Published in final edited form as:
Cognition. 2009 September ; 112(3): 457–461. doi:10.1016/j.cognition.2009.05.007.

Those Voices in your Head: Activation of Auditory Images During Reading

Christopher A. Kurby,
Washington University, St. Louis

Joseph P. Magliano, and
Northern Illinois University

David N. Rapp
Northwestern University

Abstract

Auditory imagery experiences (AIEs) occur when readers simulate character voices while reading. This project assessed how familiarity with voice and narrative contexts influences activation of AIEs. Participants listened to dialogs between two characters. Participants then read scripts with the characters, half that had been previously listened to and half that were new. During reading, participants were interrupted with an auditory recognition task, with probes presented in voices that either matched or mismatched the character associated with the current line of dialog. Faster responses to matching than mismatching voices were consistently obtained for familiar scripts, providing evidence for AIEs. Transfer to unfamiliar scripts only occurred after extended experience with character voices. These findings define factors that influence activation of speaker voice during reading, with implications for understanding the nature of linguistic representations across presentation modalities.

Imagine reading Darth Vader proclaim, “Luke! I am your father” and hearing his voice while doing so. *Auditory imagery experiences* (AIEs) occur when a reader simulates speaker voice while reading text. Readers report these types of phenomenological experiences to be easy to enact and routine (e.g., Alexander & Nygaard, 2008). But while AIEs seem commonplace, little research has examined the scope of their etiology. How do readers establish voices for story characters, and how do those voices influence text processing?

Existing theories provide explanatory accounts for any possible AIE effects, emerging as a function of debates between two major accounts of linguistic representations. The *abstractionist* view argues that speech is stripped of its indexical features (e.g., rate, prosody, pitch) in a normalization process which results in a modality-free linguistic representation of speech input (see Tenpenny, 1995 for a review). Thus, the AIEs that are constructed during reading have no basis in the activation of indexical knowledge stored in a memory representation for linguistic content. In contrast, the *multicode* view contends that linguistic representations are episodic in nature, preserving surface perceptual features of speech (Goldinger, 1998; Nygaard, 2005).

Send Correspondence to: Christopher A. Kurby, Department of Psychology, Washington University, 1 Brookings Drive, Saint Louis, MO 63130, Telephone: 314-935-4138, E-mail: ckurby@artsci.wustl.edu.

There is growing support for the multicode view. Listeners appear to encode non-linguistic features of speakers' utterances including rate (Kosslyn & Matt, 1977), gender (Monsen & Engebretson, 1977), and emotion (e.g., Scherer, Banse, Wallbott, & Goldbeck, 1991). Both real and imagined perceptual features of speakers influence speech perception (e.g., Goldinger, 1998), speech recognition (e.g., Brown & Carr, 1993; Nygaard & Pisoni, 1998), lexical decision times (Abramson & Goldinger, 1997), and memory for speech (Pilotti, Gallo, & Roediger, 2000).

Current work on knowledge representation and discourse processing also supports the notion that non-linguistic, perceptual features of text experiences (e.g., visual, auditory, etc) may be encoded into representations that influence comprehension (Barsalou, 1999, 2008; Zwaan, 2004). A similar claim might be made about the perceptual features of textual speech - readers should experience AIEs during comprehension if perceptual representations of the character's voice are present in memory to support such imagery. There is some evidence consistent with this hypothesis. Specifically, reading rates are slower for texts "written" by slower talking than faster talking speakers (Alexander & Nygaard, 2008). However, this finding does not speak to conditions that might give rise to the activation of AIEs during reading.

Readers' specific experiences with voices may establish expectations for how speakers will sound. For instance, after viewing Sean Connery as James Bond, readers might imagine the lead character of Ian Fleming's novels as speaking with Connery's voice[1]. This suggests that the ways in which readers are initially familiarized with character voice may influence the voice-specific nature of AIEs. Previous investigations of AIEs have familiarized participants with voices through individual word/voice pairings, or with brief conversations providing little insight into interlocutors' motivations or personalities. However, our everyday experiences with voices often involves extended, context-bound dialogue that encodes characters' voices as well as their goals and relationships in various situations. Understanding how, in these latter cases, familiarity influences reader simulation of voices proves crucial for outlining the ways in which AIEs are informed by naturalistic experiences, and for determining the features of linguistic events that support or enhance the activation of AIEs.

The present study investigated conditions that influence the activation of character voice during reading. At issue is whether experience with a character's voice produces a representation sufficiently robust to afford mental simulation of voice in even novel contexts. Given the possibility that perceptual representations stored in memory are situated to the learning situation (Glenberg, 1997), the experience of AIEs may only occur when the reading situation strongly overlaps with the learning episode. Another possibility is that perceptual features of character voice are applied in any situation involving those characters.

Experiment 1

In Experiment 1, we manipulated participants' familiarity with story scripts. During a listening phase, participants heard enactments of 12 script excerpts; during a reading phase they read these scripts and 12 new ones. In the reading phase, participants read text-versions of the scripts and were occasionally interrupted by an auditory word recognition task. The probes in this task were presented in voices that matched or mismatched the voice of the currently speaking character. Faster responses to matching compared to mismatching voices would provide evidence suggesting that voices were active (i.e., simulated) during reading.

[1]It is worth noting that Fleming himself doubted Connery's aptitude for the role of 007, but after viewing his performance, added a partially-Scottish background to the character, further supporting that particular voice for Bond.

Participants

Sixty-eight native English-speaking undergraduates from Northern Illinois University participated for course credit. None reported a history of hearing loss.

Materials

Scripts—Twenty-four excerpts from the 1950's radio show, *The Bickersons*, were selected from an anthology (Rapp, P., 2004) and edited for length. This show was chosen because it was unlikely that participants had heard the scripts or knew of the show prior to the experiment. Each excerpt detailed a conversation between John and Blanche Bickerson, a married couple typically engaged in a comical dispute over relationship issues. Scripts included alternating lines of dialog between John and Blanche (M = 19.75 lines in length, SD = 4.36), with each character producing from one to six sentences (M = 1.67, SD = 0.93) per line (M = 11.37 words per line, SD = 1.72). Two amateur actors recorded versions of the scripts using Audacity sound editing software (Mazzoni, 2005); the average script length was 69.34 seconds (SD = 15.99).

Probes

Critical probes: Twenty-four critical probe words were selected from the scripts; these probes were always concrete nouns from the second half of the scripts (to encourage reading of the entire script). Each probe word was explicitly mentioned in a particular line, and thus required a "yes" response in the probe recognition task. The actors playing John and Blanche each recorded audio versions of all 24 critical probes. The average probe length was 723 milliseconds (SD = 144).

Filler probes: Seventy-two filler probe words were selected, all concrete nouns. Twenty-four were words explicitly mentioned in a particular line, and thus required a "yes" response in the probe task. These were selected so that they were distributed across the length of the scripts. Forty-eight filler probes were included that did not appear in the scripts, requiring a "no" response. Filler probes did not differ from experimental probes on number of syllables, or Kucera and Francis (1967) word frequency measures (all ps > .23). Three males and three females (excluding the John and Blanche voices) each recorded 12 of the 72 filler probes, such that half of the fillers were male and half were female voices, crossed with filler probe correctness (yes or no response). The filler probes were distributed across scripts to ensure an equal number of gender matches and mismatches between probe voices and script characters. There was an average of 3 filler probes per script. Two practice scripts and probes were also constructed. Table 1 presents an example script and probe items.

Finally, a sheet of math problems was constructed to reduce participants' rehearsal of script content between the listening and reading phases.

Design

Experiment 1 employed a 2 (script familiarity: familiar vs. new) × 2 (character match: match vs. mismatch) repeated-measures design. Participants read 24 scripts, 12 previously presented in the listening phase and 12 that were new; half of the experimental probes were in the voice of the character that originally stated it, and half in the other character's voice. Four lists were constructed to counterbalance the assignment of scripts and probes to conditions. Participants were each randomly assigned to one of the counterbalanced lists.

Procedure

In phase 1, participants listened to the scripts over headphones; they were asked to listen carefully to answer subsequent questions about the scripts. Participants began with a practice script to get accustomed to the task and adjust the volume if necessary. After practice,

participants listened to 12 scripts selected from one of the counterbalanced lists; participants pressed the spacebar to begin each script. Each script played continuously until it completed. The order of script presentation was randomly determined for each participant. After this phase, participants solved math problems for five minutes.

In phase 2, participants read two practice scripts and all 24 experimental scripts on the computer screen. (One of the practice scripts had been presented during practice in phase 1.) Participants pressed the spacebar to begin each script and advance through them, one line at a time. Each line began with the name of the character speaking the line, followed by a colon and their dialogue. For probe trials, following a spacebar press, the screen was erased and a probe word was presented aurally. Participants indicated whether the probe had been mentioned in the line they just read. Participants pressed the j-key ("yes") if the word had been mentioned, and the f-key ("no") if it had not. After their judgment the next line in the script appeared. The order of script presentation was randomly determined for each participant. For half of the scripts, participants were prompted to generate a one-sentence summary on a blank piece of paper after reading the script; this ensured participants were reading for comprehension.

Results and Discussion

Eight participants were removed from the analysis because their reading times sped up at an increasing rate as they progressed through the scripts (which is indicative of non-compliance); we also removed 0.9% of the trials for reading times less than 50 ms per word, and a further 2.0% for response times 3 standard deviations above the mean in each condition. Mean reaction times (see Table 2) were computed using only correct responses. Analyses were conducted with participants (F_1) and items (F_2) as random variables using repeated measures ANOVAs.

Participants were faster to correctly identify probes in match (M = 1001, SD = 157) than mismatch conditions (M = 1034, SD = 175), significant by participants only, $F_1(1, 59) = 4.54$, MSE = 9661.88, p = .04, partial η^2= .07, $F_2(1, 23) = 2.30$, MSE = 8108.12, p = .14, partial η^2= .09. This was qualified by a significant interaction between script familiarity and character match, $F_1(1, 59) = 8.14$, MSE = 19420.63, p = .01, partial η^2= .12, $F_2(1, 23) = 5.89$, MSE = 8342.53, p = .02, partial η^2= .20. Follow-up t-tests revealed that responses were significantly faster for match than mismatch cases only with familiar scripts, $t_1(59) = 3.28$, SD = 185, p = .001, d = .35, $t_2(23) = 2.32$, SD = 155, p = .03, d = .35, with no difference observed for new scripts, ps > .22. No effects of accuracy were observed (all ps > .06).

In Experiment 1, AIEs were obtained when texts conveyed the same situations participants had already listened to, with little evidence for AIEs for new scripts. However, exposure to the voices may not have produced strong enough memory representations to allow for transfer to new scripts. The goal of Experiment 2 was to assess whether repeated exposure to scripts during the listening phase would strengthen the memory representation for the voices and encourage such transfer.

Experiment 2

Participants

One hundred eighteen native English-speaking undergraduates from Northern Illinois University participated for course credit. None reported a history of hearing loss.

Materials

The materials were identical to Experiment 1.

Design and Procedure

The design and procedure were identical to Experiment 1 with the following changes. In a preceding session, participants listened to 12 scripts. Two days later, the participants listened to the same 12 scripts again, performed the math task, and then read the full list of 24 scripts. Of these 24 scripts, 12 were listened to twice and 12 were new. Assignment of scripts to familiar and new conditions was counterbalanced.

Results and Discussion

Nine participants were removed for non-compliance; we also removed 1.3% of the trials for reading times less than 50 ms per word, and a further 1.6% of the trials for reaction time outliers. Analyses revealed a main effect of script familiarity (See Table 3): Response times to correctly identify probes were faster for familiar (M = 978, SD = 185) than new scripts (M = 1024, SD = 253), $F_1(1, 108) = 9.37$, MSE = 23664.49, p = .003, partial η^2= .08, $F_2(1, 23) = 5.54$, MSE = 4525.84, p = .03, partial η^2= .13. A main effect of character match was obtained: Participants were faster to correctly identify probes in match (M =978, SD = 204) compared to mismatch conditions (M = 1024, SD = 228), $F_1(1, 108) = 4.65$, MSE = 14493.98, $p < .001$, partial η^2= .13, $F_2(1, 23) = 3.35$, MSE = 9775.23, p = .08, partial η^2= .20.[2] In contrast to Experiment 1, the script familiarity x character match interaction was not significant (all $ps > .24$). Follow up tests revealed that response times were indeed faster for the match than mismatch condition for new scripts, $t_1(108) = 2.06$, SD = 171, p = .04, d = .14, although this effect was not significant by items, $t_2(23) = 0.67$, SD = 133, p = .51, d = .132. No effects of accuracy were observed (all $ps > .20$).

If readers transferred character voice to new scripts, the results should reveal modulation of the match effect in this condition between the two experiments. We conducted an ANOVA using Experiment as a factor, which revealed a 3-way interaction between Experiment, Familiarity, and Match, marginally significant by-subject, $F_1(1, 167) = 3.00$, MSE = 19205.53, p = .085, partial η^2= .02, but not significant by-item, $F_2(1, 23) = 1.42$, MSE = 5913.89, p = .245, partial η^2= .06. However, given the dramatic change in the direction of the match effect for new scripts from Experiment 1 (M = −24 ms) to Experiment 2 (M = +33 ms), we conducted a 2 (Match) × 2 (Experiment) ANOVA for new scripts only. This revealed a 2-way interaction significant by-subjects, $F_1(1, 167) = 4.76$, MSE = 13691.32, p = .03, partial η^2= .03, but not by-items $F_2(1, 23) = 1.17$, MSE = 6422.71, p = .292, partial η^2= .05. These results provide at least partial evidence that, with sufficient familiarity, perceptually-based knowledge of character voice obtained from particular narratives may be activated for new narratives (albeit this conclusion must be tempered by the item analyses).

These data rule out alternative explanations for the results. One possibility is that the match-mismatch paradigm might be subject to demand characteristics that encouraged participants to strategically evoke images of characters' voices. In addition, readers may have activated relatively generic images associated with speaker gender, rather than representations associated with the specific scripts that were heard. However, the fact that we did not find transfer without repeated exposures tempers both alternatives. If participants strategically evoked character voice, this should have occurred with new scripts in Experiment 1. Similarly, if participants activated general information regarding gender, match-mismatch effects should have been obtained in all conditions across the two experiments.

[2]Given that the effect size was essentially identical to the by-subject test, as revealed by Cohen's d, the nonsignificance of this by-items test is likely due to a relatively small effect size coupled with power constraints from the restricted number of items.

General Discussion

Although phenomenological reports of AIEs during reading may be ubiquitous, they are surprisingly understudied. These experiments investigated the extent to which readers enact AIEs during reading. In Experiment 1, AIEs were obtained for familiar scripts; in Experiment 2, extended experience with character voice resulted in AIEs for familiar and new scripts. Importantly, AIEs were observed in conditions involving silent reading (e.g., Alexander & Nygaard, 2008), as reading aloud can explicitly emphasize the perceptual features of speakers' voices (e.g., Kosslyn & Matt, 1977). The findings serve as a type of existence proof as well as an analysis of conditions that encourage the mental simulation of speaker voice, without unnecessarily drawing attention to such activity.

These data also speak to whether mental representations for linguistic input are abstract or multicode in nature. Some models of language perception outline an abstraction process by which linguistic input is stripped of perceptual qualities to extract the conceptual content of a message. However, a strong version of this argument currently appears untenable: Perceptual information (e.g., prosody) influences processing of linguistic content, which suggests mental representations can preserve context-specific, surface characteristics of speech (e.g., Goldinger, 1998). The current results are consistent with this view, and the view that readers activate perceptually-based knowledge during reading (Barsalou, 1999, 2008; Zwaan, 2004). However, the present study suggests that the activation of perceptual representations of voice is not likely without direct perceptual experience of that voice in context. In addition, perceptual features can be brought to bear in even unfamiliar situations if they are highly learned (i.e., following increased exposure). While linguistic representations require flexibility to allow for successful application to novel conditions, such flexibility is traditionally associated with abstract representations. The data presented here suggest that such flexibility can sustain even the perceptual features of speaker voice.

Readers' experiences of AIEs, and the quality of those AIEs, has previously been associated with factors relevant to the effects reported here. Alexander and Nygaard (2008) argued that attention to linguistic material may encourage encoding of voice information, enhancing the likelihood of AIEs. The current materials may have enhanced attention in at least two ways. First, they were written by professional scriptwriters to be humorous, which may have increased engagement with the scripts. Secondly, the scripts included puns and clever turns of phrase. These contents likely encourage attention to the conversations and indexical features of the texts to ensure successful comprehension of the jokes. Investigating the degree to which AIEs occur with more mundane materials, and the resulting quality of those representations, will provide a useful test of their scope and influence.

During our everyday entertainments, individuals learn about voices by watching television and films, listening to the radio, and interacting with others. Sometimes, material in one media format might connect with material in other formats involving similar characters and events. For example, the voice of Harry Potter, as provided by the actor in the film or the narrator of the audiobook, could be invoked during readings of Potter's adventures. AIEs offer an intriguing way to study not just the ways in which readers process texts, but generally, the multimodal nature of memory representations. Thus, AIEs may thus help us understand how diverse sensory percepts combine from multiple sources and influence our everyday experiences of language.

References

Abramson M, Goldinger SD. What the reader's eye tells the mind's ear: Silent reading activates inner speech. Perception & Psychophysics 1997;59:1059–1068. [PubMed: 9360478]

Alexander JD, Nygaard LC. Reading voices and hearing text: Talker-specific auditory imagery in reading. Journal of Experimental Psychology: Human Perception and Performance 2008;34:446–459. [PubMed: 18377181]

Barsalou LW. Grounded cognition. Annual Review of Psychology 2008;59:617–645.

Barsalou LW. Perceptual symbol systems. Behavioral and Brain Sciences 1999;22:577–660. [PubMed: 11301525]

Brown J, Carr T. Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. Journal of Experimental Psychology: Learning, Memory, & Cognition 1993;19:1277–1296.

Geiselman RE, Glenny J. Effects of imagining speakers' voices on the retention of words presented visually. Memory & Cognition 1977;5:499–504.

Glenberg AM. What memory is for. Behavioral and Brain Sciences 1997;20:1–55. [PubMed: 10096994]

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review 1998;105:251–279. [PubMed: 9577239]

Kosslyn SM, Matt MC. If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. Bulletin of the Psychonomic Society 1977;9:250–252.

Kucera, H.; Francis, WN. Computational Analysis of Present-day American English. Providence, R.I: Brown University press; 1967.

Mazzoni, D. Audacity (1.2.1). Boston, MA: Free Software Foundation, Inc; 2005 [Retrieved October 01, 2005]. Available from http://audacity.sourceforge.net/

Monsen RB, Engebretson AM. Study of variations in the male and female glottal wave. Journal of the Acoustical Society of America 1977;62:981–993. [PubMed: 911405]

Nygaard LC, Pisoni DB. Talker-specific perceptual learning in spoken word recognition. Perception & Psychophysics 1998;60:355–376. [PubMed: 9599989]

Pilotti M, Gallo DA, Roediger HL. Effects of hearing words, imagining hearing words, and reading on auditory implicit memory tests. Memory & Cognition 2000;28:1406–1418.

Rapp, P. The Bickersons scripts. Vol. 2. BearManor Media; Boalsburg, PA: 2004.

Scherer KR, Banse R, Wallbott HG, Goldbeck T. Vocal cues in emotion encoding and decoding. Motivation and Emotions 1991;15:123–148.

Tenpenny PL. Abstractionist versus episodic theories of repetition priming and word identification. Psychonomic Bulletin & Review 1995;2:339–363.

Table 1

Example excerpt of a script from P. Rapp (2004, p., 72–73), and probe words.

Character	Line	Probe Word
John:	You got any money?	
Blanche:	There's fifty cents in the sugar bowl.	cup (f)
John:	Fifty cents!	
Blanche:	You can bring me the change when you get home.	change (f)
John:	Now listen, Blanche-something's gotta be done about this. I can't go down to work like a pauper every day. A man's got to have a couple of dollars in his pocket.	wallet (f)
Blanche:	Well, don't yell at me.	
John:	I'm not gonna suffer through those lunches anymore.	
Blanche:	What's the matter with your lunches?	
John:	You ought to know-you pack 'em for me. I'm just getting sick of carrying my lunch to work in a paper sack. Why can't I go to a restaurant like the other fellows-	restaurant (f)
Blanche:	John! What are you talking about? I haven't packed your lunches in two years!	
John:	Oh, Blanche! Every morning of my life I find my lunch wrapped in brown paper on the side of the sink!	
Blanche:	Lunch? That's the garbage!	garbage (c)

Note: The letter next to each probe word denotes whether it was a filler (f) or critical (c) item. The probe word was presented after each participant completed reading the entire line.

Table 2

Mean response times and accuracies for Experiment 1 (standard deviations in parentheses).

Script Familiarity		Match	Mismatch
New	RT	1014 (163)	990 (175)
	Acc	.93 (.12)	.96 (.10)
Familiar	RT	999 (191)	1077 (249)
	Acc	.93 (.12)	.91 (.12)

Table 3

Mean response times and accuracies for Experiment 2 (standard deviations in parentheses).

Script Familiarity		Match	Mismatch
New	RT	1007 (256)	1040 (278)
	Acc	.93 (.13)	.93 (.16)
Familiar	RT	949 (184)	1008 (231)
	Acc	.94 (.12)	.94 (.12)