Supplement Article

Using Virtual Technology to Promote Functional Communication in Aphasia: Preliminary Evidence From Interactive Dialogues With Human and Virtual Clinicians

Michelene Kalinyak-Fliszar, Nadine Martin, Emily Keshner, Alex Rudnicky, Justin Shi, and Gregory Teodoro

Purpose: We investigated the feasibility of using a virtual clinician (VC) to promote functional communication abilities of persons with aphasia (PWAs). We aimed to determine whether the quantity and quality of verbal output in dialogues with a VC would be the same or greater than those with a human clinician (HC).

Method: Four PWAs practiced dialogues for 2 sessions each with a HC and VC. Dialogues from before and after practice were transcribed and analyzed for content. We

compared measures taken before and after practice in the VC and HC conditions.

Results: Results were mixed. Participants either produced more verbal output with the VC or showed no difference on this measure between the VC and HC conditions. Participants also showed some improvement in postpractice narratives. **Conclusion:** Results provide support for the feasibility and applicability of virtual technology to real-life communication contexts to improve functional communication in PWAs.

n recent years, advances in aphasia rehabilitation science reflect a new emphasis on evidence-based practice. This focus on accountability for treatment efficacy and treatment outcomes has shaped priorities for clinical practice and research protocols. The result has been a surge of highly specific and effective direct treatments of all aspects of language comprehension and production. In addition, in keeping with the paramount goal of treatment, generalization of learned language skills to functional communication situations, there has been a growing interest in developing better approaches to treating functional communication in aphasia. Holland, Fromm, DeRuyter, and Stein (1996) and others (Kagan, Simmons-Mackie et al., 2008; Martin, Thompson, & Worrall, 2007) emphasize the importance of treating the consequences of aphasia, not just the impairment alone. Advocates of this approach use the World Health Organization (WHO) model as a framework for guiding intervention approaches for persons with

aphasia (PWAs). In the study reported here, we examined the efficacy of a dialogue practice that aims to promote functional communication abilities in aphasia.

Given that more people are living longer lives with aphasia, treatment of functional communication abilities takes on new significance. PWAs reenter their community after their rehabilitation program is ended. Thus, it is incumbent on rehabilitation specialists to incorporate training in using residual language skills for functional communication (Aten, Cligiuri, & Holland, 1982). Evidence indicates that language abilities improve with continued treatment, even during chronic stages of aphasia (Raymer et al., 2008). For optimal generalization, PWAs need to practice language in everyday living situations.

The growing interest in promoting functional communication abilities of PWAs is exciting. At the same time, efforts to develop functional communication treatments are faced with the need for methods to demonstrate their efficacy (Aten et al., 1982). Direct treatments can be structured in ways that allow quantitative measures of improvement on clinical tasks and generalization to novel conditions in the clinic. These measures can be supplemented with qualitative measures of generalization to contrived functional communication situations. At present, this kind of clinical activity is carried out through role-playing with the clinician

Correspondence to Michelene Kalinyak-Fliszar: kalinyak@temple.edu

Editor: Anastasia Raymer Associate Editor: Mary Purdy Received September 16, 2014 Revision received February 27, 2015

Accepted July 9, 2015

DOI: 10.1044/2015_AJSLP-14-0160

Disclosure: The authors have declared that no competing interests existed at the time of publication.

^aTemple University, Philadelphia, PA

^bCarnegie Mellon University, Pittsburgh, PA

(Ahlsén, 2005; Holland, Frattali, & Fromm, 1998). Patientreported outcomes have been one way to quantify generalization of treatment gains to real-life situations (Cherney, Oehring, Whipple, & Rubenstein, 2010; Doyle et al., 2004; Kagan et al., 2008).

Role-playing (Holland, Frattali, & Fromm, 1999) and script training (Youmans, Holland, Munoz, & Bourgeois, 2005) have traditionally been used to improve functional communication in PWAs. Script training was developed to facilitate automatic production of spoken language in functional communication contexts. The content of the scripts, typically three to four personally relevant sentences, is generated and written by clinician and client. Drill and verbatim practice are used to facilitate automatic production of script content to improve functional communication during conversational contexts (Youmans et al., 2005). The fundamental characteristics of script training—mass practice and drillmake it particularly adaptable to virtual technology. Cherney and colleagues (Cherney, Halper, Holland, & Cole, 2008; Cherney, Halper, Holland, Lee, Babbitt, & Cole, 2007; Mannheim, Halper, & Cherney, 2009) developed a computer treatment software program, AphasiaScripts, which uses a virtual clinician (VC) for the script training. The VC, programmed to produce natural-sounding speech with correct movement of the articulators, provides hierarchical support for conversational practice (from choral responding of script sentences to simulation of the scripted conversation without cues). The unique aspect provided by the VC is that it replaces the human clinician (HC) and affords the opportunity for intensive home-based therapeutic interventions. Another important characteristic of this technology is its potential to support the successful carry-over of residual language skills to functional language abilities in everyday communication settings (Cherney et al., 2008; Garcia, Rebolledo, Metthé, & Lefebvre, 2007; Lee, Kaye, & Cherney, 2009). Last, computer-based script treatment mediated by a VC has the potential to be a cost-effective method to provide aphasia rehabilitation. Despite the current advances, the application of virtual technology for rehabilitation of aphasia is an enterprise still in its early stages.

In the present study, we add to the growing body of evidence of the feasibility of virtual technology for aphasia rehabilitation by extending the approach used in Aphasia-Scripts (Cherney et al., 2007) to conversational contexts during which the VC is the only member of the dyad who is scripted (i.e., programmed with a set of utterances, comments, and responses appropriate to the conversational situation). The PWA does not produce utterances practiced from a set of scripted sentences, as in AphasiaScripts, but generates unrehearsed replies, questions, and comments. One of our long-term goals is to develop a dialogue practice tool with a VC and virtual simulations of activities of daily living that PWA can use to help maximize and potentially improve their residual language abilities.

Although the potential of using VCs is promising, it is important to determine whether PWAs (or other language disorder) will be responsive to the VC and produce as much language in this context as they would during dialogues with HCs. Research comparing human–nonhuman interaction to human-human interaction (e.g., Fogg & Nass, 1997; Mayer, Johnson, Shaw, & Sandhu, 2006; Nass, Moon, Fogg, Reeves, & Dryer, 1995; Turkle, Taggart, Kidd, & Dasté, 2006) has shown that humans interact with nonhumans (e.g., avatars, robots, pets) in much the same way as they do with other humans. Taggart, Turkle, and Kidd (2005), for example, observed interactions of elderly individuals with robots and found that they exhibited social reactions to the robots that reflected a sense of companionship. Evidence from aphasia comes from a study reported by Cherney (2010) on the efficacy of a treatment protocol, Oral Reading for Language in Aphasia (ORLA), which was administered to 25 individuals with nonfluent aphasia by an HC and, in a second condition, delivered by a computer program. No significant difference was observed in the outcomes of the two conditions.

The present study focuses on the feasibility of using virtual technology as a rehabilitation tool for aphasia. It extends the work of Cherney and colleagues (Cherney et al., 2008; Lee et al., 2009) to conversational situations during which the VC is scripted to elicit unscripted utterances from a group of PWAs. In addition, it serves as the foundation for the development of a VC-human interaction system that can be used independently by PWAs to practice and improve communication skills. This involves the development of software that will support a spoken dialogue system (SDS) that can interact autonomously with an individual and can be configured to personalize treatment (Teodoro, Martin, Keshner, Shi, & Rudnicky, 2013). A critical first step in the development of an SDS for PWAs is to determine how interactive dialogues with a VC compare to interactive dialogues with an HC. In the present study, we do this by evaluating interactive dialogues obtained from four PWAs who participated in four practice sessions, two with an HC and two with a VC.

Our primary objective for the group of PWAs was to compare the quantity and quality of information conveyed in interactive dialogues obtained during four practice sessions, two with an HC and two with a VC. Our secondary objective was to compare the quantity and quality of information conveyed in dialogues elicited from both clinician conditions but only in pre- and postpractice testing. Last, as a measure of generalization, we administered the Nicholas and Brookshire (1993) narratives to evaluate discourse production from pre- to postpractice testing.

Method

Participants

History

Four speakers with aphasia, three men and one woman (identified by initials: CM, CN, DC, and EH), participated in this study. Participant characteristics are reported in Table 1.

Participants were at least 12 months postonset of a left middle cerebral artery cerebrovascular accident (CVA).

Table 1. Participant characteristics.

Characteristic	EH	СМ	CN	DC
Age at enrollment	49	51	49	51
Gender	Woman	Man	Man	Man
Etiology	L MCA CVA	L Parietal, basal ganglia and cerebellum	L MCA CVA	L MCA CVA, craniectomy with temporal lobectomy
Marital status	Widowed	Single	Single	Divorced
Handedness	Right	Right	Right	Right
MPO	103	63 [°]	12	44
Years of education	13	<12	<12	>12
Former occupation	Mortgage processor	Machinist	Hotel concierge	Computer technician

Note. L MCA CVA = left middle cerebral artery cerebrovascular accident; MPO = months post-onset.

Age at the time of enrollment ranged from 49 to 51 years (M = 50 years). Participants were right-handed and had at least some high school education. All were employed at the time of their stroke. All participants passed a pure-tone air audiometric screening at 25 dB HL for 1000, 2000, and 4000 Hz in at least one ear. Prior to enrollment, participants provided written informed consent as required by the Institutional Review Board of Temple University.

Language Evaluation

All participants had been previously enrolled in studies conducted in the Aphasia Rehabilitation Research Laboratory and had completed extensive standardized and laboratory-developed testing. The results of the most current administration of the Western Aphasia Battery-Revised (WAB-R; Kertesz, 2006) are reported in Table 2.

CM achieved an aphasia quotient (AQ) on the WAB-R (Kertesz, 2006) of 89.3, consistent with an anomic aphasia. CM's verbal output consisted mostly of complete, relevant sentences with occasional phonemic paraphasias. In addition, speech intelligibility was mildly compromised by an ataxic dysarthria.

EH presented with anomic aphasia. Her WAB-R AQ was 79.4. Verbal output was characterized by short subject-

verb-object utterances. Also, EH presented with an apraxia of speech. Phoneme distortions, vowel prolongations, and prosodic disruptions (slow rate, misassignment of stress) were observed during attempts to produce multisyllabic words and longer utterances.

CN presented with Broca's aphasia. His WAB-R AQ was 67.0. Verbal output was typically agrammatic, mainly consisting of nouns and occasional simple phrases (e.g., "sandals on"). Verbal output was compromised by a severe apraxia of speech characterized by distortions of on-target phonemes and phoneme substitutions in single words.

DC's language profile was consistent with a transcortical motor aphasia. His AQ on the WAB-R (Kertesz, 2006) was 57.7. Repetition was relatively spared. However, in spontaneous speech, verbal output was agrammatic, consisting of nouns, short phrases, and occasional subjectverb-object utterances and characterized by frequent, unsuccessful attempts to revise and reformulate utterances.

Prepractice Testing

The Nicholas and Brookshire (1993) narratives were administered to measure content (content information units; CIUs) and efficiency of content conveyed (CIUs/minute) in narratives. In brief, narratives consisted of 10 stimuli that

Table 2. Performance on Western Aphasia Battery-Revised (WAB-R).

CM CN	DC
(.90) 7 (.70)	5 (.50)
(.90) 4 (.40)	4 (.40)
	, ,
(.95) 57 (.95)	51 (.65)
(.97) 57 (.95)	31 (.52)
(.93) 62 (.76)	37 (.46)
(.77) 69 (.69)	77 (.77)
	, ,
(.80) 48 (.80)	43 (.72)
(.75) 6 (.30)	5 (.25)
(1.0) 7 (.70)	8 (.80)
(1.0) 7 (.70)	6 (.60)
	5 7 .7 ´
omic Broca	Transcortical motor
3	(.95) 57 (.95) (.97) 57 (.95) (.93) 62 (.76) (.77) 69 (.69) (.80) 48 (.80) (.75) 6 (.30) (.10) 7 (.70) (.10) 7 (.70) (.10) 7 (.70) (.93) 67.0

Note. Participants are identified as EH, CM, CN, and DC. Performance measures are indicated by the raw score (and proportion correct) on the WAB-R (Kertesz, 2006).

elicited spoken discourse from a number of contexts, which included describing and sequencing pictures and providing procedural (e.g., "Tell me how you go about writing and sending a letter") and personal information (e.g., "Tell me where you live and describe it to me"). At least 1 min of connected discourse was collected from each of the 10 stimuli.

Experimental Stimuli

Two scripts were developed for each of four common situations: booking a vacation through a travel agency, ordering a meal at a restaurant, ordering from a deli counter, and seeing a physician for a sick visit. The two scripts shared a common situation (e.g., booking a vacation) but were varied in content (e.g., booking trip to Las Vegas vs. Florida) so that a total of eight scripts was used as experimental stimuli: four scripts for the VC condition and four complementary variants for the HC condition. Assignment of scripts to the clinician condition was pseudorandomized to avoid a script pair being assigned to the same clinician condition. Four scripts (two for the VC, and two for the HC) were used in practice sessions to elicit dialogue (practice dialogues), and the remaining four (two for the VC, and two for the HC) were administered to elicit dialogues from scripts exposed only in pre- and postpractice sessions (pre-/ post-only dialogues). In one case, DC, only six dialogues were used in the analyses because the dialogues elicited from the script pair for booking a vacation through a travel agency were misassigned to clinician conditions in the postpractice period.

Script Development

The development of the scripts was guided by a typical sequence of events that may occur, including the authors' personal experiences, in each of the four common situations. For example, the sequence that guided the script for ordering a meal at a restaurant was: Hostess greets patron/ obtains information about reservation; hostess suggests seating; server hands out menus/recites specials/takes orders for drinks; server takes orders for appetizers, soup/salad, and dinner; server arrives with order; server checks on patrons; server takes order for dessert; server checks in/leaves check. All scripts started with a general greeting, "Hello. Welcome to the Philadelphia Travel Agency/Capital Grille/ Deli Counter," and "Hello, what seems to be the problem today?" Lines in scripts were designed to elicit spontaneous narrative responses and comments from each participant (e.g., "What airline do you want to fly?") but contained multiple choice (e.g., "Our carriers are USAir, United, and Southwest"), forced choice (e.g., "Do you want to fly USAir or United?"), and yes/no (e.g., "Do you want to fly USAir?") options if narrative responses could not be elicited. In addition, comments ("That's an excellent choice"), and acknowledgments ("OK, sure") were built into the script content as fillers to simulate a natural conversation. The language used for lines in the two scripts that shared a common situation was the same, with only minor changes in the content to match the situation. For example, the script content for ordering a meal at a restaurant was somewhat different for ordering lunch at the Hot Wok versus ordering dinner at the Capital Grille. The length of each of the two scripts that shared a common situation was the same but varied across the situations and depended on the sequence of events that guided them. The fewest script lines were developed for seeing a physician for a sick visit. This was due to the limitations of the event sequence that guided the conversation between a physician and patient during a sick visit. The scripts for booking a trip, ordering a meal, and ordering from a deli counter were longer because the sequence of events that occurred in these situations generated a greater number and variety of lines for eliciting a conversation. Table 3 shows the number of total lines per script, the proportion of lines that elicited obligatory responses, and the proportion that consisted of nonobligatory and filler utterances. The online supplemental materials show sample scripts for the dialogue pair booking a vacation through a travel agency.

Experimental Design

Microsoft Speech Application Interface (MS SAPI)

MS SAPI was used to program the VC to produce the speech output used in the scripts. MS SAPI essentially splits each word in a script line into individual phones, which are then matched to a set of 22 individual visemes. A viseme represents the sound being made and is used to animate the lip-synching of the VC. Once the phonemes and visemes are interfaced, the VC's mouth is lip-synched with the speech output. In addition, the mouth and eyes are controlled by commands built into each script so that the VC can produce simple facial expressions, such as smiling or frowning. For the present study, script lines were uploaded onto a text file, passed through MS SAPI, and converted into a computerized female voice. However, this was the limit of the capability of this software to program the VC. The VC could not independently recognize or reply to spoken utterances. Therefore, the output of the VC was controlled using a "Wizard of Oz" paradigm.

VC Condition: "Wizard of Oz" Paradigm

The VC condition was called a "Wizard of Oz" paradigm because an individual, designated the "Wizard" (the sixth author), was behind a black curtain (to give the illusion of autonomy to the VC) but controlled the speech output of the VC by inputting information on a computer keyboard. The "Wizard" selected script lines saved in the text file by a key press on the keyboard to initiate an utterance or respond to a participant's utterance. Although scripts were developed to include as many "lines" as possible to fit the communication situation, to anticipate potential participant responses, and to guide and provide structure to the dialogue, invariably, participants did not necessarily follow the script as originally developed. To account for unpredictable responses, the "Wizard" typed in ad-libbed script lines if necessary. Ad-libs included comments to guide participants back to the sequence of the script if they went off script (e.g., "Let

Table 3. Total lines for each script pair, total lines (and proportions of total) to elicit obligatory utterances, and total lines (and proportions) of nonobligatory and filler utterances.

Measure of lines in scripts	Capital Grille	Hot Wok	Garden Fresh	Deli counter	Vegas	Florida	Sore throat	Stomach virus
Total lines/script	68	62	61	44	60	61	38	39
Total lines: Obligatory utterances	35 (.51)	31 (.50)	49 (.80)	34 (.77)	37 (.62)	39 (.64)	24 (.63)	26 (.67)
Open-ended questions	15 (.43)	14 (.40)	16 (.26)	12 (.27)	26 (.43)	27 (.69)	19 (.50)	20 (.51)
Forced choice: Multiple options	3 (.09)	3 (.05)	9 (.15)	9 (.20)	3 (.05)	5 (.13)	1 (.03)	1 (.03)
Forced choice: Binary options	8 (.23)	7 (.11)	8 (.13)	6 (.14)	3 (.05)	4 (.10)	0 (.00)	1 (.03)
Yes/no questions	9 (.26)	7 (.11)	16 (.26)	7 (.16)	5 (.08)	3 (.08)	4 (.11)	4 (.10)
Total lines: Nonobligatory and filler utterances	33 (.49)	31 (.50)	12 (.20)	10 (.23)	23 (.38)	22 (.36)	14 (.37)	13 (.33)

me take your order for drinks first") and comments to verify responses ("Oh, you want the corn beef special." "Where did you say you wanted to stay?"). Some ad-libs were in response to creative requests from the participant (Participant: "Um, babysitter one night?" Clinician: "The hotel may have one."), or they were revisions of scripted lines to help with comprehension breakdowns. Other adlibs promoted the interactive nature of the dialogues (e.g., "Okay. I will find a deal and get that booked for you.") but were not intended to elicit responses. Each participant sat in front of a large flat screen television from which the VC (head-and-shoulder view) was projected. The VC was clothed (e.g., in a white coat), and the background was displayed on the flat screen (e.g., scene of an examining room) to fit the situation (e.g., seeing a physician for a sick visit).

HC Condition

The HC condition was held in a different room than that for the VC condition, but with the same set-up as in the VC condition. The participant sat across from the clinician (the first author), who followed a written script identical to the one for the VC. Similar to the "Wizard," who typed in ad-libbed script lines, the HC ad-libbed to guide participants back to the script, verify responses, or respond to creative requests. In both conditions, responses by the participant were audio- and/or video-recorded for later analysis.

Procedures for Counting Words and CIUs

The same procedures developed by Nicholas and Brookshire (1993) for counting words in narratives were used for counting words in the dialogues. However, the CIU scoring system was modified for scoring dialogues elicited from conversational situations to capture accurate, relevant information that otherwise would not have been counted with strict adherence to scoring criteria for CIUs. The term information unit (IU) was used to differentiate between the two scoring systems. Rules for coding IUs included the following:

- 1. The word and was counted as an IU when used as a conjunction.
- 2. Revisions, as these typically occur during conversational exchanges, were counted as IUs if the revision provided more specific information or indicated a change of mind:

Clinician: "Are you ready to order?" Participant: "I would like a salad. The Maine lobster salad." (9 words; 9 IUs)

Clinician: "What parks would you like to visit?" Participant: "Magic Kingdom. No, Universal Studios." (5 words; 5 IUs)

3. Revisions were not counted as IUs if they did not add new information:

Clinician: "Would you like me to book business or first class?"

Participant: "Business class, yeah, no business class." (6 words; 2 IUs).

Utterances such as yes, no, or okay, were rarely counted as CIUs and primarily considered as fillers when produced in narratives. However, yes, no, and okay were typical in the dialogues, and rules were developed for including them in the IU count:

1. If they were used in response to a direct question, they were counted as IUs:

> Clinician: "Would you like more iced tea?" Participant: "Okay, that would be good." (5 words; 5 IUs)

2. If they were repeated multiple times in response to a direct question, only one occurrence that provided the most accurate, relevant information was counted as an IU:

> Clinician: "Are you ready to order?" Participant: "Yes, yes, yes, no, no, I'm not ready." (9 words; 5 IUs)

Clinician: "Anything else from Hot Sandwiches?" Participant: "Yes, no." (2 words; 1 IU)

3. If they were used to revise a response, they were counted as IUs:

> Clinician: "What kind of bread would you like?" Participant: "Rye. No, wheat. No, rye. Yes, rye, rye, rye." (9 words; 7 IUs)

4. If they were used indiscriminately or ambiguously in response to a question or comment, they were not counted as IUs:

Clinician: "We have roast beef with provolone, turkey

with muenster cheese."

Participant: "Yes." (1 word; 0 IU)

Clinician: "How many of each?"

Participant: "50. No, no, yes." (4 words; 1 IU)

In addition, these procedures included only coding IUs in dialogues that were elicited from obligatory exchanges. This was to account for nondiscriminating responses to the clinician and some of the ad-libs that may have elicited responses but were not specifically intended to do so. A second scorer (a speech-language pathologist unfamiliar with the script content) coded lines in scripts as obligatory or nonobligatory exchanges. Interrater reliability ranged from .96 to 1.00 (M = .99).

Dependent Variables

The dependent variables measured the quantity and quality of information conveyed in practice and pre-/postonly dialogues elicited from scripts administered only in pre- and postpractice periods. Quantity of information conveyed in dialogues was assessed by using the total participant utterances with IUs in clinician-participant obligatory exchanges and the total IUs produced in these exchanges. The dependent variables for measuring quantity were (a) the proportion of participant utterances with IUs in clinician participant obligatory exchanges and (b) the mean IUs produced during these exchanges. The quality of information conveyed was assessed by using the total number of participant utterances consisting of more than one IU (>1 IU) in clinician-participant obligatory exchanges and the total number of IUs produced in these exchanges. The dependent variables for measuring quality were (a) the proportion of utterances with >1 IU in clinician-participant obligatory exchanges, (b) the proportion of IUs in utterances with >1 IU relative to the total IUs produced in obligatory exchanges; and (c) the mean number of IUs in utterances with >1 IU/ utterance. In addition, the proportions of CIUs and CIUs/ minute conveyed in pre- and postpractice narratives (Nicholas & Brookshire, 1993) were calculated.

Prepractice, Practice, and Postpractice Sessions

Table 4 illustrates a typical schedule of prepratice, practice, and postpractice sessions for CM. Order of scripts in prepractice, practice, and postpractice sessions for VC and HC conditions were counterbalanced within and across participants following an ABBA design (A = VC, and B = HC; see Table 4). For counterbalancing across participants, the clinician condition first exposed on Day 1 was: HC condition for EH, VC condition for CM, VC condition for CN, and HC condition for DC. Sessions for the prepractice, practice, and postpractice periods varied depending upon

participant availability. The pre- and postpractice sessions took 1 week to 10 days to complete. Practice sessions were completed in 1 week, with the exception of the practice sessions for CM, which took 3 weeks to complete. Postpractice sessions took 2 days to 1 week to complete.

Prepractice Sessions

Eight scripts were administered over 2 days so that four scripts were administered each day: two by the VC and two by the HC. Variants for each of the common situations were presented on different days by different clinicians.

Practice Sessions

Four practice sessions were carried out: two with the VC and two with the HC. Practice sessions with each clinician were on separate days. The HC introduced the practice session by giving background information about each conversational situation (e.g., "You have reservations for dinner at the Capital Grille") and telling participants they would be role-playing with a virtual (or human) clinician. Participants were encouraged to be creative (e.g., "Think about how many people you have with you. Where you would like to sit? What you would like to eat?"). Participants were given a brochure or menu and encouraged to use it if they could for situations in which it was expected (e.g., booking a vacation, ordering a meal). In a typical session, the VC or HC initiated the dialogue with a greeting (e.g., "Hello, welcome to the Deli Counter. How can I help you?") and then followed the script to elicit spontaneous, unscripted responses from the participant. If a script line did not elicit an open-ended response, the clinician selected alternative script lines to elicit multiple-choice, forced-choice, or yes/no responses. In the event that a participant was unable to respond with any of the alternative script lines or went off script, the clinician was allowed to respond with an ad-libbed utterance (e.g., "I'm taking your order for meats right now"). For the VC condition, the "Wizard" typed the ad-lib online. Two conversations were practiced twice within a single session, with a 10-min break after the first practice trial. The next day, the remaining two conversations were practiced following the same procedure. Practice sessions lasted 30–40 min.

Postpractice Sessions

The eight scripts (four used for practice sessions and four exposed only for pre/post testing) were re-administered over 2 days.

Pre- and Postpractice Testing

The Nicholas and Brookshire (1993) narratives were re-administered to obtain CIUs and CIUs/minute.

Transcription, Scoring, and Reliability

Practice and Pre-/Post-Only Dialogues

Participant responses in the eight dialogues were transcribed into templates with the sequence of the clinician script lines used in each conversational situation. Dialogues were verified for accuracy and content by a second transcriber

Table 4. A typical schedule of prepratice, practice, and postpractice sessions for CM.

	F	Prepractice			Prac	tice			Р	ostpractice	
Day 1	A Virtual B Human	4a (pre/post) 3a (pract)	2a (pract) 1a (pre/post)	Day 3	A Virtual	2a 10-mi	1b n break	Day 7	A Virtual B Human	1b (pract) 4b (pract)	3 (pre/post) 2b (pre/post)
					A Virtual	2a	1b				
Day 2	B Human	4b (pract)	2b (pre/post)					Day 8	B Human	3a (pract)	1a (pre/post)
•	A Virtual	1b (pract)	3b (pre/post)	Day 4	B Human	3a	4b	,	A Virtual	4a (pre/post)	2a (pract)
		u ,	u 1 /	,		10-mi	n break			4 1 /	u ,
					B Human	3a	4b				
				Day 5	B Human	4b	3a				
						10-mi	n break				
						4b	3a				
				Day 6	A Virtual	1b	2a				
				,			n break				
					A Virtual	1b	2a				

Note. Types of clinicians are identified as virtual and human. pract = scripts exposed in prepractice, practice, and postpractice periods; pre/post = scripts exposed only in pre- and postpractice periods; 1a = restaurant: Capital Grille; 1b = restaurant: Hot Wok; 2a = doctor's visit: sore throat; 2b = doctor's visit: stomach virus; 3a = travel agency: Florida; 3b = travel agency: Las Vegas; 4a = deli: supermarket deli counter (planning surprise party for friend); 4b = deli: Garden Fresh Deli (planning a barbeque).

who relistened to the digital recordings and made revisions to the transcripts as necessary in the same way as with the narratives. Words and information in dialogues were counted and scored using the procedures for counting words and scoring IUs previously described. Interrater reliability for scoring IUs ranged from .80 to .99 (M = .94).

Nicholas and Brookshire (1993)

Narratives were transcribed from digital recordings and verified for accuracy and content in the same way as the dialogues. Transcripts were uploaded into the Systematic Analysis of Language Transcripts (SALT; Miller, Andriacchi, & Nockerts, 2011) to obtain total number of words. Two research technicians, one a graduate student in Communication Sciences and Disorders (trained in the CIU scoring system by the first author), counted and scored words and CIUs. Interrater reliability for scoring CIUs ranged from .75 to 1.00 (M = .90).

Results

Fisher's exact tests with two-tailed tests were used for pre- and postcomparisons. We could not do this for the analyses of mean IUs in pre- and postpractice periods because we scored for IUs in utterances and excluded counts for non-IUs. Results are reported in Tables 5, 6, 7, and 8.

Quantity of Information Conveyed in Practice and Pre-/Post-Only Dialogues With HC and VC

Two comparisons are reported to measure the quantity of information conveyed in dialogues with the HC and VC during pre- and postpractice periods: (a) proportion of participant utterances with IUs in clinician-participant obligatory exchanges, and (b) mean IUs produced in obligatory clinician-participant exchanges. Results are reported in Tables 5 and 6, respectively.

Analysis 1: Proportion of Participant Utterances With IUs in Clinician-Participant Obligatory Exchanges in Pre- And **Postpractice Periods**

Tables 5 and 6 show the quantity measures for practice and pre-/post-only dialogues in the pre- and postpractice periods. These include the total clinician-participant obligatory exchanges, the total and proportion of participant utterances with IUs in obligatory exchanges, and the total and mean IUs produced in these exchanges.

Practice dialogues. CM showed a significant difference in clinician condition during the prepractice period only (p = .01; see Table 5). The proportion of utterances with IUs in obligatory clinician–participant exchanges was higher for the VC compared to the HC, .98 and .83, respectively. No other significant differences in clinician condition in pre- or postpractice periods were found.

Pre-/post-only dialogues. No significant differences in clinician condition in pre- and postpractice periods were found (see Table 6). For the four participants, proportion of utterances with IUs in obligatory clinician-participant exchanges was similar in HC and VC conditions, pre- and postpractice periods.

Analysis 2: Mean IUs in Clinician-Participant Obligatory **Exchanges in Pre- and Postpractice Periods**

Practice dialogues. For CM, EH, and CN, mean IUs produced in obligatory exchanges were relatively the same with the HC and VC in pre- and postpractice periods (see Table 5). In general, CM's mean IUs were lower in the postpractice period, 4.11 IUs with HC and 3.75 with VC. CN's mean IUs were similar with the HC and VC in the prepractice period, but somewhat higher with the HC compared to the VC during postpractice, 1.95 and 2.88 IUs, respectively. Mean IUs for DC were somewhat higher with the HC in pre- and postpractice periods.

Pre-/post-only dialogues. For CM and EH, mean IUs produced in obligatory exchanges were higher with the VC

 Table 5. Quantity measures for practice dialogues with a human clinician (HC) or virtual clinician (VC) during pre- and postpractice periods.

			C	М					El	н					CI	N					D	С		
	Pi	re	р	Po	st	р	Pr	·e	р	Po	st	р	Pı	·e	р	Po	st	р	Pi	re	р	Po	st	р
Quantity measure	Human	Virtual		Human	Virtual	-	Human	Virtual																
Total clinician-participant obligatory exchanges	47	48		53	44		44	39		50	37		47	53		38	60		28	59		22	84	
Total participant utterances with IUs in obligatory exchanges	39	47		51	44		38	37		47	36		43	52		36	60		27	58		22	84	
Proportion of participant utterances with IUs in obligatory exchanges	0.83	0.98	.01	0.96	1.00	NS	0.86	0.95	NS	0.94	0.97	NS	0.91	0.98	NS	0.95	1.00	NS	0.96	0.98	NS	1.00	1.00	NS
Total IUs in obligatory exchanges	242	248		218	165		257	298		289	273		69	89		74	173		80	90		54	135	
Mean IUs in obligatory exchanges	5.15	5.17		4.11	3.75		5.84	7.64		5.78	7.38		1.47	1.68		1.95	2.88		2.86	1.53		2.45	1.61	

Note. NS = not significant.

 Table 6. Quantity measures for pre-/post-only dialogues with a human clinician (HC) or virtual clinician (VC) during pre- and postpractice periods.

			С	М					Е	Н					С	N					D	С		
	Pr	е	р	Ро	st	р	Pr	е	р	Po	ost	р	Pı	re	р	Po	st	р	Pi	re	р	Po	st	р
Quantity measure	Human	Virtual		Human	Virtual	ı	Human	Virtual		Human	Virtual	ı												
Total clinician–participant obligatory exchanges	57	48		60	44		39	38		34	55		50	61		43	76		92	39		77	84	
Total participant utterances with IUs in obligatory exchanges	55	47		60	44		39	37		33	54		48	54		42	70		92	58		76	84	
Proportion of participant utterances with IUs in obligatory exchanges	0.96	0.98	NS	1.00	1.00	NS	1.00	0.97	NS	0.97	0.98	NS	0.96	0.89	NS	0.98	0.92	NS	1.00	0.98	NS	0.99	1.00	NS
Total IUs in obligatory exchanges	198	248		203	165		238	344		210	371		105	109		123	92		236	90		322	135	
Mean IUs in obligatory exchanges	3.47	5.17		3.38	3.75		6.10	9.05		6.18	6.75		2.10	1.79		2.86	1.21		2.86	2.31		2.45	1.61	

Note. NS = not significant.

 Table 7. Quality measures for practice dialogues with a human clinician (HC) or virtual clinician (VC) during pre- and postpractice periods.

			С	М					E	н					С	N					DO			
	Pr	re .	р	Po	st	р	Pr	е	р	Po	st	р	Pr	·e	р	Po	st	р	P	re	р	Po	st	р
Quality measure	Human	Virtual		Human	Virtual	l	Human	Virtual		Human	Virtual		Human	Virtual		Human	Virtual		Human	Virtual		Human	Virtual	
Total clinician–participant obligatory exchanges	47	48		53	44		44	39		50	37		47	53		38	60		28	59		22	84	
Total participant utterances with IUs >1 in obligatory exchanges	32	43		26	33		28	32		43	32		14	14		16	32		17	20		13	36	
Proportion of participant utterances with IUs >1 in obligatory exchanges	0.68	0.90	.01	0.49	0.75	.01	0.64	0.82	.09	0.86	0.86	NS	0.30	0.26	NS	0.42	0.53	NS	0.61	0.34	.02	0.59	0.43	NS
Total IUs in participant utterances with IUs >1 in obligatory exchanges	229	221		208	154		247	293		284	269		40	68		54	151		75	55		45	89	
Proportion IUs in utterances with IUs>1 per total IUs	0.95	0.89	.03	0.95	0.93	NS	0.96	0.98	NS	0.98	0.98	NS	0.58	0.76	.02	0.73	0.87	.009	0.94	0.61	.001	0.83	0.46	.02
Mean IUs in participant utterances with IUs >1 in obligatory exchanges	7.16	5.14		8.00	4.67		8.82	9.16		6.60	8.41		2.86	4.86		3.38	4.72		4.41	2.75		3.46	2.47	

Note. NS = not significant.

Table 8. Quality measures for pre-/post-only dialogues with a human clinician (HC) or virtual clinician (VC) during pre- and postpractice periods.

			(СМ					Е	Н					C	N					D	С		
	Pı	re	р	Po	st	р	Р	re	р	Po	ost	р	P	re	р	Po	st	р	Pi	·e	р	Po	st	р
Quality measure	Human	Virtual		Human	Virtual		Human	Virtual		Human	Virtua	ı	Human	Virtual		Human	Virtua		Human	Virtual	l	Human	Virtual	_
Total clinician-participant obligatory exchanges	57	48		60	44		39	38		34	55		50	61		43	76		92	39		77	84	
Total participant utterances with >1 IU in obligatory exchanges	32	37		26	37		26	33		24	44		16	17		21	20		45	20		28	36	
Proportion of participant utterances with > 1 IU in obligatory exchanges	0.56	0.77	.04	0.43	0.84	.0001	0.67	0.87	.09	0.71	0.80	NS	0.32	0.28	NS	0.49	0.26	.02	0.49	0.51	NS	0.59	0.43	NS
Total IUs in participant utterances with >1 IU in obligatory exchanges	170	221		181	154		225	340		201	361		75	91		102	76		192	55		294	89	
Proportion of IUs in utterances with >1 IU per total IUs	0.86	0.89	NS	0.89	0.93	NS	0.94	0.99	NS	0.99	0.97	NS	0.71	0.83	.05	0.83	0.83	NS	0.81	0.61	.0003	0.91	0.66	.0001
Mean IUs in participant utterances with >1 IU in obligatory exchanges	5.31	5.97		6.96	4.16		8.65	10.30		8.38	8.20		4.69	5.35		4.86	3.80		4.41	2.75		3.46	2.47	

Note. Pre/post-only dialogues are dialogues from scripts exposed only in pre- and postpractice sessions. NS = not significant.

in the prepractice period (see Table 6). By postpractice, mean IUs were relatively the same, but lower, with both clinicians. CN's mean IUs with the HC and VC were relatively the same during the prepractice period. By postpractice, mean IUs were somewhat higher with the HC compared to the VC, 2.86 and 1.21 IUs, respectively. DC's mean IUs were somewhat higher with the HC compared to the VC in preand postpractice periods.

Quality of Information Conveyed in Practice and Pre-/Post-Only Dialogues With HC and VC

Three comparisons are reported to measure the quality of information conveyed in dialogues: (a) proportion of participant utterances with >1 IU in clinician—participant obligatory exchanges; (b) proportion of IUs in utterances with >1 IU relative to the total IUs in obligatory exchanges; and (c) mean IUs in exchanges with >1 IU. Results are reported in Tables 7 and 8, respectively.

Analysis 1: Proportion of Participant Utterances With >1 IU in Clinician-Participant Obligatory Exchanges, Pre- and Postpractice Periods

Practice dialogues. CM showed a significant difference in HC and VC conditions in the pre- and postpractice periods (p = .01; see Table 7). The proportion of utterances with >1 IU in obligatory exchanges was higher for the VC compared to the HC. CN showed no significant difference in clinician conditions in pre- (p = .82) or postpractice (p = .31) periods. EH approached significance in clinician conditions in prepractice only (p = .09). Her proportion of utterances with >1 IU was .82 with the VC compared to .64 with the HC. DC showed a significant difference in HC and VC conditions in prepractice (p = .02) only. The proportion of utterances with >1 IU in obligatory exchanges was higher with the HC (.61) compared to the VC (.34).

Pre-/post-only dialogues. CM showed a significant difference in clinician condition in pre- (p = .04) and post-practice (p = .0001) periods (see Table 8). The proportion of utterances with >1 IU in obligatory exchanges was higher for the VC (.77) compared to the HC (.56). DC showed no significant difference in clinician condition in pre- (p = .85) or postpractice (p = .42) periods. EH approached significance in clinician condition in the prepractice period only (p = .09). The proportion of utterances with >1 IU in obligatory exchanges was .87 with the VC compared to .67 with the HC. CN showed a significant difference in clinician condition in postpractice only (p = .02). Proportion of utterances was higher with the HC compared to the VC, .49 and .26, respectively.

Analysis 2. Proportion of IUs in Utterances With >1 IU Relative to Total IUs in Obligatory Exchanges, Preand Postpractice Periods

Practice dialogues. CM showed a significant difference in HC and VC conditions in the prepractice period (p = .03; see Table 7). The proportion of IUs in utterances with >1 IU/ utterance produced by CM during the prepractice period was .95 with the HC compared to .89 with the VC. However,

there was no significant difference in HC compared to VC conditions during the postpractice periods (p = .50). EH showed no significant difference in either clinician condition in pre- (p = .12) or postpractice (p = 1.000) periods. CN showed a significant difference in clinician condition in the pre- (p = .02) and postpractice (p = .009) periods. Proportion IUs in utterances with >1 IU/utterance was higher with the VC compared to the HC. Likewise, DC showed a significant difference in clinician conditions in pre- (p = .001) and postpractice (p = .02) periods. However, his proportion of IUs was much greater with the HC compared to the VC.

Pre-/post-only dialogues. CM and EH showed no significant difference in HC or VC conditions during pre- or postpractice periods (p > .01; see Table 8). CN showed a significant difference in HC and VCs in the prepractice period only (p = .05). Proportion of IUs in utterances with >1 IU/utterance was higher with the VC compared to the HC, .83 and .71, respectively. For DC, there was a significant difference in clinician condition during pre- (p = .0003) and postpractice (p = .0001) periods. His proportion of IUs in pre- and postpractice periods was much greater with the HC compared to the VC.

Analysis 3: Mean IUs in Utterances With >1 IU, Pre- and Postpractice Periods

Practice dialogues. CM's mean IUs produced in utterances with >1 IU were higher with the HC compared to the VC in the pre- and postpractice periods (see Table 7). EH's mean IUs with the HC and VCs were similar but somewhat higher with the VC in the pre- and postpractice periods. CN's means, likewise, were observably higher with the VC in the pre- and postpractice periods but more so in the prepractice period, 4.86 and 2.86 for VC and HC, respectively. DC's mean IUs were higher with the HC compared to the VC in pre- and postpractice periods.

Pre-/post-only dialogues. CM's mean IUs were relatively similar with both clinicians, but slightly higher with the VC in the prepractice period (see Table 8). By postpractice, mean IUs were higher with the HC compared to the VC, 6.96 and 4.16 IUs, respectively. CN showed a similar pattern of performance. Mean IUs were similar but somewhat higher with the VC in the prepractice period. By postpractice, mean IUs were somewhat higher with the HC compared to the VC, 4.86 and 3.80 IUs, respectively. For EH, mean IUs were observably higher with the VC compared to the HC in the prepractice period. By postpractice, they were relatively the same with both clinicians. Mean IUs for DC were higher with the HC compared to the VC in pre- and postpractice periods.

Discourse Production in Nicholas and Brookshire (1993) Narratives

Proportion of CIUs and CIUs/Minute in Pre- And Postpractice Narratives

Prepractice. Proportion CIUs in prepractice narratives ranged from .40 for CN to .75 for CM (see Table 9). Efficiency of information conveyed in narratives was lowest

Table 9. Nicholas and Brookshire (1993) narratives: Total words, proportion of correct content information units (CIUs), CIUs/minute, and comparison of proportion of CIUs in pre- and postpractice periods.

		Prepractice		Po	stpractice			
Participant	Total words			Total words	Proportion of correct CIUs	CIUs/min	Pre- vs. postprac	ctice comparison
EH	993	.54	23	880	.66	24	Increased	.0020
CM	1,095	.75	52	1,540	.68	42	Decreased	.0001
CN	206	.40	4	324	.59	7	Increased	.0001
DC	402	.53	13	661	.46	19	Decreased	.0370

for CN, who also demonstrated the lowest number of narrative words. CM produced the highest proportion of CIUs/ minute (52), consistent with his milder aphasia. Although EH produced a high proportion of CIUs (similar to CM), the efficiency rating was lower than CM's (23 CIUs/minute).

Postpractice. Whereas EH and CN produced significantly greater CIUs in postpractice narratives (p < .03), CM and DC produced significantly fewer CIUs in postpractice narratives (p < .03). Nonetheless, DC showed some improvement in the efficiency measure, with his rate of CIUs per minute increasing from 13 (prepractice) to 19 (postpractice). CN also demonstrated an increase in the efficiency measure in the postpractice period (from four to seven words per minute). EH and CM showed no difference in their efficiency of conveying information in pre- and postpractice periods.

Discussion

One of our long-term goals is to develop a dialogue practice tool with a VC and virtual simulations of activities of daily living that PWAs can use to help maximize use of their residual language abilities. Our primary question in this study was whether outcomes of a dialogue practice protocol for PWAs would differ depending on whether the dialogue partner was an HC or a VC. Our participants varied in severity and type of aphasia. Two with more severe aphasia, DC and CN, presented with agrammatic verbal output. Consistent with the profile of a transcortical motor aphasia, DC also demonstrated good repetition. CN's overall language profile fit that of Broca's aphasia with a cooccurring apraxia of speech. The two participants with milder aphasia, CM and EH, produced higher rates of words, IUs, and CIUs in the pre- and postpractice assessments. CM and EH presented with a mild anomia. They typically produced relevant, syntactically well-formed utterances. Also, EH presented with an apraxia of speech particularly observed during production of multisyllabic words and longer utterances.

We used several measures to compare outcomes of dialogue practice with a HC versus a VC. For the measure of the quantity of information conveyed in dialogues, we used two measures: (a) the proportion of utterances with IUs produced in obligatory clinician-participant exchanges and (b) the mean IUs produced during these exchanges. For the measurement of quality of utterance, we used three

measures to analyze utterances with >1 IU in obligatory exchanges: (a) the proportion of utterances with >1U. (b) the proportion of IUs in these utterances relative to total IUs produced, and (c) the mean number of IUs in utterances with >1 IU/utterance.

For the quantity measures, three of the four participants demonstrated no significant difference in clinician condition in pre- and postpractice periods. CM showed a significant difference in clinician condition for the practice dialogues during the prepractice period only. The proportion of utterances with IUs in obligatory exchanges was higher with VC. For mean IUs in pre- and postpractice periods, two participants consistently produced more IUs in one clinician condition. DC's mean IUs were somewhat higher with the HC in pre- and postpractice periods. On the other hand, EH's mean IUs were somewhat higher with the VC in pre- and postpractice periods. For the quality measures, the results were more equivocal. EH showed no significant difference in clinician condition for any of the quality measures. CM showed a significant difference in clinician condition relative to the proportion of utterances with >1 IU produced in pre- and postpractice periods (the proportion of utterances was higher with the HC). By postpractice, however, he showed no significant difference in clinician condition for the proportion of IUs in utterances with >1 IU relative to total IUs produced. CN showed the opposite effect. There was no significant difference in clinician condition relative to the proportion of utterances with >1 IU produced in pre- and postpractice periods. However, the proportion of IUs in utterances with >1 IU was typically higher with the VC in both practice periods. Of the four participants, DC showed a significant difference in clinician condition for the quality measures. The proportion of utterances with >1 IU and the proportion of IUs in utterances with >1 IU were typically higher with the HC. In addition to the mean IUs in utterances with >1U/utterance, DC's means were somewhat higher with the HC in pre- and postpractice periods. In general, our results suggest that three participants were equally inclined to interact with a HC or a VC. Only DC's results for quantity and quality of information conveyed in dialogues suggest a preference for the HC condition. However, we did not collect patientreported outcome data—for example, preferred mode of delivery—which would have provided a measure of the feasibility and applicability of a VC to the treatment of functional communication. In addition, the collection of listeners'

judgments from pre- and postpractice periods would have provided a measure of social validation (Doyle, Tsironas, Goda, & Kalinyak, 1996; Hickey & Rondeau, 2005). We plan to collect these data in future studies.

As an assessment of generalization to dialogue production, we evaluated whether there was change in the content and efficiency of information conveyed by the participants in the Nicholas and Brookshire (1993) narratives exposed only in pre- and postpractice periods. The results were mixed. For two participants, EH and CN, there were significant increases in the proportion of their CIUs. For CM and DC, rates of CIUs decreased significantly in the postpractice assessment. The two participants with agrammatic verbal output, DC and CN, both showed improvement in their efficiency of information produced (CIUs/minute).

Overall, this study supports the idea that using VCs to practice functional communication abilities of PWAs could be beneficial. Although there are only four participants in this study, the improvements made by the two participants with agrammatism are especially promising. Nonetheless, it is necessary to administer this protocol with more PWAs expanding the scope of impairment profiles (e.g., including participants with input processing difficulties). In so doing, we may be able to identify the strengths and limitations of this technology for promoting functional communication abilities of PWAs.

This study, using the "Wizard of Oz" paradigm (textto-speech support to drive the VC), showed that the outcome of practicing dialogues with VCs is sometimes better, but at least similar to that with HCs. Why, then, would we want to pursue the use of virtual technology to promote functional communication skills of PWAs? There are three potential advantages that the use of a VC can provide: (a) consistency of protocol and standardized measurements of improvement within and across participants, (b) a generalization measure for standard language treatment that simulates functional communication situations, and (c) a home practice tool as a supplement to treatment.

Consistency of Measurement

It is imperative that treatment approaches to the rehabilitation of aphasia focus on both the impairment (e.g., improving access to linguistic representations) and the consequences of the impairment (e.g., maximizing use of residual language abilities in functional communication situations). Regardless of the treatment approach, direct or consequence based, it is important to systematically measure change in language performances as evidence of treatment effectiveness. Activities that facilitate functional language currently include role-playing, with the clinician playing the part of doctor, shopkeeper, or other potential interlocutor. This typically occurs as an informal exercise or with the guide of a standardized assessment battery, such as the Communication Abilities in Daily Living-Second Edition (CADL-2; Holland et al., 1999). The CADL-2 provides some structure to this stage of rehabilitation and some potential for consistency of measurement within and across

participants. Because it is programmed with a specific set of script lines to elicit and quantify spoken dialogue, a VC would build on this important component of evidencebased practice. At the same time, it would provide a means to personalize dialogues to individual needs of clients. The VC has the capability to do this. It is programmed with a set of script lines that are consistently used within and across sessions to engage the PWA in dialogue used in functional situations. The consistency of the VC's contribution to the dialogue will allow for more reliable measures of improvements in the PWA's contribution over practice sessions.

Generalization Measure

We used the Nicholas and Brookshire (1993) narratives to assess generalization of improvements from practicing the dialogues to another form of connected speech. Elicitation of spoken dialogue from narratives (Capilouto, Wright, & Wagovich, 2005; Nicholas & Brookshire, 1993) is frequently used as a generalization measure in many studies of treatments for aphasia. Dialogues with a VC could serve as another means to assess generalization of treatment effects as they are elicited in a less constrained context that simulates real communication situations. At present, collecting evidence for generalization to activities of daily living can be very challenging. Virtual technology potentially can provide opportunities for people with language impairment to practice interactive dialogues in the context of virtual environments that simulate everyday situations.

Treatment Fidelity

One of our long-term goals for this project is to develop a dialogue practice tool with a VC and virtual simulations of activities of daily living that PWAs can use to help maximize use of their residual language abilities. An important consideration in this endeavor is to ensure that this practice is in accord with the principles of treatment fidelity. Several aspects of using this technology can promote treatment fidelity. With virtual environments, there is opportunity for consistency of measurement in this stage of language treatment. Fundamental aspects of training functional communication activities can be quantified, and methods can be replicated in multiple cases. This development will, in turn, translate to more sound evidence-based practice. From a clinical perspective, the use of virtual environments to promote functional communication may serve to bridge the gap between functional communication activities in the treatment room to typical everyday communication situations.

Home Practice Tool

Although the VC as presented in this "Wizard of Oz" paradigm is not feasible to use as a means of practicing the dialogues at home, this is a primary rationale for our goal of developing speech recognition software to support a VC with the range of vocabulary to serve as a dialogue partner in a series of functional communication scenarios. Currently, there is technology to support VCs that can participate in

script practice (Cherney et al., 2007, 2008). This enterprise has been enormously successful. Our aim is to build on this by enabling the flexibility of the PWA to communicate a variety of responses to questions from the VC and for those responses to be understood. Although script training simulates dialogue, the aim is to elicit verbatim utterances, which may be automatized into a speaker's repertoire (Cherney et al., 2008; Youmans et al., 2005). The aim of practicing dialogues with the VC, however, is to elicit spontaneous, unrehearsed utterances. What we have found in the present study is that the participants respond with a bit of creativity in their dialogue, which the VC had encouraged. For example, in a travel agency dialogue, when one participant booked a room for a Las Vegas hotel, he asked the VC to reserve a room for his spouse and children, even though he was single and had no children. Thus, the creative aspect that flexibility of response allows in this dialogue task could potentially stimulate richer, more varied language output in the context of a simulated functional situation.

There are certainly challenges to the development of the VC with this kind of versatility and especially for use in aphasia rehabilitation. Aphasia can affect all stages of the communication chain, including semantics, word choices, pronunciations, acoustics, and even written input via a keyboard, touch screen, or mouse (Glosser, Wiener, & Kaplan, 1986). Although the automatic speech recognition (ASR) and SDS technology currently available is quite sophisticated, the models and interaction strategies that have been developed have always assumed that users speaking to such systems are predictable. This is not the case for those with aphasia, whose speech production and language use will not fit the assumptions typically made in ASR and SDS development (Wade, Petheram, & Cain, 2001). ASR technologies have been adapted to the treatment of speech disorders, but their use is found primarily in two areas: the diagnosis of motor speech disorders, primarily dysarthric speech (Rudzicz, 2007; Young & Mihailidis, 2010), script training (Cherney et al., 2008); and sentence processing (Thompson, Choy, Holland, & Cole, 2010). Our goal is to build on the design of spoken-language systems that incorporate contemporary SDS technology and advanced speech recognition capabilities to create a system that exhibits flexible behavior that can accommodate aphasic speech both on the level of production and on the level of language use.

Conclusions

In this study, we extended the approach by Cherney and colleagues (Cherney et al., 2007, 2008) to a unique application of virtual technology to conversational situations during which the VC is the only conversational partner who is scripted. We used a "Wizard of Oz" paradigm to test the applicability, validity, and feasibility of using virtual technology to promote the functional communication of four speakers with different types and severity of aphasia. We found that these speakers responded positively during interactive dialogues with a VC, as demonstrated by their

producing at least as much spontaneous verbal output as they did with an HC in the context of scenarios that simulated real-life communication situations. In addition, the expository nature of the task encouraged, at least with these individuals, a richness, creativity, and variety of utterances in dialogues with the VC. In future studies, we plan to extend the application of this technology to include a broader range of aphasia types and severity of language profiles, including speakers who demonstrate input processing difficulties. Our plan is to develop an SDS that interfaces with the VC to interact autonomously with a speaker with aphasia. The application of virtual technology in the rehabilitation of aphasia has the potential to provide strong evidence of treatment efficacy in terms of (a) consistency of measurement, (b) ability to assess generalization, (c) treatment fidelity, and (d) its potential to be used as a home practice tool. In the present study, we add to the growing body of evidence of the feasibility of virtual technology for aphasia. One of our long-term goals is to develop a dialogue practice tool with a VC and virtual simulations of activities of daily living that PWAs can use to help maximize and potentially improve their residual language abilities.

Acknowledgments

This study was supported by a grant from the National Institute on Deafness and Other Communication Disorders (DC 012245) awarded to Temple University (PIs: Nadine Martin, Emily Keshner). We would like to thank the coordinated efforts of our research team, without whose expertise in the fields of virtual environments, computer science, communication sciences and disorders, and neuroscience this project would not have been possible. We would like to thank Samantha Waldman-Rosenberg, Thomas Seminack, Julie Schlesinger, Lydia Spanier, Amelia Wisniewski-Barker, Katlyn Yackoski, and Julia Zabihach for their assistance with data analysis and preparation of this article. A special thank you to EH, CM, CN, and DC, who cheerfully participated in a unique application of a functional approach to aphasia rehabilitation.

References

- **Ahlsén, E.** (2005). Argumentation with restricted linguistic ability: Performing a role play with aphasia or in a second language. Clinical Linguistics & Phonetics, 19, 433-451.
- Aten, J. L., Cligiuri, M. P., & Holland, A. L. (1982). The efficacy of functional communication therapy for chronic aphasic patients. Journal of Speech and Hearing Disorders, 47, 93–96.
- Capilouto, G. J., Wright, H. H., & Wagovich, S. A. (2005). CIU and main event analyses of the structured discourse of older and younger adults. Journal of Communication Disorders, 38, 431–444.
- Cherney, L. R. (2010). Oral reading for language in aphasia (ORLA): Evaluating the efficacy of computer-delivered therapy in chronic nonfluent aphasia. Topics in Stroke Rehabilitation, 17, 423-431.
- Cherney, L. R., Halper, A. S., Holland, A. L., & Cole, R. (2008). Computerized script training for aphasia: Preliminary results. American Journal of Speech-Language Pathology, 17, 19-34.
- Cherney, L. R., Halper, A. S., Holland, A. L., Lee, J. B., Babbitt, E., & Cole, R. (2007). Improving conversational script production in aphasia with virtual therapist computer treatment software. Brain and Language, 103, 246-247.

- Cherney, L. R., Oehring, A., Whipple, K., & Rubenstein, T. (2010, May). Patient-reported outcomes following a drama class for individuals with chronic aphasia. Paper presented at the 40th Annual Clinical Aphasiology Conference, Isle of Palms, South Carolina.
- Doyle, P. J., McNeil, M. R., Mikolic, J. M., Prieto, L., Hula, W. D., Lustig, A. P., . . . Elman, R. J. (2004). The Burden of Stroke Scale (BOSS) provides valid and reliable score estimates of functioning and well-being in stroke survivors with and without communication disorders. *Journal of Clinical Epidemiology*, 57, 997–11007.
- Doyle, P. J., Tsironas, D., Goda, A. J., & Kalinyak, M. (1996). The relationship between objective measures and listeners' judgements of the communicative informativeness of the connected discourse of adults with aphasia. *American Journal of Speech-Language Pathology*, 5, 53–60.
- Fogg, B. J., & Nass, C. (1997). Silicon sycophants: The effects of computers that flatter. *International Journal of Human-Computer* Studies, 46, 551–561.
- Garcia, L. J., Rebolledo, M., Metthé, L., & Lefebvre, R. (2007).
 The potential of virtual reality to assess functional communication in aphasia. *Topics in Language Disorders*, 27, 272–288.
- **Glosser, G., Wiener, M., & Kaplan, E.** (1986). Communicative gestures in aphasia. *Brain and Language*, 27, 345–359.
- **Hickey, E., & Rondeau, G.** (2005). Social validation in aphasiology: Does judges' knowledge of aphasiology matter? *Aphasiology*, *19*, 389–398.
- Holland, A. L., Frattali, C. M., & Fromm, D. (1999). Communicative Abilities in Daily Living—Second Edition. Austin, TX: Pro-Ed.
- Holland, A. L., Fromm, F., DeRuyter, F., & Stein, M. (1996).
 Treatment efficacy: Aphasia. *Journal of Speech and Hearing Research*, 39, S27–S36.
- Kagan, A., Simmons-Mackie, N., Rowland, A., Huijbregts, M., Shumway, E., McEwen, S., . . . Sharp, S. (2008). Counting what counts: A framework for capturing real-life outcomes of aphasia intervention. *Aphasiology*, 22, 258–280.
- Kertesz, A. (2006). Western Aphasia Battery–Revised. San Antonio, TX: Pearson.
- Lee, J. B., Kaye, R. C., & Cherney, L. R. (2009). Conversational script performance in adults with non-fluent aphasia: Treatment intensity and aphasia severity. *Aphasiology*, 23, 885–897.
- Mannheim, L., Halper, A. S., & Cherney, L. R. (2009). Patient-reported changes in communication after computer-based script training for aphasia. Archives of Physical Medicine and Rehabilitation, 90(4), 623–627.
- Martin, N., Thompson, C. K., & Worrall, L. (2007). Aphasia rehabilitation: The impairment and its consequences. San Diego, CA: Plural.

- Mayer, R. E., Johnson, W. L., Shaw, E., & Sandhu, S. (2006). Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human-Computer Studies*, 64, 36–42.
- Miller, J. F., Andriacchi, K., & Nockerts, A. (2011). Assessing language production using SALT software: A clinician's guide to language sample analysis. Middleton, WI: SALT Software.
- Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995).
 Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43, 223–239.
- Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, *36*, 338–350.
- Raymer, A. M., Beeson, P., Holland, A., Kendall, D., Maher, L. M., Martin, N., . . . Gonzalez Rothi, L. J. (2008). Translational research in aphasia: From neuroscience to neurorehabilitation. *Jour*nal of Speech, Language, and Hearing Research, 51, S259–S275.
- Rudzicz, F. (2007). Comparing speaker-dependent and speaker-adaptive acoustic models for recognizing dysarthric speech. In Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 255–256). New York, NY: Association for Computing Machinery.
- Taggart, W., Turkle, S., & Kidd, C. D. (2005). An interactive robot in a nursing home: Preliminary remarks. In *Proceedings* of the CogSci Android Science Workshop (pp. 56–61). Stresa, Italy: Cognitive Science Society. Available at http://www. androidscience.com/proceedings2005/TaggartCogSci2005AS.pdf
- Teodoro, G., Martin, N., Keshner, E., Shi, J. Y., & Rudnicky, A. (2013, August). *Virtual Clinicians for the treatment of aphasia and speech disorders*. Poster presented at the International Conference on Virtual Rehabilitation, Philadelphia, PA.
- Thompson, C. K., Choy, J. J., Holland, A., & Cole, R. (2010).Sentactics: Computer-automated treatment of underlying forms.Aphasiology, 24, 1242–1266.
- Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: The complexities of cybercompanionship. *Connection Science*, 18, 347–361.
- Wade, J., Petheram, B., & Cain, R. R. (2001). Voice recognition and aphasia: Can computers understand aphasic speech? *Dis-ability and Rehabilitation*, 23, 604–613.
- Youmans, G., Holland, A., Munoz, M. L., & Bourgeois, M. (2005).Script training and automaticity in two individuals with aphasia. *Aphasiology*, 19, 435–450.
- Young, V., & Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and implications for speechbased applications used by the elderly: A literature review. Assistive Technology, 22, 99–112.