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 10 May 2016

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
0001-6918/

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