Acta Psychologica 168 (2016) 85–90

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Auditory perceptual simulation: Simulating speech rates or accents?

Peiyun Zhou *, Kiel Christianson

[a] University of Illinois, Urbana-Champaign, United States
[b] Beckman Institute for Advanced Science and Technology, United States

ARTICLE INFO

Article history:
Received 5 October 2015
Received in revised form 16 April 2016
Accepted 18 April 2016
Available online xxxx

Keywords:
Auditory Perpetual Simulation
Native and non-native speech
Speech rates
Accent
Reading
Eye-tracking

ABSTRACT

When readers engage in Auditory Perceptual Simulation (APS) during silent reading, they mentally simulate characteristics of voices attributed to a particular speaker or a character depicted in the text. Previous research found that auditory perceptual simulation of a faster native English speaker during silent reading led to shorter reading times that auditory perceptual simulation of a slower non-native English speaker. Yet, it was uncertain whether this difference was triggered by the different speech rates of the speakers, or by the difficulty of simulating an unfamiliar accent. The current study investigates this question by comparing faster Indian-English speech and slower American-English speech in the auditory perceptual simulation paradigm. Analyses of reading times of individual words and the full sentence reveal that the auditory perceptual simulation effect again modulated reading rate, and auditory perceptual simulation of the faster Indian-English speech led to faster reading rates compared to auditory perceptual simulation of the slower American-English speech. The comparison between this experiment and the data from Zhou and Christianson (2016) demonstrate further that the “speakers'” speech rates, rather than the difficulty of simulating a non-native accent, is the primary mechanism underlying auditory perceptual simulation effects.

Auditory Perceptual Simulation (APS) refers to the phenomenon when readers mentally simulate characteristics of either the voices of the characters depicted in texts or the voices of other speakers (including their own) while they read silently (e.g., imagining Daniel Radcliffe's voice from the Harry Potter films when reading the Harry Potter books; see Hubbard, 2010, for a review of auditory imagery). When participants activate auditory perceptual simulation of speech during reading, they generate a rich mental representation of a depicted or imagined speaker “saying” the words in the text. This mental representation could be a more elaborated version of the normal implicit prosody that most skilled readers generate when reading silently (Fodor, 2002). The auditory perceptual simulation representation includes both segmental and suprasegmental information about the depicted or imagined “speaker's” voice, including, e.g., speech rate (Stites, Luke, & Christianson, 2013; Yao & Scheepers, 2011), accent (Filik & Barber, 2011), and characters' perspectives (Drumm & Klin, 2011; Gunraj & Klin, 2012; Levine & Klin, 2001).

The research on auditory perceptual simulation effects (or auditory imagery) on reading date back to Kosslyn and Matt (1977), who explored whether participants would activate talker-specific auditory imagery while reading aloud after hearing some speakers' recordings. They first familiarized the readers with two speakers, one faster and the other slower, by playing recorded passages. Then, participants were asked to read aloud passages that were purportedly “written” either by the faster or slower speaker. The observed data showed that participants read the faster speaker's text more quickly than the slower speaker's text, suggesting that the participants activated the perceptual features of the corresponding speaker's voice while reading aloud.

Alexander and Nygaard (2008) extended this finding of engaging in talker-specific auditory imagery during reading aloud to silent reading. They adapted Kosslyn and Matt's paradigm in a silent reading task and manipulated the speech rates of two speakers and text difficulty. Readers were first familiarized with these two speakers' voices (fast vs. slow), then they were told to read the passages that were “written” either by the faster or slower speaker silently and answer comprehension questions. The results demonstrated that even silent reading speeds were modulated corresponding to the speech rates of the speakers. Moreover, readers were more likely to activate auditory imagery of the speakers when the texts were difficult.

Kurby, Magliano, and Rapp (2009) found that auditory imagery could be influenced by familiarity with the speakers and the texts. Participants recognized more words when read by a familiar speaker in a novel script, and repeated exposure to the text strengthened their mental representation of a character's voice, facilitating quicker recognition of the words read by the same character later. Thus, the authors concluded that readers activate perceptually based knowledge while reading even without direct experience of the voice in the particular context.

In recent studies, scholars extended the trigger of the auditory imagery from the voices alone to the photos of the speakers (Woumans et al.,

* Corresponding author at: Department of Educational Psychology, University of Illinois, Urbana-Champaign, 1310, S. Sixth Street, MC 201, Champaign, IL 61801, United States.
E-mail address: pzhou5@illinois.edu (P. Zhou).

http://dx.doi.org/10.1016/j.actpsy.2016.04.005

2015; Zhou & Christianson, in preparation). Woumans et al. familiarized Spanish-Catalan bilinguals with two speakers' photos and their corresponding languages, either in Spanish or Catalan, during simulated Skype conversations. In the later language production task, one speaker's photo was presented on a Skype interface while saying a noun in the corresponding language. The subjects were asked to produce the first verb they associated it with and in the same language as the given stimulus. Results demonstrated that subjects responded faster when speaker's photo was congruent with the corresponding language, indicating that the photo on the Skype interface cued bilingual subjects to use the corresponding language. In a follow-up experiment, the finding was replicated among the Dutch-French bilinguals with the same paradigm and task.

Zhou and Christianson (2016) used eye tracking to investigate how auditory perceptual simulation of native and non-native English speech affects sentence processing and comprehension. They also used speakers' photos as the cues to trigger the auditory perceptual simulation effects: an English speaker's photo was matched with the faster native American English speech while a Chinese speaker's photo was matched with the slower Chinese accented English. The researchers first familiarized participants with the native and non-native speech by playing separate recordings while presenting the corresponding native and non-native speakers' photos on the screen. Then, subjects were asked to read sentences (e.g., "*The policeman that chased the thief drove fast.*") and respond to a paraphrase verification probe after each sentence (e.g., *The policeman chased the thief. The policeman drove fast.* True/False). Before each sentence was presented, one of the speakers' photos appeared on the screen, and the participants were asked to imagine this speaker's voice while reading the upcoming sentence. The total sentence reading time, response accuracy, fixation durations on individual words, and saccade patterns were analyzed.

The results demonstrated that (1) participants read sentences attributed to the slower-speaking non-native speaker more slowly than sentences attributed to the faster-speaking native speaker; 2) participants who were induced to perform auditory perceptual simulation (independently of which speaker's voice had been cued) read sentences faster, in terms of total sentence reading times as well as early and late measures on individual words, than participants (in a separate session) who were not induced to perform auditory perceptual simulation; 3) there were no significant differences in comprehension probe response accuracy between auditory perceptual simulation of native speech and auditory perceptual simulation of non-native speech; and 4) participants who were induced to perform auditory perceptual simulation of either native or non-native speech showed better comprehension overall, most markedly in morphosyntactically complex (object-relative clauses) and semantically implausible sentences, such as "*The bird that the worm ate was small*," compared to those who read under normal silent reading conditions.

Zhou and Christianson argued that the online reading speed differences in the two auditory perceptual simulation conditions were modulated by the different speech rates of the native and non-native speakers—auditory perceptual simulation of the faster native speaker's voice led to faster reading speeds, whereas auditory perceptual simulation of the slower non-native speaker's voice yielded slower reading speeds. However, a large body of sociolinguistic research has shown that native speakers usually have high standards of acceptability for using their language and judge people based on how they differ from these standards (e.g., Giles & Watson, 2013; Lippi-Green, 1997; Ryan, 1983). Native English speakers often rate non-natively accented speech as less comprehensible, less favorable, less trustworthy, and less persuasive compared to native speech in various settings (Brennan & Brennan, 1981; Callen, Callois, & Forbes, 1983; Edwards, 1977; Gass & Varonis, 1984; Giles, 1972; Giles, Hewstone, Ryan, & Johnson, 1987; Gluszek & Hansen, 2013; Kinzler, Shutts, DeJesus, & Spelke, 2009; Munro & Derwing, 1995, 1998; Varonis & Gass, 1982; White & Li, 1991). In Munro and Derwing's (1995) study, when participants were asked to decide the truth-value of statements read either by Chinese-accented English speakers or native English speakers, they spent more time processing the statements read by native Chinese speakers than the ones read by native English speakers. Moreover, listeners have been observed to rate statements (e.g., *A giraffe can go without water longer than a camel can*) read by non-native speakers as less credible than statements read by native speakers (Lev-Ari & Keysar, 2010). Thus, it might be argued that, in the Zhou and Christianson study, readers' slower reading speed when activating auditory perceptual simulation of non-native speech was triggered by difficulty in simulating accented English speech, rather than the speech rate per se.

In the current study, one eye-tracking experiment was conducted to investigate whether earlier auditory perceptual simulation effects were triggered by the native and non-native speakers' speech rates or by difficulty that the English-speaking readers may have experienced in simulating an unfamiliar accent. The eye-tracking methodology was applied here for two reasons. First, we wanted to use the same paradigm from Zhou and Christianson's study, which has been shown to reliably trigger auditory perceptual simulation effects. We also wanted to ensure that the results from the current study were comparable to the previous one. Additionally, eye-tracking provides a very accurate measure of readers' fixations and reading times on the target sentences.

This experiment also manipulated the same conditions as in Zhou and Christianson (2016): plausibility, syntactic complexity, and "speaker" identity. Plausibility and structure were manipulated in the experiment because previous studies have shown consistent reading and comprehension patterns for these sentences: implausible sentences are read more slowly than plausible sentences, and object relative clauses are read more slowly than subject relative clauses (Gibson, Desmet, Grodner, Watson & Ko, 2005; Gennari & MacDonald, 2008; Traxler, Morris, & Seely, 2002; Zhou & Christianson, 2016). Readers also usually have more difficulty comprehending object-relative clauses than subject-relative clauses, and implausible sentences are more likely to me misinterpreted than plausible sentences. Importantly, when syntactic complexity and semantic plausibility are crossed in this way, comprehension errors are systematic and predictable. Specifically, readers (and listeners) tend to derive interpretations of implausible sentences with more difficult structures such that the actors (arguments) in these sentences are often reversed. This observation was originally made for passive vs. active sentences (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003; Lim & Christianson, 2013a), and more recently extended to subject- vs. object-relative clauses (Lim & Christianson, 2013b; Zhou & Christianson, 2016). For example, the misinterpretation that is frequently derived from *The bird that the worm ate was small* (1d, below) is that "the bird ate the worm." This pattern of misinterpretation has been attributed to "good-enough" processing (Christianson, 2016; Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira & Patson, 2007; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Zhou & Christianson, 2016). According to Good Enough theory, these misinterpretations are due to the use of language processing heuristics – such as plausibility and probabilistic word order cues (e.g., Townsend & Bever, 2001) – which operate along with algorithmic morphosyntactic processing. Syntactic structure is fragile, however (Sachs, 1967), and the output of the heuristic-based processing overwhelms the output of the morphosyntactic processing in a significant proportion of trials. Zhou and Christianson (2016) observed a decrease in the occurrence of this sort of misinterpretation when people read with auditory perceptual simulation. We hypothesized that this improved comprehension stemmed from a richer, more robust morphosyntactic representation that was generated with and supported by the richer, more detailed prosodic representation created via auditory perceptual simulation.

For the "speaker" identity condition, an Indian-English speaker and an American-English speaker were recruited to read the texts. Instead of a native Chinese speaker, as had been used in Zhou and Christianson (2016), an Indian-English speaker's voice was used in

Table 1
Means and standard deviations (SD) for sentence reading time (in ms) and accuracy in the current experiment.

“Speaker”	Structure	Sentence reading time		Accuracy	
		Plausible	Implausible	Plausible	Implausible
Native	Subject-RC	4067 (2211)	4555 (2972)	0.96 (0.20)	0.93 (0.25)
	Object-RC	4672 (2573)	5075 (3259)	0.88 (0.33)	0.69 (0.46)
Non-native	Subject-RC	4058 (2071)	4279 (2520)	0.95 (0.22)	0.95 (0.21)
	Object-RC	4544 (2572)	4874 (2690)	0.89 (0.32)	0.69 (0.47)

Table 2
Fixed Effects of Logit Mixed-effects Model for Response Accuracy.

Predictor	Estimate	SE	*z* value	*p* value
(Intercept)	0.71	0.18	3.85	<0.001
Plausibility	1.36	0.18	7.64	<0.001
Structure	2.2	0.2	10.74	<0.001
“Speaker”	−0.01	0.11	−0.05	0.96
Trial order	0.003	0.001	2.28	<0.05
Plausibility × structure	−1.12	0.31	−3.65	<0.001

this experiment to validate that the auditory perceptual simulation effect is not limited to a particular accent. Furthermore, the Indian-English speaker with whose voice participants were familiarized spoke faster than the American-English speaker. In this way, across the previous and current studies, the nativeness and speech rate manipulations were counterbalanced (native vs. non-native; faster vs. slower).

The hypothesis here is that if the auditory perceptual simulation effects reported in Zhou and Christianson (2016) derived from readers' difficulty in simulating an unfamiliar accent, regardless of the speech rate difference, mentally simulating the Indian-English speech here should yield longer sentence reading times compared to perceptual simulation of the American-English speech. If, on the other hand, the previous auditory perceptual simulation effects were triggered by different speech rates rather than accents, readers in this experiment should read no more slowly when they perceptually simulate the faster Indian-English speaker's voice than when they simulate the slower American-English speaker's voice. Readers might read sentences attributed to the faster Indian-English speaker more quickly than sentences attributed to the native American-English speaker; however, there may be an upper limit on how fast people can read and simulate, such that any speed advantage for the faster speaker might be relatively small. Finally, we predict no significant comprehension accuracy differences between auditory perceptual simulation of the two speakers, consistent with Zhou and Christianson (2016).

In addition, here we combine and compare the data from the current experiment to the previous experiment in Zhou and Christianson, where a faster native speaker's voice and a slower non-native speaker's voice were used.[1] The combined data include “speaker” (native vs. non-native) and speech rate (fast vs. slow) as the independent variables to predict reading speed. The prediction is that if accent modulates the auditory perceptual simulation effects, “speaker” but not speech rate, will be the significant predictor of sentence reading speed. Otherwise, speech rate but not “speaker” should be the driving force of the reading speed.

1. Method

1.1. Participants

Eighty-seven native American-English speakers with normal or corrected-to-normal vision from the University of Illinois at Urbana-Champaign community participated in the experiment. They received either $7 payment or 1 research credit. Seven subjects' data were excluded due to either fatigue or difficulty in calibrating their eye gaze. Eighty native participants[2] were retained in the data analysis.

1.2. Materials

Four 500-words texts, 48 experimental sentences, and the social attractiveness questionnaire used in Zhou and Christianson (2016) were adapted in this experiment. Four texts that were counterbalanced for length and difficulty were used as the auditory materials. A female native Indian-English speaker and a female native American-English speaker, both in their early 20s, were recruited to record the auditory materials. The speaking rates of the speakers were significantly different: the American-English speaker had a significantly slower mean reading rate than the Indian-English speaker ($t(3) = 145$, $p < 0.001$).[3] Photos were presented simultaneously with the recordings. These were not of the actual speakers, but rather non-copyrighted stock photos from the Internet. The photo of the American-English speaker was of a blonde, Caucasian woman who appeared to be in her early 30s. The photo of the Indian-English speaker was of an Indian woman who appeared to be in her early 30s. The photos showed the women from the shoulders up. Both women were smiling and dressed in a business jacket and blouse. These photos were purposely selected to suggest that the two speakers were of similar ages and had similar styles, jobs, and socioeconomic statuses. The Indian-English speaker's voice was always matched to the Indian woman's photo, and the American-English speaker's voice to the Caucasian woman's photo.

Forty-eight target sentences and 96 filler sentences were included, all of which were grammatical. Two factors, structure and plausibility, were manipulated in the target sentences, resulting in a 2 (subject-relative clause, object-relative clause) × 2 (plausible, implausible) × 2 (“speaker” identity) fully-crossed, within-participants and within-items design.

Examples of the four sentence conditions are presented in (1). Before each sentence was presented, one of the two speakers' photos appeared. After each sentence, a paraphrase verification probe was used to measure readers' comprehension. Readers were asked to decide whether the verification probe (1e) was true or false based on the sentence they had just read. The answer was always “True” for experimental items, but the number of True and False correct responses was equal across the experiment. Materials were distributed pseudo-randomly for each participant across four lists in a Latin-square design, so that every participant saw each item only once.

(1) a. “The bird that ate the worm was small.” (subject-relative, plausible)
b. “The worm that ate the bird was small.” (subject-relative, implausible)
c. “The worm that the bird ate was small.” (object-relative, plausible)
d. “The bird that the worm ate was small.” (object-relative, implausible)
e. The bird/worm ate the worm/bird. The bird/worm was small. (T/F)

The social attractiveness survey included fourteen attributes, including accent, speech rate, comprehensibility, conscientiousness, confidence,

[1] In Zhou and Christianson (2016), the duration of the native English speech: text 1 = 2 min 50 s; text 2 = 2 min 57 s; text 3 = 2 min 49 s; text 4 = 2 min 57 s. The duration of non-native speech: text 1 = 4 min 08 s; text 2 = 4 min 18 s; text 3 = 4 min 08 s; text 4 = 4 min 20s.

[2] Based on Zhou & Christianson (2016), we calculated the sample size ($N = 48.5$). Because the plan from the outset was to compare the results from this experiment to the previous study, however, we ran the same number of subjects as in the previous study ($N = 80$).

[3] Duration for American-English speech: text 1 = 3 min 43 s; text 2 = 3 min 43 s; text 3 = 3 min 43 s; text 4 = 3 min 43 s. Duration for Indian-English speech: text 1 = 2 min 53 s; text 2 = 2 min 59 s; text 3 = 2 min 57 s; text 4 = 3 min 02 s.

Table 3
Fixed-effects for linear mixed model for the sentence reading time.

	Estimate	SE	df	*t* value	*p* value
Intercept	0.42	0.09	134.4	4.47	<0.001
Plausibility	−0.14	0.05	141.2	−3.12	<0.01
Structure	−0.21	0.05	147.6	−4.65	<0.001
“Speaker”	−0.05	0.03	105.7	−1.86	0.06
Trial order	−0.003	0.0003	1408	−8.33	<0.001

dependability, education, honesty, intelligence, likability, reliability, pleasantness, prestigious, and sincerity. Participants rated each speaker on a 1–7 Likert scale (e.g., “Please rate the pleasantness of the speaker's voice on a 1–7 Likert scale”, 1 = not pleasant at all, 7 = very pleasant. See the complete survey in the Appendix A in Zhou and Christianson (in preparation). In addition, each subject reported how often they performed auditory perceptual simulation of the speakers' voices during the eye-tracking portion of the experiment (e.g. “How often were you perceptual simulating the native speaker's (Judy's) voice when you were asked to? a. All of the time (100%). b. Most of the time (80%). c. Half of the time (50%). d. Sometimes (30%). e. Rarely (10%). f. I was not able to do so (0%)).

1.3. Apparatus

Eye movement data and comprehension data were collected using an SR Research EyeLink 1000 remote desktop eye tracker. A chin rest and forehead rest were used to stabilize participants' heads. Sentences were presented in black Courier New monotype font (14 pt) on a white background, which was approximately 70 cm away from the participants. At this distance, 1° of visual angle subtended approximately three characters. All viewing was binocular, but data were recorded from the right eye. A separate computer was used to collect the survey results.

1.4. Procedure

The entire experiment lasted less than one hour. After the participant provided informed consent, he/she was calibrated on the eye tracker using a nine-point calibration procedure to start the eye-tracking portion. After the calibration, the participant listened to two recordings, one by the American-English speaker, the other by the Indian-English speaker, while the corresponding photograph was displayed on the computer screen. The order of recordings was counterbalanced across participants, with each one presented first half of the time. Then participants read and responded to the six practice items and, after any questions, proceeded to read and respond to the experiment items. In each trial, participants first saw either the American-English or Indian-English speaker's photo while listening to a recording of the corresponding speaker saying her name (“Charu”/“Judy”). Then on the next screen, a sentence was presented. Participants were instructed to imagine that the speaker whose photo had just been presented on the previous screen was saying it to them as they read it silently. Another recording set of the American-English and Indian-English speakers reading similar texts aloud was played halfway through the experiment to remind participants what each speaker sounded like. After the eye-tracking portion of the experiment, participants were asked to complete one social attractiveness survey for each speaker (Zhou & Christianson, 2016, in preparation).

2. Results

We first report the response accuracy for the target sentences and offline survey results. Then, we report the whole sentence reading time data, which is the most important result to address our research question. In addition, we compare whole sentence reading time from the current study to Zhou and Christianson (2016). Linear mixed-effects modeling was applied to analyze the continuous eye movement data, and logit mixed-effects modeling was used to analyze the binomial (response accuracy) data. All analyses were performed using maximal random effects structures, i.e., with random slopes and intercepts for participants and items. Table 1 provides means and standard deviations of sentence reading times and accuracy in all conditions in the current study and in Zhou and Christianson (2016).

2.1. Response accuracy

On average, participants achieved 86.64% comprehension accuracy for the target sentences. The logit mixed-effects model results demonstrated that comprehension did not vary as a function of which speaker's voice was being simulated (*Est.* $= -0.01$; *SE* $= 0.11$; $z = -0.05$, $p = 0.96$). Plausibility (Plausible > Implausible; *Est.* $= 1.36$; *SE* $= 0.18$; $z = 7.75$, $p < 0.001$), structure (SRC > ORC; *Est.* $= 2.2$; *SE* $= 0.20$; $z = 10.74$, $p < 0.001$), and their interaction (SRC in plausible sentences > ORC in implausible sentences; *Est.* $= -1.12$; *SE* $= 0.31$; $z = -3.65$, $p < 0.05$), as well as trial presentation order (Later trials > Earlier trails; *Est.* $= 0.003$; *SE* $= 0.001$; $z = 2.28$, $p < 0.001$) significantly influenced response accuracy. Table 2 presents the details of the LME model for the current study.

2.2. Social attractiveness survey data

Based on the self-report from the surveys, approximately 66.1% of the time, subjects performed auditory perceptual simulation of the American-English speaker's voice and 59.2% of the time, they performed auditory perceptual simulation of the Indian-English speaker's voice. Fourteen variables were used to evaluate and compare readers' attitudes towards the American-English and Indian-English speakers. ANOVA analysis demonstrated that readers' attitudes were more negative towards the Indian-English speaker than the American-English speaker on all variables except honesty ($F(1, 212) = 3.47$; $p = 0.06$) and dependability ($F(1, 212) = 3.68$; $p = 0.06$), where the differences were only marginal. Participants perceived the accent and speech rate differences between the American-English and Indian-English speakers, but only marginally so ($F(1, 212) = 3.68$; $p < 0.1$). They rated the Indian-English speech as significantly less comprehensible ($F(1, 212) = 64.15$; $p < 0.05$), less confident ($F(1, 212) = 140.9$; $p < 0.05$), less intelligent ($F(1, 212) = 51.60$; $p < 0.05$), less pleasant ($F(1, 212) = 67.08$; $p < 0.05$), less likable ($F(1, 212) = 45.69$; $p < 0.05$), and less reliable ($F(1, 212) = 6.57$; $p < 0.05$). They regarded the Indian-English speaker as less conscientious ($F(1, 212) = 14.16$; $p < 0.05$), less educated ($F(1, 212) = 37.14$; $p < 0.05$), less sincere ($F(1, 212) = 140.9$; $p < 0.05$), and less prestigious ($F(1, 212) = 55.58$; $p < 0.05$) than the American-English speaker. Nevertheless, “speaker” was not a significant predictor of comprehension accuracy based on the logistic regression model results above, suggesting that readers' biases towards the non-native speech did not affect their comprehension.

Table 4
Means of recording time for faster and slower speakers across two experiments (min:sec).

	Faster speaker	Slower speaker	Differences
Zhou and Christianson (in preparation)	2:50 (American)	4:14 (Chinese)	1:24
Current Experiment	2:58 (Indian)	3:43 (American)	45

Table 5
Means and standard deviations for sentence reading time (in ms) and accuracy in the combined data (current experiment and Zhou & Christianson, 2016).

"Speaker"	Speed	Structure	Sentence reading time		Accuracy	
			Plausible	Implausible	Plausible	Implausible
Native	Fast	Subject-RC	3825 (1764)	3497 (1644)	0.95 (0.22)	0.96 (0.19)
		Object-RC	4115 (1970)	4313 (1868)	0.70 (0.46)	0.88 (0.32)
	Slow	Subject-RC	4172. (1783)	3842 (1747)	0.94 (0.24)	0.96 (0.20)
		Object-RC	4617 (2101)	4447 (2078)	0.69 (0.46)	0.87 (0.33)
Non-native	Fast	Subject-RC	3962 (1702)	3754 (1546)	0.95 (0.21)	0.96 (0.21)
		Object-RC	4547 (2017)	4198 (1833)	0.68 (0.47)	0.88 (0.32)
	Slow	Subject-RC	3973 (1871)	3681 (1678)	0.92 (0.27)	0.96 (0.20)
		Object-RC	4645 (1702)	4385 (1546)	0.69 (0.46)	0.87 (0.34)

2.3. Whole sentence reading time

Fixations shorter than 80 ms and longer than 1200 ms were trimmed before the data analysis. Reading times that were three standard deviations away from the mean of that condition within each subject were excluded. These trimming procedures resulted in removal of 0.06% of the data. Trimmed sentence reading times were centered in the LME analyses. The LME models for all following analyses included the predictors of "speaker," plausibility, structure, and trial order.

Results reveal that auditory perceptual simulation of the American-English speech led to marginally longer sentence reading time than auditory perceptual simulation of the Indian-English speech ($t = -1.86$; $p = 0.06$; 95% CI $= -0.104$ – 0.002). Plausibility (Implausible > Plausible; $t = -3.12$; $p < 0.001$), structure (ORC > SRC; $t = -4.65$; $p < 0.01$), and trial order (Earlier trials > Later trails; $t = -8.33$; $p < 0.001$) significantly affected sentence reading times. Table 3 presents the fixed effects of the model for whole sentence reading times.

3. Combined data analysis

In the current experiment, we detected a marginally significant difference in the predicted direction between the auditory perceptual simulation of the faster-speaking non-native "speaker" and the slower-speaking native "speaker": auditory perceptual simulation of the faster non-native "speaker" led to faster reading time, despite the non-native accent. We suspect that the marginal ($p = 0.06$) statistical significance may result from a ceiling effect on how fast people can read and perceptually simulate any speech. Alternatively, the difference between slow and fast speech in the current experiment was not quite large enough to yield an effect as clearly as in Zhou and Christianson (2016), in which the difference between speaker rates was larger ($t = -5.08$, df $= 3$, $p < 0.05$; see Table 4 for the means).

To further investigate whether the auditory perceptual simulation effects were driven by accents or speech rates, we compared the current experiment to the second auditory perceptual simulation experiment[4] in Zhou and Christianson (2016), where the pattern of speech rate effects for the native and non-native speakers were opposite to the current study. In this way, the combined dataset not only included the "speaker" condition (American-English vs. Indian-English), but also the "speaker's" speech rate condition (fast vs. slow). The means and standard deviations for each condition ("speaker," speech rate, plausibility, structure) in the combined dataset are presented in Table 5. The same data trimming procedures from the current experiment were applied to the new combined dataset (4.3% of the data were removed). LME models with maximal random effects were built to analyze the results.

The results reveal that speech rates (Fast < Slow; *Est.* $= 0.08$; *SE* $= 0.02$; t = 3.99, <0.001), plausibility (Plausible < Implausible; *Est.* $= -0.14$; *SE* $= 0.03$; $t = -4.\ 59$, < 0.001), syntactic structure (SEC < ORC; *Est.* $= -0.26$; *SE* $= 0.03$; $t = -7.93$, $p < 0.001$), and trial presentation order (Earlier < Later; *Est.* $= -0.002$; *SE* $= 0.0002$; $t = -11.32$, $p < 0.001$), but not "speakers" (*Est.* $= 0.03$; *SE* $= 0.02$; $t = 1.47$, $p > 0.1$), were significant predictors of reading times when readers perceptually simulated either native or non-native speaker voices. Readers read faster when they were simulating the faster "voice," regardless of the familiarity of the accent (see Table 6 for details).

4. Discussion

In this study, we employed an auditory perceptual simulation (APS) paradigm with a faster Indian-English speaker's speech and a slower American-English speaker's speech to investigate whether previously observed effects of auditory perceptual simulation of native and non-native speech derived from speech rate differences between speakers or from difficulty simulating an unfamiliar accent. Although social attractiveness survey results clearly demonstrated that readers had more negative attitudes towards the Indian-English speech, simulation of this (faster) accented speech led to faster silent reading than simulation of the speech of the (slower) American-English speaker. By combining the data from the current experiment and the data from Zhou and Christianson (2016), we found that the auditory perceptual simulation effects on readers' silent reading times were driven by the different speech rates of the voices that were being perceptually simulated, regardless of whether they were native or non-native. Furthermore, accuracy did not differ as a function of which speaker's voice was being simulated. If the online silent reading speeds observed previously (Zhou & Christianson, 2016) and in the present study had been due to difficulty simulating an unfamiliar accent, auditory perceptual simulation of the Indian-English speech in this experiment should have resulted in slower reading times compared to the American-English speech, irrespective of the speech rates. This was not the case. Furthermore, we did not find biases against one type of accent or another to be the source of reading speed differences or comprehension accuracy effects in either study. A sociolinguistic bias

Table 6
Fixed effects of linear mixed-effects model for sentence reading time in the combined dataset.

	Estimate	SE	df	*t* value	*p* value
Intercept	0.35	0.06	235.9	5.57	<0.001
Plausibility	−0.14	0.03	294.2	−4.59	<0.001
Structure	−0.26	0.03	305.3	−7.93	<0.001
"Speaker"	0.03	0.02	175.9	1.47	>0.1
Speech rate	0.08	0.02	179.1	3.99	<0.001
Trial order	−0.002	0.0002	1476	−11.03	<0.001

[4] There were two APS experiments in Zhou and Christianson (2016). We only compared to one of them, because the two APS experiments showed the same APS effects on reading time and comprehension, and they differed only in the APS cue (photo vs. a recording of the speaker's name).

account of the results would predict readers to be more likely to choose the paraphrase of implausible sentences as false when engaging in auditory perceptual simulation of non-native or accented speech (e.g., Lev-Ari & Keysar, 2010). Yet, this was not the case: the online reading speed was in the opposite direction of this hypothesis, and auditory perceptual simulation of Indian-English speech did not lead to lower accuracy than the American-English speech.

Based on this evidence, we argue that when readers perform auditory perceptual simulation during silent reading, they focus on the prosodic characteristics of the voice(s) that they are simulating. The mental representations generated by perceptual simulation my well also contain phonological characteristics, but the relative familiarity of these characteristics do not appear to influence effects of APS on reading speed or comprehension. It is an open question as to what extent phonological characteristics can be perceptually simulated.

As it stands, however, the results reported here are consistent with the proposal by Zhou and Christianson (2016): Auditory perceptual simulation generates a prosodic representation of the text that is more detailed than the default prosodic contour that is generated by most skilled readers when APS is not cued (Fodor, 2002; Rayner, Pollatsek, Ashby, & Clifton, 2012), including speech rate. Prosodic structure correlates with syntactic and information structure (e.g., Breen, Fedorenko, Wagner, & Gibson, 2010). When a rich prosodic structure is generated via auditory perceptual simulation during reading, it “buttresses” the syntactic structure, which consequently is more likely to be maintained more robustly against intrusions from competing “good-enough” interpretations derived from non-structural heuristics.

In conclusion, the evidence presented here strongly suggests that a central aspect of the perceptual simulation of speech during reading is the speech rate of the speaker whose voice is being simulated. Differences in simulated speech rates in turn drive the reading speed differences observed in previous work and here, rather than the relative difficulty of simulating unfamiliar accents.

Acknowledgements

We appreciate the help from the members of the EdPsych Psycholinguistics Lab for data collection at UIUC and the suggestions from Susan Garnsey, Duane Watson, Shravan Vashishth and the reviewer, Celia Klin, on the project. This work was supported by Beckman Institute Cognitive Science/Artificial Intelligence Award to Peiyun Zhou.

References

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 446–459.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language & Cognitive Processes*, *25*, 1044–1098.

Brennan, E. M., & Brennan, J. S. (1981). Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language and Speech*, *24*(3), 207–221.

Callen, V., Callois, C., & Forbes, P. (1983). Evaluation to accented English. *Journal of Cross-Cultural Psychology*, *14*, 407–426.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Underspecified, shallow, or Good Enough language processing. *Quarterly Journal of Experimental Psychology*, *69*, 817–828.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 538–544.

Drumm, A. M., & Klin, C. M. (2011). When story characters communicate: Readers'representations of characters' linguistic exchanges. *Memory & Cognition*, *39*, 1348–1357.

Edwards, J. R. (1977). Students’ reaction to Irish regional accent. *Language & Speech*, *20*, 280–286.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.

Ferreira, F., & Patson, N. D. (2007). The ‘Good Enough' approach to language comprehension. *Language and Linguistics Compass*, *1*, 71–83.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*, 3–20.

Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's regional accent. *PloS One*, *6*, e25782. http://dx.doi.org/10.1371/journal.pone.0025782.

Fodor, J. D. (2002, Aprill). Psycholinguistics cannot escape prosody. *Paper presented at the meeting of Speech Prosody, Aix-en-Provence, France*.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, *34*(1), 65–87.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy and relative clause comprehension. *Journal of Memory and Language*, *58*, 161–187.

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, *16*(2), 313–353.

Giles, H. (1972). Evaluation of personality content from accented speech as a function of listeners’ social attitudes. *Perceptual and Motor Skills*, *34*, 168–170. http://dx.doi.org/10.2466/pms.1972.34.1.168.

Giles, H., & Watson, B. (2013). *The social meanings of language, dialect and accent*. New York, NY: Peter Lang.

Giles, H., Hewstone, M., Ryan, E. B., & Johnson, P. (1987). Research on language attitudes. *Sociolinguistics: An international handbook of the science of language and society*, *1*, 585–597.

Gluszek, A., & Hansen, K. (2013). Language attitudes in the Americas. *The social meanings of language, dialect and accent. International perspectives on speech styles*, 26–44.

Gunraj, D. N., & Klin, C. M. (2012). Hearing story characters' voices: Auditory imagery during reading. *Discourse Processes*, *49*(2), 137–153.

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, *136*, 302–329.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children’s social preferences. *Social cognition*, *27*(4), 623.

Kosslyn, S. M., & Matt, M. C. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society*, *9*, 250–252.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition*, *112*, 457–461.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*, 1093–1096.

Levine, W. H., & Klin, C. M. (2001). Tracking of spatial information in narratives. *Memory & Cognition*, *29*, 327–335.

Lim, J. -H., & Christianson, K. (2013a). Integrating meaning and structure in L1-L2 and L2-L1 translations. *Second Language Acquisition*, *29*, 233–256.

Lim, J. -H., & Christianson, K. (2013b). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*, 518–537.

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

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, *45*(1), 73–97.

Munro, M. J., & Derwing, T. M. (1998). The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech. *Language Learning*, *48*(2), 159–182.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of reading* (2nd ed.). New York: Psychology Press.

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker reactions to nonnative speech. *Studies in Second Language Acquisition*, *5*, 148–159.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437–442.

Stites, M., Luke, S. G., & Christianson, K. (2013). The psychologist said quickly, “dialogue descriptions modulate reading speed!”. *Memory and Cognition*, *41*, 137–151.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: evidence from eye movements. *Journal of Memory and Language*, *47*, 69–90.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits andrules*. Cambridge, MA: MIT Press.

Varonis, E. M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in second language acquisition*, *4*(02), 114–136.

White, M. J., & Li, Y. (1991). Second-language fluency and person perception in China and the United States. *Journal of Language and Social Psychology*, *10*, 99–113.

Woumans, E., Martin, C. D., Bulcke, C. V., Van Assche, E., Costa, A., Hartsuiker, R. J., & Duyck, W. (2015). Can faces prime a language? *Psychological Science*, *26*(9), 1343–1352.

Woumans, E., Martin, C., Vanden Bulcke, C., Van Assche, E., Costa, A., Hartsuiker, R., & Duyck, W. (2015). Can faces prime a language? *Psychological Science*.

Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, *121*, 447–453.

Zhou, P., & Christianson, K. (2016). I “hear” what you're “saying”. Auditory perceptual simulation, reading speed, and comprehension. *Quarterly Journal of Experimental Psychology*, *69*, 972–995.

Zhou, P., & Christianson, K. (2014). *Comparison of attitudes towards native and non-native English speech*. San Francisco, CA: Poster presented at 26th APS Annual Convention.