Atten Percept Psychophys
DOI 10.3758/s13414-017-1329-2

Does seeing an Asian face make speech sound more accented?

Yi Zheng[1] · Arthur G. Samuel[1,2,3]

Abstract Prior studies have reported that seeing an Asian face makes American English sound more accented. The current study investigates whether this effect is perceptual, or if it instead occurs at a later decision stage. We first replicated the finding that showing static Asian and Caucasian faces can shift people's reports about the accentedness of speech accompanying the pictures. When we changed the static pictures to dubbed videos, reducing the demand characteristics, the shift in reported accentedness largely disappeared. By including unambiguous items along with the original ambiguous items, we introduced a contrast bias and actually reversed the shift, with the Asian-face videos yielding lower judgments of accentedness than the Caucasian-face videos. By changing to a mixed rather than blocked design, so that the ethnicity of the videos varied from trial to trial, we eliminated the difference in accentedness rating. Finally, we tested participants' perception of accented speech using the selective adaptation paradigm. After establishing that an auditory-only accented adaptor shifted the perception of how accented test words are, we found that no such adaptation effect occurred when the adapting sounds relied on visual information (Asian vs. Caucasian videos) to influence the accentedness of an ambiguous auditory adaptor. Collectively, the results demonstrate that visual information can affect the interpretation, but not the perception, of accented speech.

Keywords Asian face · Accent · Interpretation · Perception · Ethnicity

With increasing globalization, people's exposure to accented speech is growing, especially in a culturally diverse country like the USA. In fact, all speech has an accent, either a foreign accent (e.g., a Chinese accent) or a regional accent (e.g., a Boston accent). Many factors affect a listener's judgments of how accented speech sounds, including properties of sounds (e.g., Magen, 1998; Munro, Derwing, & Morton, 2006), lexical frequency (e.g., Levi, Winters, & Pisoni, 2007), visual cues (e.g., Irwin, 2008; Kawase, Hannah, & Wang, 2014; Swerts & Krahmer, 2004), and even cultural backgrounds (e.g., Wang, Martin, & Martin, 2002). The focus of the current study is a finding that simply seeing an Asian face can make speech sound more accented (Rubin, 1992; Rubin, Ainsworth, Cho, Turk, & Winn, 1999; Rubin & Smith, 1990; Yi, Phelps, Smiljanic, & Chandrasekaran, 2013; Yi, Smiljanic, & Chandrasekaran, 2014).

In Rubin's (1992) study, American undergraduates saw a picture of a face (either an Asian or a dark-haired Caucasian, matched in physical attractiveness) while hearing a passage that had been recorded by a native speaker of American English. After the passage, the participants were given a listening comprehension test, and were asked to give judgments of how accented the speech was, the potential teaching competence of the speaker, etc. Rubin found that when the photograph had been of an Asian face, students reported hearing an accent that did not exist. Moreover, participants' listening comprehension performance was poorer in the Asian face condition than in the Caucasian face condition. In a similar study, Rubin and Smith (1990) found that the ethnicity of a

Electronic supplementary material The online version of this article (doi:10.3758/s13414-017-1329-2) contains supplementary material, which is available to authorized users.

✉ Yi Zheng
yizheng.psychology@gmail.com

[1] Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, USA

[2] Basque Center on Cognition, Brain, and Language, Donostia, Spain

[3] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Published online: 17 May 2017

static face (Asian vs. Caucasian), rather than actual accentedness of speech, affected students' attitudes toward, and comprehension of, the speaker. The authors stated that "when students perceived—whether rightly or wrongly—high levels of foreign accentedness, they judged speakers to be poor teachers" (p. 337). Similar results were found when students watched a face and listened to Dutch accented English, with negative stereotypes again associated with the Asian face, suggesting that international instructors might get unfair evaluations due to their Asian appearance (Rubin et al., 1999). The phenomenon that certain beliefs about the speakers (e.g., non-native speakers) could affect how their speech is evaluated (e.g., accentedness, intelligibility), has been called "reverse linguistic stereotyping" (Kang & Rubin, 2009).

Additional evidence has been provided by Yi and his colleagues (Yi et al., 2013, 2014). Yi et al. (2013) presented native American English speakers with audio-only and audio-visual Korean-accented English and native English. Participants were instructed to transcribe and rate the accentedness of the speech. Results showed that Korean speakers were rated as more accented in the audiovisual condition than in the audio-only condition, while the pattern was reversed for English speakers. In addition, the visual cues helped intelligibility of the native English speech more than for the Korean-accented speech.

The idea that a person's appearance affects how his or her speech is perceived has been very influential – Rubin (1992)'s study alone has been cited over 390 times to date. In the current study, we re-examine the idea, assessing not only people's *interpretation* of accentedness but also their *perception* of the speech. That is, we draw a distinction between what people judge a sound to be in terms of accentedness on a decision level and what they really hear on a perceptual level. From our perspective, what has been called perception in some previous articles, such as accent ratings or filling out a survey on a speaker's accent (Levi et al., 2007; Magen, 1998; Rubin, 1992, 1998; Scales, Wennerstrom, Richard, & Wu, 2006; Yi et al., 2013) may actually be interpretation instead. The different notions of perception can be seen in Rubin's (1992) statement that "listeners 'perceptions of the instructors' accent – whether accurate perceptions or not – were the strongest predictors of teacher ratings." (p. 513). The first use of "perception" in this statement seems to be referring to an interpretation, whereas the second seems to reflect what people were actually hearing. Firestone and Scholl (2015) have emphasized the importance of disentangling "post-perceptual judgment from actual online perception" (p. 48), a point raised previously by Norris, McQueen, and Cutler (2000); see Samuel (1997; Samuel, 2001) for studies that have done this in the area of spoken word recognition.

The distinction between interpretation and perception has potentially important practical implications. If the reported effect of seeing an Asian face is generated at a level of interpretation, it seems feasible that this could be ameliorated by social interventions (Rubin, 1998). However, if the effect occurs on a perceptual level, this is a deeper-level issue and seems less amenable to potential interventions. More generally, as just noted, there is a growing recognition in the field that it is important to be precise when assessing phenomena, and the distinction between perception and interpretation is an important aspect of this theoretical precision.

Showing pictures of faces may not be the ideal way to measure how visual information affects participants' judgments of accented speech because pictures bring with them demand characteristics. Demand characteristics, widely studied in social psychology, are present when participants believe they know the purpose of an experiment, and alter their behavior based on these beliefs (e.g., Orne, 2009). In this case, when a picture is presented with no obvious connection to the speech being heard, participants are likely to make assumptions about what the experimenter might be looking for. Therefore, in addition to a replication of the basic effect using static faces, our experiments use dubbed video clips that pair facial information with the speech in a more natural way, reducing the demand characteristics.

The current study reports six experiments that investigate how visual information (e.g., an Asian or a Caucasian face) is integrated with auditory information (e.g., accented speech). In Part 1, we presented static pictures of a speaker (Asian vs. Caucasian) in Experiment 1, and used more integrated audiovisual stimuli (i.e., videos with lip-movements) in Experiment 2. In Part 2, we tested whether a decision-level interpretation of accentedness could be shifted by experimental manipulations, by introducing a contrast bias (Experiment 3), or by switching to a mixed (Experiment 4) rather than a blocked design. In Part 3 (Experiments 5A and 5B), we used the selective adaptation procedure (Eimas & Corbit, 1973) to determine whether visually different adaptors (i.e., an ambiguous sound dubbed onto Asian and Caucasian faces with lip-movements) would shift the audiovisual percept of the adaptors and thus produce different adaptation effects.

Part 1

Experiment 1

Rubin and his colleagues (Kang & Rubin, 2009; Rubin, 1992; Rubin et al., 1999; Rubin & Smith, 1990) have reported that judgments of how accented speech sounds were affected by seeing a picture of someone with an Asian face versus someone with a Caucasian face. In Experiment 1, we sought to replicate this effect by showing static pictures of faces and playing audios in the background. Rather than playing a single passage of speech recorded by a native American English speaker, the audios used in the current study were words that had been constructed by blending a recording of a native speaker

together with a recording of an Asian-accented speaker. Creating a continuum of stimuli that range from native to strongly accented provides a platform for sensitive tests using both an identification task (Experiments 1–4) and an adaptation task (Experiments 5A and 5B). These stimuli were built with an actual foreign accent, and can reveal how visual information affects speech of varying levels of accentedness. A huge existing literature on phonetic contrasts relies on using speech continua, with the identification and adaptation paradigms. The current study extends this approach to studying accent.

Method

Participants

Stony Brook undergraduate students with self-reported normal vision and hearing participated in this experiment. Participants were members of the Psychology Department subject pool, which is 62% female and 38% male. In addition, a sample of subjects from this population showed that the majority (94%) of native English speakers speak a second language, which is usually Spanish. For Experiment 1 (as well as Experiments 2–4), based on typical sample sizes for identification studies in the speech literature, we set an a priori goal of having usable data from 24 participants. To be included in the data analyses, participants had to be native English speakers, 18 years of age or older, with self-reported normal hearing. We excluded East Asian participants from the data analyses, as well as any participants who failed to follow instructions, performed very poorly (see below), or failed to complete the task. We excluded East Asian participants to avoid a potential effect of own-race preferences when presented with stimuli that contained an East Asian face (Bar-Haim, Ziv, Lamy, & Hodes, 2006; Kelly et al., 2007; Kelly et al., 2005; see Bernstein, Young, & Hugenberg, 2007, and Sangrigoli, Pallier, Argenti, Ventureyra, & De Schonen, 2005, for analyses of the own-race bias in terms of perceptual expertise and social-categorization models). In the current study, we identified participants' ethnicity by asking them about their origins if they appeared to be Asian. All participants received partial course credit to fulfill a research requirement in psychology courses.

Twenty-nine participants were tested in Experiment 1. We excluded three participants because they did not follow the instructions to look at the computer screen in front of them during the task (subjects were observed by the experimenter through a large window in the sound proof chamber); two participants were excluded due to poor performance (see details in the Results section).

Materials

The words we chose for our stimuli met several criteria. One essential criterion was that each word must include at least one sound that is characteristically difficult for Chinese native speakers to pronounce accurately. For example, Chinese-accented speakers often mispronounce /θ/ as /s/ (e.g., "thin" as "sin"), and /æ/ as /e/ (e.g. "bat" as "bet") (Rau, Chang, & Tarone, 2009; Rogers & Dalby, 2005; Zhang & Yin, 2009). We also wanted relatively high-frequency words, and non-monosyllabic words, so that they would be recognizable, even with an accented articulation. A final criterion was that stimuli could not be lexically ambiguous in an accented form. This eliminates words like *thinking*, as an accented rendition of this would sound like a different word, *sinking*. Based on these criteria, three English words were chosen: *cancer*, *theater*, and *thousand*; *cancer* contains /æ/, and *theater* and *thousand* both have /θ/. As described below, each of these three words was used to generate a large number of experimental stimuli, and each experimental stimulus was presented many times.

Auditory stimuli We selected a female native Mandarin speaker who had a strong Chinese accent and a female native speaker of American English to record the auditory stimuli. The American speaker was chosen because the fundamental frequency (pitch) of her voice was similar to the fundamental frequency of the Chinese speaker. Each speaker recorded stimuli in a sound-attenuated booth, using a high quality microphone and digital recorder. We instructed the speakers to pronounce each of the three English words several times, ranging from a slow speed to a fast speed. From these recordings, for each of the three words we selected tokens that matched in duration across the two speakers. We used Goldwave software to pre-process the stimuli. First, we used its noise-reduction feature to minimize any background noise (the software sample a silent period, and subtracts its spectrum from the speech). Second, we matched tokens on amplitude using Goldwave's half dynamic range option, which scales the signal so that the peak amplitude fills half of the available dynamic range. After this pre-processing, we used Praat software (Boersma & Weenink, 2016) to minimize any differences in the pitch of the selected native and non-native tokens. Finally, for each of the three words, we used the TANDEM-STRAIGHT software package (Kawahara & Morise, 2011) to make an eight-step continuum that had the native token at one end and the Chinese-accented token at the other end.

Our careful matching of the timing and fundamental frequency of the tokens from the two speakers accomplished two goals. First, matching these two properties allowed the morphing software to operate cleanly. Second, when we use the resulting stimuli in our perceptual tests, listeners cannot use cues like pitch height or word duration to make judgments about how accented a token sounds. The results of the construction process sounded natural; the tokens are provided as Supplementary Materials. Across the three sets of stimuli, tokens were about 600–800 ms long and had an average fundamental frequency around 200 Hz.

Videos We videotaped the faces of two female speakers (an Asian woman and a dark-haired Caucasian woman) in front of a blackboard looking directly at the camera. They were instructed to produce each of the three words at different speeds with neutral facial expressions. We selected videos of each word for which the lip-movements of the two speakers were generally matched with each other; this selection also ensured that the durations of the two tokens in a pair (one native, one accented) were matched. Using VSDC video editing software, we deleted the original audios of the videos and replaced them with tokens from the continua. Care was taken to keep the sounds and the lip-movements temporally consistent. This procedure generated 48 videos (two apparent speakers × three words × eight continuum steps). Videos were all 720 × 480, with 44,100 Hz frequency and 29.970 fps. Sample videos are provided as Supplementary Materials.

For each apparent speaker, we cut a short clip (around 0.1 s) from a video showing only her static face with the mouth closed (Appendix 1 provides the two static images). For each of the 48 videos we had made, we made a copy in which we replaced the original video component with the silent clip, stretched to make the length of the silent clip the same as the audio component. The resulting videos with static faces are conceptually comparable to the stimuli used by Rubin (1992): static pictures of either an Asian or a Caucasian face presented while speech is played.

For Experiment 1, we selected 24 of these videos as the stimuli – the two static faces paired with continuum steps 3, 4, 5, and 6 of three words (*cancer*, *theater*, and *thousand*). We chose these four steps because they are most ambiguous in terms of accent, and thus they are the most likely to be affected by the faces. Table 1 provides a summary of the experimental designs and stimuli in Experiments 1–4.

Procedure

Participants wore headphones and were tested in a sound-attenuated booth. We tested up to three subjects at the same time. Before the task began, participants were told that they would be watching a static face while listening to English words that were slightly different each time. Their task was to determine how native-like, or how accented, the words sounded. They were told that accent refers to any kind of accent that leads to speech different from standard American English. Participants responded by pushing one of four labeled buttons on a button board: 1 = native; 2 = somewhat native (the word sounded native but they were not quite sure); 3 = somewhat accented (the word sounded non-native but they were not sure); 4 = accented. This scale essentially requires subjects to make a forced choice (accented or not accented) together with a confidence choice (very confident, or not very confident). Participants were instructed to do this task as accurately as they could without taking too much time. There was a 1-s inter-trial-interval after all subjects had responded. If one or more participants failed to press a button within 3 s after the presentation of a stimulus, the next video was presented after a 1-s delay.

The accent-rating task was run in two separate blocks: participants watched the static Asian face in one block, and the static Caucasian face in the other block. In each block, there were 15 repetitions of 12 static Asian (or Caucasian) face videos (three words × four continuum steps) randomly presented. Each block took around 12 min, with the order of the two blocks counterbalanced across subjects. There was a 5-min filler task (playing silent computer games) between the two blocks.

Results

Two participants were excluded because they failed to respond at least ten times in at least one block (i.e., ≥5.6% missing responses). We obtained complete sets of usable data from 24 non-Asian native English speakers (evenly distributed across the two counterbalancing orders).

We calculated the average accentedness rating for each video and conducted a four-way repeated measures ANOVA on these scores with three within-subject factors: Face (Asian and Caucasian), Continuum Step (3, 4, 5, and 6), and Word (*cancer*, *theater*, and *thousand*), and one between-subject factor: Presentation order (Asian face tested first or second). Figure 1 shows the overall (left panel) mean accentedness ratings for the four continuum steps, for the first Block (middle panel), and for the second Block (right panel). Figure 2 presents the data collapsed across continuum step, broken down by each of the three Words (*cancer*, *theater*, and *thousand*).

Recall that Rubin (1992) found that subjects rated speech as being more accented when it was heard while seeing a picture of an Asian person than when the picture was of a Caucasian person. That study used a between-subject design – each subject either saw one picture or the other, and provided a single set of ratings. In the blocked design used here, the overall effect of Face was not significant, F (1, 132) = .16, p = .694, $\eta 2$ = .007, consistent with the near-identical curves for the Asian and Caucasian face conditions in the left panel of Fig. 1. However, as is clear in the other two panels, this null effect was not due to the pictures not affecting the accentedness ratings. Rather, there were two different patterns – one for the first time that people did the task (with one face), and one for the second time (with the other face). The first block is essentially a between-subject test like that used by Rubin, and as the middle panel of Fig. 1 shows, we observed the same effect that he did: Subjects who saw an Asian face rated the speech as more accented than subjects who saw a Caucasian face, F (1, 22) = 9.95, p = .005, $\eta 2$ = .31.

However, as the right panel of Fig. 1 shows, when subjects did the task a second time, now with the "other" face, the pattern reversed – now, rather than giving higher accentedness ratings to speech heard while seeing an Asian face, the ratings

Table 1 An overview of the stimuli and experimental design in Experiments 1–4

	Step 1 (accented)	Steps 3–6 (ambiguous)	Step 8 (native)
Experiment 1		Caucasian and Asian faces	
static photos blocked design		3 English words	
Experiment 2		Caucasian and Asian faces	
dubbed videos blocked design		3 English words	
Experiment 3	Asian face	Caucasian and Asian faces	Caucasian face
dubbed videos blocked design	3 English words	3 English words	3 English words
Experiment 4	Asian face	Caucasian and Asian faces	Caucasian face
dubbed videos mixed design	3 English words	3 English words	3 English words

are higher while seeing a Caucasian face, F (1, 22) = 9.05, p = .006, η2 = .29. If the visual context effect is being driven by perceptual mechanisms, it is hard to imagine how this reversal could occur. On the other hand, if the effect reflects decision mechanisms, then such a reversal is easier to understand. For example, subjects may have initially reported accentedness scores that were influenced by what they guessed the experiment was about (i.e., they may have responded to the demand characteristics of the pictures), but when they then get the "other" picture they may have overcompensated in trying to provide scores that were not biased (and, as the left panel shows, the overall accentedness between the two faces was the same).

Returning to the overall ANOVA, there were three significant effects. First, the main effect of Continuum Step was significant, F (3, 132) = 127.17, p < .001, η2 = .85, an effect that simply demonstrates that our construction of the accentedness continuum was successful. Second, there was a significant main effect for Word, F (2, 132) = 30.22, p < .001, η2 = .58. Pairwise comparisons (Bonferroni) of the accentedness ratings showed that *cancer* (M = 2.83, SD = .08) > *theater* (M = 2.26, SD = .08) = *thousand* (M = 1.92, SD = .10), with *cancer* rated significantly more accented than *thousand* and *theater*, *p's* < .001, but with no significant difference between *theater* and *thousand*, p = *.058*. As Fig. 2 shows, although there were some differences among the three words in terms of how accented each sounded, the general patterns described above were consistent across the three words. Finally, there was a significant main effect of Presentation order, F (1, 22) = 10.24, p = .004, η2 = .32. Participants who watched the Asian face first and the Caucasian face second had overall higher accent rating scores (M = 2.52, SD = .08) than the participants who watched the two faces in the reverse order (M = 2.15, SD = .08).

Discussion

The results of Experiment 1 show that during an initial block of trials, speech paired with an Asian face was rated as more accented than the same speech paired with a Caucasian face. This result is consistent with the result reported by Rubin (1992), whose between-subject design matches the between-subject design of this initial block of trials. The results are similar, even though Rubin presented a short passage from a native speaker paired with two faces, whereas we tested three English words made to be ambiguous (i.e., somewhere between native and strong accented). Critically, in our second block, when subjects saw the "other" face, the speech paired with the Caucasian face was judged as having a stronger accent than the speech paired with the Asian face. We suggest

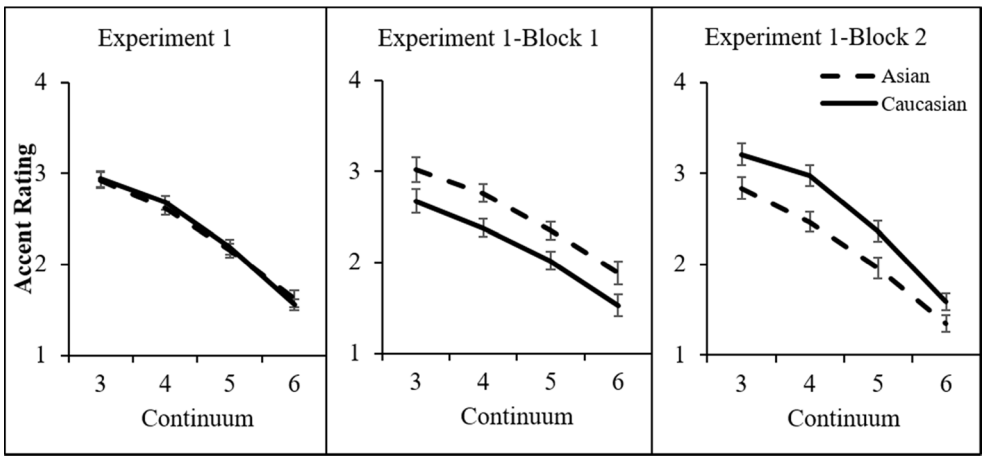


Fig. 1 Accentedness ratings of continuum steps 3–6 as a function of whether the static face was Asian versus Caucasian. Error bars represent the standard error of the mean

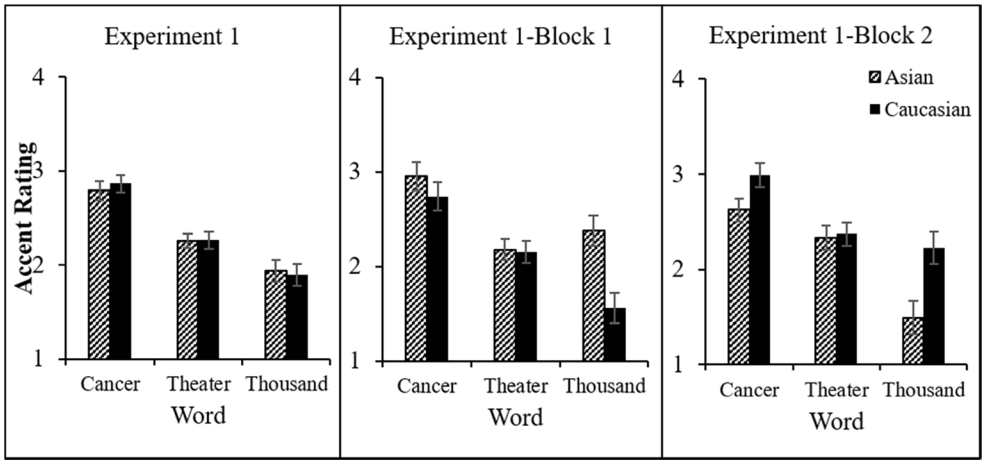


Fig. 2 Accentedness ratings of the three words separately as a function of whether the static face was Asian versus Caucasian. Error bars represent the standard error of the mean

that participants adjusted their accent rating judgments across the two blocks, producing the overall null effect of Face when the data are collapsed across the two blocks.

Our interpretation assumes that subjects were acting strategically, and participants' reports during the debriefing session support this idea. When we asked participants what they thought the experiment was about, 79% (19/24) of them correctly guessed the purpose of the study – they said that they thought we were testing their perception of the faces, and whether this affected their accent ratings. One of the 19 participants reported that she even realized that she shifted her decisions to be more accented when watching the Asian face. The remaining five participants either said that they did not know, or guessed something irrelevant (e.g., thinking that study was about the smoothness of the speech and gaps between vowels).

Experiment 2

Experiment 1 showed that static faces seem to lead participants to shift their judgments of accent, presumably because presenting static pictures during speech does not have any other obvious purpose. Videos (i.e., faces with lip-movements), in comparison, may not produce strong demand characteristics because the speech is actually integrated with the visual information. Thus, in Experiment 2, we used dubbed videos of faces, rather than static faces, to test whether judgments of accentedness differ between Asian face videos and Caucasian face videos.

Method

Participants

We tested a new set of 26 participants in Experiment 2. We excluded two participants due to a computer failure during the experiment. Participants all had self-reported normal hearing and vision. They received partial course credit for their participation.

Materials

The 24 audiovisual stimuli, eight for each of the three words, were described in Experiment 1. For each word, we dubbed steps 3, 4, 5, and 6 of the continuum onto both the Asian face and the Caucasian face videos. All videos were dubbed so that it looked as if the speakers were producing the words themselves.

Procedure

As in Experiment 1, participants wore headphones and sat in a sound-attenuated booth. On each trial, they watched a video and pressed one of four buttons, using the same rating scale as in the first experiment. Participants were instructed to do the task as accurately as possible without taking too long. Timing of the trials was as in Experiment 1.

The accent-rating task was run in two blocks. In each block, participants received 15 randomizations of 12 Asian or Caucasian face videos. Half of the participants watched the Asian face videos first, and half watched the Caucasian face videos first. As in Experiment 1, the two blocks were separated by a 5-min computer game playing filler task.

Results

For each subject, we calculated the average accentedness rating for each video. A four-way repeated measures ANOVA (Face × Continuum Step × Word × Presentation order) was conducted on these scores. For consistency with Experiment 1, a three-way repeated measures ANOVA (Face × Continuum Step × Word) was then conducted separately for the results of each Block, using Face as a between-subject variable. Figure 3

shows how the visual information (Asian vs. Caucasian face) influenced participants' judgments of the four continuum steps for the three words; Fig. 4 shows the results collapsed across the continuum steps, for each of the three words individually. Overall (left panel of Fig. 3), the main effect of Face was significant, $F(1, 132) = 4.32$, $p = .050$, $\eta 2 = .16$, reflecting a small but consistent tendency to report stimuli with the Asian face as more accented. The main effects of Continuum Step ($F(3, 132) = 119.34$, $p < .001$, $\eta 2 = .84$) and Word ($F(2, 132) = 19.32$, $p < .001$, $\eta 2 = .47$) were both significant, showing similar patterns as in Experiment 1. No other effects were significant.

A comparison of the middle and right panels of Fig. 3 to the corresponding panels of Fig. 1 makes it clear that switching to videos eliminated the reversal that occurred in Experiment 1 – judgments of accentedness with the video stimuli were much more stable. In Experiment 2, there were weak trends in both Block 1 and Block 2 towards higher accentedness ratings for the Asian face than for the Caucasian face, but in neither Block was this trend significant; the interactions of Face × Continuum Step and Face × Word were also not significant in either Block, p's > .05. As in the overall analysis, the main effect of Continuum Step and the main effect of Word were both significant in each Block individually, p's < .001.

Discussion

Experiment 2 matched Experiment 1 except for the presentation method of the faces: we changed from static pictures to videos, while playing the same sounds. Using the videos, which should reduce demand characteristics, we found a small but significant effect of Face. This result is consistent with Rubin's (1992) finding, but the effect is clearly rather weak. The absence of a reversal in the ratings from the first block to the second in Experiment 2 highlights how sensitive to response strategies the effect was when pictures were used. It is worth noting that Yi et al. (2013) also used integrated audiovisual stimuli and found a larger effect of Face. Critically, we dubbed the same ambiguous sound onto two faces whereas Yi et al. (2013) actually presented different speech with each face.

Part 2

The two experiments in Part 1 suggest that people's judgments of accentedness depend on the way that the visual stimuli (static vs. moving faces) are presented. In Part 2, we continue to use the dubbed videos, and test whether decision level interpretations of accentedness can be shifted by manipulating different aspects of the visual presentation.

Experiment 3

In Experiment 3, we added six more videos. In these additional videos, for each of the three words the native sound was paired with the Caucasian face, and the most accented sound was paired with the Asian face. The additional videos serve two purposes. First, they provide participants with an unambiguous standard to use while making judgments of the ambiguous videos. Second, they provide a test of whether the accentedness judgments are influenced by decision level factors. In particular, if the judgments are subject to decision biases, then the new unambiguously accented and unambiguously unaccented videos should produce standard contrast effects: Ambiguous words paired with Asian videos, presented in the context of strongly accented words paired with Asian videos, will be judged as less accented; ambiguous words paired with Caucasian videos, presented in the context of

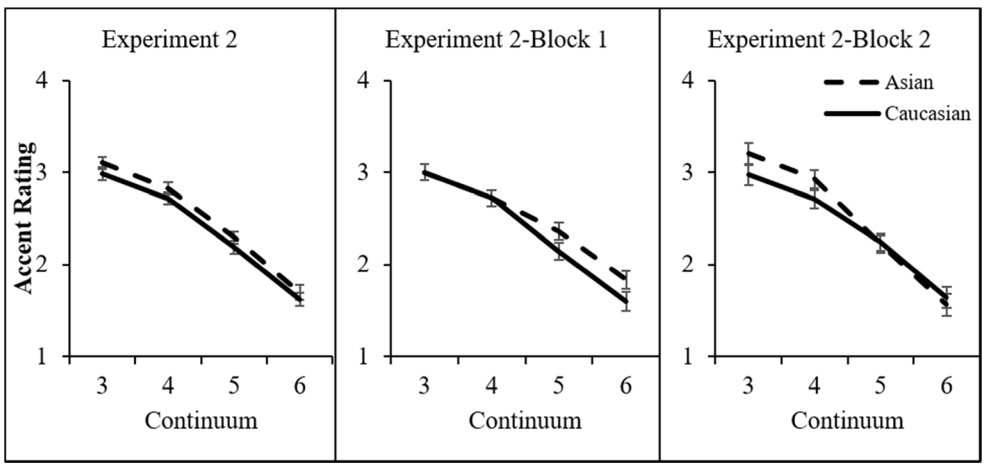


Fig. 3 Accentedness ratings of continuum steps 3–6 as a function of whether the face was Asian versus Caucasian. Error bars represent the standard error of the mean

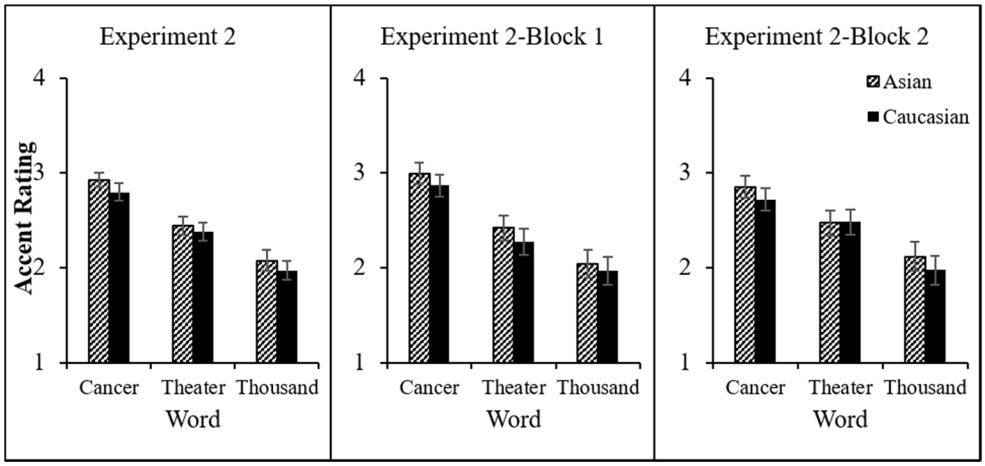


Fig. 4 Accentedness ratings of the three words separately as a function of whether the face was Asian versus Caucasian. Error bars represent the standard error of the mean

native speech paired with Caucasian videos, will be judged as more accented.

Method

Participants

Thirty students who had not been in Experiments 1 or 2 participated in Experiment 3. They all had self-reported normal hearing and vision. We excluded the data from five East Asian participants from the data analyses. Participants received partial course credit for their participation.

Materials

In addition to the 24 videos in Experiment 2, we constructed six more videos. For each word, we dubbed step 1 of the continuum (most accented) onto the Asian-face video, and we dubbed step 8 (most native) of the continuum onto the Caucasian-face video. These audiovisual tokens were intended to provide clear anchors for the participants, stimuli in which the accentedness of the audio track was consistent with the face being seen to produce it.

Procedure

The procedures were the same as in Experiments 1 and 2: The accent-rating task was run as two separate blocks, with all Asian videos in one block, and all Caucasian videos in the other block. In each block, there were 15 repetitions of 15 Asian-face (or Caucasian-face) videos randomly presented. Each block took around 15 min. The order of the two blocks was counterbalanced across subjects. The same 5-min filler task as before was used to separate the two blocks.

Results

We excluded one participant because he failed to respond to at least ten trials in at least one block. We then calculated the average rating for each video for each subject. Complete sets of usable data were obtained from 24 non-Asian native English speakers (12 in each of the two conditions).

A four-way repeated measures ANOVA was conducted: Face × Continuum Step × Word × Presentation order. The unambiguous endpoint tokens were not included in the analyses because they were only presented with one type of face (see Table 1); they were used as reference points – our focus is on the potentially movable tokens near the

Table 2 Means and standard deviations of accentedness as a function of face and word in Experiment 3

		Step 1	Step 3	Step 4	Step 5	Step 6	Step 8
Caucasian	*cancer*		3.80 (.23)	3.42 (.44)	2.83 (.64)	1.64 (.39)	1.14 (.22)
	theater		3.53 (.41)	3.23 (.45)	2.66 (.57)	1.76 (.59)	1.33 (.39)
	thousand		2.82 (.73)	2.67 (.69)	2.47 (.70)	1.97 (.61)	1.14 (.23)
Asian	*cancer*	3.79 (.55)	3.43 (.57)	2.98 (.58)	2.16 (.61)	1.41 (.39)	
	theater	3.66 (.28)	3.12 (.40)	2.73 (.47)	2.09 (.66)	1.49 (.47)	
	thousand	2.55 (.63)	2.42 (.66)	2.21 (.56)	2.12 (.61)	1.67 (.52)	

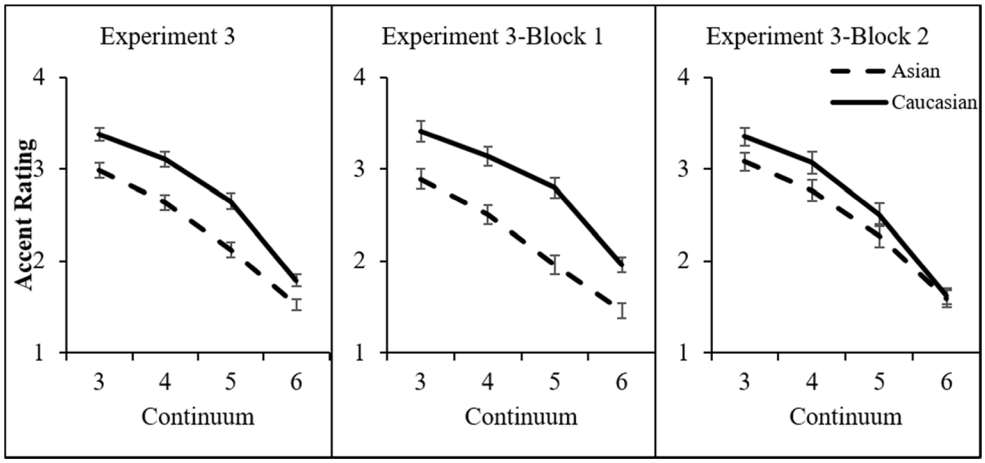


Fig. 5 Accentedness ratings of continuum steps 3–6 as a function of whether the face was Asian versus Caucasian. Error bars represent the standard error of the mean

middle of the continuum, as in Experiments 1 and 2. The means and standard deviations for all conditions, including the unambiguous tokens, are shown in Table 2. Figure 5 shows how the visual information (Asian vs. Caucasian) influenced participants' judgments of the four continuum steps for the three words; Fig. 6 shows the results collapsed across the continuum steps, for each of the three words individually.

As is clear by comparing the results in Figs. 5 and 6 to the corresponding figures from Experiment 2, adding the unambiguous endpoint stimuli drastically changed the pattern of accentedness ratings. In Experiment 3, these ratings were dominated by a contrast effect – Asian videos were rated as *less* accented ($M = 2.42$, $SD = .09$) than the Caucasian videos ($M = 2.63$, $SD = .09$), F (1, 132) = 73.71, $p < .001$, $\eta 2 = .77$. As in the previous experiments, the main effects of Continuum Step (F (3, 132) = 250.22, $p < .001$, $\eta 2 = .94$) and Word (F (2, 132) = 7.40, $p = .002$, $\eta 2 = .25$) were significant. In this case, the interaction between Continuum Step and Face was also significant, F (3, 132) = 5.60, $p = .002$, $\eta 2 = .20$, reflecting the somewhat smaller effect of Face for Step 6 than for the other Steps.

Inspection of the middle and right panels of Fig. 5 suggests that the contrast effect was stronger during the first block of the experiment than during the second block. Two three-way repeated measures ANOVAs (Face × Continuum Step × Word) were conducted to assess the effect of the videos for the first block and the second block separately, as in the previous experiments. The effect of Face was in fact significant for the first Block (F (1, 22) = 24.88, $p < .001$, $\eta 2 = .53$, Asian: $M = 2.21$, $SD = .09$; Caucasian: $M = 2.83$, $SD = .09$) but not for the second (F (1, 22) = 2.30, $p = .144$, $\eta 2 = .10$). For both blocks, the main effect of Continuum Step was significant, (Block 1, F (3, 132) = 204.89, $p < .001$, $\eta 2 = .90$; Block 2, F (3, 132) = 289.95, $p < .001$, $\eta 2 = .93$), as was the main effect of Word (Block 1, F (2, 132) = 3.068, $p = .033$, $\eta 2 = .14$; Block 2, F (2, 132) = 10.45, $p < .001$, $\eta 2 = .32$). For the first block, the

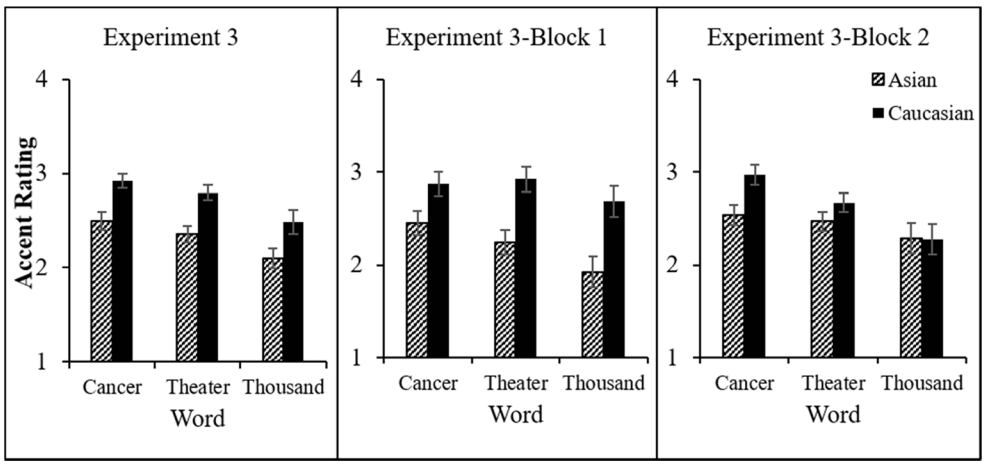


Fig. 6 Accentedness ratings of the three words separately as a function of whether the face was Asian versus Caucasian. Error bars represent the standard error of the mean

interaction of Face and Continuum Step was significant (F (3, 132) = 3.05, p = .034, η2 = .12), reflecting the slightly smaller effect on Step 6. No other effects reached significance.

Discussion

The results of Experiment 3 show that when unambiguous anchors are provided, speech heard as coming from an Asian face was rated as *less* accented than if the speech came from a Caucasian face. This pattern was due to the context effect provided by the unambiguous items. In the block with the unambiguously accented Asian videos, participants rated the ambiguous videos as less accented; in the block with unambiguously unaccented Caucasian videos, participants rated the ambiguous videos as more accented. This is a classic contrast effect, consistent with the accentedness judgments being heavily influenced by decision-level processes.

We suggested that the results in Experiment 2 differed from those in Experiment 1 because of a reduction in the demand characteristics when the speech was integrated with the visual display. That is one type of a decision-level effect. Experiment 3 has provided evidence for a second type of decision-level bias: contrast effects.

Experiment 4

In Experiment 4 we shift to a design that should minimize decision level effects by presenting the Asian and Caucasian videos in a mixed design. In general, blocking stimuli affords subjects the greatest opportunity to use strategic (decision-level) processes in their responses. By having videos with the two faces randomly presented, such strategic effects should be reduced.

Method

Participants

Forty Stony Brook students with self-reported normal vision and hearing participated in this experiment. None had participated in the previous experiments. Using the same criteria as before, we excluded 11 East Asian participants and three participants because they did not look at the screen during the task. Participants received partial course credit to fulfill a research requirement in psychology courses.

Materials

We used the same 30 videos (15 Asian, 15 Caucasian) as in Experiment 3.

Procedure

The procedures were the same as in the previous experiments. To be consistent with the procedures of the other experiments, the accent-rating task was run in two blocks, with the two blocks separated by the same filler task (i.e., computer game playing). However, because of the mixed design, there were no differences between the two blocks. Thus, half of the ten presentations of each stimulus were given in each block. Specifically, in each block, participants received five randomizations of 15 Asian videos and 15 Caucasian videos, with the two types of videos mixed and pseudo-randomly presented. Video presentation order differed for the two blocks, but the order of the stimuli within each block was the same for each participant. Each block took around 10 min.

Results

Two participants were excluded because their average ratings of the unambiguous Caucasian face videos were too similar to their average ratings of the unambiguous Asian face videos (i.e., they did not or could not pay attention to the accent). The operational definition of "too similar" was an average rating for the most native item (i.e., continuum step 8 dubbed onto the Caucasian face video) that was greater than 60% of the average rating of the most accented item (i.e., continuum step 1 dubbed onto the Asian face video) for the identification task in either block (see Samuel, 2016). We used the data from 24 participants in the analysis.

A four-way repeated measures ANOVA was conducted with four within-subject factors: Block (1 vs. 2), Face (Asian and Caucasian), Continuum Step (3, 4, 5, and 6), and Word (*cancer*, *theater*, and *thousand*). Figure 7 shows how the visual information (Asian vs. Caucasian) influenced participants' judgments of the four continuum steps for the three words; Fig. 8 shows the results collapsed across the continuum steps, for each of the three words individually. The means and standard deviations for all conditions are shown in Table 3.

As has been true in all of the experiments, the main effects of Continuum Step, F (3, 138) = 355.63, p < .001, η2 = .94, and of Word, F (2, 138) = 9.68, p < .001, η2 = .30, were significant. As would be expected by virtue of there being no difference in the stimuli or conditions across Blocks 1 and 2, performance did not differ across the two blocks, F (1, 138) = .58, p = .454, η2 = .03. The critical question is whether seeing an Asian versus a Caucasian video affected accentedness in a mixed design that minimized the opportunity for strategic effects. As Figs. 7 and 8 suggest, there was little or no such effect of Face in this mixed design, F (1, 138) = 1.31, p = .265, η2 = .05. The only hint of an effect was a significant interaction between Continuum Step and Face, F (3, 138) = 3.11, p = .032, η2 = .12. Pairwise comparisons

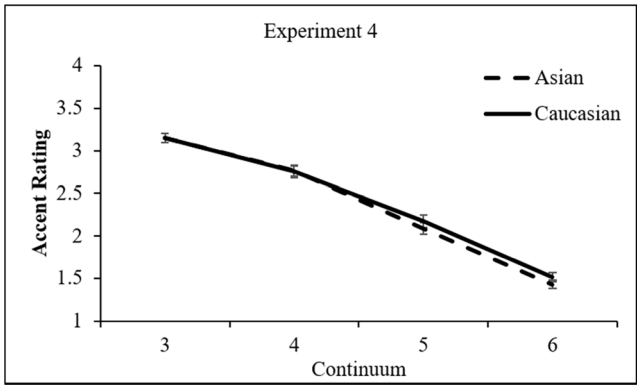


Fig. 7 Accentedness ratings of continuum steps 3–6 as a function of whether the face was Asian versus Caucasian. Error bars represent the standard error of the mean

showed that there was no effect of the Face at continuum steps 3–5 (*p's* > .05), but the Asian face was rated as more accented than the Caucasian face on continuum step 6 (mean difference = .088, $p = .032$). The overall lack of an effect was consistent across all three words, as Fig. 8 illustrates, with no interaction between Word and Face, $F(2, 138) = 1.52$, $p = .229$, $\eta 2 = .06$.

Discussion

The results of Experiment 4 showed that when we presented the same items as those in Experiment 3, but now in a mixed design, there was no overall effect of the ethnicity of the faces; there was a very small effect of Face on one step of the continuum. Overall, the results of Experiment 4 can be seen as the complement of those in Experiment 3: In one case, we designed the experiment to maximize potential decision-level factors (by including contrastive stimuli in a blocked design), whereas in the other we tried to minimize them. The quite different patterns of results for these two experiments in Part 2 demonstrate the degree to which interpretation, rather than perception, can dominate the outcome when asking listeners for judgments of accentedness.

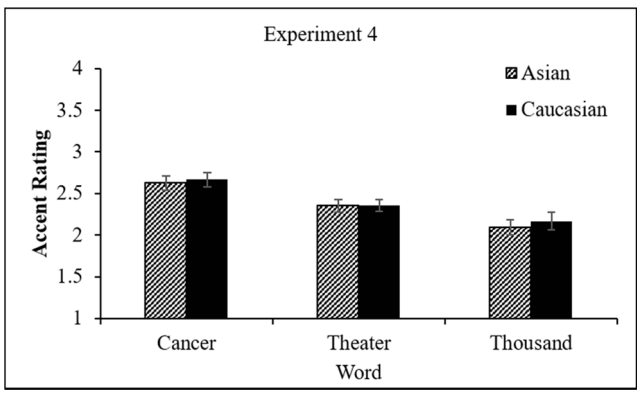


Fig. 8 Accentedness ratings of the three words separately as a function of whether the face was Asian versus Caucasian. Error bars represent the standard error of the mean

More broadly, looking across the results of the first four experiments, the systematic variation in accent ratings produced by our manipulations indicates that the "perceptual" effects of watching different faces discussed in previous studies (Levi et al., 2007; Magen, 1998; Rubin, 1992, 1998; Scales et al., 2006; Yi et al., 2013) are in fact interpretational effects. In Part 3 we use a second methodology to separate perceptual from interpretational effects.

Part 3

To isolate purely perceptual effects of accent, we used the selective adaptation paradigm. Selective adaptation is a reduction in the report of a stimulus after repeated exposure to similar stimuli. It was originally used with speech stimuli in Eimas and Corbit's (1973) study. They created a continuum between voiced and voiceless stop consonants and found that the phonemic boundary was shifted after repetitive presentation of an endpoint member of the continuum. For instance, if participants heard a repeating voiced consonant, their likelihood of reporting a voiced consonant was reduced; they reported fewer items of the continuum as voiced compared to the baseline. The selective adaptation paradigm has been used widely in later studies and has yielded strong and consistent effects for auditory stimuli (see Samuel, 1986 for a review of much of the literature). Selective adaptation is primarily sensitive to the perception of acoustic properties of the repeated sound. Its sensitivity to acoustic properties is largely unaffected by processing resource limitations, as studies have shown that a concurrent task that requires attentional resources does not lead to a reduction in the adaptation effect (Mullennix, 1986; Samuel & Kat, 1998; Sussman, 1993). In Experiments 5A and 5B, we use the selective adaptation paradigm to investigate the perception of accented speech.

Experiment 5A

The purpose of Experiment 5A is to test whether differences in accent produce adaptation; if they do, we can use adaptation to test whether audiovisually-determined accents can produce adaptation. In Experiment 5A, we used purely auditory adaptors – the endpoints of each eight-step continuum. If repeatedly hearing a clearly accented sound can generate adaptation, test words will sound less accented after hearing such accented tokens. Conversely, if hearing a clearly native sound produces adaptation, then test items will sound more accented after hearing the unaccented tokens.

Table 3 Means and standard deviations of accentedness as a function of block, face, and word in Experiment 4

		Step 1	Step 3	Step 4	Step 5	Step 6	Step 8
Caucasian Part 1	*cancer*		3.26 (.48)	3.14 (.51)	2.36 (.74)	1.65 (.48)	1.06 (.15)
	theater		3.18 (.63)	2.82 (.67)	2.06 (.61)	1.28 (.35)	1.10 (.22)
	thousand		2.62 (.59)	2.28 (.55)	1.96 (.50)	1.46 (.36)	1.02 (.08)
Asian Part 1	*cancer*	3.86 (.28)	3.63 (.36)	3.13 (.44)	2.36 (.72)	1.73 (.58)	
	theater	3.69 (.46)	3.14 (.59)	2.82 (.61)	1.95 (.65)	1.33 (.33)	
	thousand	3.02 (.60)	2.56 (.61)	2.14 (.66)	2.21 (.64)	1.60 (.48)	
Caucasian Part 2	*cancer*		3.66 (.62)	3.18 (.65)	2.31 (.63)	1.45 (.38)	1.05 (.12)
	theater		3.45 (.47)	2.75 (.50)	1.99 (.54)	1.30 (.27)	1.08 (.22)
	thousand		2.74 (.72)	2.41 (.75)	1.85 (.63)	1.43 (.40)	1.03 (.10)
Asian Part 2	*cancer*	3.81 (.44)	3.56 (.68)	3.08 (.70)	2.35 (.68)	1.49 (.44)	
	theater	3.75 (.36)	3.36 (.41)	2.96 (.51)	1.98 (.55)	1.32 (.32)	
	thousand	3.07 (.47)	2.67 (.79)	2.38 (.69)	2.18 (.70)	1.61 (.52)	

Method

Participants

For Experiment 5A (and Experiment 5B), we chose to obtain usable data from 48 participants (16 subjects for each of the three English words) using the same inclusion/exclusion criteria as in Experiments 1–4. Adaptation effects are typically relatively strong, so that a sample size of 16 per continuum is consistent with prior studies using this paradigm.

In Experiment 5A, 71 Stony Brook undergraduate students were tested; 11 were excluded because they did not return for the required second day of testing. Two of the remaining 60 participants were excluded because they were East Asian, and three participants' data were not used because of a computer failure during the experiment. Participants were drawn from the same population as in the prior experiments, and received partial course credit to fulfill a research requirement in psychology courses. Participants were tested in groups of up to three people at a time.

Materials

As noted above, we used only auditory stimuli in Experiment 5A. The test series were the eight-step continua created for the previous experiments, one continuum for each of the three words (*cancer*, *theater*, and *thousand*). The adaptors were the endpoints of the eight-step continuum of each word.

Procedure

There were two groups of participants in Experiment 5A. The first group received accented adaptors during their first testing session (i.e., on Day 1) and native adaptors during their second session (i.e., on Day 2); the order of adaptors was reversed for the second group. For each group, one-third of the participants heard only the word *cancer*, one-third heard only the word *theater*, and one-third heard only the word *thousand*, throughout the two-day experiment.

Each day, participants were instructed that there were two tasks during the session and that both tasks involved listening to simple English words and making a decision about each word that they would hear. The first task took about 5 min, and the second task took about 15 min.

On the first task (ID: baseline identification), participants listened to 20 randomizations of an eight-step continuum. They rated each sound in terms of its accentedness by pressing one of four buttons, using the same four-point scale as in the previous experiments. Participants were required to press a button within 3 s from the onset of each stimulus. One second after all participants had responded the next sound was presented. If one or more participants failed to respond within 3 s, the next item was automatically presented after 1 s.

Immediately after the first task, participants did the second task (Adapt: adaptation test). On this task, participants made the same decisions as they did on task 1, with one change in the presentation. There were periods of about 30 s during which participants just listened to a repeating word, the adaptor (30 repetitions of the adaptor, at a rate of approximately one presentation per second), without making any responses. The adaptation test consisted of 14 cycles, with each cycle including 30 repetitions of an adaptor followed by one randomization of the eight-item continuum for participants to identify. The randomization was preceded by a 500-ms pause, and the timing within the identification block was the same as in the baseline identification task (except that the maximum waiting time was 4 s, to give participants some extra time to respond as they switched from the "listening-only" condition to the "listening-and-responding" condition).

Results

On the Identification task, the first four passes of the eight-step continua were practice and were not scored. We calculated the average rating of each continuum step for the remaining 16 repetitions. On the adaptation task, we calculated the average ratings for each continuum step. We excluded six participants because their average rating of continuum step 8 was too similar to their rating of step 1. As before, "too similar" means that the average rating for the native item (continuum step 8) was greater than 60% of the average rating of the most accented item (continuum step 1) for the identification task on either day. These subjects were apparently not willing or able to judge accentedness reliably. We excluded one participant because he failed to respond at least ten times on at least one task. Complete sets of usable data were obtained from 48 participants (evenly distributed across conditions).

Figure 9 shows that when the adaptor was the native sound, participants' rating scores were higher than on the baseline identification test. Conversely, when the adaptor was accented, test items sounded less accented after adaptation. These shifts are the classic results in adaptation – a contrastive effect of the adaptor. Figure 10 shows that accent produced adaptation for each of the three words individually.

To quantify these effects, for each participant, we computed one number that was the average score across items 3, 4, 5, and 6 (the region of each continuum that was most ambiguous and thus most susceptible to shifts caused by adaptation) for both the baseline and the adaptation tasks. We conducted a four-way ANOVA on these scores: Presentation order (Accented adaptor on Day 1 vs. on Day 2) × Word (*cancer*, *theater*, and *thousand*) × Adaptor (Native vs. Accented) × Time (Baseline vs. after Adaptation). For the two within-subject factors, a significant main effect was found for Adaptor (F (1, 42) = 158.79, p < .001, η2 = .79) as well as for Time (F (1, 42) = 12.34, p = .001, η2 = .23). For the between-subject factors, there was no effect of Presentation order (F (1, 42) = .01, p = .907, η2 <.001), but the main effect for Word was significant (F (2, 42) = 9.73, p < .001, η2 = .32). See Table 4 for descriptive statistics.

The critical interaction is the one between Time and Adaptor (F (1, 42) = 194.32, p < .001, η2 = .82). The significant interaction demonstrates that adaptation worked, with the two adaptors shifting the judged accentedness differently from Baseline after adaptation. Pairwise comparisons showed that the difference between the accent ratings before and after adaptation was significant both for the accented adaptor (mean difference = .271, p < .001) and for the native adaptor (mean difference = .466, p < .001). The effect was consistent for all three words, all p's ≤ .003.

Discussion

Experiment 5A showed that accent produced adaptation, with the typical contrastive effect. This allows us to use adaptation to test whether visually different adaptors (Asian vs. Caucasian) combined with the same auditory token will produce a comparable effect. Experiment 5B provides this test.

Experiment 5B

In Experiment 5B, we aim to test whether visually different adaptors (Asian vs. Caucasian) produce different adaptation effects. If visual information affects the perception of accent, that is, if participants really perceive a sound as accented because it appears to be coming from an Asian speaker, and they hear the sound as unaccented because it appears to be coming from a Caucasian speaker, then these accented/unaccented adaptors should behave like those in Experiment 5A. If instead visual information only affects interpretation, not perception, of accent, then neither adaptor will produce an adaptation effect.

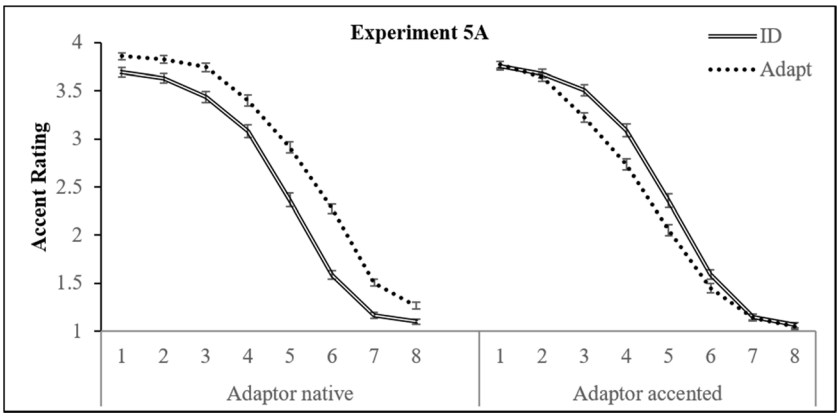


Fig. 9 Accentedness ratings of eight-step continua as a function of whether the adaptor was native versus accented. Error bars represent the standard error of the mean

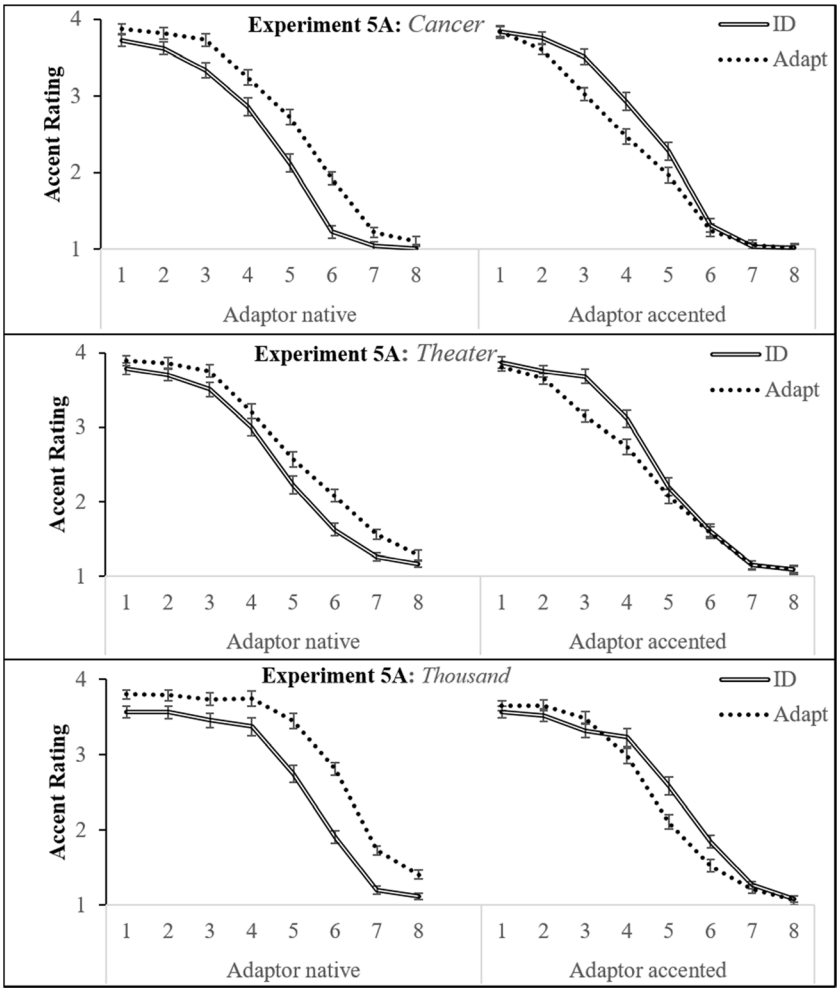


Fig. 10 Accentedness ratings of eight-step continua as a function of whether the adaptor was native versus accented for each of the three words separately. Error bars represent the standard error of the mean

The logic of Experiment 5B is similar to the logic Samuel (1997; Samuel, 2001) has used to demonstrate that lexical context can drive the perception of phonetic segments within a word. Samuel (1997) tested whether a phonetic segment produced by phonemic restoration has the same adapting properties as a phonetic segment that is acoustically present in a word. In phonemic restoration, a segment is deleted from a word and replaced by another sound, such as white noise. Listeners consistently report that the word sounds intact, indicating that they have perceptually restored the missing segment (Warren, 1970). Samuel (1997) took words like "alphabet" and "armadillo" and replaced the /b/ or the /d/ with white noise. These words were then used as adaptors, with a /b/ - /d/ test continuum. The restored phonemes produced the contrastive adaptation effect (restored /b/ reduced report of /b/, and restored /d/ reduced report of /d/), showing that they had been perceived, and were not just some decision-level interpretation. Experiment 5B uses the same logic, with videos providing the context (rather than words), and accent being the potentially perceived property (rather than /b/ or /d/).

Table 4 Means and standard deviations of accentedness as a function of adaptor, time, and word in Experiment 5A

Adaptor	Time	Word	*n*	*M*	*SD*
Accented	Baseline	*Cancer*	16	2.51	.29
		Theater	16	2.65	.32
		Thousand	16	2.75	.37
	After adaptation	*Cancer*	16	2.18	.29
		Theater	16	2.39	.28
		Thousand	16	2.53	.32
Native	Baseline	*Cancer*	16	2.39	.33
		Theater	16	2.59	.32
		Thousand	16	2.87	.31
	After adaptation	*Cancer*	16	2.91	.28
		Theater	16	2.91	.31
		Thousand	16	3.44	.23

Method

Participants

Another 61 Stony Brook undergraduate students participated in Experiment 5B. Of these, nine participants were excluded because they did not return for the second day of testing. One of the remaining participants was excluded because he was East Asian. Participants were compensated with partial course credit in a psychology course.

Materials

The same eight-step auditory-only continua were used as the test series, but we used audiovisual adaptors in Experiment 5B, rather than the purely auditory ones used in Experiment 5A. The baseline identification data of Experiment 5A showed that step 5 was the most ambiguous item for all three test words. Therefore, we used videos as adaptors in which the most ambiguous audios (step 5 for each continuum) were paired with videos of either the Asian face or the Caucasian face (6 audiovisual adaptors: 3 continua × 2 faces). Each one of these adaptors was conceptually related to an adaptor in Experiment 5A, except that in Experiment 5A the native versus accented quality of an adaptor was based on the auditory signal whereas in Experiment 5B this distinction was cued by the faces that were paired with the ambiguous auditory signal.

Procedure

In Experiment 5B, the procedures were similar to those in Experiment 5A, except that during the adaptation test, 30 repetitions of an audio-visual adaptor took about 60 s, and participants were instructed to watch the videos (instead of just listening to the sounds). Participants watched the Asian videos or the Caucasian videos as adaptors on two separate days, as they had heard accented or native adaptors on separate days in Experiment 5A. The order of the adaptors was counterbalanced across participants.

Results

The first four passes of the eight-step continua of the identification task were not scored, as before. We calculated the average accentedness rating for each step on each continuum, both for the identification task and for the adaptation task. One participant was excluded because his average rating of continuum step 8 (native) was too similar to his rating of step 1 (accented), using the same criterion as before. We excluded two participants because they failed to respond at least ten times on at least one task. Usable data were obtained from 48 participants (evenly distributed across conditions).

Figures 11 and 12 shows the results of Experiment 5B. Inspection of the figures makes it clear that unlike Experiment 5A, the adaptors here were completely ineffective. To assess the pattern statistically, we again computed the mean scores across items 3, 4, 5, and 6 to conduct a four-way ANOVA: Presentation order (Asian face adaptor on Day 1 vs. on Day 2) × Word (*cancer*, *theater*, and *thousand*) × Adaptor (a Caucasian face vs. an Asian face) × Time (Baseline vs. after Adaptation). The main effect for Word was significant, $F(2, 42) = 27.32$, $p < .001$, $\eta 2 = .57$, consistent with all of the previous experiments. No other effects even approached significance: The main effect for Presentation order was not significant, $F(1, 42) = 1.22$, $p = .275$, $\eta 2 = .03$, nor was the main effect for Time ($F(1, 42) = 1.29$, $p = .263$, $\eta 2 = .03$) or for Adaptor ($F(1, 42) = 2.32$, $p = .135$, $\eta 2 = .05$). The critical interaction of Time and Adaptor was also clearly not significant, $F(1, 42) = 1.75$, $p = .193$, $\eta 2 = .04$. The pattern – no effect – was consistent across each individual word, p's > .05, as shown in Fig. 12. The results clearly show that there was no adaptation. See Table 5 for descriptive statistics.

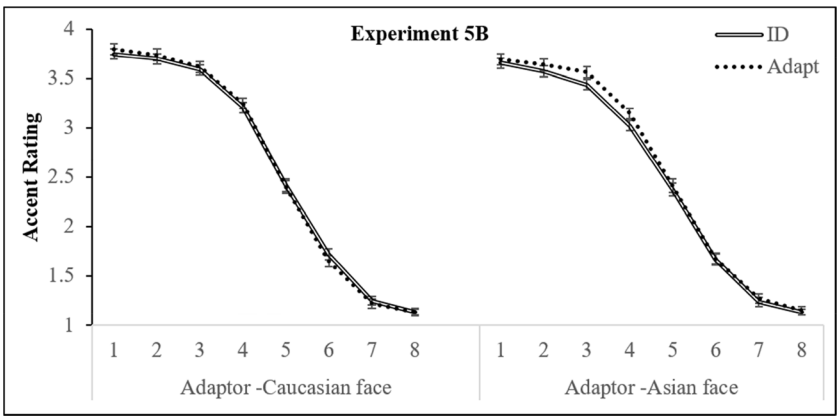


Fig. 11 Accentedness ratings of eight-step continua as a function of whether the adaptor included an Asian face versus a Caucasian face. Error bars represent the standard error of the mean

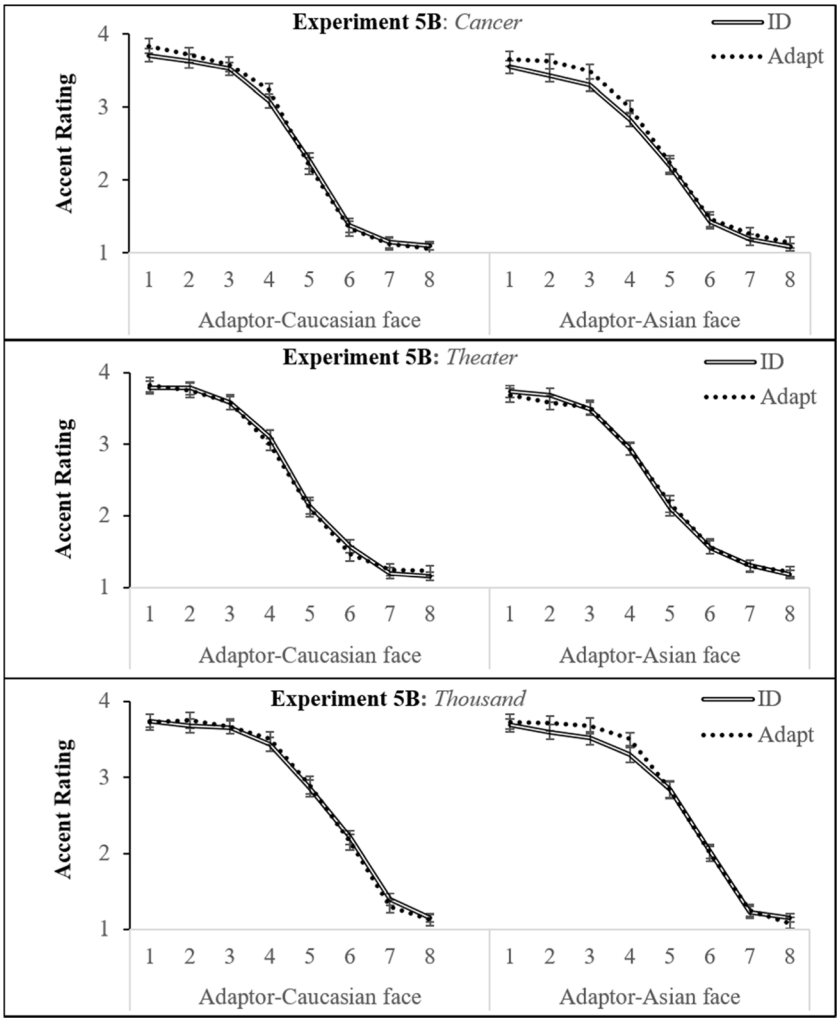


Fig. 12 Accentedness ratings of eight-step continua as a function of whether the adaptor included an Asian face versus a Caucasian face for each of the three words separately. Error bars represent the standard error of the mean

Discussion

The absence of the Time × Adaptor interaction shows that visually different adaptors failed to yield adaptation effects, as is clear in Figs. 11 and 12. Taken together with the findings of Experiment 5A that showed that differently accented adaptors produced adaptation effects, this null result demonstrates that visual information did not play a role in the perception of accent.

Previous research has shown that some types of context affect perceptual adaptation (Samuel, 1997, 2001) but others may not (Banks, Gowen, Munro, & Adank, 2015; Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014; Swerts & Krahmer, 2004). Generally speaking, lexical context has proven to be effective, while visual context has not. Swerts and Krahmer (2004) have suggested that visual information is given less weight than auditory information in participants' perception of accent. Consistent with the literature, the results of Experiments 5A and 5B show that adaptation can be driven by the auditory component of

Table 5 Means and standard deviations of accentedness as a function of adaptor, time, and word in Experiment 5B

Adaptor	Time	Word	*n*	*M*	*SD*
Accented	Baseline	*Cancer*	16	2.44	.33
		Theater	16	2.52	.21
		Thousand	16	2.92	.34
	After adaptation	*Cancer*	16	2.54	.31
		Theater	16	2.55	.29
		Thousand	16	3.00	.32
Native	Baseline	*Cancer*	16	2.56	.29
		Theater	16	2.60	.21
		Thousand	16	3.04	.30
	After adaptation	*Cancer*	16	2.59	.25
		Theater	16	2.54	.28
		Thousand	16	3.06	.34

speech (i.e., the accentedness of sounds) but not by its visual component (i.e., the ethnicity of faces).

General discussion

Previous studies showed that the ethnicity of a speaker, signaled by a picture, significantly affected people's judgments of the accent of the speaker (Rubin, 1992; Rubin et al., 1999; Rubin & Smith, 1990; Yi et al., 2013, 2014). The current study was designed to determine the nature of this effect. In particular, the goal was to test whether the effect was taking place at a perceptual level, or was instead based on later interpretation.

In Part 1, we examined the possible effect of demand characteristics produced by the pictures. With static photos, under conditions most like those in previous studies (i.e., the effectively between-subject design of the first block), we replicated the increase in judged accentedness of speech when an Asian face was shown, rather than a Caucasian face. In Experiment 2, by changing the static faces to an integral combination of visual information with the speech, the demand characteristics were reduced, largely abolishing the effect. Rubin's findings (Kang & Rubin, 2009; Rubin, 1992; Rubin et al., 1999; Rubin & Smith, 1990) have been cited in concerns about possible negative biases against non-native speakers (e.g., teaching assistants, or job applicants) based on their appearance. If we take Experiments 1 and 2 as being somewhat analogous to two versions of a real-world situation that is prone to bias, the results are potentially encouraging: If an Asian job candidate was assumed to be difficult to understand based on an application form (e.g., European resumes typically include a picture of the applicant), an actual interview (where the face and speech are integrated) could reduce the bias.

In Part 2, we varied factors that are known to affect decisions, and we found that the interpretation of accentedness while watching an Asian face is subject to these context effects. Whereas participants had a weak tendency to rate Asian videos to be more accented than the Caucasian videos in Experiment 2, with the mixed design of Experiment 4 there was no difference, and the effect could even be reversed with a contrast manipulation (Experiment 3). Collectively, the results of these identification experiments show that visual information affects the interpretation of accented speech on the decision level, rather than actually altering the way the speech sounds.

To provide a converging test of this conclusion, in Part 3 we used the selective adaptation paradigm. Experiment 5A showed that truly accented speech produces adaptation, but in Experiment 5B audiovisual adaptors (with the most ambiguous member of continuum dubbed onto an Asian face or a Caucasian face) did not. Previous studies using the same logic have demonstrated perceptual effects of lexical context (Samuel, 1997, 2001). The absence of adaption here indicates that the perception of accentedness does not differ as a function of the two faces.

Collectively, in contrast with previous claims about how the ethnicity of a face affects the perception of accentedess, the evidence provided in the current study indicates that visual information influences people's interpretation of accentedness, but not their actual perception of accentedness. We believe that the different conclusions stem from the fact that "perception" is a term that is used in two quite different ways. Here, we have used it in the restricted sense of what people actually hear. This is the more precise usage recommended by Firestone and Scholl (2015), Norris et al. (2000), and Samuel (1997, 2001). As those authors have noted, there is a more general use of "perception" that lumps together the more specific sense of perception with the decision level interpretation of stimuli. Previous authors talking about accent perception have generally used this broader sense of the term.

Even if seeing an Asian face does not truly affect people's perception of accented speech, it is important to realize that a decision level bias against Asian faces, Asian accented speech, or even speakers of that accent, matters in real social contexts. Previous studies have shown that native English speakers tend to hold negative attitudes toward Asian-accented English, and this can generalize to negative evaluations of the speakers of that accent (Cargile, 1997; Gill, 1994; Grossman, 2011; Hosoda, Stone-Romero, & Walter, 2007; Jacobs & Friedman, 1988; Lindemann, 2002, 2003, 2005). For instance, Asian-accented English speakers were perceived as poorer communicators (Hosoda et al., 2007), less likable and less competent than native English speakers (Grossman, 2011); they were also rated as less competent in the contexts of both employment interviews and college classrooms (Cargile, 1997). Kim, Wang, Deng, Alvarez, and Li (2011) showed that English proficiency among Chinese Americans was related to the speakers' depressive symptoms over time, suggesting that negative attitudes toward Chinese-accented English can have a significant impact on the speakers. The negative impact on those whose speech differs from standard American English is by no means limited to Asian accents: Spanish-accented speakers suffer at job interviews, African-American instructors face challenges from their students in building credibility and acceptance, and non-native speakers are more likely to be fired due to their accented English than native speakers (Hendrix, 1998; Lippi-Green, 1997; Rubin, 1998). Moreover, even when foreign teaching assistants' teaching was as effective as that of native teaching assistants, students' satisfaction was lower for foreign teaching assistants (Fleisher, Hashimoto, & Weinberg, 2002).

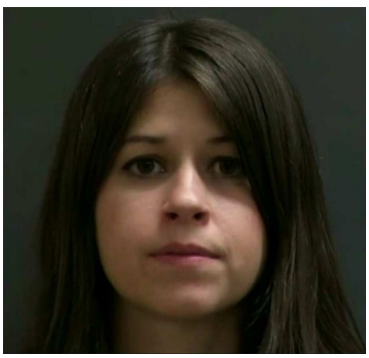
Given all of these negative consequences, Rubin (1998) has argued that university training programs should not only focus on enhancing foreign instructors' linguistic skills but also on improving students' attitudes and listening skills. The current study offers new insights into this issue by demonstrating that Asian faces do not affect accentedness of speech on a perceptual level. This fact offers hope in the sense that it should be easier to change decisions/interpretations than perception itself. As a

practical matter, our results highlight the potential demand characteristic involved in photographs of an Asian face. That is, judging a person by looking at a photo is clearly not the most accurate way to know that person; instead, face-to-face personal interactions will offer more opportunities to gain a deeper understanding of the individual, and thereby reduce decision level bias.

Acknowledgments We thank Richard Gerrig and Antonio Freitas for their valuable suggestions on the current study, and Maxwell Carmack for his great help in creating the continua using TANDEM STRAIGHT. We also appreciate the constructive suggestions of two anonymous reviewers. Support was provided by Ministerio de Ciencia E Innovacion, Grant PSI2014-53277, Centro de Excelencia Severo Ochoa, Grant SEV-2015-0490, and by the National Science Foundation under Grant IBSS-1519908.

Appendix 1

The Asian face and Caucasian face used in the experiments



References

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Audiovisual cues benefit recognition of accented speech in noise but not perceptual adaptation. *Frontiers in Human Neuroscience*, *9*.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science, 17*(2), 159–163.

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

Boersma, P. & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Cargile, A. C. (1997). Attitudes toward Chinese-accented speech an investigation in two contexts. *Journal of Language and Social Psychology, 16*(4), 434–443.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*(1), 99–109.

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, 1–72.

Fleisher, B., Hashimoto, M., & Weinberg, B. A. (2002). Foreign GTAs can be effective teachers of economics. *The Journal of Economic Education, 33*(4), 299–325.

Gill, M. M. (1994). Accent and stereotypes: Their effect on perceptions of teachers and lecture comprehension.

Grossman, L. (2011). The effects of mere exposure on responses to foreign-accented speech.

Hendrix, K. G. (1998). Student perceptions of the influence of race on professor credibility. *Journal of Black Studies, 28*(6), 738–763.

Hosoda, M., Stone-Romero, E. F., & Walter, J. N. (2007). Listeners' cognitive and affective reactions to English speakers with standard American English and Asian accents. *Perceptual and Motor Skills, 104*(1), 307–326.

Irwin, A. (2008). *Investigating the effects of accent on visual speech* (Doctoral dissertation, University of Nottingham).

Jacobs, L. C., & Friedman, C. B. (1988). Student achievement under foreign teaching associates compared with native teaching associates. *The Journal of Higher Education*, 551–563.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*.

Kawahara, H., & Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *SADHANA - Academy Proceedings in Engineering Sciences, 36,* 713–722.

Kawase, S., Hannah, B., & Wang, Y. (2014). The influence of visual speech information on the intelligibility of English consonants produced by non-native speakers. *The Journal of the Acoustical Society of America, 136*(3), 1352–1362.

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., … Pascalis, O. (2007). Cross-race preferences for same-race faces extend beyond the African versus Caucasian contrast in 3-month-old infants. *Infancy*, *11*(1), 87–95.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., … Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental science*, *8*(6), F31-F36.

Kim, S. Y., Wang, Y., Deng, S., Alvarez, R., & Li, J. (2011). Accent, perpetual foreigner stereotype, and perceived discrimination as indirect links between English proficiency and depressive symptoms in Chinese American adolescents. *Developmental Psychology, 47*(1), 289.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *The Journal of the Acoustical Society of America, 121*(4), 2327–2338.

Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society, 31*(03), 419–441.

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics, 7*(3), 348–364.

Lindemann, S. (2005). Who speaks "broken English"? US undergraduates' perceptions of non-native English1. *International Journal of Applied Linguistics, 15*(2), 187–212.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.

Magen, H. S. (1998). The perception of foreign-accented speech. *Journal of Phonetics, 26*(4), 381–400.

Mullennix, J. W. (1986). *Attentional limitations in the perception of speech*. Unpublished doctoral dissertation. Buffalo, NY: State University of New York at Buffalo.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*(01), 111–131.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(03), 299–325.

Orne, M. T. (2009). Demand characteristics and the concept of quasi-controls. *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*, *110*.

Rau, D., Chang, H. H. A., & Tarone, E. E. (2009). Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative. *Language Learning, 59*(3), 581–621.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*(4), 309–314.

Rogers, C. L., & Dalby, J. (2005). Forced-choice analysis of segmental production by Chinese-accented English speakers. *Journal of Speech, Language, and Hearing Research, 48*(2), 306–322.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511–531.

Rubin, D. L. (1998). Help! My professor (or doctor or boss) doesn't talk English. *Readings in Cultural Contexts*, 149–160.

Rubin, D. L., Ainsworth, S., Cho, E., Turk, D., & Winn, L. (1999). Are greek letter social organizations a factor in undergraduates perceptions of international instructors? *International Journal of Intercultural Relations, 23*(1), 1–12.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations, 14*(3), 337–353.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *The Journal of the Acoustical Society of America, 95*(6), 3658–3661.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology, 18*(4), 452–499.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*(2), 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348–351.

Samuel, A. G. (2016). Lexical representations are malleable for about one second: Evidence for the non-automaticity of perceptual recalibration. *Cognitive Psychology, 88,* 88–114.

Samuel, A. G., & Kat, D. (1998). Adaptation is automatic. *Attention, Perception, & Psychophysics, 60*(3), 503–510.

Samuel, A. G., & Lieblich, J. (2014). Visual speech acts differently than lexical context in supporting speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1479.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & De Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16*(6), 440–444.

Scales, J., Wennerstrom, A., Richard, D., & Wu, S. H. (2006). Language learners' perceptions of accent. *Tesol Quarterly, 40*(4), 715–738.

Sussman, J. E. (1993). Focused attention during selective adaptation along a place of articulation continuum. *The Journal of the Acoustical Society of America, 93*(1), 488–498.

Swerts, M., & Krahmer, E. (2004). Congruent and incongruent audiovisual cues to prominence. In *Speech Prosody 2004, International Conference*.

Wang, Y., Martin, M. A., & Martin, S. H. (2002). Understanding Asian graduate students' English literacy problems. *College Teaching, 50*(3), 97–101.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392–393.

Yi, H. G., Phelps, J. E., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America, 134*(5), EL387–EL393.

Yi, H. G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience, 8,* 768.

Zhang, F., & Yin, P. (2009). A study of pronunciation problems of English learners in China. *Asian Social Science, 5*(6), 141.