Rating the intelligibility of dysarthic speech amongst people with Parkinson's Disease: a comparison of trained and untrained listeners

Authors

Christina H Smith Division of Psychology and Language Science, Faculty of Brain Sciences, University College London

Smitaa Patel Birmingham Clinical Trials Unit, University of Birmingham

Rebecca L Woolley, Birmingham Clinical Trials Unit, University of Birmingham

Marian C Brady Nursing, Midwifery and Allied Health Professions Research Unit, Glasgow Caledonian University

Caroline E Rick Nottingham Clinical Trials Unit, University of Nottingham

Rhiannon Halfpenny, Division of Psychology and Language Science, Faculty of Brain Sciences, UCL

Alexia Rontiris, Division of Psychology and Language Science, Faculty of Brain Sciences, UCL

Lucy Knox-Smith, Division of Psychology and Language Science, Faculty of Brain Sciences, UCL

Francis Dowling Birmingham Clinical Trials Unit, University of Birmingham

Carl E Clarke Institute for Applied Health Research, University of Birmingham & Department of Neurology, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham

Pui Au Birmingham Clinical Trials Unit, University of Birmingham

Natalie Ives Birmingham Clinical Trials Unit, University of Birmingham

Keith Wheatley Birmingham Clinical Trials Unit, University of Birmingham

Catherine M Sackley, Life Sciences and Medicine, King's College London

Corresponding author:

Christina H Smith Division of Psychology and Language Science, Faculty of Brain Sciences, University College London. Telephone 02076794200. Email Christina.smith@ucl.ac.uk

Rating the intelligibility of dysarthic speech amongst people with Parkinson's Disease: a comparison of trained (Speech and Language Therapy students) and untrained listeners

Abstract

Intelligibility of speech is a key outcome in speech and language therapy (SLT) and research. SLT students frequently participate as raters of intelligibility but we lack information about whether they rate intelligibility in the same way as the general public. This paper aims to determine if there is a difference in the intelligibility ratings made by SLT students (trained in speech related topics) compared to individuals from the general public (untrained). The SLT students were in year 2 of a BSc programme or the first 6 months of a MSc programme.

We recorded 10 speakers with Parkinson's disease (PD) related speech reading aloud the words and sentences from the Assessment of Intelligibility of Dysarthric Speech. These speech recordings were rated for intelligibility by 'trained' raters and 'untrained' raters. The effort required to understand the speech was also reported. There were no significant differences in the measures of intelligibility from the trained and untrained raters for words or sentences after adjusting for speaker by including them as a covariate in the model.

There was a slight increase in effort reported by the untrained raters for the sentences. This difference in reported effort was not evident with the words. SLT students can be recruited alongside individuals from the general public as naïve raters for evaluating intelligibility in people with speech disorders.

Key words: Intelligibility, Rating, Parkinson's disease, Speech, Dysarthria

Introduction

Speech intelligibility relates to how easy it is for a listener to understand what has been said (Weismer, Jeng, Laures, Kent & Kent, 2001). This perception is likely to be influenced not just by the acoustic qualities of the speech but will also likely be influenced by environmental or contextual aspects. For example, speakers of non-disordered speech are generally more intelligible when the listener is given contextual information as this allows top-down processing (Hustad & Beukelman, 2001) making sentences easier to understand than single words (Grant & Seitz, 2000). In contrast, speech of people with dysarthria tends to be more difficult to understand in sentences (Dongilli, 1994; Hustad, 2007) possibly because the listener is unable to extract sufficient acoustic information from the speech in order to use the context to make assumptions (Hustad, 2007). Intelligibility is an important metric in assessing dysarthria and in evaluating the effectiveness of intervention. Within the World Health Organisation's International Classification of Functioning, Disability and Health (World Health Organisation, 2001) intelligibility is classified as an element of 'activity'. Reduced intelligibility affects an individuals' participation in society and the quality of conversations they are able to have with others. A reliable, reproducible measure of intelligibility is a valuable tool in the assessment and management of people with speech impairments and can be utilised in goal setting and as an outcome measure.

Intelligibility is complex as it often increases with familiarity with that individuals' speech production. This occurrence can be observed in clinical environments for example when family members are able to understand a person's speech which is unintelligible to a healthcare worker. The effect of this familiarity is also evident in formal assessments of intelligibility with family members and those familiar with the speech of individuals with dysarthria scoring the speech of the PwD as more intelligible than those unfamiliar with the individual (Baudonck, Buekers, Gillebert, & Van Lierde, 2009; DePaul & Kent, 2000; Tjaden & Liss, 1995). Intelligibility of dysarthric speech may be impacted not just by familiarity but by the listener, for example by the amount of effort the listener makes to understand the speech output. Intelligibility may increase with increased effort on the part of the listener. Understanding how different

listeners understand disordered speech is important in clinical and research settings where reliable measurement of change in speech production over time is essential.

Little is known about the effects that SLT professional education may have on the ability to understand disordered speech. This is relevant as SLT students are frequently recruited as participants for perception experiments related to disordered speech. Dagenais and colleagues (Dagenais, Watts, Turnage, & Kennedy, 1999; Dagenais, Garcia, & Watts, 1998) report that experienced SLT's rate the intelligibility of speech of PwD higher when compared to the ratings by naïve listeners of different ages (Dagenais et al, 1998). The young naïve listeners in this study were undergraduate students in an introductory SLT class. In both studies by Dagenais the speech of two PwD and two people without dysarthria was evaluated. In contrast, Walshe (Walshe, Miller, Leahy, & Murray, 2008) found no difference in experienced SLT's perception of intelligibility when compared to naïve listeners. Walshe used a perceptual rating scale, rather than the quantitative measure of intelligibility – Assessment of Intelligibility of Dysarthric Speech (AIDS) – used by Dagenais. No studies have examined the effect of SLT education in the rating of intelligibility of disordered speech.

The AIDS is a validated, standardised assessment, created to measure the intelligibility of PwD (Yorkston & Beukelman, 1981). The AIDS assesses speech production in frequently occurring single words from dense lexical neighbourhoods (e.g. mate, late and lake), and in sentences which increase in length from 5 to 15 words which are taken from the literature written for adults. The SLT generally carries out the assessment and a person unfamiliar with the PwD listens to an audio recording of the PwD speech. This person may not be a qualified SLT and it may be a different person who scores the person's speech at different time points. It is therefore important that there is reliability with different listeners with different backgrounds.

This study aims to determine if there is a quantitative difference in the measurement of intelligibility assessed by trained (SLT students) and untrained raters, focussing on people with dysarthria due to Parkinson's disease. Specifically, 1) are there differences in single word intelligibility scores (from AIDS)

between trained and untrained raters; 2) are there differences in sentence intelligibility scores (from AIDS) between trained and untrained raters; and 3) are there differences in the effort required to understand the speech of PwD between trained and untrained raters.

Method

This study is part of a pilot randomised control trial of speech therapy intervention for people with Parkinson's disease (PD). The published study protocol (Sackley et al., 2014) and the outcomes paper (Sackley et al., 2018) provide detailed information about the recruitment of the people with PD, the characteristics of the people with PD, and the recording of their speech. The local higher educational institution research ethics committee approved the study protocol and consent procedure

Participants

Fifty four adults (13 male and 41 female) between the ages of 19 and 62 (mean (SD) = 25.2 (7.9)) were recruited as 'raters' of the speech samples. All were monolingual English language speakers with no self-reported history of hearing impairment. Raters were allowed to participate more than once – six of the raters participated in the task twice. The 'trained raters' (n= 26) had received formal instruction in the perception and transcription of disordered speech sounds, the acoustics of speech, hearing, and phonetics, phonetics of disordered speech, and suprasegmental features of speech. These SLT students were either in year 2 of a 4 year BSc programme or in the first 6 months of a 2 year MSc programme (these are equivalent time points for the training received). These students had no experience with people with dysarthria due to PD, and limited experience with people with speech disorders. These trained raters were recruited via advertisement within the Speech and Language Therapy department. The 'untrained raters' (n=28) had no formal training in disordered speech, phonetics or acoustics, and no reported experience communicating with people with speech and language difficulties. These untrained raters were recruited via advertisement in libraries and shops in the local area and in the university, and through word of mouth.

Materials and Equipment

Ten people with disordered speech as a consequence of (PD) read individual words and sentences from the AIDS with a head mounted microphone (Monacor HSE-821SX microphone) 8 inches from the participant's lips. Each speech sample was electronically recorded Roland R-05 digital recorder & and stored in a digital file. The stimuli for the samples consisted of 50 written words and 22 written sentences which were randomly generated (following the guidelines within the AIDS manual) from AIDS which was delivered following the manual instructions. Of the 10 people with dysarthria there were nine men and one woman. The mean age was 69 (SD 11.4) years. The PD severity of the speakers was mild as assessed by the individuals medical consultant (mean Hoehn & Yahr stage 2.1 (0.8) and duration of PD was 6.5 (3.5) years. All speakers were 'on' medication for their PD at the time of the speech recording.

The electronic audio files were edited using Audacity (Audacity®, version 3.0). In order to prime the raters as to the task, two practice words were recorded by the experimenter and played prior to the 50 stimuli words. Similarly, two practice sentences were recorded by the experimenter and played prior to the 22 stimuli sentences. Each stimuli word was presented to the raters with a five second interstimulus interval to allow participants time to write the word. Each word was presented once. The sentences were presented to the raters twice with a five second interstimulus interval between the repetitions to ensure that the listener was able to recall the precise sentence. There was a 10 second interstimulus interval between each separate sentence.

Raters heard the audio files played back using Audacity through individual headphones (Touchmate TM-201) with playback at 50% connected to a computer (Dell Optiplex-GX620). The Express Scribe programme allowed keys on a standard keyboard to be programed to enable the participants to pause the replay of the repeated word or sentence whilst listening to write their response.

Procedure

Three trained (SLT students) and three untrained raters listened to each speech sample. Raters carried out the listening task through headphones in individual soundproofed rooms. Each participant listened to the

speech recordings (two sets of words and two sets of sentences) from two individuals with disordered speech The listening task lasted no more than 60 minutes. The instructions given to the raters for completion of the task were from the AIDS manual, including the provision of the AIDS word lists to review prior to commencing the listening task. The instructions are in Appendix 1.

Playback

The audio files were played to the participants who wrote the words and sentences they thought they had heard on the data collection sheet provided. The researcher read through the list of words written by the participant and asked for clarification of the word written if the handwriting was not clear. The orthographic transcription for the words and the sentences provided a percentage intelligibility score following the instructions from the AIDS manual. Whole word intelligibility was the metric utilised, that is, no credit was given for partially correct words. No analysis of phonemic errors was included.

The raters also reported the effort which they required to understand the speech. The 'effort' was reported twice – once after listening to all the words, and again after listening to all the sentences. The 'effort' rating used a visual analogue scale of 1 to 5 where one is 'normal' amount of listener effort and five is 'a lot' of effort – a cross was placed on the line indicating the amount of effort required.

Statistical Analysis

All statistical analyses were performed using Stata/SE 13.0 (Copyright 1985-2013 StataCorp LP). The intra-class correlations (ICC) for word and sentence intelligibility, both overall and by trained and untrained raters were produced using the ICC command. The mean differences between the trained and untrained raters in word and sentence intelligibility were calculated using a linear regression model, adjusting for speaker by including them as a covariate in the model. Each word was scored either correct or incorrect in both the single word and the sentence speech output.

Results

The mean scores with the SD for both words and sentences are detailed in table 1. There was no significant difference in the word intelligibility scores from trained and untrained raters (mean difference 0.3; 95% CI: -2.8 to 2.2; p= 0.8, table 1). The ICC was obtained overall and by trained and untrained groups. Overall there was a high level of agreement for word intelligibility scores (ICC=0.74; p<0.001) between the raters. When calculated by trained or untrained groups, they did not differ significantly, with the trained group having a slightly lower ICC (ICC=0.69; p<0.001) compared with the untrained group (ICC=0.75; p<0.001). Similarly, trained and untrained raters were not significantly different in their scores of sentence intelligibility (mean difference -0.4; 95% CI: -1.8 to 1.0; p=0.6, table 1). The overall ICC was slightly lower for the sentence intelligibility than word intelligibility (ICC=0.70; p<0.001). The untrained group had a lower ICC (ICC=0.59; p<0.001) than the trained group (ICC=0.73; p<0.001) though all ICC scores indicated a high level of agreement amongst the raters.

Insert table 1 about here.

Both the trained and untrained raters reported similar levels of effort to interpret the speakers' words samples (mean difference -0.2; 95% CI: -0.5 to 0.1; p = 0.1, table 2). Untrained raters recorded slightly more effort than trained raters in listening to the sentence samples (mean difference for sentences -0.4; 95% CI: -0.7 to -0.04; p=0.03, table 2).

Insert table 2 about here.

Discussion

Intelligibility is a critical component in assessing severity of an individual's speech disorder and a key component of functional communication. In turn it is an important measure of therapy effectiveness (Pennington & Miller, 2007), . We found no evidence of a difference in the trained (SLT students) and untrained raters (individuals from the general public) understanding of the disordered speech of people with PD.

Previous work has compared naïve listeners with experienced SLTs. In our study, the naïve listeners (untrained) were compared with listeners trained in SLT knowledge and skills but no clinical experience with people with PD. Despite methodological differences (percentage measure of intelligibility rather than perceptual rating scale), these findings are similar to work by Walshe (Walshe et al., 2008) who found no significant differences between trained and experienced SLTs and naïve listeners. Dagenais (Dagenais et al., 1999; Dagenais et al., 1998) using very similar methodology to the current study (i.e. AIDS), report experienced SLT's gave higher scores of intelligibility compared to the naïve listeners. Consistent with previous research (Grant & Seitz, 2000), all raters found sentences easier to understand than words, indicating knowledge of English syntax and pragmatics (i.e. top-down processing) when listening to the sentences. This top-down processing allows some words to be guessed using the context to guide the interpretation.

Our findings have important clinical and research methodological implications. SLT students evaluate intelligibility in clinical settings in keeping with the ratings of a naïve listener. We remain unclear about the similarities between SLT students and experienced clinicians ratings of intelligibility amongst people with dysarthria due to PD. From a research methodological perspective SLT students demonstrate no evidence of a difference in their ratings of intelligibility compared to naive 'untrained' raters thus suggesting the contribution they may make to such functionally relevant ratings. It also increases the potential pool of untrained raters available to contribute to future dysarthria research activities.

In this study three trained raters (SLT students), and three untrained raters (individuals from the general public) evaluated the speech of two people with disordered speech within a one hour session. A further study could have the raters evaluate the samples from all 10 speakers twice, allowing a period in between listening to the speakers so that the raters were unable to recall what they scored the first time. This would have allowed us to assess that errors introduced by the raters (some may have been more consistent with their scoring than others) and so inform on the wider population of raters. This study could be extended to separate students into year groups to determine if there is a stage in their training when they behave differently from the untrained raters.

The speakers in this study had high intelligibility scores. It is possible that differences in rater intelligibility scores between the two groups might be present with more severe speech disorders. Adding background noise to the audio data would increase the listening demands placed on the participants and be more reflective of natural communication settings. This could provide more naturalistic data for comparing listener reported intelligibility with a complex audio signal.

Conclusion

SLT students who have had formal instruction in the perception and transcription of disordered speech sounds, the acoustics of speech and hearing, and phonetics, and phonetics of disordered speech but limited clinical experience rate intelligibility in a similar way to those with no training. This means that in research settings, these SLT students can be recruited to participate in perception studies of disordered speech of people with PD alongside the general public. In clinical settings, these SLT students will evaluate disordered speech of people with PD as would the general public providing speech and language therapists an indication of how their clients are perceived outside of the clinic room.

Acknowledgements

All the people with dysarthria and all the listeners who participated in this study. Andrew Faulkner and Steve Nevard for technical support, Debbie Kelly for facilitating the trial, and Susan Jowett, Ramilla Patel and Helen Roberts for their help in the study. The anonymous reviewers who provided helpful input to enhance this manuscript. Dunhill Medical Trust for the funding for this project. MB and the NMAHP Research Unit are funded by the Chief Scientist Office, part of the Scottish Government Health and Social Care Directorate. The views expressed here are those of the authors and not necessarily those of the funders.

Baudonck, N. L. H., Buekers, R., Gillebert, S., & Van Lierde, K. M. (2009). Speech Intelligibility of Flemish Children as Judged by Their Parents. *Folia Phoniatrica et Logopaedica*, *61*(5), 288–295. doi:10.1159/000235994

Dagenais, P. A., Watts, C. R. T., Turnage, L. M., & Kennedy, S. (1999). Intelligibility and acceptability of moderately dysarthric speech by three types of listeners. *Journal of Medical Speech-Language Pathology*, *7*, 91–97.

Dagenais, P., Garcia, J., & Watts, C. (1998). Acceptability and intelligibility of mildy dysarthric speech by different listeners. In M.P. Cannito, K.M. Yorkston, D.R. Beukelman & P. H. Brookes (Eds.),*Neuromotor speech disorders: Nature, assessment and management* (pp. 229–239). Baltimore: Paul H Brookes.

DePaul, R., & Kent, R. (2000). Effects of Listener Familiarity and Proficiency on Intelligibility Judgments. *American Journal of Speech-Language Pathology*, *9*, 230–240. doi:10.1044/1058-0360.0903.230

Dongilli, P. (1994). Semantic context and speech intelligibility. In J.A. Till, K.M. Yorkston & D.R.Beukelman (Eds.), *Motor Speech Disorders: Advance in Assessment and Treatment* (pp. 175–191). Baltimore: Brookes.

Grant, K. W., & Seitz, P. F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *The Journal of the Acoustical Society of America*, *107*(2), 1000–1011. doi:10.1121/1.428280

Hustad, K. C. (2007). Contribution of Two Sources of Listener Knowledge to Intelligibility of Speakers With Cerebral Palsy. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1228–1240. doi:10.1044/1092-4388(2007/086)

Hustad, K. C. (2007). Effects of Speech Stimuli and Dysarthria Severity on Intelligibility Scores and Listener Confidence Ratings for Speakers with Cerebral Palsy. *Folia Phoniatrica et Logopaedica*, *59*(6), 306–317. doi:10.1159/000108337

Hustad, K. C., & Beukelman, D.R. (2001). Effects of Linguistic Cues and Stimulus Cohesion on Intelligibility of Severely Dysarthric Speech. *Journal of Speech, Language, and Hearing Research*, *44*(3), 497–510. doi:10.1044/1092-4388(2001/039)

Pennington, L., & Miller, N. (2007). Influence of listening conditions and listener characteristics on intelligibility of dysarthric speech. *Clinical Linguistics & Phonetics*, *21*(5), 393–403. doi:10.1080/02699200701276675

Sackley, C.M., Smith, C.H., Rick, C., Brady, M.C., Ives, N., Patel, S., … Clarke, C.E. (2014) Lee Silverman voice treatment versus standard NHS speech and language therapy versus control in Parkinson's disease (PD COMM pilot): study protocol for a randomized controlled trial. *Trial*s,*15*. 213-- XXX. DOI: 10.1186/1745-6215-15-213.

Sackley, C.M., Smith, C.H., Rick, C., Brady, M.C., Ives, N., Patel, S., … Clarke, C. E. (2018). Lee Silverman Voice Treatment versus standard speech and language therapy versus control in Parkinson's disease: a pilot

randomised controlled trial (PD COMM pilot). *Pilot and Feasibility Studies, 4*, 30 -- XXX. DOI: 10.1186/s40814-017-0222-z

Tjaden, K. K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics, 9*(2), 139–154. doi:10.3109/02699209508985329

Walshe, M., Miller, N., Leahy, M., & Murray, A. (2008). Intelligibility of dysarthric speech: perceptions of speakers and listeners. *International Journal of Language & Communication Disorders*, *43*(6), 633–648. doi:10.1080/13682820801887117

Weismer, G., Jeng, J-Y., Laures, J. S., Kent, R.D., & Kent, J.F. (2001). Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders. *Folia Phoniatrica et Logopaedia, 53*, 1-18

World Health Organisation. (2001). *International classification of functioning, disability and health (ICF).* Geneva, Switzerland.

Yorkston, K., & Beukelman, D. (1981). Assessment of Intelligibility of Dysarthric Speech. CC Publishers.

Yorkston, K., Dowden, P., & Beukelman, D. (1992). Intelligibility measurement as a tool in the clinical management of dysarthric speakers. In R. D. Kent (Ed.), *Intelligibility in Speech Disorders* (pp. 266–285). Philadelphia: John Benjamins Publishing Company.

Appendix 1 Instructions provided to the raters.

The specific instructions for the words were:

'*Please write down the word that you hear spoken. You will hear each word only once. If you are not sure of the word, write your best guess. Please only write real words. There are 50 test items in total. You will be given two practice items before the test starts'.*

The specific instructions for the sentences were:

'Please write down the sentence that you hear spoken. You will hear each sentence twice. Listen to the first sentence all the way through. You may wish to pause the repeat sentence to give you time to write your response. If you are not sure of the sentence, write your best guess. Please only write real words. There are 22 test items in total. You will be given two practice items before the test starts'.

Table 1: Summary statistics for data from word scores

	Training	N	Mean (SD)	ICC	Adjusted Mean Diff. (95% CI)	p-value	ICC
Words	Trained	48	88.8 (13.0)	0.69	0.3 (-2.8 to 2.2)	0.8	0.74
	Untrained	48	88.5 (10.6)	0.75			
Sentences	Trained	48	95.3 (6.5)	0.7	-0.4 (-1.8 to 1.0)	0.6	0.70
	Untrained	48	95.6 (6.0)	0.6			

Table 2. Summary statistics of data for effort rating for words and sentence

	Listeners	N	Mean (SD)	Adjusted Mean Diff (95% CI)	p-value
Words	Trained	46	1.8 (0.9)	-0.2 (-0.5 to 0.1)	0.1
	Untrained	48	2.1 (0.9)		
Sentences	Trained	41	1.9 (0.9)	-0.4 (-0.7 to -0.04)	0.03
	Untrained	42	2.2 (0.9)		