friendly compared to pairings with White faces. However, some results support a more positive attitude towards a

J. Acoust. Soc. Am. **157** (3), March 2025

Cao *et al.*   1627

South Asian face. The mixed results might be due to the images chosen for South Asian faces. To reduce the impact of socioeconomic differences, all the images chosen to represent South Asian minorities and White people were males wearing either a shirt or a suit. As reported by Ku *et al.* (2006), most of the South Asian minorities in Hong Kong were engaged in elementary occupations such as construction workers and service workers. The image of a South Asian male wearing a suit might not fully align with the image of Hong Kong South Asians for these Hong Kong Chinese participants. Some of their attitudes might be affected by the professional attire presented in the images.

Apart from this, this study also found a significant impact of attitudes on the intelligibility of accented Cantonese, particularly the status of faces people perceived. When the Hong Kong Chinese participants heard English-accented Cantonese, the more prestigious they rated the White faces, the more accurately they transcribed the English-accented Cantonese. In contrast, they had a lower transcription accuracy on the South Asian faces that were rated as more prestigious. The significant effect of attitudes in this study echoes the findings of Yi *et al.* (2013). Yi *et al.* (2013) found a significant effect of implicit attitudes toward White and Asian faces on perceiving Asian-accented English and American English. Babel and Russell (2015) tested the impact of implicit attitudes and explicit stereotypes towards Asian Canadians in perceiving Asian Canadian English and found no effect of both types of attitudes. Note that Babel and Russell (2015) focused more on people's ethnic bias, whereas the current project examined people's attitudes on the dimensions of status, friendliness, competence, and energetics. Different measurements of attitudes in these two studies might lead to various outcomes.

While studies like Yi *et al.* (2013) and Babel and Russell (2015) tested attitudes using an implicit association test (IAT), the current study tested attitudes using a different method of evaluating attitudes, contributing to a more comprehensive picture of how attitudes affect accent perception.

B. The dual-activation mechanism

The second research question explored the influence of face on perceiving the three accents of Cantonese and examined which of the three mechanisms (i.e., the bias mechanism, the expectation mechanism, and the adaptation mechanism) can best account for the findings. First of all, due to the mixed result on language attitudes, participants in the current study only associated the South Asian faces with negative attitudes in some aspects of attitudes. Therefore, it is difficult to interpret the non-significant effect of the face in the local Cantonese condition in relation to the bias mechanism. Second, in contrast to McGowan (2015) and Babel and Russell (2015), there was no advantage for the face-accent congruent condition (English accent + a White face) compared to the incongruent condition (English accent + a South Asian face) in the present study, which did not

support predictions of the expectation mechanism. It seems that the listeners in the current study were not bothered by the misalignment between the face and accent.

In comparison to the bias mechanism and the expectation mechanism, the adaptation mechanism offers a suitable framework for understanding the results of this study. When listeners saw a South Asian face or a White face, a general category of "non-native/foreign speakers/outgroup members" might have been activated to help them process the accents. This could account for the lack of significant difference between the incongruent and the congruent conditions in the English accent, as both a White face and a South Asian face were perceived as non-native/foreign/outgroup. This finding aligns with the general adaptation proposal put forth by Melguy and Johnson (2021) and McLaughlin and Van Engen (2023). For instance, McLaughlin and Van Engen (2023) proposed that their listeners might associate minority races with L2 accents, resulting in higher accuracy when a Mandarin accent was paired with a Middle Eastern face compared to when it was paired with a White face.

Another explanation is that Hong Kong Chinese listeners might apply a strict ingroup-outgroup distinction that Ray *et al.* (2010) reported when they perceive South Asian and White faces, viewing both ethnic groups as the same outgroup members. Ray *et al.* (2010) examined the impact of crossed-categorization on intergroup face recognition. They assigned four combinations of two social categories (i.e., political party and position on abortion) to 24 faces: Democrat and Pro-choice, Democrat and Pro-life, Republican and Pro-choice, Republican and Pro-life. When participants were Democrats and supported Pro-choice on abortion, they performed better in recognising the faces associated with Democrats and Pro-choice. For those who partially share their social membership (e.g., Democrats and Pro-life, or Republicans and Pro-choice), participants still perceived them as the same as those double outgroup members (i.e., Republicans and Pro-life). In this study, although Hong Kong Chinese listeners partially share their "Asian" membership with South Asians, they might apply a social exclusion pattern when they see non-Chinese faces.

Although the general adaptation mechanism can explain the findings in the English accent condition, a question still remains. If listeners were to perceive the South Asian face and White face as "non-native/foreign speakers/outgroup members," the enhancement effect found in the English accent should also be observed in the Mandarin accent. However, the face was not a significant predictor for the model of the Mandarin accent. Our interpretation is that when transcribing Mandarin-accented Cantonese, listeners might make good use of the phonological similarities between Cantonese and Mandarin to help them understand the Mandarin accent. As Cantonese and Mandarin are Chinese languages, they share many more phonological similarities than English and Cantonese do. This advantage might facilitate listeners' transcription. We assume that the general adaptation at the face level might be activated in the Mandarin condition too, as there was still a trend that

1628   J. Acoust. Soc. Am. **157** (3), March 2025

Cao *et al.*

listeners in the South Asian face (mean = 81%) and White face (mean = 81%) conditions gained higher accuracy of Mandarin-accented Cantonese than the listeners in the silhouette condition (mean = 79%). As the phonological similarities between Mandarin and Cantonese could help listeners transcribe the accented speech efficiently, faces might serve as secondary cues in perceiving the Mandarin accent.

What might happen in the English accent is that the face might play a larger role and contribute more to the transcription because phonological cues were relatively less useful and less familiar to the listeners. It seems that the utilisation of visual cues depends on their potential to enhance intelligibility. When relying solely on phonological cues is sufficient to handle the transcription, as observed in the local Cantonese condition of this study, visual cues may not play a prominent role in the transcription process. However, for transcribing a less familiar accent like the English accent, where phonological cues alone may be insufficient, visual cues can fill in the gaps and facilitate the task. In cases where phonological cues can account for most of the tasks, such as in the Mandarin accent, visual cues become secondary. Note that these observations are drawn from the findings of the transcription tasks in the current study. Given the inherent nature of transcription tasks,

where phonological cues inevitably play a significant role, it is unclear whether similar patterns would emerge when different tasks are used to assess the intelligibility of novel accents.

Based on the previously noted observations, we propose a dual-activation model to account for the dynamic activation of phonological and visual cues, as shown in Fig. 4. Two types of activation might have occurred in the transcription task: phonological activation by voice and general activation by face. In the silhouette condition of the Mandarin accent (the top panel of Fig. 4), solely a phonological category of the Mandarin accent would be activated by the voice. When listeners are exposed to a foreigner's face (either a South Asian face or a White face), a general category of "non-native/foreign speakers/outgroup members" would be activated in addition to the voice activation. Given the high phonological similarities between Mandarin and Cantonese, the voice activation would be stronger than the face activation.

In the English accent condition (the bottom panel of Fig. 4), two types of activation would occur when listeners see a foreigner's face (either a South Asian face or a White face) and hear English-accented Cantonese. The general category of foreigner accents including all the related exemplars of foreigner accents like the European accent, Indian accent,
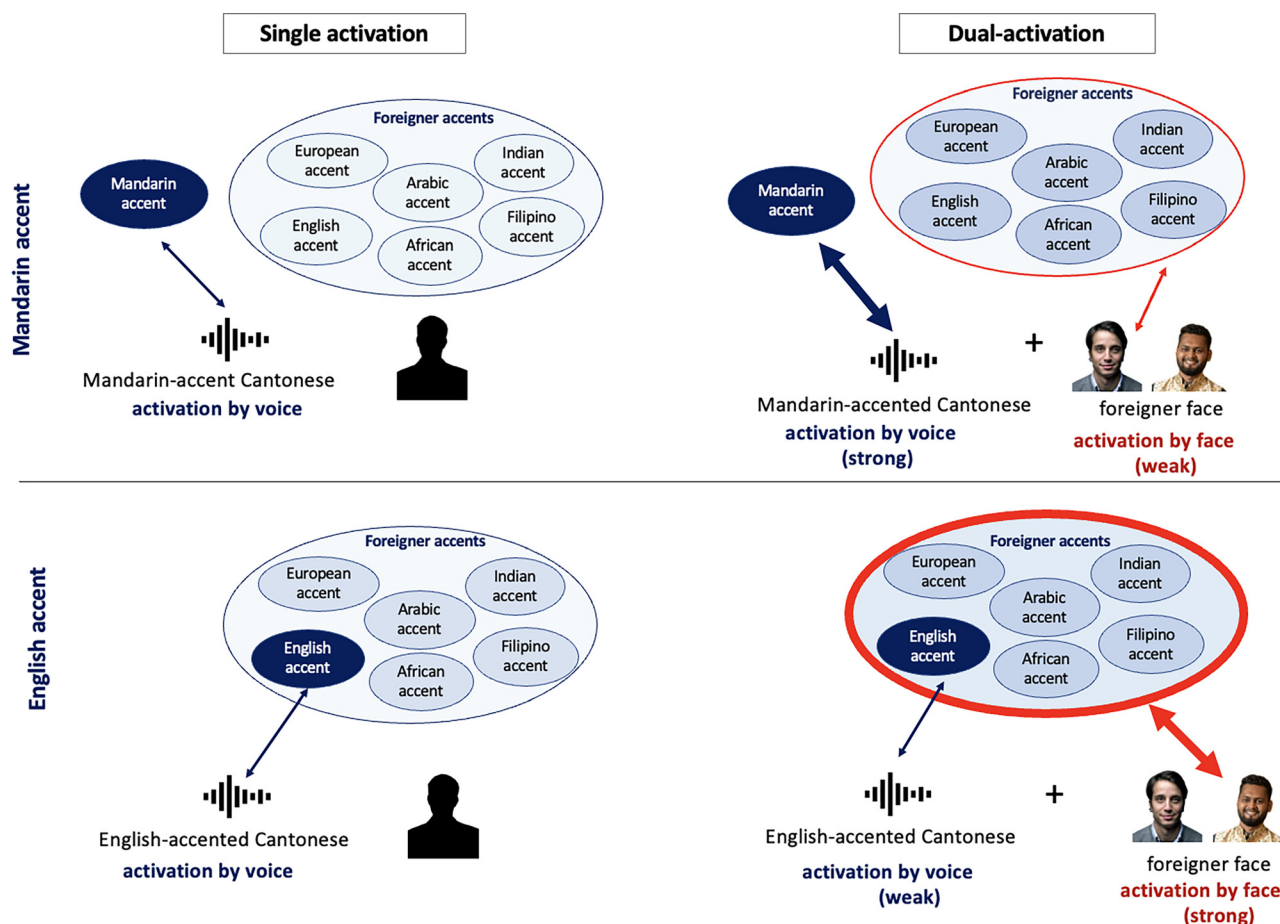

FIG. 4. Schematic of the single/dual-activation for transcribing Mandarin-accented (the top panel) and English-accented Cantonese (the bottom panel).[2]

Filipino accent, or Arabic accent would be activated and assist the transcription process. As English-accented Cantonese is less common in the Hong Kong community as compared to Mandarin-accented Cantonese, the voice activation might be weaker, and the face activation becomes stronger to make up for the gap. In other words, the voice and face activations function complementarily in accent perception.

There are several limitations in this study. First, because of the attempt to control for the social backgrounds of all images, mixed results were found in people's attitudes towards the South Asian faces. As a strong association between South Asian faces and the predicted negative attitudes was not found, interpreting the findings in relation to the bias mechanism becomes challenging. To address this issue, it would be beneficial to include images of the working class for both South Asian faces and White faces. Second, even though the 45 Cantonese sentences were placed in unpredictable contexts, the transcription accuracy in the local Cantonese condition was still quite high, resulting in a ceiling effect. To enhance the difficulty level of transcription in the local Cantonese condition, the inclusion of some background noise in the audio could be considered.

V. CONCLUSION

Using the examples of three Cantonese accents in Hong Kong, this study examined the impact of ethnic identity on accent perception and tested three mechanisms of accent perception. The results suggested that Hong Kong Chinese listeners made more errors in accent identification when the accents were mismatched with a South Asian face compared to the cases when the accents were mismatched with a White face. They also rated the Cantonese accents as more accented when the accents were paired with a South Asian face compared to a White face. Our findings supported the general adaptation mechanism (Melguy and Johnson, 2021), where listeners might perceive a South Asian face and a White face as the same general category as "non-native speakers/foreigners/outgroup members." Exemplars that related to this broad category would be activated to help process different accents. A dual-activation mechanism is proposed to illustrate how voice activation and face activation work complementarily to facilitate accent perception.

AUTHOR DECLARATIONS
Conflict of Interest

There are no conflicts to disclose.

Ethics Approval

This project contained ethics approval by the Survey and Behavioural Research Ethics at the Chinese University of Hong Kong (No. SBRE-20–636). Informed consent was obtained from all participants.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1] Many people in Hong Kong are exposed to the Southeast Asian-accented Cantonese spoken by the large number of domestic helpers in Hong Kong.

[2] The two foreigner faces are copyright-free images from unsplash.com for illustration. These are not the faces used in the experiment.

Babel, M., and Mellesmoen, G. (**2019**). "Perceptual adaptation to stereotyped accents in audio-visual speech," in *Proceedings of the 19th International Congress of Phonetic Sciences*, August 5–9, Melbourne, Australia, pp. 1004–1048.

Babel, M., and Russell, J. (**2015**). "Expectations and speech intelligibility," J. Acoust. Soc. Am. **137**(5), 2823–2833.

Bernstein, M. J., Young, S. G., and Hugenberg, K. (**2007**). "The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition," Psychol. Sci. **18**(8), 706–712.

Cao, W. (**2018**). "Short-term accommodation of Hong Kong English speakers towards native English accents and the effect of language attitudes," Ph.D. thesis, University of York, York, UK.

Cao, W. (**2024**). "One city, whose languages? Changing language attitudes in Hong Kong," in *The Politics of Language in Hong Kong*, edited by J. F. Dupré (Routledge, London), pp. 63–88.

Cao, W. (**2025**). "Supplementary material for the effect of ethnic bias and face on identification, accentedness judgments and ineligibility of Cantonese accents in Hong Kong," https://github.com/Gracecao-phonetics/JASA-visual-bias-on-Cantonese-accent-perception/blob/e07c0d0904b4361d79915a029d76dced6746f624/SuppPub1_02_2025.pdf (Last viewed February 15, 2025).

Carroll, J. (**2007**). *A Concise History of Hong Kong* (Hong Kong University Press, Hong Kong).

Chan, C. M., and Wong, H. Y. (**2005**). "Survey on public attitude and perception towards ethnic minorities 2004," APIAS Monograph Paper Series No. 9, http://commons.ln.edu.hk/apiasmp/6 (Last viewed December 2024).

Chan, J. Y. H. (**2016**). "A multi-perspective investigation of attitudes towards English accents in Hong Kong: Implications for pronunciation teaching," TESOL Quart. **50**(2), 285–313.

Drager, K. (**2011**). "Speaker age and vowel perception," Lang. Speech **54**(1), 99–121.

Erni, J. N., and Leung, L. Y. (**2014a**). "A cultural history of South Asians in Hong Kong," in *Understanding South Asian Minorities in Hong Kong* (Hong Kong University Press, Hong Kong), pp. 17–50.

Erni, J. N., and Leung, L. Y. (**2014b**). "A review of academic and policy research on ethnic minorities in Hong Kong," in *Understanding South Asian Minorities in Hong Kong*, edited by J. N. Erni and L. Y. Leung (Hong Kong University Press, Hong Kong), pp. 189–220.

Gu, M. (**2011**). "I am not qualified to be a Honkongese because of my accented Cantonese: Mainland Chinese Immigrant Students in Hong Kong," J. Multilingual Multicultural Develop. **32**(6), 515–529.

Hay, J., Warren, P., and Drager, K. (**2006**). "Factors influencing speech perception in the context of a merger-in-progress," J. Phon. **34**(4), 458–484.

HKSAR Census and Statistics Department. (**2021**). "Hong Kong 2021 population census: Population aged 5 and over by duration of residence in Hong Kong ethnicity, year and usual spoken language," (HKSAR Census and Statistics Department, Hong Kong).

HKSAR Census and Statistics Department (**2023**). "Hong Kong population statistics: Population growth by component," Hong Kong 2023 Population Census Report (HKSAR Census and Statistics Department, Hong Kong).

Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phon. **34**(4), 485–499.

Johnson, K., Strand, E. A., and D'Imperio, M. (**1999**). "Auditory-visual integration of talker gender in vowel perception," J. Phon. **27**(4), 359–384.

Kang, O., and Rubin, D. L. (**2009**). "Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation," J. Language Social Psychol. **28**(4), 441–456.

Kassambara, A., and Mundt, F. (**2020**). "Factoextra: Extract and visualize the results of multivariate data analyses, CRAN R package version 1.0.7," https://CRAN.R-project.org/package=factoextra (Last viewed September 5, 2024).

Ku, H., Chan, K.-W., and Sandhu, K. K. (**2006**). "A research report on the employment of South Asian ethnic minority groups in Hong Kong," in *Centre for Social Policy Studies* (The Hong Kong Polytechnic University, Hong Kong).

Kutlu, E., Tiv, M., Wulff, S., and Titone, D. (**2022a**). "Does race impact speech perception? An account of accented speech in two different multilingual locales," Cogn. Res. Princ. Implic. **7**(1), 7.

Kutlu, E., Tiv, M., Wulff, S., and Titone, D. (**2022b**). "The impact of race on speech perception and accentedness judgements in racially diverse and non-diverse groups," Appl. Linguist. **43**(5), 867–890.

Kwok, S., and Kirti, N. (**2003**). *Co-Prosperity in Cross-Culturalism: Indians in Hong Kong* (Commercial Press, Hong Kong).

Lakshmi, A., Wittenbrink, B., Correll, J., and Ma, D. S. (**2021**). "The India face set: International and cultural boundaries impact face impressions and perceptions of category membership," Front. Psychol. **12**, 627678.

McGowan, K. B. (**2015**). "Social expectation improves speech perception in noise," Lang. Speech **58**(4), 502–521.

McLaughlin, D. J., and Van Engen, K. J. (**2023**). "Social priming: Exploring the effects of speaker race and ethnicity on perception of second language accents," Lang. Speech **67**(3), 821–845.

Meissner, C. A., and Brigham, J. C. (**2001**). "Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review," Psychol. Public Policy Law **7**(1), 3–35.

Melguy, Y. V., and Johnson, K. (**2021**). "General adaptation to accented English: Speech intelligibility unaffected by perceived source of non-native accent," J. Acoust. Soc. Am. **149**(4), 2602–2614.

Pierrehumbert, J. B. (**2001**). "Exemplar dynamics: Word frequency, lenition and contrast," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (John Benjamins Publishing, Amsterdam), pp. 137–157.

Pierrehumbert, J. B. (**2002**). "Word-specific phonetics," in *Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 101–139.

Qualtrics (**2024**). "*Qualtrics* (2024.1). [online survey platform]," https://www.qualtrics.com (Last viewed June 21, 2024).

Ray, D. G., Way, N., and Hamilton, D. L. (**2010**). "Crossed-categorization, evaluation, and face recognition," J. Exp. Social Psychol. **46**(2), 449–452.

Revelle, W. (**2024**). *Psych: Procedures for Psychological, Psychometric, and Personality Research*, CRAN R package, https://CRAN.R-project.org/package=psych (Last viewed January 17, 2024).

Rubin, D. L. (**1992**). "Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants," Res. High. Educ. **33**(4), 511–531.

Shum, P. L., Tse, C. S., Hamamura, T., and Wright, S. C. (**2023**). "The Effects of Large-Scale Social Movements on Language Attitudes: Cantonese and Mandarin in Hong Kong," J. Lang. Social Psychol. **42**(3), 249–274.

Yi, H. G., Phelps, J. E. B., Smiljanic, R., and Chandrasekaran, B. (**2013**). "Reduced efficiency of audiovisual integration for nonnative speech," J. Acoust. Soc. Am. **134**(5), EL387–EL393.

Yip, V., and Catedral, L. (**2021**). "Perceiving (non)standardness and the Indexicality of New Immigrant Cantonese in Hong Kong," Lang. Commun. **79**, 81–94.

Zahn, C. J., and Hopper, R. (**1985**). "Measuring language attitudes: The speech evaluation instrument," J. Lang. Social Psychol. **4**(2), 113–123.

Zhang, Q. (**2009**). "Hong Kong people's attitudes towards varieties of English," Newcastle Work. Papers Linguist. **15**, 151–173.

J. Acoust. Soc. Am. **157** (3), March 2025

Cao *et al.*   1631