Miller N, Nath U, Noble E, Burn D. Utility and accuracy of perceptual voice and speech distinctions in the diagnosis of Parkinson's disease, PSP and MSA-P. *Neurodegenerative Disease Management* 2017, 7(3).

Utility and accuracy of perceptual voice and speech distinctions in the diagnosis of Parkinson's disease, PSP and MSA-P

Nick Miller (corresponding author),

Newcastle University Institute for Ageing

Speech and Language Sciences

George VI Building, Newcastle University

Newcastle-Tyne NE1 7RU, Great Britain

Email: Nicholas.Miller@ncl.ac.uk

Telephone: +44 191 20 85 60 3

Uma Nath, Consultant Neurologist, Sunderland Royal Hospital

Kyall Road, Sunderland SR4 7TP, Great Britain

Emma Noble, Speech and Language Sciences, George VI Building, Newcastle University

Newcastle-Tyne NE1 7RU, Great Britain

David Burn, Institute of Neuroscience, Professor of Movement Disorders Neurology

Newcastle University Newcastle upon Tyne NE2 4HH

Abstract

Aims

To determine if perceptual speech measures distinguish people with Parkinson's (PD), multiple system atrophy with predominant parkinsonism (MSA-P) and progressive supranuclear palsy (PSP).

Methods

Speech-language therapists blind to patient characteristics employed clinical rating scales to evaluate speech/voice in 24 people with clinically diagnosed PD, 17 with PSP, 9 with MSA-P, matched for disease duration (mean 4.9 years, SD 2.2).

Results

No consistent intergroup differences appeared on specific speech/voice variables. People with PD were significantly less impaired on overall speech/voice severity. Analyses by severity suggested further investigation around laryngeal, resonance and fluency changes may characterise individual groups

Conclusion

MSA-P and PSP compared to PD were distinguished by severity of speech/voice deterioration, but individual speech/voice parameters failed to consistently differentiate groups.

Keywords

Parkinson's; Multiple System Atrophy; Progressive Supranuclear Palsy; speech; voice; assessment; perceptual; oral-facial apraxia; language.

Introduction

Parkinson's disease (PD), multiple system atrophy (MSA) and progressive supranuclear palsy (PSP) constitute a group of closely related disorders, in particular in respect of shared parkinsonian features. Their early differentiation remains a clinical challenge. This study concerns whether perceptual evaluation of changes in communication may support diagnosis.

PD is the commonest subcortical degenerative neurological disorder, with around 1% of people over 60 years diagnosed [1]. Mean age of diagnosis falls in the mid 60s (though young onset PD in the 40s or younger occurs). Survival after diagnosis normally extends in excess of ten years. Estimates of MSA prevalence range between 5-8 per 100.000 population, with mean age of diagnosis around 58 years and mean survival from then around 5-10 years [2, 3]. PSP shows a similar, possibly slightly higher, prevalence to MSA and similar rate of decline. Mean age at diagnosis is approximately 63-65 years [4-6].

PD and MSA are synucleinopathies. The focus of pathology in PD is on the dopamine-producing neurons of the substantia nigra, but impairment is present in other neurotransmitter systems and outside nigrostriatal neuronal groups. Effects of these pathologies are widespread through the CNS, leading to the classic motor symptoms (below), but also a range of non-motor disturbances. In MSA deposition of synuclein occurs in oligodendrocytes, in the cerebellum, pons and basal ganglia. This leads to proposed sub-types depending on whether cerebellar (MSA-C) or parkinsonian (MSA-P) symptoms predominate [2, 3]. Which site is predominantly affected may vary across geographical regions [7], with implications for the prevalence of subtypes in different continents. PSP represents a tauopathy with typical distribution of pathology in the basal ganglia, brainstem and frontal lobes. Several variants are recognised [6, 8].

PD is characterised by a constellation of motor symptoms - bradykinesia, tremor, rigidity and postural instability. These affect balance, gait, and speed and sustainability of motor function. Non-motor changes are also present [9, 10], including autonomic dysfunction, cognitive changes and sleep and mood disturbance. People with MSA may present with parkinsonian features of rigidity, postural instability and bradykinesia, but signs of cerebellar and pyramidal dysfunction also arise [2, 3], together with signs of autonomic failure [11]. The hallmarks of PSP include presence of parkinsonian features - bradykinesia and postural instability (with notable akinetic, rigid features and propensity for backwards falling). However, these appear alongside symptoms of other brainstem and frontal dysfunction – e.g. vertical gaze palsy, dysexecutive disorder – and show a rapid (relative to PD) progression [4, 12].

Clinically, early differentiation of MSA, PSP and PD from one another can be problematical, with appreciable proportions of misclassification [6, 13, 14]. Although mean age of onset and diagnosis differ, the age ranges overlap considerably. Whilst the course of decline varies across MSA, PSP and PD and distinct symptoms can arise (e.g. vertical gaze palsy in PSP; stridor in MSA; early dysphagia in MSA and PSP), actual presentation and progression across patients is diverse, with heterogeneity in the relative prominence and order of appearance of symptoms [10, 12, 15-17]. When parkinsonian symptoms predominate early on clinical separation between disorders can remain especially problematic, though attempts have been made to identify possible differentiating biochemical or structural imaging markers [18].

Around 80-90% of people with Parkinson's disease (PwPD) experience voice and speech changes, even in the early stages [1, 2]. Dysarthria ranks amongst the first changes to affect people with PSP and MSA [3-6]. Understanding the nature of speech changes in these conditions therefore represents an important endeavour, to monitor progression, evaluate

impact [7, 8], measure response to pharmacological or other therapies and identify targets for intervention.

With regards to speech and voice changes, PwPD present with hypokinetic dysarthrophonia [19], typified by early effects on voice (quieter) and an emerging dysprosody. These entail loss of loudness/intensity contrasts (giving an impression of monoloudness) and narrowing pitch range (leading to impression of monopitch) resulting in problems expressing intonational and emotional content of utterances. Speech changes usually develop later [20]. Rate of speech can be altered in some speakers – to slower, or with short rushes of faster speech. Abnormally long pauses are found. Difficulty initiating phonation or arrests of phonation in mid-sentence may happen, akin to freezing of gait. Speech and voice impairment is not purely motoric. Auditory perceptual changes play a role in the dysprosody [21, 22]. Furthermore, cognitive-linguistic changes affect comprehension and production of language, even in the absence of dementia [23].

MSA and PSP portray elements of hypokinetic dysarthria [24-27]. However, they are claimed to be distinguished from the hypokinetic dysarthria of PD by the presence of additional voice and speech changes that reflect the more widespread underlying pathology. In particular, presence of cerebellar pathology in MSA (particularly MSA-C), with its dysrhythmia and dysmetria, is believed to produce a rougher voice quality, with inconsistent loudness/ intensity and pitch control and stress patterns of speech moving towards excess and equal stress [25, 27]. By contrast greater brainstem and pyramidal involvement is believed to create features of spastic dysarthria – lowered pitch, strained voice quality, slowed rate, hypernasality, laboured articulation [24, 27]. Palilalia may be a salient feature of PSP, though findings are variable [16, 26]. These presumed contrasts in speech and voice profile form the basis of claims that speech examination may play a role in the (differential) diagnosis of PD, MSA, PSP.

Previous studies have addressed this claim but have been criticised for depending on unblinded, single assessor evaluation, employing scoring procedures that potentially biased outcomes to prejudged results [24, 25], lacking comparisons between groups [24, 25, 28] and neglecting possible confounds in relation to disease duration and severity – though see [27, 29, 30]. Other variables that may separate groups (e.g. oral-facial apraxia [31] and language changes [32]) have been neglected.

Rusz et al [33] argued that a combination of acoustic measures relating to speech rate, speech fluency, pauses, and pitch and amplitude fluctuations distinguished groups with PD from atypical Parkinsonian syndromes with 95% accuracy – albeit their analyses appear to have been at group and not individual patient level and the model was not tested out on a new set of patients. A combination of voice quality, fluency, rate and voice tremor measures separated PSP from MSA with around 75% accuracy. Others also found key roles of rate and fluency in differentiating groups [30, 33-38]. However, even in studies which claimed pathognomonic features, there is variability in precisely which measures vary between groups and their differential sensitivity and specificity can be low.

Nevertheless, there is some consensus that people with PD should demonstrate hypokinetic dysarthria [19]; people with PSP show a mixed dysarthria with predominantly hypokinetic and spastic characteristics, but less prominent ataxic features [24, 26]; people with MSA have a tendency to less marked spastic elements but more prominent ataxic changes alongside hypokinetic speech [25, 27]. Few studies examined people with MSA-C vs MSA-P separately. Huh et al [36] found only selected acoustic features differentiated early (time since diagnosis around 2 years, SD1.4) PwPD and MSA-P in male speakers, but not female. No perceptual dimensions distinguished groups.

We aimed to test whether routine clinical perceptual assessment of speech and voice, attuned to potential characteristics and led by the above pointers from acoustic analyses,

can differentiate between people with PD, PSP and MSA-P, independent of duration and severity of symptoms.

Methods

Participants

We recruited patients with either PSP, MSA-P or PD via hospital specialist movement disorder outpatient clinics. PD cases were defined according to the UK PD Brain Bank diagnostic criteria [26], probable PSP according to NINDS-SPSP criteria [39], and probable MSA-P according to the Consensus and Quinn criteria [40]. Diagnosis was verified by two neurologists checking test results, blind to the diagnostic category. Recruitment was by informed, voluntary consent with right to withdraw without consequences at any point, and followed approved stipulations from a United Kingdom National Health Service Research Ethics Committee.

Potential participants were excluded if they showed dementia with or without a major depressive disorder, had a history of stroke, head injury, or developmental stuttering, were non-native English speakers, or were people with MSA-P with a tracheostomy.

Assessment

Assessment took place in one session in a quiet, sound deadened room at the patient's home first thing in the morning. Participants withheld their anti-parkinsonian medication from midnight the night before and were thus in a pragmatically defined off state. This aimed to reduce possible variability from differing drug regimes and medication cycles and thereby gain insights more closely related to underlying motor impairments.

Participants produced a prolonged /a/ at their habitual intensity. They read (after silent rehearsal) the 'Grandfather passage' [41], at their habitual intensity and rate. They described how to make a cup of tea/coffee to provide a sample of spontaneous speech. Speech was

recorded with a Roland Corporation Edirol 24-bit digital recorder attached to an AKG C420 head mounted directional MicroMic to assure constant mouth to microphone distance and minimise head movement artefact and ambient noise. Assessment also covered the UPDRS III [42] for motor status, Addenbrooke's Cognitive Assessment (ACE-R) [43], the MADRS (Montgomery-Åsberg) depression screen [44], assessment of oral-facial apraxia [45] and sentence level language comprehension (Token Test [46]).

Three speech and language therapists blind to patient identity and aims of the study independently carried out speech-voice ratings. They had at least 4 years' experience working in specialist units where perceptual voice and speech rating were part of their daily routine. They received details of each speaker's age and gender in order to judge the appropriateness of pitch and loudness. They could listen to tracks as often as necessary to form an opinion.

Voice quality was rated using the GRBAS (overall Grade, Roughness, Breathiness, Asthenia, Strain) scales [47] based on the prolonged /a/. This set of scales is widely employed in the field of voice evaluation as a valid and reliable tool [48]. Evaluation of the appropriateness and consistency of perceived loudness, pitch, speech rate, stress placement, fluency, nasality, articulation and naturalness (how normal listeners perceived output to be) were made in relation to the first three sentences of the 'Grandfather passage'. Rating of speech fluency, nasality, articulation and naturalness were based on 7 point equal appearing interval scales (EAIs) (1 representing unaffected). Where non-normality might deviate in two directions (e.g. pitch variability towards monotone vs towards inappropriate variability) EAIs were 13 point with normality graded as 7 and deviations either side representing 'less' or 'excessive' variability. .

Fluency was also quantified by counting appropriate (i.e. occurring at normal syntactic/semantic junctures) and inappropriate (i.e. in-word or interrupted verb or noun

phrase) pauses (of >200ms) and repetitions, a) when reading the entire 'Grandfather' passage and b) from the spontaneous speech sample (making a cup of tea/coffee). Immediate single (e.g. f-frock; m-my; immed-mediate) or multiple iterations of a sound, syllable, word, or phrase counted as repetitions. Three or more immediate repetitions were operationally classed as palilalia (e.g. grandfather father father father).

Diagnostic intelligibility testing followed the format of the Assessment of Intelligibility in Dysarthric Speech [49], adjusted to the accent of the local population. There were 60 items. For each item a word was randomly selected from a set of six minimally varying words (e.g. Buzz, Bus, Budge, Bun, Bud, Butt). For scoring, for each item listeners circled which word they believed they heard from the recordings from a written choice of 12 options. These consisted of the six minimally differing words plus six further foils – so for the item involving the six words above the selection was 'Bun, Moon, Budge, Botch, Bond, Buzz Bus, Bowl, Butt, Boss, Mud, Bud'. The additional words minimise possible learning effects and chance recognition.

To provide a measure of intelligibility to general listeners three raters unfamiliar with dysarthric speech but familiar with the local accent audited randomly ordered tracks independently, blind to speaker characteristics and the purpose of the study. Each item was played once only in a quiet room through a DENON 205SE amplifier and Celestion 7 speakers. The final intelligibility score was the mean number of words correctly identified across the three listeners from the 60 items.

In addition to word identification, listeners graded ease of listening (EOL) for each item (amount of perceived effort required to decipher what was said) (1 extremely easy 7 extremely difficult). EOL was calculated from the mean rating across all items per speaker.

Statistical Analysis

Groups were matched for duration (time since diagnosis) and gender proportion. Given the typical age distribution of the different disorders it was not possible to match for age. Analyses of variance (with method depending on parametricity of data for different analyses) with adjusted post-hoc comparisons were employed to look at differences between groups. Bivariate correlations were used to look at associations between factors. Significance level was set at $p = 0.05$. Tests were two-tailed. Analyses were carried out using SPSS version 22.

Three Grandfather Passage samples were repeated without the listeners' knowledge to investigate intra-rater reliability. Intraclass correlation was used to examine inter-rater agreement on the perceptual scales.

Results

Nine people with MSA-P (3 female), seventeen with PSP (6 female) and twenty-four with PD (8 female) were recruited. Summary descriptive data appear in table 1

Table 1 about here

Groups did not differ significantly for duration or gender proportions. People with MSA-P were significantly younger ($p = 0.009$) than those with PSP, but not than those with PD ($p = 0.084$). For UPDRS III all paired differences were statistically significant (MSA-P vs PSP $p = 0.04$; MSA-P vs PD: $p = 0.001$; PSP vs PD: $p = 0.006$).

The groups with MSA-P and PSP were significantly more affected on UPDRS III Speech rating (Table 1) than those with PD, but equally impaired to each other. Word identification and ease of listening scores reflected this (table 2), with no significant differences between PSP and MSA-P groups but significant differences ($p = 0.001$) between them and those with PD.

Table 2 bout here

The profile of depression (none to mild) was comparable across groups. People with PSP generated significantly fewer words than those with PD on the ACE-R Word Fluency subtest (PD mean 9.5, SD 2.9, PSP mean 6.1, SD 3.7; MSA-P mean 8.8, SD 3.0; PSP vs PD: $p < 0.01$). No other ACE-R subtest or total scores differed significantly.

On the oral-facial apraxia test, those with PSP were significantly more impaired than other groups on the upper face tasks: PSP vs MSA-P: $p = 0.014$; PSP vs PD: $p < 0.001$ - with no significant difference between people with MSA-P vs PD (Table 3). For lower face items people with MSA-P ($p < 0.01$) and with PSP ($p < 0.001$) were significantly poorer than people with PD, with no significant difference between those with MSA-P and PSP.

Five people with MSA-P, four with PSP and 20 with PD showed no or borderline upper face apraxia (score ≥38). Five with MSA-P, six with PSP, twenty-two with PD showed no/borderline lower face apraxia (score ≥400). People with PSP were proportionally significantly more likely to show oral-facial apraxia than people with PD (upper, chi = 12.30, $p = 0.001$; lower, chi = 13.23, $p < 0.001$). The PD vs MSA groups did not differ significantly on upper face, and were borderline for lower, (chi = 3.57, $p = 0.059$). The groups with PSP vs MSA-P did not differ significantly on lower face scores.

Groups overall were only mildly affected on the Token Test. There were significant raw score differences between groups (PSP poorest) but the significance disappeared when adjusted for age. There were members in all groups who showed no impairment on the Token Test.

Table 3 about here

Voice and speech perceptual ratings

The speech-language clinician raters showed high intra-rater agreement for the repeated samples (r 0.895, r 0.918, r 0.975). For inter-rater reliability intraclass correlation coefficients for average measures ranged from low for loudness consistency (0.176), roughness (0.112) and breathiness (0.176), through mid for strain (0.380), asthenia (0.462), to high and very high for grade (0.570), pitch consistency (0.586), pitch level (0.643), nasality (0.719), rate (0.787), loudness level (0.872), fluency rating (0.883), stress (0.898), naturalness (0.916) and articulation (0.943).

Outcomes for the perceptual ratings of speech and voice appear in Tables 4 and 5. All speech rating paired comparisons were non-significant except PwPD were perceived as significantly more monotone (less stress prominence) than those with PSP ($p = 0.04$); people with PSP were perceived as more dysfluent than PwPD ($p = 0.04$); PwPD were judged to have significantly clearer articulation than people with MSA-P ($p < 0.001$) and PSP ($p < 0.01$). People with MSA-P and PSP had perceived poorer loudness and pitch control, but not statistically significant. People with MSA-P were perceived to have significantly lower speech naturalness than those with PD ($p = 0.002$).

On the voice ratings mean scores consistently favoured the PwPD. However, only one comparison reached statistical significance - people with PSP were poorer than those with PD on overall Grade ($p = 0.01$). One other comparison approached significance (MSA-P vs PD on overall Grade, $p = 0.06$). All others were non-significant.

Table 4 about here

Table 5 about here

Fluency measures

Perceptual speech evaluation

In spontaneous speech there were no significant differences between groups in the number of appropriate and inappropriate pauses, nor total dysfluencies, or mean length of utterance. Four people with PSP but none with MSA-P or PD made palilalic repetitions.

On the reading passage all groups showed a similar number of appropriate pauses. Those with PSP (mean 7.2, SD 5.5) displayed significantly more inappropriate pauses ($p = 0.008$) than PwPD (mean 3.3, SD 3.1), but no significant difference from people with MSA-P (mean 6.7, SD 4.9). The difference between MSA-P and PD groups approached significance ($p = 0.06$).

For total dysfluencies the group with PSP (mean 4.1, SD 3.1) showed significantly higher occurrences ($p < 0.001$) than PwPD (mean 1.3, SD 2.3), but not significantly more than people with MSA-P (mean 2.4, SD 2.8). Palilalic repetitions occurred only in four individuals with PSP, and one single instance in someone with PD.

Speech and voice contrasts by severity

Above comparisons are based on groups matched for duration. To examine the possible influence of severity, data were examined in relation to overall motor score (UPDRS III total), and to word-recognition intelligibility score, as a measure of overall speech severity. Due to gross imbalance of severity rankings within and across groups in the cohort it was not possible to conduct reliable analyses of variance.

Taking all groups together the following variables correlated (near) significantly with UPDRS III total score (Spearman's): word-recognition intelligibility (r 0.417, $p = 0.005$), pitch consistency (r 0.295, $p = 0.055$), articulation (r 0.681, $p < 0.001$), nasality (r 0.411, $p = 0.006$), Grade (r 0.371, $p = 0.014$), Roughness (r 0.332, $p = 0.03$) and naturalness (r 0.593, $p < 0.001$).

Perceptual speech evaluation

Within the separate groups individuals with MSA-P showed significant correlations between UPDRS III and voice Breathiness (r 0.681, p <0.05) and naturalness (r 0.704, p = 0.03). Roughness (r 0.644, p = 0.061) approached significance. For the PSP group perceived speech fluency (r 0.655, p = 0.008), articulation (r 0.689, p = 0.005) and naturalness (r 0.554, p = 0.03) correlated significantly with UPDRS III total scores. There were no significant correlations between UPDRS III totals and speech and voice ratings for the group with Parkinson's.

Groups were divided into less and more affected subgroups based on the median UPDRS III total score across all groups (cut-off score 39). Comparisons (Mann-Whitney) within the PSP group between milder (n = 5) and more severe showed significant deterioration for ACE-R Word fluency (p = 0.014), with borderline (p = 0.06) differences for Token Test, ACE-R language and oral-facial apraxia. For the speech and voice measures only speech fluency (p = 0.003), articulation (p = 0.019) and naturalness (p = 0.013) differed significantly. For the group with Parkinson's (milder n = 18) the less vs more affected participants differed significantly on lower oral-facial apraxia (p = 0.036) but no other measures. As there was only one person in the milder group for MSA-P comparisons were not made.

The following variables correlated at or near significance (Spearman's) with the word-recognition intelligibility score when including all groups together: Loudness consistency (r 0.413, p = 0.006), pitch consistency (r 0.420, p = 0.005), articulation (r 0.741, p <0.001), nasality (r 0.444, p = 0.003), naturalness (r 0.532, p <0.001), voice Grade (r 0.556, p = <0.001), Breathiness (r 0.423, p = 0.005). Roughness (r 0.285, p = 0.064) and Asthenia (r 0.284, p = 0.065) approached significance.

Groups were examined independently. For those with MSA-P deterioration in loudness consistency (r 0.738, p = 0.037), articulation (r 0.857, p = 0.007), nasality (r 0.862, p = 0.006), Breathiness (r 0.970, p <0.001), Asthenia (r 0.712, p = 0.048), and naturalness (r

0.814, p = 0.014) correlated significantly with declining intelligibility. In the PSP group pitch level (r 0.526, p = 0.053), articulation (r 0.625, p = 0.017), nasality (r 0.561, p = 0.037), naturalness (r 0.573, p = 0.032) correlated (near) significantly with intelligibility. Perceived speech fluency was r 0.490, p = 0.076. For those with PD only pitch consistency (moving to more monotone) r 0.425, p = 0.055) approached significance.

The subgroups were divided by median word recognition score for the whole cohort (cut-off 49). Only two people with MSA-P and one with PSP fell into the milder group, only four with PD in the more severe group. Individual subgroup analyses were therefore not conducted. Comparing the milder versus more affected participants for all groups combined, the following differences were (near) significant: articulation ($p < 0.001$), Grade ($p = 0.002$), Roughness ($p = 0.021$), Breathiness ($p = 0.055$), naturalness ($p = 0.022$). No other differences approached significance.

Discussion

We aimed to examine whether selected perceptual voice and speech variables separate speakers with PD, PSP and MSA-P. With groups matched for duration, perceptual speech and voice ratings were unable to differentiate groups, contrary to earlier claims [24-27]. There was no indication that groups differed on rating scales that would be likely to detect differences suggestive of added elements of spastic or ataxic dysarthria in the groups with PSP and MSA-P. In particular GRBAS voice quality ratings of Strain, Roughness and Asthenia, prime candidates for variation across groups, did not differ; neither did loudness control, nasality, or perceived rate. Speech fluency was more disrupted in people with PSP both on objective counts and listener overall perception.

PSP is characterised by the development of prominent cognitive changes. Links have been made between PSP and progressive aphasia, as well as presence of a possible dynamic aphasia [50, 51]. The lower performance on ACE-R word generation may tie in with this, and

more dysfluent speech can be related to word retrieval and syntax issues. However, whilst people with PSP performed more poorly on the language measure, the differences did not reach statistical significance when controlled for age. Whether language impairments may have contributed to more speech dysfluencies is unclear here. The fact that pause types and dysfluencies did not differ significantly between groups in the spontaneous speech task, but did so in reading, suggests not, but further testing is required to settle this. For instance, dysfluency in reading may relate to emerging cognitive changes associated particularly with PSP and/or oculomotor or visual perceptual processing disturbances. Speech apraxia can disrupt fluency. We only assessed for oral-facial and not speech apraxia, oculomotor impairment or other visual disturbances.

Some studies link apraxia to PD and other movement disorders. In the current cohorts presence of oral-facial apraxia marked out the group with PSP as more affected, and may thus be an early indicator of more than just cortical basal degeneration [31]. Again, in all groups there were members who showed no evidence of oral-facial apraxia.

The overriding impression was that where groups did differ, when matched for duration, this rested on inter-related measures that capture overall severity – i.e. intelligibility scores, articulation, naturalness and GRBAS Grade scale - rather than discrete speech/voice variables (though the ‘Strain’ scale on GRBAS has been noted as producing least interrater agreement [48], so this factor cannot be excluded here). This may simply reflect that progression in PSP and MSA is faster, thus people with PD look different on speech profiles solely because their changes remain milder. The view concurs with [27] who initially claimed that perceptual features separated people with MSA vs PSP, but later concluded that this rested more on severity related differences. The lack of correspondence between UPDRS III scores and speech measures is a finding across several studies, which may derive from the dependence of speech motor control on non-dopaminergic systems. Here it may also arise from the mild nature UPDRS III and speech impairment in the PD group.

An alternative interpretation could be that much of what looks like severity difference rests on elements of spasticity or ataxia complicating the milder hypokinetic speech status. These may be what give listeners the impression, particularly for people with PSP, of less natural, slower and more dysfluent speech and more impaired articulation and intelligibility than people with PD. Possible corroboration of this comes from studies that found speech diadochokinetic rate performance [30, 33, 34] sensitive to group differences, where slower and more dysrhythmic performance may stem from spastic and ataxic elements. Tykalová et al [52] also showed that a finer grained examination of consonant production may reveal different underlying mechanisms of disruption.

There was some evidence to support this in the present speakers. Whilst measures of voice quality and specific aspects of pitch and loudness performance and resonance did not separate groups significantly based on duration, looking at changes by severity differences between less and more affected individuals hinted at possible subgroup characteristic patterns of deterioration and may tie in more closely with acoustic study findings [29, 33, 34, 36]. People with MSA-P showed increased laryngeal impairment, manifest in poorer loudness consistency and weaker, more breathy voice quality [53]. Increased nasality also featured. People with PSP showed growing difficulty controlling pitch and greater nasality - arguably suggesting increased vocal cord and soft palate spasticity, though again further investigations are required to confirm this. Poorer word generation and speech fluency was also present in people with PSP.

However, within the current cohort at least, claims are tempered by the fact that across practically all measures intra-group variability was as marked as inter-group variability. Individuals may thus display ‘classic’ added spasticity/ataxia, but equally others do not, and this leads to the negative outcomes when comparing groups as a whole. This may also

reflect the inconsistency across acoustic studies for which measures precisely separate groups.

The absence of a clear qualitative difference between groups questions whether aiming to detect such distinctions is a fruitful task. Acoustic analysis measures discrete sound variables, none of which alone corresponds to human auditory perception. Both perceptual and acoustic analyses rest on the final audible signal. This represents a complex composite of influences from multiple interacting sources within the vocal tract and the listener's perceptual sets. Whilst listeners can detect gradations of rate, intensity, pitch, precision and so forth, they can show poor agreement on presence and/or severity of changes [54-56]. This has been a chief criticism of the work that championed the assertion that there are direct localising links between speech features and site of CNS changes [41]. Attempts to replicate their dysarthria categories and the rank order of features meant to characterise them have proved elusive [55, 56]. Accuracy of syndrome/disorder detection by expert listeners based on speech judgements shows very low levels of agreement [57, 58], reflecting more ongoing issues around watertight clinical differential diagnosis [14].

Lansford et al [59] conducted a study asking listeners blind to medical aetiology or possible dysarthria type to assign speakers with motor speech disorders to classificatory categories. The resultant clusters did not match the hypokinetic, ataxic and mixed flaccid-spastic groupings predicted from the medical aetiologies. Rather, differences clustered around contrasts in rate and intelligibility of speech and vocal quality. Kim et al [60] reached similar conclusions in their study with speakers who included those with PD and MSA. They found that classification by dysarthria type produced poorest accuracy, whereas classification by severity matched disease classification much more closely. This appears to be the case in the present study where we did not find evidence that one can separate speakers on isolated dysarthria classificatory features but severity appeared to be a key variable.

Thus, earlier claims that there are distinct perceptually detectable differences between people with PD, PSP, MSA appear unsupported. Nevertheless, anecdotally clinicians do maintain they are able to detect who has PSP or MSA rather than PD. Their optimism may be due to the unblinded nature of speech/voice judgements and availability of other clues e.g. gaze palsy. However, it is also possible that they can detect features within a complex acoustic signal that have not been isolatable acoustically or on perceptual rating scales. This suggests a positive way forwards in understanding the uniqueness of speech changes in different disorders is to employ analytical methods that capture the totality of the speech signal. Promising inroads have been established using linear and non-linear analyses and machine learning methods [61-66] based on simple prolonged vowel sound data or brief mobile phone delivered speech samples. This provides a possible important route out of the acoustic/perceptual impasse.

Some aspects of the current study may have coloured results and invite clarificatory investigations. Apart from the objective fluency measures, judgments were based on a single prolonged /a/ and three sentences from a read passage. Other vowels were not included, nor comparison with results based on spontaneous speech. We did not conduct detailed narrow phonetic analyses of the nature of articulatory changes (see [52]), nor include differentiation of dysarthric from speech apraxic changes. Principally this was because even if possible differences emerged, such labour intensive exercises are unlikely to prove clinically feasible – though recently attempts have been made to employ more automatized screening (above). Participants here were predominantly in the relatively early stages of progression and we excluded dementia severe enough to preclude informed consent, as well as people with major depression. Some of the speech features alleged to distinguish groups may not develop until later in the disease course.

The fact that we included only MSA-P participants and not MSA-C may have influenced the likelihood of detecting ataxic features and detecting distinctions between MSA and PD.

Perceptual speech evaluation

Previous research points to uncertainties in speech-voice differentiation between PwPD and MSA-P [36] and suggests, along with findings here, that more optimistic conclusions of others regarding separation of syndromes may be influenced by studying an undifferentiated group of MSA-P and MSA-C. To maintain viable numbers in groups for comparison and given the ongoing flux in diagnostic criteria for pathological and speech subgroups, we did not differentiate between possible phenotypical subgroups of PSP or MSA [8, 36].

Finally, some studies [27, 29, 33, 36] have noted the possibility that female speakers may be less severely affected in voice and speech than males, especially in the earlier stages. We had similar proportions of female to male patients across groups, but this may be a variable to control or examine more closely in future studies.

The one clear conclusion is that when matched for disease duration groups with MSA-P and PSP are significantly more impaired when speaking than people with PD. It underlines the finding that even if one cannot clinically detect a qualitatively different speech profile, then an early and rapidly deteriorating speech-voice picture is indicative of atypical Parkinson's [67, 68]. It emphasizes that the presence of early articulation changes, rather than just voice deterioration, suggests atypical Parkinsonism. Clinically, this finding stresses the importance of regular planned reviews for such patients within speech-language therapy services to chart rate and patterns of change.

Higher numbers of inappropriate pauses in reading are suggestive of atypical Parkinson's. Present findings agree partially with other studies [16, 33] that suggest repetitions and palilalic iterations are more likely associated with PSP than the other conditions. The status of palilalia as a distinguishing feature, though, remains uncertain. In different studies it is not always clear precisely what counted as palilalia. Even in studies that monitored palilalia it occurred in less than 40% of speakers [9,63]. Others have not found prominent palilalia [26]. In the present study palilalia did not occur in all speakers with PSP.

Conclusion

Whilst diagnosis of PSP, MSA and PD may become obvious at later stages, in the crucial early phase differentiation remains a challenge clinically, with not insignificant rates of mis- or uncertain diagnosis associated with the prominence of shared Parkinson's features. The heterogeneity of presentation of the three disorders and possible subtypes within disorders, and the considerable individual patient variability in presentation and time-course of appearance of symptoms also pose challenges to study design – e.g. as in this study, for precise matching by duration, especially early on. The low prevalence of MSA and PSP and the proportion of participants who can supply complete data sets means that recruitment for statistically sufficiently powerful studies is difficult and points to multicentre collaborations as a way forward.

When differentiating changes to variables such as voice quality and speech rate, one must also consider the issues around possible changes to these that can happen in the normal course of ageing. Thus control groups of neurologically unaffected people should be included to test out the viability of chosen factors argued to identify neurological impairment.

Together with indications from recent acoustic evaluations, this study offers possible productive variables to pursue in such a programme: measures of speech fluency; reasons for less impaired articulation in PD compared to MSA-P and PSP; bases of poorer intelligibility and naturalness in MSA-P, with the possibility of examining prosody features and automatized detection of these in speech [66]. Employment of acoustic approaches that capture the full speech signal, rather than isolated parameters, offers another direction. Given that isolated systems analyses are seldom conclusive in differentiation clinically, and diagnosis relies on constellations of co-occurring features, discriminant analyses that combine voice, speech, language and non-speech measures (e.g. of oral-facial apraxia, handwriting, eye movement, balance and gait) should prove fruitful. Finally, longitudinal

studies of spoken output that track the gradual emergence of distinguishing speech-voice characteristics are lacking but represent a promising direction of research.

Summary points

- Differentiating between idiopathic Parkinson's patients from those with Progressive Supranuclear Palsy or Multiple System Atrophy has important prognostic and therapeutic implications.

- Some have claimed perceptual, naked ear, evaluation of speech and voice changes can assist differential diagnosis. Presence of other impairments, such as oral-facial apraxia, language impairment and other cognitive changes may alert to an atypical Parkinson's syndrome.

- To test out these possibilities we compared speech and voice output, oral-facial apraxia and screening of language comprehension in people with PD, PSP, MSA-P matched for disease duration.

- No consistent and significant differences emerged in terms of identifying predicted added spastic and ataxic dysarthria in the groups with PSP and MSA-P compared to people with PD; language performance did not differ significantly, but people with PSP were more affected by oral-facial apraxia than other groups. Groups did differ significantly on overall measures of performance (articulation, intelligibility, overall voice impairment), and variables did emerge as possible markers of severity in these groups.

- We argue that differences in our findings compared to other studies relate partly to methodological drawbacks of some earlier reports but also to the nature of the

acoustic signal and its perception by listeners.

- Investigating measures that capture the acoustic signal in its entirety rather than focus on isolated sub-features may prove a fruitful line of enquiry.

- Clinically, the presence of early and more rapid deterioration of articulation, intelligibility and naturalness of speech and voice in the context of hypokinetic changes and the higher probability of oral-facial apraxia can be taken as indicators of a possible PSP or MSA-P.

Financial disclosure

The project was funded by Grant 8044 from the Parkinson's Disease Society, Great Britain (now Parkinson's UK). No writing support was employed and the authors have no personal, academic or financial conflicts of interest to declare.

References

1. Pringsheim T, Jette N, Frolkis A, Steeves TDL. The Prevalence of Parkinson's Disease: A systematic review and meta-analysis*. Mov Disord.* 29(13), 1583-1590 (2014).
2. Kollensperger M, Geser F, Ndayisaba JP *et al..* Presentation, diagnosis, and management of multiple system atrophy in Europe: final analysis of the European multiple system atrophy registry*. Mov Disord.* 25(15), 2604-2612 (2010).
3. Roncevic D, Palma JA, Martinez J, Goulding N, Norcliffe-Kaufmann L, Kaufmann H. Cerebellar and parkinsonian phenotypes in multiple system atrophy: similarities, differences and survival*. J Neural Transmission* 121(5), 507-512 (2014).
4. Arena JE, Weigand SD, Whitwell JL *et al.*. Progressive supranuclear palsy: progression and survival*. J.Neurol.* 263(2), 380-389 (2016).
5. Glasmacher SA, Leigh PN, Saha RA. Predictors of survival in progressive supranuclear palsy and multiple system atrophy: a systematic review and meta-

analysis. *J Neurol, Neurosurg, Psychiatry*, online frst doi:10.1136/jnnp-2016-314956 (2017).

6. Respondek G, Stamelou M, Kurz C *et al.*. The phenotypic spectrum of progressive supranuclear palsy: A retrospective multicenter study of 100 definite cases. *Mov Disord.* 29(14), 1758-1766 (2014).
7. Ozawa T, Revesz T, Paviour D *et al.*. Difference in MSA phenotype distribution between populations: genetics or environment? *J Parkinsons Dis* 2(1), 7-18 (2012).
8. Lopez G, Bayulkem K, Hallett M. Progressive supranuclear palsy (PSP): Richardson syndrome and other PSP variants. *Acta Neurol Scand* 134(4), 242-249 (2016).
9. Pont-Sunyer C, Hotter A, Gaig C *et al..* The onset of nonmotor symptoms in Parkinson's disease (The ONSET PDStudy). *Mov Disord.* 30(2), 229-237 (2015).
10. Marras C, Chaudhuri KR. Nonmotor features of Parkinson's disease subtypes. *Mov Disord* 31(8), 1095-1102 (2016).
11. O'Sullivan SS, Massey LA, Williams DR *et al.*. Clinical outcomes of progressive supranuclear palsy and multiple system atrophy. *Brain* 131(5), 1362-1372 (2008).
12. Pradhan S, Tandon R. Progressive supra-nuclear palsy: frequency of cardinal extrapyramidal features at first presentation. *Postgrad Med J* 91(1075), 274-277 (2015).

*In 28 people with PSP only 14% had all six cardinal features on first presentation. Deviations from standard descriptions were common. Emphasizes the heterogeneity of clinical appearance especially in the early stages.

13. Kim HJ, Jeon BS, Jellinger KA. Diagnosis and differential diagnosis of MSA: boundary issues. *J.Neurol.* 262(8), 1801-1813 (2015).
14. Joutsa J, Gardberg M, Röyttä M, Kaasinen V. Diagnostic accuracy of parkinsonism syndromes by general neurologists. *Parkinsonism & Rel Diss* 20(8), 840-844 (2014).
15. Marras C, Lang A. Parkinson's disease subtypes: lost in translation? *J Neurol Neurosurg Psychiatry* 84(4), 409-415 (2013).

16. Testa D, Monza D, Ferrarini M, Soliveri P, Girotti R, Filippini G. Comparison of natural histories of progressive supranuclear palsy and multiple system atrophy. *Neurol Sci* 22(3), 247-251 (2001).

17. Dell'Aquila C, Zoccolella S, Cardinali V *et al.*. Predictors of survival in a series of clinically diagnosed progressive supranuclear palsy patients. *Parkinsonism & Rel Dis* 19(11), 980-985 (2013).

18. Huppertz H-J, Möller L, Südmeyer M *et al.*. Differentiation of neurodegenerative parkinsonian syndromes by volumetric magnetic resonance imaging analysis and support vector machine classification. *Mov Disord.* 31(10), 1506-1517 (2016).

19. Sapir S. Multiple Factors Are involved in the dysarthria associated with Parkinson's disease: a review with implications for clinical practice and research. *J. Speech Lang. Hear. Res.* 57(4), 1330-1343 (2014).

20. Miller N, Allcock L, Jones D, Noble E, Hildreth AJ, Burn DJ. Prevalence and pattern of perceived intelligibility changes in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 78(11), 1188-1190 (2007).

21. Albuquerque L, Martins M, Coelho M *et al.*. Advanced Parkinson disease patients have impairment in prosody processing. *J. Clin Exp. Neuropsychol.* 38(2), 208-216 (2016).

22. Clark JP, Adams SG, Dykstra AD, Moodie S, Jog M. Loudness perception and speech intensity control in Parkinson's disease. *J Comm Dis* 51, 1-12 (2014).

23. Altmann LJP, Troche MS. High-level language production in Parkinson's disease: a review. *Parkinson's Disease* 2011, 238956 (2011).

24. Kluin KJ, Foster NL, Berent S, Gilman S. Perceptual analysis of speech disorders in progressive supranuclear palsy. *Neurology* 43(3 Part 1), 563 (1993).

25. Kluin KJ, Gilman S, Lohman M, Junck L. Characteristics of the dysarthria of multiple system atrophy. *Archives of Neurology* 53(6), 545-548 (1996).

26. Metter EJ, Hanson WR. Dysarthria in progressive supranuclear palsy. (Eds C Moore, K Yorkston, D Beukelman, *Dysarthria and Apraxia of Speech: Perspectives on Management*, Brookes, Baltimore, 127-136 (1991).

27. Hartelius L, Gustavsson H, Astrand M, Holmberg B. Perceptual analysis of speech in multiple system atrophy and progressive supranuclear palsy*. J Med.. Speech-Lang Pathol* 14(4), 241-247 (2006).

28. Saxena M, Behari M, Kumaran SS, Goyal V, Narang V. Assessing speech dysfunction using BOLD and acoustic analysis in parkinsonism*. Parkinsonism & Rel Dis* 20(8), 855-861 (2014).

29. Skodda S, Visser W, Schlegel U. Acoustical analysis of speech in progressive supranuclear palsy*. J Voice* 25(6), 725-731 (2011).

30. Skodda S, Gronheit W, Schlegel U. Instability of syllable repetition in progressive supranuclear palsy*. J Neural Transmission* 119(4), 457-462 (2012).

31. Ozsancak C, Auzou P, Dujardin K, Quinn N, Destee A. Orofacial apraxia in corticobasal degeneration, progressive supranuclear palsy, multiple system atrophy and Parkinson's disease*. J.Neurol.* 251(11), 1317-1323 (2004).

32. Kim J-H, McCann CM. Communication impairments in people with progressive supranuclear palsy: A tutorial*. J Comm Dis* 56, 76-87 (2015).

33. Rusz J, Bonnet C, Klempíř J *et al.*. Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy*. J.Neurol.* 262(4), 992-1001 (2015).

** Comprehensive acoustic comparision of PD, PSP, MSA and healthy groups based on prolonged vowel and diadochokinetic (DDK) syllable repetition. Combination of results from jitter, inappropriate silences, slowed DDK, excessive instensity variation and picth fluctuation separated groups PD vs atypical groups with 95% accuracy. Harsh voice, fluency, voice tremor, slwoed rate differentiated PSP vs MSA with 75% accuracy.

34. Penner H, Miller N, Wolters M. Motor speech disorders in three parkinsonian syndromes: a comparative study*. 16th International Congress of Phonetic Sciences, Saarbruecken, August*, 1989-1992 (2007).

35. Sachin S, Shukla G, Goyal V *et al.*. Clinical speech impairment in Parkinson's disease, progressive supranuclear palsy, and multiple system atrophy*. Neurology India* 56(2), 122-126 (2008).

36. Huh YE, Park J, Suh MK *et al.*. Differences in early speech patterns between Parkinson variant of multiple system atrophy and Parkinson's disease*. Brain Lang* 147, 14-20 (2015).

*One of very few comparisons of speech in MSA-P only (n = 26) versus PD (n = 29). Employed acoustic and perceptual measures (e.g. pitch, rate, pauses, voice perturbation). Only voice pitch, speech rate and total pause time in male speakers with MSA-P differed from people with PD. No features distinguihsed female speakers.

37. Skodda S. Aspects of speech rate and regularity in Parkinson's disease*. J Neurol Sci* 310(1-2), 231-236 (2011).

38. Skodda S, Lorenz J, Schlegel U. Vocal rhythm performance in Parkinson's disease*. Mov Disord.* 25(7), S364-S365 (2010).

39. Litvan I, Agid Y, Calne D *et al.*. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome)*. Neurology* 47(1), 1-9 (1996).

40. Osaki Y, Wenning GK, Daniel SE *et al.*. Do published criteria improve clinical diagnostic accuracy in multiple system atrophy? *Neurology* 59(10), 1486-1491 (2002).

41. Darley FL, Aronson AE, Brown JR. *Motor speech disorders*. W.B. Saunders, Philadelphia. (1975).

42. Goetz CG, Poewe W, Rascol O *et al.*. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations*. Mov Disord.* 18(7), 738-750 (2003).

43. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatr* 21(11), 1078-1085 (2006).

44. Montgomery SA, Asberg M. New depression scale designed to be sensitive to change. *Br. J. Psychiatry* 134, 382-389 (1979).

45. Bizzozero I, Costato D, Sala SD, Papagno C, Spinnler H, Venneri A. Upper and lower face apraxia: role of the right hemisphere. *Brain* 123(Pt 11), 2213-2230 (2000).

46. Miller N, Willmes K, Bleser RD. The psychometric properties of the English language version of the Aachen Aphasia Test (EAAT). *Aphasiology* 14(7), 683-722 (2000).

47. Hirano M, Arnold WW. *Disorders of Human Communication, Clinical Examination of Voice.* Springer, New York (1981).

48. Webb AL, Carding PN, Deary IJ, Mackenzie K, Steen N, Wilson JA. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Oto-Rhino-Laryn* 261(8), 429-434 (2004).

49. Yorkston K, Beukelman D. *Assessment of intelligibility in dysarthric speech*. CC Publications, Tigard Oregon. (1981).

50. Rohrer JD, Paviour D, Bronstein AM, O'Sullivan SS, Lees A, Warren JD. Progressive supranuclear palsy syndrome presenting as progressive nonfluent aphasia: a neuropsychological and neuroimaging analysis. *Mov Disord.* 25(2), 179-188 (2010).

51. Robinson G, Shallice T, Cipolotti L. Dynamic aphasia in progressive supranuclear palsy: a deficit in generating a fluent sequence of novel thought. *Neuropsychologia* 44(8), 1344-1360 (2006).

52. Tykalová T, Rusz J, Klempir J, Cmejla R, Ruzicka E. Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Brain Lang* 165, 1-9 (2017).

53. Lalich IJ, Ekbom DC, Starkman SJ, Orbelo DM, Morgenthaler TI. Vocal fold motion impairment in multiple system atrophy. *Laryngoscope* 124(3), 730-735 (2014).

54. Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. *J Acoust Soc Am* 122(4), 2354-2364 (2007).

55. Zyski BJ, Weisiger BE. Identification of dysarthria types based on perceptual analysis. *J Comm Dis* 20(5), 367-378 (1987).

56. Zeplin J, Kent R: Reliability of auditory perceptual scaling in dysarthria. In: *Disorders of Motor Speech*, Robin D,Yorkston K,Beukelman D (Eds). Brookes, Baltimore 145-154 (1996).

57. Fonville S, Van Der Worp HB, Maat P, Aldenhoven M, Algra A, Van Gijn J. Accuracy and inter-observer variation in the classification of dysarthria from speech recordings. *J.Neurol.* 255(10), 1545-1548 (2008).

58. Van Der Graaff M, Kuiper T, Zwinderman A *et al.*. Clinical identification of dysarthria types among neurologists, residents in neurology and speech therapists. *Eur Neurol* 61(5), 295-300 (2009).

59. Lansford KL, Liss JM, Norton RE: Free-classification of perceptually similar speakers with dysarthria. *J Speech Lang Hear Res* 57(6), 2051-2064 (2014).

*23 student speech clinicians rated similarities and differences between 22 speakers with dysarthria, blind to diagnosis (ALS 10; Huntington's 4; PD 8). Three clusters of similarity emerged – related to perceptual and acoustic featrures, not medical diagnosis. They were articulation rate, intelligibility, voice quality.

60. Kim Y, Kent RD, Weismer G. An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *J Speech Lang Hear Res* 54(2), 417-429 (2011).

61. De Bodt MS, Huici M, Van De Heyning PH. Intelligibility as a linear combination of dimensions in dysarthric speech. *J Comm Dis* 35(3), 283-292 (2002).

62. Khan MM, Chalup SK, Mendes A. *Parkinson's* disease data classification using evolvable wavelet neural networks. In: *Artificial Life and Computational Intelligence, Acalci 2016*, Ray T,Sarker R,Li X (Eds). 113-124 (2016).

63. Mekyska J, Galaz Z, Mzourek Z *et al.*. Assessing progress of Parkinson's disease using acoustic analysis of phonation*. 2015 4th International Work Conference on Bioinspired Intelligence (IWOBI)*, 111-118 (2015).

64. Orozco-Arroyave JR, Honig F, Arias-Londono JD *et al.*. Automatic detection of Parkinson's disease in running speech spoken in three different languages*. J Acou Soc Am* 139(1), 481-500 (2016).

*Compared accuracy of machine driven recognition of PD vs healthy speaker speech in Spanish, German, Czech based on syllable repetition, readnig and sentences. Depending on language and speech task accuracies range 85% to 99%. Shows robustness of tasks and methods across languages and usefulness of automatic assessment of dysarthric speech signals.

65. Bayestehtashk A, Asgari M, Shafran I, McNames J. Fully automated assessment of the severity of Parkinson's disease from speech*. Computer Speech & Language* 29(1), 172-185 (2015).

66. Bandini A, Giovannelli F, Orlandi S *et al.*. Automatic identification of dysprosody in idiopathic Parkinson's disease*. Biomedical Signal Processing & Control* 17, 47-54 (2015).

*Twenty PD patients, 19 healthy controls. To demonstrate utility of automatic acoustic analysis in diagnosis. Significant differences between the groups on time intervals between sentence repetitions, percent of speech time in relation to sentence duration. Noise and fundamental frequency variability showed no significant difference.

67. Goetz CG, Leurgans S, Lang AE, Litvan I. Progression of gait, speech and swallowing deficits in progressive supranuclear palsy*. Neurology* 60(6), 917-922 (2003).

68. Wenning GK, Ben Shlomo Y, Hughes A, Daniel SE, Lees A, Quinn NP. What clinical features are most useful to distinguish definite multiple system atrophy from Parkinson's disease? *J.Neurol NeurosurgPsychiatry* 68(4), 434-440 (2000).

Measure	PD		PSP		MSA-P	
	Mean Range	SD	Mean Range	SD	Mean Range	SD
Age years	67.77 48-87	10.19	70.94 55-85	6.73	58.8 44-83	10.9
Years since diagnosis	5.03 3-11	1.95	5.06 2-10	2.19	4.1 1-9	2.98
UPDRS III Total 0 normal - 104	31.79 17-55	9.83	44.29 18-67	13.45	57.0 34-81	14.2
UPDRS Speech 0 normal - 4	1.04 0-2	1.03	2.47 1-3	0.62	2.8 1-4	0.97
MADRS, 0 (normal) - 60	11.71 5-25	5.30	13.71 2-27	8.42	16.44 4-27	6.39
ACE-R total 100 unaffected	86.4 68-97	8.65	81.3 53-95	11.00	89.33 75-97	7.61

Table 1: Summary group data for demographic variables and general assessments. SD, standard deviation.

	PD		PSP		MSA-P	
	Mean Range	SD	Mean Range	SD	Mean Range	SD
WR (0-60 higher better)	51.7 39-59	4.5	34.9 15-55	13.8	39.5 26-49	7.3
EOL 1 mild – 7	1.9 1-4	0.85	4.6 2-7	1.9	4.0 3-6	1.0

Table 2: Intelligibility scores by word recognition (WR), ease of listening (EOL)

	PD		PSP		MSA-P	
	Mean Range	SD	Mean Range	SD	Mean Range	SD
OFA upper Max 45; cut-off 38	42.4 28-45	5.15	29.8 8-45	11.11	39.7 23-45	7.6
OFA lower Max 435, cut-off 400	424.9 362-435	18.5	351.5 216-435	61.5	373.7 295-435	50.6
TT age adjusted Max 50 normal	48.8 25-50	5.13	46.1 35-50	6.39	49.9 49-50	0.35

Table 3: Oral-facial apraxia (OFA) and Token Test (TT) scores.

Perceptual speech evaluation

	PD		PSP		MSA-P	
	Mean Range	SD	Mean Range	SD	Mean Range	SD
Loudness level 7 normal	5.6 3-7	1.2	4.5 2-9	1.9	6.2 2-10	2.6
Loudness consistency 7 normal	6.0 4-7	1.1	4.8 3-7	1.8	5.5 2-8	1.2
Pitch level 7 normal	7.5 6-9	0.9	7.8 6-9	1.2	7.4 6-10	1.2
Pitch consistency 7 normal	6.2 4-8	1.1	4.8 2-8	1.9	5.9 4-7	1.1
Stress prominence 7 normal	2.9 1-7	2.4	4.8 1-8	2.3	3.8 2-7	1.6
Perceived speech rate 7 normal	7.0 4-10	1.0	7.0 4-10	1.7	7.1 3-10	2.2
Speech fluency 1 mild - 7	3.8 1-6	2.1	5.2 1-7	2.6	4.0 1-6	1.8
Articulation 1 mild - 7	1.4 1-3	0.5	4.0 1-7	2.0	3.6 1-7	1.8
Nasality 1 mild - 7	2.1 1-5	1.3	3.0 1-5	1.3	2.4 1-5	1.2
Naturalness 1 mild -7	2.4 1-5	1.1	3.6 1-7	2.2	4.9 2-7	1.8

Table 4: Mean group perceptual ratings across three listeners for selected speech output variables.

GRBAS 0 normal-3	PD Mean Range	 SD	PSP Mean Range	 SD	MSA-P Mean Range	 SD
Grade	0.9 0-2	0.61	1.5 1-3	0.66	1.5 1-3	0.64
Roughness	0.6 0-2	0.4	0.6 0-2	0.5	0.9 0-2	0.4
Breathiness	0.9 0-2	0.5	1.2 0-2	0.6	1.2 0-2	0.6
Asthenia	0.9 0-2	0.6	1.4 0-2	0.8	1.4 0-3	0.9
Strain	0.5 0-1	0.4	0.6 0-2	0.4	1.2 0-2	1.0

Table 5: Mean group ratings across three listeners for voice quality.