Consonant articulation acoustics and intelligibility in Swedish speakers with Parkinson’s disease: a pilot study

Inga-Lena Johansson, Christina Samuelsson & Nicole Müller

To cite this article: Inga-Lena Johansson, Christina Samuelsson & Nicole Müller (2022): Consonant articulation acoustics and intelligibility in Swedish speakers with Parkinson’s disease: a pilot study, Clinical Linguistics & Phonetics, DOI: 10.1080/02699206.2022.2095926

To link to this article: https://doi.org/10.1080/02699206.2022.2095926

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

View supplementary material

Published online: 14 Jul 2022.

Submit your article to this journal

Article views: 235

View related articles

View Crossmark data
Consonant articulation acoustics and intelligibility in Swedish speakers with Parkinson’s disease: a pilot study

Inga-Lena Johansson, Christina Samuelsson, and Nicole Müller

ABSTRACT
Imprecise consonant articulation is common in speakers with Parkinson’s disease and can affect intelligibility. The research on the relationship between acoustic speech measures and intelligibility in Parkinson’s disease is limited, and most of the research has been conducted on English. This pilot study investigated aspects of consonant articulation acoustics in eleven Swedish speakers with Parkinson’s disease and six neurologically healthy persons. The focus of the study was on consonant cluster production, articulatory motion rate and variation, and voice onset time, and how these acoustic features correlate with speech intelligibility. Among the measures in the present study, typicality ratings of heterorganic consonant clusters /spr/ and /skr/ had the strongest correlations with intelligibility. Measures based on syllable repetition, such as repetition rate and voice onset time, showed varying results with weak to moderate correlations with intelligibility. One conclusion is that some acoustic measures may be more sensitive than others to the impact of the underlying sensory-motor impairment and dysarthria on speech production and intelligibility in speakers with Parkinson’s disease. Some aspects of articulation appear to be equally demanding in terms of acoustic realisation for elderly healthy speakers and for speakers with Parkinson’s disease, such as sequential motion rate measures. Clinically, this would imply that for the purpose of detecting signs of disordered speech motor control, choosing measures with less variation among older speakers without articulation impairment would lead to more robust results.

Introduction
In Parkinson’s disease, hypokinetic dysarthria is common. Among speech characteristics associated with hypokinetic dysarthria is imprecise articulation (Duffy, 2019; Ho et al., 1999; Miller, 2017), which can affect intelligibility. Articulatory impairment is found in around 40% of speakers with Parkinson’s disease (Ho et al., 1999). Consonants have more complex phonetic and acoustic characteristics than vowels (Kent & Read, 2002). Although listeners’ perceptual judgements, ratings and transcriptions can give information about decreased speech intelligibility, the addition of acoustic analysis can provide more details about how impaired speech motor functions contribute to this (Ackermann & Ziegler, 1991).
There is a considerable body of research that has investigated acoustic aspects of dysarthria in Parkinson’s disease, but research on the relationship between acoustic measures and intelligibility is still limited. Our study focuses on the acoustics of consonant articulation, specifically on consonant cluster production, articulatory motion rate and rate variation, and voice onset time measurements, as well as on how these aspects of speech production correlate with intelligibility. We therefore restrict our review of previous research to studies relevant to these aspects.

Imprecise consonant articulation has been observed in perceptual studies on speech in Parkinson’s disease (e.g. Duez et al., 2020; Read et al., 2018). As regards acoustic studies, reduced articulatory precision in the occlusion phase of stop consonants has been noted, with a larger reduction in unstressed syllables, and this was also significantly correlated with listener ratings of severity (Ackermann & Ziegler, 1991). Consonant segment duration in speakers with Parkinson’s disease does not seem to differ significantly from healthy speakers (Weismer et al., 2001). Tjaden and Martel-Sauvageau (2017) compared consonant spectra from fricatives and stops in clear and loud speech conditions. There was no enhancement of spectral contrasts due to speech conditions and the observed subtle changes were similar for speakers with mild dysarthria due to Parkinson’s disease and healthy speakers. In a case study of a speaker with Parkinson’s disease, spectrographic analysis showed indistinct consonant transitions, and that ratings of spectrograms were moderately correlated with intelligibility (Kempler & Van Lancker, 2002).

Consonant clusters put particular demands on articulatory precision (Kim & Gurevich, 2021; Kuruvilla-Dugdale et al., 2020; Rusz, Hlavnicka et al., 2021; Schalling, 2014). They can be divided into two types: homorganic clusters, where the sounds have the same place of articulation (such as /st/), and heterorganic clusters, where the sounds have different places of articulation (such as /sp/). Effects of cluster type (homorganic versus heterorganic) on speech production has been studied in persons who stutter with somewhat ambiguous results (Huinck et al., 2004). Kim and Gurevich (2021) investigated effects of word position on consonant production in six children with spastic cerebral palsy, which included consonants in clusters in medial and final word position. Results from this study showed moderate to strong correlations between perceptual judgements of consonant production accuracy in clusters and intelligibility. Other than that, we have not been able to find studies specifically investigating consonant cluster production in dysarthria and its relationship with intelligibility.

Syllable repetition can be used to assess rate, regularity, and precision of movements as well as coordination between laryngeal and supra-laryngeal movements (Chenausky et al., 2011; Rusz, Tykalova, Ramig et al., 2021). Articulatory motion rate is commonly divided into alternating motion rate (AMR), which refers to the same syllable being repeated (e.g. /papapa/) at maximum rate, and sequential motion rate (SMR), where a sequence of different syllables is repeated (e.g. /pataka/; Duffy, 2019). Results from previous studies on Parkinson’s disease on syllable repetition rate are contradictory. Some studies report that maximum repetition rate does not seem to differ between speakers with Parkinson’s disease and healthy speakers (Connor et al., 1989; Harel et al., 2004; Skodda, 2011, 2015), and it has been suggested that a reason for the relatively unimpaired articulatory motion rate, despite bradykinesia, may be a compensatory effect by articulatory undershoot (Ackermann et al., 1997; Connor et al., 1989). Others report that speakers with Parkinson’s disease have slower repetition rates than healthy speakers.
(Chenausky et al., 2011; Lowit et al., 2018; Rusz, Tykalova, Ramig et al., 2021). As regards regularity in syllable repetition, the results from previous research from rapid repetition rate tasks are also divergent, with some studies reporting that syllable length is irregular compared to healthy speakers (Chenausky et al., 2011 Lowit et al., 2018), and others that it is not (Harel et al., 2004; Rusz, Tykalova, Ramig et al., 2021). Methodology and speaker samples differ between the studies, which could explain the contradictory findings. Speakers with Parkinson’s disease seem to have difficulties with regulating pace and keeping it steady during syllable repetition, but this has been investigated primarily in tasks with a self-paced, comfortable rate (Ackermann et al., 1997; Skodda, 2011, 2015). The relationship between syllable repetition tasks and communicative functionality is not clear. According to Rusz, Tykalova, Ramig et al. (2021), AMR or SMR is typically not related to intelligibility in cases of basal ganglia dysfunction. In studies on Swedish speakers with Parkinson’s disease, it has been found that perceptually judged articulatory imprecision and level of impairment may be predicted by acoustic measures from diadochokinetic tasks, especially from syllable repetition of /ka/ (Karlsson & Hartelius, 2019; Karlsson et al., 2020).

A measure for assessing coordination between laryngeal and supra-laryngeal movements is voice onset time (VOT), which refers to the time between the release of the oral constriction for plosive production and the onset of periodic vocal-fold vibration (Kent & Read, 2002; Kent et al., 1999). Studies across several languages have established that VOT varies with place of articulation (labial < dental < velar; Cho & Ladefoged, 1999; Helgason & Ringen, 2008; Lisker & Abramson, 1964; Lundeborg et al., 2012) and speech rate (Auzou et al., 2000; Fischer & Goberman, 2010). Measures that neutralise the effect of speech rate have been developed. One example is the VOT ratio, where the VOT is divided by the syllable or word duration (Fischer & Goberman, 2010; Tykalova et al., 2017). Results from studies investigating VOT in Parkinson’s disease are mixed. Most studies (e.g. Forrest et al., 1989; Harel et al., 2004; Rusz, Hlavnicka et al., 2021; Weismer et al., 2001) show that VOT tends to be longer for speakers with Parkinson’s disease than for healthy speakers, although this difference has not been statistically significant in all studies. In one study, shorter VOT in speakers with Parkinson’s disease compared to healthy speakers has also been found (Flint et al., 1992). It has been suggested that the varying results could emanate from physiological differences in disease-related impact: On the one hand slower or disordinated laryngeal movements resulting in longer VOT and, on the other hand, a smaller glottal opening due to rigidity leading to shorter VOT (Flint et al., 1992; Harel et al., 2004). VOT is more variable in speakers with Parkinson’s disease than in healthy controls (Chenausky et al., 2011), but healthy elderly speakers also have greater variability in VOT than younger speakers (Auzou et al., 2000). Özsancak et al. (2001) discuss challenges when measuring VOT in dysarthria. For example, the stop burst might be unidentifiable, which can occur in Parkinson’s disease and make VOT unmeasurable. These measurement difficulties tend to be more common for velar than labial or dental speech sounds.

As previously discussed, there is limited research on the relationship between acoustic speech measures and intelligibility in Parkinson’s disease, and most studies on speech acoustics in dysarthria have been conducted on English. There is a need for more studies on non-English dysarthria to widen the cross-linguistic knowledge base on factors affecting intelligibility. Although some measures may be cross-linguistically robust, speech sound inventories and phonotactic structures differ between languages, as does their contribution
to signallng contrasts that are important for intelligibility. Therefore, consequences of a motor speech impairment could depend on the speaker’s native language (Kim & Choi, 2017; Miller et al., 2014; Pinto et al., 2017). For example, timing of coarticulatory oral and laryngeal movements can differ between languages, resulting in language-dependent variations in VOT (Lisker & Abramson, 1964) or articulatory motion rate (Rusz, Hlavnicka et al., 2021). The present study was conducted in Sweden. In Swedish, voiceless stops are aspired in word-initial position, which results in longer VOT than for unaspirated stops (Engstrand, 2004; Helgason & Ringen, 2008). Like other Germanic languages, Swedish is also rich in consonant clusters (Engstrand, 2004; Riad, 2013), which are likely to be articulatorily demanding in the context of dysarthria (Kim & Gurevich, 2021; Kuruvilla-Dugdale et al., 2020; Rusz, Hlavnicka et al., 2021; Schalling, 2014).

In the field of acoustic speech analysis, quantitative acoustic measures based on temporal or spectral features have frequently been used, but visual examination of qualitative details of speech production may also contribute important details (Liss & Weismer, 1992). The visualisation of the three sound dimensions time, frequency and relative intensity in combined waveforms/spectrograms could, provided that it contributes relevant information related to speech intelligibility, potentially be useful for clinical assessment as well as intervention purposes. For example, for consonant clusters, where both features of the individual target speech sounds and the coarticulation of them need to be evaluated, a visual examination could be useful. Besides the case study by Kempler and Van Lancker (2002), we have not found studies investigating the relationship between visual examination of spectrograms and intelligibility. Therefore, we wanted to investigate whether this method had any potential to be a valid and reliable component in assessment of factors relevant to intelligibility.

The aims of the present study were to investigate (a) acoustic properties of the speech from Swedish speakers with Parkinson’s disease, focusing on consonant cluster production by visual examination of waveform diagrams and spectrograms, articulatory motion rate and variation, and VOT, and (b) how these acoustic measures correlate with speech intelligibility.

**Methods and materials**

**Participants**

Eleven speakers experiencing speech problems due to Parkinson’s disease and six neurologically healthy persons participated. The distribution regarding age, gender and regional variants of Swedish were similar for both groups. Since one aim for the present study was to investigate possible relations between acoustic features and speech intelligibility in speakers with Parkinson’s disease, a sample reflecting the known heterogeneity among speakers with Parkinson’s disease (e.g. Feenoughy et al., 2014; Read et al., 2018) was desired. Exclusion criteria were speech or language disorder not related to Parkinson’s disease or a dementia diagnosis. None of the speakers had undergone any neuro-surgical treatment, such as deep brain stimulation (DBS). Since hearing impairment is common among the elderly (Cruickshanks et al., 1998), this was not an exclusion criterion for the study. Hearing impairment was investigated through self-report, and the distribution of impaired hearing and use of hearing aids was similar in both groups. Dysarthria severity and overall disease
severity were rated by the first author. Dysarthria was assessed with the short form of the Swedish Dysarthria Assessment (Hartelius, 2015) and was mild for four participants, moderate for five participants, and moderate to severe for two participants. Overall disease severity was rated with the Hoehn and Yahr-scale (Hoehn & Yahr, 1967). The Hoehn and Yahr-scale scale ranges from stage 1 (very mild symptoms) to stage 5 (severe symptoms with constant need of nursing care). Three participants were in stage 2, five in stage 3 and three in stage 4. Other group characteristics are presented in Table 1.

Prior to data collection, all participants provided written informed consent. Ethical approval for the study was obtained from the relevant regional Ethical Review Board (2015/446-31).

### Data collection

For the audio-recordings of the participants’ speech, we used a Marantz PMD 660 and a table microphone AKG C 535 EB placed at 35 cm distance from the participant’s mouth. Settings for the recordings were a sampling rate of 48 kHz, 16 bit uncompressed PCM/WAV-format, and an input recording level set to minimise risk for sound clipping or noise in normal conversational loudness.

Seven of the participants with Parkinson’s disease were recorded in their homes and the rest in university locations, depending on the participants’ choice. Regardless of the location, it was ensured that the recording environment was quiet and without background noise. The participants also chose the preferred time, based on their experiences of what time of day they were functioning at their best. Each recording lasted about 50 minutes. Besides the collection of speech materials for analysis, which is described in more detail below, it included an interview for collection of personal and disease background data, speech materials for evaluation of intelligibility (Johansson et al., 2022), and a cognitive screening with the Mini Mental State Examination – Swedish revision (MMSE – SR; Palmqvist et al., 2012).

<table>
<thead>
<tr>
<th>Table 1. Group characteristics.</th>
<th>Speakers with Parkinson’s disease (n = 11)</th>
<th>Healthy speakers (n = 6)</th>
<th>Group comparisonsa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong></td>
<td><strong>SD</strong></td>
<td><strong>Range</strong></td>
<td><strong>M</strong></td>
</tr>
<tr>
<td>Age</td>
<td>71</td>
<td>6.84</td>
<td>61–82</td>
</tr>
<tr>
<td>MMSE-scoreb</td>
<td>28</td>
<td>1.57</td>
<td>26–30</td>
</tr>
<tr>
<td>Intelligibility (%)c</td>
<td>79</td>
<td>15</td>
<td>52–95</td>
</tr>
<tr>
<td>Time since diagnosis (yrs)</td>
<td>7</td>
<td>3.86</td>
<td>1–11</td>
</tr>
<tr>
<td>Gender</td>
<td>4 F/7 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hearing impairment</td>
<td>4/2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aMann Whitney U*

*bMaximum possible score = 30*

*cPooled percentage of scores from transcription of sentences from picture description and the Swedish Test of Intelligibility (STI; Hartelius, 2015)*

CLINICAL LINGUISTICS & PHONETICS
Speech materials

The speech materials for the acoustic analyses of articulatory aspects included (a) tasks with alternate and sequential articulatory movements (AMR and SMR), where the speakers repeated the syllables /pa/, /ta/, /ka/, and the syllable sequence /pataka/ at maximum rate for approximately five seconds, and (b) oral reading of phonetically balanced sentences from the Swedish dysarthria assessment (Hartelius, 2015), which also included words with three-consonant clusters. The intelligibility variable was a pooled percentage measure of intelligibility in transcription tasks including sentences from the Swedish Test of Intelligibility (STI; Hartelius, 2015) and picture description. The sentences were transcribed by, in total, 28 naive listeners per speaker and task. All listeners had Swedish as primary language, and most of them were university students. Intelligibility ranged between 52% and 95% for the speakers with Parkinson’s disease (M = 79; SD = 14) and between 86% and 98% for the healthy speakers (M = 95; SD = 5). The methods for the intelligibility measures are described in more detail in a separate study on intelligibility assessment (Johansson et al., 2022).

Acoustic analyses

An overview of the measures applied in the present study can be found in Table 2. Acoustic measures of dysarthric speech can be challenging, because of atypicalities in the signal (Kent et al., 1999). Therefore, measurements were made from both waveform diagrams and spectrograms obtained in the software Praat (Boersma & Weenink, 2015–2020).

Consonant cluster production

Production of speech sounds in consonant clusters was judged through typicality ratings of waveform diagrams and wide-band spectrograms for the initial consonant clusters in three target words: spräckliga (speckled), strök (passed), and skrattade (laughed). A waveform diagram together with a spectrogram per target word from

Table 2. Overview of acoustic measures.

<table>
<thead>
<tr>
<th>Speech component</th>
<th>Speech material</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonant articulation</td>
<td>Word initial consonant clusters /spr/, /str/, and /skr/ extracted from orally read sentences</td>
<td>Spectrogram ratings on six-point scale&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
<tr>
<td>Alternating motion rate (AMR)</td>
<td>Syllable repetition (/pa/; /ta/; /ka/)</td>
<td>Syllables/sec</td>
</tr>
<tr>
<td>Intersyllabic variation of duration in AMR tasks</td>
<td>Syllable repetition (/pa/; /ta/; and /ka/)</td>
<td>Intersyllabic variation quotient (IVQ)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sequential motion rate (SMR)</td>
<td>Syllable repetition (/pa-ta-ka/)</td>
<td>Syllables/sec</td>
</tr>
<tr>
<td>Intersyllabic variation of duration in SMR task</td>
<td>Syllable repetition (/pa-ta-ka/)</td>
<td>Intersyllabic variation quotient (IVQ)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Voice onset time (VOT)</td>
<td>Syllable repetition (/pa/, /ta/, and /ka/)</td>
<td>Time (ms) between stop release and onset of voicing</td>
</tr>
<tr>
<td>VOT ratio</td>
<td>Syllable repetition (/pa/, /ta/, and /ka/)</td>
<td>VOT (ms)/syllable duration (ms)</td>
</tr>
<tr>
<td>Voice onset time variability</td>
<td>Syllable repetition (/pa/, /ta/, and /ka/)</td>
<td>Intraspeaker standard deviation of VOT</td>
</tr>
</tbody>
</table>

<sup>4</sup>0 = speech sound/cluster looks very non-typical; speech sound is impossible to differentiate from neighbouring sounds or is completely missing; 5 = speech sound/cluster looks very typical and speech sounds can very easily be differentiated from neighbouring sounds.

<sup>b</sup>IVQ = SD<sub>SYLL DUR</sub>/average syllable duration
each speaker were extracted from an orally read sentence. Spectrogram settings were kept the same for all speakers with a view range 0–8000 Hz for the spectrogram, window length 0.005 sec, and dynamic range 70 dB. Three rating sets were prepared, one for each target word. In each set, the waveform diagrams/spectrograms for the target word from all speakers were included, and their ordering was randomised for each rater. Raters evaluated each cluster (/spr/, /str/, and /skr/) as a whole, as well as the individual consonants in each cluster. Adjacent to each waveform diagram/spectrogram was a table with a six-point rating scale from 0 to 5 (0 = speech sound/cluster looks very non-typical, speech sound is impossible to differentiate from neighbouring sounds or is completely missing; 5 = speech sound/cluster looks very typical and speech sounds can very easily be differentiated from neighbouring sounds), and a space where the rater could type comments, if desired. Examples of the rating form can be found in the Supplementary material. Besides the information in the rating form, the raters got a separate letter with information about the rating procedure and scale, the target words, the number of items, the appearance of the spectrograms including their view range, and variants of the Swedish pronunciation of /r/.

Raters were six English-speaking clinical linguists with expertise in acoustic phonetics who were not familiar with the speakers or the recorded material. The choice to invite English-speaking raters was based on a trial of the procedure before the main data collection, which included both Swedish- and English-speaking raters. Results from this revealed that experience of spectrogram analysis rather than native language enhanced reliability of ratings. The raters were informed about the target words, and that the rating sets included samples from both speakers with Parkinson’s disease and healthy speakers, but they were blinded to which group each speaker belonged to. The ratings were completed individually.

**Alternating and sequential articulatory movements**

Measures on alternating motion rate (AMR) and sequential motion rate (SMR) were based on ten consecutive syllables for AMRs and twelve syllables for SMRs, beginning 2–3 syllables from the start, to obtain the measure from a sequence of the best performance with most distinct and even repetitions from the speaker. Syllable duration was manually measured by the first author in waveforms and spectrograms. Rate was calculated using the formula

\[
\text{Syllables/sec} = \frac{10}{(\text{average syllable duration (sec)} \times 10)
\]

To normalise for speech rate differences, intersyllabic variation of duration was measured by calculating the intersyllabic variation quotient (IVQ; Chenausky et al., 2011):

\[
\text{IVQ} = \frac{\text{SD}_{\text{SYLL.DUR}}}{\text{average syllable duration (ms)}}
\]

**Voice onset time**

Measurements of VOT were conducted by the first author. VOT was measured for voiceless stops in ten consecutive syllables of alternating articulatory movements (/pa/, /ta/, and /ka/, respectively) extracted similarly as for evaluation of syllable repetition rate. VOT was measured from the beginning of the burst to onset of phonation, that is, the first regular phonation cycle. For speakers where double bursts occurred, time was measured from the
first burst. The measure for inter-syllabic variation was the standard deviation of VOT. VOT ratio, which neutralises the effect of speech rate on the VOT measure, was also calculated, as described by Fischer and Goberman (2010):

\[
\text{VOT ratio} = \frac{\text{VOT}(\text{ms})}{\text{syllable duration (ms)}}
\]

VOT was not measured if the stop burst was unidentifiable, which was the case in some syllables. The average number of measurable syllables for speakers with PD/healthy speakers were 8.09/9.83 for /pa/, 8.09/9.67 for /ta/, and 6.45/10.00 for /ka/.

**Statistical analyses**

The raters’ individual values were pooled into mean values for each variable (that is, each cluster as a whole and each of the individual speech sounds in the clusters) before further analysis of the waveform/spectrogram ratings. Because of the small group sizes, non-parametric statistics (Mann Whitney U, exact inference) were used to investigate differences between the groups of speakers with Parkinson’s disease and healthy speakers. Correlations between the acoustic measures and intelligibility for the speakers with Parkinson’s disease were calculated using Spearman’s rho. In addition to correlations for the separate acoustic variables, correlation between intelligibility and a pooled weighted deviation score for each speaker was also calculated. This was based on the magnitude of deviation from healthy speaker mean for each measure, where a measure deviating more than one standard deviation was scored 1, and a measure deviating more than two standard deviations was scored 2. The scores were totalled for each speaker.

**Reliability analysis**

Since rating of consonant cluster waveform diagrams/spectrograms was a new measure, inter-rater reliability was investigated in terms of both consistent ranking of ratings by calculating intraclass correlation coefficients (ICCs) and percentage agreement on the numerical values on the six-point scale (Gisev et al., 2013). Model, type and definition for calculating ICCs were selected according to Koo and Li (2016).

Inter-rater ICCs, based on a mean-rating, absolute agreement, two-way random-effects model, showed good reliability with ICC-values in the range of 0.70–0.93 (Table 3). Percentage agreement for the exact numerical scale values was low (speakers with PD: 16–36%; healthy speakers: 31–52%), while agreement ± 1 scale point was considerably higher (speakers with PD: 56–79%; healthy speakers: 64–89%; Table 4). For intra-rater reliability analysis, approximately 20% of the total material was judged twice by each rater. Intra-rater ICCs based on a single rater, absolute agreement, two-way mixed-effects model were in the range 0.64–0.97 (Table 3).

The measures for the other acoustic analyses were conducted by the first author. Reliability was explored through repeated measures intra-judge agreement of approximately 20% of the total material with at least 3 months between measurement occasions. ICCs based on a single rater, absolute agreement, two-way mixed-effects model were for syllable duration >0.97, and for VOT > 0.96.
Table 3. Inter- and intra-rater reliability for spectrogram ratings of consonant clusters: Intraclass correlations (ICC).

<table>
<thead>
<tr>
<th></th>
<th>Inter-rater reliability(^a)</th>
<th>Intra-rater reliability(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All speakers (n = 17)</td>
<td>Speakers with PD (n = 11)</td>
</tr>
<tr>
<td>(/spr^c)</td>
<td>0.880</td>
<td>0.842</td>
</tr>
<tr>
<td>(/s^d)</td>
<td>0.867</td>
<td>0.874</td>
</tr>
<tr>
<td>(/p^d)</td>
<td>0.915</td>
<td>0.932</td>
</tr>
<tr>
<td>(/t^d)</td>
<td>0.833</td>
<td>0.712</td>
</tr>
<tr>
<td>(/str^c)</td>
<td>0.830</td>
<td>0.814</td>
</tr>
<tr>
<td>(/s^d)</td>
<td>0.705</td>
<td>0.800</td>
</tr>
<tr>
<td>(/t^d)</td>
<td>0.850</td>
<td>0.848</td>
</tr>
<tr>
<td>(/r^d)</td>
<td>0.708</td>
<td>0.743</td>
</tr>
<tr>
<td>(/skr^c)</td>
<td>0.881</td>
<td>0.869</td>
</tr>
<tr>
<td>(/s^d)</td>
<td>0.774</td>
<td>0.723</td>
</tr>
<tr>
<td>(/k^d)</td>
<td>0.914</td>
<td>0.910</td>
</tr>
<tr>
<td>(/t^d)</td>
<td>0.875</td>
<td>0.849</td>
</tr>
</tbody>
</table>

\(^a\)Inter-rater reliability: Intraclass correlation coefficients (ICC) were calculated based on a mean-rating, absolute agreement, two-way random-effects model.

\(^b\)Intra-rater reliability: Intraclass correlation coefficients (ICC) were calculated based on a single rater, absolute agreement, two-way mixed-effects model.

\(^c\)Ratings of whole cluster on a six-point scale (0 = cluster looks very non-typical, none of the speech sounds are possible to differentiate from neighbouring sounds to 5 = cluster looks very typical, all speech sounds are very easily discernible with clear transitions between them)

\(^d\)Ratings of individual speech sounds on a six-point scale (0 = speech sound looks very non-typical, is impossible to differentiate from neighbouring sounds or is completely missing to 5 = speech sound looks very typical and can very easily be differentiated from neighbouring sounds)

Table 4. Percentage inter-judge agreement for spectrogram ratings of consonant clusters.

<table>
<thead>
<tr>
<th></th>
<th>Speakers with PD (n = 11)</th>
<th>Healthy speakers (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% agreement exact values</td>
<td>% agreement ± 1 scalepoint</td>
</tr>
<tr>
<td>(/spr^a)</td>
<td>28</td>
<td>68</td>
</tr>
<tr>
<td>(/s^d)</td>
<td>36</td>
<td>79</td>
</tr>
<tr>
<td>(/p^d)</td>
<td>36</td>
<td>73</td>
</tr>
<tr>
<td>(/t^d)</td>
<td>16</td>
<td>56</td>
</tr>
<tr>
<td>(/str^a)</td>
<td>32</td>
<td>70</td>
</tr>
<tr>
<td>(/s^d)</td>
<td>30</td>
<td>62</td>
</tr>
<tr>
<td>(/t^d)</td>
<td>36</td>
<td>64</td>
</tr>
<tr>
<td>(/t^d)</td>
<td>29</td>
<td>64</td>
</tr>
<tr>
<td>(/skr^a)</td>
<td>28</td>
<td>70</td>
</tr>
<tr>
<td>(/s^d)</td>
<td>38</td>
<td>75</td>
</tr>
<tr>
<td>(/k^d)</td>
<td>29</td>
<td>71</td>
</tr>
<tr>
<td>(/t^d)</td>
<td>26</td>
<td>58</td>
</tr>
</tbody>
</table>

\(^a\)Ratings of whole cluster on a six-point scale (0 = cluster looks very non-typical, none of the speech sounds are possible to differentiate from neighbouring sounds to 5 = cluster looks very typical, all speech sounds are very easily discernible with clear transitions between them)

\(^b\)Ratings of individual speech sounds on a six-point scale (0 = speech sound looks very non-typical, is impossible to differentiate from neighbouring sounds or is completely missing to 5 = speech sound looks very typical and can very easily be differentiated from neighbouring sounds)

Results

Results for speakers with Parkinson’s disease below are presented at both individual and group level, with results from a group of healthy speakers included for comparison. Correlations between the acoustic measures and intelligibility are presented in a separate section. A table showing detailed individual and group results can be found in the Supplementary material (Table S1).
Consonant cluster production

At group level, ratings of consonant clusters were lower for the speakers with Parkinson’s disease (PD) than for the healthy speakers (HS). For the clusters /spr/ and /skr/ this difference was statistically significant (/spr/: $M_{PD} = 2.69$, SD = 1.00; $M_{HS} = 3.86$, SD = 0.43, Mann Whitney U = 53.5; p = 0.037; /skr/: $M_{PD} = 2.75$, SD = 1.17; $M_{HS} = 3.81$, SD = 0.71, Mann Whitney U = 54.5, p = 0.048). An overview of individual and group results is presented in Figure 1.

For individual speech sounds in clusters, the ratings at group level were consistently lower for the speakers with Parkinson’s disease than for the healthy speakers but with considerable variation. The greatest group difference was for the liquid /t/. Similar to the global ratings, the group differences were greater for speech sounds in the heterorganic clusters /spr/ and /skr/ than for the homorganic cluster /str/. Details on individual speech sound ratings are presented in Figure 2.

Comments provided by raters

Two of the raters gave some comment to their ratings for most of the speakers. Overall, they were consistent with each other. The very few comments on the whole clusters mostly stated either that the cluster looked typical or that speech sounds were unidentifiable. Comments on the individual speech sounds in the clusters mainly concerned the stops (/p/, /t/, and /k/) or the liquid (/t/). For stops, commonly observed atypicalities were that either the closure or the release of the stop was not clear, or, in some cases, that the stop was completely absent. The /t/ was described as hard to identify or to differentiate from the following vowel. The fricative /s/ got comparatively fewer comments, of which the most common description of atypicality was a weak /s/.

![Figure 1](image-url)

**Figure 1.** Ratings of consonant /spr/, /str/, and /skr/ in waveform diagrams and spectrograms for individual speakers with Parkinson’s disease (Sp1-Sp11, grey bars) and for groups (speakers with Parkinson’s disease (PD) and healthy speakers (HS), unfilled bars).

Note. Global ratings of consonant clusters on six-point scale (0 = cluster looks very non-typical, none of the speech sounds are possible to differentiate from neighbouring sounds, to 5 = cluster looks very typical, all speech sounds are very easily discernible with clear transitions between them). Speech material: Words with initial consonant clusters extracted from oral sentence reading. Individual speakers (Sp) are numbered based on intelligibility percentage score, where Sp1 has the lowest score (52%) and Sp11 the highest (95%).

$* = p < 0.05$
Figure 2. Ratings of individual speech sounds in consonant clusters in waveform diagrams and spectrograms on six-point scale for groups of speakers with Parkinson’s disease (PD) and healthy speakers (HS). Note. Scale endpoints: 0 = speech sound looks very non-typical, speech sound is impossible to differentiate from neighbouring sounds or is completely missing, to 5 = speech sound looks very typical and can very easily be differentiated from neighbouring sounds.

* = $p < 0.05$
Alternating and sequential articulatory movements

Rate
At group level, alternating motion rate (AMR) was significantly slower for repeated lingual movements (syllables /ta/ and /ka/) for speakers with Parkinson’s disease compared to healthy speakers (/ta/: M_PD = 6.02 syll/sec, SD = 0.92; M_HS = 7.00 syll/sec, SD = 0.67, Mann Whitney U = 36.5; p = 0.027, and /ka/: M_PD = 5.68 syll/sec, SD = 0.88; M_HS = 6.51 syll/sec, SD = 0.53, Mann Whitney U = 54.5; p = 0.027, respectively). For labial movements (syllable /pa/) and sequential movements (SMR; /pa-ta-ka/) the group differences were smaller and not statistically significant. Individual and group results are presented in Figure 3.

Rate variability
AMR variability, measured as intersyllabic variation of duration (IVQ), was significantly greater for speakers with Parkinson’s disease than for healthy speakers in all three articulatory positions (/pa/: M_PD = 0.09, SD = 0.05; M_HS = 0.05, SD = 0.02, Mann Whitney U = 6, p = 0.007; /ta/: M_PD = 0.11, SD = 0.08; M_HS = 0.05, SD = 0.01, Mann Whitney U = 12, p = 0.037; /ka/: M_PD = 0.12, SD = 0.05; M_HS = 0.05, SD = 0.01, Mann Whitney U = 6, p = 0.005). Inter-speaker differences in AMR variability were greater for the speakers with Parkinson’s disease than for the healthy speakers, but the individual results show no clear relationship with the speakers’ levels of intelligibility. SMR variability was greater than AMR variability for both groups, and with no statistically significant group difference (Figure 4).

Voice onset time
VOT in alternating articulatory movements was significantly longer for speakers with Parkinson’s disease than for healthy speakers in all articulatory positions (/pa/: M_PD = 19.00, SD = 5.46; M_HS = 12.67, SD = 4.59, Mann Whitney U = 13.5, p = 0.048; /ta/: M_PD = 34.27, SD = 11.54; M_HS = 21.33, SD = 5.68, Mann Whitney U = 7.5, p = 0.007; /ka/:

Figure 3. Articular motion rate (syllables/sec) in syllable repetition for individual speakers with Parkinson’s disease (Sp1-Sp11, grey bars) and for groups (speakers with Parkinson’s disease (PD) and healthy speakers (HS), unfilled bars).
Note. Measure for alternating motion rate (AMR: /pa/, /ta/, /ka/) is the average of 10 and for sequential motion rate (SMR: /pa-ta-ka/) the average of 12 consecutive syllable repetitions.
Individual speakers (Sp) are numbered based on intelligibility percentage score, where Sp1 has the lowest score (52%) and Sp11 the highest (95%).
*p = p < 0.05
Correlations between acoustic measures and intelligibility

Correlations between the acoustic measures and intelligibility for speakers with Parkinson’s disease are displayed in Table 5.

The strongest correlations between intelligibility and acoustic measures involved consonant cluster production ($r_s = 0.706–0.797$, $p < 0.05$), where speakers with lower intelligibility tended to have lower cluster ratings than speakers with higher intelligibility. VOT variability for /pa/ showed a strong correlation with intelligibility ($r_s = -0.645$, $p < 0.05$), the negative correlation meaning that lower variability would be related to a higher intelligibility score. Correlations for measures related to syllable repetition rate and intersyllabic rate variation were in the weak to moderate range and statistically non-significant. The pooled weighted deviation score varied for the speakers with Parkinson’s disease between 2 and 25 (M = 15.09, SD = 6.79), and the correlation between this score and intelligibility was strong ($r_s = -0.749$, $p < 0.01$).
Figure 5. Voice onset time (ms) in syllable repetition for individual speakers with Parkinson’s disease (Sp1–Sp11, grey bars) and for groups (speakers with Parkinson’s disease (PD) and healthy speakers (HS), unfilled bars).

Note. Measures are mean and standard deviation of voice onset time in 10 consecutive syllables from AMR tasks (repetition of /pa/, /ta/, and /ka/). Average number of measurable syllables for speakers with PD/healthy speakers were 8.09/9.83 for /pa/, 8.09/9.67 for /ta/, and 6.45/10.00 for /ka/. Individual speakers (Sp) are numbered based on total pooled intelligibility percentage score, where Sp1 has the lowest score (52%) and Sp11 the highest (95%).

* = p < 0.05, ** = p < 0.01

Table 5. Correlations between acoustic measures and intelligibility for speakers with Parkinson’s disease (n = 11).

<table>
<thead>
<tr>
<th>Speech component</th>
<th>Variable</th>
<th>Correlation with intelligibilitya</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonant articulation (rating of whole cluster)</td>
<td>/spr/</td>
<td>0.721*</td>
</tr>
<tr>
<td>Alternating motion rate (AMR)</td>
<td>/pa/</td>
<td>0.338</td>
</tr>
<tr>
<td>Intersyllabic variation of alternating motion rate (AMRvar)</td>
<td>/pa/</td>
<td>−0.456</td>
</tr>
<tr>
<td>Sequential motion rate (SMR)</td>
<td>/pa-ta-ka/</td>
<td>0.402</td>
</tr>
<tr>
<td>Voice onset time (VOT)</td>
<td>/pa/</td>
<td>−0.509</td>
</tr>
<tr>
<td>Voice onset time variability (VOTvar)</td>
<td>/pa/</td>
<td>−0.668*</td>
</tr>
<tr>
<td>Weighted deviation scoreb</td>
<td>Pooled deviationb from healthy speakers for all variables</td>
<td>−0.749**</td>
</tr>
</tbody>
</table>

aSpearman rank order correlation, * = p < 0.05; ** = p < 0.01
b> 1 SD from healthy speaker mean = score 1; > 2 SD from healthy speaker mean = score 2
Discussion

In the present pilot study, we investigated aspects of consonant articulation acoustics in eleven Swedish speakers with Parkinson’s disease and six neurologically healthy persons. More specifically, the focus was on consonant cluster production, articulatory motion rate and variation, and VOT measurements, and how these acoustic features correlate with speech intelligibility.

According to Weismer et al. (2001), there may be at least two different ways to interpret the relationship between acoustic measures and intelligibility. One is that a specific acoustic variable reflects the overall level of impairment, and is just one affected part among others, which together affect intelligibility. The strong correlation between the pooled deviation score and intelligibility in the present study supports this suggestion. It is possible that a common underlying factor is reflected in all included measures, such as articulatory timing and/or precision, which would be expected considering the underlying sensory-motor impairment in Parkinson’s disease (Read et al., 2018).

Another interpretation suggested by Weismer et al. (2001) is that a specific acoustic variable directly influences intelligibility. The strong correlations between the consonant cluster ratings and intelligibility stand out and raise the question whether consonant cluster articulation has a particularly strong influence on intelligibility. For languages rich in consonant clusters, this could be the case. An alternative explanation could be that the consonant cluster rating measure is more sensitive to specific signs of underlying impairment that affect intelligibility, than the other measures employed in the present study. It has been suggested that more complex speech tasks, which require a higher degree of articulatory precision (such as consonant clusters), may be better for observing speech motor control than simpler speech tasks (such as CV-syllable repetition; Kuruvilla-Dugdale et al., 2020). The consonant cluster rating also differs from the time-based measures of CV-syllables in that it entails a qualitative component and comparison with a perceived standard of typicality. Further research is needed to clarify the relevance of these suggested alternative explanations.

Especially for ratings of the heterorganic consonant clusters (/spr/ and /skr/), distinct differences between the groups were observed, while the ratings for speakers with Parkinson’s disease and the healthy speakers were more similar for the homorganic cluster /str/. Differences in articulatory demands for the different types of clusters could possibly explain the divergent effects on speech production for the two groups of speakers. Homorganic clusters produced with the same articulator may be demanding because of more overlap between motor action features, but on the other hand, heterorganic clusters could be more complex, as they require coordination of different articulators (Huinck et al., 2004), and the latter may be comparatively more challenging for speakers with Parkinson’s disease than for healthy speakers.

Regarding alternating articulatory motion rate, results from previous research are ambiguous. For lingual movements, group results from the present study support observations that speakers with Parkinson’s disease tend to have slower alternating articulatory motion rate than healthy speakers (e.g. Chenausky et al., 2011; Lowit et al., 2018; Rusz, Tykalova, Ramig et al., 2021). At group level, the speakers with Parkinson’s disease in the present study also showed a significantly more variable intersyllabic duration, which corresponds to some previous studies (e.g. Chenausky et al., 2011 Lowit et al., 2018), but not to others (e.g. Harel et al., 2004; Rusz, Tykalova, Ramig et al., 2021). This could be due to effects of differences in the samples of
speakers. For example, Harel et al. (2004) studied persons at a very early stage of the disease, and Chenausky et al. (2011) studied persons with deep brain stimulation (DBS), but no medication. It is also possible that divergent results from different studies could be related to medication effects, although consonant articulation seems to be less responsive to Levodopa, possibly because it relies on non-dopaminergic neural pathways (Rusz, Tykalova, Novotny, Ruzicka et al., 2021; Rusz, Tykalova et al., 2021). In the present study, the procedure regarding medication was handled similarly to the study by Lowit et al. (2018), that is, the speech material was recorded at a time of the day when the participants, by their own experience, usually were functioning at their best. This approach was chosen to obtain speech samples which were valid examples of how the speakers’ speech motor control may function in their everyday life.

The significantly longer VOT for the group of speakers with Parkinson’s disease in the present study is in line with most previous research (e.g., Forrest et al., 1989; Harel et al., 2004; Weismer et al., 2001). Compared to studies including Swedish speakers (Helgason & Ringen, 2008; Lundeborg et al., 2012), VOT in the present study was overall shorter. This could be explained by the different speech tasks employed (word initial stops elicited by picture naming or oral reading vs. syllable repetition at maximum rate). The impact of place of articulation on VOT noted in previous research (Cho & Ladefoged, 1999; Helgason & Ringen, 2008; Lisker & Abramson, 1964; Lundeborg et al., 2012) was observed also in our data.

An overall tendency for the individual results was that the acoustic measures from the speakers with the lowest intelligibility scores tended to deviate more from the healthy speakers, but it is noteworthy that there were measures in the present study indicating that some aspects of articulation appear to be equally demanding and varying in terms of acoustic realisation for elderly healthy speakers and speakers with Parkinson’s disease, such as ratings of the homorganic consonant cluster /str/ and sequential motion rate measures.

The results from this pilot study motivate further research on waveform/spectrogram examination of speech production as a possible assessment tool in dysarthria. This study was limited to investigate scaled ratings of visually examined consonant cluster typicality in waveform diagrams and spectrograms and how the ratings relate to intelligibility. Besides further development of the rating procedure, correlations with auditory-perceptual ratings should be explored in larger samples including speakers with a wider range of speech impairment and SLT clinicians as raters. Another suggested next step would be to further investigate the contribution of specific qualitative features to judgements of atypicality and the relationship of these features with intelligibility, since the scale values in themselves do not reveal which features the raters have judged as atypical. To keep the burden on the raters within reasonable limits, comments on qualitative judgements were not mandatory in the present study, and only two of the raters provided regular comments. Another area for further research is the addition of visual feedback on articulatory features to speakers with Parkinson’s disease by waveform diagrams and spectrograms. Because of possible disease-related sensory and cognitive deficits, visualising details of speech output could be beneficial as additional feedback for speakers with Parkinson’s disease (Yunusova et al., 2017). Spectrogram analysis and feedback on articulatory features has been used in the field of second language learning (Lambacher, 1999), but we have not been able to find studies on this from the field of dysarthria.
Methodological considerations

One limitation of the present pilot study is the small sample of speakers. Because of the known individual variability among speakers with Parkinson’s disease, the sample size in this study was chosen to allow for analysis on an individual level in addition to group results. The speakers with Parkinson’s disease also had a predominantly mild to moderate speech impairment, which together with the low number of speakers limits generalisation of the results.

The method for ratings of spectrograms for investigating consonant clusters was developed for this study and has not been previously reported elsewhere. As regards the reliability of the chosen method, the percentage agreement on exact scale values was low, particularly for the speakers with Parkinson’s disease. However, the substantially higher agreement on ± 1 scale point indicates that different raters are relatively close in their judgements. The high ICC values also indicate a high level of rater consistency. It might be that a 6-point scale, as used here, gives too much room for different perceptions of location of scale points relative to the perceived level of consonant production typicality. A scale with fewer points would probably increase inter- and intra-rater reliability and could still be sufficient for clinical purposes.

The choice of using English-speaking clinical linguists as raters could be considered as a weakness of the study. However, when trying out the procedure before data collection, we did not observe any confounding effects of using English-speaking raters. The use of multiple raters also allowed us to control for both inter- and intra-rater reliability for this novel measure. The qualitative component of this measure puts demands on rater skills. In the present study, the raters had expertise in the area, but future research on the usefulness of the measure for clinical purposes should include other groups of raters, such as SLT clinicians working with dysarthria.

The other acoustic analyses were carried out by one person (first author), which could also be seen as a weakness. We chose to use measures which were well-established, well-defined and had been employed in previous research on dysarthria with reports of high inter-rater reliability (e.g. Fischer & Goberman, 2010; Flint et al., 1992; Wang et al., 2004). In the present study, repeated measures intra-judge agreement was reasonably strong, indicating consistent measuring.

Conclusions

The results of this pilot study suggest that some acoustic measures may be more sensitive than others to indicate impact on speech production and intelligibility in speakers with Parkinson’s disease. Among the measures in the present study, spectrogram ratings of typicality in the heterorganic consonant clusters /spr/ and /skr/ had the strongest correlations with intelligibility, and also differentiated speakers with Parkinson’s disease from healthy speakers. However, this measure and the scaling procedure need further development and should be investigated also in larger samples. Measures based on syllable repetition, such as AMR, SMR and VOT, showed more varying results with statistically non-significant weak to moderate correlations with intelligibility. A clinical implication of the observation that some aspects of articulation appear to be equally demanding and varying in terms of acoustic realisation for elderly healthy speakers and speakers with Parkinson’s
disease could be to choose measures where there is less variation among healthy speakers, when the purpose is to detect signs of disordered speech motor control.

**Acknowledgments**

We would like to thank all the participating speakers and raters.

**Disclosure statement**

The authors report no conflict of interest.

**Funding**

This study was supported by the Foundation for Parkinson Research at Linköping University under Grant [LiU 2015-00194], Region Östergötland under Grant [LiO 620581], and the Swedish Parkinson Foundation under Grant [1076/18].

**ORCID**

Inga-Lena Johansson [http://orcid.org/0000-0002-4782-2373](http://orcid.org/0000-0002-4782-2373)
Christina Samuelsson [http://orcid.org/0000-0003-0358-3048](http://orcid.org/0000-0003-0358-3048)
Nicole Müller [http://orcid.org/0000-0002-4405-5340](http://orcid.org/0000-0002-4405-5340)

**References**


