APHASIOLOGY, 2004, *18* (5/6/7), 521–542

Using resource allocation theory and dual-task methods to increase the sensitivity of assessment in aphasia

Malcolm R. McNeil, Patrick J. Doyle, William D. Hula, and Hillel J. Rubinsky

VA Pittsburgh Healthcare System Geriatric Research Education & Clinical Center and University of Pittsburgh, PA, USA

Tepanta R. D. Fossett

University of Pittsburgh, PA, USA

Christine T. Matthews

VA Pittsburgh Healthcare System Geriatric Research Education & Clinical Center and University of Pittsburgh, PA, USA

Background: Quantifying the severity of language impairment and measuring change in language performance over time are two important objectives in the assessment of aphasia. The notion of cognitive effort as understood from a resource allocation perspective provides a potentially useful complement to traditional constructs employed in aphasia assessment.
Aims: The series of experiments described in this paper used resource allocation theory and dual-task methodology (1) to assess whether a language comprehension task (Story Retell Procedure) and a visual-manual tracking task trade performance under dual-task conditions, and (2) to investigate the potential utility of these methods in clinical assessment of aphasia. In Experiment 1, the validity of a difficulty manipulation of the SRP was investigated. In Experiments 2 and 3, the reliability and validity of the visual-manual tracking task were evaluated. Experiment 4 investigated whether the two tasks trade performance under dual-task conditions.
Methods & Procedures: In Experiment 1, 20 normal participants listened to and retold stories presented by a normal speaker and speakers with mild, moderate, and severe aphasia. Participants' comprehension performance was measured by calculating the amount of information retold per unit time. In Experiment 2, root mean square (RMS) tracking error data were collected under fixed joystick displacement conditions. In Experiment 3, 20 normal participants performed single-task tracking across 12 trials at each of three difficulty levels, and performance was evaluated in terms of RMS error. In Experiment 4, three groups of 20 normal individuals performed the tracking task while listening to stories told by the normal speaker and speakers with aphasia. Story retell performance was evaluated between subjects across three tracking difficulty levels and tracking performance was evaluated within subjects across story difficulty (normal, mild, moderate, and severe aphasia).
Outcomes & Results: The results of Experiments 1–3 supported the reliability and validity of the difficulty manipulations for the story retell and tracking tasks. In Experiment 4, tracking performance was found to vary significantly across story difficulty, with subjects demonstrating better tracking performance while listening to stories told by a mildly aphasic

Address correspondence to: Malcolm R. McNeil, Professor and Chair, Department of Communication Science & Disorders, University of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260, USA. Email: mcneil@csd.pitt.edu

http://www.tandf.co.uk/journals/pp/02687038.html DOI: 10.1080/02687030444000138

speaker than during stories told by a speaker with moderate aphasia. There was no effect of tracking difficulty on story comprehension as measured by subsequent story retell performance.
Conclusions: The results provide qualified support for both a resource allocation view of language performance in normal individuals and the potential utility of these methods in the assessment of aphasia. These conclusions, however, are mitigated by the finding of only a unidirectional (as opposed to bidirectional) performance trade, and by the fact that the effect of story difficulty on tracking performance was observed across only two levels of aphasia severity.

One of the significant challenges of clinical aphasiology is the task of measuring severity of deficits and change in the language performance of persons with aphasia in terms that are meaningful to them and to their communication partners. Measures that do this typically include the assessment of spoken language as one component. Language samples are typically obtained through such procedures as stimulus repetition (Porch, 1981), picture description tasks (Kertesz, 1982), personal and procedural narratives (Nicholas & Brookshire, 1993), structured interviews (Goodglass, Kaplan, & Barresi, 2001), retelling of fables (Rochon, Saffran, Berndt, & Schwartz, 2003; Saffran, Berndt, & Schwartz, 2003), video narration (Dollaghan, Campbell, & Tomlin, 1990), and discourse simulations (Doyle, Thompson, Oleyar, Wambaugh, & Jackson, 1994). Performance is quantified in a variety of ways; including behaviour counts (Nicholas & Brookshire, 1993, 1995), rating scales (Goodglass et al., 2001), questionnaires (Sarno, 1969), and multidimensional scoring systems (Holland, 1980; Porch, 1981). The nature and severity of the spoken language deficit is judged based on these pre-determined behaviours. As part of these judgements each of the pre-selected behaviours is either assumed to be of equal importance and summed to form a frequency of error type, or is assigned a weight or value as in multidimensional rating scales. The untested assumptions about the individual contributions of error types threaten the interpretation of test results, because of their unknown impact on such constructs as information transfer, overall communication handicap, and the burden placed on communication partners of individuals with aphasia.

Some assessment instruments seek to measure communication handicap by having clinicians or communication partners rate the burden placed on communication partners (Goodglass et al., 2001), compare the pre- and post-morbid ability of the person with aphasia to perform in certain communication situations (Lomas, Pickard, Bester, Elbard, Finlayson, & Zoghaib, 1989), or rate the amount of assistance required by the person with aphasia to accomplish certain tasks (Frattali, Thompson, Holland, Wohl, & Ferketic, 1995).

An alternative way of approaching this measurement problem is to attempt to quantify more directly the effort expended by interlocutors when they communicate with persons with aphasia. Such a procedure would be particularly useful for demonstrating and quantifying change in the level of communication handicap experienced by speakers with aphasia in instances where traditional measures of such constructs as auditory comprehension or information transfer are insensitive. These traditional measures may be insensitive for a variety of reasons. All people, including partners of individuals with aphasia, bring to the task of communication a variety of strengths, weaknesses, and strategies. Such factors as hearing loss, familiarity with a given person's speech, and motivation conspire to affect information transfer in multifarious ways, make tenuous the assumption that treatment of aphasic language impairment will benefit all of the patient's interlocutors to the same extent. A procedure that would allow reliable measurement of the effort expended by communication partners, in addition to information transfer, would have the added benefit of taking into consideration the strengths, weaknesses, and strategies emerging from the communication dyad tested. It might also prove more sensitive to severity differences among patients and more responsive to change in either member of the dyad.

The resource allocation view of attention (Hirst & Kalmar, 1987; Kahneman, 1973; McNeil, Odell, & Tseng, 1991; Navon & Gopher, 1979) provides one of the more enduring and useful frameworks for approaching effort as a psychological construct. Within resource allocation models, the terms attention, processing resources, capacity, and cognitive effort are used interchangeably to refer to a source of fuel or activation for cognitive operations that can be flexibly allocated within and among processing domains. Kahneman (1973), in his early, influential formulation of the theory, proposed that attention (or effort) is limited in capacity and that its availability and allocation are influenced by a variety of factors including arousal, momentary and enduring dispositions, goals, priorities, and evaluation of task demands. The various resource allocation models differ in many respects, most notably on the issue of whether there is a single, undifferentiated reservoir of attentional resources (Kahneman, 1973) or whether there are multiple pools that are more or less dedicated to specific processes or domains (Gopher, Brickner, & Navon, 1982; Navon & Gopher, 1979, 1980; Wickens, 1984). The feature that they all share, however, is the notion of attention as a capacity that can be distributed among cognitive processes in a graded fashion.

Most studies supporting or employing resource allocation models have been dual-task experiments in which decrements in the performance of one task have been taken as an indicator of processing load incurred by a second, concurrently performed task (Arvedson & McNeil, 1987; Brown, 1978; Campbell & McNeil, 1985; Erickson, Goldinger, & LaPointe, 1996; Gopher et al., 1982; Granier, Robin, Shapiro, Peach, & Zimba, 2000; Hirst & Kalmar, 1987; McLeod, 1977; Murray, 2000; Murray, Holland, & Beeson, 1997a, 1997b, 1998; Navon, 1990; Payne, Peters, Birkmire, Bonto, Anastasi, & Wenger, 1994; Slansky & McNeil, 1997; Tseng, McNeil, & Milenkovic, 1993; Wickens, 1976, 1986). Dual-task experiments can be grouped into three broad categories, depending on the specific methods involved. The simplest method, single-to-dual task comparison, requires that the task(s) of interest be performed in isolation and concurrently with a secondary task. Performance decrements should be observed in the dual-task condition and when they are, sharing of a limited-capacity resource is typically inferred. This subtraction method has been productively employed in investigations of language processing in aphasia (Erickson et al., 1996; Murray, 2000; Murray et al., 1997a, 1997b, 1998), as well as of attention in general (Wickens, 1976), but it has substantial shortcomings when used in isolation to provide evidence for resource sharing. Among these are the assumptions that concurrent task demands are linear and additive combinations of the separate single-task demands and that equal attentional capacity is available in both single and dual-task conditions (Navon & Gopher, 1979). According to Kahneman (1973), the quantity of resources a person has available may fluctuate from moment to moment according to factors that include task demands. If one proposes that a person might recruit additional resources in a more demanding dual-task condition than in a less demanding single-task condition, then inferring secondary task resource demands from primary task performance decrements is problematic at best. At the very least, a failure to observe decrements in the dual-task condition should not necessarily lead to the conclusion that the two tasks do not share processing resources.

A second and even more problematic assumption of the single-to-dual-task comparison method holds that the structures and processes recruited for single-task performance are the same as those recruited during dual-task conditions (Navon & Gopher, 1979). This assumption, which is similar to the one that underlies the subtraction method in functional imaging studies, holds that the set of structures and processes utilised in dual-task performance is equivalent to the combination of the sets used in the performance of each task in isolation. It implicitly denies consideration of any qualitative changes in how the cognitive architecture is mobilised to complete the tasks concurrently versus in isolation. Primarily because of these two suspect assumptions, attempts to quantify effort or gather data supporting resource models of attention using only the single-to-dual-task subtraction method are problematic and unlikely to be successful.

A second dual-task procedure, the voluntary effort allocation method, avoids making the problematic assumptions discussed above by having subjects perform both tasks in all experimental conditions. In this method, subjects perform two concurrent tasks with explicit instructions to vary their allocation of effort between them according to the experimental condition (Arvedson & McNeil, 1987; Gopher et al., 1982; Matthews & Margetts, 1991; Navon, 1990; Slansky & McNeil, 1997; Wickens & Gopher, 1977). These instructions may take the form of allocation ratios stated as percentages of total effort, qualitative instructions to vary emphasis, or relative performance targets. For example, in one condition a subject may be instructed to give 75% effort to Task A and 25% to Task B, and then to give 50% effort to both tasks in another condition, and finally to give 25% effort to Task A and 75% to Task B. Changing allocation ratios in this way is intended to induce a trade-off between the two tasks, and performance on one may be plotted on coordinate axes against performance on the other, resulting in a performance operating characteristic or curve (POC). In theory, the POC describes the limits of joint performance on the two tasks, given the assumption that all available resources are allocated between the two tasks (Navon & Gopher, 1979).

While the voluntary effort allocation method represents an improvement over the single-to-dual-task subtraction method, it has been criticised by Navon (1984) on the grounds that observed performance trading between the two tasks may be the result of subject biases and attempts to please the experimenter by meeting performance expectations.

Given the limitations of the single-to-dual-task subtraction and voluntary effort allocation methods reviewed above, a third dual-task procedure has been described in the literature and used to investigate resource models of attention. This method is referred to here as the concurrent task difficulty manipulation method. In this procedure, subjects perform two concurrent tasks and the difficulty (or some other parameter) of each task is systematically and independently manipulated (Campbell & McNeil, 1985; Hirst & Kalmar, 1987; McLeod, 1977; Payne et al., 1994; Wickens, 1986; Wickens, Kramer, Vanasse, & Donchin, 1983). This method may be used in concert with the voluntary effort allocation method, in which case it assumes the same limitations, or it may be used exclusively with equal priority instructions. Assuming that the combined demand of the two tasks challenges the available supply of resources, and that these resources are shared, increasing the difficulty of one task should cause a performance decrement in both tasks. Of course the most interesting and useful effect is the effect of a Task A difficulty manipulation on the performance of Task B. Within a single resource model, this performance trade should be bidirectional, i.e., Task A manipulation should affect performance on Task B and vice versa (Kahneman, 1973). Within a multiple-resource model, a bidirectional effect will be observed only if the difficulty manipulations employed for both tasks affect resource pools utilised by both tasks (Navon & Gopher, 1979, 1980).

Within this theoretical framework, Doyle and McNeil (1998) developed the Resource Allocation Paradigms of Pittsburgh (RAPP). RAPP is a dual-task software environment designed to assess the relative processing resources (cognitive effort) utilized by individuals in processing and comprehending the spoken language of persons with varying competencies and performance levels, such as those with and without aphasia.

The current study had two primary goals, one theoretical and the other applied. First, it sought to investigate whether a resource view of attention could be productively used to describe the performance of non-brain-injured subjects in the domains of language and visual-manual tracking. Second, a method for quantifying the effort expended by normal listeners in comprehending stories told by speakers with varying degrees of aphasia was assessed. To this end, a story comprehension task and a visual-manual tracking task were employed in a dual-task procedure in which the difficulty of each task was independently manipulated. The investigation involved four experiments, each of which is described below. Experiments 1, 2, and 3 describe the rationale for the task choices, as well as experiments designed to validate their respective difficulty manipulations, and Experiment 4 details the methods and findings of the dual-task investigation, which used the two previously validated single tasks.

EXPERIMENT 1
VALIDATING DIFFERENT SEVERITY LEVELS IN APHASIC STORY RETELLS

The rationale for selecting auditory discourse comprehension as the language task of interest in this investigation was based on several factors. First, when people communicate in everyday life, they do so most often in multiple-sentence messages that occur within some context. Listening to and comprehending connected speech is a relatively continuous task that is a closer approximation to naturally occurring discourse than comprehending isolated sentences or words, and the latter may not be predictive of the former in some circumstances (Brookshire & Nicholas, 1997).

Second, overt motor responses are not required for auditory comprehension to proceed normally, as they are for performing language production tasks. This ability for task processing to occur without the need for overt on-line responses is desirable in designing a dual-task experiment because it minimises the opportunity for structural and computational, as opposed to capacity, interference (Kahneman, 1973; McNeil et al., 1991). This distinction, which is discussed in more detail below in Experiment 2, refers to dual-task performance decrements that occur when two tasks compete for physical structures or specific and dedicated mental computations, as opposed to those resulting from competition for shareable attention capacity. The choice of the auditory language comprehension and visual-manual tracking tasks was based on the assumption that they would eliminate structural and computational interference while allowing the opportunity to observe capacity interference.

The connected language comprehension stimuli selected for this study were the stories from the Discourse Comprehension Test (DCT: Brookshire & Nicholas, 1997). These stories are controlled for a number of important variables, including number of words, number of sentences, mean sentence length, number of subordinate clauses, number of T-units (''one main clause with all the subordinate clauses and nonclausal phrases attached to or embedded in it''; Paul, 2001, p. 514), ratio of clauses to T-units, listening difficulty, and number of unfamiliar words. Originally, an attempt was made to develop multiple-choice questions for each story as an off-line measure of information transfer that would be more sensitive than the yes-no questions published with the DCT, but preliminary findings indicated that they were psychometrically inadequate (McNeil & Doyle, unpublished data). An alternative procedure that requires subjects to provide verbal reproductions of the stories was then developed. In this Story Retell Procedure (SRP) (McNeil, Doyle, Fossett, Park, & Goda, 2001), the verbal reproductions are scored for information content and efficiency using the *percent information unit per minute* (%IU/Min) (McNeil, Doyle, Park, Fossett, & Brodsky, 2002) a metric derived from Nicholas and Brookshire's *correct information unit* (CIU) (Brookshire & Nicholas, 1997) for use specifically with these stimulus stories. It should be noted that the SRP uses not only the 10 stories comprising the DCT as originally published, but also the 2 practice stories, for a total of 12 stimulus stories.

The conceptual and psychometric development of the resulting instrument, the Story Retell Procedure (SRP) and its associated metric, the %IU/Min, has been described in a series of recent publications. Evidence has been presented to support the SRP's validity as a language sampling procedure and the linguistic equivalence of four alternate forms, comprising three stories each (Doyle et al., 2000). Further, the %IU/Min has been demonstrated to have acceptable alternate forms and inter-rater reliability (Hula, McNeil, Doyle, Rubinsky, & Fossett, 2003; McNeil et al., 2001), concurrent validity with traditional measures of both verbal production and auditory comprehension (McNeil et al., 2001), and the ability to discriminate between normal speakers and persons with aphasia with reasonable sensitivity (McNeil et al., 2002).

Experiment 1 was designed to establish the SRP as a language task whose difficulty could be manipulated by having the stimulus stories presented by speakers with a range of language impairment from none to severe aphasia. This modification of the SRP was accomplished by selecting three speakers with aphasia from the subject sample of a prior validation study (McNeil et al., 2001) to represent mild, moderate, and severe degrees of aphasia. The aphasia severity categorisations were based on their overall scores on the Porch Index of Communicative Ability (PICA: Porch, 1981) and Revised Token Test (RTT: McNeil & Prescott, 1978) and further validated perceptually by 20 normal listeners' direct magnitude estimation judgements (McNeil et al., 1999). The story retells produced by these speakers with aphasia, as well as readings of the original DCT stories by a normal speaker, were grouped into the four three-story forms that had previously shown psychometric equivalence (Doyle et al., 2000; McNeil et al., 2001). These story forms from these four speakers constituted the stimuli for the present experiment.

The specific objective of this experiment was to further validate the different difficulty levels of the stories that would serve as stimuli in the subsequent dual-task experiment. In other words, it was asked whether or not normal listeners demonstrate reduced story comprehension, as indexed by %IU/Min produced in retells of stories told by a normal speaker and speakers with aphasic language impairments of different severities.

Method

Participants. A total of 20 temporarily able-bodied subjects ranging in age from 42 to 74 years (mean = 55, SD = 10) participated in the study. All participants met the following selection criteria: Negative self-reported history of neurological, communication, or psychiatric disorders; age between 40 and 75 years, passing a pure tone hearing screening at 35 dB HL in one ear at .5, 1, 2, and 4 KHz; 20/80 vision or better measured with the reduced Snellen chart; performance above the 5th percentile for normal individuals on the Short Porch Index of Communicative Ability (SPICA: DiSimoni, Keith, & Darley, 1980) and the Revised Token Test (RTT) (McNeil & Prescott, 1978); performance at or above the 5th grade level on the reading subtest of the Wide Range Achievement Test-3 (Wilkinson, 1993); no greater than two points decline from the Immediate to the Delayed Story Recall Task from the Arizona Battery of Communication Disorders of Dementia) (ABCD: Bayles & Tomoeda, 1993); a minimum of 12 years of education. Subjects were recruited from the local community by means of fliers posted in public places, presentations to senior citizens' organisations, and by word of mouth. Subjects were paid $15 for their participation.

Procedure. The stimuli consisted of four different story forms, consisting of three stories each, which were produced by each of four different speakers: one normal speaker and three with varying degrees of aphasia. As described above, the speakers with aphasia were categorised as having mild, moderate, and severe aphasia according to standardised test scores and these rankings were validated by direct magnitude estimation (McNeil et al., 1999). Table 1 presents the clinical characteristics of the speakers with aphasia and Table 2 presents the average %IUs and %IU/Min present in the stories at each difficulty level. Each story form was quasi-randomly assigned such that no subject heard a given form more than once, and each subject heard one form produced by each of the four speakers. Following presentation of each story, subjects were instructed to immediately retell the story in their own words, following the standardised procedures of the SRP (McNeil et al., 2001). The resulting language samples were digitally recorded and later scored for %IU/Min, also using standardised procedures. An IU was defined as ''...an identified word, phrase, or acceptable alter-

TABLE 1
Clinical characteristics of the speakers with aphasia who provided story retell stimuli for Experiments 1 and 4

Subject	*Age*	*MPO*	*RTT %ile*	*RCPM*	*PICA OA %ile*	*PICA Verbal %ile*
Mild	52	17	92	36	87	91
Moderate	55	30	63	32	75	71
Severe	71	94	5	22	43	37

TABLE 2
Information content and efficiency of stimulus stories in percent information units (%IU) and percent information units per minute (%IU/Min)

	%IU	*%IU/Min*
Normal	100	64
Mild aphasia	48	40
Moderate aphasia	23	11
Severe aphasia	13	8

native from the story stimulus that is intelligible and informative and that conveys accurate and relevant information about the story'' (McNeil et al., 2001, p. 994). To derive the %IU/Min score for each retell, the number of IUs was first tallied, and then divided by the number of IUs available in the original story (as published in the *DCT*), giving %IU.[1] This measure of information transfer was then divided by the number of minutes taken to produce the retell, giving %IU/Min, a metric of the efficiency of information transfer.

Results

A significant main effect [$F(3, 57) = 162.27$; $p < .01$] of story form was obtained using a single-factor repeated measures ANOVA. Post-hoc pair-wise comparisons using the Bonferroni adjustment showed significant differences between the mild, moderate, and severe story forms generated by the persons with aphasia, as well as between the normal, moderate, and severe story forms. No significant difference was found between the normal and mild aphasic story forms. The %IU/Min produced by each subject after listening to each of the story levels are presented in Table 3.

Discussion

Three levels of story difficulty were validated by this experiment. There was a significant difference between the mild, moderate, and severe story forms, as well as the normal, moderate, and severe story forms, according to the %IU/Min measure of information transfer efficiency. As such, this experiment demonstrated that the difficulty of the SRP may be varied across three levels by presenting stimulus stories produced by speakers with differing severity of aphasic language impairment.

The lack of a significant difference between the normal and mild story forms may be attributable to the fact that the mildly aphasic story productions were not very impaired in terms of grammar and organisation. Consider data taken from an earlier study on the SRP (Doyle et al., 2000) in which these speakers with aphasia served as subjects: 96% of the clauses in the mild stories were produced correctly, compared to 100%, 77%, and 75% for the normal, moderate, and severe stories, respectively. Also, the mild stories contained only 3.6 mazes per minute—words or partial words that are unintelligible in a known context (e.g., He went to the ''frangus''), nonword fillers (e.g., um, er, uh), repetitions, revisions, and word fragments—compared to 0, 6.8, and 10.7 mazes per minute for the normal, moderate, and severe stories.

[1] The story retell data were also analysed by calculating %IU/Min relative to the number of information units available in the different stimulus stories produced by the speakers with different severity levels of aphasia (and the normal speaker). These analyses resulted in a different rank ordering of performance across conditions with subjects scoring more relative %IU/Min in the Moderate and Severe conditions. This occurred because, when subjects were presented with less information, they were able to recall a greater proportion of it. In some cases subjects' retells even contained more information than was in the aphasic-produced stimulus story because of their ability to fill in omitted information from context. We elected not to pursue these analyses because they produced unpredictable patterns of results that were unlikely to assist in the interpretation of data from the subsequent dual-task experiment (Experiment 4). This speculation was subsequently borne out, as the re-ordering of the story difficulty conditions along the lines of this relative %IU/Min analysis would not have helped to clarify the effects of story difficulty on tracking performance, and would not have changed the nonsignificant effect of tracking difficulty on story retell performance.

TABLE 3
%IU/Min produced by individual subjects at each story difficulty level in Experiment 1

Subject	*Normal*	*Mild aphasia*	*Moderate aphasia*	*Severe aphasia*
1	51.7	47.6	35.1	15.5
2	50.3	44.2	25.5	16.7
3	42.1	50.9	25.0	11.3
4	57.8	50.1	36.7	12.2
5	66.4	53.4	40.8	27.6
6	52.1	59.9	32.6	18.2
7	43.0	55.7	38.0	18.6
8	61.0	65.8	41.3	11.9
9	46.1	39.8	27.4	14.3
10	50.6	50.5	39.4	14.7
11	44.9	54.4	36.7	17.8
12	32.2	49.7	36.5	9.9
13	51.5	50.5	39.8	13.4
14	41.9	38.9	26.3	19.0
15	49.2	48.8	36.5	24.1
16	42.0	40.0	22.7	17.6
17	63.2	43.7	43.1	19.8
18	53.2	54.3	33.0	19.9
19	52.2	45.5	23.8	15.3
20	51.4	41.9	33.7	20.7
Mean	50.14	49.28	33.70	16.93
SD	8.12	6.94	6.40	4.38

EXPERIMENTS 2 AND 3: RELIABILITY OF A VISUAL-MANUAL TRACKING TASK

The literature on attention and motor control contains many dual-task studies in which visual-manual tracking has been used to induce or measure processing load in a variety of concurrent tasks, including non-verbal auditory discrimination (Backs, 1997; McLeod, 1977; Wickens, 1976), auditory verbal memory search, (Payne et al., 1994), typing (Gopher et al., 1982), mathematical reasoning (McLeod, 1977; Payne et al., 1994), speech production (Lively, Pisoni, Van Summers, & Bernacki, 1993), and auditory sentence processing (Granier et al., 2000). Visual-manual tracking was chosen as a concurrent task in the present dual-task investigation for multiple reasons. First, the continuous nature of the task makes it suitable for inducing and measuring processing load in a task requiring auditory comprehension of connected speech such as the SRP. More discrete concurrent tasks, such as visual lexical decision or form discrimination, could potentially invite subjects to allocate resources exclusively to the auditory comprehension task during inter-stimulus intervals, and to only periodically share capacity between the tasks (McLeod, 1977).

Second, visual-manual tracking has previously been shown to trade performance, and thus presumably share resources with auditory language tasks. Granier and colleagues (2000) found differences in tracking performance associated with the time course of sentence processing, with subjects demonstrating poorer tracking performance during the beginning and the end of subject-verb-object sentences than during the middle, and performance that was poorer still during post-sentence processing and answering yes-no comprehension questions. They concluded that visual-manual tracking was sensitive to the cognitive load imposed not only by off-line sentence comprehension processes, but also to demands incurred by (presumably) less automatic and more integrative on-line processes. Payne and colleagues (1994) demonstrated a performance trade between an unstable manual tracking task and an auditory Sternberg task, which required subjects to recognise spoken words from target lists of varying sizes. Subjects demonstrated less accurate and slower responses to the verbal memory search task as tracking difficulty was increased. The same study also showed an effect of Sternberg task difficulty on tracking performance, but, interestingly, the effect was in an unexpected direction, with tracking error decreasing as memory set size increased from 2 to 4.

Third, visual-manual tracking was selected as a concurrent task because it avoids structural interference when the distinction is drawn between structural and capacity interference in explaining dual-task performance decrements. According to Kahneman (1973), structural interference is a more peripheral phenomenon that results when two tasks compete for the same perceptual or effector organs. McNeil et al. (1991) further suggested that computational interference, as distinct from both structural and capacity interference, may result from competition between two tasks for the same mental representations, processing stages, or other cognitive machinery. Capacity interference, on the other hand, involves competition for resources that can be allocated flexibly between two tasks that employ independent structures and computations. When capacity interference occurs, performance changes are related to task difficulty and the intensive aspect of attention demand, rather the structural requirements. Navon and Gopher (1979) have discussed similar distinctions as *concurrence cost* versus *resource cost.* In their formulation, concurrence cost encompasses not only Kahneman's (1973) notion of structural interference, but also the state of affairs in which concurrent tasks create conditions detrimental for one another, such as increased demand on mechanisms that organise, schedule, and coordinate dual-task performance. They also pointed out that, in practice, many concurrent task pairs will present some combination of central capacity interference and non-central structural or computational interference when the terms are defined as above.

The goal of Experiments 2 and 3 was to demonstrate the reliability and validity of the motor tracking task in the RAPP software for the purposes of demonstrating performance trading in subsequent dual-task investigations. The objective of Experiment 2 was to demonstrate that the software produces target waveforms of equivalent difficulty across tracking trials. The objectives of Experiment 3 were (1) to demonstrate that three a priori chosen levels of difficulty produce reliable differences in tracking performance; (2) to describe performance changes across repeated trials for the purpose of determining how much single-task practice to give subjects in a subsequent dual-task study in order to minimise performance changes in that study due to practice and learning; and (3) to determine the reliability of tracking performance measures across repeated trials.

EXPERIMENT 2
TARGET WAVEFORM RELIABILITY UNDER FIXED DISPLACEMENT CONDITIONS

Method

Procedure. The RAPP software displays an uneven line that scrolls across the computer screen from left to right, with a circle and crosshairs whose position is controlled manually by a joystick. For the current set of experiments, the software was installed on a Dell Latitude computer with a 366 MHz processor and 128K RAM. The joystick was a Saitek Cyborg.

The target waveform is composed of a sequence of waveform segments, or wavelets, chosen from a pre-generated set of 1000 segments. Each wavelet is defined by five points or vertices, which describe the amplitude of the waveform. The points have been calculated so that each of the 1000 wavelets has the same mean and standard deviation amplitude. The starting set of vertices for each run was randomly chosen from the set of predefined wavelets. The amplitude of the displayed waveform can be adjusted from 100% to 10% in 10% steps. Two factors control the rate at which the waveform moves horizontally across the screen, the speed at which the waveform scrolls across the screen and the horizontal distance between the vertices. For the current experiments, a constant scroll speed of 15 twips[2] per ms was utilised, the waveform amplitude was set to 100%, and the distance between points was adjusted to give 10, 20, or 50 vertices per minute in the Easy, Moderate, and Hard tracking conditions. Figure 1 provides a visual example of the target waveforms produced at each of the three difficulty levels.

In order to confirm that the software produces target waveforms of consistent difficulty at each of three pre-selected difficulty levels, tracking error data were collected by generating a series of waveforms and fixing the value of the crosshair position at

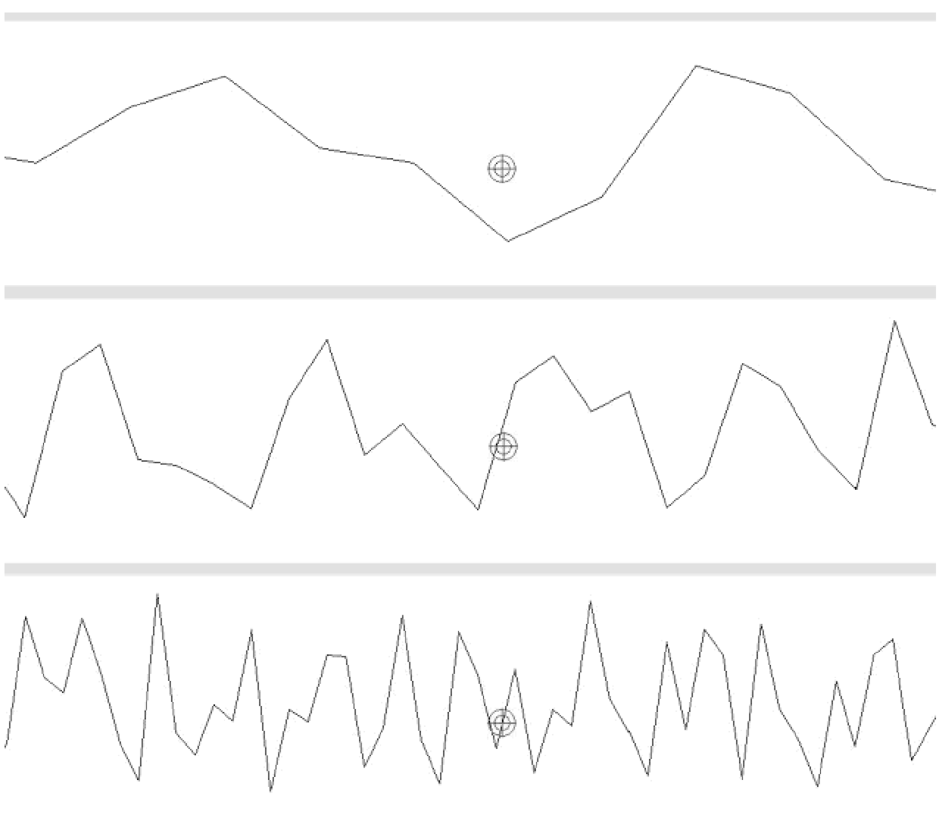

Figure 1. Examples of the target waveform for the RAPP visual-manual tracking task at the Easy, Moderate, and Hard difficulty levels.

[2] A twip is a standardised pixel, defined as 1/1140th of a logical inch.

maximum forward displacement. The waveform trials were equal in length and number to the trials that were administered to subjects in a subsequent dual-task experiment in which the manual tracking task was paired with a story listening and retelling task. Four groups of three trials were collected for each of 20 hypothetical subjects at each of the three difficulty levels. Individual trials were approximately 90 seconds in length.

Analysis. RMS tracking error was calculated for each tracking trial using the following formula, where T_i is the position of the target waveform for a given sample i, and x_i is crosshair position for the same sample, and n is the number of samples in a given trial:

$$RMS\ error = \sqrt{\left(\sum_i (T_i - x_i)^2\right) \div n}$$

Conceptually, RMS tracking error is roughly equivalent to the total area between the curves described by the target waveform and the subject's tracking response (Schmidt & Lee, 1999). It is a measure of total error that includes both bias (in terms of the current task, a tendency to respond consistently above or below the target) and variability in responding. The RMS tracking error values calculated for each tracking trial were averaged across three-trial sequences to yield four values for 20 subjects at each of the tracking difficulty levels. The means for three-trial blocks were examined because this was to be the unit of analysis of interest in the subsequent dual-task investigation.

Results and discussion

To address the question of waveform reliability independent of subject response, the tracking error values obtained from trials where the crosshair position was fixed at maximum displacement were entered into a 2-way ANOVA with a design analogous to the one to be used in the subsequent dual-task experiment: one four-level repeated factor of trial block and one between-subjects factor of tracking difficulty level. Neither main effect [$F(3, 171) = 1.107$ for trial block; $F(2, 57) = .611$ for difficulty level, $p > .3$ for both] nor their interaction [$F(6, 171) = .432$, $p > .8$] was significant, supporting the assumptions that (1) the target waveform stimulus used in subsequent visual-manual tracking experiments is of consistent difficulty across trials, and (2) differences across tracking conditions can confidently be attributed to differences in subject performance, and not to stimulus variability. The overall mean tracking error with fixed maximum joystick displacement was 1892 twips and mean differences in tracking error across trial blocks and difficulty levels ranged from 0 to 4 twips.

The objective of this experiment was to investigate whether the target waveform produced by the RAPP software produces a visual-manual tracking task of equivalent difficulty across trials, independent of human response. This was done by examining the differences in tracking error across repeated trials with the crosshair position fixed at its maximum displacement. This particular analysis revealed no differences across repeated trials in terms of RMS or variable tracking error, supporting the contention that target waveforms are of equivalent difficulty across trials of lengths equivalent to those used in the subsequent dual-task experiment.

EXPERIMENT 3
TRACKING PERFORMANCE ACROSS DIFFICULTY LEVELS AND TRIALS

Method

Subjects. A total of 20 participants (13 women, 7 men) between the ages of 40 and 74 (mean = 51, SD = 15) who met the same selection criteria as used in Experiment 1 took part in this experiment. Subjects were recruited by the same methods and were reimbursed $15 for their participation.

Procedure. Informed consent and screening measures were administered in a 60-minute session. In a second 120-minute session, conducted within a week of the first, 36 two-minute single-task tracking trials (12 each at the easy, moderate, hard tracking difficulty levels) were presented. The order of the 36 tracking trials was determined such that each sequence of three tracking trials contained one trial at each difficulty level in pseudorandom order. This was accomplished by randomly selecting without replacement from the six possible orderings of the three difficulty levels in six cycles. The time between tracking trials was approximately 30 seconds, and subjects were given a 5-minute rest after trial 12 and again after trial 24; or, stated differently, after the 4th and 8th trials had been completed for each of the three difficulty levels. Subjects were instructed to use the joystick with their dominant hand to keep the crosshairs as close to the line as possible at all times, and they were discouraged from resting their wrist on the table as they tracked.

Analysis. RMS tracking error was calculated for each 2-minute tracking trial.

Results

To address the questions of the validity of the difficulty levels and performance across repeated trials by human subjects, the RMS error values for each 2-minute tracking trial were entered into an ANOVA with two repeated factors: Difficulty (3 levels) and trial (12 levels). Because the data violated the sphericity assumption, the Huynh-Feldt correction for degrees of freedom was used. The analysis revealed a significant main effect of difficulty [$F(1.15, 21.87) = 602.19$, $p < .001$] and trial [$F(7.41, 140.82) = 42.53$, $p < .001$], as well as a significant interaction [$F(12.66, 240.45) = 5.052$, $p < .001$]. Given the presence of the interaction, post hoc analysis was carried out using the Bonferroni correction for multiple comparisons ($\alpha = .05$) to test for simple main effects of difficulty at each trial and of trial at each difficulty level. Inspection of Figure 2 reveals large differences in tracking error across difficulty levels, with all three pairwise comparisons demonstrating statistical significance at all levels of the trial factor. Analysis of the effect of trial at the hard tracking level revealed that trial 6 was the earliest trial that did not differ reliably from any of the subsequent six trials at that level, with performance at trials 1–5 all being significantly worse than at trial 9. For the moderate and easy tracking levels, trial 5 was the earliest trial that did not differ significantly from any of the seven subsequent trials at each of those levels, again demonstrating more error on trials 1–4 than on trial 9.

The standard error of measurement was also calculated for three-trial blocks within each difficulty level, using trials 7–12 at each level. Three-trial block means were used in this calculation because this was to be the unit of analysis in a subsequent dual-task experiment. The results of this analysis are presented in Table 4.

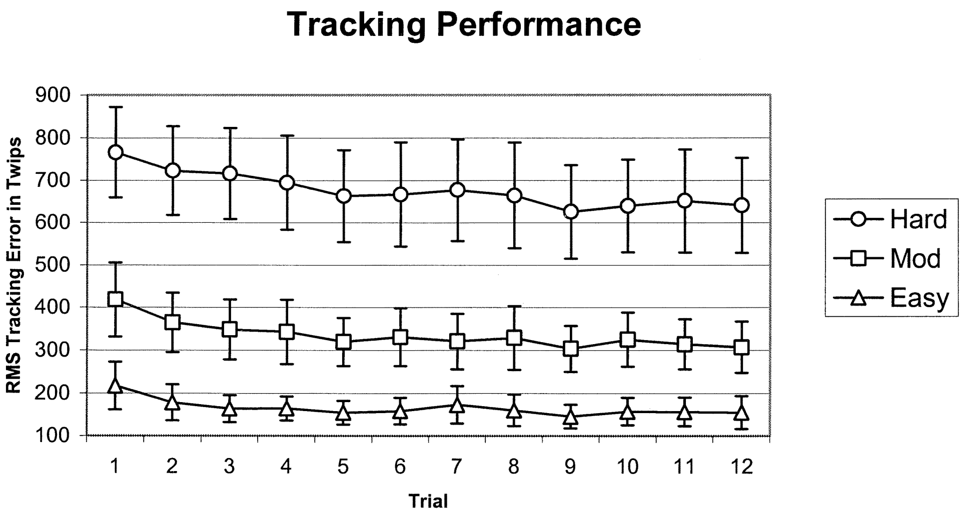

Figure 2. RMS tracking error across the 12 trials at each of the three difficulty levels in Experiment 2. Error bars indicate +/− 1 SD.

TABLE 4
Standard error of measurement for RMS tracking error in twips in single-task three-trial blocks

	Hard	*Moderate*	*Easy*
SEM	22.6	16.3	11.5

Discussion

The first objective of this experiment was to confirm that three levels of difficulty chosen for the tracking task do indeed result in differences in tracking performance. Analysis of tracking error data provided by human trackers showed large and reliable differences across the three difficulty levels chosen for this investigation. This suggests that these levels of the tracking task represent a potentially appropriate difficulty manipulation for induction of performance trade-offs in a concurrently performed task.

The second objective was to determine the amount of practice with this particular task that should be given to subjects preceding a subsequent dual-task experiment in order to minimise performance changes across dual-task conditions due to practice and learning. To accomplish this, mean performance for each trial at each difficulty level, beginning with trial 1, was compared to subsequent trials at that difficulty level until the first trial that did not differ significantly from any subsequent trials was found. This criterion was chosen because it was hypothesised that performance improvements that might not be detectable by comparing adjacent trials might, however, become apparent across a number of trials. For the hard tracking level, the first trial that met this condition was trial 6. For both the moderate and easy levels, trial 5 was the first that did not differ from subsequent trials. Thus, it was concluded that four or five single-task practice trials should be sufficient to minimise learning effects in subsequent dual-task studies with these specific tracking difficulty levels.

It is worth noting that at each difficulty level, the best average performance was obtained on trial 9 (equivalent to trials 25–27 of 36 across all conditions), and indeed in the comparisons described above, trial 9 was the only subsequent trial found to be significantly different from trial 6 for the hard condition and trial 5 for the easy and moderate conditions. A similar trend was noted for subjects to perform better on trial 5 at each level (trials 13–15 overall) than on the trials immediately preceding or following them.

We attribute these performance patterns to the fact that subjects were given a full 5-minute rest immediately prior to trials 13 and 25 of the total 36 trials. In the motor learning literature this recovery of performance following a rest period is known as the reminiscence effect. This effect has been attributed to the distinction between massed practice, where the rest time between trials is less than the time spent practising on each trial, and distributed practice, in which the rest between trials is longer than the practice trials themselves (Schmidt & Lee, 1999). In their review, Schmidt and Lee suggest that subjects receiving massed practice on pursuit rotor tasks perform worse during practice than subjects receiving distributed practice, and demonstrate less learning on retention tests.

The final objective of this study was to describe the reliability of the RMS tracking error dependent variable in terms of its standard error of measurement. It is notable that the between-subjects variability on this motor tracking task, as displayed by the standard deviation error bars in Figure 2, is large relative to the SEM values and also relative to the mean trial-to-trial differences in performance. This suggests that performance differences between individuals may be large relative to performance differences attributable to experimental conditions, such as different amounts of practice or dual-task manipulations. For this reason, examination of within-subjects effects is likely to be the most appropriate when investigating hypotheses regarding the effects of dual-task processing load on performance of this task.

To summarise, Experiments 2 and 3 provided support for the following conclusions: the RAPP tracking software produces a target waveform of consistent difficulty across repeated trials at a given difficulty level; the three a priori chosen levels of task difficulty studied here result in reliable differences in performance; and performance improvement due to practice in a single session will be negligible after four to five trials at each of the three difficulty levels.

EXPERIMENT 4
DUAL-TASK COST SHARING

Experiment 1 provided support for the validity of the difficulty manipulation of the SRP, and Experiments 2 and 3 demonstrated the reliability of the target waveform in the visual-manual tracking task and its differential difficulty manipulations. Experiment 3 investigated whether these two tasks demonstrate cost sharing under dual-task conditions. It was hypothesised that by independently manipulating the difficulty of these two concurrent tasks, a performance trade could be demonstrated, such that increasing the difficulty of one task would result in a performance decrement in the other. Such a finding would be consistent with the view that these computations share limited attentional resources and would also suggest that this dual-task procedure could be used to index the spoken language handicap in aphasia by quantifying, not only information transfer, but also the amount of effort expended by communication partners of aphasic persons in comprehending their connected spoken language.

Method

Participants. A total of 60 healthy individuals (44 females, 16 males) aged between 40 and 75 years (mean = 58, SD = 10) who were recruited by the same methods and met the same selection criteria as those used in Experiment 1, served as participants. None of the individuals participating in Experiment 4 had taken part in any prior studies of the SRP. Each participant was paid $50 for completing the study.

Procedure. The dual-task procedure involved simultaneous presentation of the SRP and visual-motor tracking tasks described above using the RAPP software program.

To minimise practice effects on tracking performance, immediately preceding the dual-task procedure, subjects performed 12 two-minute single-task tracking trials, four at each of the three tracking difficulty levels used in Experiment 2. The order of presentation of tracking difficulty levels during the 12 single-task trials was determined such that each sequence of three trials contained one trial from each difficulty level in pseudorandom order, as was done in Experiment 3.

For the dual-task conditions, the 60 subjects were randomly assigned to the three tracking difficulty levels, producing three groups of 20 subjects each. Each group performed the tracking task at one of the three difficulty levels while listening to one story form (three stories) from each of the four story difficulty levels (normal speaker and speakers with mild, moderate, and severe aphasia). At the end of each story, the subjects stopped tracking and immediately retold the story in their own words. Subjects were instructed to do their best and to devote equal effort to both tasks. Their retellings were digitally recorded by the RAPP software and later scored off-line for %IUs/Min, using the methods described in Experiment 1.

Results

In order to evaluate concurrent performance costs, two two-way ANOVAs were computed, one each with RMS tracking error and %IUs/Min as the dependent variable. For both analyses, story severity was a within-groups factor and tracking difficulty was a between-groups factor.

For the analysis of tracking performance, the results showed a significant ($p < .05$) main effect for both tracking level [$F(2, 57) = 57.62$] and story level [$F(3, 171) = 3.52$] with no significant interaction [$F(6, 171) = 1.44$]. Post-hoc analyses for both main effects were carried out at $p < .05$ using the Bonferroni adjustment for multiple comparisons. As expected, tracking performance decreased significantly as tracking difficulty increased. The analysis of the effect of story difficulty level revealed significantly better tracking performance during the ‘‘Mild’’ compared to the ‘‘Moderate’’ stories across all three tracking levels. No other comparisons reached statistical significance. The dual-task tracking data are presented numerically by tracking and story level in Table 5, and they are presented graphically, averaged across tracking levels in Figure 3.

The effects of tracking and story difficulty on story retell performance are shown in Figure 4. The main effect for story level was significant [$F(3, 171) = 319.85$]. As in Experiment 1, story retell performance was significantly different ($p < .05$; using the Bonferroni adjustment for multiple comparisons) among the three aphasic story levels (i.e., performance decreased as story difficulty increased) but not between the ‘‘Normal’’ and ‘‘Mild’’ story forms. Neither tracking difficulty level [$F(2, 57) = 1.631$], nor the interaction between story and tracking difficulty levels [$F(6, 171) = 1.226$] resulted in statistically significant changes in %IUs/Min story retell performance.

TABLE 5
Means, standard deviations, and standard errors of the mean for dual-task RMS tracking error in twips*

	Tracking level								
	Easy			*Moderate*			*Hard*		
Story level	*Mean*	*SD*	*SE*	*Mean*	*SD*	*SE*	*Mean*	*SD*	*SE*
Normal	158	43	10	333	165	37	547	128	29
Mild aphasia	153	48	11	313	152	34	540	130	29
Moderate aphasia	156	43	10	333	132	30	566	137	31
Severe aphasia	149	40	9	336	174	39	556	121	27

* A twip is a standardised pixel, defined as 1/1140th of a logical inch.

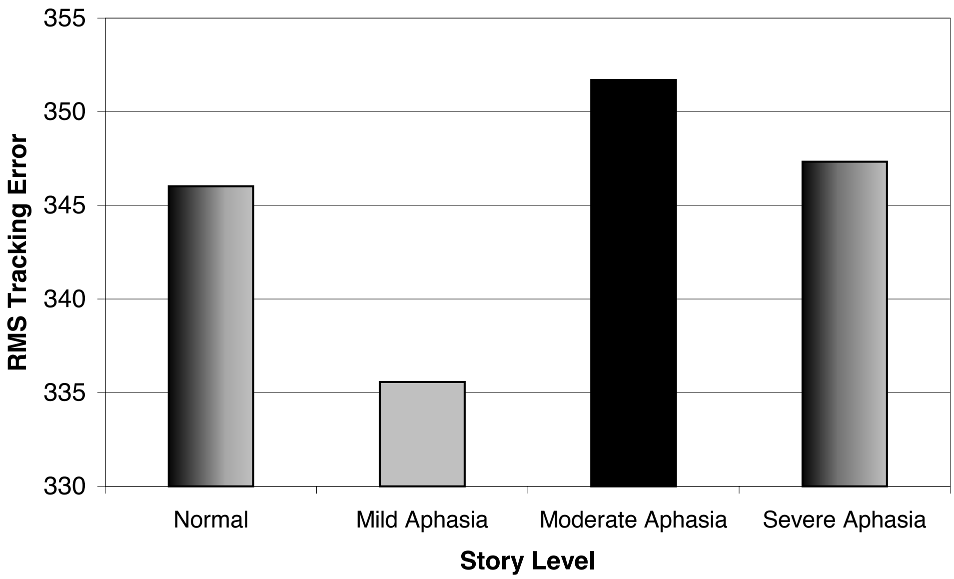

Figure 3. Dual-task tracking performance by story difficulty level, averaged across tracking difficulty levels. As indicated by the shading in the bars, RMS tracking error was significantly greater in the Moderate Aphasia than the Mild Aphasia condition ($p < .05$), and no other comparisons were significant.

Discussion

In this study, participants were required to simultaneously perform two tasks that were proposed to be computationally independent but which shared limited attentional resources under demanding conditions. It was hypothesised that a trade-off of these resources between the two tasks would affect performance on both tasks. As predicted, tracking performance was found to be significantly better under the ‘‘Mild’’ compared to the ‘‘Moderate’’ aphasic story retell condition. This finding provides qualified support for the validity of these dual tasks for augmenting the measurement of normal persons’ comprehension of connected language in persons with aphasia, in addition to those captured by traditional measures of aphasic language performance. The performance

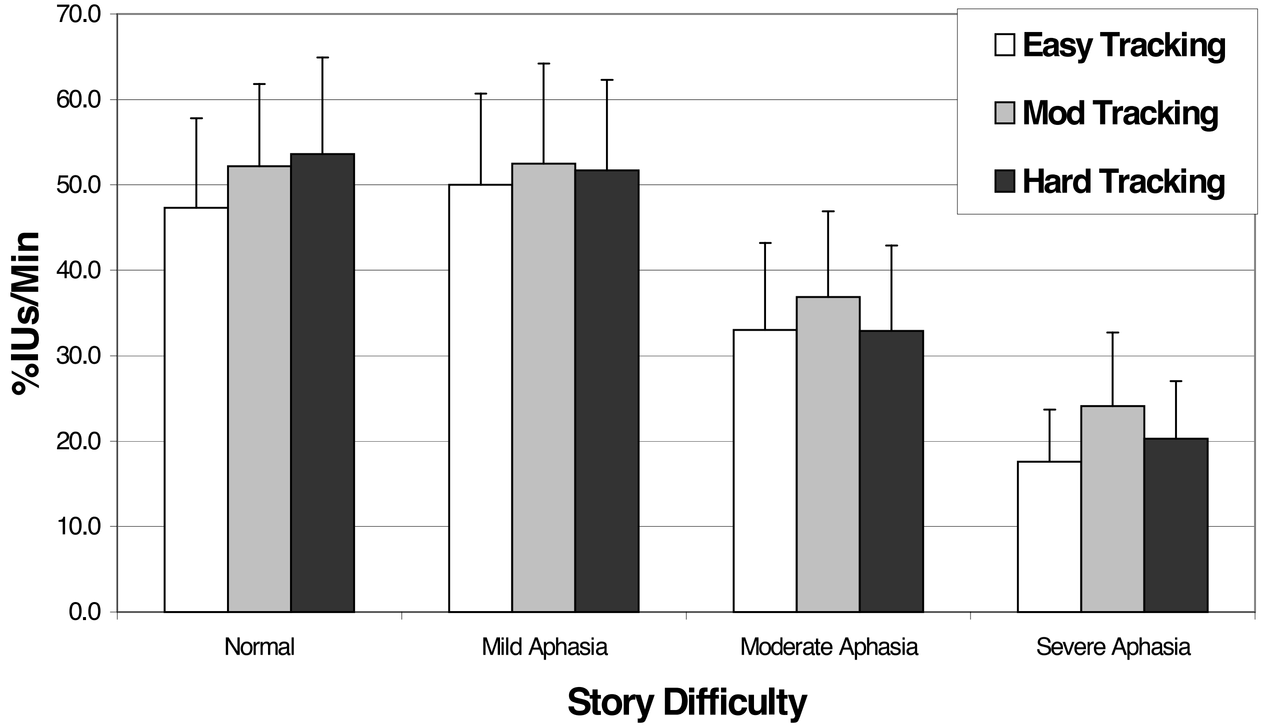

Figure 4. Dual-task story retell performance by story and tracking difficulty level.

interaction between these language comprehension and visual-manual tracking tasks also provides qualified support for resource allocation theory. Each of these qualified conclusions will be examined in turn.

The finding of a tracking performance decrement during stories told by a moderately aphasic speaker compared to stories told by a mildly aphasic speaker suggests that dual-task procedures may be used to index effort experienced by communication partners of individuals with aphasia. However, the tasks used in the present study appear to be inadequate for that purpose. First, an effect of story difficulty on tracking performance was demonstrated only across a limited severity range, from Mild aphasia to Moderate aphasia. The Normal and Severe aphasia story conditions did not show a differential effect on tracking, in spite of having resulted in significantly different %IUs/Min on retell compared to the Mild and Moderate stories. The reasons for this are not entirely clear. Perhaps, because of uncontrolled stimulus variables such as rate of presentation of new information and overall amount of information content, the attention demands of these particular story difficulty levels were not as disparate as originally presumed, based on the single-task %IUs/Min measure and normal listeners' DME judgements. For example, the Severe stimulus stories contained an average of 13% of the information units present in the Normal stories, and were delivered with an efficiency of 8%IU/Min. The overall reduced information content and the slow rate of delivery of information may have counteracted the effects of the aphasic errors and diminished content in the stories, resulting in no net change in attention demand compared to the other stories. Given the fact that participants in Experiment 1 produced fewer %IU/Min in response to the Severe stories than in response to any of the other story levels, the above speculation suggests that, in the case of the SRP, single-task performance may not be a useful measure of resource demand as indexed by dual-task performance.

Second, and more importantly, it is apparent that the dual-task procedures employed in this study will not be clinically useful because of the small effect size of aphasia severity on tracking performance. While the effect was statistically reliable, with 44 of 60 subjects demonstrating performance differences in the predicted direction, the average differences in RMS tracking error were only 26, 20, and 2 twips, respectively at the Hard, Moderate, and Easy tracking levels. These differences amount to slightly more than one standard error of measurement, as determined in Experiment 2, at the Hard and Moderate tracking levels, and considerably less than one SEM at the Easy level. A much larger effect size would be necessary in order for these methods to be clinically useful, especially on an individual basis.

The conclusions that can be drawn from the present investigation regarding resource allocation theory and the role of attention and effort in language processing, while somewhat more positive than the clinical implications of our methods, are still mitigated by the finding of a unidirectional performance trade. Although an effect of story difficulty (aphasia severity) on tracking performance was found, there was no effect of tracking difficulty on story retell performance, as most versions of resource allocation theory might predict (McNeil et al., 1991; Murray, 1999). Both methodological and theoretical factors may have contributed to this negative finding. One relevant feature of the method was the fact that SRP performance was measured off-line and, as such, may have been subject to a variety of strategies not related to on-line story comprehension during concurrent performance of the tracking task. It is also possible that the tracking task provided a more fine-grained index of the resource demand than the SRP task (D. Swinney, personal communication, 2003). The redundancy of connected language might have made it possible for subjects to divert attention away from the SRP task for brief periods while maintaining performance, while the tracking task may have required a more constant investment of resources in order to minimise error.

From a theoretical perspective, the finding of a unidirectional performance trade could result from concurrent performance of two tasks that share partially overlapping sets of resource pools (Navon & Gopher, 1980). The tracking difficulty manipulation may not have challenged the shared pools, while the story difficulty manipulation did. For example, the two tasks might share a resource pool subserving perceptual computations and it might be that it was this particular pool from which the story manipulation recruited additional resources. On the other hand, the tracking difficulty manipulation may have affected a motor-specific pool not utilised by the story comprehension task. Indeed, prior dual-task studies using event-related potential methods have shown that manipulating the frequency of the target waveform in a visual-manual tracking task degrades performance without placing additional load on perceptual and central processing, as measured by P300 amplitude elicited by an oddball secondary task (Backs, 1997; Isreal, Chesney, Wickens, & Donchin, 1980). This speculation could be investigated by introducing new difficulty manipulations into the tracking task, such as order of control, that place greater load on perceptual or central processes (Backs, 1997; Sirevaag, Kramer, Coles, & Donchin, 1989; Wickens et al., 1983).

Another potential theoretical explanation for this finding may be that the tracking levels selected for this study were not ''attention-demanding'' enough to interfere with story listening and retell. Subjects may have benefited from spatial and/or temporal prediction (Rosenbaum, 1980; Schmidt & Gordan, 1977) created by being able to view the upcoming wave. It is also possible that providing subjects with 12 single-task practice trials prior to the dual-task condition may have led to some automaticity (Brown, 1998; Brown & Carr, 1989) and too few resource demands in the subsequent tracking task. To test the first two alternative explanations, new studies are planned which will increase tracking task difficulty and use a smaller viewing window to reduce tracking-wave predictability (and presumably increase attentional demands).

REFERENCES

Arvedson, J. C., & McNeil, M. R. (1987). Accuracy and response times for semantic judgments and lexical decisions with left and right hemisphere lesions. *Clinical Aphasiology*, *17*, 188–200.

Backs, R. W. (1997). Psychophysiological aspects of selective and divided attention during continuous manual tracking. *Acta Psychologica*, *96*, 167–191.

Bayles, K. A., & Tomoeda, C. K. (1993). *Arizona Battery for Communication Disorders of Dementia*. Tucson, AZ: Canyonlands Publishing, Inc.

Brookshire, R. H., & Nicholas, L. E. (1997). *Discourse comprehension test* (2nd ed.). Minneapolis, MN: BRK Publishers.

Brown, I. D. (1978). Dual task methods of assessing workload. *Ergonomics*, *21*, 221–224.

Brown, S. W. (1998). Automaticity versus timesharing in timing and dual-task performance. *Psychological Research*, *61*, 71–81.

Brown, T. L., & Carr, T. H. (1989). Automaticity in skill acquisition: Mechanisms for reducing interference in concurrent performance. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 686–700.

Campbell, T. F., & McNeil, M. R. (1985). Effects of presentation rate and divided attention on auditory comprehension in children with acquired language disorder. *Journal of Speech and Hearing Research*, *28*, 513–520.

DiSimoni, F. G., Keith, R. L., & Darley, F. L. (1980). Prediction of PICA overall score by short versions of the test. *Journal of Speech and Hearing Research*, *23*, 511–516.

Dollaghan, C., Campbell, T. F., & Tomlin, R. (1990). Video narration as a language sampling context. *Journal of Speech and Hearing Disorders*, *55*, 582–590.

Doyle, P. J., & McNeil, M. R. (1998). *Quantifying spoken language handicap in aphasia.* Demonstration exhibit, First National Meeting of the VA Rehabilitation Research & Development Service, Washington, DC.

Doyle, P. J., McNeil, M. R., Spencer, K. A., Goda, A. J., Park, G., Lustig, A. et al. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology*, *14*, 537–549.

Doyle, P. J., Thompson, C. K., Oleyar, K. S., Wambaugh, J. L., & Jackson, A. V. (1994). The effects of setting variables on conversational discourse in normal and aphasic adults. *Clinical Aphasiology*, *22*, 135–144.

Erickson, R. J., Goldinger, S. D., & LaPointe, L. L. (1996). Auditory vigilance in aphasic individuals: Detecting nonlinguistic stimuli with full or divided attention. *Brain and Cognition*, *30*, 244–253.

Frattali, C. M., Thompson, C. K., Holland, A. L., Wohl, C. B., & Ferketic, M. M. (1995). *The American Speech-Language-Hearing Association functional assessment of communication skills for adults (ASHA FACS)*. Rockville, MD: American Speech-Language Hearing Association.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders* (3rd ed.). Baltimore, MD: Lippincott, Williams, & Wilkins.

Gopher, D., Brickner, M., & Navon, D. (1982). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 146–157.

Granier, J. P., Robin, D. A., Shapiro, L. P., Peach, R. K., & Zimba, L. D. (2000). Measuring processing load during sentence comprehension: Visuomotor tracking. *Aphasiology*, *14*, 501–513.

Hirst, W., & Kalmar, D. (1987). Characterizing attentional resources. *Journal of Experimental Psychology: General*, *116*, 63–81.

Holland, A. L. (1980). *Communicative abilities in daily living*. Baltimore: University Park Press.

Hula, W. D., McNeil, M. R., Doyle, P. J., Rubinsky, H. J., & Fossett, T. R. D. (2003). The inter-rater reliability of the Story Retell Procedure. *Aphasiology*, *17*, 523–528.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*, 259–273.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kertesz, A. (1982). *Western Aphasia Battery*. New York: Grune & Stratton.

Lively, S. E., Pisoni, D. B., Van Summers, W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *Journal of the Acoustical Society of America*, *93*, 2962–2973.

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Development and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*, 113–124.

Matthews, G., & Margetts, I. (1991). Self-report arousal and divided attention: A study of performance operating characteristics. *Human Performance*, *4*, 107–125.

McLeod, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. *Quarterly Journal of Experimental Psychology*, *29*, 651–667.

McNeil, M. R., Doyle, P. J., Fossett, T. R. D., Park, G. H., & Goda, A. J. (2001). Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology*, *15*, 991–1006.

McNeil, M. R., Doyle, P. J., Goda, A. J., Park, G. H., Szwarc, L., Spencer, K. et al. (1999). *Normal listeners' severity judgments and comprehension of aphasic discourse: Validating a dual-task instrument for the assessment of spoken language handicap.* Poster presentation to the VA Rehabilitation Research and Development Conference, Washington, DC.

McNeil, M. R., Doyle, P. J., Park, G. H., Fossett, T. R. D., & Brodsky, M. B. (2002). Increasing the sensitivity of the Story Retell Procedure for the discrimination of normal elderly subjects from persons with aphasia. *Aphasiology*, *16*, 815–822.

McNeil, M. R., Odell, K., & Tseng, C. H. (1991). Toward the integration of resource allocation into a general theory of aphasia. *Clinical Aphasiology*, *21*, 21–39.

McNeil, M. R., & Prescott, T. E. (1978). *Revised Token Test*. Austin, TX: Pro-Ed.

Murray, L. L. (1999). Attention and aphasia: Theory, research, and clinical implications. *Aphasiology*, *13*, 91–111.

Murray, L. L. (2000). The effects of varying attentional demands on the word retrieval skills of adults with aphasia, right hemisphere brain damage, or no brain damage. *Brain and Language*, *72*, 40–72.

Murray, L. L., Holland, A. L., & Beeson, P. M. (1997a). Accuracy monitoring and task demand evaluation in aphasia. *Aphasiology*, *11*, 401–414.

Murray, L. L., Holland, A. L., & Beeson, P. M. (1997b). Auditory processing in individuals with mild aphasia: A study of resource allocation. *Journal of Speech, Language, and Hearing Research*, *40*, 792–808.

Murray, L. L., Holland, A. L., & Beeson, P. M. (1998). Spoken language of individuals with mild fluent aphasia under focused and divided-attention conditions. *Journal of Speech, Language, and Hearing Research*, *41*, 213–227.

Navon, D. (1984). Resources: A theoretical soup stone? *Psychological Review*, *91*, 216–234.

Navon, D. (1990). Exploring two methods for estimating performance tradeoff. *Bulletin of the Psychonomic Society*, *28*, 155–157.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, *86*, 214–255.

Navon, D., & Gopher, D. (1980). Task difficulty, resources, and dual-task performance. In R. S. Nickerson (Ed.), *Attention and performance* (pp. 297–315). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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.

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, *38*, 145–156.

Paul, R. (2001). *Language disorders from infancy through adolescence*. St. Louis, MO: Mosby.

Payne, D. G., Peters, L. J., Birkmire, D. P., Bonto, M. A., Anastasi, J. S., & Wenger, M. J. (1994). Effects of speech intelligibility level on concurrent visual task performance. *Human Factors*, *36*, 441–475.

Porch, B. (1981). *Porch index of communicative ability*. Palo Alto, CA: Consulting Psychologists Press.

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2003). Qualitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, *72*, 193–218.

Rosenbaum, D. A. (1980). Human movement initiation: Specification of arm, direction, and extent. *Journal of Experimental Psychology: General*, *109*, 444–474.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2003). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, *37*, 440–479.

Sarno, M. T. (1969). *The functional communication profile*. New York: Howard A. Rusk Institute of Rehabilitation Medicine, New York University Medical Center.

Schmidt, R. A., & Gordan, G. B. (1977). Errors in motor responding, ''rapid'' corrections, and false anticipations. *Journal of Motor Behavior*, *9*, 101–111.

Schmidt, R. A. & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis* (3rd ed.). Champaign, IL: Human Kinetics.

Sirevaag, E. J., Kramer, A. F., Coles, M. G., & Donchin, E. (1989). Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica*, *70*, 77–97.

Slansky, B. L., & McNeil, M. R. (1997). Resource allocation in auditory processing of emphatically stressed stimuli in aphasia. *Aphasiology*, *11*, 461–472.

Tseng, C. H., McNeil, M. R., & Milenkovic, P. (1993). An investigation of attention allocation deficits in aphasia. *Brain and Language*, *45*, 276–296.

Wickens, C. D. (1976). The effects of divided attention on information processing in manual tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 1–13.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). New York: Academic Press.

Wickens, C. D. (1986). The effects of control dynamics on performance. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–60). New York: Wiley & Sons.

Wickens, C. D., & Gopher, D. (1977). Control theory measures of tracking as indices of attention allocation strategies. *Human Factors*, *9*, 365.

Wickens, C. D., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychological analysis of the reciprocity of information-processing resources. *Science*, *221*, 1080–1082.

Wilkinson, G. S. (1993). *Wide Range Achievement Test-3*. Wilmington, DE: Wide Range, Inc.