Critical Review:
Can online assessments be used to obtain an accurate measure of speech intelligibility in individuals with dysarthria?

Lorie Legge
M.Cl.Sc SLP Candidate
University of Western Ontario: School of Communication Sciences and Disorders

This critical review examines the current evidence regarding the accuracy of assessing intelligibility of dysarthric speech over the internet. Access barriers to obtaining speech and language services such as remote location and physical disability may be reduced with an online service delivery model. Overall, current research suggests that assessment results achieved face-to-face are comparable to results achieved online, however, further examination of the influence of environmental controls, various dysarthric profiles, and clinician and client technical skills are required.

Introduction

With the aging population, we expect an increase in the incidence of Parkinson's Disease, stroke, multiple sclerosis, and other neurological disorders. These disorders are often paired with communication disorders and immobility. Dysarthria, a neurologically-based motor speech disorder characterized by imprecision, weakness, slowness, and/or incoordination of movement of the speech musculature, accounts for up to 54% of these communication disorders (Darley, Aronson, & Brown, 1975; Duffy, 2005). Dysarthria can decrease speech intelligibility, which refers to how understandable one's speech is to a listener (Duffy, 2005). Intelligibility can impact an individual's ability to communicate and participate in social interactions (Hill et al., 2006; Oxtoby, 1982). In addition, individuals with reduced intelligibility may experience isolation from their community and degradation of quality of life (Oxtoby, 1982; Ziegler & Zierdt, 2008).

Because reduced intelligibility can have such a significant impact, it is important to ensure that its assessment is reliable and valid. There are several ways to assess intelligibility in a clinical setting. These include perceptual rating scales, multiple choice tasks and transcription procedures to quantify the percentage of speech understood by a listener. Some of these procedures have been included in formal intelligibility tests such as the *Assessment of Intelligibility of Dysarthric Speech* (AIDS), which was developed by Yorkston and Beukelman (1984), and the *Munich Intelligibility Profile* (MVP), which was developed by Ziegler, Hartmann, and Wiesner (1992).

Physical difficulties, financial limitations, institutional prioritization of cases, and living in a remote location from service providers can be barriers to the access of speech and language services. (Hill et al., 2006). Health care providers worldwide, face major difficulties in the recruitment and retention of speech-language pathologists (SLP) in rural areas (Pickering et al., 1998).

Alternative modes of service delivery are needed to address the increasing discrepancy between supply and demand of speech and language services (Constantinescu et al., 2010). The recent technological advances and high public availability of computers may make (telehealth and) online assessments of neurological disorders and communication impairments a potential solution to the problem of reduced access to speech and language services (Hill, Theodoros, Russell, & Cahill, 2006). Telehealth is defined as the use of information and communication technologies, such as the internet, for the delivery of healthcare at a distance (Constantinescu et al., 2010). Telehealth via the internet may be appropriate for speech and language service delivery as the focus is on auditory and visual communication rather than physical contact or manipulation (Cherney & van Vuuren, 2012). Current methods of intelligibility assessment only require auditory information, which may be easily obtained and transferred over the internet. Therefore, both the initial assessment and ongoing monitoring of speech intelligibility may be accurately administered online.

Objectives

The primary objective of this paper is to critically evaluate the current literature on assessing speech intelligibility of individuals with dysarthria in an online environment. The secondary objective is to propose evidence-based practice recommendations regarding the acceptability and accuracy of the online assessment of dysarthric speech intelligibility.

Methods

Search Strategy
Computerized databases including PubMed, Medline, Cochrane Library, and PsycINFO using the following search strategy: (dysarthria) AND (intelligibility) AND ((online) OR (telehealth) OR (telerehabilitation) OR (telespeech)).

Selection Criteria
Studies selected for inclusion in this critical review paper were required to investigate the impact of assessing intelligibility of dysarthric speech online on accuracy of results. No limits were set on the demographics of the dysarthric research participants or outcome measures.

Data Collection
Results of the search strategy yielded the following four articles congruent with the selection criteria: counterbalanced repeated measures design (1), mixed randomized clinical trial (1), and between groups randomized clinical trials (2).

Results

The accuracy of assessing speech intelligibility of dysarthric speakers online was examined in the literature using a variety of methods. All studies reviewed are experimental and provide level I or level II evidence. These types of studies typically yield valid and reliable evidence. Evidence is critically assessed based on study formulation, and validity of methods and statistical manipulations.

Hill, Theodoros, Russell, and Cahill (2006) looked at the feasibility and effectiveness of an internet-based telerehabilitation application for the assessment of motor speech disorders in adults with acquired neurological impairment. They employed a counterbalanced, repeated measures design to evaluate the accuracy of assessing speech intelligibility of 19 mild to moderate dysarthric speakers online. Each participant's speech intelligibility was assessed twice using the sentence level AIDS; once conducted by an SLP in an online environment and once conducted by a different SLP in a face-to-face (FTF) clinical environment. Environment order was randomized and separated by two to three days to minimize test-retest effects and fatigue. Readily available computers and 128 kb/s bandwidths were employed, however participant computer proficiency was not required as a non-assessing SLP set up and controlled the equipment. Two other SLPs, who were blinded to the environment and severity ratings, scored speaker intelligibility from audio recordings according to the recommended AIDS procedures (Yorkston & Beukelman, 1984).

The Bland and Altman (BA) limits of agreement technique (1986) was used to measure the magnitude of difference permissible between FTF and online assessment environments such that it would not affect the clinical assessment. Clinical criterion was set to a 8.6% change in sentence intelligibility levels as this level of variability was observed in people with dysarthria who were assessed FTF on the same day. Limits of agreement for the percentage of sentence intelligibility were marginally outside the clinical criterion at ±8.84%. The authors did not consider the 0.24% difference clinically significantly due to considerable day-to-day variation in intelligibility of dysarthric speakers. The percentage level of agreement between the two assessment environments was considered high when at least 80% of comparisons between methods were within ±8.6%. Spearman's ρ correlations revealed moderate interjudge reliability for the FTF assessment ($\rho s = .74$, $p < .01$) and high interjudge reliability for the online environment ($\rho s = .90$, $p < .05$). Pearson's r coefficient revealed high intrajudge reliability for each of the judges in both environments (FTF r= .99, .99, $p < .01$; online r= .97, .99, $p < .01$).

The researchers concluded that an online measure of sentence intelligibility is reliable and comparable to an FTF assessment. They noted limitations such as participant variability and technical issues. Furthermore, conclusions may not be appropriate for speakers with severe dysarthria and those who are not proficient with computer equipment. This study presents level II statistical evidence. Overall, when considering the high day-to-day variation of dysarthric speech and various limitations, this study provides only suggestive evidence that the online assessment of dysarthric speech intelligibility is accurate.

The accuracy of using an online application of the *Munich Intelligibility Profile (MVP-Online)* for assessing the intelligibility of dysarthric speakers was examined by Ziegler and Zierdt (2008) using a mixed randomized clinical trial.

Using readily available computers, microphones, and the online *SpeechRecorder* program (Draxler & Jänsch, 2004), 200 recordings from the *MVP-Online* assessment were obtained at a 16-bit/22,050 Hz resolution from 48 normal speakers and 110 mild to severe dysarthric speakers. The number of recordings is higher than the number of participants because

several dysarthric speakers used the system for follow-up examinations. Instructions and single word stimuli, half of which were embedded in non-predictable carrier phrases, were repeated or presented as written prompts to be read aloud. As the test was administered in a variety of clinical settings, it was unlikely participants controlled the computer.

Thirty blind listeners evaluated between 5 and 90 audio recordings each (average of 53) via 72-multiple choice tasks from which an overall intelligibility score and profile was computed. Normal speakers scored between 94.4 and 100% intelligible, with a mean of 98% ±1.5 and were normally distributed. Based on these findings, a cut off score of 95% was chosen for normal intelligibility. Dysarthric speakers scored between 20 and 100% with a median of 86% and these scores demonstrated considerable skewness (−1.3). Fourteen percent of the scores fell into the normal range and 10% of the scores were less than 50% intelligible. The most severely impaired speakers were far above the 8.70% chance for correct responses suggesting that their speech still contained information sufficient for identification of a few target words.

Between-listener agreement of two to three listeners deviated by no more than ±10% suggesting that the *MVP-Online* has good test reliability. Agreement was increased as intelligibility and number of listeners increased (±2% deviation >95% intelligibility; <±7% deviation between three listeners). These values must be interpreted with caution, as each deviation value was related to a different combination of listeners. Furthermore, circularity in the data is present as the listeners whose deviations were measured were simultaneously part of the reference samples.

Researchers concluded that the *MVP-Online* is an efficient, reliable, and valid method for assessing the speech intelligibility of a broad range of dysarthria severities. They note that MVP in its original face-to-face format is complicated to administer encompassing various constraints and efficiency issues the online version appears to eliminate.

Based on the research design, the authors' conclusions may be inappropriate. While the median intelligibility scores obtained from the *MVP-online* were significantly different for dysarthric speakers than normal speakers, a huge range (20 to 100%) with considerable skewness was observed in the intelligibility ratings of the dysarthric participants. Researchers failed to compare and analyze each participant's dysarthria severity with their intelligibility rating to determine accuracy of

intelligibility scores across this broad range. Because intelligibility ratings obtained online were not compared to ratings obtained in an FTF environment, it is difficult to conclude that the *MVP-online* yields accurate measures of speech intelligibility. It would be more appropriate to conclude that the *MVP-online* may accurately decipher abnormal from normal intelligibility. This study presents level I statistical evidence. Overall, this study provides equivocal evidence for the accuracy of assessing dysarthric speech intelligibility online.

Using a between groups randomized clinical trial, Constantinescu, Theodoros, Russell, Ward, Wilson, and Wootton (2009) examined the validity and reliability of telerehabilitation for assessing speech intelligibility of 61 participants at various stages of Parkinson's disease and mild to severe hypokinetic dysarthria. Participant's speech intelligibility was simultaneously assessed in an online and FTF environment by two SLPs. The primary mode of assessment in which all instructions were given was determined randomly. The SLPs were randomized to assessment environments and blind to severity/stage of disease. Participants were not required to have computer proficiency as all aspects of online delivery were performed by the SLPs.

The single word and sentence level sections of the AIDS were administered and audio recorded to assess speech intelligibility in both environments. Stimuli were presented on the computer screen for the online-led environment and on the test booklet for the FTF-led environment while audio recordings were made in both environments. Two independent SLPs, who were blind to dysarthria severity rating and environment, transcribed speech samples from each environment and ratings were averaged.

Within the BA limits of agreement at the 95% confidence interval, clinical criterion was set according to test-retest variability reported in the AIDS manual in a FTF environment. The limits of agreement for sentence intelligibility measures were within clinical criterion (±8.6%) and word intelligibility measures fell outside of clinical criterion (±3.2%). Less agreement on word intelligibility measures was observed on recordings of moderate and severely reduced intelligibility.

Assessing FTF and online SLPs rated overall intelligibility from a 30 second monologue sample on a five-point scale. Percent close agreement (PCA) and quadratic rated Kappa (k_w) (Landis & Koch, 1977) statistics suggested that overall speech intelligibility in the monologue sample was within

clinical criteria for both PCA (98.36) and k_w (0.79 good agreement).

Intrarater and interrater reliability was found to be comparable between environments at moderate to very good for both measures (ICC=~0.4-0.9).

It was concluded that there are comparable levels of agreement between the online and FTF environments and that the online assessment of hypokinetic dysarthria in Parkinson's disease appears to be reliable and valid. In particular, ratings of overall speech intelligibility may be made reliably online. A noted online challenge was an intermittently degraded audio signal, such as static, which was especially detrimental for transcribing those speakers with severe dysarthria. Independent control of equipment was not observed and may further contribute to technical difficulties. This study presents level I statistical evidence. Due to the limitations, this study only provides suggestive evidence that the online assessment of dysarthric speech intelligibility is accurate.

Hill, Theodoros, Russell, and Ward (2009) aimed to refine the Hill et al. (2006) study by reducing participant variability through a between groups randomized clinical trial research design. Similarly to Constentinescu et al. (2009)'s design, 24 speakers with mild to moderate dysarthria were simultaneously assessed for intelligibility in an online and FTF environment by two SLPs. The primary mode of assessment in which all instructions were given was determined randomly. The SLPs were randomized to assessment environments and blind to the participant's severity rating of dysarthria. Custom-built computer software and equipment, such as web cameras mounted on robotic arms were used on a low, readily available bandwidth (128 kb/s) for online assessment. The FTF SLP was responsible for assisting with headset microphone and orienting them to the online-SLP limiting the need for participant computer proficiency.

The sentence level section of the AIDS was administered and audio recorded in both environments to quantify speech intelligibility. A percentage level of agreement and paired-sample t-test analysis of scores obtained on the ASSIDS in the online and FTF environment revealed high agreement (95.83% at ±8.6%) and no significance difference between scores (t=1.38, p=0.17). Intraclass correlations (ICC) obtained indicated that intrarater reliability was good to very good in the FTF environment (Rater 1 ICC=0.87; Rater 2 ICC= 0.78) and high in the online environment (Rater 1

ICC=0.94; Rater 2 ICC=0.83). Interrater reliability was high for both the FTF environment (ICC = 0.94) and online environment (ICC = 0.87).

Overall perception of intelligibility in conversation in each environment was informally rated with visual and audio information by each SLP on a five-point scale. Intrarater reliability was moderate to high in the FTF environment (Rater 1 ICC=0.57; Rater 2 ICC= 1.0) and high in the online environment (Rater 1 ICC=1.0; Rater 2 ICC=1.0). Interrater reliability was moderate in the FTF environment (ICC = 0.53) and high in the online environment (ICC = 0.85). This suggests perceptual ratings of intelligibility may be more reliable when made in an online environment than in an FTF environment.

Based on the study's results, especially the high levels of agreement between environments and the strong intrarater and interrater reliability, it was concluded that valid and reliable online assessment of speech intelligibility in dysarthria is possible. However, the conclusion must be interpreted with caution because of the various limitations. As noted by the authors, technical difficulties arose and the population did not include participants with severe dysarthria. Furthermore, the influence of participant computer proficiency was not considered and may be a significant barrier to partaking in online assessments. This study presents level I statistical evidence. Due to the limitations, this study only provides suggestive evidence for the accuracy of assessing dysarthric speech intelligibility online.

Discussion

All of the reviewed studies conclude that intelligibility of dysarthric speech can accurately be assessed online. While they provide suggestive level I and level II evidence, there are several methodological limitations of these studies that must be considered.

Studies by Hill et al. (2006), Constantinescu et al. (2009), and Hill et al. (2009) were conducted by many of the same authors at the University of Queensland. This may limit variations among research design, methodology, and participants employed. For example, all of these studies compared AIDS intelligibility ratings obtained in an FTF environment and an online environment.

The study by Ziegler and Zierdt (2008), which was conducted in Bogenhausen City Clinic München, Germany, is unique in that it did not compare intelligibility ratings across environments and used a

different assessment tool, the *MVP-online*. The failure of this research design to compare online intelligibility ratings to severity or intelligibility ratings made in an FTF environment makes it difficult to determine if online scores are accurate and representative. The research designs employed by studies conducted at the University of Queensland appear more appropriate for determining the accuracy of assessing dysarthric speech intelligibility online.

Results obtained by Constantinescu et al. (2009) and Hill et al. (2009) offer the most suggestive evidence that intelligibility of dysarthric speech can accurately be assessed online. Both studies were conducted in a between groups randomized clinical trial research design. By assessing speech intelligibility simultaneously in both environments they limited the influence of day-to-day variability as observed in the counterbalanced, repeated measures design employed by Hill et al. (2006). High levels of agreement were observed between environments on the AIDS sentence intelligibility measures.

Unlike other reviewed studies, Constantinescu et al. (2009) and Hill et al. (2009) obtained perceptual ratings of overall speech intelligibility in conversation. They found these ratings to be a reliable online measure of dysarthric speech intelligibility. These findings provide valuable information to guide future research in determining which assessment tools to administer online. Based on these results, perceptual evaluations may be used alone or in conjunction with the AIDS to obtain accurate ratings of dysarthric speech intelligibility online.

All studies reviewed share common limitations that have the potential for influencing results and conclusions. The online assessment environment was highly controlled, professionally guided, and free of distractions. It is not feasible to replicate this environment in a client's home. In majority of studies, both the online SLP and individual controlling the computer were trained on specialized software and equipment. The feasibility and necessity of training SLPs must be considered. As previously noted, participants were not responsible for setting up or controlling specialized computer equipment. Realistic expectations for clients to control computer equipment must be made. As a large proportion of individuals with dysarthria are elderly and did not grow up with technology, they typically display low computer proficiency skills. Distractions combined with technical difficulties as a result of poor computer proficiency may contribute to reduced intelligibility ratings. Therefore, no conclusions can

be drawn about the success of the online assessment of dysarthric speech intelligibility if clients are required to independently run computer software.

The reviewed studies had fairly small numbers of participants and large exclusion criterion, failing to examine participants with co-occurring communication disorders. In addition, a range of dysarthria severities were not equally represented, particularly severe dysarthria. These factors limit the generalizability of findings to all people with dysarthria. Conclusions may only be appropriate for individuals with mild to moderate 'pure' dysarthria. It may be beneficial for researchers to identify specific criteria, which may or may not make an individual a suitable candidate for being assessed online for dysarthric speech intelligibility.

Conclusion

Many of the reviewed studies offer only suggestive evidence that the intelligibility of dysarthric speech can accurately be assessed online. The possible impact of the studies' methodological limitations on results requires further research for sound conclusions to be made.

Clinical Implications

Due to the methodological limitations of the studies, the clinical application of findings should be made with caution until further research is completed. Results may have important clinical implications such as allowing for easier access to SLP services for those who are immobile and/or living in rural areas. As the computer-literate population ages, clinicians may be expected to adapt to expectations of using technology in every part of their lives. In turn, the convenience of the online assessment method may contribute to both client and clinician satisfaction.

References

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 8, 307-310.

Cherney, L., & van Vuuren, S. (2012). Telerehabilitation, virtual therapists, and acquired neurologic speech and language disorders. *Seminars in Speech and Language, 33*(3), 243-257.

Constantinescu, G., Theodoros, D., Russell, T., Ward, E., Wilson, S., & Wootton, R. (2010). Assessing disordered speech and voice in Parkinson's disease: a telerehabilitation application. *International Journal of Language & Communication Disorders, 45*(6), 630-644.

Darley, F.L., Aronson, A.E., & Brown, J.R. Motor Speech

Disorders. Philadelphia, PA: WB Saunders; 1975.

Draxler, C., & Jänsch, K. (2004). SpeechRecorder – A universal platform independent multi-channel audio recording software. In *Proceedings of the IVth International Conference of Language Resources and Evaluation*.

Duffy (2005). Motor speech disorder: Substrates, differential diagnosis and management (2nd ed.). St. Louis, MO: Elsevier Mosby.

Hill, A., Theodoros, D., Russell, T., Cahill, L., & al, e. (2006). An internet-based telerehabilitation system for the assessment of motor speech disorders: a pilot study. *American Journal of Speech – Language Pathology, 15*(1), 45-56.

Hill, A., Theodoros, D., Russell, T., & Ward, E. (2009). The redesign and re-evaluation of an internet-based telerehabilitation system for the assessment of dysarthria in adults. *Journal of Telemedicine and e-Health, 15(*9), 840-850.

Oxtoby, M., 1982, *Parkinson's Disease Patients and Their Social Needs* (London: Parkinson's Disease Society).

Pickering, M., McAllister, L., Hagler, P., Whitehall, T.L., Penn, C., Robertson, S.J., et al. (1998). External factors influencing the profession in six societies. *American Journal of Speech-Language Pathology*, 7, 5-17.

Yorkston, K.M., & Beukelman, D.R. (1981). Assessment of Intelligibility of Dysarthric Speech. Austin, TX: Pro-Ed.

Ziegler, W., Hartmann, E., & Wiesner, I. (1992). Dysarthriediagnostik mit dem "Münchner Verständlichkeits-Profil" (MVP) – Konstruktion des Verfahrens und Anwendungen. *Nervenarzt*, 63, 602-608.

Ziegler, W., & Zierdt (2008). Telediagnostic assessment of intelligibility in dysarthria: a pilot investigation of MVP-online. *Journal of Communication Disorders, 41*, 553-557.