The Redesign and Re-evaluation of an Internet-Based Telerehabilitation System for the Assessment of Dysarthria in Adults

Anne J. Hill, Ph.D.,[1] Deborah G. Theodoros, Ph.D.,[1] Trevor G. Russell, Ph.D.,[2] and Elizabeth C. Ward, Ph.D.[1]

[1]Division of Speech Pathology and [2]Division of Physiotherapy, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia.

Abstract

A previous study revealed that reliable assessment of dysarthria was feasible. However, that study also revealed a number of system limitations and suggested that technological enhancements and improvements in study design and clinical assessment protocols were needed before validity and reliability of assessment of dysarthria via telerehabilitation could be confirmed. In the current study, improvements in technology, study design, and clinical assessment protocols were implemented in order to re-examine the validity and reliability of assessing and diagnosing dysarthria via the telerehabilitation medium. The aim of this study was to explore the validity and reliability of assessing dysarthria using both formal standardized and informal assessments via a purpose-built telerehabilitation system. Twenty-four participants with an acquired dysarthria were assessed simultaneously via telerehabilitation and face-to-face (FTF) on a battery of assessments. A custom-built telerehabilitation system enabled real-time telerehabilitation assessment over a 128 Kbps Internet connection. Data analysis included an analysis of strength of agreement between the two methods using percentage agreement and weighted κ statistics. Inter-rater and intrarater reliability were also examined for both the FTF and telerehabilitation-led assessments. Good strength of agreement was found between the FTF and telerehabilitation assessment methods. High intrarater and inter-rater reliability within both the FTF and telerehabilitation assessment methods supported these findings. Participants reported high overall satisfaction in the telerehabilitation environment. This study describes the improvements made to the telerehabilitation system reported previously and confirms that valid and reliable assessment of dysarthria using both standardized and informal assessments over the Internet is possible using this system.

Key words: *telehealth, Internet, telerehabilitation system, dysarthria*

Introduction

Dysarthria is a motor speech disorder due to neurological impairment of the speech mechanism. The disorder is characterized by slurred and indistinct speech that affects speech intelligibility and can have a significant effect on quality of life.[1,2] As dysarthria accounts for approximately 54% of adult neurogenic communication disorders,[3] the valid and reliable assessment of this disorder is an integral component of speech pathology practice. The demand for speech pathology services has increased in parallel with an aging population and the ensuing increases in the incidence of stroke and other neurological impairments. As a result, access to speech pathology services has become increasingly restricted. The development of telerehabilitation within speech pathology is viewed as a potential adjunct to the traditional service delivery models, and a means by which access to speech pathology services may be improved.

Few studies into the use of telerehabilitation in speech-language pathology have focused upon the assessment of dysarthria.[4,5] The results reported by Hill et al.[5] were promising with the majority of assessments demonstrating high reliably when administered via the telerehabilitation system. However, the authors acknowledged a number of limitations, including participant and rater variability, poor definition within the assessment tools, and technical issues.[5] These limitations introduced some uncertainty to the conclusions that dysarthria could be reliably and validly assessed via telerehabilitation.

DOI: 10.1089/tmj.2009.0015

The potential for participant variability impacting upon the results of the Hill et al.[5] study was significant, given that participants were assessed twice with a number of days between assessments. Considerable daily variation often occurs in the speech of people with dysarthria, which may have reduced the ability to achieve acceptable agreement across two different occasions.

An additional source of variability within the Hill et al.[5] study was the possibility of rater variability. Generally, rater variability may have a number of sources, such as increased difficulty in accurately judging mild to moderate levels of severity of dysarthric speech[6–13] or the potential for intercorrelation between speech dimensions.[6,8,11] However, another potential source of rater variability within the Hill et al.[5] study was that of ambiguous dimension definition within the assessment rating scales. The Frenchay Dysarthria Assessment (FDA),[14] used in the Hill et al.[5] study, often combines two elements of an oromotor task into one rating scale. Different raters may interpret or emphasize these dimensions in different ways, leading to lower levels of agreement between raters.

Technical difficulties cited as limitations in the Hill et al.[5] study included static Web camera position and inadequate lighting at the assessment site. Other research design limitations included the absence of a diagnosis for each participant and the lack of a participant satisfaction evaluation. It was suggested that these elements be included in any future studies so as to provide a more thorough examination of the efficacy and evidence base for the assessment of dysarthria via telerehabilitation.

The aim of the current study was to refine the telerehabilitation system used in the Hill et al.[5] study and re-evaluate this new system with a modified research design to determine validity and reliability of the assessment of acquired dysarthria in adults.

Methods

ETHICAL CLEARANCE

The study was reviewed and granted ethical clearance from the University of Queensland's human research ethics committee and two tertiary hospital human research ethics committees.

PARTICIPANTS

Participants with dysarthria. A total of 24 people diagnosed with a stable dysarthria associated with an acquired neurological impairment participated in this study. The participants ranged in age from 16 to 78 years (mean 50.2 years). In this cohort, 62.5% of the participants were male and 37.5% were female. The average time post-onset was 42.6 months (range 6 months–11 years). The majority of participants had suffered either a cerebrovascular accident (45.83%) or traumatic brain injury (TBI) (45.83%), while one participant had a progressive neurological disorder, and the other a brain tumor. One participant presented with a concomitant mild apraxia of speech. The participants were recruited from the speech pathology departments of local hospitals. The presence of dysarthria had been established prior to participation in the study by the participant's treating speech–language pathologist (SLP). The researchers were blind to the participants' level of severity of dysarthria prior to assessment, and the research SLPs had no prior clinical contact with the participants.

Exclusion criteria for the study included a history of a speech or language disturbance prior to the present disorder, a severe coexisting apraxia of speech, or a severe coexisting aphasia. Participants with significant visual or hearing impairment, post-traumatic amnesia, a positive history of alcohol abuse, and/or dementia were also excluded from participation. The participants were not required to have any knowledge or skills in the use of computers as the telerehabilitation system used automated software.

Speech–language pathologists. Two SLPs conducted simultaneous rating of the face-to-face (FTF) and telerehabilitation assessment of the participants with dysarthria. One of the 2 SLPs was randomly assigned to lead the assessment, either in the telerehabilitation environment or the FTF environment, while the other SLP became a silent scorer of the assessment in the alternative environment. This simultaneous assessment design was deemed necessary for the reduction of participant variability. The FTF scorer did not assist with the administration of the telerehabilitation-led assessment, beyond seating the participant in front of the computer, assisting with the headset microphone, and orienting them to the telerehabilitation SLP. Both SLPs led the assessments an equal number of times. Although the SLPs had varying levels of experience (2 years and 10 years post-graduation), they had both worked predominantly with adults with neurogenic communication disorders, and undergone training in the administration of the assessment tools and the use of the telerehabilitation system prior to participant assessments.

PROCEDURE

Assessment battery. The informal assessment tools described below were developed by 2 research SLPs so as to overcome some of the limitations identified within the oromotor and perceptual assessment tools used in the Hill et al.[5] study.

1. Informal oromotor assessment: This informal assessment involved the clinical evaluation of the functioning of the lips, tongue, jaw, larynx, and respiration during both speech and nonspeech tasks. The participant's performance on each task

was rated on a 5-point scale (1 = *no impairment*; 5 = *severe impairment*). Laryngeal function was the only speech subsystem that was rated on a yes/no basis. Where the task contained a timed element, the scale contained a guide to the relationship between the time taken to complete the task and severity level. The rating scales were grouped according to the musculature being assessed.

2. Informal perceptual speech assessment: This informal assessment involved the evaluation of speech and voice characteristics from a speech sample (reading aloud a standard passage, "The Grandfather Passage,"[1] and sustained phonation of "ah." A 5-point rating scale was used (1 = *no impairment*; 5 = *severe impairment*) to rate vocal quality, vocal continuity, pitch range, pitch in speech, loudness in speech, loudness variability in speech, nasality, articulatory precision, and overall intelligibility in conversation.
3. Assessment of Intelligibility of Dysarthric Speech (ASSIDS):[15] This standardized assessment was used to quantify speech intelligibility at the sentence level and determine the communication efficiency ratio (CER). Each participant was required to read/repeat after the examiner 22 sentences that were recorded onto a minidisk or saved as an acoustic wave file (telerehabilitation) and then analyzed by two independent listeners blind to the assessment environments. Twenty percent (20%) of the files were re-rated by the listeners for inter- and intrarater reliability.
4. Diagnosis: The diagnosis of type of dysarthria was determined by the assessing SLPs based on medical information and the results of the oromotor and perceptual assessments.
5. Participant Satisfaction Questionnaire: As described in associated research,[16,17] an 8-item participant satisfaction questionnaire that utilized a 5-point rating scale was completed by those participants who underwent the telerehabilitation-led assessment.

Telerehabilitation assessment.

1. System Architecture: The telerehabilitation system used in the current study contained a number of improvements on the system used by the Hill et al.[5] study. While the current system retained the videoconferencing features, the new custom-built software included a store-and-forward function that enabled the capture of high-resolution (640 × 480 pixels) video and high-quality audio files independent of the videoconference. The remote SLP was able to control the recording of these files, which were then automatically transferred from the participant's computer to the remote SLP's computer. Another improvement in the current system was the inclusion of data-sharing capabilities that allowed for the display of instructional images and video clips to the participant. Displaying instructional materials to the participant allowed for the assessment procedures to be more streamlined and efficient than that described by the Hill et al.[5] study.

Other improvements to the system architecture were applied to the participant's computer. Two Web cameras (Logitech, Pro 4000, Apples, Switzerland) were mounted on a robotic arm on the computer monitor; one camera was used for the videoconference, whereas the other was used to capture high-quality video that could be automatically delivered to the telerehabilitation SLP via a store-and-forward mechanism. To further enhance visualization of the participants, the participant was positioned in front of the monitor, with a dark gray backdrop, and additional lighting from a desk lamp. Communication over the videoconference was via a headset microphone/headphone combination. The FTF SLP also wore headphones so as to hear the same instructions.

The SLP computer was located approximately 15 km from the participant. Using this computer, the remote SLP controlled the display of text, images, and demonstration videos on the participant's computer. With respect to participant confidentiality, the system retained a number of inbuilt functions to ensure data security over the Internet.[5]

2. Procedure: Participants were randomly assigned to either a FTF-led or telerehabilitation-led assessment. The telerehabilitation SLP established the videoconference link between the two sites over a 128-Kbps Internet connection. The participant was seated in front of the computer such that the telerehabilitation SLP had a view of the participant's head and upper chest. The telerehabilitation SLP directed the entire session, giving instructions to the participant and displaying the assessment materials on the participant's computer as required. Wherever possible, the telerehabilitation SLP would score the assessment subtests live. However, if higher quality was required, the participant's performance would be rated from the stored video and audio files once the assessment session was complete. The FTF SLP present in the room positioned herself so as to have a clear view of the participant's responses to the assessment tasks. *Figure 1* shows the telerehabilitation system at the participant's end.

In the FTF environment, the assessments were video recorded using a digital video camera, and audio recordings were made using a minidisk recorder. The FTF SLP used a stopwatch to time various tasks and rated tasks live, using the video and audio recordings to check ratings after the session, as is common clinical practice. The telerehabilitation SLP used the purpose-built telerehabilitation system

Fig. 1. Participant seated in front of telerehabilitation system, with a demonstration video being screened to the participant while the face-to-face speech–language pathologist observes the participant from the far side of the monitor.

to capture audio and video recordings of the FTF assessment and scored the assessment from these recordings either live or, if higher quality was required, after the assessment session. The computer monitor at the participant's end was switched off to avoid distraction during the FTF assessment tasks.

3. Statistical Analysis: Data analysis of the oromotor and perceptual assessments was completed on grouped data, that is, the ratings given for each task were grouped according to the speech musculature subsystem. The strength of agreement between the ratings given for the informal oromotor and perceptual assessments in the FTF and telerehabilitation environments was determined by quadratic weighted κ coefficients, which corrects for the amount of agreement that can be expected to occur by chance.[18] The criteria for strength of agreement as outlined by Landis and Koch[19] was used to interpret the κ statistics (*<0.4 = poor agreement; 0.41–0.6 = moderate agreement; 0.61–0.80 = good agreement; 0.81–1.00 = very good agreement*). Percentage exact agreement (PEA) and percentage close agreement (PCA [i.e., ±1 scale point on a 5-point scale]) between the two environments were also determined for the oromotor and perceptual ratings due to the potential for the κ statistic to produce a paradoxical result with some data.[18]

Diagnosis of type of dysarthria was examined in terms of percentage of agreement between the two SLPs. Responses to the participant satisfaction questionnaires were analyzed descriptively. The results from the ASSIDS were compared between assessment methods by means of percentage of agreement within predetermined criteria and a paired *t*-test to explore any differences between methods. The clinical criteria for the ASSIDS were that at least 80% of the comparisons between the methods were within ±8.6% for the percentage sentence intelligibility and ±0.27 for the CER.[5]

Inter- and intrarater reliability was determined for each of the assessment tools in both the FTF and telerehabilitation method environments. Stored video recordings of eight telerehabilitation and eight FTF-led assessments were rated again by each SLP. PEA, PCA, and intraclass correlations (ICC) were calculated for both inter- and intrarater reliability in both environments for the oromotor and perceptual rating scales. The criteria for degree of reliability outlined by Fleiss[20] was used to interpret the ICC data (*<0.4 = poor reliability; 0.41–0.6 = moderate reliability; 0.61–0.80 = high reliability; 0.81–1.00 = very high reliability*). With regard to percentage agreement, high reliability was determined using a criterion of 80% PCA.[8,11,21] Intrarater reliability using PEA was also examined for the diagnosis of type of dysarthria within each method. For the ASSIDS, inter- and intrarater reliability was explored using ICC.

Results

OROMOTOR AND PERCEPTUAL RATING SCALES

The results of the strength of agreement analysis between the FTF and telerehabilitation methods for the oromotor and perceptual rating scales are detailed in *Table 1*. Measures of agreement resulted in a finding of strong agreement between the FTF and telerehabilitation methods (*Table 1*).

ASSESSMENT OF DYSARTHRIC SPEECH IN ADULTS (ASSIDS)

The percentage level of agreement for intelligibility in sentences was 95.83% at ±8.6%. The CER also displayed a percentage level of agreement of 95.83% at its criterion level of ±0.27. Furthermore, a paired-sample *t*-test revealed no significant difference between the scores obtained in the FTF and telerehabilitation environments for the percentage intelligibility in sentences ($t = 1.38$, $p = 0.17$), while the CER displayed a trend toward significance ($t = 2.05$, $p = 0.05$) ($p < 0.05$).

DIAGNOSIS

The percentage level of exact agreement for diagnosis of type of dysarthria across the 24 participants was 66.67% (*Table 2*).

INTRARATER RELIABILITY

Intrarater reliability for the oromotor and perceptual ratings using the FTF method is detailed in *Table 3*. The intrarater reliability for diagnosis of type of dysarthria within the FTF method was good to very good, with Rater 1 at 100% PEA and Rater 2 at 75% PEA. The intrarater reliability for the ASSIDS percentage sentence intelligibility was good to very good (Rater 1 ICC = 0.87; Rater 2 ICC = 0.78), as it was for the CER (Rater 1 ICC = 1.0; Rater 2 ICC = 0.99).

The intrarater reliability using telerehabilitation for the oromotor and perceptual ratings is detailed in *Table 4*. The intrarater reliability for the diagnosis of type of dysarthria in the telerehabilitation environment was the same as that found in the FTF assessment environment, that is, 100% for Rater 1 and 75% for Rater 2. Intrarater reliability using telerehabilitation for the ASSIDS was high for both sentence intelligibility (Rater 1 ICC = 0.94; Rater 2 ICC = 0.83) and CER (Rater 1 ICC = 0.99; Rater 2 ICC = 0.99).

Table 1. Strength of Agreement Between the Ratings Given in the Face-to-Face and Telerehabilitation Methods

RATING SCALES	FTF VS T PEA	FTF VS T PCA	FTF VS T κ	FTF VS T SE
Oromotor rating scales				
Facial symmetry at rest	78.89	96.83	0.56	0.11
Lip musculature during movement	77.08	100	0.79	0.06
Jaw musculature	83.33	100	0.39[a]	0.15
Tongue musculature at rest	73.92	100	0.30[a]	0.25
Tongue musculature during movement	81.55	100	0.87	0.03
DDK	89.58	100	0.95	0.03
Laryngeal function	83.34	100	0.65	0.11
Respiratory support	95.83	100	0.95	0.03
Perceptual rating scales				
Intelligibility in conversation	75	95.83	0.59	0.18
Articulatory imprecision	70.73	100	0.68	0.10
Vocal quality	68.75	95.83	0.66	0.06
Vocal continuity	62.5	95.83	0.63	0.08
Pitch range	75	100	0.39[a]	0.19
Pitch during speech	70.84	88.54	0.47	0.11
Loudness during speech	79.17	91.67	0.11[a]	0.15
Loudness variability	75	93.75	0.31[a]	0.19
Nasality	76.39	94.44	0.39[a]	0.14

[a]Poor strength of agreement between the two methods.
FTF, face-to-face method; T, telerehabilitation method; PEA, percentage exact agreement; PCA, percentage clinical agreement (i.e., ±1 scale point); κ, quadratic weighted κ statistic; SE, standard error for κ; DDK, diadokinetic rate.

Table 2. Diagnosis of Type of Dysarthria by the 2 Speech–Language Pathologists (SLP), with a Percentage Exact Agreement of 66.67%

PARTICIPANT	SLP 1	SLP 2
1	Mixed	Mixed
2	Spastic	Spastic
3	Mixed	Mixed
4	UUMN	Spastic
5	Mixed	Mixed
6	Mixed	Mixed
7	Mixed	Spastic
8	UUMN	Mild residual dysarthria
9	UUMN	Hypokinetic
10	Mixed	Hypokinetic
11	Ataxic	Flaccid
12	Flaccid	Flaccid
13	Mixed	Mixed
14	Mixed	Mixed
15	Spastic	Mixed
16	Mixed	Mixed
17	Mixed	Mixed
18	Mixed	Mixed
19	Mixed	Mixed
20	Spastic	Mixed
21	Mixed	Mixed
22	Mixed	Mixed
23	Mixed	Mixed
24	Flaccid	Flaccid

Mixed, mix of two or more types of dysarthria; UUMN, unilateral upper motor neuron dysarthria.

INTER-RATER RELIABILITY

The inter-rater reliability for the oromotor rating scales and perceptual rating scales in both the FTF and telerehabilitation environments is outlined in *Table 5*.

The inter-rater reliability within the FTF environment for the sentence intelligibility and CER from the ASSIDS was high with ICCs of 0.94 and 1.0, respectively. The inter-rater reliability on each measure of the ASSIDS for telerehabilitation was comparable to that found in the FTF environment with ICCs of 0.87 and 1.0, respectively.

PARTICIPANT SATISFACTION

Of the 12 participants eligible to complete the participant satisfaction questionnaire, only 11 provided responses. The majority of the participants (10 of the 11) rated both the audio and visual quality as good or excellent. All of the participants were comfortable or very happy with the telerehabilitation assessment session, and all of the participants rated their overall satisfaction as more than satisfied or very satisfied. All of the participants reported being confident with the results gained via telerehabilitation assessment, and all were willing to participate in future telerehabilitation assessments. The majority of the participants (8 of the 11) stated that they would be equally satisfied with speech pathology services delivered via telerehabilitation methods; however, only 4 of the 11 thought it would be more convenient for them to access speech pathology services in this manner. Of the remaining 7 participants, 5 felt that telerehabilitation would not be more convenient for them and 2 stated that it was not applicable as they did not have access to the Internet at home.

Discussion

Overall, the strength of agreement between the FTF and telerehabilitation assessment environments for both the oromotor and

Table 3. Intrarater Reliability Within the Face-to-Face Method for Oromotor and Perceptual Rating Scales (n = 4)

	RATER 1 FTF			RATER 2 FTF		
RATING SCALES	PEA	PCA	ICC	PEA	PCA	ICC
Oromotor rating scales						
Facial symmetry at rest	93.75	100	0.51	100	100	1.0
Lip musculature during movement	93.75	100	0.83	93.75	100	0.65
Mandibular musculature	100	100	1.0	100	100	1.0
Tongue musculature at rest	75	100	0.42	75	100	0.36[a]
Tongue musculature during movement	89.29	100	0.89	96.43	100	0.56
DDK	87.50	100	0.93	87.50	100	0.74
Laryngeal function	75	100	0.36[a]	100	100	1.0
Respiratory support	100	100	1.0	100	100	zv
Perceptual rating scales						
Intelligibility in conversation	75	100	0.57	100	100	1.0
Articulatory imprecision	75	100	0.57	100	100	zv
Vocal quality	75	100	0.82	87.50	100	0.68
Vocal continuity	66.67	100	0.65	91.67	100	0.84
Pitch range	100	100	1.0	100	100	zv
Pitch during speech	56.25	100	0.74	93.75	100	0.78
Loudness during speech	75	87.5	0.41	100	100	zv
Loudness variability	87.50	100	0.89	100	100	zv
Nasality	75	83.33	0.32[a]	91.67	100	0.85

[a]Poor reliability; zv, zero variance within one rating, therefore statistic could not be calculated.
FTF, face-to-face method; PEA, percentage exact agreement; PCA, percentage clinical agreement (i.e., ±1 scale point); ICC, intraclass correlation; DDK, diadokinetic rate.

perceptual assessment indicated that valid assessment of dysarthria is possible via telerehabilitation methods. The establishment of good reliability within both the FTF and telerehabilitation assessment environments provided a solid basis upon which the measurement of agreement between the assessment methods could be determined. Although a range of agreement from poor to very good (*Table 1*) was determined by the quadratic weighted κ statistic for the oromotor and perceptual rating scales, it is important to note that all of the quadratic weighted κ statistics were positive, indicating an above-chance level of agreement.[22] Furthermore, the quadratic weighted κ statistic can produce a paradoxical result when the data's marginal totals are highly symmetrically unbalanced.[18] This paradox, in which the κ value is low despite high percentages of agreement, occurred in the case of the jaw-movement group and the tongue-at-rest group of scales. Cicchetti and Feinstein,[23] therefore, have suggested that the quadratic weighted κ should not be reported alone. Indeed, in the current study, reporting PEA and PCA helped to provide comprehensive analysis of the data.

The PEA for the oromotor assessment ranged between 73.92% and 95.83%, indicating that strong agreement existed between the FTF and telerehabilitation assessment methods. In addition, the PCA for the oromotor rating scales was 100% for all except the facial symmetry group, which had a PCA of 93.83%. Although the quadratic weighted kappa, PEA, and PCA values for the perceptual assessment were slightly lower than the oromotor assessment, the PEA was above 70.73% for the majority of rating scales. Furthermore, the PCA for each of the groups of scales was above 88.54%, indicating high agreement (>80%) between the methods.[8,11,21] Thus, despite some low quadratic weighted κ statistics, the PEA and PCA values would indicate that telerehabilitation assessment of oromotor and perceptual speech characteristics is comparable to FTF assessment.

Table 4. Intrarater Reliability Within the Telerehabilitation Method for the Oromotor and Perceptual Rating Scales (*n* = 8)

	RATER 1 TELEREHAB			RATER 2 TELEREHAB		
RATING SCALES	PEA	PCA	ICC	PEA	PCA	ICC
Oromotor rating scales						
Facial symmetry at rest	87.50	100	0.62	87.50	100	0.76
Lip musculature during movement	87.50	100	0.84	100	100	1.0
Mandibular musculature	91.67	100	zv	83.33	100	0.65
Tongue musculature at rest	87.50	100	0.91	100	100	1.0
Tongue musculature during movement	100	100	1.0	96.43	100	0.98
DDK	62.50	100	0.16[a]	100	100	1.0
Laryngeal function	100	100	1.0	100	100	1.0
Respiratory support	100	100	1.0	100	100	1.0
Perceptual rating scales						
Intelligibility in conversation	100	100	1.0	100	100	1.0
Articulatory imprecision	75	100	zv	100	100	1.0
Vocal quality	93.75	100	0.90	68.75	100	0.86
Vocal continuity	58.33	100	0.75	75	100	0.66
Pitch range	100	100	1.0	100	100	zv
Pitch during speech	93.75	93.75	0.30[a]	75	100	0.87
Loudness during speech	87.50	100	zv	100	100	zv
Loudness variability	75	100	zv	100	100	1.0
Nasality	83.33	100	0.80	100	100	1.0

[a]Poor reliability; zv, zero variance within one rating, therefore statistic could not be calculated.

The results from the analysis of the ASSIDS data were positive, with both the percentage sentence intelligibility and CER exceeding the predetermined clinical criteria described by the Hill et al.[5] study. The results from the paired sample *t*-test, although marginal for the CER at $p = 0.05$, confirmed the positive results. This marginal result may have been due to inconsistency in the timing of the sentences by the raters, which in turn disrupts the ratio calculations. These results are not unexpected, given that both assessment methods have the ability to make high-quality audio recordings, which are used in the transcription of the ASSIDS.

The strength of agreement for the diagnosis of type of dysarthria between the FTF and telerehabilitation methods was lower than expected at 66.67% PEA. The SLPs had access to medical information for each of the participants and their own ratings on the oromotor and perceptual rating scales to aid them in the diagnosis of type of dysarthria. Nevertheless, this percentage level of agreement is moderate and may be indicative of differing clinical judgment regarding diagnosis or differing levels of experience in diagnosing types of dysarthria. Disagreement between the SLPs regarding diagnosis appeared to occur most frequently when the participant presented with a relatively mild dysarthria. It is well recognized in perceptual analysis of speech impairment that agreement is more easily obtained when participants present with moderate to severe impairments, rather than mild impairments.[8] The predominance of mild to mild–moderate severity levels within this study's sample of participants may have had an impact on the agreement level on diagnosis of type of dysarthria. Future studies should encompass larger samples with a greater variety of severity levels so as to more

Table 5. Inter-rater Reliability Within Both Face-to-Face and Telerehabilitation Methods

	FTF			TELEREHAB		
RATING SCALES	PEA	PCA	ICC	PEA	PCA	ICC
Oromotor rating scales						
Facial symmetry	90.63	100	0.68	74.11	100	0.43
Lip musculature during movement	87.50	100	0.55	90.63	100	0.85
Mandibular musculature	87.50	100	-0.06[a]	91.67	100	0.66
Tongue musculature at rest	57.14	100	-0.26[a]	75	93.75	0.30[a]
Tongue musculature during movement	89.29	96.43	0.70	89.29	100	0.93
DDK	87.50	100	0.90	87.50	100	0.94
Laryngeal function	81.25	100	0.64	100	100	1.00
Respiratory support	93.75	100	1.00	93.75	100	0.89
Perceptual rating scales						
Intelligibility in conversation	75	100	0.53	87.50	100	0.85
Articulatory imprecision	87.5	100	0.63	87.50	100	0.84
Vocal quality	75	100	0.68	71.88	100	0.76
Vocal continuity	58.33	100	0.39[a]	50	100	0.43
Pitch range	62.50	100	-0.24[a]	75	100	0.42
Pitch during speech	71.88	93.75	0.55	71.88	93.75	0.57
Loudness during speech	100	100	1.0	81.25	100	-0.10[a]
Loudness variability	81.25	93.75	zv	93.75	100	0.78
Nasality	75	95.83	0.63	79.19	95.83	0.43

[a]Poor reliability; zv, zero variance within 1 rater, therefore, statistic not able to be calculated.
FTF, face-to-face method; Telerehab, telerehabilitation method; PEA, percentage exact agreement; PCA, percentage clinical agreement (i.e., ±1 scale point); ICC, intraclass correlation statistic; DDK, diadokinetic rate.

thoroughly investigate the ability to diagnose the type of dysarthria via telerehabilitation methods.

Reliability was analyzed within both the FTF and telerehabilitation methods so as to establish the consistency of the raters within each environment. The intrarater reliability within the telerehabilitation assessment method was comparable to that found in the typical clinical FTF assessment method, for both the oromotor and perceptual assessments. While there was comparability in intrarater reliability between the two methods, there was evidence of a trend toward slightly higher values in the telerehabilitation environment (*Tables 3* and *4*). Therefore, clinicians can be confident that perceptual assessments of oromotor status and perceptual speech characteristics completed via telerehabilitation allows for consistent ratings across assessment occasions. Equivalent intrarater reliability was also established for measures of sentence intelligibility and CER on the ASSIDS, as well as for the diagnosis of type of dysarthria, between the two assessment environments.

The inter-rater reliability was comparable between the two methods for the oromotor and perceptual assessments (*Table 5*). The majority of the rating scales displayed good to very good reliability with low inter-rater reliability evident on only two rating scales within both the oromotor (jaw at rest and tongue at rest) and perceptual assessments (vocal continuity and pitch range) for the FTF method and only one rating scale within each of these assessment tools in the telerehabilitation method (tongue at rest and loudness during speech). It is important to note that although the ICC statistic was low for these parameters, the PCA for most of these groups of scales was actually 100%. This indicates that disagreement between the raters was confined to just one scale point for all except the tongue-at-rest group in the telerehabilitation method, which had a PCA of 93.75% (*Table 5*). Many researchers consider a PCA of above 80% to denote high reliability.[8,11,21] As the PCA for all of the rating scale groups in both the oromotor and perceptual assessments in both environments were above 93.75%, we feel confident that the inter-rater reliability for both environments was high.

Participant satisfaction with the telerehabilitation assessment method was high with the eligible participants rating their overall satisfaction as more than satisfied or very satisfied. High levels of participant satisfaction are important if telerehabilitation is to be accepted by both clients and clinicians as an alternative service delivery model. Two of the participants commented that they "really enjoyed the telerehabilitation experience" and that they "liked using modern ways to communicate." However, only 4 of the respondents believed that it would be more convenient for them to access speech pathology services via telerehabilitation methods. This result should be interpreted with caution, as all respondents were residents in a metropolitan area and with good access to a metropolitan hospital. Trials of telerehabilitation conducted outside of metropolitan areas may yield different results on this particular question, as access to speech pathology services are more restricted.

The results of the current study indicate that the improvements made to the telerehabilitation system used by the Hill et al.[5] study significantly enhanced the evidence of the reliability and validity of assessing dysarthria via telerehabilitation. In particular, the automation of the store-and-forward mechanism and the use of two remotely controlled Web cameras on the participant's computer allowed the assessment to proceed in a more streamlined and timely manner, more closely resembling a traditional assessment session. The use of additional lighting and a contrasting backdrop were other technical improvements suggested by the Hill et al.[5] study that were implemented in the current study, and may have contributed to the positive results indicated in the current study.

Despite the current study containing a number of technological and methodological improvements, it did contain some limitations. One such limitation was the lack of a clinician satisfaction questionnaire. However, it was felt that a clinician satisfaction questionnaire was not entirely relevant at this preliminary stage of research into the reliability and validity of the use of telerehabilitation for the assessment of dysarthria. However, future studies that involve clinicians outside of the development team should seek feedback regarding the usefulness of telerehabilitation and the clinicians' satisfaction with this form of service delivery. There is a considerable body of literature relating to the development of satisfaction questionnaires, which should be consulted thoroughly by future researchers so as to ascertain accurate levels of satisfaction for both participants and clinicians.

A consideration within the current study was the use of a relatively low bandwidth and the tendency for audio and visual breakup over the videoconferencing connection. While the use of 128 Kbps in the current study was justified in that it was the minimum bandwidth guaranteed by the public health service throughout the state of Queensland, it did impact upon assessment via telerehabilitation. Due to the tendency for audio and visual breakup, the system required demonstrations of tasks to be prerecorded and built into the telerehabilitation system. While these demonstrations were of very high quality and automated within the system, using them did disrupt the natural flow of the assessment session and lengthen the time required to complete all the assessments. The disruption and lengthening of the assessment session may have affected the

attention and thus performance of some participants, especially those with a TBI who typically present with attention deficits. Other studies[24,25] have found that participants with a TBI did not perform as well in the telerehabilitation environment as they did in the FTF environment due to difficulty attending to the tasks in the remote situation. The current study does provide evidence that valid and reliable assessments of dysarthria are possible at the low bandwidth of 128 Kbps, which may encourage the implementation of telerehabilitation in countries or regions where infrastructure precludes the use of higher bandwidths. Nevertheless, future studies should make use of advancing technology and infrastructure in order to streamline the assessment process and involve larger sample sizes that include a variety of etiologies and levels of cognitive function.

Conclusions

The results from the current study confirm that valid and reliable assessment of dysarthria via telerehabilitation methods is possible. The robust intrarater and inter-rater reliability found within both the FTF and telerehabilitation methods provided strong support for the strength of agreement found between the FTF and telerehabilitation assessment environments. Despite some low quadratic weighted κ values, descriptive data support the conclusion that valid and reliable assessment of dysarthria is possible via telerehabilitation methods. Future studies should make use of evolving technology to further enhance the telerehabilitation system, as well as expand the population sampled to include those with severe dysarthria.

Acknowledgment

The authors gratefully acknowledge funding support for this study from the National Health and Medical Research Council of Australia Project Grant No. 401604 and the National Health and Medical Research Council of Australia Public Health Postgraduate Scholarship ID 351649. The authors also acknowledge the assistance of Roy Anderson for software development of the telerehabilitation application and Jasmin Cowles for data collection. We acknowledge the assistance of the Princess Alexandra Hospital, Queen Elizabeth II Jubilee Hospital, the Royal Brisbane and Women's Hospital, and The University of Queensland Health and Rehabilitation Clinics in the recruitment of participants for this study. Finally, the authors wish to acknowledge the participants of this study for the gift of their time.

Disclosure Statement

No competing financial interests exist.

REFERENCES

1. Darley FL, Aronson AE, Brown JR. *Motor speech disorders.* Philadelphia: Saunders, **1975**.
2. Yorkston KM, Beukelman DR, Bell KR. *Clinical management of dysarthric speakers.* Austin, TX: Pro-Ed, **1988**.
3. Duffy JR. *Motor speech disorders: Substrates, differential diagnosis and management, 2nd ed.* St. Louis: Elsevier Mosby, **2005**.
4. Duffy JR, Werven GW, Aronson AE. Telemedicine and diagnosis of speech and language disorders. *Mayo Clin Proc* **1997**;72:1116–1122.
5. Hill AJ, Theordoros DG, Russell TG, Cahill LM, Ward EC, Clark KM. An Internet-based telerehabilitation system for the assessment of motor speech disorders: A pilot study. *Am J Speech Lang Pathol* **2006**;1:1–12.
6. Chenery H. Perceptual analysis of dysarthric speech. In: Murdoch BE, ed. *Dysarthria: A physiological approach to assessment and treatment.* Cheltenham: Stanley Thornes (Publishers) Ltd, **1998**:36–67.
7. Karnell MP. Nasometric discrimination of hypernasality and turbulent nasal airflow. *Cleft Palate Craniofac J* **1995**;32:145–148.
8. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *J Speech Hear Res* **1993**;36:21–40.
9. Ludlow CL, Bassich CJ. Relationships between perceptual ratings and acoustic measures of hypokinetic speech. In: McNeil MR, Rosenbek JC, Aronson AE, eds. *The dysarthrias: Physiology, acoustics, perception, management.* San Diego: College-Hill Press, **1983**:163–195.
10. McHenry MA. Aerodynamic, acoustic, and perceptual measures of nasality following traumatic brain injury. *Brain Injury* **1999**;13:281–290.
11. Sheard C, Adams RD, Davis PJ. Reliability and agreement of ratings of ataxic dysarthric speech samples with varying intelligibility. *J Speech Hear Res* **1991**;34:285–293.
12. Theodoros DG, Murdoch BE, Stokes PD, Chenery HJ. Hypernasality in dysarthric speakers following severe closed head injury: A perceptual and instrumental analysis. *Brain Injury* **1993**;7:59–69.
13. Theodoros DG, Murdoch BE, Stokes PD. Variability in the perceptual and physiological features of dysarthria following severe closed head injury: An examination of five cases. *Brain Injury* **1995**;9:671–696.
14. Enderby P. *Frenchay dysarthria assessment.* San Diego: College-Hill Press, **1983**.
15. Yorkston KM, Beukelman DR. *Assessment of intelligibility of dysarthric speech.* Austin, TX: Pro-Ed, **1981**.
16. Theodoros D, Hill A, Russell T, Ward E, Wootton R. Assessing acquired language disorders in adults via the Internet. *J Telemed e-Health* **2008**;14:552–559.
17. Hill A, Theodoros D, Russell T, Ward E, Wootton R. The effects of aphasia severity upon the ability to assess language disorders via telerehabilitation. *Aphasiology* **2008**;23:627–642.
18. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* **1990**;43:543–549.
19. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* **1977**;33:159–174.
20. Fleiss JL. *Statistical methods for rates and proportions.* New York: John Wiley & Sons, **1981**.

21. Kearns KP, Simmons NN. Interobserver reliability and perceptual ratings: More than meets the ear. *J Speech Hear Res* **1988**;31:131–136.

22. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* **1973**;33:613–619.

23. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* **1990**;43:551–558.

24. Brennan DM, Georgeadis AC, Baron CR, Barker LM. The effect of videoconference-based telerehabilitation on story retelling performance by brain-injured subjects and its implications for remote speech-language therapy. *Telemed J E Health* **2004**;10:147–154.

25. Georgeadis AC, Brennan DM, Barker LM, Baron CR. Telerehabilitation and its effect on story retelling by adults with neurogenic communication disorders. *Aphasiology* **2004**;18(5/6/7):639–652.

Address correspondence to:
Anne J. Hill, Ph.D.
Division of Speech Pathology
School of Health and Rehabilitation Sciences
The University of Queensland
Brisbane
Australia

E-mail: annejhill@gmail.com

Received: February 15, 2009
Accepted: May 13, 2009