

Research Report

Assessing disordered speech and voice in Parkinson's disease: a telerehabilitation application

Gabriella Constantinescu†, Deborah Theodoros†, Trevor Russell‡, Elizabeth Ward†, Stephen Wilson§ and Richard Wootton¶

†Division of Speech Pathology, University of Queensland, St Lucia, Brisbane, QLD, Australia

‡Division of Physiotherapy, University of Queensland, St Lucia, Brisbane, QLD, Australia

§School of Information Technology and Electrical Engineering, University of Queensland, St Lucia, Brisbane, QLD, Australia

¶Scottish Centre for Telehealth, Aberdeen, UK

(Received April 2009; accepted November 2009)

Abstract

Background: Patients with Parkinson's disease face numerous access barriers to speech pathology services for appropriate assessment and treatment. Telerehabilitation is a possible solution to this problem, whereby rehabilitation services may be delivered to the patient at a distance, via telecommunication and information technologies. A number of studies have demonstrated the capacity of telerehabilitation to provide reliable and valid assessments of speech, voice and language. However, no studies have specifically focused on assessing patients with Parkinson's disease.

Aims: To investigate the validity and reliability of a telerehabilitation application for assessing the speech and voice disorder associated with Parkinson's disease.

Methods & Procedures: Sixty-one participants with Parkinson's disease and hypokinetic dysarthria were simultaneously assessed in an online and face-to-face environment by two speech–language pathologists. The assessment protocol included perceptual measures of voice and oromotor function, articulatory precision, speech intelligibility, and acoustic measures of vocal sound pressure level, phonation time and pitch range. Online assessments were conducted via a personal computer-based videoconferencing system with store-and-forward capabilities, operating on a 128 kbit/s Internet connection. The level of agreement between the online and face-to-face ratings was determined using several different analyses, depending on the parameter. These included per cent close agreement, quadratic weighted Kappa, and the Bland and Altman limits of agreement.

Outcomes & Results: Per cent close agreement between the two environments was within a predetermined clinical criterion of 80% agreement for all voice and oromotor parameters, articulatory precision and speech intelligibility in conversation. Levels of agreement between the environments, based on quadratic weighted Kappa, ranged from poor to good for vocal parameters and from fair to very good for oromotor parameters. Bland and Altman limits of agreement analyses revealed comparability between online and face-to-face environments for vocal sound pressure level, phonation time, pitch range, sentence intelligibility and communication efficiency in reading. Intra- and inter-rater reliability scores for all tasks were comparable between the online and face-to-face environments.

Conclusions & Implications: For the majority of parameters, comparable levels of agreement were achieved between the two environments. Online assessment of disordered speech and voice in Parkinson's disease appears to be valid and reliable. The telerehabilitation application described in this study provides evidence for the delivery of online assessment for the dysarthric speech disorder associated with Parkinson's disease.

Keywords: Parkinson's disease, telerehabilitation, Internet-based assessment, speech and voice disorder.

Address correspondence to: Gabriella Constantinescu, Division of Speech Pathology, University of Queensland, St Lucia, Brisbane, QLD 4067, Australia; e-mail: gabriella@hearandsaycentre.com.au

What this paper adds*What is already known on this subject*

Recent studies have indicated the benefits of using telerehabilitation to assess motor speech, voice and language disorders in the adult neurological population. However, no studies have focused specifically on the online assessment of hypokinetic dysarthria associated with Parkinson's disease.

What this study adds

The study has demonstrated the validity and reliability of online assessment for evaluating the speech and voice disorder specifically associated with hypokinetic dysarthria and Parkinson's disease. The study provides a basis for the delivery of online services for people with Parkinson's disease.

Introduction

Parkinson's disease is a progressive neurodegenerative disorder associated with significant motor disturbance and speech difficulties. Worldwide, Parkinson's disease is present in approximately 1% of the population aged 65 years and above (De Rijk *et al.* 1997). In 2005, the prevalence rate of Parkinson's disease in Australia was estimated to be 290 cases per 100 000 population aged 55–64 years, with a total cost to the health system of AUD\$343.9 million per year (Access Economics 2007).

Hypokinetic dysarthria, the motor speech impairment associated with Parkinson's disease, is characterized by monotony of pitch and loudness, reduced loudness and stress, imprecise articulation, variable rate and short rushes of speech, inappropriate silences, and a harsh and breathy voice (Darley *et al.* 1969). The incidence of the speech disorder occurs in as many as 50–90% of individuals during the course of their disease (Ramig *et al.* 2004), with the severity of the dysarthria increasing with disease progression (Hartelius and Svensson 1994). Impaired speech intelligibility may result in decreased involvement in communicative exchanges, isolation within the family and community, and a subsequent degradation in the person's quality of life (Oxtoby 1982).

Patient access to speech-language pathology services for appropriate assessment and treatment of this condition is limited. Grimm *et al.* (2004) in a survey of 250 people with Parkinson's disease across Queensland, Australia, identified certain barriers impacting on service access. The barriers included: low priority placed on speech pathology services in the public health system; limited availability of speech–language pathologists (SLPs) trained to administer the effective Lee Silverman Voice Treatment® (LSVT) for Parkinson's disease; physical incapacity of the individuals; difficulties with transport and travel, and the large distances to the health service facilities. Of the people with Parkinson's disease who were surveyed, only 36.7% reported having access to speech pathology services.

Reduced services for people with Parkinson's disease have also been noted in studies conducted in the United Kingdom and Sweden, where relatively few participants surveyed (as little as 2.0 to 14.2%) had been able to access a SLP (Hartelius and Svensson 1994, Mutch *et al.* 1986, Oxtoby 1982, Peto *et al.* 1997). These findings are in contrast to the 49% of people with Parkinson's disease who were identified as having speech difficulties (Oxtoby 1982) and 33% of people with Parkinson's disease presenting with speech difficulties who were dissatisfied with their level of service access (Peto *et al.* 1997).

The disparity between the supply and demand of speech pathology services for people with Parkinson's disease suggests the need for an additional or alternate mode of service delivery for this population. One possible solution is the use of telerehabilitation, whereby telecommunication and information technologies are used in the delivery of healthcare at a distance. Telerehabilitation is an emerging field in speech pathology and research has included various communication technologies for the assessment and treatment of motor speech, fluency, voice and language disorders. The earlier use of the telephone, closed-circuit television and satellite-based videoconferencing, are gradually being replaced by Internet-based videoconferencing via a personal computer which are now accessible to many individuals.

Regardless of the technology used, valid and reliable assessment procedures need to be established to ensure effective telerehabilitation services. Recent studies using Internet-based videoconferencing via a personal computer have highlighted the benefits of telerehabilitation for the assessment of motor speech performance, story retell, language comprehension and expression for adults with neurological impairments (Brennan *et al.* 2004, Georgeadis *et al.* 2004, Hill *et al.* 2006, 2008a, 2008b, Palsbo 2007, Theodoros *et al.* 2008). The studies commonly reported high levels of agreement between the online and face-to-face ratings for the majority of parameters investigated on the informal and

standardized assessments. High participant satisfaction with the online environment was also obtained (Brennan *et al.* 2004, Georgeadis *et al.* 2004, Hill *et al.* 2008a, 2008b, Theodoros *et al.* 2008). Furthermore, it was encouraging to note that severity of aphasia and apraxia did not significantly impact on the accuracy of the online assessments. However, the occasional audio-visual disturbances online caused by heavy traffic on the network did make it more difficult for the SLPs to conduct the apraxia assessment for the more severe participants, and rate the aphasia parameters of naming and paraphasia (Hill *et al.* 2008a, 2008b). The ratings of some motor speech parameters (palatal movement in speech, laryngeal volume, tongue elevation and lateral tongue movements), were also more difficult online for all participants involved, and this was a result of audio-visual difficulties, lighting, camera positioning and lack of zoom focus (Hill *et al.* 2006).

As previous studies have demonstrated the capacity of telerehabilitation to provide reliable and valid assessment of speech, voice and language, it was proposed that an Internet-based application to provide services to people with Parkinson's disease might lessen the access issues that exist for this population. In order to ensure that valid assessments underpin treatment programs, the current study aimed to investigate the validity and reliability of an Internet-based assessment protocol specifically designed to evaluate the speech and voice disturbances associated with Parkinson's disease, by comparison with clinical face-to-face assessment. It was hypothesized that online assessment of the speech and voice disturbances in Parkinson's disease can be achieved to a level comparable with standard face-to-face assessment. The current study forms the first validation stage to determine the feasibility of online delivery as a complete assessment and treatment service delivery model for Parkinson's disease.

Method

Participants

Before commencement of the study, ethical clearance was obtained from the Behavioural and Social Sciences Ethical Review Committee of The University of Queensland, Brisbane. Sixty-one participants with Parkinson's disease and hypokinetic dysarthria (42 males, 19 females) aged between 52 and 89 years (mean = 69.23 years; standard deviation (SD) = 8.60) volunteered for the study. Participants were diagnosed as having Parkinson's disease by a neurologist experienced in movement disorders. Time post-diagnosis ranged from 6 months to 30 years (mean = 6.52 years; SD = 6.53).

Fifty-seven of the participants had been diagnosed with Idiopathic Parkinson's disease, of which nine participants had undergone surgical treatment for Parkinson's disease including deep brain stimulation (seven participants) and pallidotomy (two participants). The remaining four participants in the cohort had been diagnosed with Parkinson-plus syndromes, including progressive supranuclear palsy (three participants) and multiple system atrophy (one participant). Stages of Parkinson's disease as per the Hoehn and Yahr (1967) scale for the participants ranged from I to IV with 48 participants rated at Stages I and II, and 13 participants rated at Stages III and IV. For all participants, an overall severity level for hypokinetic dysarthria was determined by the investigators from clinical judgement. The dysarthria levels ranged from mild to severe, with 41% of participants considered as mild, 48% of participants as moderate and 11% of participants as severe. Participants were recruited from various support groups of Parkinson's Queensland Incorporated, public hospitals and from private neurologists in Brisbane, Australia. Proficiency in the use of computers was not a requirement for inclusion in the study as all aspects of the online assessment delivery were performed by the online assessing SLP. Exclusion criteria included a speech and/or language disturbance or a co-existing neurological disorder inconsistent with Parkinson's disease, a severe uncorrected auditory and/or visual disturbance, a cognitive disturbance inconsistent with the capacity to provide informed consent, a respiratory dysfunction unrelated to the neurological disorder and a positive history of alcohol abuse. The primary mode of assessment (online or face-to-face led) was randomly selected for each participant and 31 assessments were led face-to-face and 30 led online. A computerized random-number generator was used for the randomization.

Assessors

Three SLPs experienced in the assessment of motor speech disorders and Parkinson's disease took part in the study. Assessments were conducted at The University of Queensland. Prior to the commencement of the study, SLP training was conducted in a 3-hour session which covered the administration of all assessments in both the online and face-to-face assessment environments. The SLPs were deemed competent with online administration when they could adequately deliver a mock session within a 1-hour time frame and also agree on the level of severity of five dysarthric speakers who were not involved in the study. The speakers were judged on the perceptual measures of voice, overall articulatory precision, overall speech intelligibility in conversation and oromotor function. During the study, two of the three SLPs took part in each assessment session, where

one SLP led the session, while the second SLP acted as a silent rater and did not interact with the participant. One SLP assessed the participant in the face-to-face environment (within the same room as the participant), while the second SLP conducted the assessment in the online environment, through a videoconferencing link via the Internet. The SLPs were also randomized to the assessment environments and were blind to the participant and their level of hypokinetic dysarthria prior to assessment. In total, SLP 1 took part in 25 of the online assessments (16 as leader and 9 as silent rater) and 20 of the face-to-face assessments (9 as leader and 11 as silent rater); SLP 2 took part in 21 online sessions (7 as leader and 14 as silent rater) and 24 face-to-face assessments (11 as leader and 13 as silent rater); and SLP 3 took part in 14 online assessments (7 as leader and 7 as silent rater) and 18 face-to-face assessments (12 as leader and 6 as silent rater).

Assessment battery

Each participant underwent a 1-hour assessment on one occasion on a battery of perceptual and acoustic measures specifically designed for this study. The battery consisted of perceptual ratings of voice and oromotor parameters, overall articulatory precision, speech intelligibility in reading and conversation, and an instrumental evaluation of sound pressure levels, duration of vowel prolongation and pitch range. These measures were chosen for the study as they are commonly used to diagnose and define the level of severity of hypokinetic dysarthria associated with Parkinson's disease. Furthermore, as this study forms part of a larger validation trial that also evaluates online treatment, the measures were chosen as they have been used in the LSVT[®] literature as sensitive predictors of treatment change (Ramig *et al.* 1995a).

Perceptual measures

Perceptual voice parameters

In the absence of a standardized voice assessment available at the time of the study, vocal parameters were evaluated using a five-point rating scale developed for the study (1 = normal, 5 = severely impaired). The reading of a standard passage, The Grandfather Passage (Darley *et al.* 1975) was used for perceptual ratings of breathiness, roughness (lack of clarity), strain-strangled vocal quality, vocal tremor, pitch and phonation breaks. Modal pitch and loudness levels and pitch and loudness variability were rated on a 30 s conversational monologue about a topic of interest such as family, hobbies or a recent holiday trip.

Oromotor function

An informal assessment of non-speech oromotor function was developed to evaluate specific parameters using a five-point rating scale (1 = normal, 5 = severely impaired). The parameters included masked facial expression, lip movement (retraction, pucker, seal, alternate movement), tongue movement (symmetry, protrusion, elevation/depression, lateral and alternate movement), breath support and diadochokinetic (DDK) rates (alternate motion rate [AMR]/pʌ/pʌ/and sequential motion rate [SMR]/pʌtʌkʌ/).

Overall articulatory precision

A perceptual rating of each participant's articulatory precision was made from the speech sample obtained during the reading of The Grandfather Passage. Articulatory precision was rated on a five-point scale (1 = normal, precise production of sounds, 5 = severe distortion or imprecision that interferes with speech intelligibility).

Measures of speech intelligibility

The Assessment of Intelligibility of Dysarthric Speech (ASSIDS) (Yorkston and Beukelman 1981) was used to measure speech intelligibility at the single word and sentence level, as well as communication efficiency. For this task, participants read or repeated a series of 50 words and 22 sentences of increasing length. The words and sentences had been randomly generated prior to the assessment, in accordance with test procedure. The reading material was displayed on the participant's screen or presented as per the test booklet, depending on the assessment environment. Copyright approval was obtained from the publishers (PRO-ED, Austin, TX, USA) to enable conversion of test materials to an online format. Audio recordings of participant speech samples were made in both environments. Following assessments of all participants, the speech samples obtained face-to-face and online were numerically coded, randomized and saved to CD for analysis and scoring. Two independent SLPs who did not participate in the study and who were blinded to the assessment environment and participants transcribed the speech samples obtained in each environment. Following the ASSIDS ratings, the values given by the two SLPs from the online and face-to-face recordings were averaged to express a single mean value for each sample obtained in that environment. Scores for the word and sentence intelligibility tasks were expressed as per cent correct. The communication efficiency ratio was determined by dividing the participant's rate of intelligible speech (intelligible words per minute) by the mean rate of intelligible words per minute for

normal speakers (190 words per minute) (Yorkston and Beukelman 1981).

An additional rating of the participant's overall speech intelligibility in conversation was made from the 30 s monologue sample using a five-point scale (1 = normal, completely intelligible speech, 5 = severely unintelligible speech with difficulties deciphering many words). For this task, the rating was made by the two SLPs who took part in the online and face-to-face assessment.

Acoustic measures

The LSVT[®] Evaluation Protocol (Ramig *et al.* 1995b) was used to assess the participant's sound pressure levels (SPL), duration of vowel prolongation and pitch range during several speaking tasks. This protocol has been widely used in the LSVT[®] literature as a routine assessment.

Sound pressure levels and duration of vowel prolongation

The SPLs (dB-C) of the participant's speech were recorded during six maximum sustained vowel phonation of /a/, readings of the Rainbow Passage (Fairbanks 1960) and The Grandfather Passage, and during a 30 s conversational monologue. The duration of each vowel phonation was also measured in seconds. For all tasks, the participants were instructed to speak in a comfortable voice and no reference was made to their loudness level. Following the assessment, the SPL and duration levels were then averaged to provide mean levels for each participant.

Pitch range

Each participant performed a series of six vocal glides, reaching their highest and lowest pitch levels respectively. No reference was made to their loudness level. The average highest and lowest frequency levels (Hz) obtained for each participant were then converted to a maximum range in semitones (ST) (de Pijper 2007).

Participant satisfaction questionnaire

The 30 participants in the online-led assessments completed a brief questionnaire. On a five-point scale, the questionnaire evaluated the level of participant satisfaction with: (1) the online assessment sessions (possible responses ranging from would not participate again to would prefer these types of sessions to face-to-face sessions); (2) the audio and video quality during the sessions (responses ranging from poor to excellent); and (3) overall satisfaction with the online modality (ranging from not at all satisfied to very satisfied). Please refer to the Appendix.

Assessment environment

Online environment

Two personal computer-based videoconferencing systems developed at The University of Queensland were used for the online assessment. The applications operated on a 128 kbit/s Internet connection which was the minimum connection speed available in Queensland's public health systems at the time of the study. Videoconferencing at 320 × 240 pixel resolution was conducted between the online SLP's computer and that of the participant. Additional features of the system which were used in this study included the ability: to display printed material and instructional video clips on the participant's screen; to control the remote camera with the use of a robotic arm and adjust its alignment for optimal viewing of the participant's head and upper torso; to capture high-quality video (640 × 480 pixel resolution compressed with the windows media video CODEC Version 8 at 384 kbit/s) and audio recordings (windows media audio CODEC Version 8 at 368 kbit/s) independent of videoconferencing for the oromotor tasks, and then to store-and-forward these audio and video files back to the online SLP for later review.

For tasks requiring acoustic measures, both the online and face-to-face SLPs were able to view and sample real-time calibrated average recordings of SPL (dB-C), peak frequency (Hz) and duration (s) data via the system's acoustic speech processor specifically developed for this study. The validity of the speech processor as an acoustic measurement device was examined in a series of calibration trials. These trials involved the generation of pure-tones by a Function Generator (Topward Electronic Instruments Model TFG-462) at varying levels of SPL (55–95 dB) and pitch (100–975 Hz), and comparing measures from the acoustic speech processor with those of the commercially available Visi-Pitch II (Kay Elemetrics Model No. 3300). Statistical analyses using paired *t*-tests revealed no significant differences ($p > 0.05$) in SPL and pitch measures for pure-tones between the two devices. Furthermore, verification trials comparing SPL measures using the speech processor with those from a Digital Sound Level Meter (Radio Shack[®] Model No. 23-553) using voice samples (66–91 dB) also revealed no significant differences ($p > 0.05$) between the two devices (table 1). To standardize the acoustic measures across the two assessment environments, the system's acoustic speech processor was used as the objective measurement tool in both the online and face-to-face environments for all acoustic measures.

During the online assessment, the online SLP wore a headset microphone attached to the telerehabilitation system for communication with the participant.

Table 1. Calibration trial for measures of pitch and sound pressure levels

Task variable	Speech processor, mean (SD)	Visi-Pitch II, mean (SD)	SLM, mean (SD)	MAD speech processor and Visi-Pitch (SD)	MAD speech processor and SLM (SD)	T-value (<i>t</i> -test)	p-value (<i>p</i> -test)
Pure-tone pitch (Hz)	535.94 (263.32)	536.85 (263.86)	–	1.46 (3.66)	–	– 0.015	0.988
Pure-tone SPL (dB-C)	77.95 (9.70)	77.71 (9.77)	–	0.36 (0.19)	–	0.077	0.939
Voice samples SPL (dB-C)	79.36 (7.67)	–	79.43 (7.53)	–	0.5 (0.51)	– 0.025	0.980

Note: Dashes (–) correspond to data not obtained. Speech processor is the system's online acoustic speech processor. Visi-Pitch II (Kay Elemetrics Model No. 3300). SLM is a Digital Sound-Level Meter (Radio Shack® Model No. 23-553). MAD is maximum average difference. Pure-tone pitch is the pitch calibration trial of the speech processor with the Visi-Pitch II using pure-tones. Pure-tone SPL is the sound pressure level calibration trial of the speech processor with the Visi-Pitch II using pure-tones. Voice samples SPL is the sound pressure level verification trial comparing sound pressure level measures using the speech processor with the Digital Sound Level Meter for voice samples. Measurements are in Hertz. dB-C is measurements in decibels-C weighted. SD, standard deviation.

The SLP controlled all displays on the participant's screen, without the need for the participant to operate the system. For standardization purposes, the participant was seated in front of the system at a distance of approximately 50 cm from the monitor and wore a headset microphone to enable interaction with the online SLP during videoconferencing. The microphone distance was set at 5 cm from the corner of the participant's mouth in order to reduce sound distortion, maximize visibility of the participant's face, and allow for accurate recordings of pitch and SPL.

During the online-led assessment, the online SLP administered the various tasks and interacted with the participant over the 128 kbit/s Internet videoconferencing link. At this bandwidth, live ratings of pitch and SPL were possible, however, judgements of fine movements and precision on the oromotor assessment were more difficult due to a low picture frame rate and resolution picture quality. In addition, the real-time detection of subtle features of speech production for perceptual ratings of voice, overall articulatory precision and speech intelligibility was also more difficult on occasion due to the degradation of audio quality. Therefore, to improve the video and audio quality for rating, the online SLP used the store-and-forward features of the system to record the task and store the video and audio files for later viewing and analyses. The store-and-forward feature was used routinely by all online SLPs (leading and silent assessors). To standardize the perceptual measures, SLPs in both environments rated the assessments live (where possible) and then reviewed the sessions off-line using the equipment available in that environment. A summary of the online assessment procedure is displayed in table 2.

In addition, the effects of the audio-visual difficulties with videoconferencing on the participants' ability to follow task instructions were minimized with the use of pre-recorded task demonstrations of the oromotor assessment. Where necessary, these demonstrations were

displayed on the participant's screen by the online SLP. Throughout the online-led assessment, the face-to-face SLP acted as the silent rater at the participant site. The face-to-face SLP wore headphones and was able to follow the assessment instructions given to the participant. The online assessment environment is represented in figures 1a and 1b.

Face-to-face environment

During the face-to-face led assessment, the participant was seated in front of the telerehabilitation system with the monitor turned off. Standard face-to-face test administration procedures were used. The online SLP became the silent rater and viewed, listened, and recorded tasks while the face-to-face SLP interacted with the participant. In keeping with the online procedure, the face-to-face SLP obtained real-time measures of SPL, duration and pitch via the system's acoustic speech processor. The SLP also used a video camera and minidisk recorder to collect video and audio data respectively for later analyses, as per standard clinical practice. The video camera was positioned as close as possible behind the web cameras and a microphone on a stand was connected to a minidisk recorder and placed 30 cm from the participant. The face-to-face SLP wore headphones and was able to hear the online SLP if there was a need for a task repetition or further online recording. A summary of the face-to-face assessment procedure is also displayed in table 2.

Statistical analyses

Online and face-to-face ratings and measurements for all participants were compared on each assessment task to determine the level of agreement between the two environments. For those parameters consisting of ordinal data (perceptual ratings of voice and oromotor parameters, overall articulatory precision and speech

Table 2. Assessment procedure for online and face-to-face environment

Assessment measure	Online instrument	Online scoring procedure	FTF instrument	FTF scoring procedure
<i>Perceptual Measures</i>				
Perceptual voice parameters	Online store and forward audio	Real-time where possible and reviewed off-line	Minidisk recorder	Real-time where possible and reviewed off-line
Oromotor parameters	Online store and forward video	Real-time where possible and reviewed off-line	Video camera	Real-time where possible and reviewed off-line
Overall articulatory precision	Online store and forward audio	Real-time where possible and reviewed off-line	Minidisk recorder	Real-time where possible and reviewed off-line
Overall speech intelligibility in conversation	Online store and forward audio	Real-time where possible and reviewed off-line	Minidisk recorder	Real-time where possible and reviewed off-line
ASSIDS	Online store and forward audio	Off-line	Minidisk recorder	Offline
<i>Acoustic measures</i>				
Sound pressure levels	Online acoustic speech processor	Real-time	Online acoustic speech processor	Real-time
Duration of phonation	Online acoustic speech processor	Real-time	Online acoustic speech processor	Real-time
Pitch range	Online acoustic speech processor	Real-time	Online acoustic speech processor	Real-time
Participant satisfaction questionnaire	Paper based	End of online-led assessment session	n.a.	n.a.

Note: ASSIDS = assessment of intelligibility of dysarthric speech; FTF, face-to-face assessment environment; n.a., not applicable.

intelligibility in conversation), percent close agreement (PCA) and the quadratic weighted Kappa (κ_w) statistic (Landis and Koch 1977) were calculated. Analysis of the ASSIDS and acoustic parameters (SPL tasks, duration of vowel phonation and pitch range) were performed using the Bland and Altman (1986) 'limits of agreement' method for continuous data.

Percent close agreement

PCA was chosen as it is commonly used to quantify agreement in perceptual ratings of dysarthria. PCA was also selected for the present study to further verify κ_w as it has been reported in some instances that non-linear distribution of data can negatively impact on κ_w creating a paradox (Cicchetti and Feinstein 1990). PCA expressed the percentage of ratings where differences were within ± 1 scale point on the perceptual rating scales (Kearns and Simmons 1988). In keeping with previous studies that examined dysarthric speech using perceptual rating scales, the clinical criterion for an acceptable level of agreement in the present study was considered to be equal to or greater than 80% agreement within ± 1 scale point (Kearns and Simmons 1988).

Quadratic weighted Kappa statistic

The κ_w is widely used in telerehabilitation studies for ordinal data and provides an indication of agreement between raters (Landis and Koch 1977). In the present study, the statistic provided a measure of agreement

beyond chance between the online and face-to-face measures. The κ_w assigned weights to the observed and chance agreement and presented levels of agreement where κ_w less than 0.20 is interpreted as poor; 0.21–0.40 is fair; 0.41–0.60 is moderate; 0.61–0.80 is good, and 0.81–1.00 indicates very good agreement (Landis and Koch 1977). For this study, the clinical criterion for an acceptable level of agreement was set at $\kappa_w > 0.6$ (good agreement).

Bland and Altman (1986) limits of agreement

This statistic establishes the limits of agreement (LA) within which 95% of differences between the two environments are predicted to lie. If the LA are found to be within a predetermined clinical criterion, the new method can be considered an acceptable measurement tool and the two methods can be used interchangeably (Bland and Altman 1986). In the absence of a reported minimal, clinically important difference for word and sentence intelligibility of the ASSIDS assessment, the clinical criterion was established on the test–retest variability reported in the manual for dysarthric speakers assessed in the face-to-face environment (Yorkston and Beukelman 1981). Values of $\pm 3.2\%$ and $\pm 8.6\%$ were set for these respective measures. In addition, the clinical criterion for the communication efficiency ratio was set at ± 0.27 which was consistent with the criterion determined by Hill *et al.* (2006) for dysarthric speakers. Furthermore, as the assessment battery used in the study was designed ultimately to



Figure 1a. Online-led assessment by online SLP and equipment at site. Note (1) the videoconferencing system displaying the participant; (2) pitch (Hz) and SPL (dB-C) data via the system's acoustic speech processor; and (3) the web camera.

determine treatment outcome, the clinical criteria for SPL, vowel duration and pitch range tasks were set at levels below the minimal improvement expected following the LSVT[®]. For all SPL tasks (maximum vowel phonation, reading and monologue loudness), the clinical criterion was set at ± 4 dB difference between the two raters, a level below the 4.5–14.03 dB mean level of improvement reported with the LSVT[®]

(Ramig *et al.* 1995a). The clinical criterion for the duration of vowel phonation was set at ± 3 s, as the minimal change in phonation time expected with the LSVT[®] has been reported to be a mean of 3.72 s (Ramig *et al.* 1995a). For measures of pitch range, the clinical criterion was set at ± 3 ST, which was below the 4 ST minimum improvement in fundamental frequency range post-LSVT[®] (Ramig *et al.* 1994). This clinical



Figure 1b. Online-led assessment at participant site with face-to-face SLP as the silent rater (left). Note (1) the videoconferencing system displaying the online SLP; (2) the web cameras; (3) the video camera; and (4) pitch (Hz) and SPL (dB-C) data via the system's acoustic speech processor displayed for the face-to-face clinician.

criterion was also in keeping with the level of subject variability in healthy adults that can range from 2 to 4 ST (Gelfer 1986).

Reliability

Reliability between the online and face-to-face environments was conducted for all the perceptually based assessments and the acoustic pitch measure. Although the pitch data was objectively obtained via the acoustic speech processor, the SLPs were required to select a sample pitch level from a section of the vocal glide thus introducing a subjective element to this task. The SLPs used the audio, video and pitch files captured during the assessment session in the respective environments to rate and score the various parameters. Intra- and inter-rater reliability between the online and face-to-face assessors was calculated using two-way, random effect intra-class correlations ($ICC_{(2,1)}$) for 20% ($n = 13$) of participants in each environment. For inter-rater reliability, the third SLP who did not take part in a particular assessment session became the additional rater and was randomly assigned to the online or face-to-face ratings. Intra- and inter-rater reliability was calculated collectively using the ratings from each of the three SLPs in the particular environment. ICC values below 0.40 corresponded to poor-to-fair reliability; between 0.40 and 0.75 to moderate-to-good reliability; and values above 0.75 represented very good reliability (Fleiss 1981).

Results

Perceptual measures

Perceptual voice parameters

All individual voice parameters met the predetermined clinical criterion of 80% agreement for PCA (table 3). However, when using κ_w , seven of the ten parameters (breathiness, roughness, strained-strangled, pitch

Table 3. Perceptual voice parameters between face-to-face and online environments

Voice parameters	PCA	κ_w
Breathiness	91.80	0.36 (fair)*
Roughness	95.08	0.33 (fair)*
Strained-strangled	95.08	0.41 (moderate)*
Vocal tremor	100	0.69 (good)
Pitch breaks	96.66	0.11 (poor)*
Phonation breaks	95.00	0.37 (fair)*
Modal pitch	96.66	0.38 (fair)*
Pitch variability	96.72	0.63 (good)
Loudness level	100	0.69 (good)
Loudness variability	98.36	0.49 (moderate)*

Note: κ_w , quadratic weighted Kappa statistic; *achieved lower than the clinical criterion of $\kappa_w > 0.6$; PCA, per cent close agreement.

breaks, phonation breaks, modal pitch and loudness variability) were below the clinical criterion of good agreement ($\kappa_w > 0.6$).

Oromotor function parameters

Analyses revealed that all individual oromotor parameters reached the clinical criterion for PCA (table 4). The κ_w indicated that only two parameters (masked facial expression and lip retraction) fell outside of the clinical criterion of good agreement.

Overall articulatory precision

For ratings of overall articulatory precision between the online and face-to-face environments, PCA (100) and κ_w (0.67 good agreement) were within the clinical criteria.

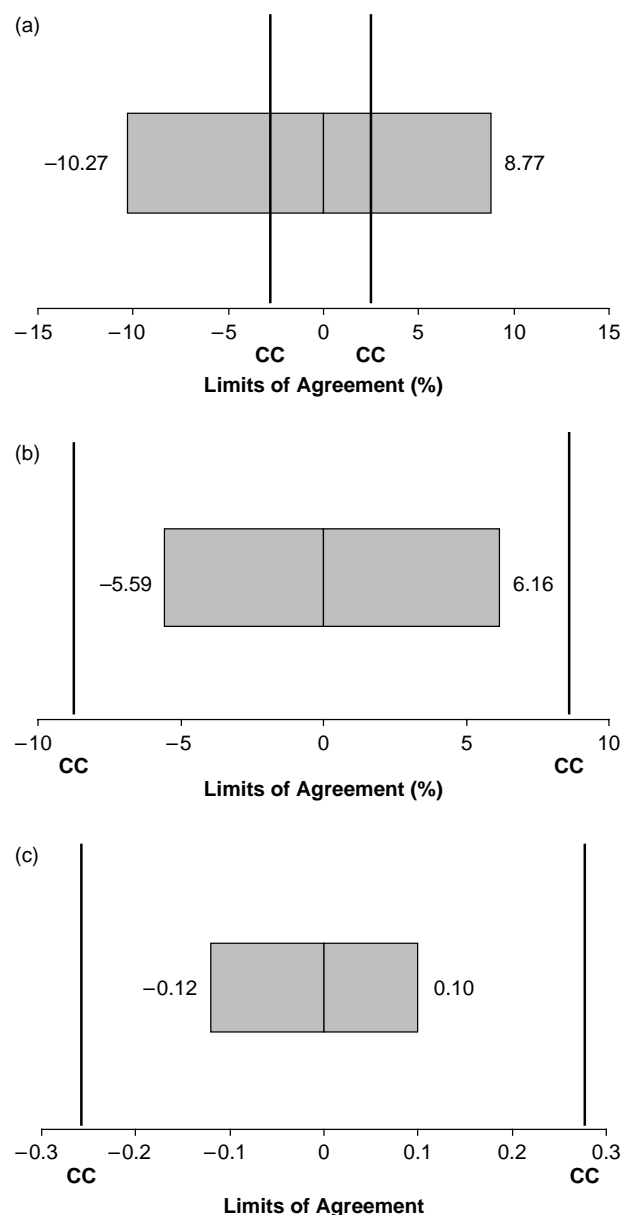
Measures of intelligibility

For the ASSIDS assessment, the Bland and Altman (1986) LA at the 95% confidence interval are displayed in figures 2a–2c for word (LA = -10.27% to 8.77%) and sentence intelligibility reading tasks (LA = -5.59% to 6.16%), and the communication efficiency ratio (LA = -0.12 to 0.10). The LA for sentence intelligibility and communication efficiency ratio were within the respective clinical criterion ($\pm 8.6\%$ and ± 0.27), while the word intelligibility LA fell outside of the clinical criterion of $\pm 3.2\%$. In addition, perceptual ratings of overall speech intelligibility in conversation were within the clinical criteria for PCA (98.36) and κ_w (0.79 good agreement).

Table 4. Perceptual oromotor parameters between face-to-face and online environments

Oromotor parameters	PCA	κ_w
Breath support	96.72	0.83 (very good)
Masked facial expression	86.89	0.31 (fair)*
<i>Lip movement</i>		
Retraction	98.36	0.56 (moderate)*
Pucker	98.36	0.77 (good)
Seal	95.08	0.66 (good)
Alternate	100	0.95 (very good)
<i>Tongue movement</i>		
Symmetry	100	0.66 (good)
Protrusion	100	0.94 (very good)
Elevation/depression	98.36	0.93 (very good)
Lateral	100	0.89 (very good)
Alternate	100	0.85 (very good)
<i>Diadochokinetic (DDK)</i>		
AMR /pʌpʌ/	100	0.75 (good)
SMR /pʌtʌkʌ/	100	0.87 (very good)

Note: κ_w = quadratic weighted Kappa statistic; *achieved lower than the clinical criterion of $\kappa_w > 0.6$; PCA, per cent close agreement.



Figures 2a–c. Bland and Altman (1986) limits of agreement for word intelligibility, sentence intelligibility, and communication efficiency ratio relative to the clinical criteria: (a) word intelligibility: clinical criterion (CC) = ± 3.2 percentage points; (b) sentence intelligibility: clinical criterion (CC) = ± 8.6 percentage points; and (c) communication efficiency ratio: clinical criterion (CC) = ± 0.27 .

Acoustic measures

Sound pressure levels

The Bland and Altman (1986) LA are displayed in figure 3 for sustained vowel phonation (LA = -1.97 to 1.35 dB), the reading of the Rainbow Passage (LA = -1.05 to 1.04 dB) and The Grandfather Passage (LA = -1.18 to 1.11 dB), and monologue loudness

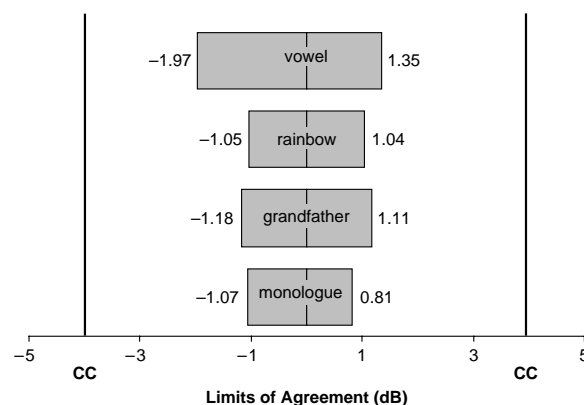


Figure 3. Bland and Altman (1986) limits of agreement for sustained vowel phonation, reading of the Rainbow and Grandfather Passages and monologue task relative to the clinical criteria. Clinical criteria (CC) for all SPL tasks = ± 4 dB.

(LA = -1.07 to 0.81 dB). For all SPL tasks, the LA were within the predetermined clinical criterion of ± 4 dB. Similarly, the LA for the duration of sustained vowel phonation task (LA = -2.74 to 2.70 s) were within the clinical criterion of ± 3 s for online and face-to-face ratings (figure 4a).

Pitch range

Figure 4b represents the LA for the pitch range (LA = -2.03 to 2.19 ST) that were within the clinical criterion of ± 3 ST.

Reliability

Intra-class correlations ranged from moderate to very good intra-rater reliability in both assessment environments (ICC = 0.43 – 0.99 face-to-face; ICC = 0.48 – 0.99 online), indicating comparable intra-rater reliability between environments. Inter-rater reliability was also found to be comparable between environments, with reliability values between moderate to very good for the majority of face-to-face (ICC = 0.43 – 0.99) and online (ICC = 0.48 – 0.99) assessments (table 5).

Participant satisfaction questionnaire

On the participant satisfaction questionnaire, the majority of participants in the online-led assessments felt comfortable while participating in the online session (56.67%) or very happy with the session (36.67%). The audio quality during videoconferencing was largely rated as excellent (40%), adequate (30%) and more than adequate (23.33%), while the video quality was primarily found to be adequate (33.33%) or more than adequate (30%).

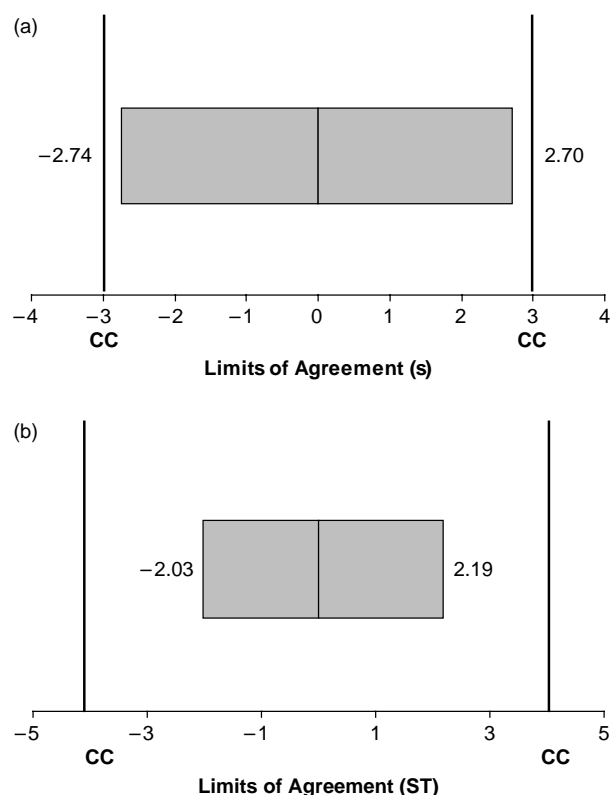


Figure 4a, b. Bland and Altman (1986) limits of agreement for the duration of sustained phonation and pitch range relative to the clinical criterion: (a) duration of sustained phonation: clinical criterion (CC) = ± 3 s; and (b) pitch range: clinical criterion (CC) = ± 4 ST.

Participant satisfaction with the online modality overall ranged from very satisfied (50%) to more than satisfied (30%) and satisfied (20%).

Discussion

The results of the present study indicated that an Internet-based assessment of the disordered speech and voice associated with Parkinson's disease was generally reliable and valid. For the majority of the perceptual and acoustic parameters, the face-to-face and online ratings were within the clinical criteria as reported in previous face-to-face studies.

Perceptual measures

Perceptual voice parameters

Analysis of individual voice parameters showed PCA between the online and face-to-face ratings to be within the clinical criterion. This suggested that an Internet-based videoconferencing application is a valid tool for administering a voice evaluation for Parkinson's disease and the audio store-and-forward capabilities of the

system were sensitive enough for the SLP to determine the presence and level of severity of specific voice parameters online. However, at the more stringent level of analysis (κ_w), the online and face-to-face assessments of several of the individual voice parameters failed to reach acceptable levels of agreement. The κ_w revealed greater variability than PCA for seven of the ten voice parameters (breathiness, roughness, strained-strangled vocal quality, pitch breaks, phonation breaks, modal pitch and loudness variability) that fell below clinical criterion (achieving poor to moderate agreement). These lower values may reflect rater variability commonly seen in perceptual rating of voice and/or the nature of κ_w .

Findings of lower agreement for vocal parameters in an online assessment were similarly reported by Hill *et al.* (2006) in their study, where consensus agreement was below 80% for ratings of strained-strangled vocal quality (73.68%), breathiness (68.42%) and roughness parameters (63.16%). These authors attributed aspects of the lower agreement to possible inter-rater variability that is inherent in perceptual ratings of voice. It has been acknowledged that naturally occurring variability is commonly associated with traditional face-to-face evaluations and reflects the subjective nature of perceptual rating scales. The listener often applies variable internal standards of pathological voice qualities from their own experiences to the evaluation (Kreiman and Gerratt 1998). Consequently, achieving high agreement between different raters in relation to voice qualities is often problematic. A further explanation for the lower κ_w relates to the statistic per se. It has been found that κ_w may be negatively influenced by the data distribution, where despite high inter-rater agreement, calculations using non-linear data can result in low κ_w (Cicchetti and Feinstein 1990). For example, in the present study, strain-strangled vocal quality (non-linear data), where although receiving comparable PCA to the vocal tremor parameter (linear data), obtained only moderate agreement according to κ_w , in contrast to good agreement for vocal tremor (table 3).

Further support for the comparability of the online and face-to-face assessment environments in the evaluation of all vocal parameters was provided by the reliability measures obtained for these ratings. Intra- and inter-rater reliability were found to be similar in each assessment environment (table 5), suggesting that the online assessment was as reliable as the traditional face-to-face assessment on this task.

Oromotor parameters

For the purpose of assessing oromotor function, the two assessment environments were found to be clinically comparable (PCA between 86.89% and 100%) (table 4).

Table 5. Intra-class correlation (ICC) values for intra-rater and inter-rater reliability for online and face-to-face ratings

Assessment task	FTF intra-rater reliability	Online intra-rater reliability	FTF inter-rater reliability	Online inter-rater reliability
Oromotor parameters	0.85	0.81	0.74	0.76
Perceptual voice parameters	0.60	0.68	0.44	0.56
OAP	0.43	0.48	0.43	0.48
OIC	0.63	0.69	0.75	0.82
Pitch range	0.99	0.99	0.99	0.99
ASSIDS				
WI	0.93	0.99	0.89	0.94
SI	0.97	0.77	0.94	0.82
CER	0.99	0.97	0.97	0.98

Note: CER, communication efficiency ratio; FTF, face-to-face environment; OAP, overall articulatory precision; OIC, overall speech intelligibility in conversation; ASSIDS, assessment of intelligibility of dysarthric speech; SI, percentage sentence intelligibility; WI, percentage word intelligibility; ICC values below 0.40 = poor-to-fair; from 0.40 to 0.75 = moderate-to-good, above 0.75 = very good reliability. Reliability was calculated for three assessors.

Two of the 14 variables (masked facial expression and lip retraction), however, were below the clinical criterion using κ_w . As noted with the vocal parameters, a certain level of variability inherent in perceptual ratings may have contributed to the lower ratings for these oromotor parameters. Previous face-to-face evaluations of facial expression in Parkinson's disease using a range of rating scales and statistical analyses have shown varying levels of intra- and inter-rater reliability including fair, moderate and substantial agreement (Goetz *et al.* 1995, Martinez-Martin *et al.* 1994). The authors of these studies attributed some level of the variability to the subjective interpretation of severity levels, the possible inexperience of a few of the raters, and some variability in consensus prior to rating.

Similarly, labial judgements which have been investigated predominantly in the cleft palate literature, have been associated with rater variability (Morrant and Shaw 1996, Ritter *et al.* 2002). Face-to-face evaluations of lip retraction in participants with repaired unilateral cleft lip have shown poor (Morrant and Shaw 1996) and moderate levels of inter-rater agreement (Ritter *et al.* 2002). The subjective interpretation of severity levels has also been reported to affect rater agreement in these studies (Morrant and Shaw 1996, Ritter *et al.* 2002).

Analyses of the oromotor parameters including masked facial expression and lip retraction using κ_w may also require a level of cautious interpretation due to the non-linear distribution of the data, and results may need to be interpreted alongside PCA. On the whole, the high intra- and inter-rater reliability obtained for the oromotor parameters collectively, and the comparable levels between the two environments were very encouraging.

Overall articulatory precision

The complete agreement obtained between the two assessment environments for articulatory precision is

consistent with a previous study by Hill *et al.* (2006) where 89.47% consensus agreement was achieved between online and face-to-face ratings for consonant precision. Direct comparison of outcomes between the two studies is not possible, however, due to the different assessment procedures used by Hill *et al.* (2006) including the use of a four-point rating scale, different evaluation criteria and the non-simultaneous assessment of participants in the online and face-to-face environments. In the present study, the high level of agreement between online and face-to-face ratings and the comparable intra- and inter-rater reliability values (moderate agreement) between environments lends further support to the validity of an online application.

Measures of speech intelligibility

The Bland and Altman (1986) LA were used for online and face-to-face scores of the ASSIDS assessment (reading tasks). For the sentence tasks, both the LA for sentence intelligibility and communication efficiency ratio were within the clinical criterion, indicating that comparable measures of speech intelligibility can be achieved between the online store-and-forward method and traditional face-to-face audio recordings using this assessment. Hill *et al.* (2006) similarly reported comparable values for the communication efficiency ratio in their study, while sentence intelligibility was just outside the clinical criterion. In the present study, the LA for word intelligibility were outside the clinical criterion of ± 3.2 percentage points between the environments. It is possible that speaker severity may have influenced these results. Differences of three or more words between raters, which were outside the clinical criterion, occurred predominantly for participants with moderate and severely reduced intelligibility (66.66% of the time), as identified on the overall speech intelligibility in conversation rating scale. The reduced

speaker intelligibility may have contributed to the differences in ratings between the two environments. Yorkston and Beukelman (1981) acknowledge that transcription of word tasks (as used in this study) in the traditional face-to-face environment is difficult with more severely dysarthric speakers. For all the ASSIDS tasks, intra- and inter-rater reliability was largely comparable between the two environments, and showed very good agreement overall. These values are in keeping with the very good reliability measures reported for the ASSIDS assessment (Yorkston and Beukelman 1981). Comparable levels between assessment environments and previous literature lend further support to the use of this assessment in an online application.

For the additional ratings of overall speech intelligibility in conversation, PCA within the clinical criterion was achieved. This finding is in keeping with previous reports of high inter-rater agreement within one scale point for overall speech intelligibility in traditional face-to-face ratings (Sheard *et al.* 1991). The κ_w further reflected good agreement within the clinical criterion. Together with the comparable reliability measures, these findings suggest that ratings of overall speech intelligibility in conversation can be made reliably online.

Acoustic measures

Objective measures of SPL, vowel duration and pitch range were obtained in real-time via the system's acoustic speech processor in both environments. Although the calibration trials demonstrated the validity of the speech processor as an objective tool (table 1), it was important to assess the performance of the speech processor during each of the different assessment modes to determine if there was an effect of transmission across the Internet. Acoustic measures were analysed using the Bland and Altman (1986) LA. For the SPL tasks (sustained vowel phonation, reading and conversational loudness), all measures were within the clinical criterion of ± 4 dB. This finding is not surprising as the acoustic speech processor provided objective measures of SPL which were consistent between environments. The minor differences in values most likely reflected the slightly unsynchronized start of SPL sampling by the two SLPs in each environment, while transmission across the Internet appeared to have little effect on the task (figure 3). Measures of sustained vowel duration were also within the clinical criterion of acceptable differences (± 3 s), and the minor differences in duration may also have reflected the subjective element in initiating the sampling (figure 4a).

For the pitch range task, the LA were also within the predetermined clinical criterion of ± 3 ST (figure 4b). This further demonstrates that SLPs in both environments were able to obtain comparable pitch values

within a clinically acceptable level, and the subjective element of selecting a sample from a section of the vocal glide for analysis did not impact greatly on the results. Moreover, reliability measures revealed very good intra- and inter-rater reliability for the pitch task between the two environments. Collectively, the comparable values obtained for all acoustic measures suggested that the acoustic speech processor used in the online environment is a sensitive assessment tool that can be used to detect minimal changes in SPL, duration and pitch in a Parkinson's disease assessment battery.

Audio-visual challenges and the online environment

Although the current trials largely support the feasibility of an online assessment, it is acknowledged that a number of challenges were experienced with this modality. Firstly, the assessments were conducted over a 128 kbit/s Internet videoconferencing connection, which at this bandwidth, made the real-time evaluation of a number of assessment items difficult. This included the detection of fine motor movements and precision on the informal oromotor assessment due to the frame rate and pixelated image, especially with movement. The more subtle features of speech production on the overall articulatory precision and speech intelligibility tasks were also difficult to rate due to the less than optimal audio quality that occurred intermittently. As mentioned previously, this was especially evident for participants with more severe dysarthria. The use of the store-and-forward capabilities of the online system did allow for high quality audio and video recordings and helped to minimize the audio-visual difficulties associated with real-time videoconferencing.

Despite the advantages of the store-and-forward method, other difficulties were encountered that affected the ratings. Factors such as shadowing on the participant's face or reduced contrast of facial features due to background lighting and/or lack of webcam zoom function, and considerable pixelation of a few video recordings on occasion did impact upon the online ratings of lip and tongue symmetry, tongue deviation and general facial features. Such difficulties have also been reported in other online studies and it has been proposed that web cameras with greater zoom and focus capabilities and higher Internet bandwidth would possibly enhance the online ratings in real-time (Hill *et al.* 2006). Additional audio disturbances such as intermittent static in the recordings of the ASSIDS and reading passages were occasionally present within the store-and-forward modality, making some judgements of these parameters more difficult. Furthermore, participant factors such as head and body dystonias and stooped forward posture also made it more difficult on occasions to view the participant's entire face and

judge aspects of lip retraction, pucker, and tongue movements. However, these difficulties in judgement occurred independently of the assessment environment. Overall, the audio-visual disturbances that occurred with the store-and-forward method were infrequent and did not impact significantly on rater-agreement and reliability between the two environments.

In relation to the participant–clinician rapport, the occasional audio delays of up to three seconds over the videoconferencing link had the potential to affect general communicative interactions and turn-taking in the online environment. However, the SLPs and participants were able to quickly and effectively compensate for any disturbances by actively waiting until the other had clearly finished speaking before replying. Thus, participants did not report that this delay impacted greatly on the assessment. Overall, the online assessments were rated favourably by the participants, the application was user-friendly, and the features of the application were conducive to assessment.

Conclusion

The comparable ratings achieved for the majority of parameters between the online and face-to-face environments and high rater reliability have demonstrated the validity and reliability of the online assessment tool for evaluating the speech and voice parameters associated with hypokinetic dysarthria and Parkinson's disease. The telerehabilitation application described in this study provides a basis for the delivery of online assessment for the dysarthric speech and voice disorder associated with Parkinson's disease. Online service delivery may prove to be a necessary alternative or addition for people with Parkinson's disease, whereby lessening the difficult access issues that exist for this population. Further research should involve analyses of online rater variability involving a larger group of assessors and possible effects on assessment outcomes; provide a comprehensive analyses of participant and SLP satisfaction with the online modality; examine the effects of dysarthria severity and variance of Parkinson's disease on the online assessments; and provide an in-depth cost analysis of the use of telerehabilitation applications in the assessment and treatment of people with Parkinson's disease.

Acknowledgements

This research was funded by a National Health and Medical Research Council Project (Grant Number 301029). The research forms part of the doctoral thesis of Gabriella Constantinescu at The University of Queensland, Australia. The authors would like to acknowledge Parkinson's Queensland Incorporated, all participants, Roy Anderson, Anne Hill, Monique Waite, Jasmin

Cowles, and Christina Iezzi for their contribution to the research. **Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- ACCESS ECONOMICS, 2007, *Living with Parkinson's Disease: Challenges and Positive Steps for the Future*. Report for Parkinson's Australia (Access Economics).
- BLAND, J. M. and ALTMAN, D. G., 1986, Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, **i**, 307–310.
- BRENNAN, D. M., GEORGEADIS, A. C., BARON, C. R. and BARKER, L. M., 2004, The effect of videoconference-based telerehabilitation on story retelling performance by brain-injured subjects and its implications for remote speech–language therapy. *Telemedicine Journal and e-Health*, **10**, 147–154.
- CICCHETTI, D. V. and FEINSTEIN, A. R., 1990, High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, **43**, 551–558.
- DARLEY, F. L., ARONSON, A. E. and BROWN, J. R., 1969, Clusters of deviant speech dimensions in the dysarthrias. *Journal of speech and hearing research*, **12**, 462–496.
- DARLEY, F. L., ARONSON, A. E. and BROWN, J. R., 1975, *Motor Speech Disorders* (Philadelphia, PA: W. B. Saunders).
- DE PIJPER, J. R., 2007, *Semitone conversions* (available at: <http://users.utu.fi/jyrtuoma/speech/semitone.html>) (accessed 1 March 2007).
- DE RIJK, M. C., TZOURIO, C., BRETELER, M. M., DARTIGUES, J. F., AMADUCCI, L., LOPEZ-POUSA, S., MANUBENS-BERTRAN, J. M., ALPEROVITCH, A. and ROCCA, W. A., 1997, Prevalence of parkinsonism and Parkinson's disease in Europe: The EUROPARKINSON collaborative study. European community concerted action on the epidemiology of Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry*, **62**, 10–15.
- FAIRBANKS, G., 1960, *Voice and Articulation Drill Book* (New York, NY: Harper & Brothers).
- FLEISS, J. L., 1981, *Statistical Methods for Rates and Proportions* (New York, NY: Wiley).
- GELFER, M. P., 1986, The stability of total phonational frequency range. *Journal of the Acoustical Society of America*, **79**, 83.
- GEORGEADIS, A., BRENNAN, D., BARKER, L. M. and BARON, C., 2004, Telerehabilitation and its effect on story retelling by adults with neurogenic communication disorders. *Aphasiology*, **18**, 639–652.
- GOETZ, C. G., STEBBINS, G. T., CHMURA, T. A., FAHN, S., KLAUANS, H. L. and MARSDEN, C. D., 1995, Teaching tape for the motor section of the unified Parkinson's disease rating scale. *Movement Disorders*, **10**, 263–266.
- GRIMM, N., PAUL, J. and WAKEHAM, A., 2004, Access to speech pathology services by people with Parkinson's disease in Queensland, Australia (Unpublished report: University of Queensland, Australia).
- HARTELIUS, L. and SVENSSON, P., 1994, Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey. *Folia Phoniatrica et Logopaedica*, **46**, 9–17.
- HILL, A., THEODOROS, D. G., RUSSELL, T. G., CAHILL, L. M., WARD, E. C. and CLARK, K. M., 2006, An Internet-based telerehabilitation system for the assessment of motor speech disorders: a pilot study. *American Journal of Speech–Language Pathology*, **15**, 45–56.
- HILL, A., THEODOROS, D., RUSSELL, T. and WARD, E., 2008a, Using telerehabilitation to assess apraxia of speech in adults.

- International Journal of Language and Communication Disorders*, 1–17.
- HILL, A., THEODOROS, D., RUSSELL, T., WARD, E. and WOOTTON, R., 2008b, The effects of aphasia severity on the ability to assess language disorders via telerehabilitation. *Aphasiology*, [epub ahead of print] 1–16.
- HOEHN, M. and YAHR, M., 1967, Parkinsonism: onset, progression and mortality. *Neurology*, **17**, 427–442.
- KEARNS, K. P. and SIMMONS, N. N., 1988, Interobserver reliability and perceptual ratings: more than meets the ear. *Journal of Speech and Hearing Research*, **31**, 131–136.
- KREIMAN, J. and GERRATT, B. R., 1998, Validity of rating scale measures of voice quality. *Journal of the Acoustical Society of America*, **104**, 1598–1608.
- LANDIS, J. R. and KOCH, G. G., 1977, The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- MARTINEZ-MARTIN, P., GIL-NAGEL, A., GRACIA, L. M., GOMEZ, J. B., MARTINEZ-SARRIES, J. and BERMEJO, F., 1994, Unified Parkinson's Disease Rating Scale characteristics and structure. The Cooperative Multicentric Group. *Movement Disorders*, **9**, 76–83.
- MORRANT, D. G. and SHAW, W. C., 1996, Use of standardized video recordings to assess cleft surgery outcome. *Cleft Palate Craniofacial Journal*, **33**, 134–142.
- MUTCH, W. J., DINGWALL-FORDYCE, I., DOWNIE, A. W., PATERSON, J. G. and ROY, S. K., 1986, Parkinson's disease in a Scottish city. *British Medical Journal (Clinical Research Edn)*, **292**, 534–536.
- OXTOBY, M., 1982, *Parkinson's Disease Patients and Their Social Needs* (London: Parkinson's Disease Society).
- PALSBO, S. E., 2007, Equivalence of functional communication assessment in speech pathology using videoconferencing. *Journal of Telemedicine and Telecare*, **13**, 40–43.
- PETO, V., FITZPATRICK, R. and JENKINSON, C., 1997, Self-reported health status and access to health services in a community sample with Parkinson's disease. *Disability and Rehabilitation*, **19**, 97–103.
- RAMIG, L. O., BONITATI, C. M., LEMKE, J. H. and HORII, Y., 1994, Voice treatment for patients with Parkinson disease: development of an approach and preliminary efficacy data. *Journal of Medical Speech–Language Pathology*, **2**, 191–209.
- RAMIG, L. O., COUNTRYMAN, S., THOMPSON, L. L. and HORII, Y., 1995a, Comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research*, **38**, 1232–1251.
- RAMIG, L. O., FOX, C. and SAPIR, S., 2004, Parkinson's disease: speech and voice disorders and their treatment with the Lee Silverman Voice Treatment. *Seminars in Speech and Language*, **25**, 169–180.
- RAMIG, L. O., PAWLAS, A. A. and COUNTRYMAN, S., 1995b, *The Lee Silverman Voice Treatment (LSVT®): A Practical Guide to Treating the Voice and Speech Disorders in Parkinson Disease* (Iowa City, IA: National Centre for Voice and Speech).
- RITTER, K., TROTMAN, C. A. and PHILLIPS, C., 2002, Validity of subjective evaluations for the assessment of lip scarring and impairment. *Cleft Palate-Craniofacial Journal*, **39**, 587–596.
- SHEARD, C., ADAMS, R. D. and DAVIS, P. J., 1991, Reliability and agreement of ratings of ataxic dysarthric speech samples with varying intelligibility. *Journal of Speech and Hearing Research*, **34**, 285–293.
- THEODOROS, D., HILL, A., RUSSELL, T., WARD, E. and WOOTTON, R., 2008, Assessing acquired language disorders in adults via the Internet. *Telemedicine Journal and e-Health*, **14**, 552–559.
- YORKSTON, K. M. and BEUKELMAN, D. R., 1981, *Assessment of Intelligibility of Dysarthric Speech* (Austin, TX: PRO-ED).

Appendix: Online Participant Satisfaction Questionnaire

This questionnaire has been developed to determine your satisfaction with your assessment across the Internet. Please CIRCLE the answer that you feel is most appropriate based on your experience.

- How did you feel while participating in this session on the Internet?
 - Would prefer this type of session to face-to-face session
 - Very happy with this session
 - Comfortable
 - Uneasy
 - Would not participate again
- What is your opinion of the audio quality (what you were able to hear) during the session?
 - Excellent
 - More than adequate
 - Adequate
 - Inadequate
 - Poor
- What is your opinion of the visual quality (what you were able to see) during the session?
 - Excellent
 - More than adequate
 - Adequate
 - Inadequate
 - Poor
- Please rate your overall satisfaction with the Internet session.
 - Very satisfied
 - More than satisfied
 - Satisfied
 - Less than satisfied
 - Not at all satisfied