AFFECTIVE ATTITUDES TOWARDS ASIANS INFLUENCE PERCEPTION OF ASIAN-ACCENTED VOWELS

Nhung Nguyen[a], Jason A. Shaw[a,b], Michael D. Tyler[a,c], Rebecca T. Pinkus[a,c], Catherine T. Best[a,b]

[a] The MARCS Institute, University of Western Sydney, Sydney, Australia;
[b] School of Humanities and Communication Arts, University of Western Sydney, Sydney, Australia;
[c] School of Social Sciences and Psychology, University of Western Sydney, Sydney, Australia
nhung.nguyen@uws.edu.au, j.shaw@uws.edu.au, m.tyler@uws.edu.au, r.pinkus@uws.edu.au, c.best@uws.edu.au

ABSTRACT

Previous studies have shown that speech perception can shift because of cognitive attitudes about the speaker, but little is known about affective attitudes in speech perception. In our study, we investigated how affective attitudes towards Asians relate to the perception of Asian-accented vowels by native Australian English listeners. Affective attitudes were assessed with an established scale from the social psychology literature adapted to our specific purpose. Vowel perception was assessed using a vowel categorization task. Results show that the degree of dislike towards Asians negatively correlates with listeners' accuracy in vowel categorization. The results also provide evidence that cognitive attitudes elicit affective attitudes, and suggest the appropriateness of using social psychological tools to explicitly evaluate the role of affective attitudes in speech perception.

Keywords: sociophonetics, speech perception, social cognition, affective attitudes, modelling

1. INTRODUCTION

According to Allport, 'most of the business of life can go on with less effort if we stick together with our own kind. Foreigners are a strain...' [1]. Attitudes towards outgroups are known to influence a wide range of behaviour. In discussing how intergroup attitudes influence behaviour, social psychologists distinguish among three components: a cognitive component called stereotypes ('exaggerated [beliefs] associated with a category' [1]), an affective component called prejudice ('[a] feeling, favorable or unfavorable, toward a person or thing, prior to, or not based on, actual experience' [1]), and a behavioural component called discrimination ('[prejudice]-driven behavior' [13]). There is a growing body of evidence that the cognitive component (e.g., activation of beliefs about groups via priming) influences low level speech perception [7, 8, 10]. For example, although listening to the same Detroit speaker, Detroit listeners in [10] perceived 'standard' vowels when they were told to expect a Detroit speaker but perceived raised vowels (as the vowels actually were) when they were told to expect a Canadian speaker. Similarly, New Zealand listeners exposed to the concept of Australia by the word 'Australia' written on the test sheets [8], or stuffed kangaroos and koalas surreptitiously presented in the test rooms [7], reported more Australian-like vowels on a vowel matching task, even when they knew the voice was from a New Zealander. Therefore, extra-linguistic factors should also be considered for inclusion in speech perception models [7].

The studies above suggest that vowel perception can be affected by the cognitive component of intergroup attitudes. The other two components - the affective component and the behavioural component - have not been extensively investigated in speech perception. Cognitive attitudes have been shown to elicit affective attitudes which then shape other types of behavior such as recall tasks in [4]. After going through lists of traits describing English Canadians and French Canadians, English Canadians with negative affective attitudes towards French Canadians recalled more negative traits for the French Canadian outgroup and more positive traits for their own group. The categories 'English Canadian' and 'French Canadian' were activated in English Canadians, which in turn elicited their negative affective attitudes towards French Canadians to influence their performance on the recall task. Given this finding, we ask here whether cognitive attitudes might also elicit affective attitudes which then relate to performance in speech perception tasks.

There are few demonstrations of this association in speech perception. [10] is perhaps the closest although affective information was not directly measured. In this study, we measured affective attitudes towards the speaker group and assessed the relation between these affective attitudes and speech perception behaviour. Affective attitudes towards Asians

were measured by the Scale of Anti-Asian American Stereotypes [9] modified for the Australian context (SAAAS). These attitudes were then elicited by revealing the Vietnamese identity of the speaker. Listeners' perception of this Asian-accented English speech was tested via a standard vowel categorization task [14].

2. METHOD

2.1. Participants

Thirty-one participants from the Greater Sydney community took part in the experiment for either course credit or cash payment. Data from seven participants were excluded because they did not pass the training phase of the vowel categorization task. The analyses were conducted on the remaining 24 participants who were all native listeners of English and born in Australia.

2.2. SAAAS survey

SAAAS was constructed based on the Stereotype Content Model, which contains two dimensions - Competence and Sociability [5]. The combination of the stereotypes along these two dimensions invokes different ingroups' mixed feelings towards outgroups. For example: outgroups that are perceived as high on Competence but low on Sociability are respected but disliked, which is the case for Asian Americans [6, 9]. Participants were asked to rate 25 SAAAS items [9] (12 Competence items and 13 Sociability items) from 'strongly disagree' (coded as 0) to 'strongly agree' (coded as 4). The survey was hosted online by Qualtrics Survey Software on the platform of the University of Western Sydney.

Several filler items were included to mask the purposes of the survey. Firstly, to mask the target Asian group, 16 SAAAS items were used for five distractor groups - females, young people, Aborigines, Africans, and Arabs. Secondly, to mask the investigation on affective attitudes, eight other scales were included: handedness, religious ideologies, disgust, self-esteem, blatant and subtle prejudice, right-wing authoritarianism, implicit theories, and state self-esteem. The whole survey took approximately 30 min to complete.

For data analyses, the SAAAS items were recoded and summed, following [12], to yield a composite SAAAS score for each participant. Higher SAAAS scores indicated stronger dislike towards the Asian outgroup. The SAAAS scores were normally distributed and ranged from -24 to 24 (M = -0.38, SD = 11.97).

2.3. Vowel categorization task

2.3.1. Speakers

There were two accents in the categorization task: Australian-accented English for the training phase and Vietnamese-accented English for the test phase. The Australian-accented English speaker was a female in her 20s. She was born and raised in Western Sydney, and spoke Australian English only. The Vietnamese-accented English speaker was a female in her 30s. She learned English in Vietnam with Vietnamese teachers, had lived in Australia for 19 years at the time of the recording, and rated herself as having intermediate fluency in English.

2.3.2. Nonce word auditory stimuli

Auditory stimuli were recorded in a sound-attenuated booth using a Shure SM10A-CN headset microphone and a MOTU 896 mk3 sound card. The sampling rate was 44.1 kHz. Materials were 13 Australian English monophthongs (as classified in [3]) embedded in the /hVdə/ context: heeda /hiːdə/, hidda /hɪdə/, hedda /hedə/, hadda /hædə/, harda /hɐːdə/, hudda /hɐdə/, hodda /hɔdə/, horda /hoːdə/, hoodda /hʊdə/, who'da /hɯːdə/, hurda /hɜːdə/, heerda /hɪədə/, and hairda /heːdə/. In each accent, four tokens out of 10 repetitions per monophthong were selected based on subjective judgements of similarity in speaking rate and loudness. The only exception was Australian-accented 'hudda' where only two tokens were chosen (and therefore repeated twice in the training phase) since pilot results confirmed that the rest of Australian-accented 'hudda' tokens were confused with Australian-accented 'hadda' tokens.

2.3.3. Choice word visual display

Participants were presented with a grid of 13 choice words: 'bad,' 'bard,' 'bead,', 'beard,' 'bed,' 'bid,' 'bird,' 'book,' 'bored,' 'bud,' 'paired,' 'pod,' and 'rude.' The letter(s) corresponding to the monophthong in each word were highlighted to maximize clarity. The position of the words on the screen was randomized across participants.

2.4. Procedure

At the lab, participants were greeted by either of two associate researchers who were Australians and not Asians. They were asked to do the online survey, followed by the vowel categorization task, in which

they listened to the /hVdə/ nonce words and selected a choice word that contained the same vowel as the nonce word. After making a selection, their chosen word reappeared with the question 'How good is the match? (1 = very poor, 7 = excellent)' and the seven numbers from 1 to 7 beneath it. Participants then selected a number to indicate the goodness of fit.

There were 52 trials per phase. In the training phase, participants heard the Australian-accented English stimuli, and they were given feedback on incorrect responses. Participants continued cycling through 50 vowel tokens (four tokens per vowel, with the exception of 'hudda' as mentioned in 2.3.2) until they answered at least 10 out of 13 vowels correctly. Those who failed to achieve that criterion after four cycles of training were excluded from subsequent analysis. In the test phase, participants heard Vietnamese-accented English (four tokens per vowel) and no feedback was provided. They could take as long as they wished before responding, but the auditory stimuli were played only once for the choice word selection step. For the rating step, the stimuli were played only once also, but participants could click anywhere on the screen (except the rating numbers) to hear the stimuli again.

After finishing the tasks, participants were debriefed on the purposes of the speech task only. The entire lab session lasted from 50 min to 1.5 hrs, depending on how quickly participants completed the training phase.

3. RESULTS

In discussing the results, we refer to vowels according to the lexical sets of Wells [15]. Figure 1 shows mean accuracy by vowel for the Vietnamese-accent test phase. Performance was near ceiling for two vowels (START and KIT) and near floor for two others (NORTH and TRAP). The remaining nine vowels show more variation across participants. Of these nine, the GOOSE, STRUT, and NEAR vowels had a large number of inaccurate responses (37.5% of participants had at least two inaccurate responses for GOOSE tokens, 41.7% for STRUT, and 41.7% for NEAR tokens) in the Australian-accent training phase. Therefore, these were excluded from further analyses. Categorization accuracy for the remaining six vowels (FLEECE, DRESS, NURSE, FOOT, SQUARE, and LOT) was very high in the training phase (with no more than one inaccurate response for the majority of participants) but exhibited a large amount of cross-participant variation in the test phase, indicated by the error bars in Figure 1. We focus on these vowels to evaluate how affective attitudes relate to vowel perception accuracy.

Figure 1: Mean categorization accuracy by vowel. Error bars indicate 95% confidence intervals.

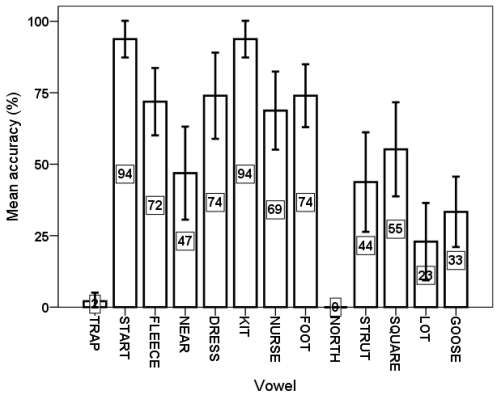
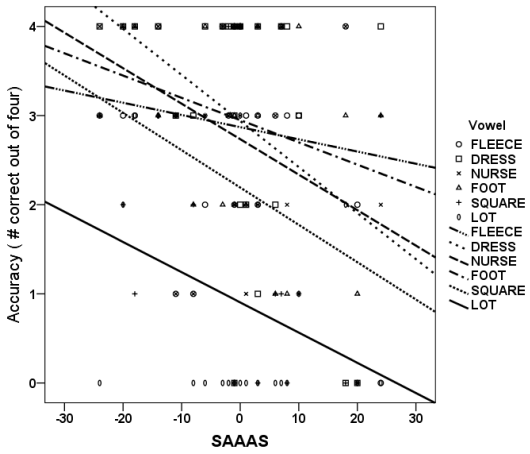


Figure 2 shows a negative correlation between SAAAS scores and speech perception accuracy - the higher the SAAAS score, the poorer the speech perception performance. Regression lines were fitted to each vowel separately. The downward slope of these lines indicates that people who score high on the SAAAS generally have lower vowel categorization accuracy than people who score low on the SAAAS. It can also be seen that this correlation is fairly uniform across the six vowels.

Figure 2: Categorization accuracy in the test phase as a function of SAAAS.



Even though we excluded participants who failed to reach the accuracy criterion in training and we eliminated vowels that showed a high number of errors in training (GOOSE, STRUT, and NEAR), there was still some variation across participants in vowel categorization accuracy in training. Some partici-

pants passed only 10 out of 13 vowels (n = 11) in training, while others passed 11 (n = 4), 12 (n = 7), or 13 out of 13 (n = 2). Because variation in the number of vowels passed in training (V_Passed) is likely to correlate with categorization accuracy in test, we also included this as a factor in the model.

We fitted a binomial mixed effects model to the accuracy data in R (version 3.1.2) [11] using lme4 package [2] with optimizer NM2 (the nloptr package version of Nelder-Mead):

(1) Accuracy $\sim$ $SAAAS + Vowel + V\_Passed + (1|Participant) + (1 + SAAAS|Token)$

SAAAS, Vowel, and V_Passed were the three fixed factors in the model. The random factors were Participant (random intercepts) and Token (both random slopes and intercepts). Table 1 lists the significant predictors in the best-fitting model. The β values indicate the magnitude and direction of the associations. The Z values and Pr(>|z|) indicate statistical significance. SAAAS has a statistically significant association with accuracy. The negative β value (-0.04) for SAAAS indicates that, for every unit increase in dislike towards Asians, the odds of accurately categorizing Vietnamese-accented English monophthongs decrease by a multiplicative factor of 0.96 ($e^{-0.04}$). The table also shows that accuracy on LOT was worse than the other five vowels. Finally, and as expected, the positive β coefficient for V_Passed indicates that participants who passed more vowels in training had higher accuracy in test. The interaction between SAAAS and individual vowels was also tested but did not lead to significant improvement over the model.

Table 1: Summary of significant effects in the best-fitting model.

predictor	β	S.E.	Z value	Pr(>|z|)
(intercept)	-6.06	2.12	-2.85	< 0.01
SAAAS	-0.04	0.02	-2.21	0.03
LOT	-2.68	0.49	-5.53	< 0.001
V_Passed	0.64	0.19	3.35	< 0.001

4. DISCUSSION

Cognitive attitudes are known to elicit affective attitudes which then shape intergroup behaviour. We have presented evidence that suggests that the same is true for speech perception. In our study, the category 'Asians' was presumably activated in listeners when they were told they would be hearing Vietnamese-accented speech. This priming aspect of the experimental set-up was quite similar to [10]. This set-up also elicited our listeners' affective attitudes towards Asians, which were directly measured by the SAAAS survey beforehand. The SAAAS scores allowed us to quantify affective attitudes towards Asians so that we could directly evaluate whether they correlated with vowel perception. As has been observed in other types of intergroup behaviour, attitudes towards Asians had a significant negative correlation with the accurate perception of Asian-accented speech. Although we are careful about interpreting our results in terms of direct causation, the findings are consistent with the conclusion drawn in other areas of social psychology [4], that activation of beliefs about a group also involves activation of favourability towards the group and that both of these components can influence speech perception.

To our knowledge, this is the first direct demonstration of the role of affective attitudes towards speaker groups in a speech perception task. We are optimistic about the prospects of using tools such as SAAAS to evaluate the role of affective attitudes in speech perception. To facilitate future research in this area, we would like to close the discussion with a methods point that may be helpful. According to the Stereotype Content Model, the combination of high Competence and low Sociability traits elicits dislike towards outgroup members [9]. However, these two stereotypic components do not have equal power in predicting the levels of liking and attention that outgroup members receive. An Asian who is perceived to be less sociable is more disliked and paid less attention to [9]. Accordingly, we explored the separate Competence and Sociability subscales of the SAAAS (instead of the composite score) and found that the Sociability score was a significant predictor of vowel accuracy whereas Competence was not.

5. CONCLUSIONS

We have established that affective attitudes, just like cognitive attitudes, influence speech perception. This result supports the conclusion in [7] that speech perception models should take both linguistic and extra-linguistic factors into account. Moreover, it goes further by demonstrating that, in addition to beliefs about a group, listeners' favourability towards a group can also be informative for speech perception. The result also suggests the appropriateness of using SAAAS and vowel categorization tasks to assess the role of affective attitudes in speech perception.

6. REFERENCES

[1] Allport, G. W. 1954. *The Nature of Prejudice*. Cambridge: MA: Addison-Wesley.

[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. _lme4: Linear mixed-effects models using Eigen and S4_. http://CRAN.R-project.org/package=lme4.

[3] Cox, F. 2006. The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Aust. J. of Ling.* 26, 147–179.

[4] Dutta, S., Kanungo, R. N., Freibergs, V. 1972. Retention of affective material: Effects of intensity of affect on retrieval. *Journal of Personality and Social Psychology* 23, 64–80.

[5] Fiske, S. T., Cuddy, A. J. C., Glick, P., Xu, J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82, 878–902.

[6] Glick, P., Fiske, S. T. 2001. Ambivalent sexism. In: Zanna, M. P., (ed), *Advances in Experimental Social Psychology Vol. 33*. Thousand Oaks, CA: Academic Press 115–188.

[7] Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Linguistics* 48, 865–892.

[8] Hay, J., Nolan, A., Drager, K. 2006. From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review* 23, 351–379.

[9] Lin, M. H., Kwan, V. S. Y., Cheung, A., Fiske, S. T. 2005. Stereotype content model explains prejudice for an envied outgroup: Scale of anti-Asian American stereotypes. *Personality and Social Psychology Bulletin* 31, 34–47.

[10] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18, 62–85.

[11] R Core Team. 2014. R: A language and environment for statistical computing. http://www.R-project.org/.

[12] Schweitzer, A., Lewandowski, N. 2013. Convergence of articulation rate in spontaneous speech. *Proc. 14th Interspeech* Lyon. 525–529.

[13] Smith, E. R. 1993. Social identity and social emotions: Toward new conceptualizations of prejudice. In: Mackie, D. M., Hamilton, D. L., (eds), *Affect, Cognition, and Stereotyping Interactive Processes in Group Perception*. Academic Press 297–315.

[14] Tyler, M. D., Best, C. T., Faber, A., Levitt, A. G. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71, 4–21.

[15] Wells, J. C. 1982. *Accents of English I: An Introduction*. Cambridge: Cambridge University Press.