HAL open science

Analysing spectral changes over time to identify articulatory impairments in dysarthria

Anneke Slis, Nathalie Lévêque, Cécile Fougeron, Michaela Pernon, Frédéric Assal, Leonardo Lancia

To cite this version:

Anneke Slis, Nathalie Lévêque, Cécile Fougeron, Michaela Pernon, Frédéric Assal, et al.. Analysing spectral changes over time to identify articulatory impairments in dysarthria. Journal of the Acoustical Society of America, 2021, 149 (2), pp.758 - 769. 10.1121/10.0003332 . hal-03099517

HAL Id: hal-03099517

https://hal.science/hal-03099517v1

Submitted on 3 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

JASA ARTICLE

Analysing spectral changes over time to identify articulatory impairments in dysarthria

A. Slis,[1] N. Lévêque,[2,a)] C. Fougeron,[1,b)] M. Pernon,[3,a)] F. Assal,[3] and L. Lancia[1]
[1]*LPP, UMR 7018, CNRS/University Sorbonne-Nouvelle, Paris, France*
[2]*APHP, Department of Neurology, Pitié-Salpêtrière Hospital, ALS Reference Center, France*
[3]*Department of Clinical Neurosciences, Geneva University Hospital, Switzerland*

ABSTRACT:
Identifying characteristics of articulatory impairment in speech motor disorders is complicated due to the time-consuming nature of kinematic measures. The goal is to explore whether analysing the acoustic signal in terms of total squared changes of Mel-Frequency Cepstral Coefficients (TSC_MFCC) and its pattern over time provides sufficient spectral information to distinguish mild and moderate dysarthric French speakers with Amyotrophic Lateral Sclerosis (ALS) and Parkinson's Disease (PD) from each other and from healthy speakers. Participants produced the vowel-glide sequences /ajajaj/, /ujujuj/, and /wiwiwi/. From the time course of TSC_MFCCs, event-related and global measures were extracted to capture the degree of acoustic change and its variability. In addition, durational measures were obtained. For both mild and moderately impaired PD and ALS speakers, the degree of acoustic change and its variability, averaged over the complete contour, separated PD and ALS speakers from each other and from healthy speakers, especially when producing the sequences /ujujuj/ and /wiwiwi/. Durational measures separated the moderate ALS from healthy and moderate PD speakers. Using the approach on repetitive sequences targeting the lingual and labial articulators to characterize articulatory impairment in speech motor disorders is promising. Findings are discussed against prior findings of articulatory impairment in the populations studied.
© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1121/10.0003332
(Received 12 September 2020; revised 25 November 2020; accepted 17 December 2020; published online 2 February 2021)
[Editor: Paavo Alku] Pages: 758–769

I. INTRODUCTION

To better understand processes underlying articulatory behavior, scientists rely on a variety of techniques to extract vocal tract information from the speaker. Articulatory behavior is measured kinematically with, for example, ultrasound or electromagnetic articulography, or supra-laryngeal characteristics extrapolated from the acoustic output. Kinematic data are more difficult to acquire compared to acoustic data. Articulatory studies on a large set of speakers demand time and expertise for the acquisition and the analyses of the data, and specialized equipment, which is frequently not on location in, for example, hospitals. Consequently, acoustic data as a source of information about vocal tract changes during speech are a desirable and fast alternative to study this population. It has been recognized, however, that frequently no reliable one-to-one relationship exists between the actual movements in the vocal tract and the acoustic outcomes and that articulation affects the acoustic output in a nonlinear manner, making it difficult to relate acoustic information to articulatory events (see, e.g., Whitfield and Goberman, 2014; Yunusova *et al.*, 2012). The current study uses Mel Frequency Cepstral Coefficients (MFCC) to capture vocal tract shape information (Davis and Mermelstein, 1980; Khan *et al.*, 2014; Rudzicz *et al.*, 2016). These coefficients represent the information contained in the acoustic spectrum. They are computed by applying a discrete cosine transform to the coefficients of a power spectrum whose frequency bands are equally spaced on the Mel scale (see Sec. II C 1 for a more detailed explanation). The changes over time of MFCCs are theorized to acoustically represent articulatory movement behavior (see, e.g., Goldstein, 2019). The objective of the current study, therefore, is to extract indirect information about vocal tract fluctuations from the acoustic output in terms of the total squared change of Mel Frequency Cepstral Coefficients (TSC_MFCCs) and their change over time to capture global articulatory impairment, and to validate whether the proposed method provides sufficient information to distinguish dysarthric speakers at different levels of articulatory impairment from each other and from healthy speakers.

Parkinson's Disease (PD) and Amyotrophic Lateral Sclerosis (ALS) are both neurogenerative diseases that eventually result in dysarthria, which is a motor speech disorder characterized by impairments at the respiratory, laryngeal, and articulatory level (Darley *et al.*, 1975; Duffy, 2014). Speech of individuals with PD is identified,

[a)]Also at: LPP, UMR 7018, CNRS/U. Sorbonne-Nouvelle, Paris, France.
[b)]Electronic mail: cecile.fougeron@sorbonne-nouvelle.fr, ORCID: 0000-0002-1192-5848.

 ©Author(s) 2021.

among other things, by reduced loudness, hoarse and breathy voice, momentary rushes of speech, variable speech rate, less variability of pitch, imprecise articulation, and dysfluent speech production, resulting in hypokinetic dysarthria (Darley *et al.*, 1969; Skodda *et al.*, 2013; Tjaden, 2008). Speech of ALS patients is marked by a strained, breathy, and hoarse voice, excessive prolonged speech segments, and imprecise consonants and hypernasality, resulting in mixed flaccid-spastic dysarthria (Darley *et al.*, 1969; Duffy, 2014; Hanson *et al.*, 2011; Mefferd and Dietrich, 2020; Rong *et al.*, 2019; Tomik and Guiloff, 2010; Yunusova *et al.*, 2019). Although both PD and ALS individuals develop motor speech disorders, the underlying etiology is different. PD is a slowly progressing neurodegenerative disease caused by deficient cortico-basal ganglia-thalamic neural circuits (Duffy, 2014). ALS is a quickly progressing neurogenerative disorder involving deteriorating upper and lower motor neurons in the cerebral cortex, brainstem, and spinal cord (Kiernan *et al.*, 2011). The strained and hoarse voice quality and slow speaking rate in ALS speakers are associated with an impairment in the function of the upper motor neurons (UMN) and are referred to as spastic dysarthria. The breathy and weak voice quality, hypernasality, nasal emissions and articulatory imprecision indicate an impairment in the function of the lower motor neurons (LMN) and are referred to as flaccid dysarthria (Tomik and Guiloff, 2010). Despite their different etiology and the fact that their speech productions at the level of breathing, voice and articulation are perceived as different, many of the acoustic and kinematic characteristics of dysarthric speech of PD and ALS individuals are similar. Kinematically, smaller movement amplitudes and slower peak velocity of articulatory movements are typical in PD and ALS speakers (Ackermann and Ziegler, 1991; Ackermann *et al.*, 1997; Caligiuri, 1989; Connor *et al.*, 1989; Delvaux *et al.*, 2016; Forrest and Weismer, 1995; Forrest *et al.*, 1989; Lansford and Liss, 2014; Lee *et al.*, 2017; Mefferd, 2015; Shellikeri *et al.*, 2016; Walsh and Smith, 2012; Yunusova *et al.*, 2008; Yunusova *et al.*, 2012). Acoustically, reduced $F2$ slopes and smaller vowel space have been measured in both PD and ALS speakers (Forrest *et al.*, 1989; Kim *et al.*, 2009; Lansford and Liss, 2014; Lee *et al.*, 2017, Mefferd, 2015; Skodda *et al.*, 2012; Walsh and Smith, 2012; Whitfield and Goberman, 2014; Yunusova *et al.*, 2012). Consequently, even though speech productions of individuals with PD and ALS are perceptually very distinct, the impairments at the articulatory level of different types of dysarthria have been very difficult to classify objectively based on the "traditional" acoustic markers such as acoustic vowel space, $F2$ slope, or $F0$ range (Kim *et al.*, 2011). One of the factors that potentially affects the outcomes of acoustic studies is the lack of sensitivity of the individual measures employed. For instance, isolating and relying solely on formant frequency contours is not only time consuming and susceptible to error (Shadle *et al.*, 2016; Whitfield and Goberman, 2014), but also ignores other potentially relevant sources of acoustic information reflecting changes in vocal tract shape. This study explores the feasibility of considering the complete shape of the spectral envelope and its patterns of change over time by extracting MFCCs (Davis and Mermelstein, 1980; Fletcher *et al.*, 2017; Goldstein, 2019; Rudzicz *et al.*, 2016), thus avoiding *a priori* assumptions about the important features of a spectrum. Although other techniques using the complete shape of the spectrum possibly describe articulatory behavior successfully as well, the strength of MFCCs lies in the fact that they provide a non-redundant representation of spectral information that captures the relevant perceptual, articulatory, and phonological characteristics of the speech signal (Davis and Mermelstein, 1980; Khan *et al.*, 2014; Rudzicz *et al.*, 2016). This success is related, among others, to the fact that MFCCs separate the source from the filter, the latter containing acoustic information about the vocal tract resonator (Rudzicz *et al.*, 2016). MFCCs are commonly used in the field of speech technology (Huang *et al.*, 2001; Khan *et al.*, 2014; Lokesh and Devi, 2019; Rudzicz *et al.*, 2012; Rudzicz *et al.*, 2016) and have been promising in identifying and classifying speech disorders (Fletcher *et al.*, 2017; Khan *et al.*, 2014; Rudzicz *et al.*, 2016).

Recently, it has been shown that the time course of TSC_MFCC correlates consistently across speakers with the degree of vocal tract change, as represented by the changes over time of articulator marker positions in X-ray microbeam images (Goldstein, 2019). To the authors' knowledge, the TSC_MFCC measure has not been applied for clinical or research purposes. Based on evidence that TSC_MFCCs closely represents the degree of change of vocal tract shape, the working hypothesis for the current study is that the resulting patterns of successive TSC_MFCCs are associated with kinematic events in the spatial domain.[1] Consequently, it is expected that several derivatives of this acoustic measure successfully describe articulatory impairment in ALS and PD speech.

II. METHODS

A. Participants

A total of 129 French speakers participated in the study in the context of a larger project.[2] Twenty-six French speakers with PD were recruited at the Hôpitaux Universitaires in Geneva, Switzerland, 33 French speakers with ALS were recruited at the Pitié-Salpêtrière University Hospital in Paris, and 70 healthy French speakers (40 from Paris, 30 from Geneva) were recruited as control participants. The healthy speakers were matched in age and gender on a group level.

All ALS patients were diagnosed by neurologists of the "ALS and other rare MNDs center" of the Pitié-Salpêtrière University Hospital (Paris, France). ALS patients were diagnosed based on the El Escorial criteria (Brooks *et al.*, 2000). Functional signs of bulbar involvement according to the

bulbar subscore of the ALSFRS-R scale had to be present (Cedarbaum *et al.*, 1999). Subjects were not included if they showed evidence of cognitive impairment and/or had respiratory issues according to a routine neurologist examination. Patients with severe dysarthria were also excluded (Perceptual Score, PS > 14, see below). All PD patients were taking Levodopa and two (one mild and one moderate) on a combination of Levodopa and Carbidopa. During the recordings, all the PD speakers were on the Med-ON stage.

The PD and ALS speakers were diagnosed with dysarthria by certified speech-language pathologists, responsible for the recruitment, and only participants identified with mild to moderate dysarthria were included in the cohorts. Two expert speech-language pathologists judged the severity of dysarthria on five dimensions on a 5-point scale: voice quality, phonetic realisation, prosody, intelligibility, and naturalness of speech (Auzou and Rolland-Monnoury, 2019). This evaluation was done separately by the two judges on 3–5-min audio recordings of a spontaneous speech sample of each patient, presented in a random order. The five sub-scores were then summed into a PS ranging from 0 to 20 (0 = normal to 20 = severe).

The interrater measurements were performed at the level of the overall/20 score. Intraclass Correlation (ICC) estimates and their 95% confidence intervals were calculated based on a single-rating, absolute-agreement, two-way random-effects model (Koo and Li, 2016). The overall reliability score between the two raters was high (ICC = 0.93, $p < 0.001$, 95%-Confidence Interval: $0.89 < \text{ICC} < 0.96$). For ALS speakers the reliability was excellent (ICC = 0.94, $p < 0.001$, 95%-Confidence Interval: $0.87 < \text{ICC} < 0.97$) and for the PD population the reliability was good (ICC = 0.86, $p < 0.01$, 95%-Confidence Interval: $0.71 < \text{ICC} < 0.93$). Because of the high reliability scores, the two PS scores were averaged. Following Auzou and Rolland-Monnoury (2019), the PS score was used to define a group with mild dysarthria (PS 1–6) and a group with moderate dysarthria (PS 6–14). The distribution of the population per subgroups is described in Table I.

B. Task and stimuli

The participants produced onomatopoeic expressions made of glides and vowels, which are quite naturally produced in French: "aïe aïe aïe" /ajajaj/ and "ouille ouille ouille"/ujujuj/ (both meaning "ouch ouch ouch") and "oui oui oui" /wiwiwi/ ("yes yes yes"). These sequences are characterized by continuous articulatory movements and are rich in transitional acoustic information. These three sequences also targeted different alternative vocal tract shapes: successive opening-closing with the tongue body and the jaw moving from the open /a/ to the palatal /j/ in /ajajaj/, successive back-front movements of the tongue and rounding-unrounding of the lips from the back rounded /u/ to the unrounded palatal /j/ for /ujujuj/ and the labio-velar /w/ to the unrounded front /i/. These last two sequences also differed in syllabic structure—a nucleus-coda structure for /ujujuj/ and onset-nucleus structure for /wiwiwi/. Because of this set of sequences targeting different articulatory and acoustic patterns, the material was also tested for its ability to capture different distorted articulations.

Speakers were instructed to produce the vowel-glide expressions presented in their orthographic forms (e.g., "oui oui oui") as naturally as possible, at a comfortable speaking rate, and in a continuous manner (i.e., with no pause between the three "oui"). Examples were given if this last instruction was not understood. The three sequences were presented four times in a random order resulting in 12 trials per speaker. All the productions were recorded with a microphone (SM35-XLR, Shure) located at a distance of 2–3 cm (sampling rate of 44.1 kHz), *via* an external audio card (scarlett-2i4, Focusrite) in a quiet room and stored on a laptop.

C. Acoustic analysis

1. Extracting the TSC_MFCC as a measure of acoustic change

The onset and offset of the complete sequences /ajajaj/, /ujujuj/, and /wiwiwi/ were manually segmented in PRAAT (Boersma and Weenink, 2020) and if speakers produced more than one sequence during one trial, the first production was selected. Next, MATLAB scripts, developed at the Laboratoire de Phonétique et Phonologie (LPP), analyzed the selected sequences as follows. The annotated sound files were first resampled to 16 kHz and the resulting signal was

TABLE I. Distribution of the population (ALS, PD, and HEALTHY), severity (Mild, Moderate), and sex [Female (F) and Male (M)], with mean age (Age), and perceptual score (PS). Standard deviations of age are indicated in parentheses and minimum and maximum age are indicated between brackets. N = number of speakers in each group.

	ALS				PD				HEALTHY	
	Mild		Moderate		Mild		Moderate			
	F	M	F	M	F	M	F	M	F	M
N	6	4	7	16	6	13	2	5	41	29
Age	67.33 (4.22) [61-73]	62.75 (12.71) [50-79]	64 (5.07) [56-72]	62.38 (13.2) [36-86]	70.69 (10.07) [55-80]	72.63 (9.14) [5088]	84 (9.19) [75-93]	77.8 (5.00) [70-83]	62.94 (19.1) [23-90]	58.77 (20.75) [25-90]
PS	4.21 (1.1)	4.38 (1.13)	9.79 (1.87)	9.59 (2.22)	2.83 (1.22)	3.23 (1.39)	7.75 (1.41)	8.05 (0.97)		

https://doi.org/10.1121/10.0003332

pre-emphasised. To obtain the TSC_MFCCs (see also Goldstein, 2019, for a similar approach), the Fourier spectra were calculated from the individual segmented waveforms with a 25 ms window sliding over time in 2 ms steps. To target spectral information maximally relevant to the chosen linguistic material, only spectral coefficients between 300 and 4500 Hz were retained for further analyses. The trimmed spectra were processed with a filter bank composed of 20 triangular filters spaced evenly on the Mel-frequency scale, which is based on the filter characteristics of the human ear (O'Shaughnessy, 1987). The individual steps are illustrated in Fig. 1.

To obtain the final MFCCs, the resulting spectral energy values were log-transformed and submitted to a discrete cosine transform (Davis and Mermelstein, 1980). Changes in the first coefficient reflect changes in average spectral energy and this coefficient was discarded, leaving only the coefficients 2–13. Each MFCC was low pass filtered with a Butterworth filter of 12 Hz. Finally, to obtain the TSC_MFCCs, squared differences of cepstral energy values in consecutive frames were computed and summed at each time step according to the following equation:

$$\mathrm{MFCC}_k = \sum_{i=2}^{13} (f(i,k+1) - f(i,k))^2. \quad (1)$$

In which $f(i,k)$ is the ith MFCC at frame k.

Plotting the TSC_MFCC values over time results in a TSC_MFCC contour representing the degree of spectral energy change from one 25 ms frame to the next, in 2 ms steps (Fig. 2). A large degree of energy change between two windows, i.e., a high TSC_MFCC value, shows up as a peak in the contour, which is expected to coincide with transitions from vowel-to-glide and glide-to-vowel. Because the working hypothesis for the current study was that the resulting patterns of successive TSC_MFCCs were associated with vocal tract changes, the magnitudes of maxima and minima in this TSC_MFCC contour were theorized to correlate with the speed at which the vocal tract changes and consequently are treated as velocity values in the current study (see for a similar approach using the amplitude envelop: He and Dellwo, 2017). To capture the relevant maximum and minimum values from this velocity contour and ignore other irrelevant events due to small instabilities during the vowel, the resulting contours were low pass filtered with a cut off frequency of 14 Hz.[3]

The lowest velocity values (minima) coincide with steady states of the vowel and glide and the highest values (maxima) coincide with the points of maximum spectral change during the transition from vowel to glide and *vice versa* (see Fig. 2).

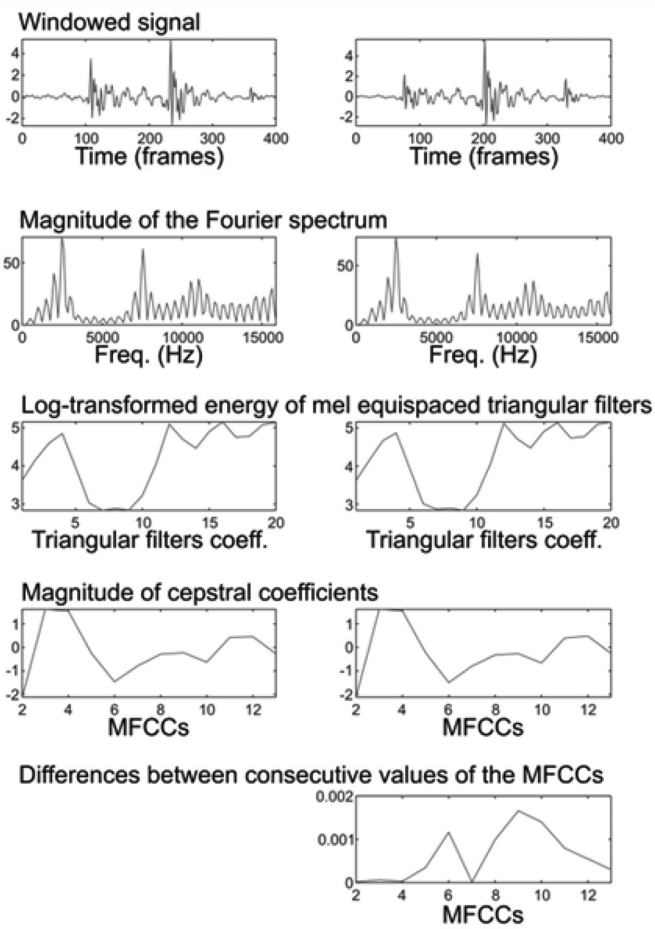

FIG. 1. Individual steps in the calculating the final envelop of total squared spectral energy changes. The topmost row displays two consecutive portions of 25 ms (400 samples) windowed signal 2 ms apart. The second row from the top displays the magnitude of the Fourier transform coefficients, while the third line from the top displays the log transformed energy levels captured by the MEL scale filter bank. The fourth row from the top displays the magnitude of the cepstral coefficients, while the bottom row displays the 12 squared differences between the corresponding values obtained from the analysis of the two portions of signal. The total degree of spectral change is obtained by summing these 12 squared differences [Eq. (1)].

2. Measures extracted from the TSC_MFCC contour

Two sets of measures were extracted from the velocity contours. The first set of "event-related measures"(i) was established by identifying the individual local minima and maxima in the velocity contour, while the second set of "global measures" (ii) was extracted from the complete velocity contour, without any prior sub-segmentation.

(i) Maximum and minimum velocity values in the contour, defined as 'events' in the contour, were automatically determined by a peak-picking algorithm in MATLAB. Values were manually removed when two maxima occurred close together in time (see Fig. 2). In this case, the highest value was selected. The first maximum in the contour marks an energy change from silence to onset of the first segment and this value was disregarded as well (see Fig. 2). In addition to the maximum velocity values, local minima were automatically identified on either side of a maximum and were theorized to coincide with the steady state of a segment.

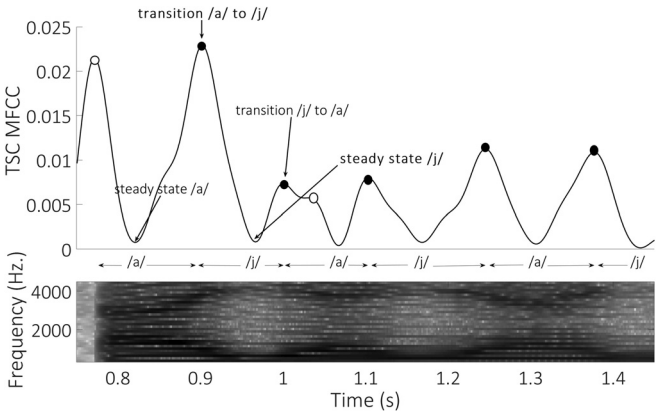

FIG. 2. Example of /ajajaj/ produced by a healthy speaker. Upper part: The vertical axis represents the degree of spectral energy change from one 25 ms frame to the next in 2 ms steps. The lower part displays the corresponding spectrogram in the range from 300 to 4500 Hz; horizontal axis shows time (s). Clear maxima in the contour occur during the formant transitions between the segment's steady states, visible on the spectrogram. The first maximum coincides with the onset of the first vowel and is disregarded in the analysis. The filled dots indicate selected maxima and open dots the disregarded maxima. The corresponding vowel and glide are shown between the two parts with arrows indicating the start and end.

Two measures were extracted from the maximum and minimum values, resembling articulatory measures in the spatial dimension of movement tracking data, and measures of the speed of acoustic change in He and Dellwo (2017):

- *Positive dynamics (POSDYN)* captured the steepness of degree of acoustic changes from acoustic steady-states, i.e., the minimum events, to subsequent points of maximal acoustic changes, i.e., the maximum events (similar to a positive velocity slope). The measure corresponded to the difference between a minimum and the following maximum velocity value divided by the duration from the interval between these minimum and maximum. The steepness was calculated instead of the actual maximum value to correct for the time it takes to reach this maximum. Some speakers reached a maximum over a longer stretch of time, which was captured in the steepness measure. In Fig. 3, this has been indicated with "4a"/"1a."
- *Negative dynamics (NEGDYN)* captured the steepness of the degree of acoustic energy changes from the point of maximal acoustic change (maxima) to the following acoustic steady-state (minima), similar to a negative velocity slope. It was computed as the difference between maximum and following minimum velocity value divided by the time from this maximum to minimum event. In Fig. 3, this has been indicated with "4b"/"1b."

(ii) Global measurements were based on the complete velocity contour from the first minimum (segment 1 steady-state) to the last minimum (segment 6 steady-state). In this approach, the average degree of acoustic change in the sequence, taken as the mean of the complete contour values (*MEANCH*) and their coefficient of variation (standard deviation calculated from the complete contour values divided by *MEANCH*: COVAR), captured the modulation in acoustic change and its variability within the sequence.

Two additional measures related to duration of the event- and sequence production were extracted:

- *eventDUR*: from the "event-related measures," the duration from each minimum to the next minimum was calculated and corresponded to the time between one segment steady state to the next steady state to capture segmental transition duration. This interval corresponded to "2" in Fig. 3.
- *globalDUR:* the total duration from first to the last minimum was computed to capture the duration of the whole sequence by taking the sum of the five eventDUR values in a sequence.

D. Statistical analysis

First, values below the first percentile and over the 99th percentile were removed from the event-related measures.

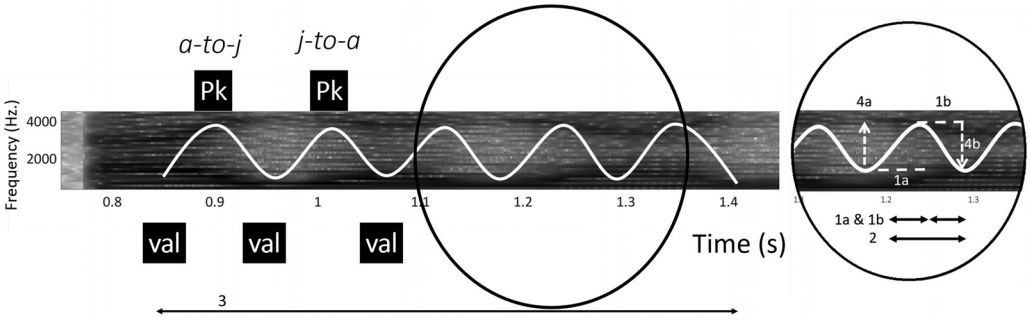

FIG. 3. Spectrogram of /ajajaj/ with a manually added "expected" velocity contour. The 5 peaks (Pk) coincide with the 5 transitions from segment to segment: a1 to j1, j1 to a2, a2 to j2, j2 to a3, and a3 to j3. The isolated circle explains the extracted measures: "1a" and "1b" are the durations from minimum to maximum velocity and maximum to minimum velocity respectively and are used to calculate: 4a/1a = positive dynamics (POSDYN) and 4b/1b = negative dynamics (NEGDYN). "2" marks the steady state to steady state duration (val-to-val: eventDUR). The long arrow "3" indicates the duration of the complete contour.

https://doi.org/10.1121/10.0003332

TABLE II. LMM results for the fixed effects "population" and "sequence" and their interaction for the three dependent variables "POSDYN," "NEGDYN," and "eventDUR," extracted from the event-related measures, and the three variables "MEANCH," "COVAR," and "globalDUR," extracted from the global measures.

	Event-related measures					
	POSDYN		NEGDYN		eventDUR	
	X^2	p	X^2	p	X^2	p
Population	91.06	<0.001	93.24	<0.001	62.90	<0.001
Sequence	1299.50	<0.001	1159.80	<0.001	27.88	<0.001
Population:Sequence	256.82	<0.001	214.08	<0.001	18.72	0.01
	Global measures					
	MEANCH		COVAR		globalDUR	
	X^2	p	X^2	p	X^2	p
Population	100.55	< 0.001	50.42	<0.001	69.98	<0.001
Sequence	866.04	<0.001	26.38	<0.001	20.74	<0.001
Population:Sequence	247.03	<0.001	69.91	<0.001	26.01	<0.01

Linear mixed models (LMM) were performed using R (R Core Team, 2020) to examine how the event-related and global measures derived from the velocity contour were predicted by the "population" (ALS_mild, ALS_moderate, PD_mild, PD_moderate, Healthy[4]), and "sequence" types (/ajajaj/, /ujujuj/, /wiwiwi/), with a random intercept per "speaker," according to the model: lmer [(AM∼population +sequence+population:sequence + (1|Speaker)].

P-values estimates for linear regressions were based on Satterthwaite approximations through the *lmerTest*-function (Kuznetsova *et al.*, 2017). The threshold at which an effect was considered significant was set to $p < 0.05$. Likelihood ratio tests, as implemented in the *anova*-function, were performed to check main effects of each fixed factor and interactions. *Post hoc* comparisons were computed using the *lsmeans*-function from the library "emmeans" (Lenth, 2016). The following six contrasts were selected: Healthy versus ALS_mild, Healthy versus ALS_moderate, Healthy versus Parkinson_mild, Healthy versus Parkinson_moderate, ALS_mild versus Parkinson_mild, and ALS_moderate versus Park_moderate. *P*-values were adjusted by adopting the False Discovery Rate criterion (Benjamini and Hochberg, 1995).[5]

III. RESULTS

For all the event-related and global measures, the fixed effects "population," "sequence," and the interaction between the two were significant. Table II summarizes the individual X^2 and p values of the fixed effects. As expected, the three sequences tested showed different TSC_MFCC patterns. /ajaja/ was characterized by smaller values of "MEANCH," "COVAR," "POSDYN," and "NEGDYN," meaning that the degree of spectral change was smaller and less steep than the other two sequences, while /wiwiwi/ differred form /ujujuj/ with higher "POSDYN" and "NEGDYN" values and shorter duration. More interestingly, there is a population and population by sequence type interaction on most of the measures. These results will be reported in detail below. Pairwise comparisons between subgroups are summarized in Tables III and IV and presented in more detail in Tables V and VI in the Appendix.

TABLE III. Results of the LMM model, testing the effects of population (Healthy, ALS_mild, ALS_moderate, PD_mild, and PD_moderate) and sequence [/ajajaj/ (AJ), /ujujuj/ (UJ), and /wiwiwi/ (WI)] on the variables "POSDYN," and "NEGDYN," for event-related analyses, and "MEANCH," and "COVAR," for the global analyses. The results are indicated with ns = not significant, (*) = trend, * = p < 0.05, ** = p < 0.01, *** = p < 0.001. < and > specify the directions of significance based on the two groups in the header. The light gray boxes indicate the variables that separated all the groups from the healthy and ALS mild from PD mild successfully; dark gray boxes mark the variables that separated mild and moderate ALS speakers from the other groups.

Spatial measures							
Event-related							
		Healthy ALS_mild	Healthy ALS_moderate	Healthy PD_mild	Healthy PD_moderate	PD_mild ALS_mild	PD_moderate ALS_moderate
POSDYN ($R^2 = 0.45$)	AJ	>*	> ***	ns	>*	>*	ns
	UJ	> ***	> ***	ns	> ***	> ***	ns
	WI	> ***	> ***	> **	> ***	> ***	ns
NEGDYN ($R^2 = 0.42$)	AJ	>*	> ***	ns	> *	> *	ns
	UJ	> ***	> ***	ns	> ***	> ***	ns
	WI	> ***	> ***	> *	> ***	> ***	ns
global							
MEANCH ($R^2 = 0.77$)	AJ	>(*)	> ***	ns	>(*)	ns	ns
	UJ	> ***	> ***	ns	> ***	> ***	ns
	WI	> ***	> ***	> **	> ***	> ***	ns
COVAR ($R^2 = 0.40$)	AJ	<*	<*	ns	ns	<*	ns
	UJ	<**	<***	ns	ns	<**	<**
	WI	ns	<***	ns	ns	<*	<**

TABLE IV. Results of the LMM model testing the effects of population on the variables "eventDUR" and "globalDUR" for each sequence /ajajaj/ (AJ), /ujujuj/ (UJ) and /wiwiwi/ (WI). The results are indicated with ns = not significant, (*) = trend, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. < and > specify the directions of significance based on the two groups in the header.

		Durational measures					
		Healthy ALS_mild	Healthy ALS_moderate	Healthy PD_mild	Healthy PD_moderate	PD_mild ALS_mild	PD_moderate ALS_moderate
EventDUR ($R^2 = 0.27$)	AJ	ns	<***	ns	ns	ns	<***
	UJ	ns	<***	ns	ns	ns	<*
	WI	ns	<***	ns	ns	ns	<**
GlobalDUR ($R^2 = 0.65$)	AJ	ns	<***	ns	ns	ns	<**
	UJ	ns	<***	ns	ns	ns	ns
	WI	ns	<***	ns	ns	ns	<**

A. Event-related measures

The two measures "POSDYN" and "NEGDYN" separated mild and moderately dysarthric ALS and PD speakers from each other and from the healthy speakers, especially in the productions of /wiwiwi/. Decreasing values were observed in the order: Healthy speakers > mild PD > mild ALS, and Healthy speakers > moderate PD and ALS. This means that the speech of healthy speakers showed a steeper degree of spectral energy change than the other groups, indicating that the vocal tract changes during the transitions were larger and/or reached the target position faster. Compared to mild ALS speakers, the mild PD speakers showed higher values in steepness of spectral energy change. Moderate ALS and PD showed the lowest values but were not different from each other. Although productions of /ujujuj/ and /ajajaj/ resulted in higher values for the healthy speakers than most disordered speakers, these sequences did not result in significantly different values for healthy speakers compared to mild PD, and moderate PD compared to moderate ALS.

B. Global measures

Inspecting the results in Table III and Fig. 4, similar patterns emerged for the global measures taking the complete contour into account as the event-related measures. The mild and moderate PD and ALS speakers differed consistently from healthy speakers in "MEANCH" values in the productions of /wiwiwi/ as follows: Healthy speakers > mild PD > mild ALS and Healthy speakers > moderate PD and ALS. The values for healthy speakers are in these instances significantly higher than the ones for the dysarthric population, meaning that there is more overall acoustic change in the productions of the healthy speakers. The moderate PD an ALS speakers showed the lowest "MEANCH" values, meaning the least overall acoustic change. And again, the measures that separated the healthy speakers from the PD and ALS speakers, also differentiated between mild PD and mild ALS in the production of /ujujuj/ and /wiwiwi/; the speakers with mild ALS showed significantly lower "MEANCH" than the mild PD speakers.

COVAR (The Coefficient of variation) separated mild and moderate ALS (which POSDYN, NEGDYN and MEANCH failed to do) from healthy, mild, and moderate PD speakers, especially during productions of /ujujuj/. COVAR values increased as follows: Healthy, mild, and moderate PD speakers < mild and moderate ALS. This implies that both groups of ALS speakers showed more variable TSC_MFCC values than the PD and healthy speakers.

C. Durational measures

Differences related to duration of the speech productions were found mostly for the ALS speakers, at a moderate level of impairment, as compared to the healthy and PD speakers.

At a moderate level of impairment, ALS speakers showed significantly longer event durations (eventDUR) for all the sequences, indicating that the complete transition from steady state vowel to steady state glide was longer (see Table IV and Fig. 5). The whole duration of the sequence (globalDUR) captured also the altered speech rate of the moderate ALS speakers as compared to the healthy speakers for all sequence types, but only for /ajajaj/ and /wiwiwi/ as compared to the moderate PD speakers.

IV. DISCUSSION

Our working hypothesis was that the contours derived from the changeover time of TSC_MFCC values indirectly conveyed information about underlying vocal tract shape changes, and that these changes would be different in the speech of healthy, mildly, and moderately dysarthric speakers with ALS and PD. Our findings show that the proposed measure successfully reveals differences between the groups and the results provide encouraging evidence that TSC_MFCCs captures acoustic subtleties in the signal that allow us to differentiate mildly and moderately dysarthric ALS and PD speech from the healthy ones and from each other effectively.

Before discussing the important findings, it is noted that the study distinguished between the event-related measures

"POSDYN" and "NEGDYN," which were corrected for duration from steady-state to steady-state, and the global measure "MEANCH." It should be noted, however, that the values for "POSDYN" and "NEGDYN" likely were not affected by the duration from steady-state to steady-state because the outcomes for "MEANCH" were identical to the outcomes for "POSDYN" and "NEGDYN." In addition, durational data indicated that only moderate ALS speakers differed from healthy speakers on this dimension. Thus, one may argue that "NEGDYN" and "POSDYN," which resulted both in higher values for healthy speakers compared to other speakers, indicated that the speech of healthy speakers differed from that of other populations only in the magnitude of the spectral change peaks. Therefore, the following discussion focusses on the results from "POSDYN," "NEGDYN," and "MEANCH" in terms of degree of spectral change.

Because the acoustic measure is not directly describing kinematic events, findings from earlier kinematic studies are first discussed to show that the proposed measure indeed captured properties linked to articulatory impairment in these dysarthric populations. First, the most consistent differences between groups were found when speakers produced the sequences /ujujuj/ and /wiwiwi/, which involves a front-back lingual and rounding/unrounding movements. For ALS speakers, the finding that the degree of spectral change was much lower than for the healthy speakers in these sequences, even at a mild level of severity, suggests smaller vocal tract changes and is consistent with earlier kinematic findings that demonstrated that lingual articulation is limited, especially in the posterior-anterior dimension (Lee *et al.*, 2017; Lee and Bell, 2018). Second, our study showed that mild PD speakers also realised /wiwiwi/ productions differently from healthy speakers, suggesting difficulties with these sequences at the earlier stages of dysarthria for this group as well. Individuals with PD have been shown to experience perioral stiffness (Caligiuri, 1987; Chu *et al.*, 2015). This likely affected the production of /wiwiwi/ more than the other sequences in the current study, as /wiwiwi/ involves both lip rounding and spreading. Although producing /ujujuj/ also involves a degree of lip rounding and unrounding, PD speakers did not differ from healthy speakers. A possible explanation is linked to the coarticulatory differences between these sequences. While /j/ is clearly sensitive to labial coarticulation allowing the full /ujujuj/ sequence to be produced on a single rounding gesture, the labial coarticulation is minimal in the /wiwiwi/ and its articulation requires a more complex alternation between rounded and unrounded lip positions. It would be interesting to examine the effect of articulatory complexity by employing alternating motion rate task in a follow-up study. This task is frequently used to test maximum motor performance to reveal neurological deficits and it has been shown, for example, that speakers with ALS face difficulty in adapting to the higher motor commands on the articulatory system (Eshghi *et al.*, 2019). Depending on the type of neurogenerative disease, different levels of adaptation are expected. Finally, the differences in degree of spectral change during productions of /ajajaj/ between mild ALS patients and mild PD and healthy speakers were not as salient as for the other sequences. This finding could be explained by the fact that the jaw serves as a compensating articulator for ALS speakers (Green *et al.*, 2013; Lee *et al.*, 2020; Mefferd, 2015; Yunusova *et al.*, 2013) and the deteriorating effects on lingual and bilabial articulatory control are thus longer obscured due to the efficiency of this articulator in supporting compensatory strategies /ajajaj/ sequences. The inherent lower degree of spectral changes for /ajajaj/ compared to the other two sequences might have played a role as well, causing a "floor" effect; the effect of the vocal tract changes between /a/ and /j/ on the extracted measures might not have been large enough to result in a significant difference between the different groups.

The type of sequence chosen to separate the different groups of dysarthric speakers is crucial and the study showed that especially /wiwiwi/ productions were affected in ALS and PD speakers. This stimulus dependent result is in accordance with Kim *et al.* (2009) who also found that the magnitude of the $F2$ slope differences between healthy and dysarthric PD speakers depended on the chosen material. Yunusova *et al.* (2008) revealed differences in kinematic measures as well depending on the spoken material suggesting differences in the role of individual articulators when producing words with diverse complexity of movement behavior.

Although the mild PD speakers also differed from the healthy speakers in the sense that they showed smaller degree of spectral change, they still showed larger degrees of change than the mild ALS speakers. The current metric of spectral energy change is successful in separating dysarthric speech of PD and ALS, possibly because the metric does not disregard any acoustic information which potentially reflects changes in vocal tract shape.

Whereas the degrees of acoustic change failed to distinguish moderate ALS from moderate PD, mild and moderate ALS showed much higher variability in acoustic change, and thus more variable vocal tract shape changes, compared to healthy and mild and moderate PD speakers, especially when producing /ujujuj/. This is an interesting finding which differs from findings of Kuruvilla-Dugdale and Mefferd (2017). In their kinematic study they found that mild and moderately dysarthric ALS speakers had less variable articulatory movements compared to healthy speakers, while only severely dysarthric speakers showed more spatiotemporal articulatory variability. Kuruvilla-Dugdale and Mefferd (2017) calculated spatiotemporal index variability across 12 repetitions of a sentence. Compared to healthy speakers, ALS patients in their study were thus less variable in their movement patterns across

repetitions of complete sentences and the method did not capture within-stimulus variability. Our study measured variability within a given sequence and the higher variability values likely reflect impairment at the level of execution. Methodological differences thus likely account for these different findings and more research is needed to examine variability within and across stimuli in speakers with motor speech disorders such as ALS.

Although moderate ALS speakers showed longer acoustic transitions between the individual segments and longer overall sequence measures than moderate PD and healthy speakers, duration values did not differ in the mild and healthy groups. Theses findings confirm evidence that dysarthric speech of ALS, especially at a more severe stage of dysarthria, is characterised by slower articulation and speaking rate, (Green *et al.*, 2013; Liss *et al.*, 2009; Yunusova *et al.*, 2008; Yunusova *et al.*, 2010; Yunusova *et al.*, 2012) and thus did not make it an efficient measure in our study to separate mild ALS from mild PD and healthy speakers.

It is concluded that TSC_MFCC measures correlated well with the findings observed on articulatory data in earlier studies investigating the two groups of motor speech disorders. Thus, it is tentatively assumed that the TSC_MFCC measure captures differences in vocal tract changes linked to articulatory impairments successfully. Whereas the durational measures were a less sensitive measure, the measures of variability and degree of spectral change show a promising potential in differentiating mild and moderate dysarthric speakers from each other and from the healthy population. Using these measures as a potential diagnostic tool for these two dysarthria types showing quite different speech distortion types is relatively straightforward. The vowel-glides sequences are globally easy to produce for the patients and are well suited to capture acoustic changeover time.

Future studies will focus on the feasibility of this approach to separate other dysarthria types and/or speech impairments. In addition, the results need to be validated cross-linguistically by including other varieties of French and other languages.

ACKNOWLEDGMENTS

The study was funded by the Swiss National Science Foundation, Sinergia program (CRSII5_173711, 10.2017-9.2020) and by a grant from Assistance Publique—Hôpitaux de Paris, Délégation de la Recherche Clinique et de l'Innovation (DRCI) to the second author. We thank the reviewers of the manuscript for their inspiring comments and all the participants of this study.

APPENDIX

TABLE V. β, standard error (SE), t, and p values for the event-related measures for the six comparisons between groups.

	POSDYN											
	AJ				UJ				WI			
	β	SE	t.ratio	p	β	SE	z.ratio	p	β	SE	z.ratio	p
Healthy ALS_mild	10.72	4.16	2.57	0.03	19.84	4.15	4.78	<0.0001	30.71	4.17	7.37	<0.0001
Healthy ALS_moderate	16.14	2.96	5.46	<0.0001	34	2.96	11.47	<0.0001	36.09	2.96	12.19	<0.0001
Healthy Park_mild	−0.84	3.18	−0.26	0.79	−1.57	3.18	−0.49	0.62	8.54	3.19	2.68	<0.01
Healthy Park_moderate	11.15	4.87	2.29	0.03	24.65	4.87	5.07	<0.0001	29.06	4.88	5.95	<0.0001
ALS_mild PD_mild	−11.56	4.81	−2.40	0.03	−21.41	4.8	−4.46	<0.0001	−22.16	4.82	−4.60	<0.0001
ALS moderate PD_moderate	−4.99	5.31	−0.94	0.42	−9.35	5.3	−1.76	0.09	−7.03	5.32	−1.32	0.19
	NEGDYN											
	β	SE	t.ratio	p	β	SE	z.ratio	p	β	SE	z.ratio	p
Healthy ALS_mild	11.75	4.17	2.82	0.01	20.33	4.15	4.90	<0.0001	31.6	4.17	7.58	<0.0001
Healthy ALS_moderate	16.46	2.96	5.56	<0.0001	32.75	2.97	11.03	<0.0001	37.01	2.96	12.48	<0.0001
Healthy Park_mild	−1.01	3.19	−0.32	0.75	−2.32	3.18	−0.73	0.46	7.56	3.2	2.36	0.02
Healthy Park_moderate	11.1	4.88	2.28	0.03	24.17	4.87	4.96	<0.0001	29.11	4.89	5.96	<0.0001
ALS_mild PD_mild	−12.76	4.82	−2.65	0.02	−22.66	4.8	−4.72	<0.0001	−24.04	4.82	−4.99	<0.0001
ALS moderate PD_moderate	−5.36	5.31	−1.01	0.38	−8.59	5.31	−1.62	0.13	−7.9	5.32	−1.49	0.14
	eventDUR											
	β	SE	t.ratio	p	β	SE	z.ratio	p	β	SE	z.ratio	p
Healthy ALS_mild	−0.01	0.01	−1.98	0.10	−0.01	0.01	−1.71	0.13	−0.01	0.01	−1.19	0.35
Healthy ALS_moderate	−0.03	0.00	−9.11	<0.0001	−0.03	0.00	−6.90	<0.0001	−0.03	0.00	−8.53	<0.0001
Healthy Park_mild	0.00	0.00	−0.35	0.73	0.00	0.00	−0.32	0.75	0.00	0.00	−0.20	0.85
Healthy Park_moderate	−0.01	0.01	−1.24	0.26	−0.01	0.01	−2.19	0.09	−0.01	0.01	−1.76	0.16
ALS_mild PD_mild	0.01	0.01	1.48	0.21	0.01	0.01	1.27	0.24	0.01	0.01	0.90	0.44
ALS moderate PD_moderate	0.03	0.01	3.94	<0.0001	0.01	0.01	1.84	0.13	0.02	0.01	3.15	<0.0001

https://doi.org/10.1121/10.0003332

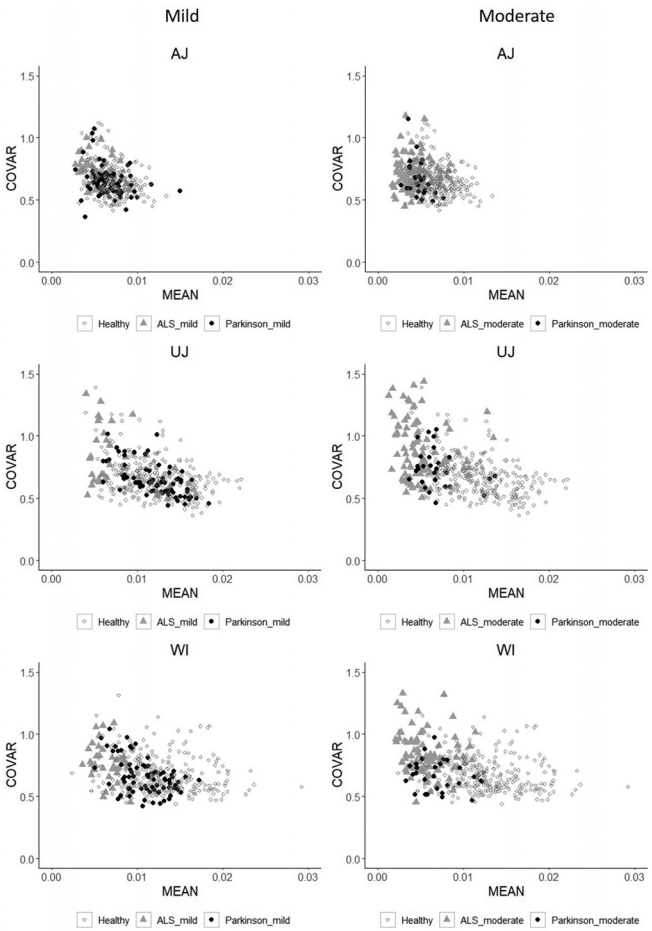

FIG. 4. Plot of mean values on the horizontal axis and the COVAR values on the vertical axis comparing the mild and moderate PD and ALS with Healthy speakers.

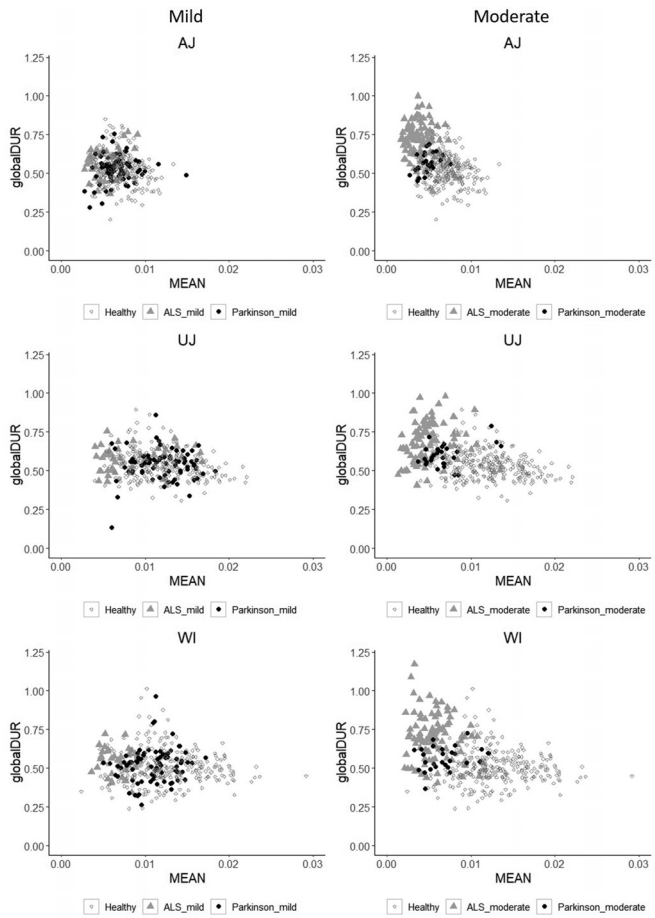

FIG. 5. Plot of MEANCH values of TSC_MFCC values over one sequence on the horizontal axis and the globalDUR values (seconds) on the vertical axis comparing the mild and moderate PD and ALS with Healthy speakers.

TABLE VI. β, SE, t, and p values for the global measures for the six comparisons between groups.

	MEANCH											
	AJ				UJ				WI			
	β	SE	t.ratio	p	β	SE	z.ratio	p	β	SE	z.ratio	p
Healthy ALS_mild	171.93	78.2	2.20	0.06	355.00	78.20	4.54	<0.0001	599.30	78.2	7.67	<0.0001
Healthy ALS_moderate	294.78	55.6	5.30	<0.0001	692.83	55.6	12.46	<0.0001	714.48	55.6	12.46	<0.0001
Healthy Park_mild	26.73	59.8	0.44	0.66	−9.43	59.80	−0.16	0.87	178.46	59.9	2.98	<0.01
Healthy Park_moderate	214.15	91.7	2.34	0.06	505.55	91.70	5.52	<0.0001	580.73	91.7	6.34	<0.0001
ALS_mild PD_mild	−145.19	90.5	−1.61	0.16	−364.42	90.5	−4.03	<0.001	−420.83	90.5	−4.65	<0.0001
ALS moderate PD_moderate	−80.63	99.9	−0.81	0.50	−187.28	99.9	−1.87	0.07	−133.75	99.9	−1.34	0.18
	COVAR											
	β	SE	t.ratio	p	β	SE	z.ratio	p	β	SE	z.ratio	p
Healthy ALS_mild	−0.09	0.03	−2.58	0.03	−0.12	0.03	−3.42	<0.01	−0.07	0.03	−2.03	0.07
Healthy ALS_moderate	−0.06	0.02	−2.38	0.04	−0.23	0.02	−9.21	<0.0001	−0.16	0.02	−6.44	<0.0001
Healthy Park_mild	0.01	0.03	0.58	0.67	0.01	0.03	0.54	0.59	0.02	0.03	0.79	0.51
Healthy Park_moderate	−0.02	0.04	0.39	0.69	−0.08	0.04	−1.89	0.07	0.01	0.04	0.26	0.79
ALS_mild PD_mild	0.10	0.04	2.61	0.03	0.13	0.04	3.31	<0.01	0.09	0.04	2.27	0.05
ALS moderate PD_moderate	0.04	0.04	0.96	0.51	0.15	0.04	3.38	<0.01	0.17	0.04	3.82	<0.001

	globalDUR											
	β	SE	t.ratio	p	β	SE	z.ratio	p	β	SE	z.ratio	p
Healthy ALS_mild	−0.05	0.03	−1.85	0.13	−0.04	0.03	−1.51	0.20	−0.03	0.03	−1.12	0.39
Healthy ALS_moderate	−0.19	0.02	−9.40	<0.0001	−0.14	0.02	−7.01	<0.0001	−0.19	0.02	−9.52	<0.0001
Healthy Park_mild	−0.01	0.02	−0.32	0.75	0.00	0.02	−0.19	0.85	−0.01	0.02	−0.66	0.59
Healthy Park_moderate	−0.04	0.03	−1.29	0.24	−0.07	0.03	−1.99	0.10	−0.06	0.03	−1.76	0.16
ALS_mild PD_mild	0.04	0.03	1.39	0.24	0.04	0.03	1.19	0.28	0.02	0.03	0.54	0.59
ALS moderate PD_moderate	0.14	0.04	4.04	<0.001	0.07	0.04	2.08	0.10	0.13	0.04	3.68	<0.001

[1]Spectral energy changes are not spatial in the sense that these are acoustic descriptors and not kinematic; however, in this study, we refer to them as "spatial" as they correlate well with kinematic measures (Goldstein, 2019).

[2]See https://www.unige.ch/fapse/mospeedi/. For the patient data collection, ethics were approved by the "commission cantonale d'éthique de la Recherche (CCER), Direction Générale de la Santé, République et Canton de Genève" (PD) and the APHP Ethics Committee (NCT03560661) in accordance with the 1964 Helsinki declaration and its later amendments (ALS). For the healthy control speakers, ethics were approved by the Faculty of Psychology (Switzerland), University of Geneva, and by the CPP SudOuest et OutreMer (ID_RCB: 2019-A02553-54) (France). Prior to inclusion, study participants provided informed consent.

[3]Examining different cut-off values showed that a higher frequency cut-off frequently resulted in a contour with too many valleys and peaks, not relevant for the vowel-glide and glide-vowel productions; a lower value, on the other hand, frequently resulted ininstances in which relevant peaks and valleys were missed.

[4]To exclude the possibility that the healthy French speakers from Paris differed from the French speakers in Geneva due to regional differences, an analysis was run on the dependent variables "POSDYN", "NEGDYN" and "eventDUR" with the speaker's origin as fixed effect. No differences on any of these measures were found between the speakers from Paris and Geneva.

[5]The False Discovery Rate is the proportion of tests that may give a significant result by chances (i.e., the number of false discoveries) over (divided by) the number of tests in the experiments (the maximum number of discoveries). If p^i is the ith smallest P-value out of N total P-values, the quantity $q^i = p^i\ N/i$ represents the number of expected false positives based on the p-value, divided by the total number of positives accepted at that same P-value. To correct for the lack of monotonicity between q and p, each q^i is substituted by the smallest q^k with $i > k$ (i.e., by the smallest q obtained by a P-value which was ranked after p^i).

Ackermann, H., Konczak, J., and Hertrich, I. (**1997**). "The temporal control of repetitive articulatory movements in Parkinson's disease," Brain Lang. **56**(2), 312–319.

Ackermann, H., and Ziegler, W. (**1991**). "Articulatory deficits in parkinsonian dysarthria: An acoustic analysis," J. Neurol. Neurosurg. Psychiatr. **54**(12), 1093–1098.

Auzou, P., and Rolland-Monnoury, V. (**2019**). *BECD: Batterie D'évaluation Clinique de la Dysarthrie* [*Clinical Dysarthria Assessment Battery*], Ortho Edition (Isbergues, France).

Benjamini, Y., and Hochberg, Y. (**1995**). "Controlling the false discovery rate: A practical and powerful approach to multiple testing," J. R. Statistical Soc. Ser. B (Methodological) **57**(1), 289–300.

Boersma, P., and Weenink, D. (**2021**). "Praat: Doing phonetics by computer [computer program]," Version 6.0.13, http://www.fon.hum.uva.nl/praat/manual/FAQ__How_to_cite_Praat.html (Last viewed 2 January 2021).

Brooks, B. R., Miller, R. G., Swash, M., and Munsat, T. L, and World Federation of Neurology Research Group on Motor Neuron Diseases (**2000**). "El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis," Amyotroph. Lateral Sclerosis Other Motor Neuron Disord. **1**(5), 293–299.

Caligiuri, M. P. (**1987**). "Labial kinematics during speech in patients with parkinsonian rigidity," Brain **110**(4), 1033–1044.

Caligiuri, M. P. (**1989**). "The influence of speaking rate on articulatory hypokinesia in Parkinsonian dysarthria," Brain Lang. **36**(3), 493–502.

Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., and Nakanishi, A. (**1999**). "The ALSFRS-R: A revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III)," J. Neurolog. Sci. **169**(1–2), 13–21.

Chu, S. Y., Barlow, S. M., and Lee, J. (**2015**). "Face-referenced measurement of perioral stiffness and speech kinematics in Parkinson's disease," J. Speech Lang. Hear. Res. **58**(2), 201–212.

Connor, N. P., Abbs, J. H., Cole, K. J., and Gracco, V. L. (**1989**). "Parkinsonian deficits in serial multiarticulate movements for speech," Brain **112**(4), 997–1009.

Darley, F. L., Aronson, A. E., and Brown, J. R. (**1969**). "Differential diagnostic patterns of dysarthria," J. Speech Hear. Res. **12**(2), 246–269.

Darley, F. L., Aronson, A. E., and Brown, J. R. (**1975**). *Motor Speech Disorders* (Saunders, Philadelphia, PA).

Davis, S. B., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust. Speech Signal Process. **28**(4), 357–365.

Delvaux, V., Roland, V., Huet, K., Piccaluga, M., Haelewyck, M. C., and Harmegnies, B. (**2016**). "The production of intervocalic glides in non dysarthric parkinsonian speech," in *Proceedings of the Annual Conference of the International Speech Communication Association*, INTERSPEECH, September 8–12, San Francisco, CA, pp. 253–256.

Duffy, J. R. (**2014**). *Motor Speech Disorders*, 4th ed. (Elsevier, Amsterdam).

Eshghi, M., Stipancic, K. L., Mefferd, A., Rong, P., Berry, J. D., Yunusova, Y., and Green, J. R. (**2019**). "Assessing oromotor capacity in ALS: The effect of a fixed-target task on lip biomechanics," Front. Neurol. **10**, 1288.

Fletcher, A. R., Wisler, A. A., McAuliffe, M. J., Lansford, K. L., and Liss, J. M. (**2017**). "Predicting intelligibility gains in dysarthria through automated speech feature analysis," J. Speech Lang. Hear. Res. **60**(11), 3058–3068.

Forrest, K., and Weismer, G. (**1995**). "Dynamic aspects of lower lip movement in parkinsonian and neurologically normal geriatric speakers' production of stress," J. Speech Lang. Hear. Res. **38**(2), 260–272.

Forrest, K., Weismer, G., and Turner, G. S. (**1989**). "Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults," J. Acoust. Soc. Am. **85**(6), 2608–2622.

Goldstein, L. (**2019**). "The role of temporal modulation in sensorimotor interaction," Front. Psychol. **10**, 2608.

Green, J. R., Yunusova, Y., Kuruvilla, M. S., Wang, J., Pattee, G. L., Synhorst, L., Zinman, L., and Berry, J. D. (**2013**). "Bulbar and speech motor assessment in ALS: Challenges and future directions," Amyotroph. Lateral Sclerosis Frontotemp. Degen. **14**(7–8), 494–500.

Hanson, E., Yorkston, K., and Britton, D. (**2011**). "Dysarthria in Amyotrophic Lateral Sclerosis: A systematic review of characteristics, speech treatment, and augmentative and alternative communication," J. Med. Speech Lang. Pathol. **19**, 12–30.

He, L., and Dellwo, V. (**2017**). "Amplitude envelope kinematics of speech: Parameter extraction and applications," J. Acoust. Soc. Am. **141**(5), 3582–3582.

Huang, X., Acero, A., Hon, H.-W., and Reddy, R. (**2001**). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, 1st ed. (Prentice Hall, Englewood Cliffs, NJ).

Khan, T., Westin, J., and Dougherty, M. (**2014**). "Classification of speech intelligibility in Parkinson's disease," Biocybernet. Biomed. Eng. **34**(1), 35–45.

Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R., and Zoing, M. C. (**2011**). "Amyotrophic lateral sclerosis," Lancet **377**(9769), 942–955.

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

Kim, Y., Weismer, G., Kent, R. D., and Duffy, J. R. (**2009**). "Statistical models of F2 slope in relation to severity of dysarthria," Folia Phoniatr. Logopaed. **61**(6), 329–335.

Koo, T. K., and Li, M. Y. (**2016**). "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," J. Chiropr. Med. **15**, 155–163.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2017**). "lmerTest Package: Tests in linear mixed effects models," J. Stat. Softw. **82**(1), 1–26.

Kuruvilla-Dugdale, M., and Mefferd, A. (**2017**). "Spatiotemporal movement variability in ALS: Speaking rate effects on tongue, lower lip, and jaw motor control," J. Commun. Disorders **67**, 22–34.

Lansford, K. L., and Liss, J. M. (**2014**). "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification," J. Speech Lang. Hear. Res. **57**(1), 57–67.

Lee, J., and Bell, M. (**2018**). "Articulatory range of movement in individuals with dysarthria secondary to amyotrophic lateral sclerosis," Am. J. Speech Lang. Pathol. **27**(3), 996–1009.

Lee, J., Littlejohn, M. A., and Simmons, Z. (**2017**). "Acoustic and tongue kinematic vowel space in speakers with and without dysarthria," Int. J. Speech Lang. Pathol. **19**(2), 195–204.

Lee, J., Rodriguez, E., and Mefferd, A. (**2020**). "Direction-specific jaw dysfunction and its impact on tongue movement in individuals with dysarthria secondary to amyotrophic lateral sclerosis," J. Speech Lang. Hear. Res. **63**(2), 499–508.

Lenth, R. V. (**2016**). "Least-squares means: The R Package lsmeans," J. Stat. Softw. **69**(1), 1–33.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., and Caviness, J. N. (**2009**). "Quantifying speech rhythm abnormalities in the dysarthrias," J. Speech Lang. Hear. Res. **52**(5), 1334–1352.

Lokesh, S., and Devi, M. R. (**2019**). "Speech recognition system using enhanced Mel frequency cepstral coefficient with windowing and framing method," Cluster Comput. **22**(5), 11669–11679.

Mefferd, A. S. (**2015**). "Articulatory-to-acoustic relations in talkers with dysarthria: A first analysis," J. Speech, Lang. Hear. Res. **58**(3), 576–589.

Mefferd, A., and Dietrich, M. (**2020**). "Tongue- and jaw-specific articulatory changes and their acoustic consequences in talkers with dysarthria due to amyotrophic lateral sclerosis: Effects of loud, clear, and slow speech," J. Speech Lang. Hear. Res. **63**, 2625–2636.

O'Shaughnessy, D. (**1987**). *Speech Communication: Human and Machine* (Addison-Wesley, Boston, MA).

R Core Team (**2020**). *R: A* Language *and* Environment *for* Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria, https://www.R-project.org/ (Last viewed 12 September 2020).

Rong, P., Yunusova, Y., Eshghi, M., Rowe, H. P., and Green, J. R. (**2019**). "A speech measure for early stratification of fast and slow progressors of bulbar amyotrophic lateral sclerosis: Lip movement jitter," Amyotrophic Lateral Sclerosis Frontotemp. Degen. **21**, 1–8.

Rudzicz, F., Frydenlund, A., Robertson, S., and Thaine, P. (**2016**). "Acoustic-articulatory relationships and inversion in sum-product and deep-belief networks," Speech Commun. **79**, 61–73.

Rudzicz, F., Hirst, G., and van Lieshout, P. (**2012**). "Vocal tract representation in the recognition of cerebral palsied speech," J. Speech Lang. Hear. Res. **55**(4), 1190–1207.

Shadle, C. H., Nam, H., and Whalen, D. H. (**2016**). "Comparing measurement errors for formants in synthetic and natural vowels," J. Acoust. Soc. Am. **139**(2), 713–727.

Shellikeri, S., Green, J. R., Kulkarni, M., Rong, P., Martino, R., Zinman, L., and Yunusova, Y. (**2016**). "Speech movement measures as markers of bulbar disease in Amyotrophic Lateral Sclerosis," J. Speech Lang. Hear. Res. **59**(5), 887–899.

Skodda, S., Grönheit, W., Mancinelli, N., and Schlegel, U. (**2013**). "Progression of voice and speech impairment in the course of Parkinson's disease: A longitudinal study," Parkinson's Disease **2013**, 1.

Skodda, S., Grönheit, W., and Schlegel, U. (**2012**). "Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease," PLoS ONE **7**(2), e32132.

Tjaden, K. (**2008**). "Speech and swallowing in Parkinson's disease," Top. Geriatric Rehabil. **24**(2), 115–126.

Tomik, B., and Guiloff, R. J. (**2010**). "Dysarthria in amyotrophic lateral sclerosis: A review," Amyotrophic Lateral Scler. **11**(1–2), 4–15.

Walsh, B., and Smith, A. (**2012**). "Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease," Movement Disord. **27**(7), 843–850.

Whitfield, J. A., and Goberman, A. M. (**2014**). "Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease," J. Commun. Disord. **51**, 19–28.

Yunusova, Y., Green, J. R., Greenwood, L., Wang, J., Pattee, G. L., and Zinman, L. (**2012**). "Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis," Folia Phoniatr. Logopaed. **64**(2), 94–102.

Yunusova, Y., Green, J. R., Lindstrom, M. J., Ball, L. J., Pattee, G. L., and Zinman, L. (**2010**). "Kinematics of disease progression in bulbar ALS," J. Commun. Disord. **43**(1), 6–20.

Yunusova, Y., Green, J. R., Lindstrom, M. J., Pattee, G. L., and Zinman, L. (**2013**). "Speech in ALS: Longitudinal changes in lips and jaw movements and vowel acoustics," J. Med. Speech Lang. Pathol. **21**(1), 1–13.

Yunusova, Y., Plowman, E. K., Green, J. R., Barnett, C., and Bede, P. (**2019**). "Clinical measures of bulbar dysfunction in ALS," Front. Neurology **10**, 106.

Yunusova, Y., Weismer, G., Westbury, J. R., and Lindstrom, M. J. (**2008**). "Articulatory movements during vowels in speakers with dysarthria and healthy controls," J. Speech Lang. Hear. Res. **51**(3), 596–611.