Multidisciplinary stroke rehabilitation delivered by a humanoid robot: Interaction between speech and physical therapies

Yu-kyong Choe¹, Hee-Tae Jung², Jennifer Baird³, and Roderic A. Grupen²

¹Department of Communication Disorders, University of Massachusetts Amherst, Amherst, MA, USA
²Department of Computer Science, University of Massachusetts Amherst, Amherst, MA, USA
³Department of Kinesiology, University of Massachusetts Amherst, Amherst, MA, USA

Background: A great number of stroke patients pursue rehabilitation services in multiple domains (e.g., speech, physical, occupational). Although multidisciplinary and interdisciplinary approaches to stroke rehabilitation are considered desirable, it is largely unknown how the intervention in one domain affects the progress in others.

Aims: The current study investigated the interaction between speech therapy and physical therapy. Additionally, the feasibility of utilising a humanoid robot in stroke rehabilitation was described.

Methods & Procedures: A 72-year-old male chronically challenged by aphasia and hemiparesis completed speech and physical therapy tasks in the sole condition (Speech Only, Physical Only) and in the sequential condition (Speech & Physical). The therapy activities were delivered by a humanoid robot.

Outcomes & Results: Greater gains in speech and physical functions were obtained during the sole condition than in the sequential condition, suggesting a competitive interaction between speech and physical therapies.

Conclusions: The cross-domain competition can be accounted for by fatigue, participant characteristics, and task characteristics. Objective data on speech and physical functions and subjective data on perceived quality of life indicate positive outcomes in this single case. These findings warrant further research on the feasibility and utility of humanoid robots in stroke rehabilitation.

Keywords: Aphasia; Hemiparesis; Multidisciplinary; Interdisciplinary; Robotics.
Stroke is the leading cause of severe and complex disability (Adamson, Beswick, & Ebrahim, 2004). A great number of stroke survivors are chronically challenged by deficits in mobility, dexterity, and ability to communicate (Kelley-Hayes et al., 2003). Such impairments in multiple domains require rehabilitation from various disciplines (Miller et al., 2010). To provide coordinated care a team of professionals may take a multidisciplinary approach (i.e., individual clinicians with strong autonomy provide services in parallel) or an interdisciplinary approach (i.e., clinicians collaborate in determining treatment plans and conducting intervention) (Korner, 2010; Mickan & Rodger, 2000). Past research reported that, compared to routine care, multidisciplinary and interdisciplinary stroke rehabilitation resulted in less mortality and better functional state (Indredavik et al., 1991), greater independence in medication management (Purdy, 2007), improved activities of daily living (Yagura, Miyai, Seike, Suzuki, & Yanagihara, 2003), and greater emotional support felt by stroke patients and less strain felt by caregivers (Lincoln, Dixon, & Knights, 2004).

Despite the importance of working together with other professionals, it is largely unknown how the intervention in one domain affects the progress in other domains. A more specific research question investigated in the current study was: How would effects of speech-language therapy interact with the effects of physical therapy in a multidisciplinary approach? To answer this question the current study compared two treatment schedules: sole (e.g., only speech therapy was provided for 6 weeks, and then only physical therapy was provided for 6 weeks), and sequential (e.g., speech and physical therapies were provided in back-to-back sessions 6 weeks). At least three different patterns of potential outcomes were predicted: No interaction, Synergistic interaction, and Competitive interaction.

If there is no interaction between speech and physical therapies, the sole and sequential conditions would yield equivalent outcomes. Presumably, speech and physical therapies would have only independent domain-specific effects without affecting progress in the other area. This pattern of recovery is supported by the modular view of cortical and cognitive functions (Fodor, 1983). In this view, speech-language production and limb movements are controlled by two autonomous modules. Thus impairments and functional recovery in one module do not interact with those in the other.

A second possibility is that the interventions on speech-language and limb functions may yield a synergistic outcome. As a result stroke patients would make greater gains from the sequential condition than from the sole condition. Contrary to the modularity hypothesis, recent research proposes a common cortical area for speech and limb movements (Rizzolatti & Craighero, 2004). Broca’s area, previously known as the motor speech area, is now considered to play a role in understanding and planning hand gestures (Baumgaertner, Buccino, Lange, McNamara, & Binkofskii, 2007; Binkofskii & Buccino, 2004). The close link between speech and limb motor control can be exemplified by an involuntary contraction of lip muscles during precision grips (Higginbotham, Isaak, & Domingue, 2008) and a strong correlation between mouth and hand movements in velocity and amplitude (Gentilucci, Benuzi, Gangitano, & Grimaldi, 2001). Furthermore, vocalising “yeah” has been reported to significantly enhance the velocity and smoothness of hand movements in stroke rehabilitation (Maitra, Telage, & Rice, 2006). Given the tight connection between speech and hand movements at the cortical and behavioural levels, it is likely that effective intervention in one domain has a positive impact on the other. Nonetheless, this synergistic interaction between speech and physical therapies may be more relevant to interdisciplinary
intervention (i.e., simultaneous speech and arm movements) than multidisciplinary intervention (i.e., sequential therapy sessions as in the current study).

A third possibility states that the speech and physical functions may compete for limited resources available in the damaged brain. Hence back-to-back sessions of speech and physical therapies may be detrimental, yielding less gain than the sole treatment condition. This pattern of recovery is supported by the fact that a great number of stroke survivors chronically experience physical and mental fatigue, which is qualitatively differentiated from typical tiredness (Kirkevold, Christensen, Andersen, Johansen, & Harder, 2012). Such persisting fatigue may account for the findings of Marshall and King (1973) that individuals with stroke-induced aphasia had significantly lower scores in the Porch Index of Communicative Ability after 30 minutes of lower-extremity exercises than after 30 minutes of rest. Similarly, the sequential condition of speech and physical therapies may cause certain stroke patients to experience a heightened sense of mental and physical fatigue which, in turn, adversely affects the treatment outcome. For those individuals the sole condition may be more beneficial.

The secondary aim of this study was to explore the feasibility of utilising a humanoid robot (i.e., a robot that resembles the shape of a human body) in stroke rehabilitation. To address this aim the sole and sequential conditions of speech and physical therapies were implemented through the use of a humanoid robot. Its morphology and its physical embodiment enabled the robot to present therapy activities in an interactive manner, promoting the participant’s motivation to complete the given tasks. This paper reports a single-case study that investigated the interaction between speech and physical therapies in robot-mediated stroke rehabilitation by comparing sole and sequential therapy schedules.

**METHOD**

**Participant**

A 72-year-old male (Tom) with chronic aphasia and hemiparesis participated in the study. Tom had a left-hemisphere stroke 9 years prior to participation in the present study. He was a monolingual English speaker and premorbidly right-handed. He spoke in short phrases and sentences, and wrote single words and numbers. Tom had right hemiparesis, and ambulated with a cane. He had received a doctoral degree and formerly worked as a school superintendent. Tom was provided verbal and written explanations of the study in the presence of his wife and gave informed consent in written form. The procedures for this study were approved by University of Massachusetts Amherst Institutional Review Board.

During the pre-treatment phase Tom’s speech-language and cognitive functions were assessed by administering the Western Aphasia Battery – Revised (Kertesz, 2007), the Cognitive Linguistic Quick Test (Helm-Estabrooks, 2001), and the Apraxia Battery for Adults – Second Edition (Dabul, 2000). Table 1 provides an overview of the test results. Tom presented with moderate transcortical motor aphasia, mild verbal apraxia, and mild cognitive-linguistic deficits. Tom’s upper-extremity functions were assessed by the Fugl-Meyer Assessment (FMA; Deakin, Hill, & Pomeroy, 2003; Fugl-Meyer, Jaasko, Leyman, Olsson, & Steglund, 1975) and the Wolf Motor Function Test (WMFT; Wolf et al., 2010; Wolf, Lecraw, Barton, & Jann, 1989). The FMA and the
**TABLE 1**
Overview of the standardised test results

<table>
<thead>
<tr>
<th>Test Instrument</th>
<th>Aphasia Quotient</th>
<th>Aphasia Type</th>
<th>Verbal Apraxia</th>
<th>Limb/Oral Apraxia</th>
<th>Composite Severity Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Aphasia Battery – Revised</td>
<td>73.7</td>
<td>Transcortical Motor</td>
<td>None – Mild</td>
<td>None</td>
<td>Mild</td>
</tr>
<tr>
<td>Apraxia Battery for Adults – Second edition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Linguistic Quick Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WMFT are the two most widely used instruments in stroke rehabilitation. These two tests in combination can provide a comprehensive profile of a stroke survivor’s upper-extremity functions. The FMA consists of 33 items to evaluate upper limb motor performance of stroke patients. Items tested include reflex activity, motor synergies, voluntary movement, grasp, and coordination. Performance on each item (e.g., *Touch your ear with your weaker hand*) is rated by a clinician (e.g., 0 – cannot be performed, 1 – detail partly performed, 2 – detail is performed faultlessly). Tom received 31 out of 66 potential points on the FMA. The WMFT assesses both gross and fine motor function during performance of 15 timed tasks (e.g., folding a towel). All 15 items are highly relevant to daily activities. Some of the tasks require only single-joint control and one-step actions, whereas other tasks require coordination of multi-joint control. Up to 120 seconds are given to complete each task. Tom completed 3 out of the 15 tasks, and his mean response time for the 3 tasks was 4.95 seconds. The tasks Tom completed were considered to be the least difficult of the 15 WMFT items (Woodbury et al., 2010). Results of the two tests indicated that Tom presented with moderate right-sided hemiparesis characterised by minimal active motion in the right shoulder and elbow, and no functional use of the right hand.

**Design**

Figure 1 outlines the study design that compares two treatment conditions: sole (Speech Only; Physical Only) and sequential (Speech & Physical). Tom completed 14 speech therapy sessions in 4 weeks during the Speech Only period, and he completed 14 speech therapy sessions in 5 weeks during the Speech & Physical period. The Physical Only period and the Speech & Physical period each consisted of 18 sessions for upper-limb exercises. Tom’s speech and physical functions were assessed before and after each therapy period.

![Figure 1. Study design.](image-url)
Equipment and stimuli

Speech and physical therapy activities were presented through uBot-5, a humanoid robot developed at the Laboratory for Perceptual Robotics (LPR) at the University of Massachusetts Amherst. As depicted in Figure 2, the uBot-5 is a bi-manual mobile manipulator that is 86 cm tall and weighs 16 kg. Each arm has 4 degrees of freedom. The robot can move and dynamically balance on two wheels. The uBot-5’s behaviours during the therapy sessions are implemented using control basis and action schema frameworks, which have been developed by the LPR. These frameworks are especially powerful at describing interaction between a robot and its environment.

The task for the robot-mediated speech practice was confrontation naming. To make the task functionally relevant, word stimuli were selected by Tom and his wife. The word list submitted by them included the names of family members and friends, food items, and objects related to daily activities (e.g., Jon, oyster, paint brush). The baseline assessment on the words consisted of: (1) target picture naming; (2) importance rating; and (3) articulatory demand score. Tom’s ability to name target pictures prior to the treatment phase was scored using a modification of the Porch Index of Communicative Ability scoring system (Porch, 2001). Table 2 summarises the multidimensional 16-point scoring system. The scores range from 16, “correct production without any support or delay” to 0, “no relevant response”. These 16 points are classified into five groups: Spontaneous (16–12), Cueing (11–7), Modeling (6–4), Tactile (3), Not produced (2–0). Tom named most of the items once the target words were presented in written form, and thus his scores were 9 (Whole Word Written Cue) or above. Any words with scores of 16–12 in the baseline assessment were eliminated from target word selection to avoid potential ceiling effects in subsequent assessments.

Even if any two items had the same baseline naming score, progress on the two words could differ depending on functional relevance and motoric demands. To control for those factors, importance ratings and articulatory demand scores were obtained. Tom rated each target item (1 – somewhat important, 2 – very important,
### TABLE 2
Hierarchical scoring system for naming responses

<table>
<thead>
<tr>
<th>Score</th>
<th>Response type</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Complete</td>
<td>Spontaneous</td>
<td>Correctly says the target word without any cue within 3 seconds.</td>
</tr>
<tr>
<td>15</td>
<td>Delayed</td>
<td>Spontaneous</td>
<td>Correctly says the target word without any support but after more than 3 seconds’ delay.</td>
</tr>
<tr>
<td>14</td>
<td>Phonemic error</td>
<td>Spontaneous</td>
<td>Incorrect phonemes are pronounced but spontaneously corrected (e.g., “tat . . . . . cat” for a target “cat”).</td>
</tr>
<tr>
<td>13</td>
<td>Delayed phonemic error</td>
<td>Spontaneous</td>
<td>After more than 3 seconds’ delay, incorrect phonemes are pronounced but spontaneously corrected.</td>
</tr>
<tr>
<td>12</td>
<td>Self-corrected</td>
<td>Spontaneous</td>
<td>Responds with a wrong word and then self-corrects (e.g., “dog . . . . . cat” for a target “cat”).</td>
</tr>
<tr>
<td>11</td>
<td>Semantic cue</td>
<td>Cueing</td>
<td>Correctly says the target word after a phrase or sentence providing a semantic cue (e.g., “it meows” for a target “cat”).</td>
</tr>
<tr>
<td>10</td>
<td>Word shape cue</td>
<td>Cueing</td>
<td>Correctly says the target word when an initial letter and total number of letters are given in a written form (e.g., “c__” for a target “cat”).</td>
</tr>
<tr>
<td>9</td>
<td>Whole word written cue</td>
<td>Cueing</td>
<td>Correctly says the target word when the whole word is presented in a written form.</td>
</tr>
<tr>
<td>8</td>
<td>Initial sound cue</td>
<td>Cueing</td>
<td>Requires initial sound cue (e.g., the /k/-sound for a target “cat”) before correctly producing the target word. The cue can be repeated once on request.</td>
</tr>
<tr>
<td>7</td>
<td>Lip shape cue</td>
<td>Cueing</td>
<td>Requires seeing a clinician silently mouth the word before correctly producing the target word. The cue can be repeated once on request.</td>
</tr>
<tr>
<td>6</td>
<td>Whole word spoken cue</td>
<td>Modelling</td>
<td>Correctly says the target word after it has been spoken by the clinician. Modelling may be repeated once by request.</td>
</tr>
<tr>
<td>5</td>
<td>Repeated presentation</td>
<td>Modelling</td>
<td>Correctly says the target word after watching the clinician repeat the word five times.</td>
</tr>
<tr>
<td>4</td>
<td>Simultaneous production</td>
<td>Modelling</td>
<td>Correctly produces the target word during five times of in-unison repetitions with the clinician.</td>
</tr>
<tr>
<td>3</td>
<td>Tactile cue</td>
<td>Tactile</td>
<td>Correctly produces the target word with touch cues in conjunction with in-unison repetitions from the clinician.</td>
</tr>
<tr>
<td>2</td>
<td>Incomplete</td>
<td>Not produced</td>
<td>Produces an approximation but cannot completely produce the word.</td>
</tr>
<tr>
<td>1</td>
<td>Incorrect</td>
<td>Not produced</td>
<td>Produces none of the phonemes in the target word.</td>
</tr>
<tr>
<td>0</td>
<td>No response</td>
<td>Not produced</td>
<td>Produces no relevant response (e.g., stereotypic utterance).</td>
</tr>
</tbody>
</table>

3 – extremely important) based on the level of functional relevance of the item in his everyday activities. Articulatory demand scores were obtained by asking Tom to repeat a syllable consisting of a consonant and the schwa after a clinician’s model. His responses were scored as follows: 1 – correct repetition after 1 or 2 presentation(s); 2 – correct repetition after 3 to 5 presentations; 3 – correct repetition with tactile cueing; 4 – incorrect. The articulatory demand score for each word was calculated by adding up the scores of constituting consonants. Thus this measure could control for word length as well.
Based on the initial assessments, 60 words were divided into three experimental conditions: sole (Speech Only), sequential (Speech & Physical), and no practice (control). The three conditions had approximately the same mean and median scores in baseline naming performance, importance ratings, and articulatory demand scores. Friedman tests indicated that the three conditions were statistically equivalent in naming, $\chi^2(2) = 1, p = .607$, importance, $\chi^2(2) = 1.351, p = .509$, and articulatory demand, $\chi^2(2) = 0.8731, p = .646$.

The 40 words in the two practice conditions were programmed for computerised practice using Microsoft PowerPoint software. Detailed descriptions of the practice program are provided in Choe and Stanton (2011). The main feature of the therapy program was the use of video clips of a speaker giving cues (e.g., semantic, phonemic) and models for verbal naming. Each target word was presented on six consecutive slides with increasing levels of support. Figure 3 presents examples of practice slides. The speech practice program was visually presented on the uBot-5’s monitor screen. Sound output was played through speakers placed at the level of the robot’s waist.

The robot-mediated physical therapy focused on the dominant right arm and hand functions with the following goals: to enhance ranges of motion, to improve dexterity, and to increase the use of the impaired arm and hand in daily activities. Through an extensive discussion among a physical therapist (third author), Tom, and his wife, it was decided that Tom would engage in three therapy exercises: Task 1 – holding his hands together and stretching his arms to reach for the robot’s hand presented at various points on the vertical plane, as demonstrated in Figure 4 (assisted movement); Task 2 – flexing and extending the elbow joint to touch the robot’s hand presented at various points on the horizontal plane (unassisted movement); and Task 3 – lifting the forearm to touch the robot’s hand presented above Tom’s hand and at various points on the horizontal plane (unassisted movement against gravity).

![Figure 3](image-url)
Procedure

During the Speech Only period Tom attended fourteen 30-minute sessions. The initial criterion for completion was either correctly naming 18 out of 20 items at Slide #1 (i.e., 90% spontaneous naming) or completing 18 sessions, whichever came first. Tom reached 85% at the fourth session, after which his performances continued to hover around the 65–85% range. After the 14th session Tom and his wife felt that a sufficient amount of practice had been given to the 20 target items. Thus it was decided that the speech practice would end. To ensure the equivalent amount of practice in the two speech-practice conditions, the number of speech sessions in the Speech & Physical period was set at 14 as well.

During the 6 weeks of the Physical Only period, Tom attended a 1-hour session three times per week. Each of the three upper-limb tasks was presented in two 5-minute blocks. The order of the three tasks was always the same (i.e., Task 1 → Task 2 → Task 3 → Task 1 → Task 2 → Task 3). Throughout the practice session the robot presented video files of the physical therapist providing various instructions and remarks (e.g., “Thanks for coming in today,” “Clasp your hands together and reach out to touch my hand,” “Nice work! It’s time to take a break.”) Between tasks Tom took a 5-minute break where the uBot-5 played a short video clip of classical music or a TV show of Tom’s choice. Thus a 1-hour session typically consisted of 30 minutes of practice (= 5 minutes × 3 tasks × 2 repetitions), 25 minutes of break (= 5 minutes × 5 breaks), and 5 minutes for Tom to get ready for tasks and wrap up to leave.

During the 6-week Speech & Physical period, Tom attended a therapy session three times per week. The first 14 sessions included 30 minutes of speech practice and 1 hour of upper-limb exercises. The order of speech and physical therapies was counterbalanced across sessions: upper-limb exercises were completed first in odd-numbered sessions (e.g., Sessions 1, 3, 5) and speech practice was completed first in even-numbered sessions (e.g., Sessions 2, 4, 6). Between speech and physical therapy activities, Tom typically took a 5-minute break. The last four sessions of the Speech & Physical period only included 1 hour of physical therapy.

Throughout the therapy periods Tom primarily interacted with the uBot-5 during the practice sessions. The researcher quietly monitored Tom’s performance and did
not provide any feedback on Tom’s performance, except for general comments (e.g., “I appreciate that you completed today’s practice”).

Analysis

Daily probes of naming during the practice phase. As depicted in Figure 3, each target word was presented in six slides. Because the slides provided increasing levels of support for naming, Tom’s progress on target items was easily observed. During the practice the researcher recorded at which of the six slides Tom was able to produce the target word. The slide number (e.g., Slide # 3 that provided phonemic cue) quantified how much support Tom needed from the practice program to name the target item. This level of required support was monitored throughout the Speech Only period and the Speech & Physical period.

Daily probes of upper-limb functions during the practice phase. Tom’s performance in the three upper-limb exercises was measured by frequency (i.e., number of trials completed within a set duration)\(^2\) and range of motion (i.e., distance to the target). Task difficulty was regularly adjusted by moving the targets further from Tom. At the same time, feedback from Tom was used to ensure that the tasks were challenging but attainable with effort. An increase in the range of motion was directionally defined by three components (Figure 5): \(x\) – forward reach of the movement, \(y\) – height of the movement, and \(|z|\) – lateral reach of the movement in both directions. Tom’s progress in frequency and range of motion was documented during the Physical Only period and the Speech & Physical period.

Progress in naming over the five assessments. As delineated in Figure 1, Tom’s ability to name 60 target words was tested five times. Tom’s responses were scored using the 16-point scale (Table 2). The data were divided into three groups based on practice condition: 20 words practised during the Speech Only period, 20 words practised

---

\(^2\)The number of trials was automatically logged by the custom-built software that operated the uBot-5’s behaviours. At the same time a researcher manually tallied the number as a backup.
during the Speech & Physical period, and 20 words never practised. The three groups were separately analysed to detect the direct treatment effect on trained items, generalisation to untrained items, and maintenance of treatment effect. To determine significance of the gains, a statistical analysis compared baseline performance to scores in the four subsequent assessments. Given the non-parametric nature of the data (Table 2), Wilcoxon signed-ranks tests were utilised with a Bonferroni corrected alpha ($p < .0125$) for the multiple comparisons. In addition to the statistical analysis based on the score changes in the 16-point scales, the number of items Tom named spontaneously without the clinician’s support was counted. Such spontaneous responses received scores ranging from 16 to 12 (Table 2).

Progress in upper-limb functions over the four assessments. The Fugl-Meyer Assessment (FMA; Deakin et al., 2003; Fugl-Meyer et al., 1975) and the Wolf Motor Function Test (WMFT; Wolf et al., 1989, 2010) were used to measure the changes in Tom’s upper-limb functions.

Changes in quality of life over the five assessments. To document Tom’s perception on his quality of life, the Quality of Communication Life Scale (QCLS; Paul et al., 2004) and the Stroke and Aphasia Quality of Life Scale – 39 (SAQOL – 39; Hilari, Byng, Lamping, & Smith, 2003) were administered at the five assessments. Non-parametric data from the two scales were analysed using Wilcoxon signed-ranks tests with a Bonferroni corrected alpha ($p < .0125$) for the multiple comparisons.

RESULTS

Daily probes of naming during the practice phase

Figure 6 presents the average amount of support required by Tom at each practice session. He needed slightly less support (i.e., named more items spontaneously) during the Speech Only period than in the Speech & Physical period. Day-to-day fluctuations in Tom’s performances were noted especially during the Speech & Physical period. On certain days when speech practice followed the upper-limb exercises (Sessions 5,
7, 9, & 11), Tom required more support for naming than in the immediately previous sessions when speech practice preceded the upper-limb exercises (Sessions 4, 6, 8, & 10). These patterns of Tom’s progress during the practice periods suggest more efficient re-acquisition of target items in the sole condition and an adverse fatigue effect of physical tasks on speech practice in the sequential condition.

**Daily probes of upper-limb functions during the practice phase**

Figure 7 presents cumulative increases in the range of motion for the three tasks. Tom made notable gains in the height he could reach with his upper extremities. The average increase across the tasks was 14.9 cm during the Physical Only period and 6.4 cm during the Speech & Physical period. Additionally, the number of trials completed in each 5-minute block of the tasks was recorded. The average across the three tasks was 7.0/minute in the Physical Only period and 6.8/minute in the Speech & Physical period. These data on the frequency and range of motion together suggest that Tom made greater improvements in the range of motion without sacrificing the frequency of movements during the Physical Only period, as compared to the Speech & Physical period.

**Progress in naming over the five assessments**

Figure 8 depicts the data from the five assessments organised by practice condition. A visual inspection of the scores suggests treatment-specific gains during the two speech practice periods (Speech Only, Speech & Physical) with little generalisation. Table 3 summarises the results of Wilcoxon signed-ranks tests. Tom’s ability to name the 20 words practised during the Speech Only period significantly improved from baseline to Assessment 1 ($Z = 3.312, p = .001, r = .523$) and Maintenance Assessment ($Z = 2.625, p = .009, r = .415$). His gains in the 20 words practised during the Speech & Physical period were significant at Assessment 3 ($Z = 3.557, p < .001, r = .566$) and Maintenance assessment ($Z = 2.692, p = .007, r = .426$). Tom’s scores
Figure 8. Median naming scores of the words assigned to three speech practice conditions (Speech Only, Speech & Physical, Control – no practice) at five assessments. * indicates significant at $p < .0125$ (Bonferroni corrected alpha).

**TABLE 3**

Results of Wilcoxon Signed Tests comparing each assessment to baseline performance

<table>
<thead>
<tr>
<th>Practice condition</th>
<th>Assessments</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Only</td>
<td>$Z = 3.312$</td>
<td>$Z = 1.802$</td>
<td>$Z = 1.807$</td>
<td>$Z = 2.625$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = .001^*$</td>
<td>$p = .072$</td>
<td>$p = .071$</td>
<td>$p = .009^*$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = .524$</td>
<td>$r = .285$</td>
<td>$r = .286$</td>
<td>$r = .415$</td>
<td></td>
</tr>
<tr>
<td>Speech &amp; Physical</td>
<td>$Z = 0.677$</td>
<td>$Z = 0.962$</td>
<td>$Z = 3.557$</td>
<td>$Z = 2.692$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = .498$</td>
<td>$p = .336$</td>
<td>$p &lt; .001^*$</td>
<td>$p = .007^*$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = .105$</td>
<td>$r = .152$</td>
<td>$r = .566$</td>
<td>$r = .426$</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>$Z = 1.069$</td>
<td>$Z = 1.134$</td>
<td>$Z = 0.862$</td>
<td>$Z = 2.658$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$p = .285$</td>
<td>$p = .257$</td>
<td>$p = .389$</td>
<td>$p = .008^*$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r = .169$</td>
<td>$r = .179$</td>
<td>$r = .136$</td>
<td>$r = .420$</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at $p < .0125$ (Bonferroni corrected alpha).

on the 20 control items, which were not practised during the study periods, were significantly better at Maintenance Assessment than at baseline ($Z = 2.658, p = .008, r = .420$).

Figure 9 presents the number of spontaneous responses in each practice condition over the five assessments, as well as incremental changes in the total number of spontaneous responses in each therapy period. Similar to the statistical analysis stated above, this examination of spontaneous responses suggests treatment-specific gains, little generalisation during therapy periods, and maintenance of treatment effect and delayed generalisation at 4 weeks post-treatment.

Progress in upper-limb functions over the four assessments

The FMA scores showed little changes over the four assessments (Table 4). Tom consistently completed the same 3 out of the 15 WMFT tasks over the assessments. The three
Figure 9. Number of spontaneous responses to items assigned to three speech practice conditions (Speech Only, Speech & Physical, Control – no practice) at five assessments. The number in parentheses indicates the incremental change in total spontaneous responses during each therapy period.

TABLE 4
Results of the Fugl-Meyer Assessment: Scores in total and in the subsections

<table>
<thead>
<tr>
<th>Assessments</th>
<th>Baseline</th>
<th>2</th>
<th>3</th>
<th>Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (/66)</td>
<td>31</td>
<td>30</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>Shoulder/Elbow/Forearm (/36)</td>
<td>18</td>
<td>16</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Wrist (/10)</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Hand (/14)</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Coordination/Speed (/6)</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

tasks were: placing the forearm on the table, placing the hand on the table, and pulling a 1-lb weight across the table. The mean response time of the tasks fluctuated as shown in Figure 10. Tom’s movements became notably faster over the Physical Only period, slower over the Speech & Physical period, and again faster over the 4-weeks of No

Figure 10. Average response time of completed tasks in the Wolf Motor Function Test at four assessments.
Therapy. This pattern of progress in upper-limb functions mirrors his improvements in naming ability: more gains in the sole condition than in the sequential condition.

Changes in quality of life over the five assessments

Figures 11 and 12 present Tom’s responses to the QCLS and the SAQOL – 39 respectively. Compared to baseline, Tom’s responses to the QCLS was significantly higher at Assessment 1 ($Z = 2.714, p = .007, r = .480$) and Maintenance Assessment ($Z = 2.970, p = .003, r = .525$). When the incremental changes during each therapy period were evaluated, significant gains were observed only during the Speech Only period. Tom’s responses to the SAQOL – 39 significantly improved from baseline to the Assessment 2 ($Z = 2.677, p = .007, r = .303$), Assessment 3 ($Z = 3.129, p = .002, r = .354$), and

![Figure 11. Overall quality of communication life (QCL) scores and general quality of life (QOL) scores from the Quality of Communication Life Scale (Paul et al., 2004).](image)

![Figure 12. The mean scores of the four sub-domains (physical, communication, psychosocial, energy) and the overall mean scores from the Stroke and Aphasia Quality of Life Scale – 39 (Hilari et al., 2003).](image)
Maintenance Assessment ($Z = 2.696, p = .007, r = .305$). Incremental gains were significant only over the Physical Only period ($Z = 2.684, p = .007, r = .304$). These data from the QCLS and the SAQOL – 39 indicate that the most notable gains in Tom’s subjective perception of communicative life occurred during the Speech Only period and that the most notable gains in his perception of overall quality of life occurred during the Physical Only period.

Tom’s and his wife’s perception of the robot-mediated therapy

Upon completing Assessment 3, Tom and his wife independently responded to a 23-item questionnaire on the robot-mediated therapy. It used 5-point scales for responses with 1 being the most positive and 5 being the most negative. The items were grouped into four subsections: Overall Experience (e.g., I enjoyed working with the robot; Yes = 1, No = 5), Feelings and Impressions (e.g., Working with the robot was ...; Pleasant = 1, Unpleasant = 5), Comments on the Robot (e.g., I like that the robot has arms; Like = 1, Dislike = 5), and Future Plans (e.g., I would come back for more speech and physical therapies with the robot; Yes = 1, No = 5). The median scores of Tom’s and his wife’s responses were both 1, indicating overall positive perception of the robot-mediated intervention. One of the items included in the Overall Experience subsection was I would rather work with clinicians than with the robot (No = 1, Yes = 5). To this item, Tom gave 2, and his wife gave 3.

DISCUSSION

Tom, a stroke survivor chronically experiencing aphasia and right-sided hemiparesis, completed a robot-mediated rehabilitation program in the sole condition (i.e., only speech therapy for 4 weeks and then only physical therapy for 6 weeks) and in the sequential condition (i.e., speech and physical therapies for 5/6 weeks). Tom’s naming scores at the five assessments demonstrate practice effects on trained items during the therapy periods, and maintenance of the gains and delayed generalisation at 4-week post-treatment. Tom made notable gains in the frequency and range of the upper-limb movements. Tom’s responses to the Quality of Communication Life Scale (Paul et al., 2004) and the Stroke and Aphasia Quality of Life Scale – 39 (Hilari et al., 2003) suggest significant improvements in perceived quality of life during the practice periods and at 4-week post-treatment.

To the authors’ knowledge this study was the first attempt to implement a humanoid robot in delivering speech and physical therapy services to stroke patients. The objective data on speech and physical functions and the subjective data on quality of life that have been documented by the current study warrant further investigations of utilising a humanoid robot in stroke rehabilitation. If future research supports feasibility and utility of certain therapy robots, potential placements of such robots at clinical and residential settings can enhance the availability and accessibility of intensive and interactive therapy services, especially for patients residing in rural areas.

The main objective of the current project was to investigate the interaction between speech and physical therapies. Data from this single case suggest a competitive interaction between the two domains. Tom mastered target items for verbal naming more efficiently and improved frequency and range of motion to a greater extent in the sole condition (Speech Only, Physical Only) than in the sequential condition (Speech & Physical). Daily progress in his naming ability during the Speech & Physical period
suggests a fatigue effect (Figure 5). The significant practice effect obtained during the Speech Only period vanished during the Physical Only and Speech & Physical periods and reappeared at 4-week post-treatment (Table 3, Figure 8, Figure 9). Similarly, the notable gain in the speed of movement achieved during the Physical Only period was not maintained during the Speech & Physical period but reappeared at 4-week post-treatment (Figure 10). This advantage of the sole therapy schedule over the sequential schedule suggests that speech and physical functions competed for limited resources.

The adverse effect of upper-limb exercises on subsequent naming tasks observed during the Speech & Physical period (Figure 5) is consistent with the fatigue effect that stroke survivors demonstrated in speech-language performances after 30 minutes of lower-extremity exercises (Marshall & King, 1973). In addition, the significant performance decline in communicative ability from morning sessions to afternoon sessions (Marshall, Tompkins, & Phillips, 1980) can likewise be attributed to accumulated fatigue that stroke survivors experience from their daily activities. The current study has replicated these previous findings and expanded them by demonstrating that the fatigue experienced during practice can ultimately affect the treatment outcomes in speech-language and physical functions. It is likely that several factors escalated the mental and physical fatigue and, as a result, amplified the competitive interaction between speech and physical therapies. Such factors include the participants’ age, severity, and type of speech program.

Advanced age is strongly associated with reduced cognitive capacity (Gilchrist, Cowan, & Naveh-Benjamin, 2008), sensorimotor decline (Lindenberger & Baltes, 1994; Wohler & Smith, 1998), and inefficiency in learning new skills (Daselaar, Rombouts, Veltman, Raaijmakers, & Jonker, 2003). The participant of the current study was a 72-year-old male, whose reduced capacity in cognitive and motor functioning presumably limited the level of recoveries in speech and physical functions. It is likely that the adverse effect of age-related deterioration was amplified by the more demanding therapy schedule of the sequential condition, as compared to the sole condition. The hypothesised effect of age-related capacity decline warrants a research study that compares younger and older patient groups in their performances in sole and sequential therapy schedules.

Severity is another factor that likely contributed to the competitive interaction between speech and physical therapies. The initial assessments revealed that Tom presented with moderate aphasia, mild verbal apraxia, and moderate-to-severe hemiparesis. It is possible that patients with milder deficits in speech and physical functions may not experience the cross-domain competition.

The speech practice program used in this study provided increasing levels of support (i.e., least-to-most cues) for naming and promoted effortful learning of target items. Although effortful learning is as beneficial as errorless learning (Fillingham, Sage, & Ralph, 2005; McKissock & Ward, 2007), the effort exerted in word retrieval and speech production may consume more resources for rehabilitation and cause a higher level of fatigue. Thus a practice program that provides decreasing levels of support (i.e., most-to-least cues) can reduce the competitive interaction between speech and physical therapies. Yet another alternative is a lexical-perceptual training program that requires no effort for overt naming and that has been shown to improve verbal naming (Fridriksson et al., 2009). Eliminating motor requirements for speech practice (Fridriksson et al., 2009) can further reduce overall fatigue and enhance treatment outcomes.
Potential effects of these participant characteristics (i.e., age, severity) and task characteristics (i.e., effortful vs errorless vs no articulation) should be investigated in future research. Moreover, therapy activities can be further refined in future studies. In the current study the uBot-5 dynamically interacted with Tom during physical therapy sessions whereas it remained static during speech therapy sessions. In a follow-up study the robot can be programmed to play a board game that incorporates the speech practice program. The interactive exchange with the humanoid robot can enhance the participant's motivation to complete therapy tasks.

The current robotics technology enabled the uBot-5 to deliver therapy activities in a highly consistent and structured manner across sessions and across treatment conditions. Because of this consistency, objective data of the participant’s progress was easily and instantly collected. At the same time, the uBot-5 was lacking the flexibility that a human clinician would demonstrate in response to Tom’s verbal and non-verbal reactions to therapy activities. This lack of flexibility and adaptability to the participant’s immediate needs might have contributed to an increased level of fatigue. To overcome this shortcoming, next generation of therapy robots need to be equipped with sophisticated sensor technology, such as a computer vision system that monitors facial expression, eye gaze, and head posture (Ji, Zhu, & Lan, 2004; Lang, Wachsmuth, Wersing, & Hanheide, 2010; Sanchez et al., 2011; Tian, Kanade, & Cohn, 2001). This study took a multidisciplinary approach in stroke rehabilitation and observed a competitive interaction between speech and physical therapies. The treatment schedule of the present study reflected routine clinical practice where speech and physical therapies would be separately provided in back-to-back sessions. As a result, the average duration of an entire session was 30–60 minutes for the sole condition and 90 minutes for the sequential condition. To eliminate this confounding factor of session length and, more importantly, to examine the effect of simultaneously provided therapies, a future study could take an interdisciplinary approach by embedding speech-language tasks into physical exercises. A participant could be required to reach his/her impaired arm for the picture of a target object and simultaneously say the name of the object. A potential comparison between back-to-back multidisciplinary sessions and a single interdisciplinary session will yield useful information with clinical implications.

REFERENCES

Dabul, B. (2000). Apraxia Battery of Adults (2nd ed.). Austin, TX: Pro-Ed.


