MDPI

Article

On the Primary Influences of Age on Articulation and Phonation in Maximum Performance Tasks

Fredrik Karlsson [1,*] and Lena Hartelius [2]

[1] Department of Clinical Science, Umeå University, 907 36 Umeå, Sweden
[2] Department of Health and Rehabilitation, University of Gothenburg, 405 30 Gothenburg, Sweden; lena.hartelius@neuro.gu.se
* Correspondence: fredrik.k.karlsson@umu.se

Abstract: Maximum performance tasks have been identified as possible domains where incipient signs of neurological disease may be detected in simple speech and voice samples. However, it is likely that these will simultaneously be influenced by the age and sex of the speaker. In this study, a comprehensive set of acoustic quantifications were collected from the literature and applied to productions of sustained [a] productions and Alternating Motion Rate diadochokinetic (DDK) syllable sequences made by 130 (62 women, 68 men) healthy speakers, aged 20–90 years. The participants were asked to produce as stable (sustained [a] and DDK) and fast (DDK) productions as possible. The full set of features were reduced to a functional subset that most efficiently modeled sex-specific differences between younger and older speakers using a cross-validation procedure. Twelve measures of [a] and 16 measures of DDK sequences were identified across men and women and investigated in terms of how they were altered with increasing age of speakers. Increased production instability is observed in both tasks, primarily above the age of 60 years. DDK sequences were slower in older speakers, but also altered in their syllable and segment level acoustic properties. Increasing age does not appear to affect phonation or articulation uniformly, and men and women are affected differently in most quantifications investigated.

Keywords: oral diadochokinesis; sustained phonation; age-related changes; cross-validated acoustic model

1. Introduction

Our age is one of the characteristics that is conveyed through voice. People are usually very good at determining a person's age by listening to samples of their voice, accuracy depending on the type of speech sample and age of the speaker (Ptacek and Sander 1966). The reason for this is that voice and speech are audible reflections of age-related physiological changes. These may also be influenced by the development of disease- or disorder-related changes such as functional or organic dysphonia or neurological disease. Disease-related and as age-related changes may have similar characteristics and the effects of a disease can, therefore, only be accurately assessed once age-related changes are well understood.

The physiological factors contributing to the changes characteristic of an aging voice (presbyphonia) are numerous, including decreased pulmonary function, changes in the neuromuscular properties of laryngeal musculature as well as the lamina propria of the vocal folds. Age-related changes within the lamina are reported to be primarily tissue loss and decreased glandular secretions, which affect the elasticity of the vocal folds and thus, the vibratory pattern (Kahane 1983; Tomita et al. 2006; Bloch and Behrman 2001). Muscle atrophy is one of the described neuromuscular changes, and when affecting the thyroarytenoid muscle, a vocal fold bowing is produced, resulting in incomplete closure during phonation, which, in turn, contributes to changes in overall voice quality (Linville 2002) rather than resulting in a simple change in f_0 (Titze 2011). The main perceptual char-

acteristics of presbyphonia are vocal weakness, hoarseness, breathiness, and instability (Leeuw and Mahieu 2004).

The physiological changes of the voice are, to a certain degree, sex-specific and conditioned by hormonal changes. The changes of the female larynx include thickening of the vocal folds and increased prominence of the vocal process, leading to greater vocal fold contact and closing of the posterior gap characteristic of the phonation of younger women (Pontes et al. 2005; Pontes et al. 2006). This in turn explains the audible decrease in breathiness and the lowering of f_0 described in aging women (Linville 1992; Gorham-Rowan and Laures-Gore 2006; Eichhorn et al. 2017). Aging men show an opposite change, where the vocal folds gradually tend to thin, which is thought to contribute to vocal fold bowing, increased breathiness, and higher fundamental frequency (f_0) (Gorham-Rowan and Laures-Gore 2006; Gugatschka et al. 2010).

The perceived hoarseness, breathiness, and instability and their relationships with selected acoustic measures were investigated in the study by Gorham-Rowan and Laures-Gore (2006) and it was found that f_0 standard deviation, amplitude perturbation quotient, and noise-to-harmonic ratio were significantly influenced by age of the speaker. No significant differences in perceived hoarseness were found between young and old speakers, but the younger women were perceived as significantly more breathy compared to the older women. The correlations between perceptions of hoarseness and breathiness were moderate. A recent systematic review and meta-analysis by Rojas et al. (2020) summarizes the findings of 47 studies including almost 4000 participants and conclude that f_0 and signal-to-noise-ratio as well as measures of instability in f_0 and amplitude (jitter, shimmer) were acoustic parameters that differed between age groups, with the correlating perceptual parameters being severity of dysphonia, roughness, breathiness, strain, instability, and presence of loudness and pitch abnormalities.

The findings regarding changes in sustained phonation reported in the review by Rojas et al. (2020) are confirmed by Tucker et al. (2021). In addition, Tucker et al.'s overview of the literature also includes data on other kinds of speech production. They report both longitudinal and cross-sectional studies of speech changes, as reflected by formant frequency changes and durational changes. Several studies report a decrease in F_1 and F_2 for older adults and conflicting findings regarding F_3 and F_4. The changes of F_1—in both men and women—may be explained by increased oral cavity length and vocal tract volume. Concerning durational changes in articulatory rather than speech rate level, syllable and segment durations, although phoneme-dependent, are longer for older compared to younger adults. These changes contribute to the overall finding of slower speech rate with increasing age.

A speaker's maximum ability to produce syllables fast and at a stable pace may be assessed using an oral diadochokinetic task. Diadochokinesis is generally described as the ability to make antagonistic movements in quick succession, and is used in speech–language pathology testing to provide valuable clinical information regarding neuromuscular control and coordination. The task is also widely used as a base for detecting early signs of disease-related speech impairment (Karlsson and Hartelius 2019; Karlsson et al. 2020; Novotny et al. 2020; Rong 2020; Solomon et al. 2021). How age affects oral diadochokinesis has not been thoroughly investigated, but some results suggest that syllable repetitions/papapa, tatata, kakaka/(Alternating Motion Rates, AMR) and/patakapataka/ (Sequential Motion Rates, SMR) produced by older adults are slower and more variable both in duration and amplitude compared to younger adults (Amerman and Parnell 1992; Parnell and Amerman 1987). There are also some previous results that suggest a link between increasing age and the ability to alternate between voiced and voiceless states, specifically. Lombard and Solomon (2019) investigated a large group of neurologically and vocally healthy men and women rapidly repeating strings of /glottal plosive + a/ and /hahahaha/. They found that syllable repetition rate decreased significantly with age in both tasks, which implies that the age-related change is related to the phonated nucleus portion of the syllable. Variability did not increase in older speakers, and neither rate nor

variability differed significantly between men and women. Thus, it cannot currently be concluded based on the research available that a deterioration in phonatory control similar to that observed to be a possible marker of disease-related speech deterioration (Goberman and Blomgren 2008; Eklund et al. 2014; Karlsson et al. 2014; Tanaka et al. 2015; Karlsson and Hartelius 2019) is not also simultaneously increased in speakers due to increased age.

In summary, physiologically conditioned speech changes reflect healthy as well as disease-related processes and help us identify speakers' age, sex and health condition. The changes are multidimensional and any attempt to describe speech evolution needs to be multiparametric and include several types of speech tasks that challenge the speech production system. The present study brought together the most comprehensive set of acoustic measures applicable to sustained vowel productions and DDK sequences, and for which there exists a publicly available implementation currently available, in an attempt to explore and define the optimal set of measures that account for variation in age, determined separately for men and women.

2. Materials and Methods

Recordings of 130 participants (women aged 57.0 ± 17.9 years; men aged 57.5 ± 17.1 years) were included in this study. The participants had been recorded as healthy control speakers in two studies of the effects of neurological diseases on speech and had reported no known neurological disease or problems affecting voice, speech, or ability to understand speech. An overview of the age and sex distribution of the speakers is presented in Table 1. The participant ages have been grouped into age categories in the table for the purpose of providing a concise presentation of the participants; the actual age of the participants was used in all parts of the analysis procedures in the current study. The two studies differed in their number of participants (100 and 30, respectively) and in the average age of participants (54.7 ± 18.4 and 65.8 ± 10.2 years) but provided a consistent sex distribution for this study when combined (Set 1: 54.6 ± 18.6 (women), 54.8 ± 18.4 (men); Set 2: 66.4 ± 11.7 (women), 65.3 ± 9.3 (men)). All participants were native speakers of Swedish.

The participants were asked to perform all the tasks of Swedish dysarthria standardized test, including a sustained and stable vowel [a] and productions of DDK sequences ([pa . . .], [ta . . .], and [ka . . .]) as fast and evenly as they could. The participant's performances of the tasks were recorded using either an external RolandQuad-Capture sound card connected to a Sennheiser HSp 4 (Sennheiser, Wennebostel, Wedemark, Germany) headset microphone with an MZA 900 P phantom adapter (Sennheiser, Wennebostel, Wedemark, Germany), or either a digital audio tape recorder (Panasonic SV 3800, Panasonic Corporation, Osaka, Japan) or digital audio flash recorder (Marantz PMD 660, Marantz, Kanagawa, Japan) connected to a head-mounted microphone (Sennheiser MKE 2 P-C, Sennheiser, Wennebostel, Wedemark, Germany). The recordings made using the Marantz PMD 660 recorder were down-sampled from 48 kHz to 44.1 kHz before performing the acoustic analyses to keep the frequency ranges consistent. All recordings were made in either a quiet room or a sound-treated booth.

Table 1. An overview of the number of participants divided into in ten-year age ranges.

Sex	20–29	30–39	40–49	50–59	60–69	70–79	≥80 yrs	Total
Women	9	3	5	13	13	16	3	62
Men	6	8	5	12	19	16	2	68
Total	15	11	10	25	32	32	5	130

The sustained [a] recordings were subjected to a manual markup procedure where start and end points were identified. Each sustained [a] was then analyzed in terms of its acoustic properties using the Voice Analysis Toolbox (version 1.0), which is a MATLAB™ toolbox. This toolbox was developed for early detection and telemonitoring of speakers with Parkinson's disease and computes 339 measures, including a range of acoustic quantifications of f_0 and amplitude perturbation, harmonic-to-noise ratios, the linear predictive

coding coefficients, the Mel Frequency Cepstrum (MFCC, 32 filters), and a (db8) wavelet decomposition of the signal. The toolbox further computes specific measures of dysphonia (Glottal to Noice Excitation ratio, Detrended Fluctuation Analysis, Recurrence Periodicity Density Entropy, Pitch Period Entropy, Empirical Mode Decomposition Excitation Ratios, Vowel Fold Excitation Ratios, and Glottal Quotient). The implementation details of these measures have been discussed in detail elsewhere (Tsanas et al. 2010b, 2011; Tsanas 2012). Due to the extensive number of measures being computed by the toolkit, we introduce a terminology and description for measures that are of primary importance for the current investigation when presenting the results of the feature selection procedure.

In addition, each sustained vowel was quantified using the Voice Report features (Maryn et al. 2009) of Praat (Maryn et al. 2009) and of PraatSauce (Boersma and Weenink 2001), which computes (using the notation of Titze et al. (2015)) the f_0, the frequency (F_1, F_2, and F_3) and bandwidths (B_1, B_2, and B_3) of the first three formants, with bandwidths computed according to the formulation of Hawks and Miller (1995). The amplitudes of the first, second and fourth harmonic (L_1, L_2, and L_4) were computed and returned in two forms—the original estimate and an additional estimate that compensates for the contribution of nearby formants (Iseli et al. 2007). Further, the amplitudes of harmonics closest to F_1, F_2, and F_3 and closest to 2000 Hz and 5000 Hz, respectively, were extracted. A standardized notation for these harmonic amplitude measurements was not established by Titze et al. (2015) but are referred to as A1, A2, A3, H2K, and H5K in the software (please note the lack of subscripts in this notation). The software further computed several difference measures (L_1–L_2, L_2–L_4, L_1–A1, L_1–A2, L_1–A3, L_4–H2K, and H2K–H5K) on the harmonic amplitude measurements, in formant-corrected (Iseli et al. 2007) and uncorrected forms, as well as the Cepstral Peak Prominence (CPP) (Hillenbrand et al. 1994), as measures of voice quality. In addition, we computed the slope and standard deviation of the f_0 and amplitude measurements across the sustained [a] divided into 5 ms analysis windows.

The produced DDK sequences were analyzed by marking the start and end times of each syllable, along with the start and end of the consonant and vowel portions. The identification of start and end points of the syllables, vowels and consonants was conducted by an experienced transcriber and based on the auditory impression combined with the information conveyed in the wideband spectrogram and the shape of the waveform. Each syllable sequence was then summarized by their durational and phonation properties in accordance with the procedure described in previous reports where the effect of Parkinsonian dysarthria was investigated (Karlsson and Hartelius 2019; Karlsson et al. 2020). The stability of syllable durations and amplitudes across a sequence, both in terms of short-term fluctuations and as overall trends across a sequence (Wang et al. 2009; Schmitz-Hübsch et al. 2011; Skodda 2011; Karlsson et al. 2020) were assessed. Further, the fraction of a syllable's duration made up by the vowel, and the amplitude relationship between consonant and vowel portions (Karlsson and Hartelius 2019; Karlsson et al. 2020), were analyzed in terms of averages and stability across a sequence. The release transients of consonants were assessed by their prominence relative to their acoustic surroundings (Karlsson et al. 2014). The voicing of the voiceless consonant (Karlsson et al. 2012) and devoicing of the vowel, both in overall terms and at the beginning and end of the segment, are marked in terms of their average percentage of segment durations affected, as well as their stability across the sequence. All measures used to quantify DDK sequences are presented in Table 2.

The acoustic summaries of each speaker's performance in the two speech tasks were analyzed in two steps to meet the aims of this investigation. The sustained vowel and DDK tasks were analyzed separately throughout the procedure to assess their association with the age and sex of the speaker independently.

Table 2. The duration and amplitude measures used to quantify diadochokinetic (DDK) sequences.

Measure Type	Abbreviations	Description
Segmental measures	%N, O/N Ampl.	The proportion between the nucleus duration/amplitudes compared to that of the syllable onset
	RTP, RTP (sd)	Mean and standard deviations of the Release transient prominences of plosives
	mRTP, mRTP (sd)	Difference between the amplitude of the release transient and the median amplitude of the syllable onset
	%Phon, Prog. %Phon	The average and slope of the percentage of the duration of the syllable onset in which phonation was detected
	%Phon_init, %Phon_med, %Phon_final	Percentage of the first, the second and the third of the syllable onset in which phonation was detected
	%NPhon, Prog. %Nphon	Percentage of the duration of the syllable nucleus in which phonation was detected
	%NPhon_init, %NPhon_med, %NPhon_final	Percentage of the first, the second and the third of the syllable nucleus in which phonation was detected
Syllable measures	Rate, Syll Ampl	The average of syllable rate and syllable amplitude
	relStab5–12, Ampl. relStab5–12	Average duration/amplitude of syllables 5–12 in a sequence as a percentage of that of syllables 1–4
	relStab13–20, Ampl. relStab13–20	Average duration/amplitude of syllables 13–20 in a sequence as a percentage of that of syllables 1–4
	%PA, %AD	Change in durations/amplitudes from syllables 5–12 to syllables 13–20 in a sequence
	Rate slope, Ampl. slope	Slope of a regression line fitted to syllable durations or amplitudes across a sequence
Measures of short-term variability	Jitter, Shimmer nPVI, nPVI_A	Sum of changes in duration or amplitude from one syllable to the next Sum of syllable-to-syllable changes, normalized to the local average
	AP, APQ3	Sum of per-syllable deviances from the local average (for the preceding syllable, the syllable itself and the following syllable)
	PPQ5, APQ5	Sum of per-syllable deviances from the local average (for the 2 preceding syllables, the syllable itself and the 2 following syllables)
	DDP, DDP_A	Sum of changes in differences going into the next syllable compared with the change coming into the current syllable
Measures of overall variability	Rate (sd), Syll Ampl (sd)	The standard deviations of syllable rates and amplitudes
	COV5–20, COV5–20_A	The standard deviation of rate and amplitudes of syllables 5–20 normalized by the average duration/amplitudes of syllables 1–4

In the first step of the statistical analysis procedure, a feature selection procedure was performed to reduce the large number of measurements of each task to the most efficient subset with minimal inter-measure correlation. It was reasoned that the most functional subset of acoustic measurements of a task would be those which, when used in a model, would provide the best prediction of the age of speakers on which the model had not been trained. To determine the features of interest, therefore, cross-correlation procedures were performed for each speech task (sustained [a] and DDK) to determine the best sex-specific set of features to include in the model. In the cross-correlation, 20% of men and women, respectively, were randomly assigned to a validation set. The acoustic measures for the remaining 80% of speakers were used to train an L1 penalized regression model (Friedman et al. 2010) in a 10-fold cross-validation procedure, in which the penalization parameter lambda was selected that resulted in minimal deviance when predicting age of the speaker in the cross-validation fold. The set of measures that remained after penalization of the statistical models of men's and women's performances of a task were collected to form the set of measures of particular interest in terms of their sex-specific differences between younger and older speakers. The value of the collection of measures for describing sex-specific age differences was evaluated using the percent explained variance of the age of speakers in the validation set.

In the second step of the analysis procedure, the identified measures of interest for each speech task were analyzed in terms of their sex-specific change with age of the speaker. All were used in this analysis.

3. Results

An illustration of the feature selection procedure performed using L1 regularization of a linear model of the speakers' age based on sustained [a] measurements is presented in Figure 1. The horizontal axis shows the value of the penalization parameter lambda that may be selected to zero out the predictors which add the least value to the model. The vertical axis indicates the mean-squared error (MSE) of the model when the given lambda is applied to reduce the number of predictors, with standard error bands across the 10 folds used in model training. The number of non-zero predictors left when applying a specific lambda is indicated on the top of the graph. Two vertical dashed lines indicate lambda values of particular interest. The rightmost vertical line indicates the number of predictors that provide the smallest MSE when predicting age of the speaker in the holdout fold. The leftmost vertical dashed line indicates the lambda, which increases the model error at the most standard error compared to the optimal model but using a smaller number of predictors.
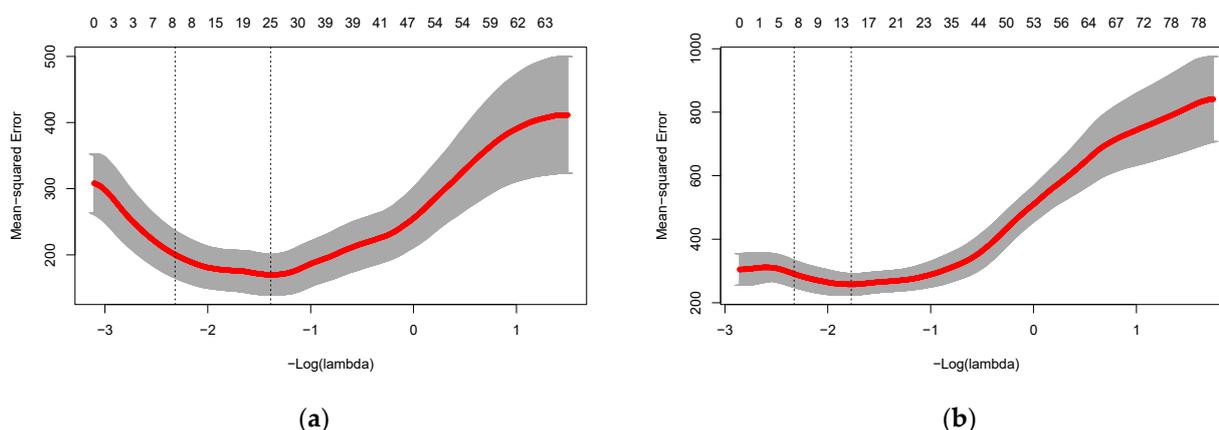

(**a**)   (**b**)

Figure 1. Visualization of the procedure used to train a model of participants' age based on the acoustic parameters of the sustained vowel [a]. The regularization parameter lambda was selected so that the mean-squared error (MSE) was within one standard error (left dashed horizontal line) from the model with the smallest MSE. (**a**) The model of participating women's ages; (**b**) The training of the model for men. The number of acoustic measures of [a] that were identified to primarily contribute to the prediction of speakers' ages is indicated at the top of each subfigure.

Across men and women, twelve unique predictors were identified to contribute to the prediction of a speakers' age from acoustic measures of a sustained [a]. The measures identified were the frequency and bandwidths of the second formant (F_2 and B_2), the difference in uncorrected amplitudes of the harmonics closest to 2 kHz and 5 kHz frequency regions (H2KH5Ku), jitter (Jitter->F0_abs_diff), and the standard deviation of Harmonic to Noise ratio (HNR_std). The standard deviation of the Glottal to Noise Excitation Ratio (Kalwa and Patil 2015) was also selected, as well as the average of the 1st, 5th and 9th Mel Frequency Cepstral Coefficient (MFCC; e.g., mean_MFCC_1st coef) and the standard deviations of the 10th and 12th MFCC across the vowel (std_MFCC_10th coef and std_MFCC_12th coef). The 1st, 5th 9th, 10th and 12th Mel frequency spaced bands correspond to the 76–160 Hz, 473–601 Hz, 1074–1267 Hz, 1267–1481 Hz, and 1719–1982 Hz frequency bands, respectively. Finally, the standard deviation of the Teager–Kaiser Energy Operator (TKEO) of the first-level wavelet decomposition of the f_0 signal track (app_TKEO_std_1 coef) (Tsanas et al. 2010a) was indicated to contribute to the model of speaker age. The pattern of change in these predictors is presented in Figure 2, where all measures are indicated on the horizontal axis in accordance with their names as indicated

by PraatSauce, the Praat Voice Report, or Voice Analysis Toolbox. The models explained 6% (men) and 26% (women), respectively, of the variance in the validation set of speakers. The average error of predicted age was 0.9 ± 17.1 years for men and 0.14 ± 14.3 years for women in the testing set. If evaluated within the same data on which it was trained (training set), the models explained 23% (men) and 45% (women) of the variance, respectively. If applied to all speakers, the models predicted age with an average error of 0.16 ± 15.1 for men and 0.02 ± 13.1 for women. The accuracies of age prediction across the range of speaker ages based on sustained vowel measures are presented in Figure 3a. All pairwise correlations between acoustic vowel measures and age of the speaker are provided for both men and women in correlation matrices in Supplementary Material A.
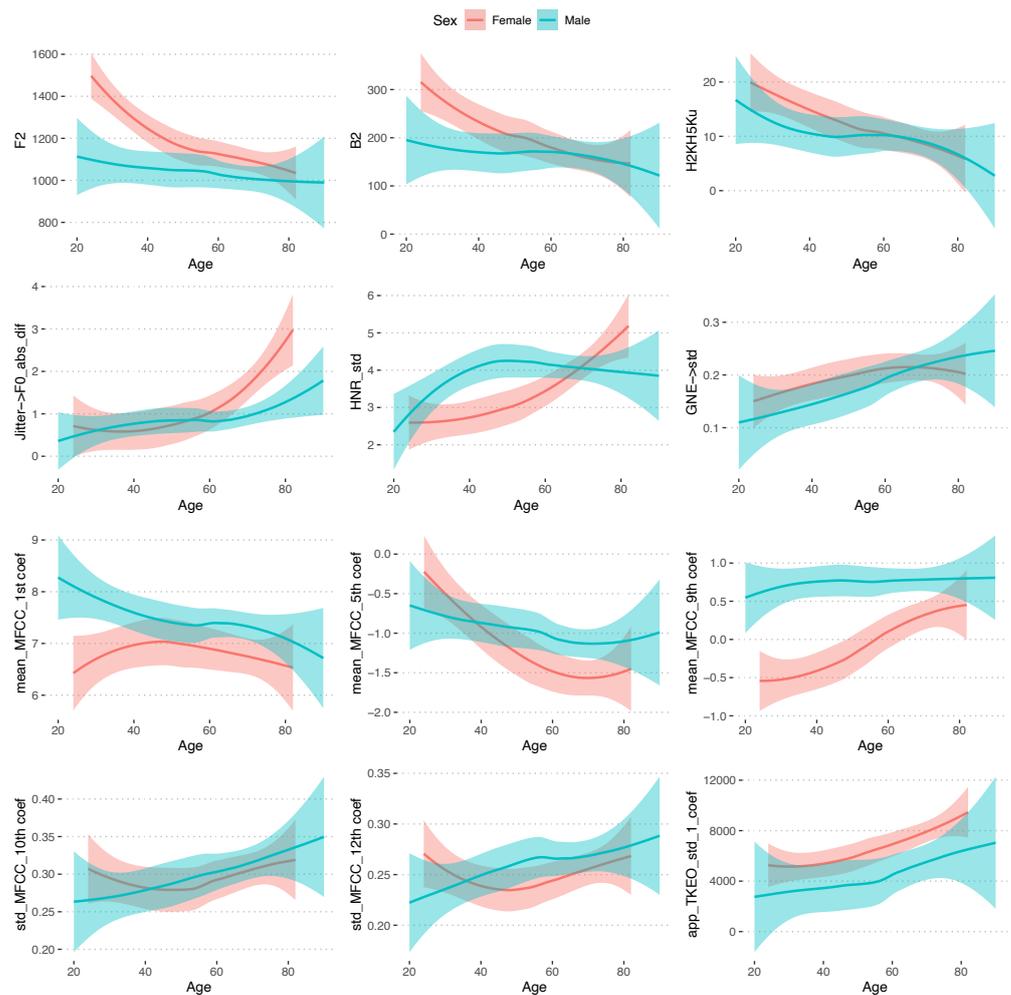
Figure 2. The sex-specific pattern of change with age for identified acoustic measures of a sustained [a]. The trend lines were computed as locally smoothed regression lines (LOESS) using a span of 0.75.

For DDK sequences, 16 unique acoustic measures were identified to contribute to the prediction of a speaker's age. The sex-specific differences in these measures between younger and older speakers are presented in Figure 4, along with the confidence region of the trend line. The DDK measures that were identified to contribute to the accurate prediction of sex-specific age of the speaker were DDK rate, variability in DDK rate (Rate (sd)), the average absolute difference between consecutive differences between consecutive syllable durations (DDP), the variability in syllable durations 5–12 compared to the average syllable duration of syllables 1–4 (relStab5–12), the percent of the syllable duration made up of the nucleus (%N), the average and standard deviation of the relative amplitude of the syllable onsets and nucleus (O/N Ampl., O/N Ampl. (sd)), the amplitude of syllables 13–20 compared to syllables 5–12 (AP), the average and standard deviation of the Release

Transient Prominence of syllable onsets (RTP, RTP (sd)), and variability in the degree of voicing spread from the following vowel (%Phon_final (sd)). Further, the average, variability, and trend in devoicing the vowel, both overall (%NPhon, %NPhon (sd), Progr. %NPhon) and in the final portions (%NPhon_final, %NPhon_final (sd)), were observed to contribute to a sex-specific model of age.
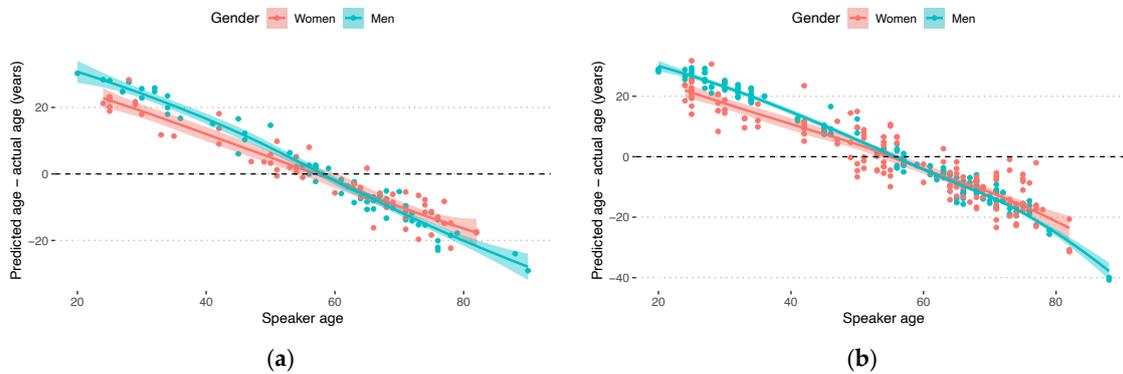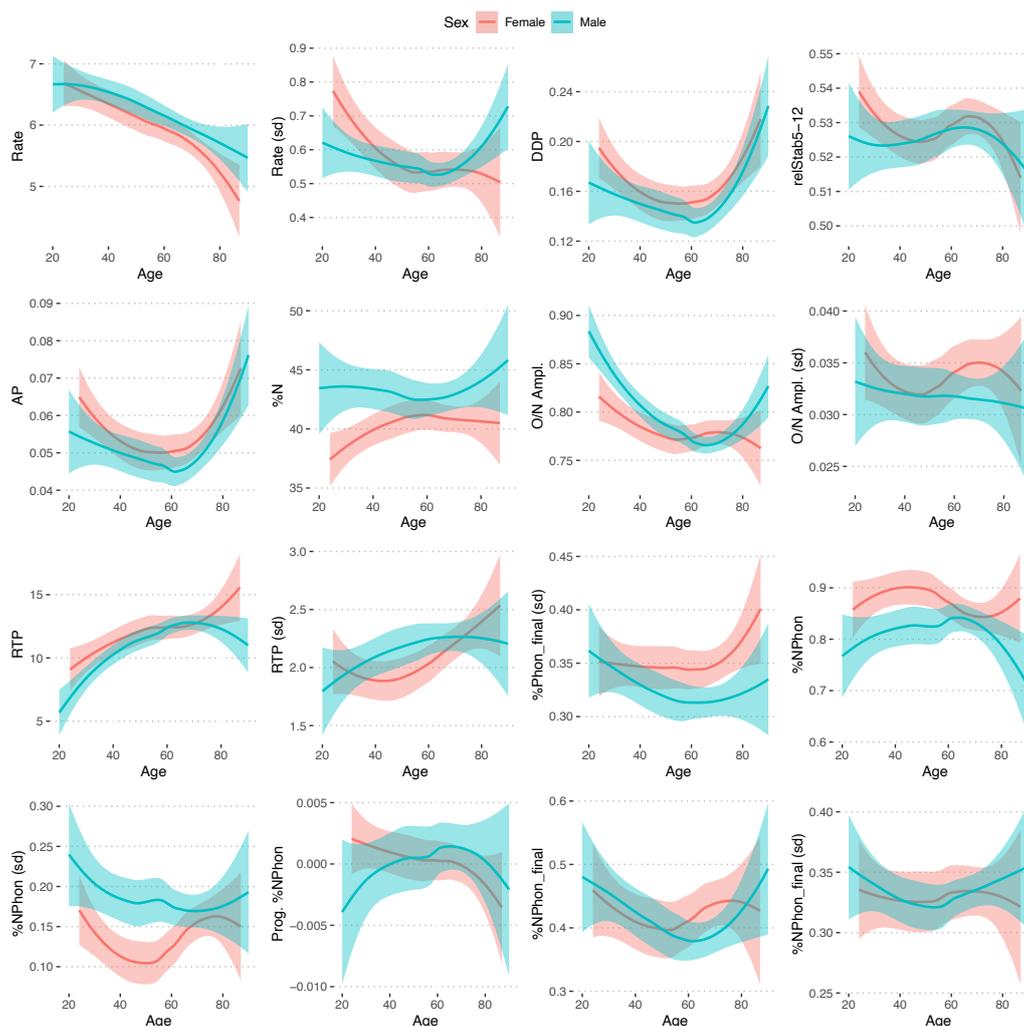

(a)　　　　　　　　　　　　　　　　　(b)

Figure 3. Errors in predicting a speakers' age based on (**a**) the cross-validated model using acoustic measures of a sustained [a] as predictors, and (**b**) the cross-validated model in which DDK measures were used. The known age of the speaker is shown on the horizontal axis and the vertical axis shows the prediction error.

Figure 4. The sex-specific pattern of change with age for identified acoustic measures of DDK sequences. The trend lines were computed as locally smoothed regression lines (LOESS) using a span of 0.75.

The models predicted 4% of the variance for men and 33% for women when applied to predict the age of speakers in the validation set. The average error of predicted age was -0.3 ± 17.6 years for men and -2.0 ± 16.0 years for women in the testing set. If evaluated within the same data on which it was trained (training set), the models explained 14% (men) and 39% (women) of the variance, respectively. If applied to all speakers, the models predicted age with an average error of -0.05 ± 15.8 and -0.39 ± 14.3 for men and women, respectively. The accuracy of age prediction across the range of speaker ages based on DDK measures is presented in Figure 3b. All pairwise correlations between acoustic DDK measures and age of the speaker are provided for both men and women in correlation matrices in Supplementary Material B.

4. Discussion

To detect disease-related changes in speech and voice, it is increasingly important to be able to discriminate them from changes due to aging. Establishing the acoustic consequences of aging is a complicated endeavor as motor changes may be assumed to have more than one acoustic consequence, with varying degrees of consistency between speakers. We applied the most comprehensive set of acoustic measures available to us to the sustained vowel phonation and oral–motor diadochokinesis tasks. We investigated which subset of predictors provided the best sex-specific and cross-validated prediction of the age of the speaker. In this way, we argue that we were able to find and describe the primary acoustic measures associated with sex-specific differences between younger and older speakers.

We used a collected set of measures that originate both from an understanding of age-related voice and speech changes as perceived by human listeners, as well as from an understanding that all changes due to age may not be completely observable using human auditory perception. It is important to acknowledge that limiting the quantification of motor actions to consider only features that may have a communicative role introduces bias into the analysis. Further, the preselection of measures limits the utility of the outcomes as a foundation for future efforts directed towards finding acoustic markers of onset of neurological diseases. Non-linear analysis techniques are used more often in other fields to quantify motor actions (Kalwa and Patil 2015; Turner and Joseph 2015). The Voice Analysis Toolbox (Tsanas et al. 2010a; Tsanas et al. 2012; Tsanas and Gómez-Vilda 2013; Gómez-Vilda et al. 2015) include quantifications for which there are no established perceptual model, and the DDK measures of Karlsson et al. (Karlsson and Hartelius 2019; Karlsson et al. 2020) include temporal quantifications of a scale that is likely smaller than what is perceivable by humans. It may, therefore, not be assumed that all measures that are established to differ between older and younger speakers have an established interpretation in terms of what the quantity correlates with in articulation, or indeed in perception. Therefore, we delimit our discussion here to those identified features for which an understanding of their cause has already been established in previous reports or for which a tentative interpretation may be suggested based on our current knowledge of speech production.

When speakers are asked to produce a sustained [a] for as long as possible, they are given the most optimal context that we can provide them with for reaching the articulatory target for the tongue. Instead, the primarily aim of the task is to stress the phonatory and respiratory systems, and the ability to keep the tongue stable. Similarly, when we ask a person to perform a DDK syllable sequence with the same syllables repeated as quickly as possible, we reduce the high-level processing and planning required and focus on the efficiency and stability in motor execution. Increasing age has been associated with decreasing phonatory stability and it is, therefore, not surprising to see that cycle-to-cycle instability in phonation (Jitter) and the instability in glottal cycle synchronous and general harmonic-to-noise measures (GNE->std and HNR_std) increase in older compared to younger speakers. Speakers are able to keep a progressively lower maximum pace when performing the DDK syllables but show a more complex pattern with age in their ability to keep the syllable rate stable across the sequence. The overall variability in syllable

rate (Rate (sd)) decreases with age from a relatively high level in younger women until it plateaus around the age of 40 at the same variability as men. Standard deviation may encode a wide range of variability, including quite different phenomena such as changes from one cycle to the next and steady deceleration, but the change in differences between consecutive syllables (DDP) indicates that at least some of the changes may be due to changes from cycle to cycle. Simultaneously, our young women show a larger deceleration in the initial portion of the syllable sequence compared to both men and older women, which is evidenced by the relstab 5–12 measure. It should, however, be kept in mind that our sample of speakers is much smaller in the 20–40 years group, so these observations can, to a higher degree, be due to individual differences and should be interpreted with some caution. In speakers older than 50–60 years, variability increases again in cycle-to-cycle durations across both tasks and for both men and women. Here, the sample of speakers is considerably larger, and the observation of an increased cycle-to-cycle variability in older speakers may, therefore, be considered more robust.

More surprising than the observed age-associated instability is the observation that the second formant frequency and bandwidth are higher in younger compared to older women. Simultaneously, a reduction in uncorrected voice source spectral tilt (H2KH5u) is observed, but these features are changed in older speakers of both sexes. The F_2 is usually attributable to a more fronted vowel production when assessed within the same speaker and could also be the effect of a smaller resonance tract overall. The current investigation is a cross-speaker comparison, but it was assumed that the characteristics of the second formant would not show the strong change in older speakers that were indeed observed. The cross-sectional design and the smaller sample of speakers should be kept in mind and changes observed may well be caused by sociophonetic factors. The spectral tilt of the source spectrum is, however, reduced in older speakers compared to younger speakers, and while the measure identified in the analysis was not corrected for by the impact of neighboring formants, the F_2 and associated bandwidth and the overall F_3 observed for the women (Average F3 = 2920 ± 203 Hz and average B3 = 110 ± 258 Hz) do not appear to be sufficiently close to influence the amplitude estimate of the harmonic closest to 2000 Hz and the H2KH5Ku measure in a systematic way. We, therefore, conclude that the spectral tilt of sustained vowels is reduced in older speakers compared to younger for both sexes, which could possibly be identified perceptually as weaker resonance.

While phonatory changes may be discussed reasonably well from the current dataset, some aspects of the DDK tasks would require longitudinal data and information from other tasks to tease apart the different strategies that a given speaker may choose to meet the demands of the task. There is a trade-off between rate and precision. The speaker may choose to focus on maintaining a high articulatory precision when encouraged to increase articulation rate and sacrifice some precision to maintain a higher rate than what would otherwise be possible. The obstruent in the simple syllables produced in the DDK task is the most articulatorily dynamic element and has been observed to be more strongly affected when the speaker is challenged (Karlsson and Hartelius 2019; Karlsson et al. 2014, 2020). Younger women show a strong presence of a stop closure (reduced %N) but align with men's values later on. In older men (>70 years), the vowel takes a larger proportion of the syllable duration, which is reminiscent of what is observed in hypokinetic dysarthria for which the measures were first developed (Karlsson and Hartelius 2019; Karlsson et al. 2020). Both men and women show a simultaneous decrease in consonant to vowel amplitude ratio (O/N Ampl.) and an increase in the prominence in release transient against the acoustic background in the consonant. The release transient is not strong in consonants produced at a high DDK rate, so what is observed here is a reduction in post-release burst energy. Other effects could, however, also be possible and likely require cycle-to-cycle analysis in a longitudinal study to analyze further. The release transient is produced more variably across a sequence in older speakers compared to younger across both sexes. These signs of weakened high-pressure consonants are also typical of dysarthria and occur in different

degrees depending on severity, from slightly indistinct consonant articulation to imprecise or slurred.

In addition, the DDK sequences show some tentative signs of increasing devoicing of vowels in older compared to younger speakers. The signs are not strong and more present in speakers aged above 60 years. For men, the overall percentage of the nucleus being voiced (%NPhon) decreases around that age; a continually strong presence of voicing percentage variability in phonation (%NPhon (sd)) and a pattern of progressive devoicing of nuclei across a sequence (Prog. %NPhon) are further observed. Further, there is an increased presence of voicing in the final part of the consonant (%Phon_final (sd)) present for both men and women above the age of approximately 60 years. The data, therefore, suggest that men and women aged above 60 years old may show signs of a decreasing phonatory control similar to, but less prominently, than what has been suggested to occur in speakers with dysarthria (Goberman and Blomgren 2008; Karlsson et al. 2012; Tanaka et al. 2015). Bearing in mind the relatively small number of speakers included here in the age range of interest, and the cross-sectional rather than longitudinal design, more research is needed to further substantiate this observation. The present data only afford the conclusion that the ability to regulate phonation may not be assumed to be intact in speakers above 60 years of age.

The set of acoustic measures used to analyze sustained [a] productions was extensive and may be regarded as a superset of measures compiled from many fields of research investigating how neurological diseases may affect phonation. Therefore, not all measures have a clear correspondence to a known aspect of phonatory function. We may, therefore, simply observe that the average 1st and 5th MFCCs (corresponding to the 76–160 Hz and 473–601 Hz frequency band amplitudes) decreased in older speakers of both sexes above the age of 40. The average 9th MFCC (1074–1267 Hz frequency band) increased in older women throughout the investigated age range, and the variability of the 10th and 12th MFCCs (corresponding to the 1267–1481 Hz and 1719–1982 Hz frequency bands) over the duration of the vowel further increased with age for all speakers. A progressive increase in the app_TKEO_std_1_coef measure of the Voice Analysis Toolbox was found in older speakers, for which there currently is no clear interoperation. The TKEO is, in this measure, applied to the first order wavelet decomposition of the f_0 curve and may, therefore, tentatively be suggested to capture aspects of the non-linear low-level fluctuations of f_0 frequency over the sustained vowel, but this interpretation requires further validation in more controlled data.

The research described here employed a methodology of using cross-validation of a statistical model of speaker age as a selection process for determining the measures of primary interest and was, therefore, able to include and competitively evaluate a substantial number of measures in a single study. We further opted to focus on the acoustic predictors which provided the most efficient description considering the number of described acoustic quantities, rather than the optimal model providing the best possible description of speaker age in the training set. This methodological decision, while justified, does come with some implications that demand highlighting. First, while we report on and discuss the measures which may most effectively be employed to describe the differences between older compared to younger speakers of each sex, other acoustic measures may also show a high correlation with age and could also have been employed to provide a reasonably effective model. To reduce the impact of this selective reporting of outcomes, we provide both correlation matrices and age-stratified descriptive statistics for the complete set of measures in Supplementary Materials A and B. The difficulties in interpreting cross-sectional data in terms of an effect of aging should also be acknowledged, as what is analyzed here is also the effect of between-speaker differences in how maximum performance tasks are performed. The risk of over-interpreting the effect of aging should, in part, be reduced by the analysis of overall trends but should be validated in further research.

5. Conclusions

The results presented here indicate that older speakers perform the sustained [a] and oral–motor diadochokinetic tasks differently than younger speakers. In the sustained [a], an increased instability in f_0 and frequency regions and above the second formant was observed, together with a reduction in source spectrum amplitude in low-frequency regions. In the DDK task, the articulation rate is reduced in older speakers, but the results also indicate that age influences how syllables are produced within the DDK task in a manner which for older speakers are similar to acoustic signs of dysarthria. Increasing age does not appear to affect phonation or articulation uniformly, and men and women are affected differently in most quantifications investigated here.

References

Amerman, James D., and Martha M. Parnell. 1992. Speech timing strategies in elderly adults. *Journal of Phonetics* 20: 65–76. [CrossRef]

Bloch, Isac, and Alison Behrman. 2001. Quantitative Analysis of Videostroboscopic Images in Presbylarynges. *The Laryngoscope* 111: 2022–27. [CrossRef]

Boersma, Paul, and David Weenink. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5: 341–45.

Eichhorn, Julie T., Raymond D. Kent, Diane Austin, and Houri K. Vorperian. 2017. Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women. *Journal of Voice* 32: 644. [CrossRef]

Eklund, Elisabeth, Johanna Qvist, Lena Sandström, Fanny Viklund, Jan van Doorn, and Fredrik Karlsson. 2014. Perceived articulatory precision in patients with Parkinson's disease after deep brain stimulation of subthalamic nucleus and caudal zona incerta. *Clinical Linguistics and Phonetics* 29: 150–66. [CrossRef]

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33: 1–20. [CrossRef]

Goberman, Alexander M., and Michael Blomgren. 2008. Fundamental Frequency Change During Offset and Onset of Voicing in Individuals with Parkinson Disease. *Journal of Voice* 22: 178–91. [CrossRef]

Gómez-Vilda, Pedro, Agustín Álvarez-Marquina, Athanasios Tsana, Carlos Alfredo Lázaro-Carrascosa, Victoria Rodellar-Biargem, Víctor Nieto-Lluis, and Rafael Martínez-Olalla. 2015. Phonation Biomechanics in Quantifying Parkinson's Disease Symptom Severity. In *Recent Advances in Nonlinear Speech Processing*. Berlin: Springer, pp. 93–102.

Gorham-Rowan, Mary M., and Jacqueline Laures-Gore. 2006. Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders* 39: 171–84. [CrossRef]

Gugatschka, Markus, Karl Kiesler, Barbara Obermayer-Pietsch, Bernadette Schoekler, Christoph Schmid, Andrea Groselj-Strele, and Gerhard Friedrich. 2010. Sex Hormones and the Elderly Male Voice. *Journal of Voice* 24: 369–73. [CrossRef]

Hawks, John, and James D. Miller. 1995. A formant bandwidth estimation procedure for vowel synthesis. *The Journal of the Acoustical Society of America* 97: 1343–44. [CrossRef]

Hillenbrand, James M., Ronald A. Cleveland, and Robert L. Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37: 769–78. [CrossRef]

Iseli, Markus, Yen-Liang Shue, and Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America* 121: 2283–95. [CrossRef]

Kahane, Joel C. 1983. A survey of age-related changes in the connective tissues of the human adult larynx. *Vocal Fold Physiology*, 44–49.

Kalwa, Shravanti, and H. T. Patil. 2015. Neuromuscular Disease Classification by Wavelet Decomposition Technique. Paper present at the 2015 International Conference on Communications and Signal Processing (ICCSP), Melmaruvathur, India, April 2–4; pp. 602–6. [CrossRef]

Karlsson, Fredrik, and Lena Hartelius. 2019. How Well Does Diadochokinetic Task Performance Predict Articulatory Imprecision? Differentiating Individuals with Parkinson's Disease from Control Subjects. *Folia Phoniatrica et Logopaedica* 71: 251–60. [CrossRef] [PubMed]

Karlsson, Fredrik, Patrik Blomstedt, Katarina Olofsson, Jan Linder, Erik Nordh, and Jan van Doorn. 2012. Control of phonatory onset and offset in Parkinson patients following deep brain stimulation of the subthalamic nucleus and caudal zona incerta. *Parkinsonism and Related Disorders* 18: 824–27. [CrossRef] [PubMed]

Karlsson, Fredrik, Katarina Olofsson, Patrik Blomstedt, Jan Linder, Erik Nordh, and Jan van Doorn. 2014. Articulatory closure proficiency in patients with Parkinson's disease following deep brain stimulation of the subthalamic nucleus and caudal zona incerta. *Journal of Speech, Language and Hearing Research* 57: 1178–90. [CrossRef] [PubMed]

Karlsson, Fredrik, Ellika Schalling, Katja Laakso, Kerstin M. Johansson, and Lena Hartelius. 2020. Assessment of speech impairment in patients with Parkinson's disease from acoustic quantifications of oral diadochokinetic sequences. *The Journal of the Acoustical Society of America* 147: 839–51. [CrossRef]

Leeuw, Irma M. V., and Hans F. Mahieu. 2004. Vocal aging and the impact on daily life: A longitudinal study. *Journal of Voice* 18: 193–202. [CrossRef]

Linville, Sue Ellen. 1992. Glottal Gap Configurations in Two Age Groups of Women. *Journal of Speech, Language, and Hearing Research* 35: 1209–15. [CrossRef]

Linville, Sue Ellen. 2002. Source Characteristics of Aged Voice Assessed from Long-Term Average Spectra. *Journal of Voice* 16: 472–79. [CrossRef]

Lombard, Lori, and Nancy Pearl Solomon. 2019. Laryngeal Diadochokinesis Across the Adult Lifespan. *Journal of Voice* 34: 651–56. [CrossRef]

Maryn, Yuri, Paul Corthals, Marc De Bodt, Paul van Cauwenberge, and D. Deliyski. 2009. Perturbation measures of voice: A comparative study between Multi-Dimensional Voice Program and Praat. *Folia Phoniatrica et Logopaedica* 61: 217–26. [CrossRef]

Novotny, Michal, Jan Melechovsky, Kriss Rozenstoks, Tereza Tykalova, Petr Kryze, Martin Kanok, Jiri Klempir, and Jan Rusz. 2020. Comparison of Automated Acoustic Methods for Oral Diadochokinesis Assessment in Amyotrophic Lateral Sclerosis. *Journal of Speech, Language, and Hearing Research* 63: 3453–60. [CrossRef]

Parnell, Martha M., and James D. Amerman. 1987. Perception of oral diadochokinetic performances in elderly adults. *Journal of Communication Disorders* 20: 339–51. [CrossRef]

Pontes, Paulo, Alcione Brasolotto, and Mara Behlau. 2005. Glottic Characteristics and Voice Complaint in the Elderly. *Journal of Voice* 19: 84–94. [CrossRef] [PubMed]

Pontes, Paulo, Rosiane Yamasaki, and Mara Behlau. 2006. Morphological and Functional Aspects of the Senile Larynx. *Folia Phoniatrica et Logopaedica* 58: 151–58. [CrossRef] [PubMed]

Ptacek, Paul H., and Eric K. Sander. 1966. Age Recognition from Voice. *Journal of Speech, Language, and Hearing Research* 9: 273–77. [CrossRef]

Rojas, Sandra, Elaina Kefalianos, and Adam Vogel. 2020. How Does Our Voice Change as We Age? A Systematic Review and Meta-Analysis of Acoustic and Perceptual Voice Data from Healthy Adults Over 50 Years of Age. *Journal of Speech, Language, and Hearing Research* 63: 533–51. [CrossRef]

Rong, Panying. 2020. Automated Acoustic Analysis of Oral Diadochokinesis to Assess Bulbar Motor Involvement in Amyotrophic Lateral Sclerosis. *Journal of Speech, Language, and Hearing Research* 63: 59–73. [CrossRef]

Schmitz-Hübsch, Tanja, Oleksandr Eckert, Uwe Schlegel, Thomas Klockgether, and Sabine Skodda. 2011. Instability of syllable repetition in patients with spinocerebellar ataxia and Parkinson's disease. *Movement Disorders* 27: 316–19. [CrossRef]

Skodda, Sabine. 2011. Aspects of speech rate and regularity in Parkinson's disease. *Journal of the Neurological Sciences* 310: 231–36. [CrossRef]

Solomon, Nancy Pearl, Douglas S. Brungart, Jessica R. Wince, Jordan C. Abramowitz, Megan M. Eitel, Julie Cohen, Sara M. Lippa, Tracey A. Brickell, Louis M. French, and Real T. Lange. 2021. Syllabic Diadochokinesis in Adults with and Without Traumatic Brain Injury: Severity, Stability, and Speech Considerations. *American Journal of Speech-Language Pathology* 30: 1400–9. [CrossRef] [PubMed]

Tanaka, Yasuhiro, Takashi Tsuboi, Hirohisa Watanabe, Yasukazu Kajita, Yasushi Fujimoto, Reiko Ohdake, Noritaka Yoneyama, Machihito Masuda, Kazuhiro Hara, Joe Senda, and et al. 2015. Voice features of Parkinson's disease patients with subthalamic nucleus deep brain stimulation. *Journal of Neurology* 262: 1–9. [CrossRef] [PubMed]

Titze, Ingo R. 2011. Vocal Fold Mass Is Not A Useful Quantity for Describing F0 in Vocalization. *Journal of Speech, Language, and Hearing Research* 54: 520–22. [CrossRef]

Titze, Ingo R., Ronald J. Baken, Kenneth W. Bozeman, Svante Granqvist, Nathalie Henrich, Christian T. Herbst, David M. Howard, Eric J. Hunter, Dean Kaelin, Raymond D. Kent, and et al. 2015. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America* 137: 3005–7. [CrossRef] [PubMed]

Tomita, Hideichiro, Taashi Nakashima, Akiteru Maeda, Hirohito Umeno, and Kiminori Sato. 2006. Age related changes in the distribution of laryngeal glands in the human adult larynx. *Auris Nasus Larynx* 33: 289–94. [CrossRef] [PubMed]

Tsanas, Athanasios. 2012. Accurate Telemonitoring of Parkinson's Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning. Ph.D. Thesis, University of Oxford, Oxford, UK.

Tsanas, Athanasios, and Pedro Gómez-Vilda. 2013. Novel robust decision support tool assisting early diagnosis of pathological voices using acoustic analysis of sustained vowels. In *Multidisciplinary Conference of Users of Voice, Speech and Singing*. Berlin/Heidelberg: Springer, pp. 3–12.

Tsanas, Athanasios, Max A. Little, Patrick McSharry, and Lorraine Ramig. 2010a. Accurate Telemonitoring of Parkinso's Disease Progression by Noninvasive Speech Tests. *IEEE Transactions on Biomedical Engineering* 57: 884–93. [CrossRef]

Tsanas, Athanasios, Max A. Little, Patrick McSharry, and Lorraine Ramig. 2010b. New nonlinear markers and insights into speech signal degradation for effective tracking of Parkinson's disease symptom severity. *IEICE Proceedings Series* 44: 457–60. [CrossRef]

Tsanas, Athanasios, Max A. Little, Patrick McSharry, and Lorraine Ramig. 2011. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of the Royal Society, Interface/the Royal Society* 8: 842–55. [CrossRef]

Tsanas, Athanasios, Max A. Little, Patrick E. McSharry, Jennifer L. Spielman, and Lorraine Olson Ramig. 2012. Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering* 59: 1264–71. [CrossRef]

Tucker, Benjamin V., Cathrine Ford, and Stephanie Hedges. 2021. Speech aging: Production and perception. *Wiley Interdisciplinary Reviews: Cognitive Science* 12: e1557. [PubMed]

Turner, Claude, and Anthony Joseph. 2015. A Wavelet Packet and Mel-Frequency Cepstral Coefficients-Based Feature Extraction Method for Speaker Identification. *Procedia Computer Science* 61: 416–21. [CrossRef]

Wang, Yu-Tsai, Ray D. Kent, Joseph R. Duffy, and Jack E. Thomas. 2009. Analysis of Diadochokinesis in Ataxic Dysarthria Using the Motor Speech Profile Program^TM. *Folia Phoniatrica et Logopaedica* 61: 1–11. [CrossRef] [PubMed]