npj | parkinson's disease

www.nature.com/npjparkd

ARTICLE OPEN

Check for updates

Articulatory undershoot of vowels in isolated REM sleep behavior disorder and early Parkinson's disease

Dominik Skrabal[1], Jan Rusz[1,2,3], Michal Novotny[2], Karel Sonka[1], Evzen Ruzicka[1], Petr Dusek[1] and Tereza Tykalova[2]✉

Imprecise vowels represent a common deficit associated with hypokinetic dysarthria resulting from a reduced articulatory range of motion in Parkinson's disease (PD). It is not yet unknown whether the vowel articulation impairment is already evident in the prodromal stages of synucleinopathy. We aimed to assess whether vowel articulation abnormalities are present in isolated rapid eye movement sleep behaviour disorder (iRBD) and early-stage PD. A total of 180 male participants, including 60 iRBD, 60 de-novo PD and 60 age-matched healthy controls performed reading of a standardized passage. The first and second formant frequencies of the corner vowels /a/, /i/, and /u/ extracted from predefined words, were utilized to construct articulatory-acoustic measures of Vowel Space Area (VSA) and Vowel Articulation Index (VAI). Compared to controls, VSA was smaller in both iRBD ($p = 0.01$) and PD ($p = 0.001$) while VAI was lower only in PD ($p = 0.002$). iRBD subgroup with abnormal olfactory function had smaller VSA compared to iRBD subgroup with preserved olfactory function ($p = 0.02$). In PD patients, the extent of bradykinesia and rigidity correlated with VSA ($r = -0.33$, $p = 0.01$), while no correlation between axial gait symptoms or tremor and vowel articulation was detected. Vowel articulation impairment represents an early prodromal symptom in the disease process of synucleinopathy. Acoustic assessment of vowel articulation may provide a surrogate marker of synucleinopathy in scenarios where a single robust feature to monitor the dysarthria progression is needed.

npj Parkinson's Disease (2022)8:137 ; https://doi.org/10.1038/s41531-022-00407-7

INTRODUCTION

Isolated rapid eye movement sleep disorder (iRBD) is a parasomnia characterized by dream-enactment behavior and loss of physiologic muscle atonia during the rapid eye movement sleep phase. iRBD is considered a prodromal stage of neurodegeneration as more than 80% of diagnosed patients developed alfa-synuclein-aggregation disorders such as Parkinson's disease (PD), Lewy body dementia, or multiple system atrophy[1,2]. Considering the development of Parkinson's disease-modifying treatment[3], a multicentre study including 1280 iRBD patients identified quantitative fine motor skill testing as the strongest predictor for conversion[4]. Another study by Postuma et al. [5] revealed that voice and face akinesia represent the earliest prodromal motor manifestations in iRBD subjects preceding the onset of parkinsonism by a mean 9.8 years. This is likely a consequence of motor speech complexity and its sensitiveness to neural damage[6].

Hypokinetic dysarthria of PD, which is mainly characterised by articulatory, phonatory and prosodic alterations, occurs in up to 90% of patients over the course of the disease[6,7]. Moreover, speech impairment is present in a majority of newly diagnosed PD patients[8,9]. Considering that patients with iRBD are at high risk of developing PD, the speech behavior assessment in iRBD is subjected to thorough investigation. Recent multilanguage research based on fully automated analysis of seven distinctive speech dimensions of hypokinetic dysarthria[10], including harsh voice, slow sequential motion rates, imprecise consonants, monoloudness, monopitch, prolonged pauses, and articulation rate performed on 150 iRBD patients revealed that only monopitch was able to significantly differentiate iRBD patients from controls[11]. Interestingly, monopitch was found in iRBD subjects with impaired olfactory function before the nigrostriatal dopaminergic transmission is affected[12], that is, in Braak stage 2 before the substantia nigra is affected by synucleinopathy[13]. Among monopitch, vowel articulation impairment represents one of the core deficits contributing to dysarthric speech, as it reflects the range of articulatory movements and strongly correlates with overall intelligibility[14–16]. The potential of imprecise vowel articulation to serve as an early biomarker can also be supported by a previous pilot study where deficits in vowel articulation were detected in a small sample of 20 patients with de novo PD[17]. However, potential changes of vowel articulation in iRBD have never been investigated. Also, no previous research independently related articulation impairment to other essential prodromal features of synucleinopathy, such as olfactory dysfunction.

The purpose of this study is to investigate vowel articulation in iRBD and early-stage PD patients compared to healthy controls in order (i) to verify the prospect of using measurement of vowel articulation as a biomarker for the detection of prodromal PD and (ii) to investigate the relationship between articulation measures and the degree of motor and smell dysfunction.

RESULTS

Group differences

Normative values of the first two formants for iRBD, PD and healthy control (HC) groups are summarized in Table 1. Vowel Space Area (VSA) was found to be the best parameter for differentiating between groups [$F(2,177) = 7.4$, $p = 0.001$, $\eta^2 = 0.08$] (Fig. 1). Post hoc comparisons revealed significantly smaller VSA in both iRBD ($p = 0.01$) and PD ($p = 0.001$) compared to HC individuals. In addition, group differences were also

[1]Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic. [2]Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic. [3]Department of Neurology & ARTORG Center, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. ✉email: tykalova.tereza@gmail.com

Table 1. Normative values of formant frequencies in male groups of iRBD, PD and HC.

	iRBD mean/SD (range)	PD mean/SD (range)	HC mean/SD (range)
F1/a/ (Hz)	548/43 (448−627)	548/53 (452−676)	582/49 (497−667)
F2/a/ (Hz)	1331/78 (448−627)	1328/81 (1172−1567)	1350/86 (1119−1535)
F1/i/ (Hz)	331/16 (300−376)	332/18 (293−380)	348/23 (304−408)
F2/i/ (Hz)	1980/108 (1680−2179)	1931/128 (1703−2190)	2012/103 (1750−2226)
F1/u/ (Hz)	331/19 (294−377)	337/20 (295/401)	340/20 (299−405)
F2/u/ (Hz)	770/64 (661−957)	786/52 (680−894)	762.9/54 (666−903)

iRBD, isolated rapid eye movement sleep behaviour disorder, *PD* Parkinson's disease, *HC* Healthy control, *F1* First formant frequency, *F2* Second formant frequency

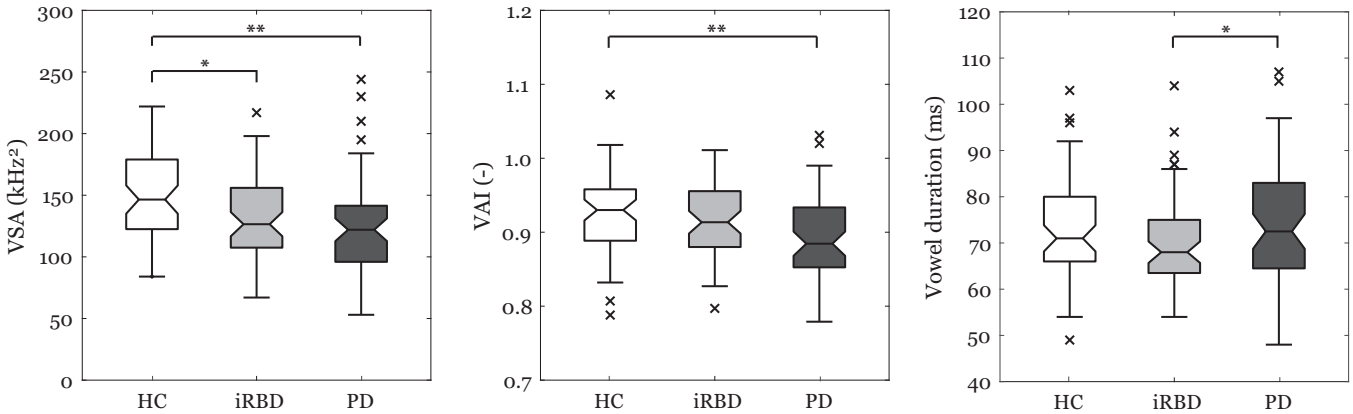


Fig. 1 Comparison of vowel measurements including VSA, VAI and vowel duration between HC, iRBD and PD using boxplots. The centre line indicates the median, and the bounds of the box indicate the 25th and 75th percentiles. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the 'x' symbol. VSA, vowel space area, VAI vowel articulation index, "Asterisks" indicate significant differences after Bonferroni correction: *$p < 0.05$; **$p < 0.01$.

detected for Vowel Articulation Index (VAI) [$F(2,177) = 6.3$, $p = 0.002$, $\eta 2 = 0.07$], as the PD group manifested significantly smaller VAI ($p = 0.002$) compared to HC group. Slight vowel duration differentiation was also observed across groups [$F(2,177) = 3.2$, $p = 0.04$, $\eta^2 = 0.04$), associated with differences between PD and iRBD groups ($p = 0.04$). The sub-experiment concerning olfactory function in iRBD showed that iRBD group with preserved olfactory function (iRBD-POF) had greater VSA than iRBD group with abnormal olfactory function (iRBD-AOF) [$F(1,54) = 5.4$, $p = 0.024$, $\eta^2 = 0.094$] (Fig. 2a). In addition, iRBD-AOF with normal dopamine transporter single-photon emission computed tomography (DAT-SPECT) showed greater VSA than iRBD-AOF with abnormal DAT-SPECT [$F(1,31) = 4.2$, $p = 0.049$, $\eta^2 = 0.140$] (Fig. 2b). No significant differences for VAI and vowel duration were found.

Correlations between speech and motor variables

Movement Disorder Society Unified Parkinson's Disease Rating Scale motor part (MDS-UPDRS III) total in PD patients showed negative correlation with VSA ($r = -0.29$, $p = 0.03$) and VAI ($r = -0.29$, $p = 0.03$). In addition, bradykinesia and rigidity subscore in PD patients showed negative correlation with VSA ($r = -0.33$, $p = 0.01$) and VAI ($r = -0.34$, $p < 0.01$) while neither correlation between postural instability and gait difficulty (PIGD) subscore and VSA ($r = -0.04$, $p = 0.75$) or VAI ($r = -0.12$, $p = 0.75$) nor between tremor subscore and VSA ($r = 0.01$, $p = 0.96$) or VAI ($r = 0.06$, $p = 0.64$) was detected (Fig. 3).

Regarding brain imaging, the putamen binding ratio in iRBD showed positive correlation with VSA ($r = 0.35$, $p = 0.01$). No other significant correlations were found between vowel articulation parameters and clinical scales in PD or iRBD.

DISCUSSION

Results of our study revealed that subtle impairment in vowel articulation due to the reduced articulatory range of motion is already evident in prodromal synucleinopathy. Articulatory impairment in iRBD was detectable through objective acoustic analysis despite almost no perceptual dysarthria severity was noted during clinical examination. Since the extent of articulatory undershoot was related to bradykinesia and rigidity in our PD cohort, we may hypothesize that vowel articulation abnormalities in parkinsonism result mainly as a consequence of nigrostriatal degeneration. The strength of this study is that vowel articulation features were evaluated in a large sample of iRBD and de-novo PD patients. Examining drug-naïve patients is especially important as dopaminergic treatment may improve certain aspects of speech disorder such as vowel articulation[18]. Since the vowel articulation performance of iRBD subjects intermediated between controls and de-novo PD patients and since it was more severe in iRBD with hyposmia compared to iRBD with preserved smell function, acoustic assessment of vowels could be potentially useful as a diagnostic and prognostic biomarker in α-synuclein-aggregation disorders.

Our findings of lowered vowel space in iRBD demonstrate that vowel articulation is already altered in the prodromal stages of synucleinopathy. In particular, vowel articulation was more impaired in iRBD subgroup with severe hyposmia, which is one of the most common and earliest non-motor prodromal features

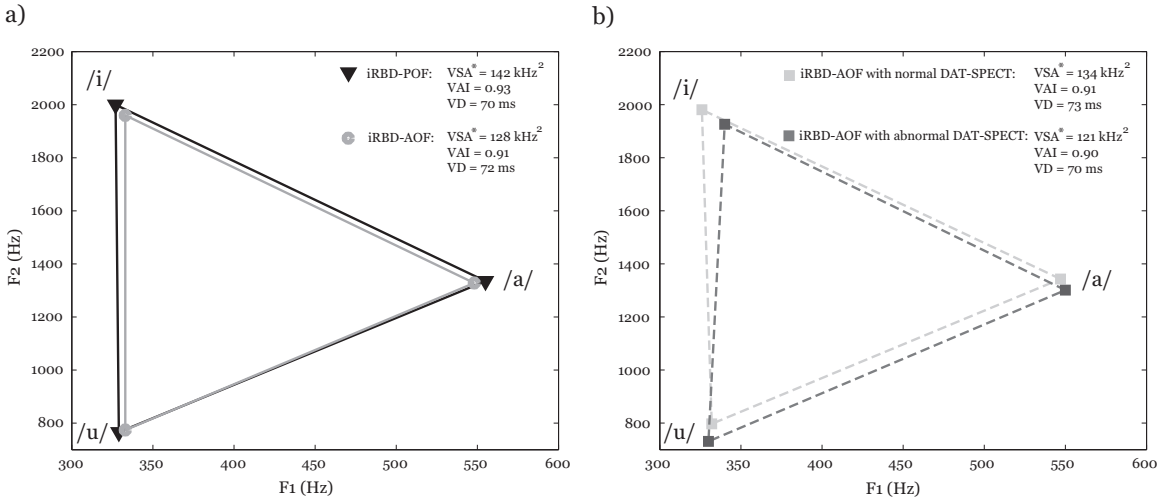


Fig. 2 Mean F1 and F2 values and vowel space area for a) iRBD-POF compared iRBD-AOF subgroups and b) iRBD-AOF with normal DAT-SPECT compared to iRBD-AOF with abnormal DAT-SPECT. VSA vowel space area, VAI vowel articulation index, VD vowel duration, iRBD-POF isolated rapid eye movement sleep behaviour disorder patients with preserved olfactory function, iRBD-AOF isolated rapid eye movement sleep behaviour disorder patients with abnormal olfactory function, DAT-SPECT dopamine transporter single-photon emission computed tomography; "Asterisks" indicate significant differences: *$p < 0.05$.

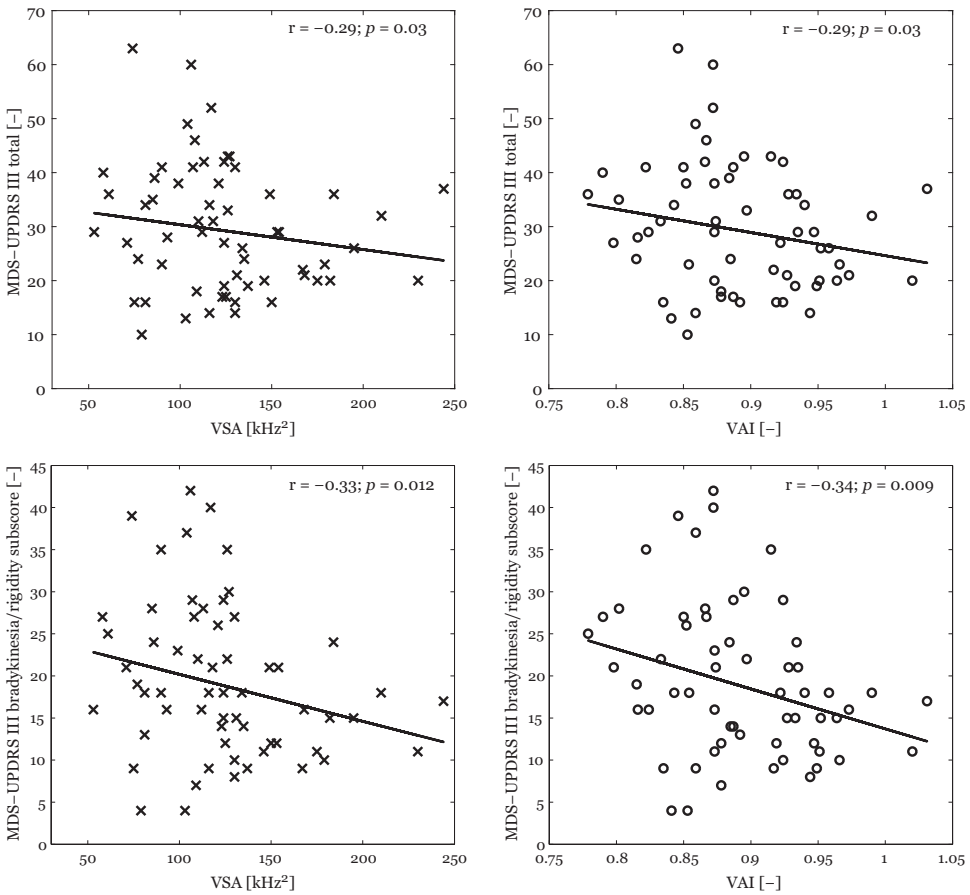


Fig. 3 Significant correlations between clinical motor scales and acoustic data plotted to 2D space with the trend of averaged data (black line) in the PD group. VSA Vowel space area, VAI Vowel articulation index, MDS–UPDRS III Movement Disorder Society – Unified Parkinson's Disease Rating Scale part III.

to emerge in PD[2,4]. Thus, we might assume that vowel articulation impairment is an early prodromal symptom that progresses along with olfactory dysfunction in the disease process of synucleinopathy. Accordingly, recent research demonstrated that dysprosody is already present in iRBD subjects with impaired olfactory function but still intact nigrostriatal pathway[12]. Together, these findings might indicate that speech production is already slightly affected in Braak's stage 2, which is associated with Lewy

pathology within brainstem nuclei[13], a brain region crucial for controlling vocal fold tension[19].

Furthermore, our findings in iRBD subgroup with impaired olfactory function implies that articulatory undershoot is also result of nigrostriatal neurodegeneration, as greater vowel deficits were found in the subgroup of patients with abnormal compared to those with normal DAT-SPECT. This assumption can be further supported by the observed link between the extent of vowel articulation decline and bradykinesia and rigidity but not axial gait symptoms in our de-novo PD group. Also, the previous pilot study discovered a positive correlation between amelioration of vowel articulation and dopaminergic treatment-related improvement in bradykinesia and rigidity[18]. On the other hand, degeneration in non-dopaminergic brain regions may further contribute to the worsening of vowel articulation performance as PD progresses. This is in agreement with a previous longitudinal study reporting a further decline of vowel articulation in the course of the disease in PD patients with an average disease duration of 6 years after the diagnosis[20]. Indeed, it is well known that vowel articulation impairments strongly correlate with overall intelligibility[14–16], which tends to be affected in the later stages of PD. A recent study showed that articulatory deficits in de-novo PD patients are indicative of more widespread brain damage affecting extranigral cortical or subcortical regions[8]. In accordance with this assumption, a former study reported that aggravation of dysarthria during PD progression results particularly from the increasing severity of cerebral non-dopaminergic lesions[21]. Finally, in PD patients treated with bilateral subthalamic nucleus deep brain stimulation, the severity of the residual parkinsonian speech score was predictive of a poor postoperative outcome, likely due to the presence of non-dopaminergic lesions within the brain[22]. However, we cannot exclude that vowel articulation impairment in advanced PD may be also related at least in part to levodopa-induced dyskinesia[23], together with the neurodegeneration of dopaminergic and non-dopaminergic brain areas. In summary, given the existing evidence in literature, we may hypothesize that vowel articulation deficits, along with limb bradykinesia, are primarily related to dopaminergic involvement in the early stages, whereas nondopaminergic lesions further contribute to the worsening of vowel articulation in the later stages of PD. Therefore, acoustic assessment of vowel articulation may provide a surrogate marker of neurodegeneration from prodromal to more advanced synucleinopathy for scenarios where a single robust feature to monitor the dysarthria progression is desired.

In our study, both VSA and VAI reflected speech impairment in drug-naive PD patients, which is in accordance with a previous pilot study investigating a small sample of 20 de-novo PD patients[17]. Contrary to our results on de-novo PD patients, several authors suggested VAI is superior over VSA in moderate to advanced stages of PD[20,24]. The principle of VAI construction focuses on formant centralization in order to minimize the effect of interspeaker variability; as a consequence, VAI may not reflect subtle articulatory spatial modifications[25]. On the other hand, VSA is calculated out of the maximal extent of vowel working space, which might better mirror subtle speech changes[18,26]. In other words, we may hypothesize that PD vowel articulatory impairment is tongue-dominant and emerge initially in the posterior parts of an articulatory organ[26,27]. Since the individual vowel /u/, which is characterised with tongue positioning high and backward[28], seem to be the most affected vowel in early stages of PD[17], VSA construction shall significantly acknowledge single-vowel frequency shift, whereas the sensitivity of VAI might be lowered as it accounts for centralization of all corner vowel that does not need to be affected in the early disease process.

Previous studies assumed that speaking rate may influence vowel articulation performance in dysarthrias[29]. However, we did not find differences for vowel duration at the group level between HC and both patient cohorts, likely as a consequence of the very early stages of synucleinopathy investigated. Despite inconclusive results regarding articulation rate in later stages of synucleinopathies, including no changes in the speech rate, decreased speech rate, or even an accelerated speech rate[30–32], our findings are in agreement with a recent study showing no speech rate changes in de-novo PD and only a trend toward slower speech rate in iRBD[33]. Faster speech observed in advanced PD is likely to reflect a physiological tendency to accelerate speech due to the impaired motor planning (oral festination)[30,31] whereas the tendency toward a slower speech rate may be theoretically attributed to the degeneration of non-dopaminergic pathways[8].

This study has certain limitations. We enrolled exclusively male participants primarily because of the strong predominance of male subjects within iRBD patients[34]. Anatomical dispositions of the vocal tract reflect sexual dimorphism[35,36], and we therefore cannot exclude a potential sex effect on our results. On the other hand, no sex-specific speech dysfunction in de-novo PD was found[9]. The findings of current study are based on a cross-sectional design. Future research is needed to estimate the sensitivity of vowel articulation features in long-term follow-up and in predicting phenoconversion from iRBD to established parkinsonism.

In conclusion, the vowel articulation deficits in male subjects significantly differentiated both iRBD and de-novo PD patients from controls. Future studies should elaborate our findings in the female population, and vowel articulation analysis might then have the potential to serve as a speech biomarker indicating the prodromal stage of synucleinopathy. Introducing novel acoustic biomarkers enable to design a speech assessment battery allowing for valid and easy evaluation of a large number of subjects in clinical trials as speech assessment is inexpensive, noninvasive and digitally storable. Additionally, vowel articulation analysis may facilitate better speech phenotype classification of PD and has, thus, the potential utility for personalized medicine.

METHODS

Participants

From 2015 to 2021, a total of 180 male Czech participants, including a separate group of iRBD, PD, and HC participants, were recruited. Each participant provided written informed consent. The study received approval from the Ethics Committee of the General University Hospital in Prague, Czech Republic, and has been performed in accordance with the ethical standards established in the 1964 Declaration of Helsinki.

The iRBD group consisted of 60 male patients aged 65.6 (SD 7.1) years diagnosed according to the third edition of the International Classification of Sleep Disorders[37]. None of the iRBD patients suffered a significant communication disorder or had a history of treatment with antiparkinsonian medication, nor were they taking any medication affecting sleep, cognition or motor function. The PD group consisted of 60 untreated drug-naïve male patients aged 61.8 (SD 11.6) years, fulfilling the Movement Disorder Society clinical diagnostic criteria for PD[38]. At the time of examination, none of the PD patients had a history of communication disorder unrelated to PD, nor were they taking any medication affecting sleep or cognition function.

Both iRBD and PD subjects were examined by a movement disorder specialist (P.D.) using the MDS-UPDRS III[39]. Symptom duration was assigned based on the self-reported occurrence of the first motor symptoms. To define specific movement disorder manifestations that may influence speech disorder, three subscores from the MDS-UPDRS III scale were calculated as follows: (i) PIGD subscore (MDS-UPDRS part III, 3.9 Arising from the chair, 3.10 Gait, 3.11 Freezing of gait, 3.12 Postural stability, 3.13 Posture), (ii) Bradykinesia and Rigidity subscore (MDS-UPDRS part III, 3.3 Rigidity, 3.4 Finger tapping, 3.5 Hand movements, 3.6 Pronation-

Table 2. Patient clinical and demographic characteristics.

	PD, $n = 60$ Mean/SD (range) or n (%)	iRBD, $n = 60$ Mean/SD (range) or n (%)	iRBD-POF, $n = 23$ Mean/SD (range) or n (%)	iRBD-AOF, $n = 32$ Mean/SD (range) or n (%)
Clinical characteristics				
Male gender	60 (100%)	60 (100%)	23 (100%)	32 (100%)
Age (years)	61.8/11.6 (34−81)	65.6/7.1 (46−81)	63.7/8.5 (46−77)	67.6/5.7 (57−81)
Symptom duration (years)	1.1/1.5 (0.3−5.9)	6.5/5.5 (1−28)	5.8/4.0 (1−13)	8.5/8.4 (1−39)
MoCA	25.5/3.1 (17−30)	24.1/2.7 (18−30)	24.5/2.5 (20−29)	23.8/3.0 (18−30)
SCOPA-AUT	8.8/5.2 (0-23)	12.4/7.6 (1-33)	15.3/9.4 (1-33)	11.0/5.6 (2-23)
UPSIT[a]	20.9/7.5 (2-35)	22.1/7.9 (9-37)	30.2/3.4 (26-37)	16.3/4.5 (9-24)
MDS-UPDRS III total	22.5/11.8 (10−63)	6.8/6.2 (0−25)	6.0/5.9 (0−22)	7.3/6.4 (0−25)
MDS-UPDRS III speech item	0.5/0.5 (0−2)	0.1/0.3 (0−1)	0.0/0.0 (0−0)	0.1/0.3 (0−1)
MDS-UPDRS III tremor subscore	6.3/4.0 (0−15)	2.3/2.5 (0−11)	2.3/3.0 (0−11)	2.1/2.4 (0−9)
MDS-UPDRS III PIGD subscore	2.5/1.5 (0−6)	0.7/0.9 (0−4)	0.5/0.7 (0−2)	0.9/1.0 (0−4)
MDS-UPDRS III bradykinesia and rigidity subscore	12.5/9.1 (4−42)	0.6/0.8 (0−3)	2.7/3.1 (0−11)	3.7/4.4 (0−20)
iRBD presence	17 (28%)	60 (100%)	23 (100%)	32 (100%)
Brain imaging (DAT-SPECT)				
Caudate binding ratio	2.8/0.6 (1.3−3.8)	3.6/0.7 (2.3−5.4)	3.8/0.7 (2.5−5.3)	3.4/0.6 (2.3−5.4)
Putamen binding ratio	1.5/0.4 (0.8−2.5)	2.8/0.7 (1.3−4.5)	3.1/0.6 (1.9−4.1)	2.6/0.7 (1.3−4.5)
Abnormal DAT-SPECT	60 (100%)	16 (27%)	3 (13%)	12 (38%)

PD Parkinson's disease, *iRBD* isolated rapid eye movement sleep behaviour disorder, *iRBD-POF* iRBD patients with preserved olfactory function, *iRBD-AOF* iRBD patients with abnormal olfactory function, *MDS-UPDRS* Movement Disorder Society Unified Parkinson disease rating scale, *PIGD* Postural instability and gait difficulty, *MoCA* Montreal cognitive assessment, *SCOPA-AUT* Scales for Outcomes in Parkinson's Disease - Autonomic Dysfunctionm, *UPSIT* University of Pennsylvania Smell Identification Test, *DAT-SPECT* dopamine transporter single-photon emission computed tomography, [a]UPSIT was not available in five subjects

supination movements of hands, 3.7 Toe-tapping, 3.8 Leg agility, 3.14 Body bradykinesia) and Tremor subscore (MDS-UPDRS part III, 3.15 Postural tremor of the hands, 3.16 Kinetic tremor of the hands, 3.17 Rest tremor amplitude, 3.18 Constancy of rest tremor). Item 3.1 Speech of the MDS-UPDRS III was used for the perceptual description of overall dysarthria severity. In addition, patients were examined using the University of Pennsylvania Smell Identification Test (UPSIT)[40], Montreal Cognitive Assessment and Scales for Outcomes in Parkinson's Disease - Autonomic Dysfunction[41]. iRBD patients were categorized into two groups based on their olfactory function: iRBD group with preserved olfactory function (iRBD-POF) consisted of patients with UPSIT > 25 (i.e. normosmic to moderately microsmic) and the iRBD group with abnormal olfactory function (iRBD-AOF) of patients with UPSIT ≤ 25 (i.e. severely microsmic or anosmic). Patient clinical and demographic characteristics are summarised in Table 2.

The HC group consisted of 60 male volunteers of comparable aged 64.1 (SD 12.8) years, with no history of significant neurological or communication disorder. No group differences in age distribution were revealed between iRBD, PD and HC groups (ANOVA, $p = 0.15$).

Speech examination

The audio data were recorded in a quiet room with a low level of ambient noise using a head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed approximately 5 cm from the subject's lips. The speech signals were sampled at 48 kHz with 16–bit resolution. The recordings were obtained during one session with a speech specialist (D.S., J.R., M.N. or T.T.) who conveyed instructions to the participants. Each participant completed a series of speaking tasks, including reading twice a standardised passage composed of 80 words as part of a longer protocol lasting about 20 minutes. A reading passage (Fig. 4), written by famous Czech writer Karel Capek, was chosen as it is a standardised speaking task representing a natural condition of connected speech that was further shown to possess the highest accuracy for discriminating iRBD from HC[11]. The second realization of the reading passage was utilized to minimize effect of patient's natural ability to read aloud[42]. Neither signs of fatigue nor any changes in the quality of voice from the beginning to the end of the session were observed in any participant.

Když člověk poprvé vsadí do země sazeničku, chodí se na ni
dívat třikrát denně: tak co, povyrostla už nebo ne? I tají dech,
naklání se nad ní, přitlačí trochu půdu u jejích kořínků,
načechrává jí lístky a vůbec ji obtěžuje různým konáním, které
považuje za užitečnou péči. A když se sazenička přesto ujme a
roste jako z vody, tu člověk žasne nad tímto divem přírody,
má pocit čehosi jako zázraku a považuje to za jeden ze svých
největších osobních úspěchů.

Fig. 4 The text of the reading passage. Ten occurrences of each corner vowel used for the acoustic analysis are depicted by red (vowel /a/), green (vowel /i/) and blue (vowel /u/) colour.

Speech analyses

Acoustic measures were performed using the widely used speech-analysis software PRAAT (available at www.praat.org). Using PRAAT, both the combined wideband spectrographic display and the power spectral density were used to determine the first (F1) and the second (F2) formant frequencies in Hz. A total of 30 vowels per passage were studied, including 10 occurrences of /a/, 10 occurrences of /i/, and 10 occurrences of /u/ (Fig. 4). The

formant frequencies of vowels /a/, /i/, and /u/ were extracted from a 30-ms segment at the temporal midpoint of the stable part of each vowel (to avoid the influence of vowels preceding or following); this method has been previously validated[43] and proved to be reliable when applied to different languages[17,44,45]. The values of F1 and F2 frequencies were separately averaged for the individual corner vowel of each participant. The measurement of VSA and VAI were used. VSA is traditional and probably the most used articulatory-acoustic measure[46,47], and can be easily calculated using the following formula:[48]

$$VSA = 0.5 \times \left(\left[F2_{/u/} + F2_{/i/} \right] \times \left[F1_{/u/} - F1_{/i/} \right] - \left[F2_{/a/} + F2_{/u/} \right] \times \left[F1_{/a/} - F1_{/u/} \right] - \left[F2_{/a/} + F2_{/i/} \right] \times \left[F1_{/a/} - F1_{/i/} \right] \right). \quad (1)$$

The measurement of VAI is another commonly used measure[46] that was introduced by Roy et al [49]. and can be expressed using the following formula[49]:

$$VAI = \left(F2_{/i/} + F1_{/a/} \right) / \left(F1_{/i/} + F1_{/u/} + F2_{/u/} + F2_{/a/} \right). \quad (2)$$

In addition, vowel duration was measured as the difference between the onset and offset of each vowel according to previously published methodology[50]. The final value used for statistical analysis was averaged across all 30 occurrences available.

Analysis of inter- and intra-judge reliability was not performed as the stability of methodology used for both healthy as well as dysarthric speech was thoroughly validated in previous studies[17,50].

Dopamine transporter imaging

In PD and RBD patients, we performed DAT-SPECT using the [123I]-2-b-carbomethoxy-3b-(4-iodophenyl)-N-(3-fluoropropyl) nortropane (DaTscan®, GE Healthcare) tracer according to European Association of Nuclear Medicine procedure guidelines[51], using common acquisition and reconstruction parameters described in detail previously[52]. Automated semi-quantitative analysis was applied using the BasGan V2 software[53], and specific binding ratios in both caudate nuclei and putamina relative to background binding were calculated; the lower value from both hemispheres was used for further analyses. Specific binding ratio values below the 95% reference prediction interval were considered abnormal.

Statistical analyses

As the Kolmogorov-Smirnov test for independent samples showed that all acoustic variables were normally distributed, we used analysis of covariance (ANCOVA) with post hoc Bonferroni adjustment to assess group differences. Since there was wide variability in age among PD participants, age was considered as a covariate. The Pearson's partial correlation analysis controlled for age was performed to test for significant relationships between the clinical and acoustic data. The level of significance was set to $p < 0.05$. All statistical analyses were performed in MATLAB (MathWorks, Natick, MA, USA).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Individual participant data that underlie the findings of this study are available upon request to the corresponding author by qualified researchers (i.e., affiliated to a respected university or research institution/hospital). The speech data are not publicly available due to their contain of information that could compromise the privacy of study participants.

Received: 13 April 2022; Accepted: 4 October 2022;
Published online: 20 October 2022

REFERENCES

1. Boeve, B. F. et al. Clinicopathologic correlations in 172 cases of rapid eye movement sleep behavior disorder with or without a coexisting neurologic disorder. *Sleep. Med.* **14**, 754–762 (2013).
2. Miglis, M. G. et al. Biomarkers of conversion to α-synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. *Lancet Neurol.* **20**, 671–684 (2021).
3. Dawson, V. L. & Dawson, T. M. Promising disease-modifying therapies for Parkinson's disease. *Sci. Transl. Med.* **11**, eaba1659 (2019).
4. Postuma, R. B. et al. Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. *Brain* **142**, 744–759 (2019).
5. Postuma, R. B., Lang, A. E., Gagnon, J. F., Pelletier, A. & Montplaisir, J. Y. How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. *Brain* **135**, 1860–1870 (2012).
6. Duffy, J. R. *Motor Speech Disorders - E-Book: Substrates, Differential Diagnosis, and Management*. (Elsevier Health Sciences, 2013).
7. Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L. & Gates, S. Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurol.* **11**, 131–137 (1999).
8. Rusz, J. et al. Defining speech subtypes in De Novo Parkinson Disease: response to long-term levodopa therapy. *Neurology* **97**, e2124–e2135 (2021).
9. Rusz, J. et al. Automated speech analysis in early untreated Parkinson's disease: relation to gender and dopaminergic transporter imaging. *Eur. J. Neurol.* **29**, 81–90 (2022).
10. Darley, F. L., Aronson, A. E. & Brown, J. R. Differential diagnostic patterns of dysarthria. *J. Speech Hearing Res.* **12**, 246–269 (1969).
11. Rusz, J. et al. Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. *Ann. Neurol.* **90**, 62–75 (2021).
12. Rusz, J. et al. Dysprosody in isolated REM sleep behavior disorder with impaired olfaction but intact nigrostriatal pathway. *Mov. Disord.* **37**, 619–623 (2022).
13. Braak, H. et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**, 197–211 (2003).
14. Weismer, G., Jeng, J.-Y., Laures, J. S., Kent, R. D. & Kent, J. F. Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders. *Folia Phoniatr. Logop.* **53**, 1–18 (2001).
15. Lee, J. & Hustad, K. C. A preliminary investigation of longitudinal changes in speech production over 18 months in young children with cerebral palsy. *Folia Phoniatr. Logop.* **65**, 32–39 (2013).
16. Carl, M. & Icht, M. Acoustic vowel analysis and speech intelligibility in young adult Hebrew speakers: Developmental dysarthria versus typical development. *Int. J. Lang. Commun. Disord.* **56**, 283–298 (2021).
17. Rusz, J. et al. Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task. *J. Acoustical Soc. Am.* **134**, 2171–2181 (2013).
18. Rusz, J. et al. Evaluation of speech impairment in early stages of Parkinson's disease: a prospective study with the role of pharmacotherapy. *J. Neural Transm.* **120**, 319–329 (2013).
19. Zhang, S. P., Bandler, R. & Davis, P. J. Brain stem integration of vocalization: role of the nucleus retroambigualis. *J. Neurophysiol.* **74**, 2500–2512 (1995).
20. Skodda, S., Grönheit, W. & Schlegel, U. Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS ONE* **7**, e32132 (2012).
21. Bonnet, A.-M., Loria, Y., Saint-Hilaire, M.-H., Lhermitte, F. & Agid, Y. Does long-term aggravation of Parkinson's disease result from nondopaminergic lesions? *Neurology* **37**, 1539–1539 (1987).
22. Tripoliti, E. et al. Predictive factors of speech intelligibility following subthalamic nucleus stimulation in consecutive patients with Parkinson's disease. *Mov. Disord.* **29**, 532–538 (2014).
23. Cavallieri, F. et al. Dopaminergic Treatment Effects on Dysarthric Speech: Acoustic Analysis in a Cohort of Patients With Advanced Parkinson's Disease. *Front. Neurol.* **11**, 616062 (2021).
24. Caverlé, M. W. J. & Vogel, A. P. Stability, reliability, and sensitivity of acoustic measures of vowel space: A comparison of vowel space area, formant centralization ratio, and vowel articulation index. *J. Acoustical Soc. Am.* **148**, 1436–1444 (2020).
25. Thies, T., Mücke, D., Dano, R. & Barbe, M. T. Levodopa-based changes on vocalic speech movements during prosodic prominence marking. *Brain Sci.* **11**, 594 (2021).

26. Thies, T. et al. Prominence marking in parkinsonian speech and its correlation with motor performance and cognitive abilities. *Neuropsychologia* **137**, 107306 (2020).
27. Mefferd, A. S. & Dietrich, M. S. Tongue- and jaw-specific articulatory underpinnings of reduced and enhanced acoustic vowel contrast in talkers with Parkinson's disease. *J. Speech Lang. Hearing Res.* **62**, 2118–2132 (2019).
28. Hasegawa-Johnson, M., Pizza, S., Alwan, A., Alwan, J. S. & Haker, K. Vowel category dependence of the relationship between palate height, tongue height, and oral area. *J. J. Speech Lang. Hearing Res.* **46**, 738–753 (2003).
29. Tjaden, K., Rivera, D., Wilding, G. & Turner, G. S. Characteristics of the lax vowel space in dysarthria. *J. Speech, Lang., Hearing Res.* **48**, 554–566 (2005).
30. Skodda, S. & Schlegel, U. Speech rate and rhythm in Parkinson's disease: Speech Rate and Rhythm in Parkinson's Disease. *Mov. Disord.* **23**, 985–992 (2008).
31. Delval, A. et al. Freezing/festination during motor tasks in early-stage Parkinson's disease: a prospective study. *Mov. Disord.* **31**, 1837–1845 (2016).
32. Konstantopoulos, K., Vogazianos, P., Christou, Y. & Pisinou, M. Sequential motion rate and oral reading rate: normative data for Greek and clinical implications. *Logopedics Phoniatrics Vocology* 1–6 (2021).
33. Illner, V. et al. Towards automated articulation rate analysis via connected speech in dysarthrias. *J. Speech Lang. Hearing Res.* **65**, 1386–1401 (2022).
34. Hu, M. T. REM sleep behavior disorder (RBD). *Neurobiol. Dis.* **143**, 104996 (2020).
35. Vorperian, H. K. et al. Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: an imaging study. *J. Speech Lang. Hearing Res.* **54**, 995–1010 (2011).
36. Hamdan, A.-L., Khalifee, E., Ziade, G. & Semaan, S. Sexual dimorphism in laryngeal volumetric measurements using magnetic resonance imaging. *Ear, Nose Throat J.* **99**, 132–136 (2020).
37. Mansukhani, M. P., Kolla, B. P. & Ramar, K. International classification of sleep disorders 2 and American Academy of Sleep Medicine Practice Parameters for Central Sleep Apnea. *Sleep. Med. Clin.* **9**, 1–11 (2014).
38. Postuma, R. B. et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591–1601 (2015).
39. Martinez-Martin, P. et al. Expanded and independent validation of the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS). *J. Neurol.* **260**, 228–236 (2013).
40. Doty, R. L., Shaman, P., Kimmelman, C. P. & Dann, M. S. University of pennsylvania smell identification test: a rapid quantitative olfactory function test for the clinic. *Laryngoscope* **94**, 176–178 (1984).
41. Kopecek, M. et al. Montreal cognitive assessment (MoCA): normative data for old and very old Czech adults. *Appl. Neuropsychology: Adult* **24**, 23–29 (2017).
42. Rusz, J., Tykalova, T., Ramig, L. O. & Tripoliti, E. Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. *Mov. Disord.* **36**, 803–814 (2021).
43. Sapir, S., Spielman, J. L., Ramig, L. O., Story, B. H. & Fox, C. Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings. *J. Speech Lang. Hearing Res.* **50**, 899–912 (2007).
44. Sapir, S., Ramig, L. O., Spielman, J. L. & Fox, C. Formant Centralization Ratio: A Proposal for a New Acoustic Measure of Dysarthric Speech. *J. Speech Lang. Hearing Res.* **53**, 114–125 (2010).
45. Skodda, S., Visser, W. & Schlegel, U. Vowel Articulation in Parkinson's Disease. *J. Voice* **25**, 467–472 (2011).
46. Kent, R. D. & Vorperian, H. K. Static measurements of vowel formant frequencies and bandwidths: A review. *J. Commun. Disord.* **74**, 74–97 (2018).
47. Fant, G. *Speech Sounds and Features* (MIT Press, Cambridge, MA,1973).
48. Liu, H.-M., Tsao, F.-M. & Kuhl, P. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *J. Acoustical Soc. Am.* **117**, 3879–3889 (2005).
49. Roy, N., Nissen, S. L., Dromey, C. & Sapir, S. Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *J. Commun. Disord.* **42**, 124–135 (2009).
50. Tykalova, T. et al. Effect of ageing on acoustic characteristics of voice pitch and formants in Czech vowels. *J. Voice* **35**, 931.e21–931.e33 (2020).
51. Darcourt, J. et al. EANM procedure guidelines for brain neurotransmission SPECT using (123)I-labelled dopamine transporter ligands, version 2. *Eur. J. Nucl. Med. Mol. Imaging* **37**, 443–450 (2010).
52. Dusek, P. et al. Relations of non-motor symptoms and dopamine transporter binding in REM sleep behavior disorder. *Sci. Rep.* **9**, 15463 (2019).
53. Calvini, P. et al. The basal ganglia matching tools package for striatal uptake semi-quantification: description and validation. *Eur. J. Nucl. Med Mol. Imaging* **34**, 1240–1253 (2007).

ACKNOWLEDGEMENTS

This study was supported by the Ministry of Health of the Czech Republic, grants no. NU20-08-00445 and MH CZ–DRO-VFN64165, and by National Institute for Neurological Research (Programme EXCELES, ID Project No. LX22NPO5107) - Funded by the European Union – Next Generation EU).

AUTHOR CONTRIBUTIONS

Research project: (A) Conception: J.R., K.S., P.D., and T.T. (B) Organization: P.D., and T.T. (C) Execution: D.S., J.R., M.N., P.D., and T.T. Statistical analysis: (A) Design: J.R. (B) Execution: J.R. (C) Review and critique: J.R. M.N., K.S., E.R., P.D., and T.T. Manuscript preparation: (A) Writing of the first draft: D.S. (B) Review and critique: J.R., M.N., K.S., E.R., P.D., and T.T.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41531-022-00407-7.

Correspondence and requests for materials should be addressed to Tereza Tykalova.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Neural Transmission
https://doi.org/10.1007/s00702-020-02171-5

NEUROLOGY AND PRECLINICAL NEUROLOGICAL STUDIES - ORIGINAL ARTICLE

Dysarthria enhancement mechanism under external clear speech instruction in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy

Dominik Skrabal[1] · Tereza Tykalova[2] · Jiri Klempir[1,3] · Evzen Ruzicka[1] · Jan Rusz[1,2]

Received: 13 January 2020 / Accepted: 5 March 2020

Abstract

Clear speech refers to intentionally modifying conversational speech to maximise intelligibility. This study aimed to compare the speech behaviour of patients with progressive supranuclear palsy (PSP), multiple system atrophy (MSA) and Parkinson's disease (PD) under conversational and clear speech conditions to gain greater pathophysiological insight. A total of 68 participants including 17 PD, 17 MSA, 17 PSP and 17 healthy controls (HC) performed two readings of the same standardized passage. During the first reading, participants were instructed to read the text in an ordinary way, while during the second reading to read the text as clearly as possible. Acoustic analyses were based upon measurements of mean loudness, loudness variability, pitch variability, vowel articulation, articulation rate and speech severity. During clear speech production, PD patients were able to achieve improvements mainly in loudness ($p < 0.05$) and pitch variability ($p < 0.001$), leading to a reduction in overall speech severity ($p < 0.001$), whereas PSP and MSA patients were able to modulate only articulation rate ($p < 0.05$). Contrary to HC and PD groups, which slowed or maintained articulation rate, PSP and MSA groups employed a markedly faster articulation rate under the clear speech condition indicating an opposing approach to speech adaptation. Patients with atypical Parkinsonism showed a different strategy to intentionally improve their speech performance following a simple request to produce speech more clearly compared to PD, suggesting important therapeutic implications for speech rehabilitation management.

Keywords Parkinson's disease · Atypical parkinsonian syndromes · Clear speech · Dysarthria · Speech impairment · Acoustic analysis

Introduction

Idiopathic Parkinson's disease (PD) is a neurological disorder characterized by a combination of motor and non-motor manifestations such as bradykinesia, rigidity, resting tremor, postural instability, REM sleep behaviour disorder, depression and cognitive decline (Poewe et al. 2017). Progressive supranuclear palsy (PSP) and multiple system atrophy (MSA) represent atypical parkinsonian syndromes (APS), a group of closely related disorders featuring parkinsonism combined with various overlapping symptoms such as oculomotor disturbances and cognitive deficit in PSP, and cerebellar ataxia and autonomic dysfunction in MSA. APS differ from PD by more widespread detrimental neurodegeneration, resulting in additional clinical signs, more rapid disease progression and poor levodopa response (O'Sullivan et al. 2008; Wenning et al. 2011).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00702-020-02171-5) contains supplementary material, which is available to authorized users.

✉ Tereza Tykalova
tykalova.tereza@gmail.com

[1] Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University, Katerinska 30, 120 00, Prague 2, Czech Republic

[2] Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27, Prague 6, Czech Republic

[3] Institute of Anatomy, First Faculty of Medicine, Charles University, U nemocnice 3, 128 00 Prague 2, Czech Republic

Published online: 19 March 2020

Speech and voice impairment represent cardinal clinical features developing in 90–100% of patients with PD, PSP and MSA during the course of the disease (Ho et al. 1998; Kluin et al. 1993, 1996; Rusz et al. 2015). Compared to PD, dysarthria is generally more severe in PSP and MSA (Rusz et al. 2015; Tykalova et al. 2017). The majority of PD patients exhibit pure hypokinetic dysarthria (Darley et al. 1969; Ho et al. 1998), while PSP and MSA patients typically develop a mixed dysarthria with a combination of hypokinetic, ataxic and spastic components (Kluin et al. 1993, 1996; Rusz et al. 2015) as a result of more widespread neurodegeneration. In particular, MSA patients typically manifest predominant ataxic patterns of dysarthria due to cerebellar dysfunction, while in PSP, spastic elements prevail as a result of damage to the corticobulbar pathways. These divergent dysarthria patterns associated with differing pathophysiology including α-synucleinopathy in MSA and tauopathy in PSP may have important prognostic and therapeutic implications, especially for speech rehabilitation management.

Conversational speech represents spontaneous, effortless and habitual speech production. On the contrary, clear speech refers to a speaking style where talkers intentionally modify their conversational speech to maximise intelligibility, generally by hyperarticulation, a reduction of speaking rate and increased loudness. Healthy speakers naturally adopt clear speech strategies in day-to-day life when talking in a noisy environment or with someone with a hearing impairment (Payton et al. 1994). Nevertheless, regular application of clear speech techniques has also been recommended as a suitable behavioural therapy for speakers with dysarthria secondary to various neurological disorders, including PD (Beukelman et al. 2002; Duffy 2013; Hustad and Weismer 2007). Acoustic analyses of clear speech production have revealed that PD patients are able to use some of the same clear strategies as non-impaired speakers (Goberman and Elmer 2005; Kearney et al. 2017; Lam and Tjaden 2016; Tjaden et al. 2013, 2014; Whitfield and Goberman 2014). In particular, PD speakers produce significantly slower articulation rate (Goberman and Elmer 2005; Kearney et al. 2017; Tjaden et al. 2013, 2014), increased loudness (Kearney et al. 2017; Tjaden et al. 2013, 2014), increased pitch variability (Goberman and Elmer 2005), increased pitch level (Goberman and Elmer 2005) and increased articulatory-acoustic vowel space (Whitfield and Goberman 2014) under the clear speech condition. The positive association between these acoustic alterations in PD speech and increased clarity (Whitfield and Goberman 2014), intelligibility (Kearney et al. 2017; Tjaden et al. 2014) as well as movement size of the articulators including the jaw, tongue blade and tongue dorsum (Kearney et al. 2017) has been documented by perceptual and kinematic studies. However, the ability of PSP or MSA patients to intentionally improve their speech performance remains to be investigated.

Therefore, the current study aimed to investigate speech behaviour in PSP and MSA patients compared to PD and healthy speakers under conversational and clear speech conditions to gain greater insight into speech changes that would improve speech therapy management and facilitate differential diagnosis.

Methods

Participants

From 2011 to 2016, a total of 51 consecutive patients were recruited for the present study including 17 with a medical diagnosis of probable PSP (12 men and 5 women), 17 with a medical diagnosis of probable MSA (8 men and 9 women) and 17 with a medical diagnosis of idiopathic PD (7 men and 10 women). The PSP group included 14 patients diagnosed with PSP-Richardson syndrome, 2 with PSP-parkinsonism, and 1 with PSP-pure akinesia with gait freezing. The MSA group consisted of 13 patients diagnosed with MSA-parkinsonian subtype and 4 patients with MSA-cerebellar subtype. The clinical diagnoses of all patients were established by a movement disorders specialist (JK) according to the UK Parkinson's Disease Society Brain Bank Criteria for PD (Hughes et al. 1992), the NINDS-PSP clinical diagnostic criteria for PSP (Litvan et al. 1996) and the consensus diagnostic criteria for MSA (Gilman et al. 2008). At the time of examination, all patients were on stable medication for at least 4 weeks consisting of various doses of levodopa alone or in combination with different dopamine agonists and/or amantadine. PD patients were in their best on-medication state and did not manifest dyskinesia at the time of examination. None of the participants received antipsychotic drugs or benzodiazepines. Disease duration was determined based on the self-reported occurrence of the first motor symptoms. PSP and MSA patients underwent scoring according to the Neuroprotection and Natural History in Parkinson Plus Syndromes (NNIPPS) scale (Payan et al. 2011), while PD patients were rated by the Unified Parkinson's Disease Rating Scale (UPDRS) motor subscore. Item 18 of the UPRDS III was used for the perceptual description of speech severity. None of the patients reported a history of speech-language disorders apart from possible neurologic disease manifestations. No statistically significant differences between PSP and MSA groups for disease duration, medication dose, cognitive status, motor or speech severity were found (Mann–Whitney U test: $p = 0.14$–0.91).

The dysarthria presence, type, and severity estimates were evaluated based on the auditory-perceptual judgment of a speech language specialist experienced in movement

disorders using audio recordings of vowel prolongation, /pa/-/ta/-/ka/ syllable repetition, and monolog following the perceptual criteria outlined by Darley et al. (1969). Patient clinical and demographic characteristics are summarised in Table 1.

The healthy control (HC) group consisted of 17 participants (7 men and 10 women) of comparable age (mean age 63.3, SD 7.4, range 53–74). No HC subjects reported a history of neurological disorders or other disorders that may affect speech, language or hearing. No significant differences in age distribution were detected among PSP, MSA, PD and HC groups (analysis of variance: $p = 0.37$).

All subjects in the present study were Czech native speakers, and none manifested marked depressive or cognitive deficits that would conflict with the recording procedure. The study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All participants provided written, informed consent to the neurological examination and recording procedure.

Recording procedure

The audio data were recorded in a quiet room with a low level of ambient noise using a head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed approximately 5 cm from the subject's lips. The speech signals were sampled at 48 kHz with 16–bit resolution. The recordings were obtained during one session with a speech specialist who conveyed instructions to the participants. Each participant completed a series of speaking tasks, including the reading of two different standardised passages (Supplementary material Fig. S1 and Fig. S2) as part of a longer protocol lasting about 20 min. A reading passage was preferred to create a more natural condition of connected speech and, at the same time, to maintain a standardised speaking task. The reading passage in Supplementary material Fig. S2 was read under clear and conversation instructions and used for evaluation of the clear speech condition, while the reading passage in Supplementary material Fig. S1 was read twice and was included to assess the effect of repeated reading. The first passage was comprised of 80 words (Supplementary material Fig. S1) and was presented at the beginning of the recording session. During the

Table 1 Patient clinical and demographic characteristics

	PSP ($n = 17$; 12 men, 5 women) Mean/SD (range)	MSA ($n = 17$; 8 men, 9 women) Mean/SD (range)	PD ($n = 17$; 7 men, 10 women) Mean/SD (range)
General			
Age (years)	66.2/5.0 (54 – 72)	62.3/6.3 (52 – 72)	64.5/7.3 (49 – 74)
Disease duration[a] (years)	4.3/2.2 (2 – 11)	3.9/1.4 (2 – 7)	8.8/6.5 (2 – 25)
Levodopa equivalent (mg/day)	468/527 (0 – 1500)	444/524 (0 – 1700)	840/536 (0 – 1680)
Amantadine (mg/day)	171/172 (0 – 500)	121/123 (0 – 300)	117/133 (0 – 300)
NNIPPS total	71.2/27.7 (19 – 132)	69.5/24.9 (41 – 123)	–
NNIPPS mental function subscore	7.6/4.6 (0–17)	5.9/3.8 (0–14)	–
UPDRS III total	–	–	15.1/8.5 (7 – 31)
UPDRS III speech item	2.0/0.7 (1 – 3)	1.6/0.6 (1 – 3)	0.8/0.8 (0 – 2)
Dysarthria severity			
None (%)	0 ($n = 0$)	0 ($n = 0$)	23 ($n = 4$)
Mild (%)	18 ($n = 3$)	41 ($n = 7$)	65 ($n = 11$)
Moderate (%)	53 ($n = 9$)	35 ($n = 6$)	12 ($n = 2$)
Severe (%)	29 ($n = 5$)	24 ($n = 4$)	0 ($n = 0$)
Dysarthria type			
Hypokinetic (%)	6 ($n = 1$)	6 ($n = 1$)	100 ($n = 13$)
Hypokinetic–spastic (%)	41 ($n = 7$)	6 ($n = 1$)	0 ($n = 0$)
Hypokinetic–ataxic (%)	6 ($n = 1$)	41 ($n = 7$)	0 ($n = 0$)
Hypokinetic–spastic–ataxic (%)	35 ($n = 6$)	29 ($n = 5$)	0 ($n = 0$)
Ataxic (%)	6 ($n = 1$)	12 ($n = 2$)	0 ($n = 0$)
Ataxic–spastic (%)	6 ($n = 1$)	6 ($n = 1$)	0 ($n = 0$)

PSP progressive supranuclear palsy; *MSA* multiple system atrophy; *PD* Parkinson's disease; *NNIPPS* Natural History and Neuroprotection on Parkinson plus syndromes-Parkinson plus scale; *UPDRS* Unified Parkinson Disease Rating Scale

[a]Based on self-reported occurrence of first motor symptoms

first attempt, speakers were asked to read the passage "in a habitual manner which they employ in everyday life" (hereafter, the first reading). During the second attempt, speakers were asked to read the passage "once again" (hereafter, the second reading). The second passage consisted of 137 words (Supplementary material Fig. S2) and was acquired approximately 10 min later. During the first performance, participants were instructed to read the passage "in a habitual manner which they employ in everyday life" (hereafter, conversational speech). During the second performance, participants were asked to read the passage "as clearly as possible, as if they were presenting in front of a large audience" (hereafter, clear speech).

Acoustic analyses

Audio samples were analysed using PRAAT® (Boersma and Weenink 2001) specialised speech software. Five different acoustic measures including intensity level (intensity), intensity variability (intensity SD), fundamental frequency variability (F0 SD), vowel space area (VSA), and articulation rate (AR) were applied to objectively investigate speech changes in the reading utterances. These specific measures were chosen as they have previously been reported to be affected in patients with both PD and APS (Rusz et al. 2013a, 2015), employed during clear speech production in PD (Goberman and Elmer 2005; Lam and Tjaden 2016), and can be obtained from the reading passage.

Average loudness of speech was established as the mean value of the intensity contour. Loudness variability was defined as the standard deviation of the intensity contour after removing periods of silence exceeding 60 ms (Rusz et al. 2011). Fundamental frequency variability, which characterises intonation, was calculated by defining the standard deviation of the fundamental frequency contour that was established via the autocorrelation-based procedure in PRAAT® (Boersma and Weenink 2001), after manual adjustment of pitch range for each speaker. Moreover, the conversion of fundamental frequency contour from Hertz to semitone scale was accomplished in order to compensate for gender-related differences (Rusz et al. 2011). Vowel space area, which mirrors precision of vowel articulation, is the traditional acoustic parameter defined as Euclidean distances between the first (F1) and second (F2) formant frequency coordinates of the corner vowels /a/, /i/, /u/ in the triangular F1– F2 vowel space and can be easily calculated by following formula: $VSA = 0.5 \times |F1i \times (F2a - F2u) + F1a \times (F2u - F2i) + F1u \times (F2i - F2a)|$. The F1 and F2 frequencies were extracted from a 30 ms segment at the temporal midpoint of the stable part of each vowel. Each of the enlisted vowels /a/, /i/, and /u/ were represented by 10 occurrences predefined within each standardised reading passage (Supplementary material Fig. S1 and Fig. S2). Detailed methodological information has been described previously (Rusz et al. 2013a). Articulation rate, which reflects proper speech timing ability, was calculated as the number of words per second after removing periods of silence exceeding 60 ms (Rusz et al. 2011).

The dysarthria index, based upon quantitative acoustic analysis of four individual speech dimensions including loudness, loudness variability, pitch variability, and vowel articulation, was calculated as the mean z-score of the four individual measures, which were converted to z-scores using the HC means and standard deviations. For those measures in which the lower raw scores were associated with greater dysarthria, the z-scores were reversed, allowing all results to be interpreted as higher z-scores indicating more speech impairment. The articulation rate was not included in the dysarthria index as slow rate has been reported to be common in PSP, manifest occasionally in MSA and rarely in PD (Rusz et al. 2015); thus both increased as well as decreased speaking rate might be expected to enhance speech performance under the clear speech condition.

Statistical analyses

All acoustic features were found to be normally distributed by the Kolmogorov–Smirnov test. Statistical analyses were performed using repeated measures analysis of variance (RM-ANOVA) with GROUP (PSP vs. MSA vs. PD vs. HC) treated as a between-group factor and TASK (clear vs. conversational) treated as a within-group factor. Post-hoc significance was assessed with a paired t test for effect of TASK among individual groups and by the Fisher least-squares difference for effect of GROUP. Bonferroni correction for multiple comparisons was applied for the six tests performed with a corrected p threshold equal to $p < 0.0083$ (i.e., 0.05/6) for a $p < 0.05$ level of significance.

Results

Clear vs. conversational speech

Figure 1 depicts the results of acoustic analyses for clear vs. conversational speech among PSP, MSA, PD and HC subjects.

For loudness, RM-ANOVA showed a significant effect for TASK [$F(1,64) = 8.5$, $p = 0.03$, $\eta^2 = 0.12$], particularly as the PD group increased loudness (average change from conversational to clear speech 1.36 dB, $p = 0.03$). No significant effect was found for GROUP [$F(3,64) = 2.9$, $p = 0.25$, $\eta^2 = 0.12$] or GROUP × TASK [$F(3,64) = 3.0$, $p = 0.21$, $\eta^2 = 0.13$].

For loudness variability, we detected a significant effect for TASK [$F(1,64) = 7.6$, $p = 0.046$, $\eta^2 = 0.11$], particularly

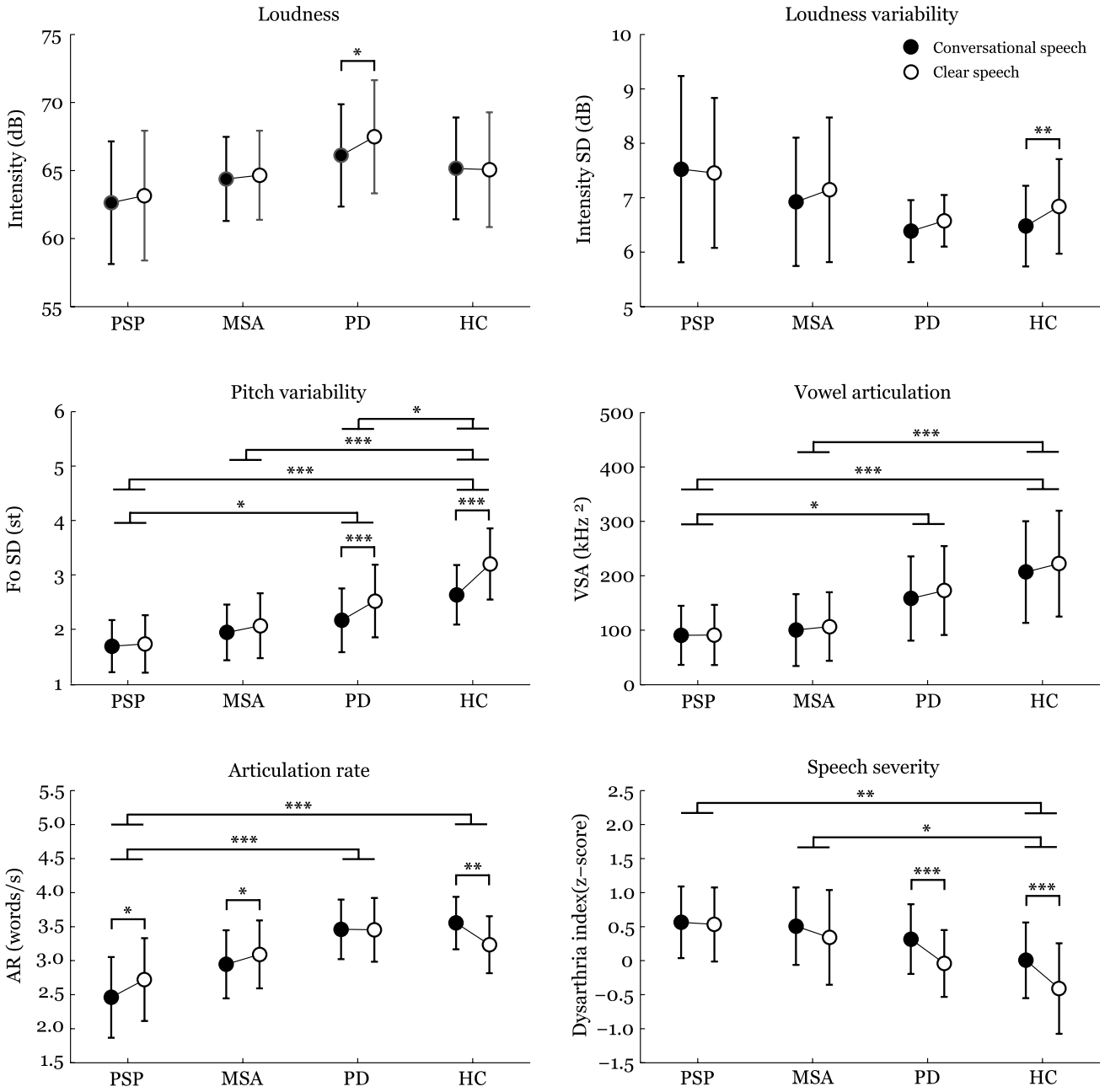


Fig. 1 Comparison of results of acoustic analyses among PSP, MSA, PD and HC groups under conversational and clear speech conditions. *Intensity SD* intensity variability; *F0 SD* fundamental frequency variability, *VSA* vowel space area, *AR* articulation rate, *PSP* progressive supranuclear palsy, *MSA* multiple system atrophy, *PD* Parkinson's disease, *HC* healthy controls. "Asterisks" indicate significant differences after Bonferroni correction: $*p<0.01$; $**p<0.001$; $***p<0.0001$

as the HC group increased loudness variability (average change from conversational to clear speech 0.36 dB, $p=0.002$). No significant effect was found for GROUP [$F(3,64)=2.9$, $p=0.24$, $\eta^2=0.12$] or GROUP × TASK [$F(3,64)=2.0$, $p=0.75$, $\eta^2=0.09$].

For pitch variability, a significant effect was found for TASK [$F(1,64)=85.6$, $p<0.001$, $\eta^2=0.57$], mainly as the PD group increased pitch variability (average change from conversational to clear speech 0.36 st, $p<0.001$) as well as HC group increased pitch variability (average change from conversational to clear speech 0.57 st, $p<0.001$). In addition, a significant effect was found for GROUP [$F(3,64)=14.4$, $p<0.001$, $\eta^2=0.40$], mainly due to differences between PSP and HC ($p<0.001$), MSA and HC ($p<0.001$), PD and HC ($p=0.02$), and PSP and PD ($p=0.01$) groups. Finally, significant interaction was revealed for GROUP × TASK

[$F(3,64) = 16.3$, $p < 0.001$, $\eta^2 = 0.43$] as both PD and HC groups ($p < 0.001$) were able to increase pitch variability in the clear condition while PSP ($p = 1$) and MSA ($p = 0.23$) groups were not.

For vowel articulation, we revealed a significant effect for TASK [$F(1,64) = 7.9$, $p = 0.04$, $\eta^2 = 0.11$] as well as GROUP [$F(3,64) = 10.3$, $p < 0.001$, $\eta^2 = 0.33$]. Post-hoc tests showed significant differences between HC and both APS groups ($p < 0.001$) as well as PSP and PD groups ($p = 0.03$). No significant effect was found for GROUP × TASK [$F(3,64) = 1.1$, $p = 1.0$, $\eta^2 = 0.05$].

For articulation rate, a significant effect was detected for GROUP [$F(3,64) = 12.3$, $p < 0.001$, $\eta^2 = 0.37$], reflecting differences between PSP and both PD and HC groups ($p < 0.001$). Importantly, a significant effect was revealed for GROUP × TASK [$F(3,64) = 11.0$, $p < 0.001$, $\eta^2 = 0.34$], as the PSP group increased articulation rate (average change from conversational to clear speech 0.26 words/s, $p = 0.01$) as well as MSA group increased articulation rate (average change from conversational to clear speech 0.15 words/s, $p = 0.04$), whereas the HC group decreased articulation rate (average change from conversational to clear speech − 0.32 words/s, $p = 0.004$). No significant effect was detected for TASK [$F(1,64) = 0.3$, $p = 1$, $\eta^2 = 0.01$].

For dysarthria index, we revealed a significant effect for TASK [$F(1,64) = 72.8$, $p < 0.001$, $\eta^2 = 0.53$], reflecting increased speech performance (average change from conversational to clear speech) in PD (z-score change of − 0.36, $p < 0.001$), HC (z-score change of − 0.42, $p < 0.001$), and partially MSA (z-score change of − 0.17, uncorrected $p = 0.03$) groups. In addition, a significant effect was found for GROUP [$F(3,64) = 5.9$, $p = 0.007$, $\eta^2 = 0.22$], mainly attributed to differences between PSP and HC ($p = 0.001$), and MSA and HC ($p = 0.01$) groups. Finally, a significant interaction was revealed for GROUP × TASK [$F(3,64) = 9.7$, $p < 0.001$, $\eta^2 = 0.31$] as both PD and HC ($p < 0.001$) groups were able to improve speech performance while PSP ($p = 1$) and MSA ($p = 0.18$) groups were not.

First vs. second reading

Figure 2 shows the results of acoustic analyses for first vs. second reading among PSP, MSA, PD and HC subjects.

For loudness, RM-ANOVA revealed no significant effect for TASK [$F(1,62) = 0.01$, $p = 1$, $\eta^2 = 0.00$], GROUP [$F(3,62) = 2.0$, $p = 0.71$, $\eta^2 = 0.09$], or GROUP × TASK [$F(3,62) = 1.0$, $p = 1$, $\eta^2 = 0.05$].

For loudness variability, we observed no significant effect for TASK [$F(1,62) = 0.2$, $p = 1$, $\eta^2 = 0.00$], GROUP [$F(3,62) = 2.8$, $p = 0.29$, $\eta^2 = 0.12$], or GROUP × TASK [$F(3,62) = 3.3$, $p = 0.16$, $\eta^2 = 0.14$].

For pitch variability, a significant effect was found for GROUP [$F(3,62) = 10.5$, $p < 0.001$, $\eta^2 = 0.34$], mainly attributed to differences between PSP and HC ($p < 0.001$), MSA and HC ($p < 0.001$), and PD and HC ($p = 0.049$) groups. No significant effect was found for TASK [$F(1,62) = 2.3$, $p = 0.82$, $\eta^2 = 0.04$] or GROUP × TASK [$F(3,62) = 0.8$, $p = 1$, $\eta^2 = 0.04$].

For vowel articulation, we revealed a significant effect for GROUP [$F(3,62) = 6.9$, $p = 0.003$, $\eta^2 = 0.25$]. Post-hoc tests showed significant differences between PSP and HC ($p = 0.004$), MSA and HC ($p = 0.045$) as well as PSP and PD ($p = 0.005$) groups. No significant effect was revealed for TASK [$F(1,62) = 0.0$, $p = 1$, $\eta^2 = 0.00$], or GROUP × TASK [$F(3,62) = 1.8$, $p = 1$, $\eta^2 = 0.08$].

For articulation rate, a significant effect was detected for TASK [$F(1,62) = 8.7$, $p = 0.03$, $\eta^2 = 0.12$] as all groups slightly increased their articulation rate at the second reading. In addition, a significant effect was revealed for GROUP [$F(3,62) = 10.5$, $p < 0.001$, $\eta^2 = 0.34$], reflecting differences between PSP and both PD and HC groups ($p < 0.001$). No significant effect was observed for GROUP × TASK [$F(3,62) = 0.7$, $p = 1$, $\eta^2 = 0.03$].

For dysarthria index, a significant effect was found for GROUP [$F(3,62) = 6.1$, $p = 0.006$, $\eta^2 = 0.23$], mainly attributed to differences between PSP and HC ($p = 0.004$), MSA and HC ($p = 0.02$), and PSP and PD ($p = 0.03$) groups. No significant effect was observed for TASK [$F(1,62) = 0.3$, $p = 1$, $\eta^2 = 0.00$], or GROUP × TASK [$F(3,62) = 3.3$, $p = 0.16$, $\eta^2 = 0.14$].

Discussion

Although previous studies (Goberman and Elmer 2005; Kearney et al. 2017; Lam and Tjaden 2016; Tjaden et al. 2013, 2014; Whitfield and Goberman 2014) have examined the acoustic characteristics of clear speech in PD, the present study represents the first attempt to clarify potential differences in clear speech strategies among atypical parkinsonian syndromes. During clear speech production, PD patients were able to achieve improvements mainly in loudness and pitch variability, leading to enhanced overall speech performance, whereas PSP and MSA patients were able to modulate only articulation rate. One potential explanation is that patients with PD are able to dramatically improve their performance when cued externally (Sapir 2014). Improvement under external cueing has also been documented in handwriting and gait (Ford et al. 2010; Oliveira et al. 1997). The ability to generate externally cued movements is largely mediated via the cerebellum (Brown and Marsden 1998), which may be affected in both PSP and MSA (Gilman et al. 2008; Höglinger et al. 2017). Moreover, at least some speech dimensions in PD such as pitch and loudness were found to be responsive to the administration of dopaminergic therapy (Rusz et al.

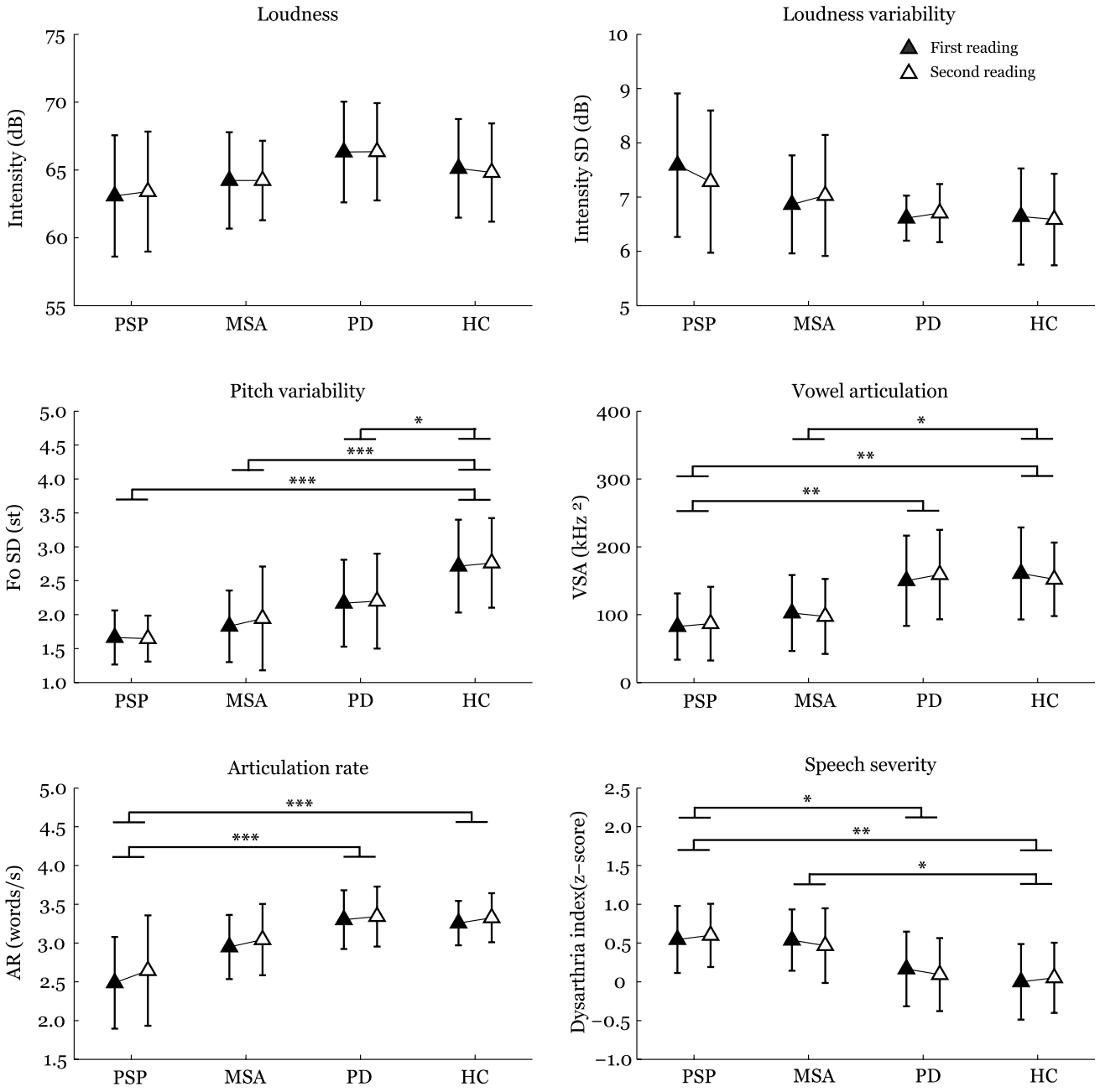


Fig. 2 Comparison of results of acoustic analyses among PSP, MSA, PD and HC groups for first vs. second reading. *Intensity SD* intensity variability, *F0 SD* fundamental frequency variability, *VSA* vowel space area, *AR* articulation rate, *PSP* progressive supranuclear palsy, *MSA* multiple system atrophy, *PD* Parkinson's disease, *HC* healthy controls. "Asterisks" indicate significant differences after Bonferroni correction: $*p < 0.01$; $**p < 0.001$; $***p < 0.0001$

2013b), which might lead to temporarily better speech outcomes compared to APS manifesting more axial dysfunction unresponsive to levodopa (Armstrong 2018; Fanciulli et al. 2019). Furthermore, postsynaptic receptors in the striatum are preserved in PD while in APS they are associated with reduced binding as a reflection of postsynaptic as well as presynaptic dopaminergic degeneration (Antonini et al. 1997; Schreckenberger et al. 2004), which might lead to various adaptive or compensatory changes (Navntoft and Dreyer 2016).

Contrary to HC and PD groups, in which slowed or stable articulation rate was observed, PSP and MSA groups employed markedly faster articulation rate under the clear speech condition. Although this alteration of articulation rate did not lead to a significant reduction of speech severity, it cannot be considered a simple effect of repeated

reading, as no significant increase of articulation rate in APS was observed in the comparison of habitual reading of the same text twice. Interestingly, we did not reveal a statistically significant difference in speech severity measurement between APS groups, although a trend toward better speech production during clear speech was found in MSA ($p = 0.03$, uncorrected). The reason behind worse PSP performance in clear speech may be related to different pathophysiology, where PSP is a tauopathy while both MSA and PD are α-synucleinopathies. Moreover, worse performance in PSP might be attributed to earlier development of executive dysfunction (Gerstenecker et al. 2013). However, no group difference in NNIPPS mental function subscore between PSP and MSA was observed. All together, these findings may imply a potential beneficial effect of speech therapy, especially in MSA, although a significant effect of longitudinal speech therapy such as SPEAK OUT or Lee Silverman Voice Treatment has been reported in both MSA and PSP cohorts (Park 2018; Sale et al. 2015).

On the task level, our results indicate that participants were able to improve their speech performance with the simple request to produce speech clearly by intended adjustments in loudness level, loudness variability, pitch variability, vowel formants, and articulation rate. Notably, no differences were found between two repeated readings, supporting the direct effect of clear speech instructions. However, the specific clear speech strategies differed between groups. While clear speech in the HC group was characterised mainly by increased loudness variability, increased pitch variability and slower articulation rate, speech in the PD group was defined by elevated pitch variability and loudness, while no significant changes were revealed for loudness variability, VSA or articulation rate. Notably, monopitch and monoloudness are thought to be the most distinctive features of hypokinetic dysarthria in PD (Darley et al. 1969). These findings are further in general agreement with previous reports investigating PD, which also found increased pitch variability (Goberman and Elmer 2005), higher loudness (Kearney et al. 2017; Tjaden et al. 2013, 2014) and no change in VSA (Goberman and Elmer 2005) in clear compared to habitual speech. However, contrary to previous studies (Goberman and Elmer 2005; Kearney et al. 2017; Tjaden et al. 2013, 2014), we did not observe slower articulation rate in the clear speech of PD subjects. This discrepancy may be due to differences among PD cohorts. In particular, studies published by Tjaden et al. (2013, 2014) investigated PD patients with longer post-diagnosis disease duration and thus faster habitual articulation rate compared to controls, while we included patients in earlier stages of the disease with the same articulation rate as controls. Therefore, one might expect that our PD patients did not need to compensate for faster articulation rate to such an extent.

To date, clear speech has been expected to result in rate reduction which allows speakers enough time for vocal tract arrangement leading to more precise articulation (Yorkston et al. 2010) in both controls as well as patients with different neurological conditions such as PD, multiple sclerosis or traumatic brain injury (Beukelman et al. 2002; Goberman and Elmer 2005; Tjaden et al. 2013, 2014). Somewhat unanticipated, under the clear speech condition, we found faster articulation rate in our MSA and PSP groups, indicating an opposing approach to speech adaptation in APS. It should be mentioned that on a group level, our PSP and MSA patients manifested a significantly slower articulation rate compared to HC and PD subjects. In agreement, a significantly slower speaking rate was previously observed in both APS groups (Huh et al. 2015; Skodda et al. 2011; Rusz et al. 2015), though more expressed in PSP patients (Rusz et al. 2015). Thus, one might expect that the markedly slow articulation rate in APS could be even so slow as to not be profitable concerning speech performance, thereby the acceleration of articulation rate would be beneficial. Indeed, Tjaden et al. (2013, 2014) revealed that artificially decreased articulation rate does not contribute to better speech performance. Specifically, the authors (Tjaden et al. 2013, 2014) compared four different speaking styles including habitual, clear and slow conditions in patients with PD and multiple sclerosis, and found that clear speech maximized peripheral and nonperipheral vowel space areas (Tjaden et al. 2013) as well as overall perceived intelligibility (Tjaden et al. 2014), despite the most significant rate reduction observed in the slow condition.

On the group level, pitch variability seems to be the most sensitive biomarker separating HC from PSP, MSA and PD groups as well as PD from PSP group. Indeed, in landmark work published by Darley et al. (1969) monopitch was identified as the most prevalent aspect of hypokinetic dysarthria in PD. In addition, the previous study (Harel et al. 2004), which conducted a retrospective analysis of speech in two PD patients on samples of speech produced over a 10-year period surrounding the time of disease diagnosis, reported reduced intonation variability detectable several years before the onset of the first PD motor manifestations. Interestingly, the occurrence of monopitch was revealed even in patients with rapid eye movement sleep behaviour disorder (Rusz et al. 2018), which is considered to be one of the most important clinical phenotypes for predicting future conversion to PD.

Due to restricted opportunities in the recruitment of participants with rare diagnoses, the presented findings may be limited by partial gender-imbalance between groups, especially by the lower percentage of women in the PSP group. Therefore, we cannot exclude the possibility that the observed changes in speech parameters are partially affected by gender-specific aspects of speech. We also cannot exclude

the effect of possible differences in the cognitive abilities of patients on presented results as we did not perform thorough cognitive testing such as the Montreal Cognitive Assessment.

Conclusion

In conclusion, our results objectively confirmed that patients with PD as well as APS are capable of intentionally improving their speech performance with a simple request to produce speech more clearly. In particular, PD patients achieved enhanced overall speech performance mainly by greater loudness and pitch variability, while MSA and PSP patients employed faster articulation rate, contrary to HC and PD patients, which slowed or maintained articulation. Clear speech could therefore represent an alternative therapeutic method that maximizes intelligibility in parkinsonian patients, such as the most commonly-used speech-treatment program for PD called Lee Silverman Voice Treatment (Ramig et al. 2001). This speech-treatment program (Ramig et al. 2001) is based only on instructing patients to concentrate on performing loud speech. Future studies should assess the effect of long-term intensive clear speech therapy in parkinsonian patients to determine whether clear speech techniques improve speech performance only within a single session or with carryover to real-life communication contexts. To date, there have not been any studies investigating clear speech as a regular treatment technique for dysarthria in PD or APS, however a study published by Park et al. (2016) suggests that the "Be clear" intensive treatment method may potentially be an effective intervention with carryover to real-life situations for nonprogressive dysarthria in patients with traumatic brain injury or stroke.

Acknowledgements This study was supported by the Czech Ministry of Health project NV19-04-00120. All rights reserved.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical standards Each participant of this study provided written, informed consent. The study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic, and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

References

Antonini A, Leenders KL, Vontobel P, Maguire RP, Missimer J, Psylla M, Günther I (1997) Complementary PET studies of striatal neuronal function in the differential diagnosis between multiple system atrophy and Parkinson's disease. Brain 120:2187–2195. https://doi.org/10.1093/brain/120.12.2187

Armstrong MJ (2018) Progressive supranuclear palsy: an update. Curr Neurol Neurosci Rep 18:12. https://doi.org/10.1007/s11910-018-0819-5

Beukelman DR, Fager S, Ullman C, Hanson E, Logemann J (2002) The impact of speech supplementation and clear speech on the intelligibility and speaking rate of people with traumatic brain injury. J Med Speech Lang Pathol 10:237–242

Boersma P, Weenink D (2001) Praat, a system for doing phonetics by computer. Glot Int 5:341–345

Brown P, Marsden C (1998) What do the basal ganglia do? Lancet 351:1801–1804. https://doi.org/10.1016/S0140-6736(97)11225-9

Darley FL, Aronson AE, Brown JR (1969) Clusters of deviant speech dimensions in the dysarthrias. J Speech Hear Res 12:462–496

Duffy JR (2013) Motor speech disorders: substrates, differential diagnosis, and management, 3rd edn. Mosby, St. Louis, MO

Fanciulli A, Stankovic I, Krismer F, Seppi K, Levin J, Wenning GK (2019) Multiple system atrophy. Int Rev Neurobiol 149:137–192. https://doi.org/10.1016/bs.irn.2019.10.004

Ford M, Malone L, Nyikos I, Yelisetty R, Bickel C (2010) Gait training with progressive external auditory cueing in persons with Parkinson's disease. Arch Phys Med Rehabil 91:1255–1261. https://doi.org/10.1016/j.apmr.2010.04.012

Gerstenecker A, Mast B, Duff K, Ferman TJ, Litvan I (2013) Executive dysfunction is the primary cognitive impairment in progressive supranuclear palsy. Arch Clin Neuropsychol 28:104–113. https://doi.org/10.1093/arclin/acs098

Gilman S, Wenning GK, Low PA, Brooks DJ, Mathias CJ et al (2008) Second consensus statement on the diagnosis of multiple system atrophy. Neurology 71:670–676. https://doi.org/10.1212/01.wnl.0000324625.00404.15

Goberman AM, Elmer LW (2005) Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease. J Commun Disord 38:215–230. https://doi.org/10.1016/j.jcomdis.2004.10.001

Harel BT, Cannizzarom MS, Cohen H, Reilly N, Snyder PJ (2004) Acoustic characteristics of Parkinsonian speech: a potential biomarker of early disease progression and treatment. J Neurolinguistics 17:439–453. https://doi.org/10.1016/j.jneuroling.2004.06.001

Ho AK, Iansek R, Marigliani C, Bradshaw JL, Gates S (1998) Speech impairment in a large sample of patients with Parkinson's disease. Behav Neurol 11:131–137

Höglinger GU, Respondek G, Stamelou M, Kurz C, Josephs KA et al (2017) Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. Mov Disord 32:853–864. https://doi.org/10.1002/mds.26987

Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55:181–184. https://doi.org/10.1136/jnnp.55.3.181

Huh YE, Park J, Suh MK, Lee SE, Kim J, Jeong Y, Kim HT, Cho JW (2015) Differences in early speech patterns between Parkinson variant of multiple system atrophy and Parkinson's disease. Brain Lang 147:14–20. https://doi.org/10.1016/j.bandl.2015.04.007

Hustad KC, Weismer G (2007) Interventions to improve intelligibility and communicative success for speakers with dysarthria. In: Weismer G (ed) Motor speech disorders. Plural, San Diego, pp 217–228

Kearney E, Giles R, Haworth B, Faloutsos P, Baljko Y (2017) Sentence-Level Movements in Parkinson's Disease: Loud, Clear, and Slow Speech. J Speech Lang Hear Res 60:3426–3440. https://doi.org/10.1044/2017_JSLHR-S-17-0075

Kluin KJ, Foster NL, Berent S, Gilman S (1993) Perceptual analysis of speech disorders in progressive supranuclear palsy. Neurology 43:563–566. https://doi.org/10.1212/wnl.43.3_part_1.563

Kluin KJ, Gilman S, Lohman M, Junck L (1996) Characteristics of the dysarthria of multiple system atrophy. Arch Neurol 53:545–548. https://doi.org/10.1001/archneur.1996.00550060089021

Lam J, Tjaden K (2016) Clear Speech Variants: An Acoustic Study in Parkinson's Disease. J Speech Lang Hear Res 59:631–646. https://doi.org/10.1044/2015_JSLHR-S-15-0216

Litvan I, Hauw JJ, Bartko JJ, Lantos PL, Daniel SE, Horoupian DS, McKee A, Dickson D, Bancher C, Tabaton M, Jellinger K, Anderson DW (1996) Validity and reliability of the preliminary NINDS neuropathologic criteria for progressive supranuclear palsy and related disorders. J Neuropathol Exp Neurol 55:97–105. https://doi.org/10.1097/00005072-199601000-00010

Navntoft CA, Dreyer JK (2016) How compensation breaks down in Parkinson's disease: insights from modeling of denervated striatum. Mov Disord 31:280–289. https://doi.org/10.1002/mds.26579

Oliveira R, Gurd J, Nixon P, Marshall J, Passingham R (1997) Micrographia in Parkinson's disease: the effect of providing external cues. J Neurol Neurosurg Psychiatry 63:429–433. https://doi.org/10.1136/jnnp.63.4.429

O'Sullivan SS, Massey LA, Williams DR, Silveira-Moriyama L, Kempster PA, Holton JL, Revesz T, Lees AJ (2008) Clinical outcomes of progressive supranuclear palsy and multiple system atrophy. Brain 131:1362–1372. https://doi.org/10.1093/brain/awn065

Park S, Theodoros D, Finch E, Cardell E (2016) Be clear: a new intensive speech treatment for adults with nonprogressive dysarthria. Am J Speech Lang Pathol 25:97–110. https://doi.org/10.1044/2015_AJSLP-14-0113

Park Y (2018) Efficacy of intensive treatment of dysarthria for people with multiple system atrophy. Phonetics Speech Sci 10:163–171

Payan CA, Viallet F, Landwehrmeyer BG, Bonnet AM, Borg M, Durif F et al (2011) Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: validation of the NNIPPS–Parkinson Plus Scale. PLoS ONE 6:e22293. https://doi.org/10.1371/journal.pone.0022293

Payton KL, Uchanski RM, Braida LD (1994) Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. J Acoust Soc Am 95:1581–1592. https://doi.org/10.1121/1.408545

Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, Schrag AE, Lang AE (2017) Parkinson disease. Nat Rev Dis Primers 3:17013. https://doi.org/10.1038/nrdp.2017.13

Ramig LO, Sapir S, Countryman S et al (2001) Intensive voice treatment (LSVT) for patients with Parkinson's disease: a 2 year follow up. J Neurol Neurosurg Psychiatry 71:493–498. https://doi.org/10.1136/jnnp.71.4.493

Rusz J, Cmejla R, Ruzickova H, Ruzicka E (2011) Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J Acoust Soc Am 129:350–367. https://doi.org/10.1121/1.3514381

Rusz J, Cmejla R, Tykalova T, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E (2013a) Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. J Acoust Soc Am 134:2171–2181. https://doi.org/10.1121/1.4816541

Rusz J, Cmejla R, Ruzickova H, Klempir J, Majerova V, Picmausova J, Roth J, Ruzicka E (2013b) Evaluation of speech impairment in early stages of Parkinson's disease: a prospective study with the role of pharmacotherapy. J Neural Transm 120:319–329. https://doi.org/10.1007/s00702-012-0853-4

Rusz J, Bonnet C, Klempir J, Tykalova T, Baborova E, Novotny M, Rulseh A, Ruzicka E (2015) Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. J Neurol 262:992–1001. https://doi.org/10.1007/s00415-015-7671-1

Rusz J, Hlavnicka J, Tykalova T, Novotny M, Dusek P, Sonka K, Ruzicka E (2018) Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng 26:1495–1507. https://doi.org/10.1109/tnsre.2018.2851787

Sale P, Castiglioni D, De Pandis MF, Torti M, Dall'armi V, Radicati FG, Stocchi F (2015) The Lee Silverman Voice Treatment (LSVT®) speech therapy in progressive supranuclear palsy. Eur J Phys Rehabil Med 51:569–574

Sapir S (2014) 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 57:1330–1343. https://doi.org/10.1044/2014_JSLHR-S-13-0039

Schreckenberger M, Hägele S, Siessmeier T, Buchholz HG, Armbrust-Henrich H, Rösch F, Gründer G, Bartenstein P, Vogt T (2004) The dopamine D2 receptor ligand 18F-desmethoxyfallypride: an appropriate fluorinated PET tracer for the differential diagnosis of parkinsonism. Eur J Nucl Med Mol Imaging 31:1128–1135. https://doi.org/10.1007/s00259-004-1465-5

Skodda S, Visser W, Schlegel U (2011) Acoustical analysis of speech in progressive supranuclear palsy. J Voice 25:725–731. https://doi.org/10.1016/j.jvoice.2010.01.002

Tjaden K, Lam J, Wilding G (2013) Vowel acoustics in Parkinson's disease and multiple sclerosis: comparison of clear, loud, and slow speaking conditions. J Speech Lang Hear Res 56:1485–1502. https://doi.org/10.1044/1092-4388(2013/12-0259)

Tjaden K, Sussman JE, Wilding GE (2014) Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. J Speech Lang Hear Res 57:779–792. https://doi.org/10.1044/2014_JSLHR-S-12-0372

Tykalova T, Rusz J, Klempir J, Cmejla R, Ruzicka E (2017) Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Brain Lang 165:1–9. https://doi.org/10.1016/j.bandl.2016.11.005

Wenning GK, Litvan I, Tolosa E (2011) Milestones in atypical and secondary Parkinsonisms. Mov Disord 26:1083–1095. https://doi.org/10.1002/mds.23713

Whitfield JA, Goberman AM (2014) Articulatory-acoustic vowel space: application to clear speech in individuals with Parkinson's disease. J Commun Disord 51:19–28. https://doi.org/10.1016/j.jcomdis.2014.06.005

Yorkston K, Beukelman DR, Strand E, Hakel M (2010) Clinical management of speakers with motor speech disorders, 3rd edn. Pro-Ed, Austin, TX

ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Automated Vowel Articulation Analysis in Connected Speech Among Progressive Neurological Diseases, Dysarthria Types, and Dysarthria Severities

Vojtech Illner,[a] Tereza Tykalova,[a] Dominik Skrabal,[b] Jiri Klempir,[b] and Jan Rusz[a,b,c]

[a] Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic [b] Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic [c] Department of Neurology and ARTORG Center, Inselspital, Bern University Hospital, University of Bern, Switzerland

ARTICLE INFO

Article History:
Received September 9, 2022
Revision received January 3, 2023
Accepted April 20, 2023

Editor-in-Chief: Cara E. Stepp
Editor: Jun Wang

https://doi.org/10.1044/2023_JSLHR-22-00526

ABSTRACT

Purpose: Although articulatory impairment represents distinct speech characteristics in most neurological diseases affecting movement, methods allowing automated assessments of articulation deficits from the connected speech are scarce. This study aimed to design a fully automated method for analyzing dysarthria-related vowel articulation impairment and estimate its sensitivity in a broad range of neurological diseases and various types and severities of dysarthria.
Method: Unconstrained monologue and reading passages were acquired from 459 speakers, including 306 healthy controls and 153 neurological patients. The algorithm utilized a formant tracker in combination with a phoneme recognizer and subsequent signal processing analysis.
Results: Articulatory undershoot of vowels was presented in a broad spectrum of progressive neurodegenerative diseases, including Parkinson's disease, progressive supranuclear palsy, multiple-system atrophy, Huntington's disease, essential tremor, cerebellar ataxia, multiple sclerosis, and amyotrophic lateral sclerosis, as well as in related dysarthria subtypes including hypokinetic, hyperkinetic, ataxic, spastic, flaccid, and their mixed variants. Formant ratios showed a higher sensitivity to vowel deficits than vowel space area. First formants of corner vowels were significantly lower for multiple-system atrophy than cerebellar ataxia. Second formants of vowels /a/ and /i/ were lower in ataxic compared to spastic dysarthria. Discriminant analysis showed a classification score of up to 41.0% for disease type, 39.3% for dysarthria type, and 49.2% for dysarthria severity. Algorithm accuracy reached an F-score of 0.77.
Conclusions: Distinctive vowel articulation alterations reflect underlying pathophysiology in neurological diseases. Objective acoustic analysis of vowel articulation has the potential to provide a universal method to screen motor speech disorders.
Supplemental Material: https://doi.org/10.23641/asha.23681529

Imprecise vowels represent one of the core articulatory deficits contributing to reduced intelligibility due to dysarthria (H. Kim, Hasegawa-Johnson, & Perlman, 2011). Impairment of vowel articulation reflects reduced amplitude and velocity of articulators, including lips, tongue, and jaw (the so-called undershooting of articulatory gestures; Robertson & Hammerstad, 1996). Previous studies have documented the presence of vowel articulation abnormalities in a number of progressive neurological diseases (Whitfield, 2019), particularly in Parkinson's disease (PD; Lam & Tjaden, 2016; Skodda et al., 2011; Tjaden et al., 2013; Whitfield & Goberman, 2014; Whitfield & Mehta, 2019) and sporadically in progressive supranuclear palsy (PSP), multiple-system atrophy (MSA), Huntington's disease (HD), essential tremor (ET), cerebellar ataxia (CA), multiple sclerosis (MS), and amyotrophic lateral sclerosis (ALS; Rusz et al., 2014, 2015; Tjaden et al., 2005; Tykalova et al., 2016; Yunusova et al., 2013).

Correspondence to Jan Rusz: rusz.mz@gmail.com. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

In addition, distinctive progressive neurological diseases typically comprehend differing subtypes of dysarthria, with the most prevalent hypokinetic, hyperkinetic, spastic, ataxic, or flaccid variant (Duffy, 2019). These dysarthria subtypes reflect the underlying pathophysiology of the disease and may give us clues for differential diagnosis (Duffy, 2019). In some cases, such as PD, where most patients develop pure hypokinetic dysarthria (Ho et al., 1999), there is good correspondence between the type of disease and type of dysarthria. Contrary, the correspondence might be weaker in other cases as multiple dysarthria subtypes may occur for a single disease type due to more than one component of the motor system being affected. For instance, patients with atypical parkinsonism such as MSA or PSP typically manifest various combinations of hypokinetic, spastic, and ataxic dysarthria components (Rusz et al., 2015). However, previous studies have not addressed whether the vowel articulation impairment is differentially valuable by directly comparing several disease etiologies or dysarthria subtypes. Moreover, the previous evidence is limited due to the small sample sizes available and different methodologies used for analysis (Lam & Tjaden, 2016; Rusz et al., 2014, 2015; Skodda et al., 2011; Tjaden et al., 2005; Tykalova et al., 2016; Whitfield & Goberman, 2014; Whitfield & Mehta, 2019; Yunusova et al., 2013).

Additionally, dysarthria severity varies across neurological diseases depending on their stage and rate of progression (Y. Kim, Kent, & Weismer, 2011). In particular, higher dysarthria severity could be expected in disorders with faster disease progression (Rusz et al., 2015). Nevertheless, there is no standard measure of speech severity in dysarthria. Estimates of speech intelligibility are frequently used to estimate the extent to which neurological disease affects the speech mechanism (Y. Kim, Kent, & Weismer, 2011). Since the relationships between the severity of vowel articulation impairment and the perceptual impression of unintelligibility in dysarthric speakers have been widely documented (H. Kim, Hasegawa-Johnson, & Perlman, 2011; H. M. Liu et al., 2005; Weismer et al., 2001), automated vowel articulation analysis may have a potential to provide such a measure of speech severity in dysarthria. However, there is a lack of relevant vowel articulation studies with a sufficiently large number of dysarthric speakers on various levels of severity.

A reliable and automatic method applicable to natural, spontaneous speech without any cost or burden to the patient or investigator is necessary to facilitate the use of vowel articulation assessment in common clinical practice. The intelligibility and quality of each vowel can be determined particularly by the distinct acoustic energy peak of the first (F_1) and second (F_2) formant frequency. The acoustic–articulatory relationship is defined such that the F_1 frequency varies inversely with tongue height and the F_2 frequency varies directly with tongue advancement (Kent et al., 1999). The limited articulatory range of motion due to dysarthria may result in various shifts in formant frequencies; most typically, formants with naturally higher frequencies tend toward lower frequencies, whereas formants with naturally lower frequencies tend toward higher frequencies (Kent & Kim, 2003; Roy et al., 2009; Shimon et al., 2010). However, most current methods for evaluating vowel articulation via formants in dysarthrias rely on precise and time-consuming hand-labeling of predefined speech utterances (Shimon et al., 2010; Skodda et al., 2011). Only two attempts have been made to evaluate vowel articulation employing automated acoustic analysis (Y. Liu et al., 2021; Sandoval et al., 2013); these were limited by analysis of only predefined reading sentences obtained from a sample predominantly composed of healthy controls (HCs) and PD patients with mild severity of hypokinetic dysarthria.

Therefore, we aimed to design a fully automated method for analyzing vowel articulation impairment due to dysarthria via detecting formant frequencies from corner vowels. Based on this approach and a large sample of patients with various progressive neurological diseases, we quantitatively assessed the sensitivity of imprecise vowel articulation to different (a) types of neurological disease, (b) types of dysarthria, and (c) severity of dysarthria.

Method

Subjects

Each participant provided written informed consent. This study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic, in accordance with the ethical standards established in the 1964 Declaration of Helsinki.

Between 2011 and 2021, a total of 459 successive native Czech speakers with Central Bohemia accent were recruited for this study. Considering progressive neurodegenerative diseases, 20 patients with PD (10 women, 10 men; de-novo PD examined before antiparkinsonian treatment was started), 15 with PSP (five women, 10 men; 11 with Richardson's syndrome, two with PSP-parkinsonism, and two with PSP-pure akinesia with gait freezing), 20 with MSA (12 women, eight men; 17 with parkinsonian and three with cerebellar variant), 20 with HD (10 women, 10 men), 20 with ET (10 women, 10 men), 18 with CA (eight women, 10 men; 11 with sporadic late-onset CA other than MSA, seven with spinocerebellar ataxia [Type 1, 2, 7, or 8]), 20 with MS (11 women, nine men; 10 with relapsing–remitting MS, five with primary progressive

MS, five with secondary progressive MS), and 20 with ALS (14 women, six men) were recruited (see Table 1). All patients were examined by a neurologist with an experience in movement, demyelinating, or neuromuscular disorders. The diagnosis of PD was established by the Movement Disorders Society clinical diagnostic criteria (Postuma et al., 2015); PSP by the Movement Disorder society diagnostic criteria for PSP (Höglinger et al., 2017); MSA by the consensus diagnostic criteria for MSA (Gilman et al., 2008); HD by clinical and genetic testing (Huntington Study Group; 1996); ET by published clinical research criteria (Louis et al., 2007); CA by genetic testing or results of neurological, neuropsychological, and magnetic resonance imaging testing; MS by the revised McDonald Criteria (Thompson et al., 2018); and ALS according to the El Escorial Criteria from the World Federation of Neurology (Brooks et al., 2000). Additionally, 306 healthy subjects (158 women, 148 men) with a mean age of 59.1 (SD = 13.2, range: 31–87) years with no history of neurological or communication disorders participated as HCs to match the wide age and gender range of investigated neurodegenerative diseases.

Clinical Evaluation

The disease severity of PD was assessed according to the motor score of the Movement Disorders Society–Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III (Goetz et al., 2008), PSP and MSA by The Natural History and Neuroprotection in Parkinson Plus Syndromes–Parkinson Plus Scale (NNIPPS-PPS; Payan et al., 2011), HD by the motor score of the Unified Huntington's Disease Rating Scale (UHDRS; Huntington Study Group, 1996), ET by the Tremor Research Group Essential Tremor Rating Assessment Scale (TETRAS; Elble et al., 2012), CA by the Scale for the Assessment and Rating of Ataxia (SARA; Schmitz-Hübsch et al., 2006), MS by the Expanded Disability Status Scale (EDSS; Kurtzke, 1983), and ALS by the ALS Functional Rating Scale–Revised (ALSFRS-R; Cedarbaum et al., 1999). Disease duration was estimated based on the self-reported occurrence of the first motor symptoms.

Speech Examination

Each subject was recorded during a single session accompanied by a speech specialist who guided through the standardized protocol. No time limits were imposed during the recording. All participants were willing to cooperate and could repeat their performance if necessary. The participants were instructed to present a monologue about an arbitrary, emotionally neutral topic for at least 90 s (M = 128.3, SD = 27.5, range: 74–312). In addition, all subjects performed a reading passage task of a standardized text of 80 words (Supplemental Material S1). The same settings were applied to subjects in all groups. Speech recordings were performed in a quiet room with a low ambient noise level using a head-mounted condenser microphone (Beyerdynamic Opus 55) placed approximately 5 cm from the subject's mouth. Speech signals were sampled at 48 kHz with a 16-bit resolution.

Auditory–Perceptual Estimates of Dysarthria Presence, Type, and Severity

The dysarthria presence and type, including severity, were made by the consensus auditory–perceptual judgment of two speech-language pathologists with more than 10 years of experience in movement disorders who were aware of each patient's medical diagnosis. The judgment was based on offline audio recordings following the perceptual criteria outlined by Darley et al. (1969a, 1969b). The dysarthria types identified across eight neurological conditions included hypokinetic, hyperkinetic, ataxic, spastic, flaccid–spastic, spastic–ataxic, hypokinetic–spastic, hypokinetic–ataxic, and hypokinetic–spastic–ataxic (see Table 1). In addition, the severity of dysarthria was rated on a 4-point scale (0 = *none*, 1 = *mild*, 2 = *moderate*, 3 = *severe*). The lower average dysarthria severity with a dominant occurrence of mild dysarthria was observed only for PD and MS groups (see Table 1). Potential participants without the presence of perceptual severity of dysarthria, with the presence of language disorders or apraxia of speech, or with a speech dysfunction not related to the diagnosed neurological disorder were excluded from this study.

Automatic Algorithm for Vowel Articulation Features

The algorithm utilizes a formant tracker in combination with a phoneme recognizer and subsequent signal processing analysis (see Figure 1). It processes the connected speech utterance for reading passages and monologues separately, and estimates F_1 and F_2 formant values for each corner vowel /a/, /i/, and /u/. These corner vowels are essential to form a vowel triangle (i.e., triangular F_1–F_2 vowel space), which reflects extreme placements of the tongue. (H. Kim, Hasegawa-Johnson, & Perlman, 2011; Rusz et al., 2013; Skodda et al., 2011).

Formants and Phonemes Estimation (Step A)

The speech input was processed in parallel by a formant tracker and a phoneme recognizer (see Figure 1A). Burg algorithm (Childers, 1978) implementation in Praat (Boersma, 2001) was used for the first two formants contour estimation resulting in $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ vectors over the utterance. After the trial testing, the window length was

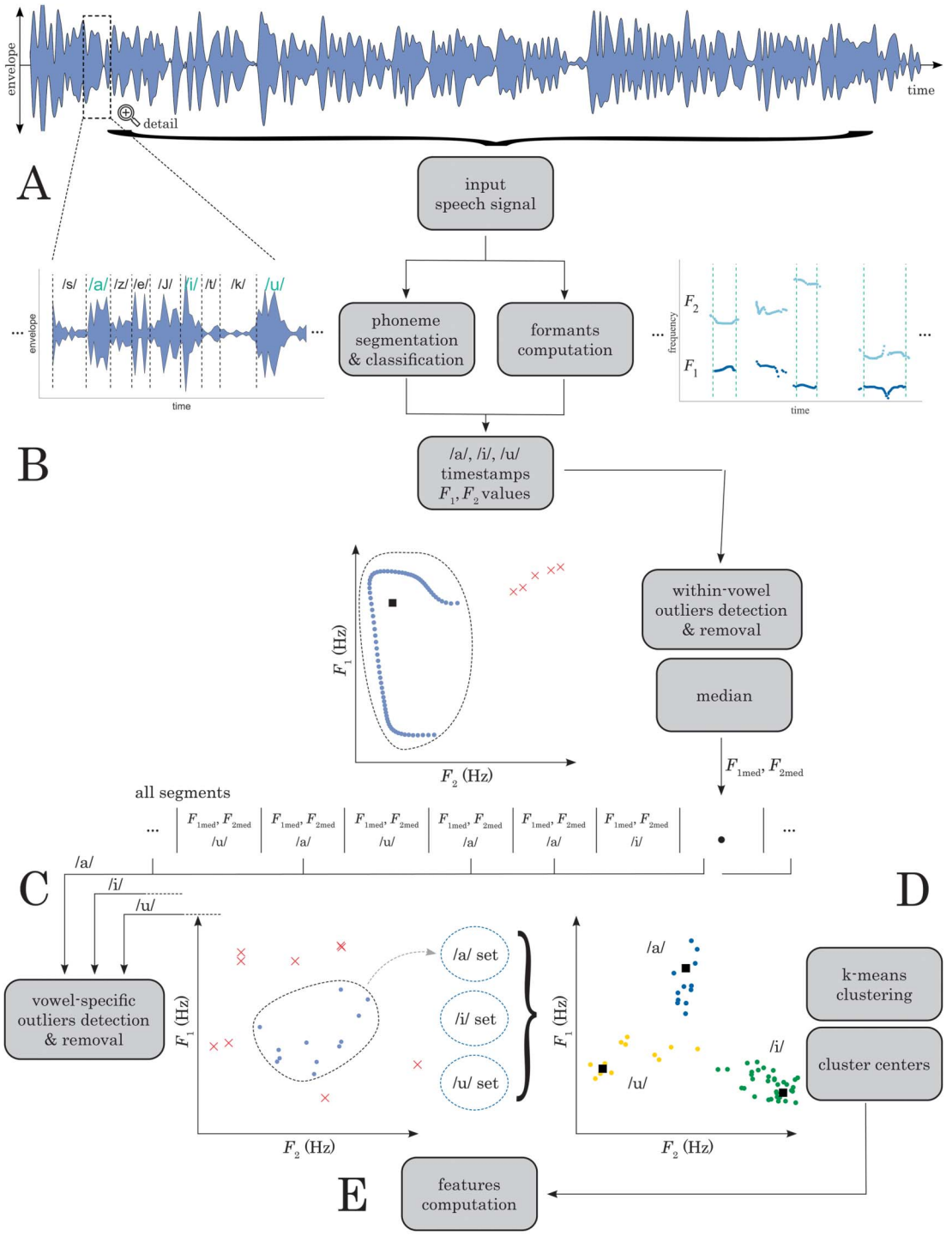
Table 1. Clinical characteristics of the investigated subjects.

Disease	Sex	Motor score (disease severity) *M/SD* (range)	Age (years) *M/SD* (range)	Symptom duration (years) *M/SD* (range)	Dysarthria type (auditory–perceptual)	Dysarthria severity (auditory–perceptual)
PD	F = 10	38.7/14.7[a]	63.5/8.9	1.6/1.3	Hypokinetic (*n* = 20)	Mild (*n* = 13)
	M = 10	(18–70)	(42–79)	(0.3–5.9)		Moderate (*n* = 7) Severe (*n* = 0) Mean severity: 1.35[h]
PSP	F = 5	65.7/28.9[b]	66.0/5.1	4.7/2.7	Hypokinetic (*n* = 3)	Mild (*n* = 3)
	M = 10	(19–132)	(54–71)	(2.0–11.0)	Hypokinetic–spastic (*n* = 4)	Moderate (*n* = 5)
					Hypokinetic–ataxic (*n* = 3)	Severe (*n* = 7) Mean severity: 2.27
					Hypokinetic–spastic–ataxic (*n* = 5)	
MSA	F = 12	79.1/21.1[b]	62.0/7.0	4.4/1.8	Hypokinetic (*n* = 3)	Mild (*n* = 1)
	M = 8	(35–115)	(45–73)	(2.0–7.5)	Spastic–ataxic (*n* = 1)	Moderate (*n* = 12)
					Hypokinetic–spastic (*n* = 8)	Severe (*n* = 7)
					Hypokinetic–ataxic (*n* = 3)	Mean severity: 2.30
					Hypokinetic–spastic–ataxic (*n* = 5)	
HD	F = 10	24.8/9.9[c]	53.1/11.0	5.2/3.6	Hyperkinetic (*n* = 20)	Mild (*n* = 1)
	M = 10	(8–42)	(34–69)	(1.0–16.0)		Moderate (*n* = 13)
						Severe (*n* = 6) Mean severity: 2.25
ET	F = 10	17.5/7.6[d]	64.3/11.1	28.9/17.5	Hyperkinetic (*n* = 18)	Mild (*n* = 5)
	M = 10	(6–35)	(40–82)	(3.0–60.0)	Hypokinetic (*n* = 1)	Moderate (*n* = 9)
					Spastic (*n* = 1)	Severe (*n* = 6) Mean severity: 2.05
CA	F = 8	13.9/4.8[e]	54.7/12.6	11.0/8.5	Ataxic (*n* = 5)	Mild (*n* = 5)
	M = 10	(4–24)	(34–72)	(0.5–28.0)	Spastic (*n* = 1)	Moderate (*n* = 7)
					Spastic–ataxic (*n* = 11)	Severe (*n* = 6) Mean severity: 2.06
					Hypokinetic–ataxic (*n* = 1)	
MS	F = 11	4.6/0.8[f]	52.2/10.1	17.8/8.6	Ataxic (*n* = 7)	Mild (*n* = 16)
	M = 9	(4–7)	(33–74)	(6.0–32.0)	Spastic (*n* = 3)	Moderate (*n* = 3)
					Spastic–ataxic (*n* = 10)	Severe (*n* = 1) Mean severity: 1.25[h]
ALS	F = 14	35.6/6.5[g]	62.1/11.1	1.9/1.2	Spastic (*n* = 4)	Mild (*n* = 5)
	M = 6	(22–45)	(37–85)	(0.5–5.0)	Flaccid–spastic (*n* = 16)	Moderate (*n* = 6)
						Severe (*n* = 9) Mean severity: 2.20
Total	F = 80		59.7/9.6	9.4/5.7	Hypokinetic (*n* = 27)	Mild (*n* = 49)
	M = 73		(33–85)	(0.3–60.0)	Hyperkinetic (*n* = 38)	Moderate (*n* = 62)
					Ataxic (*n* = 12)	Severe (*n* = 42)
					Spastic (*n* = 9)	Mean severity: 2.09
					Flaccid–spastic (*n* = 16)	
					Spastic–ataxic (*n* = 22)	
					Hypokinetic–spastic (*n* = 12)	
					Hypokinetic–ataxic (*n* = 7)	
					Hypokinetic–spastic–ataxic (*n* = 10)	

Note. PD = Parkinson's disease; F = female; M = male; PSP = progressive supranuclear palsy; MSA = multiple system atrophy; HD = Huntington's disease; ET = essential tremor; CA = cerebellar ataxia; MS = multiple sclerosis; ALS = amyotrophic lateral sclerosis; MDS-UPDRS = Movement Disorders Society–Unified Parkinson's Disease Rating Scale; NNIPPS-PPS = Natural History and Neuroprotection in Parkinson Plus Syndromes–Parkinson Plus Scale; UHDRS = Unified Huntington's Disease Rating Scale; TETRAS = Tremor Research Group Essential Tremor Rating Assessment Scale; SARA = Scale for the Assessment and Rating of Ataxia; EDSS = Expanded Disability Status Scale; ALSFRS-R = Amyotrophic Lateral Sclerosis Functional Rating Scale–Revised.

[a]MDS-UPDRS Part III total scale. [b]NNIPPS-PPS total scale. [c]UHDRS total scale. [d]TETRAS score scale. [e]SARA total scale. [f]EDSS total scale. [g]ALSFRS-R total scale. [h]This group was found to have significantly lower disease severity compared to PSP, MSA, HD, ET, CA, and ALS groups with $p < .01$.

Figure 1. Illustrative schema of the automated method for formants estimation. F_1 = first formant frequency; F_2 = second formant frequency; VSA = vowel space area; FRI = formant ratio index; SFRI = second formant ratio index.



$$F_{1a}, F_{2a}, F_{1i}, F_{2i}, F_{1u}, F_{2u}$$

$$\mathrm{VSA} = \tfrac{1}{2}|F_{1i}(F_{2a} - F_{2u}) + F_{1a}(F_{2u} - F_{2i}) + F_{1u}(F_{2i} - F_{2a})|$$

$$\mathrm{FRI} = \tfrac{F_{1a}+F_{1i}+F_{1u}+F_{2a}+F_{2i}}{F_{2u}} \quad \mathrm{SFRI} = \tfrac{F_{2a}+F_{2i}}{F_{2u}}$$

set to 50 ms with a 1-ms overlap and the formant ceiling was set to 5250 Hz, as these values achieved the best precision of the estimates. The maximum number of formants was set to 5 as recommended by the method documentation, even though only the first two formants were subsequently extracted. A phoneme recognizer was employed based on split temporal context feature extraction (Schwarz & Černocký, 2008), pretrained on the Czech version of the SpeechDat-E database (Pollak et al., 2000). The recognizer is available pretrained for several languages with an error rate of 24.2% for both Czech and English (Schwarz et al., 2022). The recordings were subsampled to 8 kHz beforehand to match the training data. The output is represented by recognized phonemes with timestamps marking the corresponding speech segment.

Outlier Detection Across Individual Phonemes Segments (Step B)

The consecutive phoneme segments were further analyzed (see Figure 1B). If the frame was classified as corresponding to either /a/, /i/, or /u/ vowel, the $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ values within were extracted. These might be burdened with formant tracker errors, and thus, an outlier analysis is performed in each segment. Outliers were identified and discarded based on Mahalanobis distance (Mahalanobis, 1936), which calculates the distance of a given point from a chosen distribution. For normally distributed data, the squared distance follows χ^2 distribution. The procedure consists of two phases and is as follows.

First, normalized versions of each formant vector were computed by extracting the mean and dividing by the standard deviation. Then, the Mahalanobis distance was computed between each point on the normalized [$\boldsymbol{F}_1$, $\boldsymbol{F}_2$] grid and χ^2 distribution with two degrees of freedom since we have two formant contours. If the distance was greater then $\chi^2(q)$, where q is a chosen quantile value, it was marked as an outlier. In the second phase, the nonoutlier points formed a new distribution, and Mahalanobis distance was calculated between previously identified outlier points. If the distance was less than $\chi^2(q + 0.1)$, the corresponding point was withdrawn from the outliers set. After the conducted trial testing, the value of q was set to 0.8, making the procedure more benevolent in the outlier decision. It achieved higher effectiveness yet not suffering a decrease in accuracy than choosing harsher settings, that is, lower values of q.

The procedure ensures that the extreme outliers are correctly recognized while preserving most of the information around the formant contour. From each segment, a median value was computed from the first and second formants of the nonnormalized, nonoutlier points resulting in $\boldsymbol{F}_{1\text{med}}$ and $\boldsymbol{F}_{2\text{med}}$ vectors over the whole utterance with information about the particular vowel on each index.

Outlier Detection Across All /A/, /I/, and /U/ Vowels (Step C)

The medians from the segments might still contain false values, for example, when the phoneme recognizer misclassifies a consonant as a vowel. Therefore, the $\boldsymbol{F}_{1\text{med}}$ and $\boldsymbol{F}_{2\text{med}}$ values are grouped to either /a/, /i/, or /u/ set, and another outlier analysis was performed in each group (see Figure 1C).

The procedure is the same as described in the previous section; however, the value of the quantile q is set to 0.5, making the method less benevolent to any deviations, which was found to provide more accurate outcomes while maintaining a reasonable throughput. The nonoutliers for each vowel were then put together for final cluster analysis.

Vowels Clustering (Step D)

The described method is still prone to error when the phoneme recognizer misclassifies the vowel as another, for example, /u/ as /i/. The misclassified vowel might have the formant frequencies close to the original one and thus will not be detected in the outlier analysis.

For this reason, the vowel points were partitioned using the k-means algorithm into three clusters representing the single vowels (see Figure 1D). The distance metric was set to square Euclidean distance, and initial cluster centroid positions were chosen as the maximum value of F_1 and median of F_2 of the vowel /a/ (hence, cluster /a/), the minimum of F_1 and maximum of F_2 of the vowel /i/ (hence, cluster /i/), and the minimum of F_1 and the minimum F_2 of the vowel /u/ (hence, cluster /u/). The resulting clusters /a/, /i/, and /u/ then consisted of [$\boldsymbol{F}_{1a}$, $\boldsymbol{F}_{2a}$], [$\boldsymbol{F}_{1i}$, $\boldsymbol{F}_{2i}$], and [$\boldsymbol{F}_{1u}$, $\boldsymbol{F}_{2u}$] points, respectively. The misclassified vowel should be included in its corresponding cluster in this process.

In the final step, one pair of $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ values was calculated from the points of each cluster. For the /a/ cluster, $\boldsymbol{F}_1$ was calculated as an upper (0.75) quantile of the $\boldsymbol{F}_{1a}$ values and F_2 as the median of the $\boldsymbol{F}_{2a}$ values. For the /i/ cluster, F_1 was computed as a lower (0.25) quantile of the $\boldsymbol{F}_{1i}$ values and F_2 as an upper quantile of the $\boldsymbol{F}_{2i}$ values. For the /u/ cluster, F_1 and F_2 were selected as lower quantiles of $\boldsymbol{F}_{1u}$ and $\boldsymbol{F}_{2u}$ values, respectively. The choice of the particular quantiles was designed to reflect the corner vowel characteristics (Y. Liu et al., 2021), and the values were tuned in pretesting to achieve maximum estimates precision.

Vowel Articulation Features (Step E)

The outcome of the process is the pair of F_1 and F_2 values for each vowel from which the vowel articulation features were derived (see Figure 1E). Subsequently, the most commonly used features that represent complex

vowel articulation characteristics are vowel space area (VSA) and measures representing various shifts in formant frequencies (Kent & Kim, 2003; Roy et al., 2009; Shimon et al., 2010; Skodda et al., 2011). VSA, expressed in Hz^2, was calculated using the Euclidean distances between the F_1 and F_2 coordinates of the corner vowels /a/, /i/, and /u/ in the triangular [F_1, F_2] vowel space as

$$\mathrm{VSA} = \frac{1}{2}|F_{1i}(F_{2a} - F_{2u}) + F_{1a}(F_{2u} - F_{2i}) + F_{1u}(F_{2i} - F_{2a})|. \quad (1)$$

Formant ratio index (FRI) reflects the shift in formant frequencies based on all corner vowels and can be expressed using the following formula (i.e., expected trend is lowering of F_{1a}, F_{1i}, F_{1u}, F_{2a}, and F_{2i} and rising of F_{2u} due to the presence of dysarthria):

$$\mathrm{FRI} = \frac{F_{1a} + F_{1i} + F_{1u} + F_{2a} + F_{2i}}{F_{2u}}. \quad (2)$$

Finally, the second formant ratio index (SFRI) reflects the shift of the second formants only and was computed using the following formula (i.e., expected trend is lowering of F_{2a} and F_{2i} and rising of F_{2u} due to the presence of dysarthria)

$$\mathrm{SFRI} = \frac{F_{2a} + F_{2i}}{F_{2u}}. \quad (3)$$

All analyses were conducted in MATLAB (MathWorks).

Reference Hand Labels

The hand-labeled reference values of F_1 and F_2 formant frequencies and time event of the vowel occurrence for each corner vowel were obtained from 20 randomly selected recordings of the reading passage (1,760 vowels; 660 vowels of /a/, 720 vowels of /i/, and 380 vowels of /u/) with the representative distribution regarding gender, etiology, and dysarthria severity (11 men and nine women; four HC, two PD, two PSP, two MSA, two HD, two ET, two CA, two MS, and two ALS speakers; four *none*, seven *mild*, five *moderate*, and four *severe dysarthria* severity). All corner vowels of /a/, /i/, and /u/ were selected; the position of the selected vowel for the reading passage is in bold in Supplemental Material S1. Formants were extracted according to widely accepted previously published methodology validated in several languages (Roy et al., 2009; Rusz et al., 2013; Shimon et al., 2010; Skodda et al., 2011); F_1 and F_2 frequencies were determined by employing a 30-ms segment at the temporal midpoint of the stable part of each vowel (in order to avoid the influence of vowels preceding or following). The corresponding timestamps including the start and end times of the segment were recorded. The formant frequencies were not possible to extract in 23 cases of /a/, 43 of /i/, and 77 of /u/ due to (a) coarticulation with other phonemes leading to the indistinct formants in the target band (68%), (b) coarticulation with other phonemes leading to too many formants in the target band (10%), (c) the word with target vowel is not pronounced properly (16%), and (d) the vowel duration is shorter than 30 ms (6%). All analyses were performed in the Praat software (Boersma, 2001) using both the combined wideband spectrographic display and the power spectral density.

Algorithm Performance Metrics

F-score was used as the primary outcome to assess the algorithm accuracy and was defined as

$$F = 2 \times \frac{\mathrm{precision} \times \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}, \quad (4)$$

where

$$\mathrm{precision} = \frac{\mathrm{true\ positives}}{\mathrm{true\ positives} + \mathrm{false\ positives}}, \quad (5)$$

and

$$\mathrm{recall} = \frac{\mathrm{true\ positives}}{\mathrm{true\ positives} + \mathrm{false\ negatives}}. \quad (6)$$

In other cases, normalized root-mean-square error (NRMSE) and the Spearman correlation coefficient r were utilized. The NRMSE enables a description of several variables by describing the error as a fraction of the observed variable range and is defined as

$$\mathrm{NRSME} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_l - x_i)^2}}{(\max(\hat{x}_l) - \min(\hat{x}_l))}, \quad (7)$$

where N is the number of utterances, $\hat{x}_i$ represents an estimated feature, and x_i is its respective reference. The r coefficient was computed as a nonparametric measure of the rank correlation between the estimated and reference values.

Algorithm Validation Steps

The algorithm incorporates several steps in its procedure. In each step, it can make a different type of error. Therefore, to uncover potential error sources for each step separately, a three-step validation that corresponds to the steps in algorithm design is provided: (a) validation of vowel identification via phoneme recognizer and independent parallel validation of formant values estimation via

formant tracker (the result of the algorithm's Step A), (b) validation of combined accuracy via phoneme recognizer and formant tracker based on outlier detection and vowel clustering (the result of the algorithm's mutual Steps B–D), and (c) validation of algorithm total accuracy via resulting formant features (the result of the algorithm's Step E).

First, the performances of the phoneme recognizer and the formant tracker (the result of the algorithm's Step A) were compared to reference hand labels. The validation was performed across the corner vowels of /a/, /i/, and /u/. The vowel from automatic recognition was searched for within the 30-ms segment corresponding to manual timestamps with a 5-ms tolerance for the start and the end. To evaluate the reliability of phoneme recognizer, the accuracy was evaluated in terms of F-score. *True positive* cases were if the vowel was correctly detected (e.g., /a/ was detected as /a/). *False positive* cases were if the vowel was incorrectly detected (e.g., /a/ was detected as /i/). *False negative* case was if the vowel was not detected (e.g., /a/ was missed). To evaluate the reliability of the formant tracker, a median formant frequency was calculated from automatically obtained formant estimates via a 30-ms window corresponding to the start and end of the hand-label timestamps. These medians were compared to reference hand values in terms of NRMSE and Spearman correlation.

Second, the accuracy of the combination of phoneme recognizer and formant tracker (the result of the algorithm's mutual Steps B–D) was validated using F-score; this evaluation corresponds with the mutual outlier detection and vowel class correction mechanism performed by the algorithm. *True positive* cases were if (a) the vowel was correctly detected (e.g., /a/ was detected as /a/), (b) the vowel was detected as another vowel but automatically corrected (e.g., /a/ was detected as /i/ but corrected back to /a/), and (c) formants were found impossible to estimate by both hand-labeling and automated detection (e.g., formants of /a/ were impossible to determine by hand labels and the automated algorithm was not able to estimate them as well). *False positive* cases were if (a) the vowel was incorrectly detected and not corrected (e.g., /a/ was detected as /i/ and not corrected), (b) the vowel was correctly detected but incorrectly reclassified to a different vowel (e.g., /a/ was detected as /a/ but corrected to /i/), and (c) vowel formants were found impossible to estimate by hand-labeling but were still calculated automatically (e.g., formants of /a/ were impossible to determine by hand labels but automated algorithm produced an estimate). *False negative* case was if the vowel was not detected (e.g., /a/ was missed).

Third, the final averaged formant estimates (i.e., one F_{1a}, F_{2a}, F_{1i}, F_{2i}, F_{1u}, and F_{2u} value per subject/speaking task), as well as complex formant features (i.e., one VSA, FRI, and SFRI value per subject/speaking task) by both automated (the result of the algorithm's step E) and manual analysis were compared using NRMSE and Spearman correlation.

Statistical Analysis

Data extracted from reading passages and monologues were analyzed separately; data related to monologues are presented within the article, whereas data for reading passages can be found in Supplemental Material S1. Data normality was verified via the Shapiro–Wilcoxon and Bartlett (equality of variance) tests. One-way analysis of covariance with post hoc Fisher's least significant difference test was applied to evaluate group differences. All analyses were controlled for age and sex (covariates); intergroup differences among diseases and dysarthria types were in addition controlled for dysarthria severity.

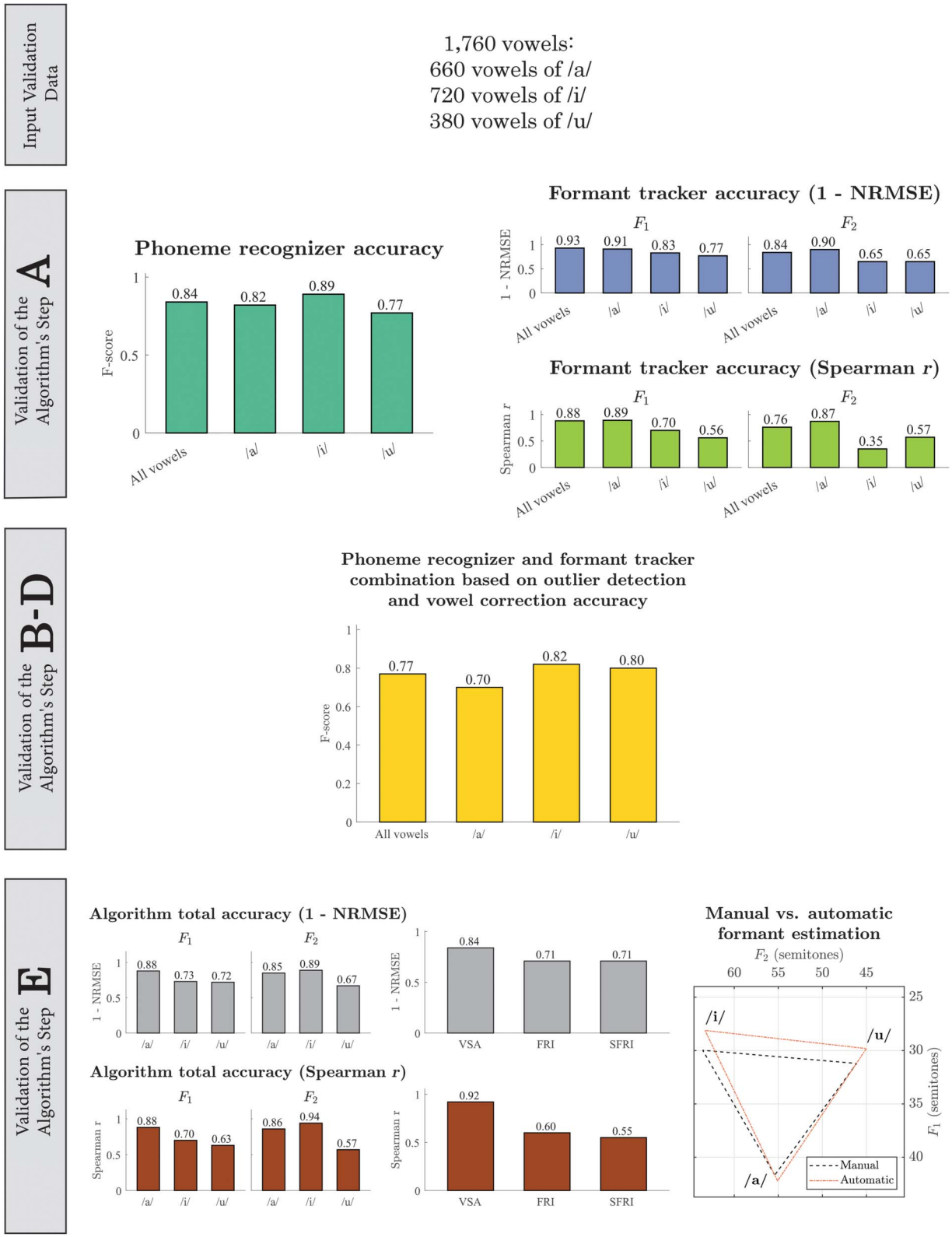
Prompted by primary hypothesis results, we performed a classification experiment based on the discriminant analysis followed by a leave-one-out cross-validation scheme to assess whether the vowel articulation features are best suited to differ between (a) type of neurological disease, (b) type of dysarthria, or (c) severity of dysarthria. In addition, to identify the probability of correct factor identification by chance, we generated a random vector of values ranging from 0 to 100 to substitute vowel articulation features across 459 hypothetical speakers; the average performance was calculated across 100 repetitions.

Results

Algorithm Performance

Compared to manual hand labels (based on 1,760 vowels), the phoneme recognizer attained an F-score of 0.84, whereas the formant tracker achieved 1-NRMSE of 0.93 for F_1 and 0.84 for F_2 across all vowels (see Figure 2, the results of the algorithm's Step A). After combining the error rate of the phoneme recognizer and formant tracker (based on 1,760 vowels), the F-score for all vowels was 0.77 (see Figure 2, the results of the algorithm's Steps B–D). Concerning the final averaged vowel articulation features (based on 20 utterances), the estimation of individual formants achieved 1-NRMSE of 0.88 for F_{1a}, 0.85 for F_{2a}, 0.73 for F_{1i}, 0.89 for F_{2i}, 0.72 for F_{1u}, and 0.67 for F_{2u}, leading to the 1-NRMSE of 0.84 for VSA, 0.71 for FRI, and 0.71 for SFRI (see Figure 2, the results of the algorithm's Step E). In summary, considering the final shape of vowel areas (see Figure 2, VSA plots), the most notable difference between automated and manual labels is due to

Figure 2. Illustrative scheme depicting step-by-step performance results between automated and manual analysis based on 1,760 hand-labeled vowels. NRMSE = normalized root-mean-square error; F_1 = first formant frequency; F_2 = second formant frequency; VSA = vowel space area; FRI = formant ratio index; SFRI = second formant ratio index.



lower estimates of F_1 frequencies of vowel /i/ and /u/ and F_2 of /u/ by the automated approach.

Effect of Neurological Disease Type

Compared to controls, the change in vowel articulation due to neurodegeneration in monologues was primarily demonstrated by trends toward the shift of formants across vowels /i/ and /u/, including an increase in F_{2u} and decrease in F_{1i}, F_{1u}, and F_{2i} frequencies across PD, PSP, MSA, HD, and ALS (see Figure 3 and Table 2). Among diseases, MSA tended to decrease F_1 and CA tended to increase F_1 compared to other neurological diseases, leading to a significantly lower F_1 for MSA than CA across all corner vowels (see Figure 4). Considering complex formant measures, compared to controls, VSA was significantly decreased for MSA, whereas FRI and SFRI were decreased for all neurological diseases except ET and MS (see Figure 5).

Effect of Dysarthria Type

Compared to controls, the trends toward the shift of formants across vowels /i/ and /u/ including increase in F_{2u} and decrease in F_{1i}, $F_{1u,}$ and F_{2i} frequencies in monologues were demonstrated mainly for hypokinetic and hyperkinetic dysarthria, mixed dysarthrias involving hypokinetic components, and flaccid–spastic subtype (see Figure 6 and Table 3). Among dysarthrias, there was a particular difference between ataxic dysarthria manifested by the decrease of F_{1a}, F_{2a}, and F_{2i} compared to spastic dysarthria (and its mixed variants with ataxic and flaccid elements) and in addition by a trend toward increase of F_{1u} to hypokinetic dysarthria (see Figure 7). Additionally, spastic–ataxic dysarthria showed a trend toward increase of F_{1a}, F_{1i}, and F_{1u} compared to hypokinetic dysarthria (and its mixed variants with spastic elements).

Considering complex formant measures, compared to controls, VSA was significantly decreased for ataxic and hypokinetic–spastic dysarthria (see Figure 8). FRI was decreased for hypokinetic, hyperkinetic, ataxic, flaccid–spastic, spastic–ataxic, hypokinetic–spastic, and hypokinetic–spastic–ataxic dysarthria. Finally, SFRI was decreased for hypokinetic, hyperkinetic, ataxic, flaccid–spastic, spastic–ataxic, and hypokinetic–spastic. Among dysarthrias, VSA of ataxic dysarthria was significantly lower than in spastic or spastic–ataxic dysarthria. FRI and SFRI of hypokinetic–spastic dysarthria were lower compared to hyperkinetic, flaccid–spastic, spastic–ataxic, and hypokinetic–ataxic dysarthria.

Effect of Dysarthria Severity

Compared to controls, the shift of formants across vowels /i/ and /u/ in dependence on auditory–perceptual dysarthria severity in monologues was observed, including an increase in F_{2u} and a decrease in F_{1i}, F_{1u}, and F_{2i} frequencies (see Figure 9 and Table 4). Considering complex formant measures, both measures of FRI and SFRI were reduced across all dysarthria severities (see Figure 10).

Classification Analysis

The classification analysis among vowel articulation features in monologues manifested accuracy of up to 39.7% for disease type, 37.3% for dysarthria type, and 49.2% for dysarthria severity (see Table 5); the probability of correct factor identification by chance using a random vector showed 5.3% accuracy for disease type, 4.2% for dysarthria type, and 19.8% for dysarthria severity. Acoustic metrics reflecting the shift in formant frequencies of FRI and SFRI were more sensitive to capturing the change of vowel articulation than VSA.

Effect of Speaking Task Type

The trends toward decrements in complex measures of VSA, FRI, and SFRI in reading passages were demonstrated similarly to those observed in monologues, except for the PD group where imprecise vowels articulation was not affected in reading passages (Supplemental Material S1); the classification experiment showed similar accuracy of up to 41.0% for disease type, 39.3% for dysarthria type, and 47.4% for dysarthria severity.

Discussion

This study is the first to demonstrate a fully automated objective approach to assessing the quality of vowel articulation in a large cohort of 459 speakers, including controls and patients with various neurological diseases and different types and severity of dysarthria, using the natural, unconstrained speech recordings. Based on complex formant measures, we showed that imprecise vowel articulation was presented in a broad spectrum of progressive neurodegenerative diseases, including PD, PSP, MSA, HD, ET, CA, MS, and ALS. Similarly, vowel articulation impairment was presented in all dysarthria subtypes such as hypokinetic, hyperkinetic, ataxic, spastic, and their mixed variants, including the flaccid–spastic subtype. In addition, the extent of vowel articulation impairment was influenced by dysarthria severity. However, we still observed divergent patterns of vowel articulation abnormalities across certain etiologies and dysarthria types independent of dysarthria severity. F_1 of all corner vowels were significantly lower for MSA than CA. In addition, F_2 of vowel /a/ and /i/ was lower in ataxic compared to

Figure 3. Corner vowel production triangles estimated from monologues for individual neurological disease types compared to healthy controls. The arrows indicate significant differences in the values to healthy controls adjusted by age and sex, with three, two, and one arrows referring to $p < .001$, $p < .01$, and $p < .05$, respectively. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). F_1 = first formant frequency; F_2 = second formant frequency; PD = Parkinson's disease; PSP = progressive supranuclear palsy; MSA = multiple system atrophy; HD = Huntington's disease; ET = essential tremor; CA = cerebellar ataxia; MS = multiple sclerosis; ALS = amyotrophic lateral sclerosis.

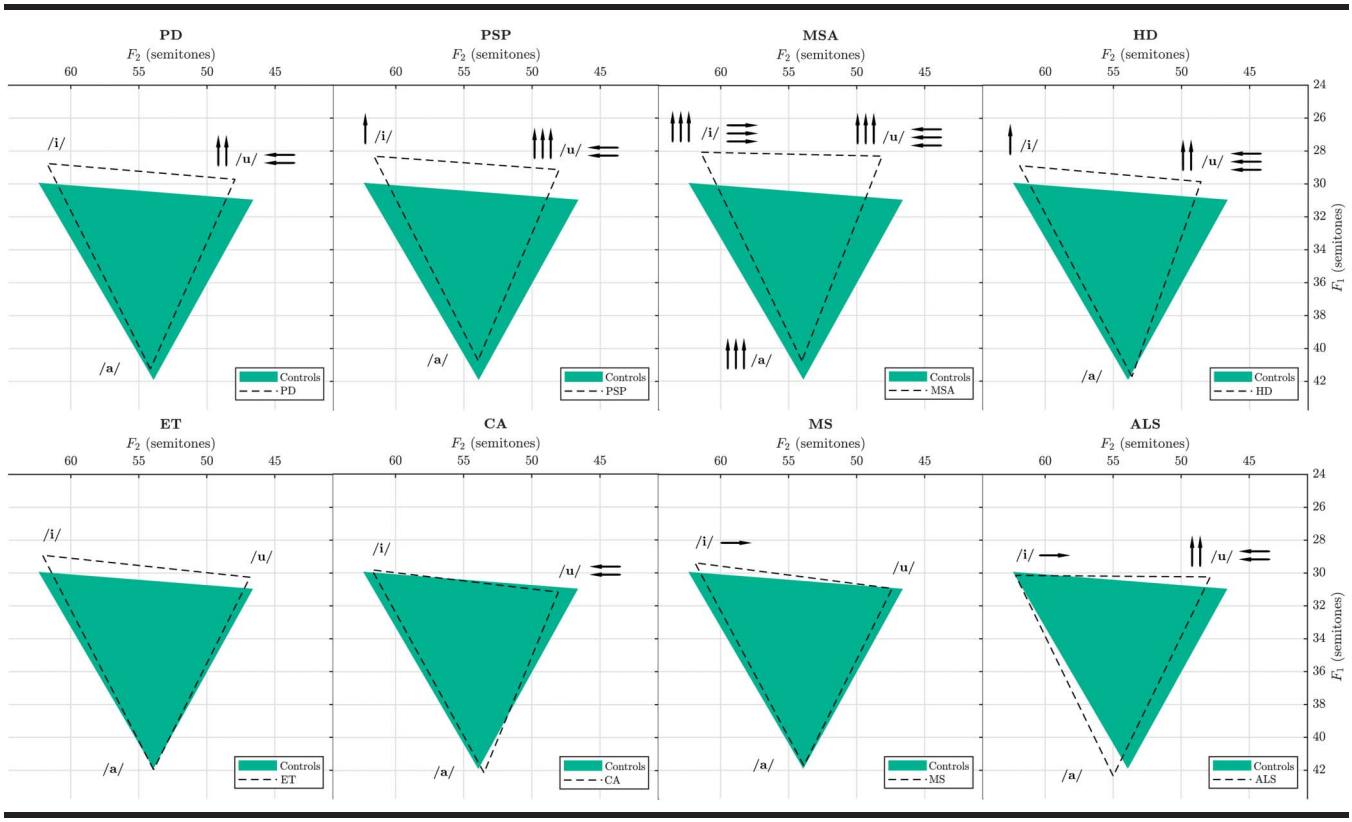
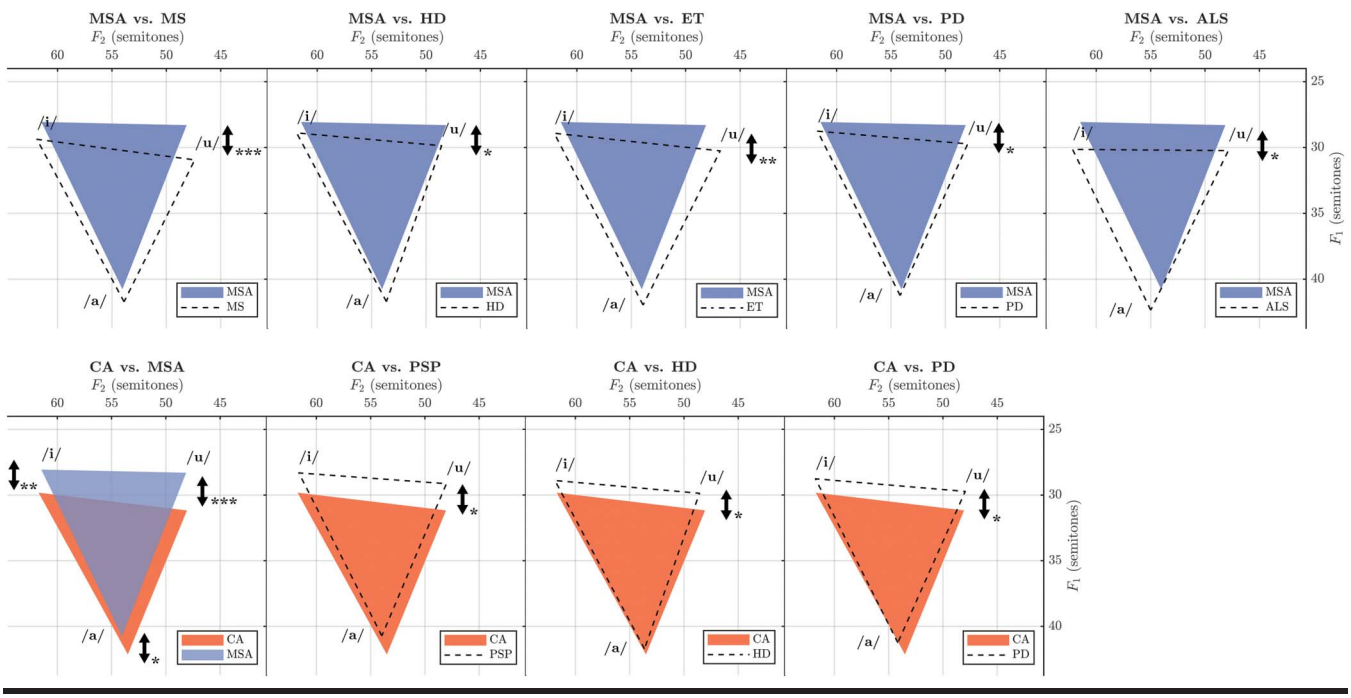


Table 2. Formant frequencies of corner vowels estimated from monologues for individual neurological disease types compared to healthy controls.

Neurological disease type	/a/ *M* (*SD*) Semitones		/i/ *M* (*SD*) Semitones		/u/ *M* (*SD*) Semitones	
	F_1	F_2	F_1	F_2	F_1	F_2
Controls	41.91 (2.6)	53.92 (2.1)	29.93 (2.3)	62.37 (1.6)	30.96 (1.6)	46.60 (1.8)
PD	41.25 (2.9)	54.16 (2.9)	28.76 (2.9)	61.75 (1.6)	29.72 (2.5)	47.95 (1.7)
PSP	40.72 (4.0)	53.97 (1.5)	28.32 (2.6)	61.61 (1.8)	29.13 (2.3)	48.02 (2.3)
MSA	40.76 (3.1)	54.03 (2.0)	28.06 (2.2)	61.48 (1.7)	28.30 (2.5)	48.13 (2.8)
HD	41.71 (3.2)	53.64 (2.0)	28.89 (2.6)	61.88 (1.7)	29.86 (2.3)	48.57 (2.6)
ET	41.97 (3.5)	53.91 (1.7)	28.92 (2.5)	62.08 (1.7)	30.26 (1.7)	46.83 (1.9)
CA	42.14 (3.1)	53.50 (1.7)	29.82 (1.6)	61.72 (1.5)	31.17 (1.6)	48.05 (1.9)
MS	41.72 (2.9)	53.88 (1.8)	29.38 (2.5)	61.89 (1.5)	30.94 (1.7)	47.39 (2.0)
ALS	42.35 (3.8)	54.99 (1.2)	30.15 (2.2)	62.17 (1.4)	30.24 (2.0)	47.86 (2.2)

Note. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). Hertz to semitone formula: f(semitone) = 12*((log*f(Hz)/60)/log(2)). F_1 = first formant frequency; F_2 = second formant frequency; PD = Parkinson's disease; PSP = progressive supranuclear palsy; MSA = multiple system atrophy; HD = Huntington's disease; ET = essential tremor; CA = cerebellar ataxia; MS = multiple sclerosis; ALS = amyotrophic lateral sclerosis.

Figure 4. Corner vowel production triangles estimated from monologues across two pairs of neurological disease types. The double-headed arrows indicate significant differences across diseases adjusted by age, sex, and dysarthria severity with ***, **, * referring to $p < .001$, $p < .01$, and $p < .05$. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). F_1 = first formant frequency; F_2 = second formant frequency; MSA = multiple system atrophy; MS = multiple sclerosis; HD = Huntington's disease; ET = essential tremor; PD = Parkinson's disease; ALS = amyotrophic lateral sclerosis; CA = cerebellar ataxia; PSP = progressive supranuclear palsy.

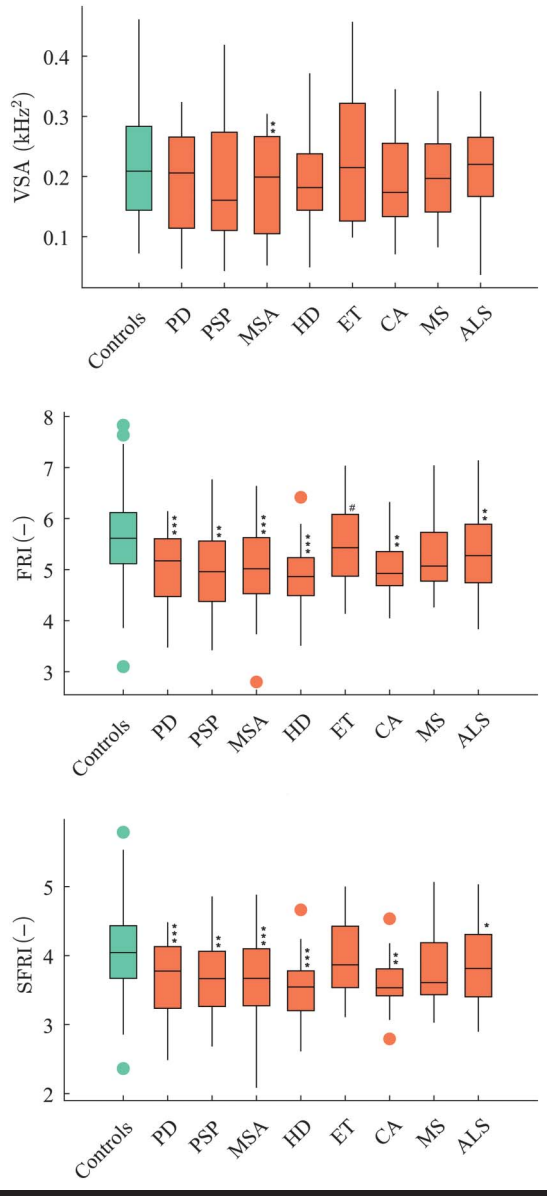


spastic dysarthria. Therefore, objective analysis of vowel articulation has the potential to provide a universally applicable method to screen neurological diseases affecting movement abilities that can be obtained from everyday speech without any cost or burden to the patient and investigator. In the future, vowel articulation deficits could be analyzed via smartphone calls (Kouba et al., 2022), thus significantly aiding in improving innovative neuroprotective therapies' stratification and monitoring effect.

Effect of Neurological Disease Type

Our results confirmed that vowel articulation impairment is exhibited in multiple types of neurological diseases. This finding follows previous acoustic studies (Rusz et al., 2014, 2015; Tjaden et al., 2005; Tykalova et al., 2016; Yunusova et al., 2013), although these have described vowel articulation pertaining to a specific disease rather than comparing these characteristics across diseases. In fact, the presence of vowel articulation deficits across various neurological diseases is not surprising because articulatory impairments represent the most common and distinct characteristics of most dysarthrias (Darley et al., 1969b). However, in this study, certain disease-specific patterns of imprecise vowel articulation have been observed. In general, vowel articulation impairment appeared to be more pronounced in parkinsonian disorders and HD. We might thus assume that the greater extent of vowel articulatory deficits due to tongue movement restriction, reflected mainly by the decrease of F_{2i} and the increase of F_{2u}, is associated with bradykinesia, which represents a common motor sign not only in parkinsonism but also in HD (Reilmann, 2019). Indeed, the previous study on a rat PD model has shown that even unilateral deficits to the nigrostriatal dopamine system leading to bradykinesia substantially contribute to tongue movement restriction responsible for imprecise vowel articulation (Ciucci et al., 2011). Interestingly, the parallel decrease in F_1 of all corner vowels was able to statistically separate MSA from CA even after adjustment for dysarthria severity, presumably as a consequence of damage to basal ganglia structures in addition to cerebellar dysfunction that is typical in both diseases. This finding might have important clinical implications as the differentiation of the cerebellar variant of MSA from idiopathic late-onset CA early in the disease course remains a major diagnostic challenge (Lin et al., 2016). However, although the extent of articulatory disorder appeared to be similar for both MSA subtypes (Rusz et al., 2019), the

Figure 5. Statistically significant group differences for estimated articulation features in monologues among the different types of neurological disease types compared to healthy controls adjusted by age and sex with ***, **, * referring to $p < .001$, $p < .01$, and $p < .05$, respectively. # indicates significant differences to MSA ($p < .05$) after adjusting for age, sex, and dysarthria severity. Middle bars represent median, and rectangles represent the interquartile range. Maximum and minimum values are by error bars. Outliers are marked as dots. VSA = vowel space area; PD = Parkinson's disease; PSP = progressive supranuclear palsy; MSA = multiple system atrophy; HD = Huntington's disease; ET = essential tremor; CA = cerebellar ataxia; MS = multiple sclerosis; ALS = amyotrophic lateral sclerosis; FRI = formant ratio index; SFRI = second formant ratio index.



utility of vowel articulation analysis as such a potential diagnostic marker has to be verified in future studies as the current sample was composed dominantly of the parkinsonian variant of MSA.

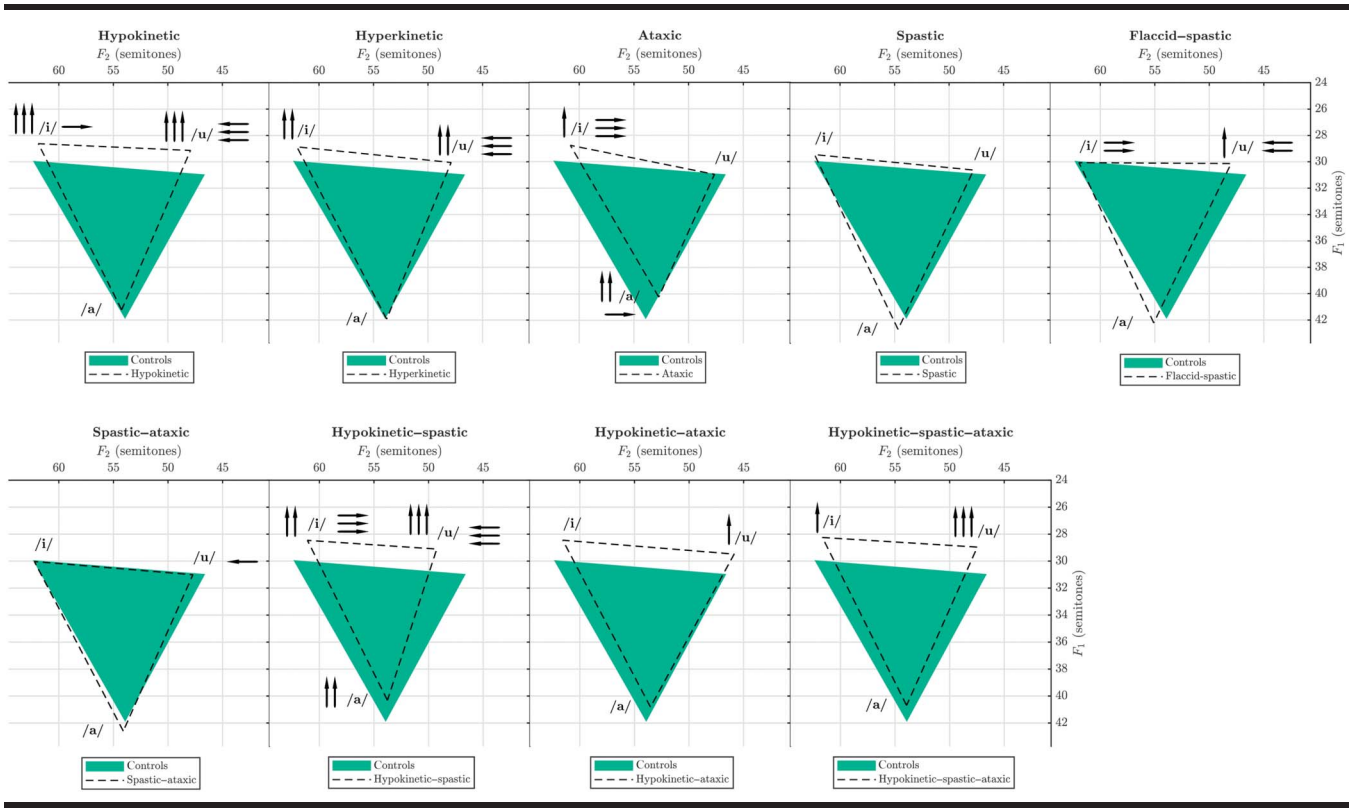
Effect of Dysarthria Type

In line with findings across multiple types of neurological diseases, vowel articulation impairment was observed across various dysarthria types. This result is not surprising as dysarthria type is frequently linked with disease type, and vowel articulation impairment was found in all investigated etiologies. The shift in formants for vowels /i/ and /u/ showed strong similarities for all investigated dysarthrias. This finding follows previous research demonstrating that complex formant-based measures are not sensitive to distinguishing between dysarthria subtypes (Lansford & Liss, 2014). However, one potential phenomenon that might be helpful in the differential diagnosis of dysarthrias is a shift in vowel frequencies for vowel /a/. While both formants of vowel /a/ remained relatively unchanged in patients with hypokinetic or hyperkinetic dysarthria, they tend to be decreased in ataxic dysarthria and increased in spastic dysarthria (as well as in mixed dysarthrias involving spastic elements). However, this finding should be interpreted with caution due to the relatively low number of samples for pure spastic and ataxic speakers in this study. Although the studies on vowel articulation in spastic and ataxic dysarthrias are rare, the shift toward higher vowel /a/ formants in spastic dysarthria seems to align with previous research on patients with poststroke spastic dysarthria (Ge et al., 2021; Mou et al., 2018). This shift might be hypothesized as a consequence of spasticity or weakness of tongue muscles, leading to lower tongue advancement. In addition, a decrease of F_2 for vowel /a/ has been previously reported in patients with spinocerebellar ataxia (Skodda et al., 2014), which might be hypothesized to be a result of inconsistency over the range of tongue movement (Saigusa et al., 2006).

Effect of Dysarthria Severity

Our findings showed that auditory–perceptual dysarthria severity was another factor contributing to the extent of vowel articulation impairment. The result agrees with previous research demonstrating a strong relationship between vowel formant measures and perceptual ratings of dysarthria severity (Fletcher et al., 2017). Further support comes from a recent study that showed a progressive pattern of vowel articulation impairment from the prodromal stages of parkinsonism (Skrabal et al., 2022). Compared to VSA, formant indexes were more effective in capturing dysarthria severity, which follows the previous study showing that vowel articulation index based on changes in individual formants was more stable and reliable over repeated assessments compared to VSA (Caverlé & Vogel, 2020). The effectiveness of formant indexes in contrast to the low sensitivity of VSA suggests that

Figure 6. Corner vowel production triangles estimated from monologues for different dysarthria types compared to healthy controls. The arrows indicate significant differences in the values to healthy controls adjusted by age and sex, with three, two, and one arrows referring to $p < .001$, $p < .01$, and $p < .05$, respectively. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). F_1 = first formant frequency; F_2 = second formant frequency.



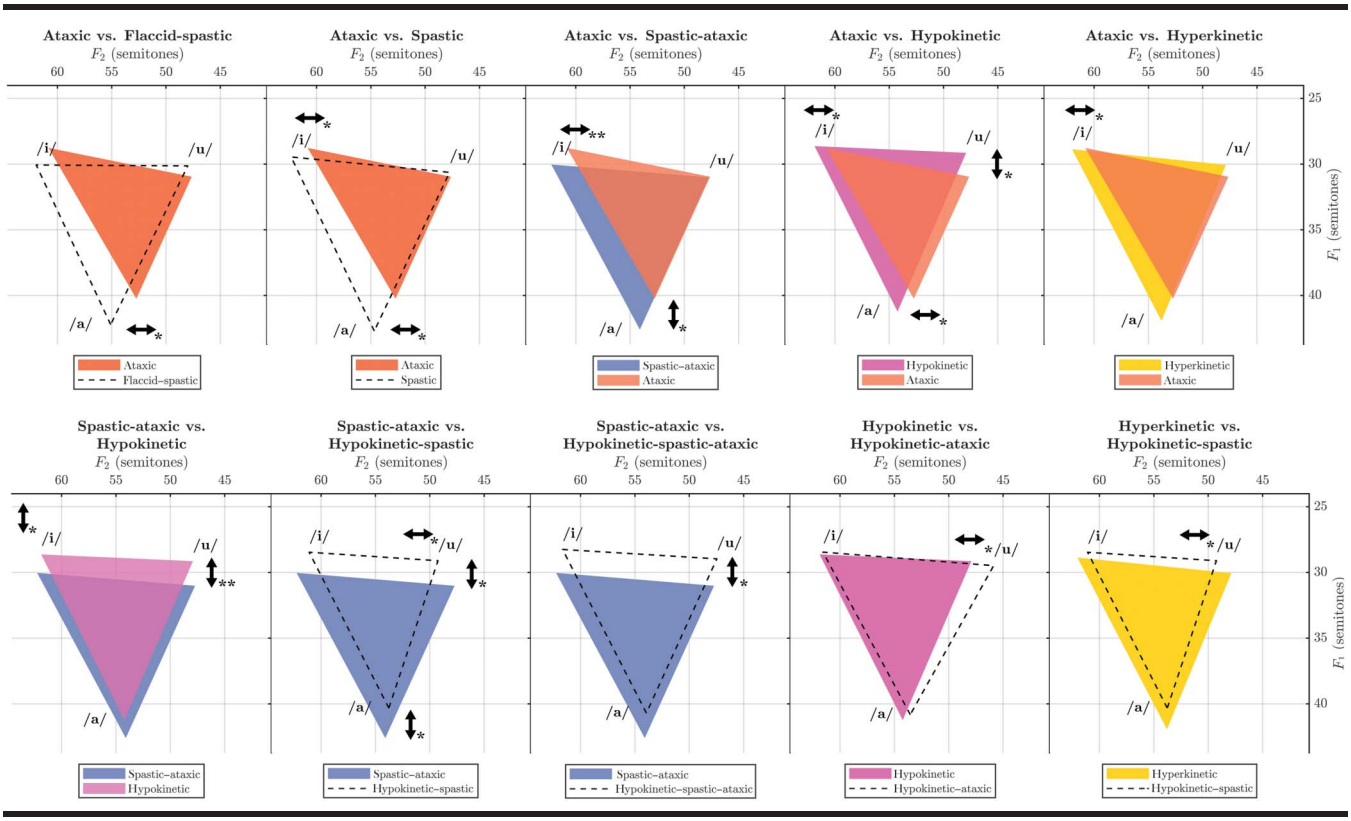
articulatory deficits are due mainly to alterations of the vowel /u/, followed by the vowel /i/, with the vowel /a/ remaining most resistant to change due to dysarthria. This behavior might be a result of different tongue positions and lip posture during individual corner vowels production, where the tongue is positioned low for the vowel /a/, high and forward for the vowel /i/, and high and backward for the vowel /u/, whereas lip posture is spread for

Table 3. Formant frequencies of corner vowels estimated from monologues for different dysarthria types compared to healthy controls.

	/a/ *M* (*SD*) Semitones		/i/ *M* (*SD*) Semitones		/u/ *M* (*SD*) Semitones	
Dysarthria type	F_1	F_2	F_1	F_2	F_1	F_2
Controls	41.91 (2.6)	53.92 (2.1)	29.93 (2.3)	62.37 (1.6)	30.96 (1.6)	46.60 (1.8)
Hypokinetic	41.23 (2.9)	54.24 (2.7)	28.63 (2.8)	61.87 (1.5)	29.14 (2.7)	47.92 (2.3)
Hyperkinetic	41.93 (3.3)	53.79 (1.9)	28.87 (2.6)	62.02 (1.7)	30.05 (2.0)	47.88 (2.3)
Ataxic	40.24 (2.9)	52.73 (2.0)	28.76 (2.2)	60.81 (1.3)	30.96 (1.8)	47.64 (1.6)
Spastic	42.71 (3.5)	54.68 (1.1)	29.45 (2.3)	62.34 (1.7)	30.63 (1.4)	47.75 (2.8)
Flaccid–spastic	42.22 (4.0)	55.09 (1.3)	30.06 (2.3)	61.95 (1.2)	30.13 (2.1)	48.00 (2.3)
Spastic–ataxic	42.60 (2.4)	54.10 (1.4)	30.03 (2.3)	62.24 (1.4)	31.01 (1.9)	47.7 (1.9)
Hypokinetic–spastic	40.32 (3.8)	53.77 (2.0)	28.46 (2.0)	61.09 (1.7)	29.11 (2.8)	49.25 (2.8)
Hypokinetic–ataxic	40.82 (3.1)	53.54 (2.2)	28.45 (1.7)	61.57 (1.7)	29.48 (1.5)	45.87 (1.6)
Hypokinetic–spastic–ataxic	40.70 (3.9)	53.94 (1.5)	28.25 (2.6)	61.65 (1.9)	28.96 (2.1)	47.46 (2.1)

Note. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). Hertz to semitone formula: f(semitone) = 12*((log*f(Hz)/60)/log(2)). F_1 = first formant frequency; F_2 = second formant frequency.

Figure 7. Corner vowel production triangles estimated from monologues across two pairs of dysarthria types. The double-headed arrows indicate significant differences across dysarthria types adjusted by age, sex, and dysarthria severity with ***, **, * referring to $p < .001$, $p < .01$, and $p < .05$, respectively. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). F_1 = first formant frequency; F_2 = second formant frequency.



both the /a/ and /i/ vowels and rounded for the vowel /u/ (Hasegawa-Johnson et al., 2003). Therefore, we might assume that the production of the vowel /a/ is less demanding than the production of the vowels /i/ and /u/. Moreover, in comparison to the vowel /i/, the articulation of the vowel /u/ requires more challenging involvement of the orofacial muscles to produce and maintain a tightly rounded lip posture (Hasegawa-Johnson et al., 2003), and its restrictions might also be linked to swallowing deficits in dysarthria (Sapir et al., 2008; Tjaden, 2008).

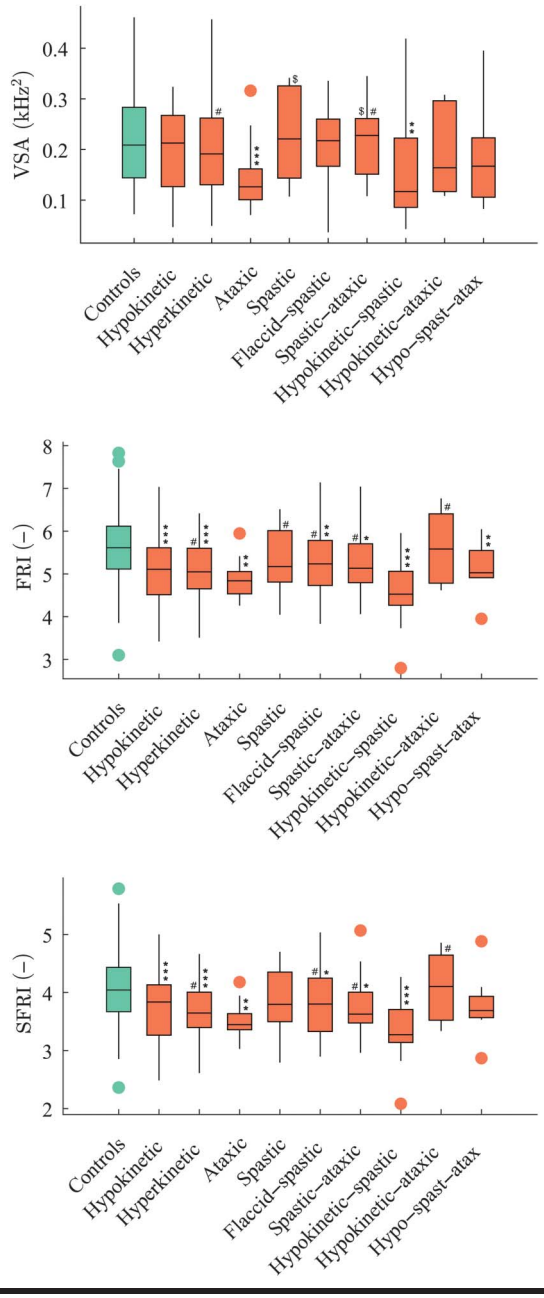
Effect of Speaking Task Type

The results showed that both monologue and reading speech are appropriate for assessing articulation deficits in neurological diseases with similar sensitivity. One notable difference was that only reading passages showed a significant difference between dysarthria severities, although the classification accuracy for dysarthria severity across both tasks was similar. We may thus hypothesize that standardized reading passages might be a better speaking material if capturing speech progression via vowel articulation is the primary endpoint. Considering individual diseases, the only evidence available is from PD, where previous studies have reported the occurrence of a more notable alteration of vowel articulation performance in spontaneous speech compared to nonspontaneous speech (Kempler & Lancker, 2002; Rusz et al., 2013; Weismer, 1984). Indeed, PD was the only group in this study that showed considerably better performance in vowel articulation in reading than monologues. In fact, persons with PD are often highly intelligible in prepared utterances but significantly less intelligible in spontaneous speech, whereas persons with other types of neuromotor disease might be equally intelligible in both forms of utterance (Y. Kim, Kent, & Weismer, 2011). Therefore, this finding might have important implications for future clinical trials in which PD participants should be assessed via spontaneous speech if vowel articulation represents an outcome measure.

Which of the Factors Most Contributes to the Vowel Articulation Impairment?

One of this study's goals was to answer whether vowel articulation impairment is most sensitive to disease

Figure 8. Statistically significant group differences for estimated articulation features in monologues among the different dysarthria types compared to healthy controls adjusted by age and sex with ***, **, * referring to $p < .001$, $p < .01$, and $p < .05$, respectively. # indicates significant differences to hypokinetic–spastic dysarthria ($p < .05$), whereas $ indicates significant differences to ataxic dysarthria ($p < .05$) after adjusting for age, sex, and dysarthria severity. Middle bars represent median, and rectangles represent the interquartile range. Maximum and minimum values are by error bars. Outliers are marked as dots. VSA = vowel space area; FRI = formant ratio index; SFRI = second formant ratio index; Hypo-spast-atax = Hypokinetic–spastic–ataxic dysarthria.

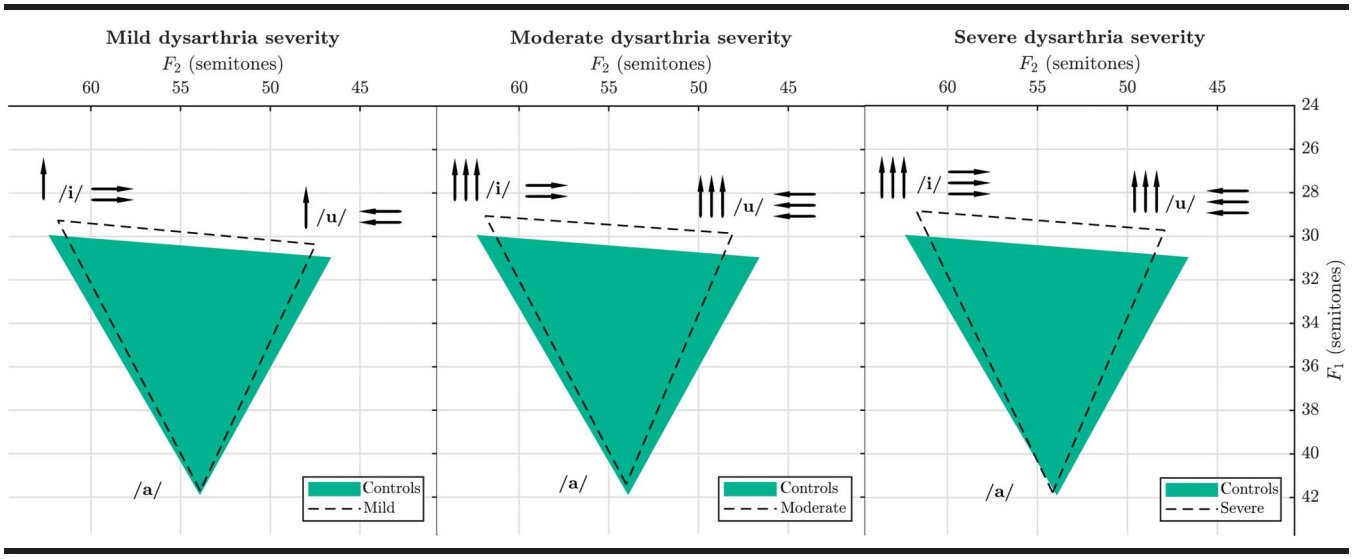


type, dysarthria type, or dysarthria severity. The discriminant analysis classification showed a score of up to 41.0% for the type of neurological disease, 39.3% for dysarthria type, and 49.2% for dysarthria severity. One might thus assume that vowel impairment appears to be more distinctive to dysarthria severity compared to a specific diagnosis of disease or dysarthria subtype. However, these results need to be put in context with the probability of correct factor identification by chance, which showed 5.3% accuracy for disease type, 4.2% for dysarthria type, and 19.8% for dysarthria severity (i.e., approximately equal to the number of groups across each investigated factor). Bearing this in mind, the best ratio between correct classification and identification by chance could be obtained for dysarthria type, although none of the three factors gained superior classification performance. Despite some differences observed in this study that might contribute to the differential diagnosis of dysarthria or disease etiology, we may assume that imprecise vowels represent a universal sign of articulatory disorder showing severity-related variations within several different types of dysarthria. This finding is perhaps not surprising as acoustic similarity across etiologies and types of dysarthria has already been assumed not only for vowel space but also for other acoustic measures such as speaking rate or voice onset time (Weismer, 2006). Indeed, accumulating evidence supports the view that various neuropathologies might similarly affect neuromotor control of speech production, leading to similar manifestations for certain speech aspects at the acoustic surface (Y. Kim, Kent, & Weismer, 2011). On the other hand, the combination of vowel articulation characteristics with other distinct cues that are pathognomic for a specific type of dysarthria, such as strained-strangled voice, slow rate, and reduced loudness variability in spastic dysarthria or normal rate and excessive loudness variability in ataxic dysarthria, might considerably increase correct classification to dysarthria type or disease etiology.

Algorithm Performance

Although articulatory deficits represent the main speech impairment characteristic of most dysarthrias, automated methods for assessing articulatory deficits from connected speech are scarce. In this study, we provided a fully automated approach to assessing the "undershoot of vowels" applicable across various neurological diseases, different dysarthrias, and a wide range of severity, from healthy speech to severe dysarthria. In particular, there are two main sources of errors including incorrect phoneme recognition (16% error based on 1–F-score) and incorrect formant tracking (7% error for F_1 and 16% error for F_2 based on NRMSE). However, the combination of both these error sources leads to an even lower accuracy of the algorithm. Therefore, to provide reliable vowel

Figure 9. Corner vowel production triangles estimated from monologues for different dysarthria severities compared to healthy controls. The arrows indicate significant differences in the values to healthy controls adjusted by age and sex, with three, two, and one arrows referring to $p < .001$, $p < .01$, and $p < .05$, respectively. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). F_1 = first formant frequency; F_2 = second formant frequency.



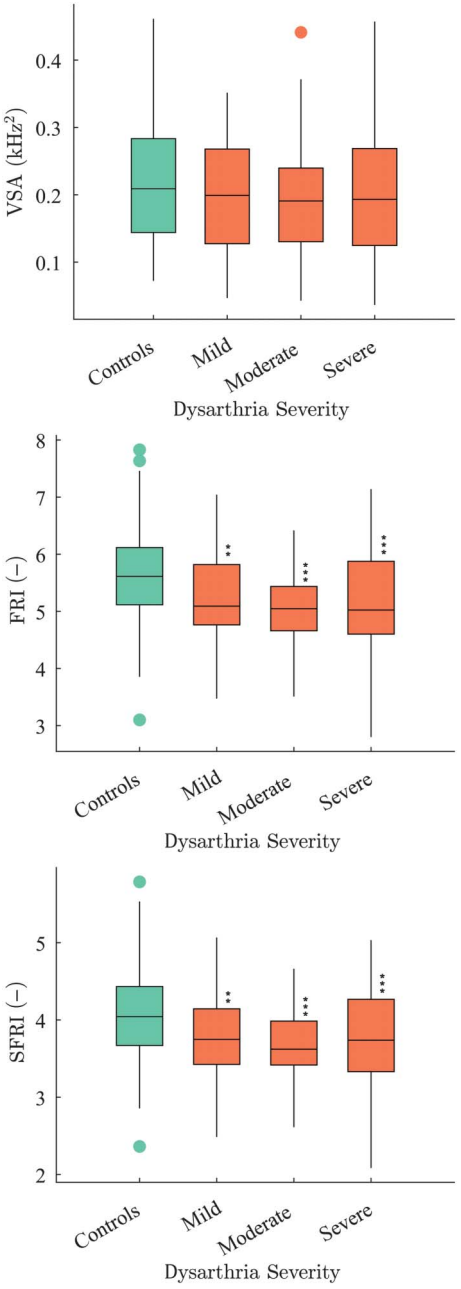
articulation metrics, our algorithm involved multiple levels of error correction such as outlier exclusion and correction of vowel identification by clustering. As a result, we reached the resulting accuracy of 77% (i.e., 23% error based on 1–F-score), which we believe is a very promising accuracy given a large number of etiologies and dysarthria severities involved. In addition, there are limitations in the accuracy of available technologies for phoneme recognition and formant tracking, even for healthy speech. For instance, it might be assumed that the solution toward better accuracy would be to change the formant tracker, yet all available open-source formant trackers were found to have similar detection performance (Schiel & Zitzelsberger, 2018). Considering the shape of the resulting vowel areas, the most considerable discrepancy between automated and manual labels was for F_1 estimation across vowels /i/ and /u/. Whereas the automated method tended to capture lower F_1 of vowels /i/ and /u/ with increasing dysarthria severity, the hand-labeled method did not find any change in F_1 or even increased F_1 due to dysarthria (Roy et al., 2009; Rusz et al., 2013; Skodda et al., 2011). Therefore, in comparison to the vowel articulation index that is most widely used in the literature (Roy et al., 2009; Rusz et al., 2013; Skodda et al., 2011), we proposed an alternative FRI that reflects the dysarthria-related lowering of F_1 in vowels /i/ and /u/ as captured by automated method. The inconsistency between manual and automated labels might be caused by an incidental formant tracker confusion of F_1 as the fundamental frequency and its harmonics close to F_1 of vowel /i/ and /u/. The SFRI, based only on F_2 values showed similar classification accuracy to detect neurological disease type or dysarthria type and even slightly better

Table 4. Formant frequencies of corner vowels estimated from monologues for different dysarthria severities compared to healthy controls.

Dysarthria severity	/a/ *M* (*SD*) Semitones		/i/ *M* (*SD*) Semitones		/u/ *M* (*SD*) Semitones	
	F_1	F_2	F_1	F_2	F_1	F_2
Controls	41.91 (2.6)	53.92 (2.1)	29.93 (2.3)	62.37 (1.6)	30.96 (1.6)	46.60 (1.8)
Mild	41.71 (2.9)	53.90 (1.6)	29.26 (2.0)	61.84 (1.6)	30.36 (1.6)	47.47 (2.2)
Moderate	41.41 (3.2)	54.03 (2.3)	29.05 (2.6)	61.89 (1.7)	29.86 (2.5)	48.09 (1.9)
Severe	41.78 (3.9)	54.15 (1.8)	28.84 (2.8)	61.73 (1.5)	29.72 (2.4)	47.93 (2.6)

Note. To minimize the effects of sex between individual speakers, the estimated formant frequencies were converted into a logarithmic tonal scale (semitones). Hertz to semitone formula: f(semitone) = 12*((log*f(Hz)/60)/log(2)). F_1 = first formant frequency; F_2 = second formant frequency.

Figure 10. Statistically significant group differences for estimated articulation features in monologues among the different dysarthria severities compared to healthy controls adjusted by age and sex with ***, **, * referring to $p < .001$, $p < .01$, and $p < .05$, respectively. Middle bars represent median, and rectangles represent the interquartile range. Maximum and minimum values are by error bars. Outliers are marked as dots. VSA = vowel space area; FRI = formant ratio index; SFRI = second formant ratio index.



accuracy to detect dysarthria severity compared to FRI based on both F_1 and F_2, suggesting SFRI as a suitable alternative to measure vowel articulation deficits. Thus, impairment of vowel articulation in neurodegenerative diseases can be tracked solely by changes in F_2 frequencies that are related to particular deficits in frontward/backward tongue movements. However, automated method achieved an increased inaccuracy in F_2 estimation of vowel /u/, with tendency to capture lower values compared to hand-labeling. Since the observed effects of etiology, dysarthria subtype, and severity were largely reflected by shifts in both formants of vowel /u/, we cannot exclude that these changes could be partially attributed to artifacts related to inaccurate formant tracking rather than actual disease effects. On the other hand, we believe that the automated method's error bias is not specific for etiology or dysarthria subtype and is generally the same for dysarthric and healthy speech, therefore not significantly accounting for the group differences.

Limitations of This Study

This study has certain limitations. Eight groups of patients were selected to cover a wide range of the common movement disorders associated with different pathophysiology responsible for the occurrence of vowel articulation impairment. Several of these etiologies showed different types of mixed dysarthria, leading to a smaller sample size for specific dysarthria subtypes. Additional better-sampled investigations with participants having different disease types, and possibly different dysarthria types, are required to confirm and further extend our findings. The study is based solely on the Czech language; thus, the language independence of the applied methods should be verified in future studies. Nonetheless, a recent multilanguage trial in PD revealed broadly similar profiles of dysarthria across multiple languages (Rusz et al., 2021). In addition, the formant tracker utilizing Burg's algorithm is considered language independent. The phoneme recognizer used in this study can be easily substituted for a universal recognizer that supports most of the world's languages (Li et al., 2020; Y. Liu et al., 2021). Subsequently, it is noteworthy to point out that shifts in formant frequencies and reductions in vowel space might occur due to other conditions than dysarthria such as differing dialect (Williams & Escudero, 2014), behavioral accent (Kamiloğlu et al., 2020), or stuttering-like behavior (Blomgren et al., 1998). We strived to minimize these effects by investigating subjects of the same dialect via an emotionally neutral context of monologue. From the etiologies investigated, the stuttering-like behavior is common only in PSP and very rare in de-novo PD (Rusz et al., 2015; Tykalová et al., 2015). However, the severity of vowel articulation impairment in PSP was not principally different from MSA, which is also atypical parkinsonism without the occurrence of dysfluency but with a similar dysarthria type and severity (Rusz et al., 2015), suggesting that affected vowels are mainly a consequence of dysarthria itself. Finally, our

Table 5. Classification analysis for the formant features for monologues.

%	VSA	FCI	SFCI	Random vector
Neurological disease type	5.0	**39.7**	38.8	5.3
Dysarthria type	5.0	37.0	**37.3**	4.2
Dysarthria severity	39.9	46.8	**49.2**	19.8

Note. The numbers indicate the percentage of subjects correctly identified by the discriminant analysis as original groups. Bold numbers indicate the best accuracy across neurological disease type, dysarthria type, and dysarthria severity. Random vector refers to the experimental results regarding probability of correct factor identification by chance. VSA = vowel space area; FRI = formant ratio index; SFRI = second formant ratio index.

algorithm was tested only with data acquired via a professional microphone without any disruptive noise. Therefore, future studies should evaluate the vowel articulation algorithm performance via a low-quality microphone, such as within smartphones in natural environments (Rusz et al., 2018).

Conclusions

This study represents an insight into the imprecise vowel articulation as a consequence of impairment of fine voluntary movements in a wide range of progressive neurological diseases with various etiologies and stages. We found that an automatized approach could reliably estimate vowel articulation features from natural connected speech regardless of the disease localization in the nervous system (pyramidal tract, basal ganglia, cerebellum, and cranial nerves), etiology (neurodegeneration and autoimmune disorder), and different degrees of disability. However, the specific tongue movement reflected by formant measures differed across some etiologies and dysarthria types independently on dysarthria severity. Therefore, acoustic analysis of vowel articulation may provide a practical tool not only for monitoring the efficacy of future experimental disease-modifying treatments and speech therapy but also for delivering clues for differential diagnosis. Future longitudinal studies should corroborate the sensitivity of vowel articulation deficits to disease progression among progressive disorders.

Authors Contributions

Vojtech Illner: Conceptualization (Equal), Data curation (Lead), Formal analysis (Lead), Methodology (Equal), Software (Lead), Validation (Lead), Visualization (Lead), Writing – original draft (Equal). **Tereza Tykalova:** Investigation (Supporting), Formal analysis (Supporting), Project administration (Equal), Validation (Supporting), Visualization (Supporting), Writing – review & editing (Lead). **Dominik Skrabal:** Formal analysis (Supporting), Investigation (Supporting), Writing – review & editing (Supporting). **Jiri Klempir:** Investigation (Lead), Writing – review & editing (Supporting). **Jan Rusz:** Conceptualization (Equal), Investigation (Supporting), Methodology (Equal), Project administration (Equal), Validation (Supporting), Visualization (Supporting), Funding acquisition (Lead), Writing – original draft (Equal).

Data Availability Statement

Individual participant data that underlie the findings of this study are available upon reasonable request from the corresponding author. The speech data are not publicly available because they contain information that could compromise the privacy of study participants.

Acknowledgments

This study was supported by the Czech Ministry of Health (Grants MH CZ-DRO-VFN64165 to Jiri Klempir and Jan Rusz and NU-20-08-00445 to Vojtech Illner, Tereza Tykalova, Dominik Skrabal, and Jan Rusz), National Institute for Neurological Research (Programme EXCELES; ID Project No. LX22NPO5107), funded by the European Union–Next Generation EU to Tereza Tykalova and Jan Rusz, Czech Technical University in Prague (Grant SGS23/170/OHK3/3 T/13 to Vojtech Illner), and by the Cooperation Program, research area Neuroscience to Jiri Klempir and Jan Rusz. We are obliged to speech-language pathologists Hana Ruzickova and Tereza Listvanova for auditory–perceptual evaluation of dysarthrias.

References

Blomgren, M., Robb, M., & Chen, Y. (1998). A note on vowel centralization in stuttering and nonstuttering individuals. *Journal of Speech, Language, and Hearing Research, 41*(5), 1042–1051. https://doi.org/10.1044/jslhr.4105.1042

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Brooks, B., Miller, R. G., Swash, M., & Munsat, T. (2000). El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and*

Other Motor Neuron Disorders, 1(5), 293–299. https://doi.org/10.1080/146608200300079536

Caverlé, M. W. J., & Vogel, A. P. (2020). Stability, reliability, and sensitivity of acoustic measures of vowel space: A comparison of vowel space area, formant centralization ratio, and vowel articulation index. *The Journal of the Acoustical Society of America, 148*(3), 1436–1444. https://doi.org/10.1121/10.0001931

Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., & Nakanishi, A. (1999). The ALSFRS-R: A revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences, 169*(1–2), 13–21. https://doi.org/10.1016/S0022-510X(99)00210-5

Childers, D. G. (1978). *Modern spectrum analysis*. IEEE Computer Society Press.

Ciucci, M. R., Russell, J. A., Schaser, A. J., Doll, E., Vinney, L. M., & Connor, N. (2011). Tongue force and timing deficits in a rat model of Parkinson disease. *Behavioural Brain Research, 222*(2), 315–320. https://doi.org/10.1016/j.bbr.2011.03.057

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*(3), 462–496. https://doi.org/10.1044/jshr.1203.462

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246–269. https://doi.org/10.1044/jshr.1202.246

Duffy, J. R. (2019). *Motor speech disorders: Substrates, differential diagnosis, and management* (4th ed.). Elsevier.

Elble, R., Comella, C., Fahn, S., Hallett, M., Jankovic, J., Juncos, J. L., LeWitt, P., Lyons, K., Ondo, W., Pahwa, R., Sethi, K., Stover, N., Tarsy, D., Testa, C., Tintner, R., Watts, R., & Zesiewicz, T. (2012). Reliability of a new scale for essential tremor. *Movement Disorders, 27*(12), 1567–1569. https://doi.org/10.1002/mds.25162

Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2017). Assessing vowel centralization in dysarthria: A comparison of methods. *Journal of Speech, Language, and Hearing Research, 60*(2), 341–354. https://doi.org/10.1044/2016_JSLHR-S-15-0355

Ge, S., Wan, Q., Yin, M., Wang, Y., & Huang, Z. (2021). Quantitative acoustic metrics of vowel production in Mandarin-speakers with post-stroke spastic dysarthria. *Clinical Linguistics & Phonetics, 35*(8), 779–792. https://doi.org/10.1080/02699206.2020.1827295

Gilman, S., Wenning, G. K., Low, P. A., Brooks, D. J., Mathias, C. J., Trojanowski, J. Q., Wood, N. W., Colosimo, C., Durr, A., Fowler, C. J., Kaufmann, H., Klockgether, T., Lees, A., Poewe, W., Quinn, N., Revesz, T., Robertson, D., Sandroni, P., Seppi, K., & Vidailhet, M. (2008). Second consensus statement on the diagnosis of multiple system atrophy. *Neurology, 71*(9), 670–676. https://doi.org/10.1212/01.wnl.0000324625.00404.15

Goetz, C., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A., Lees, A., Leurgans, S., LeWitt, P., Nyenhuis, D., . . . LaPelle, N. (2008). Movement Disorder Society-Sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders, 23*(15), 2129–2170. https://doi.org/10.1002/mds.22340

Hasegawa-Johnson, M., Pizza, S., Alwan, A., Cha, J. S., & Haker, K. (2003). Vowel category dependence of the relationship between palate height, tongue height, and oral area. *Journal of Speech, Language, and Hearing Research, 46*(3), 738–753. https://doi.org/10.1044/1092-4388(2003/059)

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology, 11,* Article 327643. https://doi.org/10.1155/1999/327643

Höglinger, G., Respondek, G., Stamelou, M., Kurz, C., Josephs, K., Lang, A., Mollenhauer, B., Müller, U., Nilsson, C., Whitwell, J., Arzberger, T., Englund, E., Gelpi, E., Giese, A., Irwin, D., Meissner, W., Pantelyat, A., Rajput, A., van Swieten, J., . . . Movement Disorder Society-endorsed PSP Study Group. (2017). Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. *Movement Disorders, 32*(6), 853–864. https://doi.org/10.1002/mds.26987

Huntington Study Group. (1996). Unified Huntington's Disease Rating Scale: Reliability and consistency. *Movement Disorders, 11*(2), 136–142. https://doi.org/10.1002/mds.870110204

Kamiloğlu, R. G., Fischer, A. H., & Sauter, D. (2020). Good vibrations: A review of vocal expressions of positive emotions. *Psychonomic Bulletin & Review, 27*(2), 237–265. https://doi.org/10.3758/s13423-019-01701-x

Kempler, D., & Van Lancker, D. (2002). Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease. *Brain and Language, 80*(3), 449–464. https://doi.org/10.1006/brln.2001.2602

Kent, R. D., & Kim, Y. J. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics, 17*(6), 427–445. https://doi.org/10.1080/0269920031000086248

Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders, 32*(3), 141–186. https://doi.org/10.1016/S0021-9924(99)00004-0

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica, 63*(4), 187–194. https://doi.org/10.1159/000318881

Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54*(2), 417–429. https://doi.org/10.1044/1092-4388(2010/10-0020)

Kouba, T., Illner, V., & Rusz, J. (2022). Study protocol for using a smartphone application to investigate speech biomarkers of Parkinson's disease and other synucleinopathies: SMART-SPEECH. *BMJ Open, 12*(6), Article e059871. https://doi.org/10.1136/bmjopen-2021-059871

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology, 33*(11), 1444–1452. https://doi.org/10.1212/WNL.33.11.1444

Lam, J., & Tjaden, K. (2016). Clear speech variants: An acoustic study in Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 59*(4), 631–646. https://doi.org/10.1044/2015_JSLHR-S-15-0216

Lansford, K. L., & Liss, J. M. (2014). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research, 57*(1), 57–67. https://doi.org/10.1044/1092-4388(2013/12-0262)

Li, X., Dalmia, S., Li, J., Lee, M., Littell, P., Yao, J., Anastasopoulos, A., Mortensen, D. R., Neubig, G., Black, A. W., & Metze, F. (2020). Universal phone recognition with a multilingual allophone system. In *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8249–8253). IEEE. https://doi.org/10.1109/ICASSP40776.2020.9054362

Lin, D. J., Hermann, K. L., & Schmahmann, J. (2016). The diagnosis and natural history of multiple system atrophy, cerebellar type. *The Cerebellum, 15*(6), 663–679. https://doi.org/10.1007/s12311-015-0728-y

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America, 117*(6), 3879–3889. https://doi.org/10.1121/1.1898623

Liu, Y., Penttila, N., Ihalainen, T., Lintula, J., Convey, R., & Räsänen, O. (2021). Language-independent approach for automatic computation of vowel articulation features in dysarthric speech assessment. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29,* 2228–2243. https://doi.org/10.1109/TASLP.2021.3090973

Louis, E. D., Faust, P. L., Vonsattel, J.-P. G., Honig, L. S., Rajput, A., Robinson, C. A., Rajput, A., Pahwa, R., Lyons, K. E., Ross, G. W., Borden, S., Moskowitz, C. B., Lawton, A., & Hernandez, N. (2007). Neuropathological changes in essential tremor: 33 cases compared with 21 controls. *Brain, 130*(12), 3297–3307. https://doi.org/10.1093/brain/awm266

Mahalanobis, P. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta), 2,* 49–55.

Mou, Z., Chen, Z., Yang, J., & Xu, L. (2018). Acoustic properties of vowel production in Mandarin-speaking patients with post-stroke dysarthria. *Scientific Reports, 8*(1), Article 14188. https://doi.org/10.1038/s41598-018-32429-8

Payan, C. A. M., Viallet, F., Landwehrmeyer, B. G., Bonnet, A. M., Borg, M., Durif, F., Lacomblez, L., Bloch, F., Verny, M., Fermanian, J., Agid, Y., Ludolph, A. C., Leigh, P. N., Bensimon, G., & NNIPPS Study Group. (2011). Disease severity and progression in progressive supranuclear palsy and multiple system atrophy: Validation of the NNIPPS–PARKINSON PLUS SCALE. *PLOS ONE, 6*(8). https://doi.org/10.1371/journal.pone.0022293

Pollak, P., Cernocky, J., Boudy, J., Choukri, K., Kochanina, J., Majewski, W., Ostroukhov, V., Rusko, M., Sadowski, J., Siemund, R., Staroniewitz, P., Trnka, M., Tropf, H., Vicsi, K., Heuvel, H., & Virag, A. (2000). SpeechDat(E) - Eastern European Telephone Speech Databases. *Proceedings LREC'2000 Satellite Workshop XLDB - Very Large Telephone Speech Databases*, 20–25.

Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., Halliday, G., Goetz, C., Gasser, T., Dubois, B., Chan, P., Bloem, B. R., Adler, C. H., & Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders, 30*(12), 1591–1601. https://doi.org/10.1002/mds.26424

Reilmann, R. (2019). Parkinsonism in Huntington's disease. *International Review of Neurobiology, 149,* 299–306. https://doi.org/10.1016/bs.irn.2019.10.006

Robertson, L. T., & Hammerstad, J. P. (1996). Jaw movement dysfunction related to Parkinson's disease and partially modified by levodopa. *Journal of Neurology, Neurosurgery, & Psychiatry, 60*(1), 41–50. https://doi.org/10.1136/jnnp.60.1.41

Roy, N., Nissen, S. L., Dromey, C., & Sapir, S. (2009). Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of Communication Disorders, 42*(2), 124–135. https://doi.org/10.1016/j.jcomdis.2008.10.001

Rusz, J., Bonnet, C., Klempíř, J., Tykalová, T., Baborová, E., Novotný, M., Rulseh, A., & Růžička, E. (2015). Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Journal of Neurology, 262*(4), 992–1001. https://doi.org/10.1007/s00415-015-7671-1

Rusz, J., Cmejla, R., Tykalova, T., Ruzickova, H., Klempir, J., Majerova, V., Picmausova, J., Roth, J., & Ruzicka, E. (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task. *The Journal of the Acoustical Society of America, 134*(3), 2171–2181. https://doi.org/10.1121/1.4816541

Rusz, J., Hlavnička, J., Novotný, M., Tykalová, T., Pelletier, A., Montplaisir, J., Gagnon, J., Dušek, P., Galbiati, A., Marelli, S., Timm, P., Teigen, L., Janzen, A., Habibi, M., Stefani, A., Holzknecht, E., Seppi, K., Evangelista, E., Rassu, A., . . . Sonka, K. (2021). Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. *Annals of Neurology, 90*(1), 62–75. https://doi.org/10.1002/ana.26085

Rusz, J., Hlavnicka, J., Tykalova, T., Novotny, M., Dusek, P., Sonka, K., & Ruzicka, E. (2018). Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26*(8), 1495–1507. https://doi.org/10.1109/TNSRE.2018.2851787

Rusz, J., Klempíř, J., Tykalová, T., Baborová, E., Čmejla, R., Růžička, E., & Roth, J. (2014). Characteristics and occurrence of speech impairment in Huntington's disease: Possible influence of antipsychotic medication. *Journal of Neural Transmission, 121*(12), 1529–1539. https://doi.org/10.1007/s00702-014-1229-8

Rusz, J., Tykalová, T., Salerno, G., Bancone, S., Scarpelli, J., & Pellecchia, M. (2019). Distinctive speech signature in cerebellar and parkinsonian subtypes of multiple system atrophy. *Journal of Neurology, 266*(6), 1394–1404. https://doi.org/10.1007/s00415-019-09271-7

Saigusa, H., Saigusa, M., Aino, I., Iwasaki, C., Li, L., & Niimi, S. (2006). M-Mode color Doppler ultrasonic imaging of vertical tongue movement during articulatory movement. *Journal of Voice, 20*(1), 38–45. https://doi.org/10.1016/j.jvoice.2005.01.003

Sandoval, S., Berisha, V., Utianski, R. L., Liss, J., & Spanias, A. (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America, 134*(5), EL477–EL483. https://doi.org/10.1121/1.4826150

Sapir, S., Ramig, L., & Fox, C. (2008). Speech and swallowing disorders in Parkinson disease. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*(3), 205–210. https://doi.org/10.1097/MOO.0b013e3282febd3a

Schiel, F., & Zitzelsberger, T. (2018, May). Evaluation of automatic formant trackers. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga, (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (pp. 2843–2848). European Language Resources Association.

Schmitz-Hübsch, T., du Montcel, S., Baliko, L., Berciano, J., Boesch, S., Depondt, C., Giunti, P., Globas, C., Infante, J., Kang, J., Kremer, B., Mariotti, C., Melegh, B., Pandolfo, M., Rakowicz, M., Ribai, P., Rola, R., Schols, L., Szymanski, S., . . . Fancellu, R. (2006). Scale for the assessment and rating of ataxia: Development of a new clinical scale. *Neurology, 66*(11), 1717–1720. https://doi.org/10.1212/01.wnl.0000219042.60538.92

Schwarz, P., & Černocký, J. (2008). *Phoneme recognition based on long temporal context* [Doctoral dissertation, Brno University of Technology].

Schwarz, P., Matejka, P., Burget, L., & Glembek, O. (2022). *Phoneme recognizer based on long temporal context*. Brno

University of Technology. https://speech.fit.vutbr.cz/software/phoneme-recognizer-based-long-temporal-context

Shimon, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research, 53*(1), 114–125. https://doi.org/10.1044/1092-4388(2009/08-0184)

Skodda, S., Schlegel, U., Klockgether, T., & Schmitz-Hübsch, T. (2014). Vowel articulation in patients with spinocerebellar ataxia. *International Journal of Speech and Language Pathology and Audiology, 1,* 63–71. https://doi.org/10.12970/2311-1917.2013.01.02.3

Skodda, S., Visser, W., & Schlegel, U. (2011). Vowel articulation in Parkinson's disease. *Journal of Voice, 25*(4), 467–472. https://doi.org/10.1016/j.jvoice.2010.01.009

Skrabal, D., Rusz, J., Novotny, M., Sonka, K., Ruzicka, E., Dusek, P., & Tykalova, T. (2022). Articulatory undershoot of vowels in isolated REM sleep behavior disorder and early Parkinson's disease. *NPJ Parkinson's Disease, 8*(1), Article 137. https://doi.org/10.1038/s41531-022-00407-7

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A., Miller, D. H., Montalban, X., . . . Cohen, J.A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology, 17*(2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2

Tjaden, K. (2008). Speech and swallowing in Parkinson's disease. *Topics in Geriatric Rehabilitation, 24*(2), 115–126. https://doi.org/10.1097/01.TGR.0000318899.87690.44

Tjaden, K., Lam, J., & Wilding, G. (2013). Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions. *Journal of Speech, Language, and Hearing Research, 56*(5), 1485–502. https://doi.org/10.1044/1092-4388(2013/12-0259

Tjaden, K., Rivera, D., Wilding, G., & Turner, G. (2005). Characteristics of the lax vowel space in dysarthria. *Journal of Speech, Language, and Hearing Research, 48*(3), 554–566. https://doi.org/10.1044/1092-4388(2005/038)

Tykalova, T., Pospisilova, M., Cmejla, R., Jerabek, J., Mares, P., & Rusz, J. (2016). Speech changes after coordinative training in patients with cerebellar ataxia: A pilot study. *Neurological Sciences, 37*(2), 293–296. https://doi.org/10.1007/s10072-015-2379-7

Tykalová, T., Rusz, J., Čmejla, R., Klempíř, J., Růžičková, H., Roth, J., & Růžička, E. (2015). Effect of dopaminergic medication on speech dysfluency in Parkinson's disease: A longitudinal study. *Journal of Neural Transmission, 122*(8), 1135–1142. https://doi.org/10.1007/s00702-015-1363-y

Weismer, G. (1984). Articulatory characteristics of parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal–supraglottal coordination. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 101–130). College-Hill.

Weismer, G. (2006). Philosophy of research in motor speech disorders. *Clinical Linguistics & Phonetics, 20*(5), 315–349. https://doi.org/10.1080/02699200400024806

Weismer, G., Jeng, J. Y., Laures, J. S., Kent, R., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica, 53*(1), 1–18. https://doi.org/10.1159/000052649

Whitfield, J. A. (2019). Exploration of metrics for quantifying formant space: Implications for clinical assessment of Parkinson disease. *Perspectives of the ASHA Special Interest Groups, 4*(2), 402–410. https://doi.org/10.1044/2019_PERS-SIG19-2018-0004

Whitfield, J. A., & Goberman, A. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of Communication Disorders, 51,* 19–28. https://doi.org/10.1016/j.jcomdis.2014.06.005

Whitfield, J. A., & Mehta, D. D. (2019). Examination of clear speech in Parkinson disease using measures of working vowel space. *Journal of Speech, Language, and Hearing Research, 62*(7), 2082–2098. https://doi.org/10.1044/2019_JSLHR-S-MSC18-18-0189

Williams, D., & Escudero, P. (2014). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America, 136*(5), 2751–2761. https://doi.org/10.1121/1.4896471

Yunusova, Y., Green, J., Lindstrom, M. J., Pattee, G. L., & Zinman, L. (2013). Speech in ALS: Longitudinal changes in lips and jaw movements and vowel acoustics. *Journal of Medical Speech-Language Pathology, 21*(1), 1–13.

Effect of Ageing on Acoustic Characteristics of Voice Pitch and Formants in Czech Vowels

*Tereza Tykalova, *,†Dominik Skrabal, ‡Tomas Boril, *Roman Cmejla, ‡Jan Volin, and *Jan Rusz, *†‡*Prague, Czech Republic*

Summary: Background. The relevance of formant-based measures has been noted across a spectrum of medical, technical, and linguistic applications. Therefore, the primary aim of the study was to evaluate the effect of ageing on vowel articulation, as the previous research revealed contradictory findings. The secondary aim was to provide normative acoustic data for all Czech monophthongs.

Methods. The database consisted of 100 healthy speakers (50 men and 50 women) aged between 20 and 90. Acoustic characteristics, including vowel duration, vowel space area (VSA), fundamental frequency (f_o), and the first to fourth formant frequencies (F_1-F_4) of 10 Czech vowels were extracted from a reading passage. In addition, the articulation rate was calculated from the entire duration of the reading passage.

Results. Age-related changes in pitch were sex-dependent, while age-related alterations in F_2/a/, F_2/u/, VSA, and vowel duration seemed to be sex-independent. In particular, we observed a clear lowering of f_o with age for women, but no change for men. With regard to formants, we found lowering of F_2/a/ and F_2/u/ with increased age, but no statistically significant changes in F_1, F_3, or F_4 frequencies with advanced age. Although the alterations in F_1 and F_2 frequencies were rather small, they appeared to be in a direction against vowel centralization, resulting in a significantly greater VSA in the older population. The greater VSA was found to be related partly to longer vowel duration.

Conclusions. Alterations in vowel formant frequencies across several decades of adult life appear to be small or in a direction against vowel centralization, thus indicating the good preservation of articulatory precision in older speakers.

Key Words: Aging−Czech−Formant−Vowel−Acoustic analysis−Fundamental frequency.

INTRODUCTION

The quality and intelligibility of each vowel can be described mainly by its formant structure but also by the fundamental frequency (f_o) and vowel duration.[1] While the first (F_1) and second (F_2) formant frequencies are essential for phonemic recognition of various vowels,[2] the higher third (F_3) and fourth (F_4) formants contribute mainly to the expression of emotions.[1,3] From a physiological point of view, F_1 and F_2 frequencies reflect primarily tongue position and lips rounding[1] while F_3 and F_4 are thought to be related mainly to lip spreading or protrusion.[3,4] Since F_1 and F_2 frequencies have a well-defined acoustic-articulatory relationship, they were used for the definition of several derived metrics (see[2] for an overview). Among these, the most frequently reported acoustic measure was probably the vowel space area (VSA), typically calculated as the area of a triangular or quadrilateral polygon formed by the corner vowels in the two-dimensional formant plane ($F_1 \times F_2$), three-dimensional formant space ($F_1 \times F_2 \times F_3$), or potentially the four-dimensional formant hypercube ($F_1 \times F_2 \times F_3 \times F_4$), but the first of these is the most commonly used.[2] VSA is supposed to reflect the articulatory extrema of vowel production.[2] Therefore, vowel centralization, which is caused due to the limited articulatory range of motion (ie, formants with naturally higher frequencies tend toward lower frequencies, and formants with naturally lower frequencies tend toward higher frequencies), can be captured well by a reduced size of the VSA. From a clinical perspective, the evaluation of formant frequencies, VSA, or similar formant-based measures has proved its feasibility in many fields; specifically, they might be used to evaluate the effect of voice and speech therapy,[5−7] to serve as an early marker of Parkinson's disease[8] or other neurological conditions,[9] to ease the diagnosis of obstructive sleep apnoea,[10] and to monitor disease progression or the effect of drug introduction in neurodegenerative conditions.[11,12] Furthermore, the relevance of vowel formant measurements has also been noted across a spectrum of technical and linguistic applications, including automatic speech recognition,[13] age and sex identification,[14] forensic science,[15] dialect assessment,[16,17] and second-language studies.[18]

AGE-DEPENDENT ACOUSTIC CHARACTERISTICS

The elderly population is increasing dramatically across the world; as a result, the number of elderly subjects with speech

Accepted for publication February 26, 2020.

This study was supported by the Czech Science Foundation (https://gacr.cz/en/) GACR 16-19975S (TT, DS, TB, RC, JV, JR) and GACR 19-20887S (TT, RC), by the Czech Ministry of Health (http://www.azvcr.cz/en) NV19-04-00120 (TT, DS, JR), and by the OP VVV MEYS (http://www.msmt.cz/) project "Research Center for Informatics" - Grant no. CZ.02.1.01/0.0/0.0/16_019/0000765 (RC, JR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

From the *Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic; †Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University, Prague, Czech Republic; and the ‡Institute of Phonetics, Faculty of Arts, Charles University, Prague, Czech Republic.

Address correspondence and reprint requests to Tereza Tykalova, Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27, Praha 6, Czech Republic. E-mail: tykalova.tereza@gmail.com

Journal of Voice, Vol. 35, No. 6, pp. 931.e21−931.e33
0892-1997

https://doi.org/10.1016/j.jvoice.2020.02.022

and/or language disorders has also increased rapidly. Nevertheless, natural alterations in voice and speech also occur in healthy populations, and are attributed to anatomical and physiological changes in the larynx and other speech-related structures. For example, the speech performance of elderly people might be affected by worsening physiological conditions represented by factors such as longer processing times, reduced sensory feedback, general neuromuscular slowing, or peripheral degeneration of the speech mechanism.[19] In addition, sex-related differences might also occur between the sexes as a result of different laryngeal lowering and vocal tract lengthening, different rates of ageing, or the hormonal effects of menopause in women.[20,21]

Knowledge about the magnitude of physiological voice and speech alterations is essential for differentiating normal from pathological utterances. In fact, numerous studies have investigated the effect of ageing on speech production in order to identify specific speech alterations that commonly occur in healthy older adults.[2,19–32] It has been noted that, regardless of sex, people tend to have slower speaking and reading rates as they age,[19,26,29] and produce longer vowel segments.[22,25,26] Based on the results of both cross-sectional and longitudinal studies, the f_o in older women has been shown to decrease significantly.[21,24,27,31] However, the findings for the male population are somewhat inconsistent, as f_o is reported to decrease significantly,[23,27] remain unchanged[24] or even increase markedly[21,26,31] with advanced age. Considering age-related changes to F_1 and F_2 frequencies, the majority of the previous studies have observed vowel-specific alterations that are unique to male and female populations.[20,21,24,30] These changes in spectral patterns have been further hypothesized to be related to vowel centralization[22,30] that is present due to neuromuscular changes affecting the rate and precision of articulatory movements, or a general decrease in F_1 and F_2 frequencies that is mainly attributed to vocal tract lengthening.[28] However, more recent studies[24,25,31] do not appear to confirm these assumptions, as they have revealed no changes in F_1 and F_2 frequency for men or women over the age of 60,[25,31] nor of the trend towards VSA reduction.[24,25] In addition, the study published by Fletcher et al[25] revealed a significant relationship between speakers' average vowel durations and their VSA, indicating that speakers who had longer vowel durations produced vowels that were more spectrally distinct. Thus, some authors have suggested that a habitually slower speaking rate may be an effective compensatory mechanism that some speakers use to maintain acoustically distinct speech segments.[25] Finally, to the best of our knowledge, the age-related changes in F_3 and F_4 have not yet been investigated thoroughly, with only a very few studies generally reporting no age-related changes in higher formants.[20,24] Therefore, further research should be conducted in order to clarify these ambiguous findings.

CHARACTERISTICS OF CZECH VOWELS

Czech is a western Slavic language of the Indo-European family, and is spoken by nearly 10 million people in the Czech Republic and about two million people living abroad.[16] Compared to most other European countries, the Czech Republic does not exhibit large variations in local or social accents. Absence of such variations can be explained by the geographical compactness and relatively small size of the entire territory, combined with the traditionally unrestrained mobility of the population. Only the borders of the Czech territory, which are relatively sparsely populated, can claim to host true accents. The political development in recent decades and the influence of the media have also prevented the development of salient sociolects. However, remnants of local pronunciation features can still be traced in some regions such as southern Moravia or Silesia. Standard Czech pronunciation is based on the original accents of Central Bohemia—the region most densely populated and with the greatest political power.

The Czech vowel inventory contains 10 monophthongs and three diphthongs. The monophthongs consist of five different vowel qualities /a/, /ɛ/, /ɪ/, /o/ and /u/, occurring in two quantities as short and long. The long vowels are about 1.7 times longer than their short counterparts.[33] With the exception of /ɪ/ and /i:/, in which the short vowel is noticeably less close and more central than is the long one,[34] the pairs of short and long vowels are assumed to have similar spectral patterns.[16] However, only a few previous studies[35–37] have focused on the acoustic investigation of formant frequencies in Czech. These studies are further limited by the inclusion of a small number of nonrepresentative groups of subjects of a similar age and work status, such as university students[35,37] or professional actors.[36] Moreover, these studies have only analyzed F_1 and F_2 frequencies while neglecting higher formants,[35,37] have only considered short vowels,[35,36] and have only provided reference values for men.[37] Since there is tremendous phonetic diversity in the languages of the world and Czech belongs to underdocumented languages, there is a need for the definition of representative, normative acoustic data.

AIMS OF THE STUDY

The aim of the current study was to examine the acoustic characteristics of vowels in healthy Czech native speakers aged from 20 to 90 in order to evaluate the effect of ageing on vowel articulation. An additional aim was to provide normative data for Czech vowels. We decided to use a reading passage to create a more natural condition of connected speech while simultaneously maintaining a standardized speaking task. The set of acoustic characteristics chosen included f_o, F_1, F_2, F_3, F_4, VSA, vowel duration, and articulation rate. Since most of the previous studies,[21,22,24–27,32] with the exception of the study published by Sebastian et al[31] who focused only on participants aged from 60 to 80, have reported that the f_o in women and the vowel duration in both sexes were age-dependent, we expected the same trend in our data. With regard to formant frequencies, some of the previous studies indicated vowel-specific alterations in F_1 and F_2 frequencies as a function of age in at least 50%

of the formants investigated,[20,21,30] while more recent studies[24,25] do not appear to confirm these findings. Since these studies[24,25] investigated a larger cohort of participants, thus allowing for a more appropriate statistical design, we hypothesized that F_1 and F_2 frequencies would be age-independent for most of the vowels.

METHODS

Participants

A total of 100 Czech native speakers (50 men and 50 women) were recruited for the study. All the participants provided written, informed consent for the recording procedure, and the study was approved by the Ethics Committee of the Faculty of Biomedical Engineering at the Czech Technical University in Prague, Czech Republic. The age distribution in both sexes was balanced, with the males' ages ranging from 20 to 87 (mean 52.7 ± SD 20.1) years, and the females' ages ranging from 20 to 89 (53.2 ± 19.8) years. In addition, the age of each speaker in the male and/or female groups was different in order to provide a greater diversity of ages. The separate age distributions for the sexes are presented in Figure 1. The frequency of man and woman was similar in each age group. In particular, the percentage of men was 53% in 20−29 age group, 50% in 30−39 age group, 46% in 40−49 age group, 53% in 50−59 age group, 44% in 60−69 age group, and 52% in 70−89 age group. All the participants had completed 8 years of elementary education as a minimum, but most of the participants had higher educational levels (there were sociodemographic development disadvantages for the older generations—higher education was generally unavailable to them due to the political situation). All the participants were from the middle or upper-middle socioeconomic class, and had been living permanently or studying in Prague or in the Central Bohemian region for a minimum of 4 years at the time of recording. None of the participants was employed in professions that required the professional use of the voice such as acting, singing, or speech-language pathology. None of the participants suffered from depression or cognitive deficits that could have interfered with the recording procedure. To ensure a relatively homogenous database, all the participants were subject to a short interview and a careful, auditory-based dialect assessment was performed by a Czech phonetic specialist (JV) based on the reading text and on a monologue. Speakers who displayed clear traces of regional pronunciation were excluded from the analyses. All the participants spoke the standard language. The exclusion criteria for the participants were:

- a strong regional dialect;
- a history of developmental stuttering or other speech and/or language disorders;
- the use of hearing aids or medically diagnosed hearing loss;
- a history of neurological disorders;
- the current use of antidepressants or antipsychotics; and
- a history of excessive smoking (defined as more than 20 cigarettes per day for at least 3 years).[38]

Recording procedure

The audio data were recorded in a quiet room with a low level of ambient noise (< 40 dBA) using a head-mounted condenser microphone (Beyer dynamic Opus 55, Heilbronn, Germany) that was placed approximately 5 cm from the corner of the subject's mouth with 70° angle. The speech signals were sampled at 48 kHz with16-bit resolution. The recordings were collected during one session with a speech specialist (TT or DS) who explained the instructions to the subjects. Each participant was required to complete a series of speaking tasks including the standardized reading of a text as part of a longer protocol lasting about 25 minutes. There were no time limits during the recordings. All participants were asked to repeat their performance at any time if they or the examiner were not fully satisfied with their initial attempt. To ensure good concentration of speakers and to minimize fatigue, the complete reading text was divided into two passages. The first passage, which consisted of 257 words (Appendix A), was presented at the beginning of the recording session, while the second, consisting of 313 words Appendix B), was presented approximately 10 minutes later. The chosen passages (see the Appendix B and the first paragraph in the Appendix A) were extracts taken from books written by the famous Czech writer Karel Čapek. To facilitate the reading, the final text that was used was changed slightly from the original in some places in order to provide familiar, up-to-date vocabulary and grammatical structures. Since the long vowel /o:/ occurs very rarely in

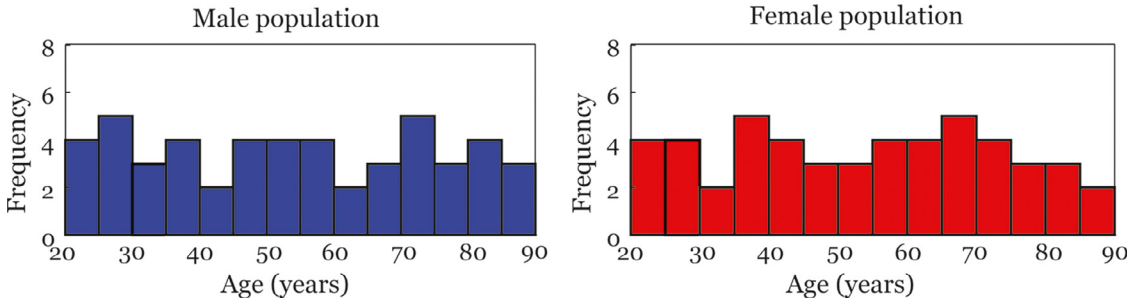


FIGURE 1. The age distribution of the participants.

Czech and almost exclusively in loanwords,[16] two additional, specially designed paragraphs were added in the middle of the reading text (see the second and third paragraphs in the Appendix A). During the recording session, each speaker was instructed to read the passages in a habitual manner with natural tempo and volume.

Selection of the target vowels

For the purposes of this study, 10 monophthongs, including five short vowels /a/, /ɛ/, /ɪ/, /o/, and /u/ and their long counterparts /aː/, /ɛː/, /iː/, /oː/ and /uː/, were of interest. Ten occurrences of each of these vowels were predefined within the reading passages (see the underlined vowels in Appendix A and Appendix B). To preserve the high diversity of the extracted vowels in order to represent the Czech language well while simultaneously maintaining good conditions for the evaluation of acoustic characteristics, the specific words and/or vowels were chosen according to the following criteria:

(1) The target words were selected from the entire duration of the passages at various positions within the sentences and intonation phrases to balance the influence of prosodic structure.
(2) Only one vowel in any given word was analyzed.
(3) The vowels were obtained equally from both stressed and unstressed syllables because Czech has a fixed stress on the first syllable of the word, but no direct reduction of vowel duration or vowel quality due to the occurrence in unstressed syllables.[39,40]
(4) To minimize the effect of coarticulation with surrounding phonemes, as well as the effect of the place of articulation, the vowels were chosen as much as possible
(5) to follow different voiceless plosives, fricatives, affricates, or no phoneme (only words at the beginning of intonation phrases).

In order to summarize the characteristics of the vowels analyzed, Table 1 presents the manner of articulation of the preceding consonant, the place of articulation of the preceding consonant, the syllable stress, the position of the target syllable within the word, and the position of the target word within the intonation phrase for the 10 occurrences of each of the vowels investigated.

Acoustic analysis

Six acoustic parameters, including vowel duration, f_o, F_1, F_2, F_3, and F_4, were evaluated for each vowel by means of specialized, widely used speech software PRAAT version

TABLE 1.
The Characteristics of the Preceding Consonant, As Well As the Syllable and Word Positions Related to the 10 Occurrences of Each Monophthong Investigated

	a	ɛ	ɪ	o	u	aː	ɛː	iː	oː	uː
Manner										
Stop	7	5	2	7	5	4	8	5	7	3
Fricative	2	2	3	1	3	3	1	3	3	3
Affricative	-	3	3	1	1	2	-	2	-	2
Liquid	-	-	-	-	-	1	1	-	-	-
No consonant	1	-	2	1	1	-	-	-	-	2
Place of articulation										
Bilabial	3	2	1	2	1	-	1	2	2	2
Alveolar	4	5	4	4	4	8	5	2	5	4
Postalveolar	-	2	2	-	-	1	-	3	-	-
Palatal	-	-	1	-	-	-	-	3	-	-
Velar	2	1	-	3	4	1	3	-	2	2
Glottal	-	-	-	-	-	-	1	-	1	-
No consonant	1	-	2	1	1	-	-	-	-	2
Syllable stress										
Stressed	5	5	5	5	4	3	4	5	5	4
Unstressed	5	5	5	5	6	7	6	5	5	6
Position of the Syllable										
Beginning	5	5	5	5	5	4	4	6	5	4
Middle	2	2	2	3	-	4	2	-	5	1
End	3	3	3	2	5	2	4	4	-	5
Position of the word										
Beginning	3	2	3	2	2	3	3	1	4	-
Middle	6	4	6	7	5	4	3	4	3	4
End	1	4	1	1	3	3	4	5	3	6

5.4.04.[41] The duration of a vowel was measured as the difference between the onset and offset of a vowel according to the criteria summarized in.[42] Specifically, the vowel onset was defined as the point of the abrupt onset of a periodic signal where the onset of f_o, F_1, and F_2 frequencies was evident, while the offset of a vowel was defined as the point of the abrupt offset of the periodic signal where the F_2 offset was mainly considered to be the indicator.[42] The f_o in Hz was calculated as a mean value from the entire vowel duration following the manual adjustment of the f_o range for each speaker. F_1, F_2, F_3, and F_4 frequencies in Hz were determined from 30-ms segment close to the middle section of a vowel where F_1 and F_2 formant patterns were visible and stable. In rare cases, when the steady-state segment of a vowel was not present but F_1 and F_2 formants were clearly visible and continuous, the 30-ms segment around the midpoint of the vowel duration was used. All the formant frequencies were extracted manually using a wide-band spectrogram with the formant contours depicted and the power spectral density displayed on the screen. The formant contours were analyzed using PRAAT default settings including the Burg method, 0.025 second duration of window length, and a maximum formant of 5000 Hz for men or 5500 Hz for women and five depicted formants. All the values obtained were checked with regard to phonetic knowledge in order to search for errors in the formant analysis, such as the merged or missing formants that commonly occur[2,43] due to reasons such as very strong first harmonic that hinders detection of a closely spaced F_1 or close proximity of formants (eg, F_2 and F_3 formants in vowel /i:/).[2] If the examiner concluded that there was a probable formant merge, no value was recorded for any of the higher formants frequencies (ie, the values for F_3 and F_4 were not considered in the event of an F_2−F_3 merger). A similar approach was applied for missing formants for example, the values of F_3 and F_4 were not considered in the event of a probable F_3 missing formant).

The VSA was calculated based on /a/, /ɪ/ and /u/ corner vowels using the following formula: VSA = ABS $((F_1/ɪ/ \times (F_2/a/ - F_2/u/) + F_1/a/ \times (F_2/u/ - F_2/ɪ/) + F_1/u/ \times (F_2/ɪ/ - F_2/a/)/2)$.[2]

To determine the potential effect of speech tempo on vowel duration, the articulation rate was calculated from the entire reading passage as the number of words per second after removing periods of silence that exceeded 60 milliseconds.[44]

Nonmeasurable data

Some acoustic variables could not be obtained from the complete database that included 100 vowels for each subject due to various methodological constraints. Specifically, vowel duration was not assessed for 1% of the target vowels due to misreadings. The f_o, F_1, and F_2 frequencies were not found in 3%−4% of the data, mainly due to the short duration of a vowel (< 30ms), sudden pitch drops, or the overall weak energy of the signal. Finally, F_3 and F_4 formants were judged to be nonmeasurable in 15% and 23% of the data, respectively, mainly due to the presence of formants that were assumed to be merged or missing. As the unmeasurable data were distributed evenly across the reading passages, at least five occurrences of each monophthong were always available for further analysis.

Measurement reliability

Intrajudge reliability was assessed following a reanalysis of 10% of the recordings by the same investigator (TT) who performed the original set of measurements. A Pearson correlation analysis calculated across individual vowel qualities showed significant, positive correlations for f_o ($r = 0.99$, $P < 0.001$), F_1 ($r = 0.97$−0.98, $P < 0.001$), F_2 ($r = 0.93$−0.99, $P < 0.001$), F_3 ($r = 0.95$−0.98, $P < 0.001$), and F_4 ($r = 0.93$−0.99, $P < 0.001$), as well as for vowel duration ($r = 0.97$−0.99, $P < 0.001$). The mean intrajudge standard error of measurement calculated across individual vowel qualities was 5 ± 0 Hz for f_o, 8 ± 1 Hz for F_1, 21 ± 5 Hz for F_2, 33 ± 14 Hz for F_3, 65 ± 23 Hz for F_4, and 2 ± 1 milliseconds for vowel duration.

Interjudge reliability was evaluated based on reanalysis of 10% of the recordings by the second investigator (DS), who was well trained in analyzing procedure. A Pearson correlation analysis calculated across individual vowel qualities indicated significant, positive correlations for f_o ($r = 0.99$, $P < 0.001$), F_1 ($r = 0.86$−0.97, $P < 0.001$), F_2 ($r = 0.77$−0.98, $P < 0.001$), F_3 ($r = 0.75$−0.99, $P < 0.001$), and F_4 ($r = 0.93$−0.98, $P < 0.001$), as well as for vowel duration ($r = 0.95$−0.98, $P < 0.001$). The mean interjudge standard error of measurement calculated across individual vowel qualities was 5 ± 0 Hz for f_o, 13 ± 2 Hz for F_1, 35 ± 15 Hz for F_2, 57 ± 38 Hz for F_3, 68 ± 19 Hz for F_4, and 3 ± 1 ms for vowel duration.

Statistical analysis

For the subsequent investigation, the final data from all the available occurrences were averaged separately for each speaker, acoustic variable, and monophthong. The averaging of vowel formants was applied because

(a) it reduces the variability that is typical of formant frequencies and controls for lexical factors of phonological neighbourhood density, thus ensuring that all the available occurrences within the utterance have similar importance;
(b) it is a standard procedure that is commonly used in many applications[5,8,9,12,45−48]; and
(c) it allows for a comparison with the majority of the previous research in the area of vowel articulation.[21,25,30]

All the relevant data used for the statistical analyses are available in supplementary material S1 Table. The Kolmogorov-Smirnov test for independent samples did not reject the null hypothesis of normal distribution. In order to

determine the age-dependent acoustic characteristics of vowels, we applied a 6 × 2 × 2 repeated measure analysis of variance (RM-ANOVA) with AGE (20−29, 30−39, 40−49, 50−59, 60−69, 70−89) and SEX (men, women) being treated as between-group factors and VOWEL (short, long) being treated as a within-group factor. *Post hoc* significance was assessed by the Fisher least-squares difference for the effect of AGE. The Bonferroni correction for multiple comparisons was applied for six tests that were conducted for each vowel quality individually, with a corrected P threshold equal to $P < 0.0083$ for $P < 0.05$. With regard to the articulation rate and VSA, the 6 × 2 ANOVA involving the factors of AGE (20−29, 30−39, 40−49, 50−59, 60−69, 70−89) and SEX (men, women) was applied. *Post hoc* significance was assessed by the Fisher least-squares difference for the effect of AGE. The nominal alpha level was set at 0.05. Statistical analyses were performed using Matlab (Mathworks, Massachusetts). The Pearson coefficient was calculated to determine correlations among the average vowel duration calculated across all monophthongs, the articulation rate, and VSA.

RESULTS

Age-dependent acoustic characteristics

The results of the acoustic analysis of the corner vowels /a/, /i/, and /u/ for the male and female populations are presented in Figures 2−4. The comparison of speech measurements for the vowels /ɛ/ and /o/ are included in supplementary S3 File.

For the vowel /a/, the RM-ANOVA showed a significant effect for AGE in F_2 [$F(5,88) = 5.1$, $P = 0.002$, $\eta 2 = 0.23$] and in vowel duration [$F(5,88) = 9.7$, $P < 0.001$, $\eta 2 = 0.36$]. *Post hoc* comparisons revealed significantly higher F_2 in 20−29 age group compared to 40−49 ($P = 0.03$), 50−59 ($P < 0.001$), 60−69 ($P = 0.004$) and 70−89 ($P < 0.001$) age groups as well as significantly increased vowel duration in 70−89 age group compared to 20−29 ($P < 0.001$), 30−39 ($P < 0.001$), 40−49 ($P < 0.001$), 50−59 ($P = 0.007$), and 60−69 ($P < 0.001$) age groups. The significant main effect for SEX was detected in f_o, F_1, F_2, F_3, and F_4 [$F(1,88) = 90−397$, $p < 0.001$, $\eta 2 = 0.51−0.82$], as well as for VOWEL in f_o, F_1, F_2, F_3, and for vowel duration [$F(1,88) = 38−2396$, $P < 0.001$, $\eta 2 = 0.31−0.97$]. Importantly, significant interaction was

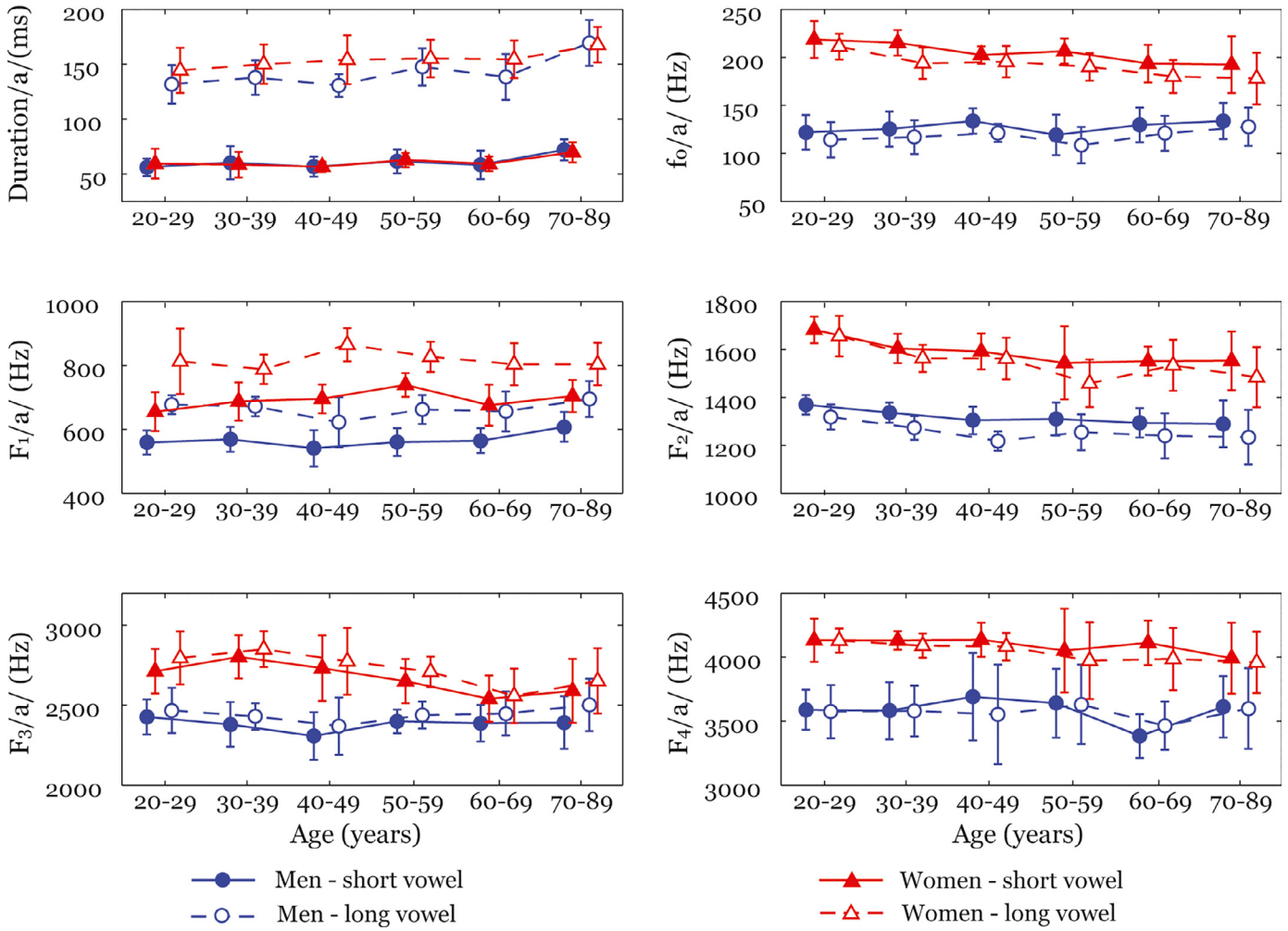


FIGURE 2. The comparison of speech measurements of the vowel /a/. Mean values and standard deviations (error bars) are depicted for both sexes (men, women), and vowel quantities (short, long), presented as a function of age.

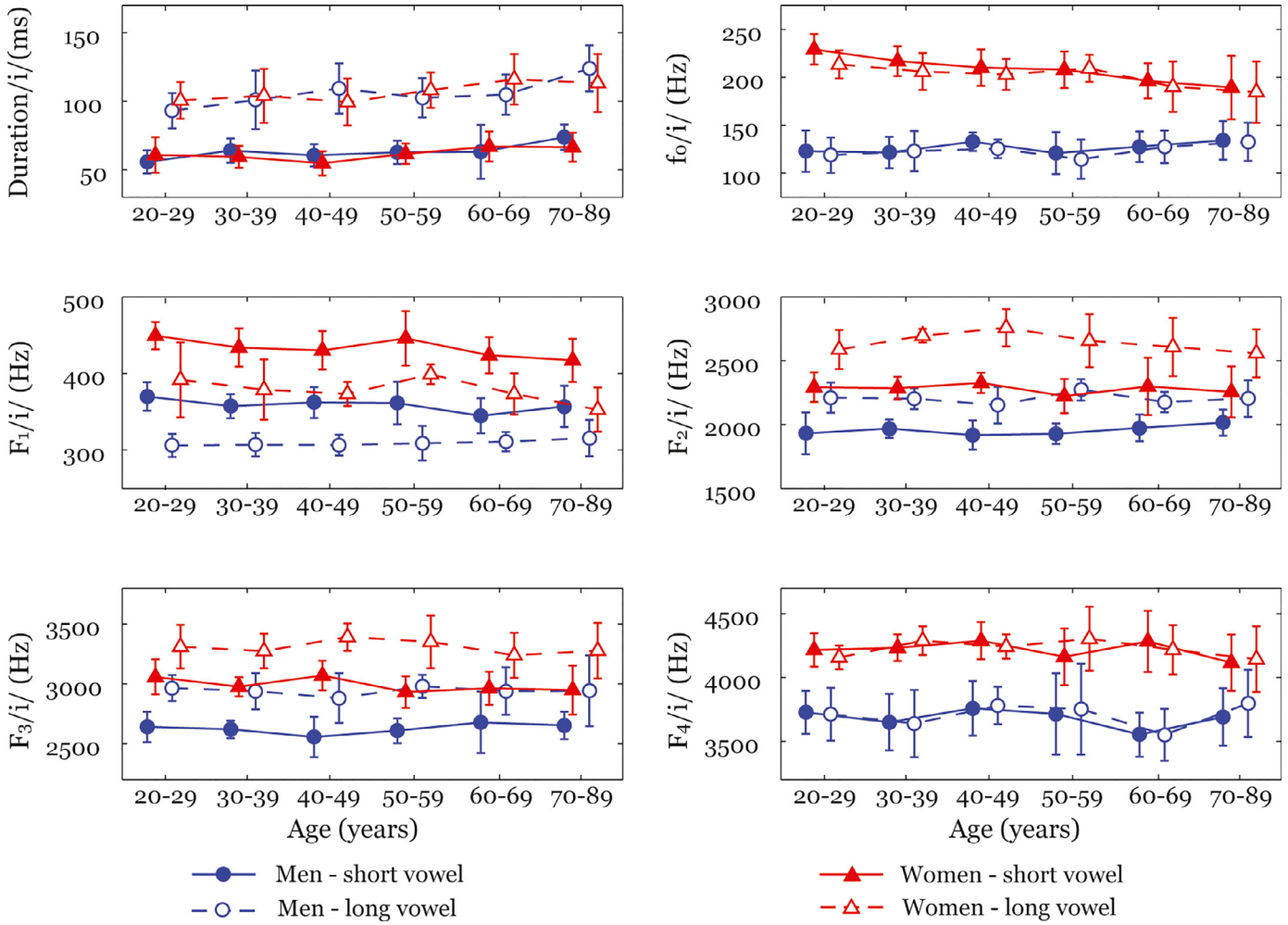


FIGURE 3. The comparison of speech measurements of the vowel /i/. Mean values and standard deviations (error bars) are depicted for both sexes (men, women), and vowel quantities (short, long), presented as a function of age.

revealed for AGE × SEX in f_o [$F(5,88) = 3.5$, $P = 0.04$, $\eta 2 = 0.17$]. In addition, we observed a significant interaction for VOWEL × SEX in F_1 [$F(1,88) = 8.3$, $P = 0.03$, $\eta 2 = 0.09$] and for vowel duration [$F(1,88) = 10.3$, $P = 0.01$, $\eta 2 = 0.11$].

For the vowel /i/, a significant effect for AGE was found in the vowel duration [$F(5,88) = 5.0$, $P = 0.003$, $\eta 2 = 0.22$]. *Post hoc* comparisons revealed significantly increased vowel duration in 70−89 age group compared to 20−29 ($P < 0.001$), 30−39 ($P = 0.02$), 40−49 ($P = 0.009$), and 50−59 ($P = 0.05$) age groups. In addition, a significant main effect was revealed for SEX in f_o, F_1, F_2, F_3 and F_4 [$F(1,88) = 141-365$, $P < 0.001$, $\eta 2 = 0.62-0.81$], as well as for VOWEL in f_o, F_1, F_2, F_3, and vowel duration [$F(1,88) = 27-829$, $P < 0.001$, $\eta 2 = 0.23-0.90$]. Importantly, we observed a significant interaction of AGE × SEX in f_o [$F(5,88) = 3.6$, $P = 0.03$, $\eta 2 = 0.17$]. We also found a significant interaction of AGE × VOWEL in F_2 [$F(5,88) = 4.6$, $P = 0.006$, $\eta 2 = 0.21$] associated with increase of F_2 in long vowels in 50−59 male age group and of VOWEL × SEX in F_2 [$F(1,88) = 30$, $P < 0.001$, $\eta 2 = 0.26$].

For the vowel /u/, the RM-ANOVA showed a significant effect for AGE in F_2 [$F(5,88) = 4.3$, $P = 0.009$, $\eta 2 = 0.20$] and for vowel duration [$F(5,88) = 5.1$, $P = 0.002$, $\eta 2 = 0.23$]. *Post hoc* comparisons revealed significantly higher F_2 in 20−29 age group compared to 50−59 ($P = 0.05$), 60−69 ($P = 0.006$), and 70−89 ($P < 0.001$) age groups as well as significantly increased vowel duration in 70−89 age group compared to 20−29 ($P < 0.001$), 30−39 ($P = 0.01$), 40−49 ($P < 0.001$), and 50−59 ($P = 0.02$) age groups. A significant main effect was revealed for SEX in f_o, F_1, F_3, F_4 [$F(1,88) = 106-348$, $P < 0.001$, $\eta 2 = 0.55-0.80$] and F_2 [$F(1,88) = 15.1$, $P = 0.001$, $\eta 2 = 0.15$], as well as for VOWEL in F_1, F_2, and for vowel duration [$F(1,88) = 34-489$, $P < 0.001$, $\eta 2 = 0.28-0.85$]. Interestingly, a significant interaction was observed for AGE × SEX in f_o [$F(5,88) = 4.2$, $P = 0.01$, $\eta 2 = 0.19$]. Finally, we also found a significant interaction of VOWEL × SEX in F_2 [$F(1,88) = 8.5$, $P = 0.03$, $\eta 2 = 0.09$] and in vowel duration [$F(1,88) = 7.8$, $P = 0.04$, $\eta 2 = 0.08$].

The results of the statistical analysis for the articulation rate and VSA are presented in Figure 5. For the articulation

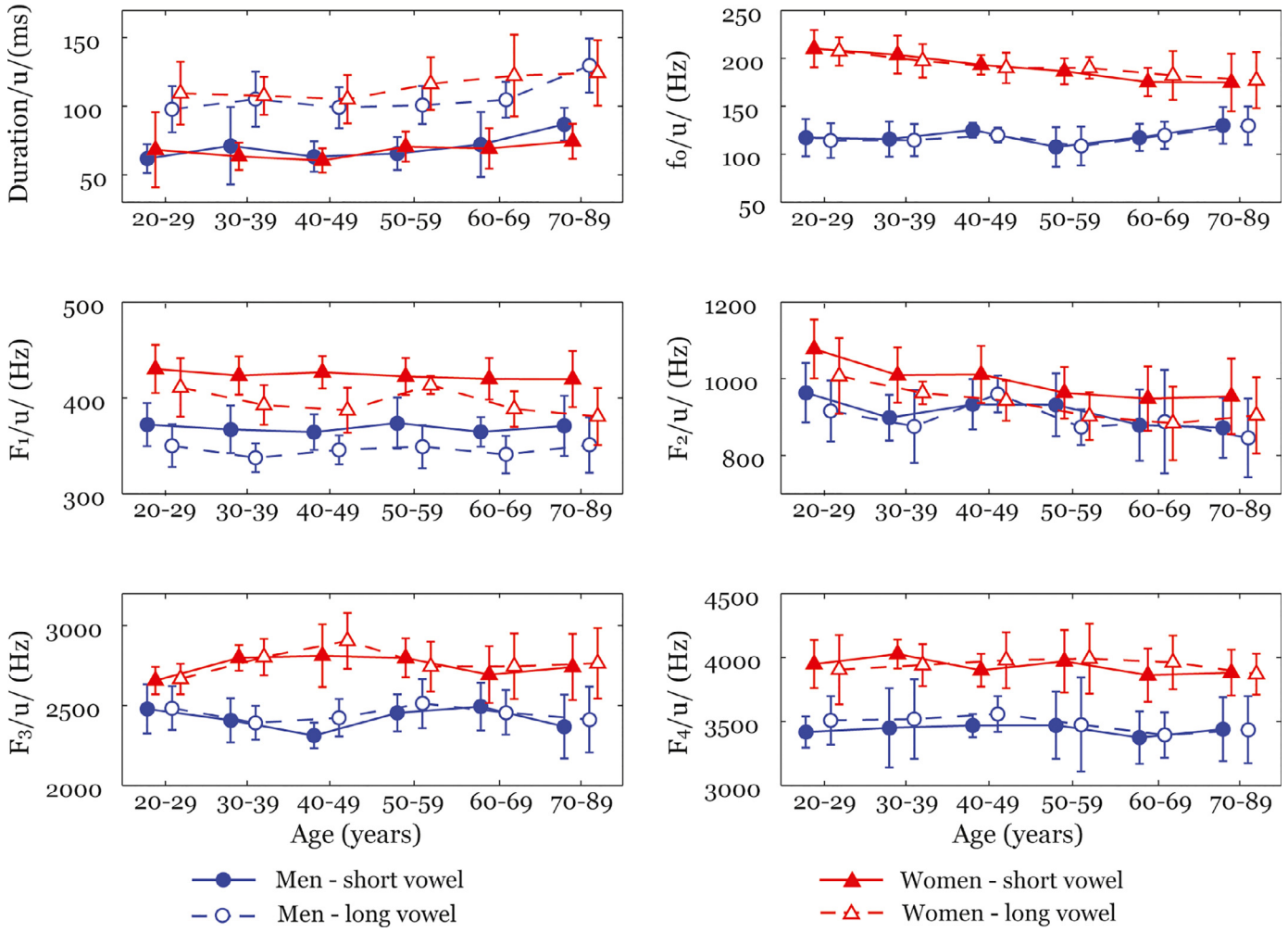


FIGURE 4. The comparison of speech measurements of the vowel /u/. Mean values and standard deviations (error bars) are depicted for both sexes (men, women), and vowel quantities (short, long), presented as a function of age.

rate, the ANOVA showed a significant effect of AGE [$F(5,88) = 7.5$, $P < 0.001$]. *Post hoc* comparisons revealed significantly slower articulation rate in 70–89 age group compared to 20–29 ($P < 0.001$), 30–39 ($P < 0.001$), 40–49 ($P < 0.001$), 50–59 ($P < 0.001$), and 60–69 ($P = 0.004$). With regard to the VSA, there was a significant effect of AGE [$F(5,88) = 2.8$, $P = 0.02$] and SEX [$F(1,88) = 53$, $P < 0.001$]. *Post hoc* comparisons revealed significantly greater VSA in 70–89 age group compared to 20–29 ($P < 0.001$) and 40–49 ($P = 0.04$) age groups. In addition, we found statistically significant correlations between the articulation rate and the average vowel duration ($r = -0.83$, $P < 0.001$) calculated across all monophthongs, as well as between the VSA and the average vowel duration ($r = 0.40$, $P < 0.001$).

There were no other statistically significant findings.

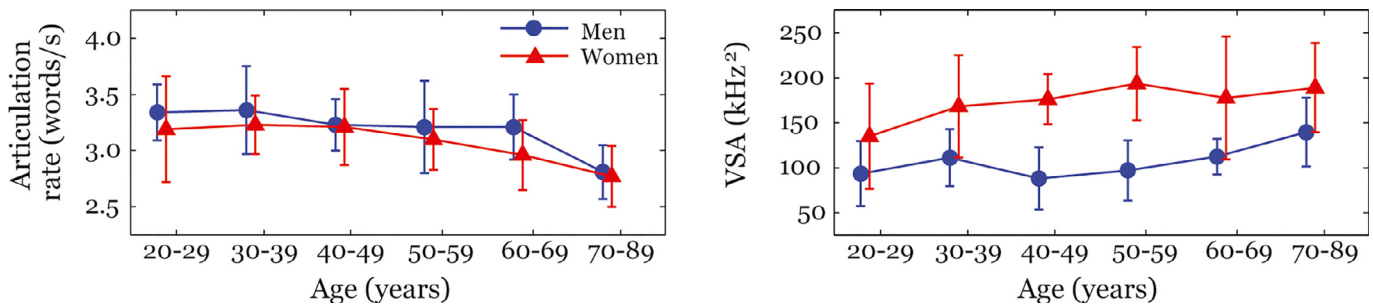


FIGURE 5. The results of the articulation rate and VSA. Mean values and standard deviations (error bars) are shown for both sexes (men, women), presented as a function of age.

TABLE 2.
The Acoustic Characteristics of Czech Vowels for the Adult Male Population

	Duration (ms) Mean/SD (range)	f_o (Hz) Mean/SD (range)	F_1 (Hz) Mean/SD (range)	F_2 (Hz) Mean/SD (range)	F_3 (Hz) Mean/SD (range)	F_4 (Hz) Mean/SD (range)
Short						
/a/	62/12 (39–84)	128/18 (89–173)	572/47 (458–674)	1316/72 (1139–1425)	2387/129 (2101–2700)	3586/241 (3116–4191)
/ɛ/	63/12 (42–97)	123/19 (81–169)	469/32 (371–538)	1658/74 (1514–1842)	2469/123 (2254–2835)	3609/218 (3211–4351)
/ɪ/	64/12 (40–106)	127/19 (88–173)	359/23 (304–410)	1962/112 (1659–2191)	2632/143 (2383–3207)	3685/223 (3268–4348)
/o/	63/12 (35–93)	126/20 (85–174)	460/32 (389–556)	1052/80 (894–1263)	2391/162 (2009–2735)	3424/217 (3079–4247)
/u/	72/19 (38–124)	120/19 (78–164)	369/24 (329–452)	910/81 (755–1085)	2418/158 (2128–2779)	3437/215 (2876–4088)
Long						
/a:/	146/23 (96–208)	119/18 (82–158)	669/54 (524–768)	1257/84 (1086–1442)	2452/138 (2146–2769)	3573/269 (3112–4235)
/ɛ:/	127/18 (80–165)	122/18 (85–158)	525/42 (402–610)	1659/77 (1531–1835)	2495/117 (2306–2949)	3648/246 (3185–4327)
/i:/	107/19 (66–155)	124/19 (83–165)	310/18 (278–366)	2206/115 (1871–2402)	2945/193 (2526–3562)	3717/255 (3316–4532)
/o:/	133/23 (94–194)	121/20 (77–172)	465/35 (353–585)	930/78 (787–1177)	2432/164 (2062–2793)	3369/231 (3020–4314)
/u:/	109/21 (65–173)	119/18 (79–161)	347/22 (320–422)	887/94 (667–1096)	2446/154 (2144–2858)	3476/250 (2870–4259)

Characteristics of Czech monophthongs

The acoustic characteristics, including the vowel duration and f_o, F_1, F_2, F_3, and F_4 frequencies across 10 Czech monophthongs, are listed separately for the adult male and female populations in Tables 2 and 3. There were no statistically significant differences in vowel duration between the sexes (two sample *t* test: $P = 0.52$). The long vowels were 2.0 ± 0.3 (range 1.5–2.5) times longer than were their short counterparts. The average values of the F_1 and F_2 frequencies across 10 Czech vowels with ellipses fit to the data are presented in Figure 6. The marked difference in the spectral patterns of long compared to short vowel counterpart was only found for F_1/a:/, F_1/ɛ:/ F_1/i:/, F_2/i:/, F_2/o:/ and F_3/i:/ formants (supplementary material S2 Table).

DISCUSSION

This study examined the age-related acoustic characteristics of vowels derived from a reading passage across a group of 100 healthy Czech speakers aged between 20 and 90. The primary aim of the study was to evaluate the effect of ageing on vowel articulation, as the previous literature provided

TABLE 3.
The Acoustic Characteristics of Czech Vowels for the Adult Female Population

	Duration (ms) Mean/SD (range)	f_o (Hz) Mean/SD (range)	F_1 (Hz) Mean/SD (range)	F_2 (Hz) Mean/SD (range)	F_3 (Hz) Mean/SD (range)	F_4 (Hz) Mean/SD (range)
Short						
/a/	62/10 (43–86)	203/22 (127–246)	692/57 (543–780)	1584/104 (1336–1841)	2659/181 (2218–2985)	4084/213 (3476–4671)
/ɛ/	66/10 (48–100)	196/25 (121–247)	551/39 (464–625)	1935/105 (1626–2137)	2787/149 (2375–3057)	4103/186 (3656–4440)
/ɪ/	62/11 (45–87)	206/25 (122–256)	432/28 (384–485)	2279/155 (1885–2648)	2988/152 (2554–3287)	4209/193 (3770–4632)
/o/	66/12 (40–98)	202/25 (119–254)	526/36 (447–621)	1206/91 (1022–1429)	2701/185 (2179–3037)	3890/183 (3534–4435)
/u/	68/15 (47–129)	189/24 (110–238)	423/22 (374–476)	990/90 (814–1189)	2745/163 (2344–3074)	3925/183 (3492–4342)
Long						
/a:/	155/19 (110–203)	190/21 (117–233)	815/68 (631–925)	1539/114 (1308–1805)	2711/189 (2264–3092)	4029/205 (3404–4369)
/ɛ:/	137/17 (106–179)	194/24 (116–250)	651/57 (531–772)	1952/113 (1727–2210)	2820/147 (2437–3113)	4171/179 (3719–4532)
/i:/	108/18 (61–151)	199/24 (116–243)	376/34 (306–457)	2633/180 (2186–2991)	3302/187 (2902–3713)	4216/191 (3715–4607)
/o:/	145/21 (106–191)	194/22 (115–240)	537/33 (457–608)	1073/92 (914–1295)	2738/204 (2193–3070)	3833/209 (3334–4427)
/u:/	115/23 (79–169)	189/23 (117–240)	394/26 (311–476)	930/89 (758–1163)	2767/179 (2370–3172)	3935/210 (3496–4338)

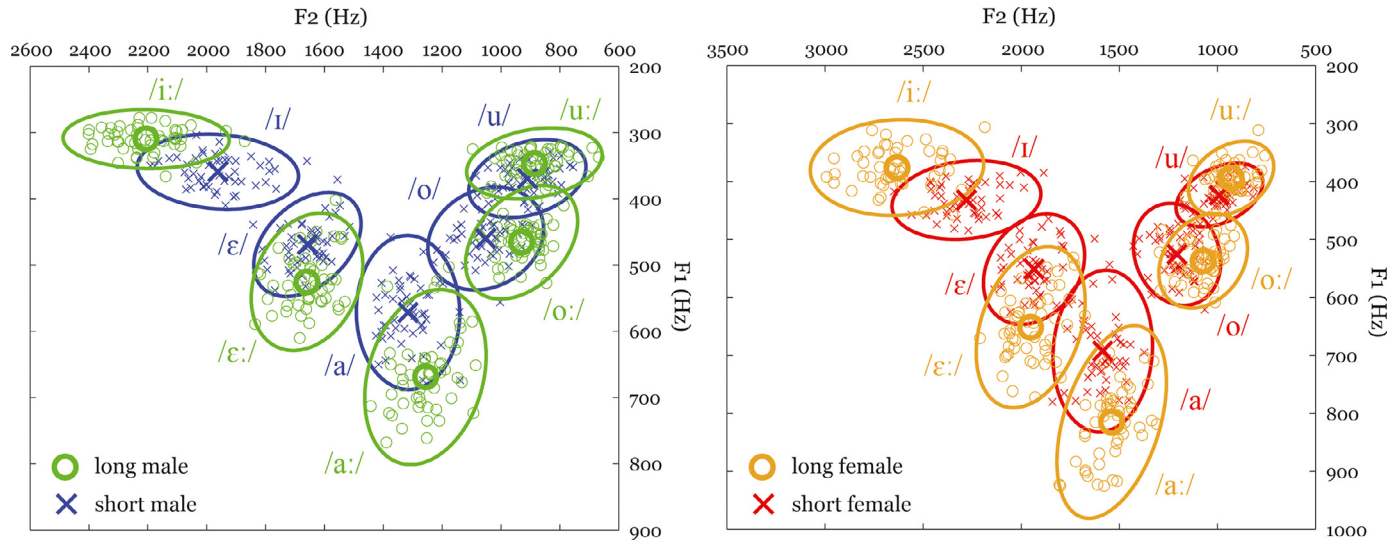


FIGURE 6. Average values of F_1 and F_2 across 10 Czech vowels with ellipses fit to the data presented separately for the male and female populations. Ellipses represent 95% confidence intervals

somewhat inconclusive findings.[2] In fact, knowledge about the typical changes in voice and speech parameters is not only essential for understanding the process of ageing, but may also help to differentiate normal from pathological speech. Since Czech belongs to underdocumented languages, the secondary aim was to provide normative data for the f_o, F_1, F_2, F_3, and F_4 frequencies of all Czech monophthongs.

The findings of this study indicate that age-related changes in pitch are sex-dependent, while age-related alterations in F_2/a/, F_2/u/, VSA, and vowel duration seem to be more consistent in both sexes. Specifically, we observed a clear lowering of f_o with age for women, but no change for men. With regard to formants, we found the lowering of F_2/a/ and F_2/u/ with increasing age, but no statistically significant changes in F_1, F_3, or F_4 frequencies with advanced age. Interestingly, although the alterations in F_1 and F_2 frequencies were rather small, they appeared to be in a direction against vowel centralization, resulting in a significantly greater VSA in the older population. However, it seems that a greater VSA is related partly to longer vowel duration.

Age-dependent acoustic characteristics

In line with the previous literature,[21,24,27,31,32] we observed a significant age-related lowering of f_o in women. By contrast, no alteration of f_o was found in men; this finding is consistent with a study published by Eichhorn et al,[24] but is inconsistent with other studies that reported significant decreases or increases in the male pitch with age.[21,23,26,27,31] The f_o changes in women may be related to a number of age-related physiologic changes, including hormonal changes after menopause, decrease in size of the laryngeal muscles, hardening and possible ossification of the laryngeal cartilages, decreased glandular function, and thickening of the vocal folds.[24] Given that only women showed a significant effect of ageing in this study, we hypothesized that a decrease in f_o for the women may be a consequence of the increase in vocal fold mass due to hormonal changes that occur during menopause.[49]

Across the different corner vowels investigated in our study, a statistically significant increase in vowel duration was revealed with advanced age for both sexes, which is in accordance with earlier studies that reported a longer segmental vowel duration in older men[22,25,26,29] as well as in women.[22,25] The lengthening of vowel duration was most prominent in the oldest group (aged 70–89). A strong, negative correlation between the articulation rate and the average vowel duration was found, indicating that a longer vowel duration is associated with a slowing down of the overall speech tempo. Similar findings were reported by Harnsberger et al,[26] who observed the lengthening of sentence, word, and diphthong durations as a function of age. Nevertheless, the effect of other factors such as preservative coarticulation on the lengthening of vowels in older speakers cannot be excluded.

Most of the previous studies found statistically significant, age-related, and vowel-specific alterations of F_1 and F_2 frequencies that were unique to male and female populations in at least 50% of the formants investigated.[20,21,30] By contrast, our results indicated that F_1 frequency was an age- and sex-independent parameter across all the vowels investigated, while F_2/a/ and F_2/u/ decreased for both sexes and F_2/i/ remained unchanged. These findings are generally in accordance with recent research[24] that investigated native English speakers, in which a marked decline was only reported for F_1/u/ and F_1/æ/ in women and for F_2/u/ in both men and women. Although the observed alterations of F_1 and F_2 frequencies in our cohort were rather small, they appeared to be in the direction against vowel centralization, resulting in a significantly greater VSA in the older population. Nevertheless, as shown previously, the greater VSA is partially related to a longer vowel duration.[25,50] In fact, we

also observed a positive correlation between the size of the VSA and the average vowel duration in our speakers, thus supporting this hypothesis. With regard to the higher formants, former studies investigating F_3 frequency[20,21,24] reported no change[21] or a decline in only a small number of the vowels elicited.[20,24] As no alterations in F_3 or F_4 frequencies were revealed in our study, we agree with previous studies[20,24] neglecting the hypothesis of age-related vocal tract lengthening, which should result in a decrease in all formant frequencies. Since higher formants such as F_3 are thought to be related mainly to the vocal expression of emotions,[3] we hypothesized that no age-related alteration of F_3 and F_4 might be related to the preserved ability of spontaneous use of emotion regulation tactics in older persons.[51]

VOWEL × SEX interactions for F_1/a/, F_2/i/, F_2/u/, vowel duration /a/, and vowel duration /u/ were revealed, indicating vowel specific physiologic differences in measurement ranges between male and female sexes. No consistent AGE × VOWEL interactions were observed to enable thorough discussion.

Characteristics of Czech vowels

The average mean duration across all the Czech vowel qualities was 65 ± 3 milliseconds for short vowels and 128 ± 18 milliseconds for long vowels, with no statistically significant differences between the sexes, resulting in a duration ratio of 1:2 for short and long monophthongs. The observed ratio is slightly higher than it was in the previous study,[33] in which the long vowels were documented as being about 1.7 times longer than their short counterparts. However, the study by Podlipsky, Skarnitzl and Volin[33] analyzed the speech of six professional newsreaders employed by a public broadcaster, while we examined 100 healthy speakers with different professions. Therefore, one might expect professional newsreaders to read more quickly than would normal speakers. With regard to the formant structure of Czech vowels, in accordance with previous literature reporting the short /ɪ/ to be noticeably less close and more central than the long /i:/,[34] we revealed a marked difference in the spectral pattern between /ɪ/ and /i:/ that was associated with changes in F_1, F_2, and F_3 formant frequencies. Although we observed some alterations in the formant structure of F_1/a:/, F_1/ɛ:/, and F_2/o:/, our results tended to confirm the earlier perceptual findings that reported minor qualitative differences in the formant structure of the short and long counterparts of /a/, /ɛ/, /o/, and /u/ vowels.[16]

Limitations of the study

One potential limitation of this study is that the results were based solely on an analysis of a reading passage; thus, the current findings may differ from those obtained via different speaking tasks. We decided to use a reading passage because it represents a more natural task with regard to the influence of lexical and syntactic variables compared to sustained vowel phonation, reading a word list or reading meaningless words in a carrier sentence. Moreover, compared to more complex speaking tasks such as monologues, a reading passage maintains strongly standardized conditions and enables the inclusion of less frequently occurring monophthongs in the investigation. Admittedly, each of the speaking tasks has both advantages as well as disadvantages. For example, the analysis of sustained vowel phonation enables better interlingual comparison but does not reflect common connected speech well. Indeed, a previous study[8] showed two times greater VSA calculated from sustained phonation compared to sentence repetition or the reading passage. Nevertheless, while using the reading passage, we still cannot exclude the influence of prosodic structures and the coarticulatory context on the acoustic characteristics of vowels. Therefore, we decided to use diverse speech materials, including various places of articulation for the preceding consonant, and various positions of the syllables within the word or target words elicited from different positions within the sentence. Notably, we were also heavily limited by the natural structure of Czech; for example, the long vowels in Czech occur 3.5 times less frequently than do their short counterparts, and the vowel /o:/ occurs almost exclusively in loanwords.[16] Finally, the mixed group of healthy speakers aged between 20 and 90 may not have been the optimal age group for the definition of normative Czech formant data due to the possible effects of biological ageing or sociolinguistic differences between the age groups. The sociolinguistic development of Czech within the past 80 years has been traced in the lexical domain, but has not involved the sound patterns of the language. Both the postwar and the post-communist periods were more or less egalitarian rather than being divisive in terms of language use. Not a single account of any generational differences in the pronunciation of vowels or consonants exists. In addition, a comparison of the pronunciation norms across decades suggests an era of relative stability in the vocalic and consonantal systems of Czech.[52–55]

CONCLUSIONS

The acoustic properties of all 10 Czech monophthongs were defined, thus allowing for a comparison with the data reported by other investigators in relation to different languages. With regard to the effect of ageing, the alterations in the vowel formant frequencies across several decades of adult life appear to be small or in a direction against vowel centralization, either because physiological ageing has little effect on formant patterns or because individuals manage to develop a compensatory mechanism for age-related changes in their anatomy and physiology. Our results indicated that an extension of the vowel duration might be such a compensatory mechanism, which helps older subjects to maintain articulatory precision. Future longitudinal research is necessary in order to identify possible compensatory mechanisms for imprecise articulation. From a clinical point of view, as we did not observe any age-related trends towards the reduction or centralization of the VSA in older speakers, and as the decreased vowel area has been documented previously in

the early stages of Parkinson disease[8] and in other neurological conditions,[9,56] the analysis of individual differences in vowel articulation and its variability may be suitable in future for the early detection of neurodegeneration.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at doi:10.1016/j.jvoice.2020.02.022.

APPENDIX A

The first reading passage with the labeled short and long vowels that were used in the acoustic analyses.

I na tom, že člověk si opatří psa, aby nebyl sám, je mnoho pravdy. Pes opravdu nechce být sám. Jen jednou jsem nechal Mindu o samotě v předsíni; na znamení protestu sežrala všechno, co našla, a bylo jí pak poněkud nedobře. Po druhé jsem ji zavřel do sklepa s tím výsledkem, že rozkousala dveře. Od té doby nezůstala sama ani po jedinou minutu.Když píši, chce, abych si s ní hrál. Když si lehnu, považuje to za znamení, že si mně smí lehnout na prsa a kousat mě do nosu. Přesně o půlnoci s ní musím provádět Velkou Hru,při níž se s velikým hlukem honíme, koušeme a kutálíme po zemi. Když se uříti,jde si lehnout; pak si smím lehnout i já, ovšem s tou podmínkou, že nechám dveře do ložnice otevřené, aby se Mindě nestýskalo.

Laktóza je mléčný cukr skládající se z glukózy a galaktózy, který se vyskytuje v mateřském mléce všech savců včetně lidského. Laktóza v mateřském mléce slouží kojencům k tvorbě nervových buněk především pro rychle rostoucí mozek. Hlavním zdrojem laktózy jsou mléko, jogurt, tvaroh, smetana, pudink, sýr a máslo. Výrobky ze sóji a všechny další mléčné náhražky z ořechů či obilovin laktózu neobsahují.

Když jsem se blížil k náměstí, již z dálky jsem slyšel hudbu. Na pódiu umístěném ve středu náměstí tančilo a zpívalo několik dívek. Před pódiem postávaly hloučky lidí, další sledovali vystoupení z okolních balkónů. Od přihlížejících lidí jsem se dozvěděl, že zde probíhá celostátní soutěž v sólovém a chórovém zpěvu. Rozhodl jsem se chvíli zůstat a vychutnal atmosféru.

APPENDIX B

The second reading passage with the labeled short and long vowels that were used in the acoustic analyses.

Když člověk poprvé vsadí do země sazeničku, chodí se na ni dívat třikrát denně: tak co, povyrostla už nebo ne? **I** tají dech, naklání se nad ní, přitlačí trochu půdu u jejích kořínků, načechrává jí lístky a vůbec ji obtěžuje různým konáním, které považuje za užitečnou péči. **A** když se sazeničkapřesto ujme a roste jako z vody, tu člověk žasne nad tímto divem přírody, má pocit čehosi jako zázraku a považuje to za jeden ze svých největších osobních úspěchů.

Později je to už jiné; později člověk osadí svůj záhon s expertní nedbalostí, tak, a teď ukaž, co dovedeš. Když se některá sazenička nepovede, pokrčí nad ní rameny; je to její vina. A že ty druhé rostou, inu, to je samozřejmé; udělal jsem jim dobrou půdu, tak co; byl by jen holý nevděk, kdyby nerostly.

Když člověk jede poprvé v životě za hranice své vlasti, cítí především strach z toho neznáma, do kterého se vrhá, ale nedává to příliš najevo. Za druhé cítí ohromnou odvahu k dobrodružství, pýchu dobyvatele a statečnost objevitele; má v sobě malou dušičku, ale nesmírně jaksi načepýřenou a nadouvající se téměř bolestně. Abyste věděli, já jedu do širého a cizího světa; já nejsem jen tak někdo, nýbrž veliký dobrodruh.

A když člověk takto jede po desáté nebo po dvacáté, stáhne si cestovní čepici do očí, založí ruce a oddává se jakémusi sebelitování. Bože, jaká otrava, jaká obtíž! Zas abych se tloukl po všech čertech a ďáblech, měl nepříjemnosti s celníky, musel měnit peníze a hledal nocleh v hotelu, který neznám. Čert mi byl dlužen tuhle cestu.

A tak je to se vším, krom narození a smrti; to oboje má Člověk, bohudík, odbyto hned napoprvé. V tom pak je celá rovnováha a vyváženost života: že totéž, co jedna část lidí dělá po prvé, s objevitelským zápalem a úžasem, druhá část dělá po sté, zamlkle, nerada a s rutinou starého návyku.

REFERENCES

1. Kent RD, Weismer G, Kent JF, et al. Acoustic studies of dysarthric speech: methods, progress, and potential. *J Commun Disord*. 1999;32:141–186.
2. Kent RD, Vorperian HK. Static measurements of vowel formant frequencies and bandwidths: a review. *J Commun Disord*. 2018;74:74–97.
3. Waaramaa T, Alku P, Laukkanen AM. The role of F3 in the vocal expression of emotions. *Logoped Phoniatr Vocol*. 2006;31:153–156.
4. Titze IR. *Principles of Voice Production*. Iowa City, USA: National Center for Voice and Speech; 2000. Second printing.
5. Roy N, Nissen SL, Dromey C, et al. Articulatory changes in muscle tension dysphonia: evidence of vowel space expansion following manual circumlaryngeal therapy. *J Commun Disord*. 2009;42:124–135.
6. Sapir S, Spielman JL, Ramig LO, et al. Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: acoustic and perceptual findings. *J Speech Lang Hear R*. 2007;50:899–912.
7. Takatsu J, Hanai N, Suzuki H, et al. Phonologic and acoustic analysis of speech following glossectomy and the effect of rehabilitation on speech outcomes. *J Oral Maxil Surg*. 2017;75:1530–1541.
8. Rusz J, Cmejla R, Tykalova T, et al. Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. *J Acoust Soc Am*. 2013;134:2171–2181.
9. Rusz J, Hlavnicka J, Tykalova T, et al. Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder. *Sleep Med*. 2016;19:141–147.
10. Ng AK, Koh TS, Baey E, et al. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Med*. 2008;9:894–898.
11. Rusz J, Cmejla R, Ruzickova H, et al. Evaluation of speech impairment in early stages of Parkinson's disease: a prospective study with the role of pharmacotherapy. *J Neural Transm*. 2013;120:319–329.
12. Skodda S, Gronheit W, Schlegel U. Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *Plos One*. 2012;7: e32132.
13. Plonkowski M. Using bands of frequencies for vowel recognition for Polish language. *Int J Speech Technol*. 2015;18:187–193.
14. Li M, Han KJ, Narayanan S. Automatic speaker age and gender recognition using acoustic and prosodic level information fusion. *Comput Speech Lang*. 2013;27:151–167.
15. Morrison GS. Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. *J Acoust Soc Am*. 2009;125:2387–2397.

16. Simackova S, Podlipsky VJ, Chladkova K. Czech spoken in Bohemia and Moravia. *J Int Phon Assoc*. 2012;42:225–232.
17. Themistocleous C. Dialect classification using vowel acoustic parameters. *Speech Commun*. 2017;92:13–22.
18. McCarthy KM, Evans BG, Mahon M. Acquiring a second language in an immigrant community: the production of Sylheti and English stops and vowels by London-Bengali speakers. *J Phonetics*. 2013;41:344–358.
19. Ramig LA. Effects of physiological aging on speaking and reading rates. *J Commun Disord*. 1983;16:217–226.
20. Xue SA, Hao GJP. Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. *J Speech Lang Hear R*. 2003;46:689–701.
21. Torre P, Barlow JA. Age-related changes in acoustic characteristics of adult speech. *J Commun Disord*. 2009;42:324–333.
22. Albuquerque L, Oliveira C, Teixeira A, et al. Impact of age in the production of European Portuguese vowels. Presented at: *Interspeech 2014: 15th Annual Conference of the International Speech Communication Association*. 2014:940–944.
23. Cox VO, Selent M. Acoustic and respiratory measures as a function of age in the male voice. *J Phonet Audiol*. 2015;1:105.
24. Eichhorn JT, Kent RD, Austin D, et al. Effects of aging on vocal fundamental frequency and vowel formants in men and women. *J Voice*. 2018;32.
25. Fletcher AR, McAuliffe MJ, Lansford KL, et al. The relationship between speech segment duration and vowel centralization in a group of older speakers. *J Acoust Soc Am*. 2015;138:2132–2139.
26. Harnsberger JD, Shrivastav R, Brown WS, et al. Speaking rate and fundamental frequency as speech cues to perceived age. *J Voice*. 2008;22:58–69.
27. Harrington J, Palethorpe S, Watson CI. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. Presented at: *Interspeech 2007: 8th Annual Conference of the International Speech Communication Association, Vols 1-4*. 20071081.
28. Linville SE, Fisher HB. Acoustic characteristics of womens voices with advancing age. *J Gerontol*. 1985;40:324–330.
29. Liss JM, Weismer G, Rosenbek JC. Selected acoustic characteristics of speech production in very old males. *J Gerontol*. 1990;45:P35–P45.
30. Rastatter MP, Jacques RD. Formant frequency structure of the aging male and female vocal-tract. *Folia Phoniatr*. 1990;42:312–319.
31. Sebastian S, Babu S, Oommen NE, et al. Acoustic measurements of geriatric voice. *J Laryngol Voice*. 2012;2:81–84.
32. S Schötz, Acoustic analysis of adult speaker age. In Müller C, ed. *Speaker Classification I. Fundamentals, Features, and Methods*. Berlin: Springer. 2007;88−107.
33. Podlipsky VJ, Skarnitzl R, Volin J. High front vowels in Czech: a contrast in quantity or quality? Presented at: *Interspeech 2009: 10th Annual Conference of the International Speech Communication Association, Vols 1-5*. 2009:144.
34. Dankovicova J. The domain of articulation rate variation in Czech. *J Phonetics*. 1997;25:287–312.
35. Grepl M, Furst T, Pesak J. The F1−F2 vowel Chart for Czech whispered vowels a, e, i, o, u. *Biomed Pap*. 2007;151:353–356. Available at: http://biomed.papers.upol.cz/magno/bio/2007/mn2.php. Accessed March 23, 2020.
36. Pribil J, Pribilova A, Matousek J. Comparison of formant features of male and female emotional speech in Czech and Slovak. *Elektron Elektrotech*. 2013;19:83–88.
37. Skarnitzl R, Volin J. *Referencni hodnoty vokalickych formantu pro mlade dospele mluvci standardni cestiny*. 2012. Akustické listy18:7-11. Available at: https://www.czakustika.cz/misc/aklisty/18-1.pdf. Accessed March 23, 2020.
38. Ayoub MR, Larrouy-Maestri P, Morsomme D. The effect of smoking on the fundamental frequency of the speaking voice. *J Voice*. 2019;33:802.e11–802.e16.
39. Skarnitzl R, Eriksson A. The acoustics of word stress in Czech as a function of speaking style. Presented at: *Interspeech 2017: 18th Annual Conference of the International Speech Communication Association, Vols 1-4*. 2017:3221–3225.
40. Palkova Z, Volin J. The role of F0 contours in determining foot boundaries in Czech. Presented at: *ICPhS 2003: 15th International Congress of Phonetic Sciences*. 2003:1783–1786.
41. Boersma P. Praat, a system for doing phonetics by computer. *Glot Int*. 2002;5:341–345.
42. Novotny M, Rusz J, Cmejla R, et al. Automatic evaluation of articulatory disorders in Parkinson's disease. *Ieee-Acm T Audio Spe*. 2014;22:1366–1378.
43. Hillenbrand J, Getty LA, Clark MJ, et al. Acoustic characteristics of American english vowels. *J Acoust Soc Am*. 1995;97:3099–3111.
44. Rusz J, Cmejla R, Ruzickova H, et al. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J Acoust Soc Am*. 2011;129:35–67.
45. Lam J, Tjaden K, Wilding G. Acoustics of clear speech: effect of instruction. *J Speech Lang Hear Res*. 2012;55:1807–1821.
46. Liu HM, Kuhl PK, Tsao FM. An association between mothers' speech clarity and infants' speech discrimination skills. *Dev Sci*. 2003;6:F1–F10.
47. Liu H, Ng M. Formant characteristics of vowels produced by Mandarin Esophageal Speakers. *J Voice*. 2009;23:255–260.
48. Skodda S, Visser W, Schlegel U. Vowel articulation in Parkinson's disease. *J Voice*. 2011;25:467–472.
49. Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. *J Voice*. 1999;13:424–446.
50. Gendrot C, Adda-Decker M. Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. Presented at: *Interspeech 2005: 6th Annual Conference of the International Speech Communication Association, Vols 1-4*. 2005:2483−2456.
51. Livingstone KM, Isaacowitz DM. Age similarities and differences in spontaneous use of emotion regulation tactics across five laboratory tasks. *J Exp Psychol Gen*. 2019;148:1972–1992.
52. Hala B. *Vyslovnost Spisovne Cestiny I*. Praha: Academia; 1967.
53. Romportl M. *Vyslovnost Spisovne Cestiny II*. Praha: Academia; 1978.
54. Hurkova J. *Ceska Vyslovnostni Norma*. Praha: Scientia; 1995.
55. Stepanova V. *Vyvoj Kodifikace Ceske Vyslovnosti*. Praha: Academia; 2019.
56. Rusz J, Bonnet C, Klempir J, et al. Speech disorders reflect differing pathophysiology in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *J Neurol*. 2015;262:992–1001.