brain sciences

MDPI

Article

Acoustic Identification of Sentence Accent in Speakers with Dysarthria: Cross-Population Validation and Severity Related Patterns

Viviana Mendoza Ramos [1,2,*], Anja Lowit [3], Leen Van den Steen [1], Hector Arturo Kairuz Hernandez-Diaz [4], Maria Esperanza Hernandez-Diaz Huici [2], Marc De Bodt [1,2] and Gwen Van Nuffelen [1,2,5]

1 Department of Otorhinolaryngology, Head and Neck Surgery and Communication Disorders, University Hospital of Antwerp, Wilrijkstraat 10, 2650 Edegem, Belgium; Leen.VandenSteen@uza.be (L.V.d.S.); Marc.DeBodt@uza.be (M.D.B.); Gwen.VanNuffelen@uza.be (G.V.N.)
2 Faculty of Medicine and Health Sciences, Antwerp University, Wilrijkstraat 10, 2650 Edegem, Belgium; mariahuici@gmail.com
3 School of Psychological Sciences and Health, University of Strathclyde, 40 George Street, Glasgow G1 1QE, Scotland, UK; a.lowit@strath.ac.uk
4 Faculty of Electrical Engineering, Central University Marta Abreu of Las Villas, C. Camajuani km 5.5, Santa Clara 50100, Cuba; akairuz1985@gmail.com
5 Faculty of Medicine and Social Health Sciences, University of Ghent, De Pintelaan 185, 9000 Gent, Belgium
* Correspondence: Viviana.MendozaRamos@uza.be

Citation: Mendoza Ramos, V.; Lowit, A.; Van den Steen, L.; Kairuz Hernandez-Diaz, H.A.; Hernandez-Diaz Huici, M.E.; De Bodt, M.; Van Nuffelen, G. Acoustic Identification of Sentence Accent in Speakers with Dysarthria: Cross-Population Validation and Severity Related Patterns. *Brain Sci.* **2021**, *11*, 1344. https://doi.org/10.3390/brainsci11101344

Academic Editors: Diego Iacono and Paola Marangolo

Received: 21 August 2021
Accepted: 6 October 2021
Published: 13 October 2021

Abstract: Dysprosody is a hallmark of dysarthria, which can affect the intelligibility and naturalness of speech. This includes sentence accent, which helps to draw listeners' attention to important information in the message. Although some studies have investigated this feature, we currently lack properly validated automated procedures that can distinguish between subtle performance differences observed across speakers with dysarthria. This study aims for cross-population validation of a set of acoustic features that have previously been shown to correlate with sentence accent. In addition, the impact of dysarthria severity levels on sentence accent production is investigated. Two groups of adults were analysed (Dutch and English speakers). Fifty-eight participants with dysarthria and 30 healthy control participants (HCP) produced sentences with varying accent positions. All speech samples were evaluated perceptually and analysed acoustically with an algorithm that extracts ten meaningful prosodic features and allows a classification between accented and unaccented syllables based on a linear combination of these parameters. The data were statistically analysed using discriminant analysis. Within the Dutch and English dysarthric population, the algorithm correctly identified 82.8 and 91.9% of the accented target syllables, respectively, indicating that the capacity to discriminate between accented and unaccented syllables in a sentence is consistent with perceptual impressions. Moreover, different strategies for accent production across dysarthria severity levels could be demonstrated, which is an important step toward a better understanding of the nature of the deficit and the automatic classification of dysarthria severity using prosodic features.

Keywords: sentence accent; dysarthria; acoustic features; prosody

1. Introduction

Prosody forms part of the suprasegmental characteristics of speech and reflects meaningful variations in pitch, loudness, length, and pause across words, phrases, and sentences. Prosodic disturbances are considered a perceptual hallmark of dysarthria, a neurological motor speech disorder [1–3]. Common manifestations include a slowed or accelerated speech rate, monopitch and monoloudness, rhythmic disturbances and reduced ability to vary stress and accent [2–10]. Stress is a structural, linguistic property of a word that specifies which syllable in the word is more prominent than the others, and it is determined by the language system [11–13]. Accentuation is associated with the communicative

intention of highlighting important information within utterances (dependent on language behaviour), also known as focus [4,11–15].

Changes in pitch, loudness and/or duration—represented acoustically by changes in fundamental frequency (F0), intensity (I) and duration [4,16]—enable the speaker to make a clear distinction between the more and less important parts of his utterance. Consequently, effective accent or focus placement on an utterance is essential for the efficient conveyance of meaning [17], and disturbances can lead to reductions in intelligibility and naturalness of speech [1,6,8,16,18].

Previous case studies suggest that healthy speakers and speakers with dysarthria rely on different combinations of changes in F0, intensity, and duration to achieve sentence accent [9,10,19–24], but more profound insight into these strategies is lacking. Objective analysis and detection of valid acoustic descriptors of sentence accent may increase our understanding and support the clinical assessment of speech in patients with dysarthria [3].

Studies of acoustic correlates of sentence accent have provided valuable insight into this domain [12,25–29]. Acoustic accent production descriptors have been studied in healthy speech [30–37] and in dysarthria [9,10,16,24,38,39]. Currently, there is general agreement in the literature that syllable duration, pitch pattern, and intensity (or sub-band energy) correlate with accentuation [26,40]. These acoustic parameters have been used in systems for automatic accent detection; however, there is a lack of validated automatic analysis techniques to investigate accent production in disordered populations.

A recent study by Mendoza et al. [41] analysed the speech samples of 30 healthy control participants (HCP) and 50 participants with dysarthria, including different aetiologies and severity levels (ranging from mild to severe), who are all native Dutch speakers. The study demonstrated that sentence accent production could be characterised using a set of ten acoustic features. The selected features included not only the traditional values of F0, intensity, and duration measured within the target syllables, but also the differences of these three parameters with the values of their preceding syllable and with the median values of the entire utterance. These acoustic features demonstrated how a speaker manipulated F0, intensity, and duration to accentuate the target syllable within their prosodic capabilities. Furthermore, the combination of these features also allowed a reliable classification between accented and unaccented syllables in healthy and pathological speech. The study demonstrates the value of considering a more comprehensive range of variables and how they interact with each other in investigations of sentence accent production. However, further validation of the set of variables and the developed automatic analysis is necessary across different speaker populations varying in the type and severity of their speech disorders and spoken language.

Consequently, the purpose of this study is twofold. First, it aims to validate the methodology used by Mendoza et al. [41] with a new sample of British English speakers. Although most Germanic languages tend to produce sentence accent similarly [11,12,19,20,42], there might be subtle differences. Therefore, it is important to investigate whether Mendoza et al. [41] feature pool can equally distinguish accented from unaccented syllables in impaired speech in other languages. Second, the study aims to investigate the extent to which a more detailed analysis of accent production has the potential to reflect the severity of dysarthria.

2. Materials and Methods

2.1. Speech Samples

Samples of adult native speakers of Dutch and English were analysed. For the Dutch samples, 30 HCP and 50 participants with different types of dysarthria (spastic, flaccid, ataxic, hypokinetic, unilateral upper motor neuron (UUMN), mixed) and all severity levels were selected from the 'Computerized Assessment and Treatment of Rate, Intonation, and Stress' (CATRIS) corpus [43], which was composed for prosody research. It contains samples from 36 control and 55 speakers with dysarthria and different types of speech tasks. They all reported sufficient visual and auditory abilities to participate in the study. Cognitive

skills were not explicitly screened, but all participants demonstrated sufficient abilities to understand and perform the assessment instructions appropriately. For the present study, only speech samples from the focus communicative function were selected, 1 of the 55 participants did not perform the focus task, and other samples from 10 speakers (6 HCP and 4 with dysarthria) were not included due to poor acoustic quality. The group with dysarthria included 31 male and 19 female participants with an age range between 30 and 87 years (mean = 61 years, std = 13 years). The control group included 10 male and 20 female participants with an age range between 18 and 75 years (mean = 40 years, std = 15 years).

For the English samples, eight out of ten speakers with hereditary ataxia and dysarthria first described in Lowit et al. [9] were selected. Two speakers were excluded due to background noise in the audio recordings, making them unsuitable for the automatic analysis. The hearing and vision of all participants were normal or corrected-to-normal, and they had no significant cognitive deficits. The group included 3 male and 5 female participants with an age range between 28 and 72 years (mean = 52 years, std = 16 years).

The dysarthria severity level of all individuals ranged from mild over moderate to severe and was rated with a four-point grading scale (0 = normal; 1 = mild; 2 = moderate; 3 = severe) by three experienced speech and language pathologists (SLPs). Tables 1 and 2 summarise the selected speakers' characteristics, based on the severity of dysarthric features and perceptually rated intelligibility.

Table 1. Characteristics of the selected Dutch speakers.

			Gender		Dysarthria Severity		
Dysarthria Type	**Number of Subjects**	**Aetiology**	**F**	**M**	**1**	**2**	**3**
UUMN	2	Stroke-1, TBI-1		2	1	1	
Spastic	2	Stroke-2		2	1		1
Flaccid	7	Encephalopathy-1, Stroke-5, TBI-1		7	5	1	1
Ataxic	2	Neuropathy-1, Stroke-1	1	1	1	1	
Hypokinetic	31	PD-28, Stroke-3	17	14	21	8	2
Mixed	3	ALS-1, Encephalopathy-1, Stroke-1		3	2	1	
Undetermined	3	BT-1, Stroke-2	1	2	3		
Total	50		19	31	34	12	4

Note: M = male; F = female; PD = idiopathic Parkinson's disease; TBI = Traumatic brain injury; BT = brain tumour; ALS = amyotrophic lateral sclerosis; UUMN = Unilateral upper motor neuron; Dysarthria severity scale: 1 = mild, 2 = moderate, 3 = severe.

Table 2. Characteristics of the selected English speakers.

			Gender		Dysarthria Severity		
Dysarthria Type	**Number of Subjects**	**Aetiology**	**F**	**M**	**1**	**2**	**3**
Ataxic	8	CA-3, FA-3, SCA6-1, SCA8-1	5	3	2	4	2

Note: M = male; F = female; CA = cerebellar ataxia of undefined type; SCA = spinocerebellar ataxia; FA = Friedreich's ataxia; Dysarthria severity scale: 1 = mild, 2 = moderate, 3 = severe.

2.2. Speech Production Tasks

The speech material consisted of a comparable set of sentences in both languages based on the standard paradigm to elicit sentence accent, i.e., repetitions of the same sentence with varying accent positions depending on the asked question [9,39,44–46]. The questions were structured to elicited new information rather than contrastive focus. For the Dutch speakers, the sample included 3 different sentences. Each sentence was elicited twice,

with the focus occurring either in the initial (one case), medial (two cases), or final sentence position (three cases), resulting in six productions per participant, Table 3. A total of 180 and 300 sentences were available for the control and dysarthric group, respectively. For the English sample, speakers produced a set of 10 sentences, in which the focus was either in the initial, medial, or final position (10 cases for each position), resulting in 30 productions per speaker. A total of 240 sentences were included for this group. Table 3 summarises the focus sentences used in each language. In all cases, the participants were instructed to accent the typographically highlighted word.

Table 3. Focus sentences used in each experiment.

Dutch Task	English Task
Ze **wil** geen **telefoon** meer krijgen. *She does not want to get any more calls.*	The **gardener** grew **roses** in **London**.
Luc werkt in het **ziekenhuis**. *Luke works at the hospital.*	The **minister** has a **nanny** from **Norway**.
Misschien heeft **Piet vakantie**. *Maybe Pete is on holiday.*	The **model** wrote her **memoirs** in **Lima**.
	The **diva** made a **movie** in **Venice**.
	The **lawyer** met the **model** in **London**.
	The **widow** bought a **villa** in **Ealing**.
	The **neighbour** plays **melodies** on her **mandolin**.
	The **milliner** got a **memo** from **Melanie**.
	The **murderer** met his **lover** in **Limerick**.

Note. English translations in italics for the sentences in Dutch.

In the Dutch samples, the perceived accent of each sentence was assessed by three experienced clinicians, who independently marked the accented syllables without knowledge of the target word. If at least two judges had assigned the label 'accented' to a particular syllable, that label was retained. For the perceptual analysis of the English speakers, five SLPs judged the samples, deciding whether single or multiple elements were accented and indicating their respective locations [9]. They had the option of selecting a single or multiple accented word. Decisions on which syllable(s) was accented were again made by majority rule. Native speakers with intact hearing completed all the evaluations. Table 4 summarises the number of perceived accented syllables included in each group's analysis.

Table 4. Perceptually detected accented and unaccented syllables for each group of speakers.

Syllables	Speaker Population		
	Control	Dysarthric Dutch	Dysarthric English
Accented	197	338	259
Unaccented	1160	1891	1988
Total	1357	2229	2247

2.3. Acoustic Analysis

2.3.1. Set of Parameters Used to Detect and Describe Sentence Accent

The acoustic analysis was performed with MatLab software (version 2019b). An automatic algorithm described by Mendoza et al. [41] was used to detect the syllable nuclei and extract fundamental frequency (F0), energy, and duration. F0 values and intensity values were then normalised to make them speaker-independent, i.e., F0 was transposed into semitones (ST) [47], and intensity was normalised with respect to the maximum amplitude value of each sentence. Duration values (D) are reported as absolute values in milliseconds (ms). The algorithm automatically detected syllable nucleus boundaries, energy envelopes,

and F0 and plotted them over the spectrogram of each utterance, where they could be visually and auditorily inspected as well as manually corrected if necessary (see [41], for a detailed description of the automatic algorithm). Syllable nucleus boundaries had been manually modified when the algorithm had erroneously marked the boundaries, and corrections of F0 were required in 25 sentences of the total population (3.47%). Subsequently, a total of ten acoustic features were calculated for each syllable nucleus within an utterance (Table 5). Three parameters were inherent to each syllable, four were calculated in comparison with the previously uttered syllable and three others in comparison with the median of the entire sentence (the median was selected because it is a good estimator of central tendency in small samples). These parameters were then used to determine the acoustic differences between perceptually identified accented and unaccented syllables.

Table 5. Set of parameters derived from F0, Duration, and Intensity used in this study [41].

Parameters inherent to the syllable
ΔF0—the difference between the initial and the final value of F0 within the syllable nucleus in semitones
Int—maximum intensity of the syllable nucleus, relative to the overall utterance amplitude envelope
F0 & Int—the interaction of F0max (the maximum F0 value within the syllable nucleus) minus F0min (the minimum F0 value within the syllable nucleus) multiplied by Int
Parameters in comparison with the preceding syllable
dF0max—the difference between the F0max of each syllable with that of the preceding one
dF0min—the difference between the F0min of each syllable with that of the preceding one
dInt—the difference between Int of each syllable and that of the preceding one
dDrange—the difference between the duration of each syllable and that of the preceding one (dD) normalised to the range of all dD values in the sentenceNote. For initial syllables, the second syllable was used to calculate the difference
Parameters in comparison with the utterance
F0maxM—the difference between F0max and the median F0 of the utterance
IntM—the difference between Int and the median Int of the utterance
DM—the normalised duration (D, associated with the time length of the syllable nucleus) with respect to the median of all the D values in the sentence

2.3.2. Accent Detection in Dysarthric Speech and Description by Severity Levels

The set of independent acoustic features outlined in Table 5 was used as a predictor of accent placement. Discriminant analysis was performed to determine a linear combination of the ten parameters that enables the identification of the following two categories: accented and unaccented syllables [48,49]. The coefficients for the linear equation were calculated based on the group of Dutch speakers with dysarthria (Table 4). The discriminant analysis was performed using the Statistical Package for the Social Sciences (SPSS) software (version 21). The discriminative capacity of the equation was then validated with the English corpus of speakers with ataxic dysarthria, which included a total of 2247 syllables classified as accented or unaccented.

To investigate the strategies used for accent production across the different dysarthria severity levels (DSL), we performed a discriminant analysis using the SPSS software (version 21). For this analysis, the Dutch and English data were merged. The front-end processing used the set of ten acoustic features previously defined in Table 5, the independent variables were used together, and the Wilks's Lambda criteria was selected [48]. The different severities of dysarthria (mild, moderate, and severe) were analysed separately; the aim was to identify the contribution of each acoustic feature to accent production for each severity level in order to look for possible differences in the accentuation patterns.

3. Results

3.1. Validation of the Acoustic Features for Accent Detection

The discriminant analysis was initially performed with the samples of the Dutch-speaking population with dysarthria. As a result, the unstandardised discriminant function coefficients were obtained (Table 6). They were used to construct the following actual prediction Equation (1), which was used to classify the new English language cases in this study:

$$Y = \beta_1 * \Delta F0 + \beta_2 * Int + \beta_3 * F0 \,\&\, Int + \beta_4 * dF0min + \beta_5 * dF0max + \ldots + C \quad (1)$$

where Y is the discriminant score, β's are the unstandardised discriminant function coefficients, and C is a constant. Y is the score obtained from the linear combination of the β coefficients (listed in Table 6) multiplied by each discriminant feature. For $Y \geq 0.86$, the syllable is classified as accented; for $Y < 0.86$, the syllable is unaccented. The cut-off value (0.86) is the mean of the two centroids (Table 7), which are the mean value of the discriminant score for a given category (un/accented) of the dependent variable.

Table 6. Canonical Discriminant Function Coefficients.

Parameters	Function
ΔF0	$\beta_1 = 0.043$
Int	$\beta_2 = 0.605$
F0 & Int	$\beta_3 = 0.119$
dF0min	$\beta_4 = -0.042$
dF0max	$\beta_5 = -0.119$
dInt	$\beta_6 = 0.192$
dDrange	$\beta_7 = 0.336$
F0maxM	$\beta_8 = 0.041$
DM	$\beta_9 = -0.378$
IntM	$\beta_{10} = 1.095$
Constant	$C = -0.595$

Note: Unstandardized coefficients.

Table 7. Functions at Group Centroids.

Category	Function
Accented	2.094
Unaccented	−0.374

Note: Unstandardised canonical discriminant functions evaluated at group means.

The linear combination of the ten acoustic parameters was then applied to classify accented and unaccented syllables in the English sample. As previously reported by Mendoza et al. [41], the results for the Dutch speakers showed a percentage of correct classification of 82.8% for accented syllables and 90.5% for unaccented syllables for the speakers with dysarthria and 87.3 and 96.6% for the control group, respectively. Table 8 shows the confusion matrix with the results (in %) of correct classification for the two categories of the dependent variable (accented versus unaccented syllables) for the newly analysed English corpus, indicating that the approach worked equally well across the two speaker populations. The Receiver Operating Characteristic (ROC) curve was represented in Figure 1; this is a graphical representation of the equation's performance, representing the true-positive rate against the false-positive rate. The area under the ROC curve (AUC) is a measure of how well the equation can discriminate between the two outputs (un/accented) syllables; for our study, AUC = 0.964, 95% confidence interval: 0.952–0.975, $p < 0.001$.

Table 8. Confusion matrix for the classification results of the English speakers with dysarthria.

	Predicted Group Membership	
Target Syllables	**Accented**	**Unaccented**
Accented	91.9%	8.1%
Unaccented	7.8%	92.2%

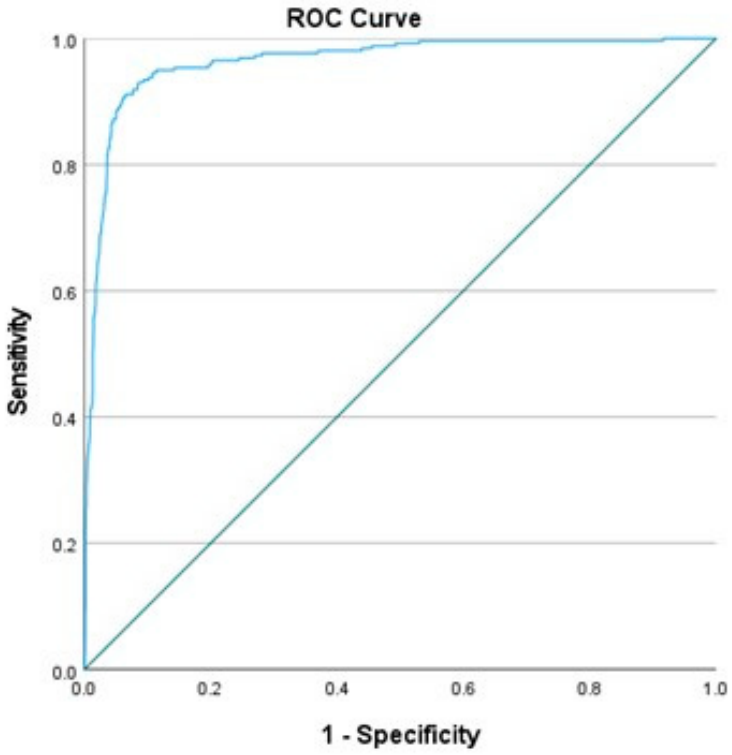

Figure 1. ROC curve representing the discriminatory ability of Equation (1) on the task of detecting un/accented syllables.

3.2. Impact of Different Dysarthria Severity Levels on Production Patterns for Sentence Accent

The discriminant analysis was applied individually to the different severity levels, showing the standardised canonical discriminant function coefficients per group. The magnitude of these coefficients indicates the relative importance of each independent acoustic feature in predicting the accent. They also allow for a comparison of the parameters measured on different scales (F0, Intensity, Duration). Coefficients with large absolute values correspond to acoustic features with greater discriminating ability. The discriminative coefficients are listed in Table 9.

Table 9. Standardised canonical discriminant function coefficients per group of dysarthria severity levels (mixed groups of English and Dutch speakers).

	Severity Levels			
	Control (30)	**Mild (36)**	**Moderate (16)**	**Severe (6)**
Parameters Inherent to the syllable				
ΔF0	0.261	0.200	0.184	0.060
Int	0.015	0.104	0.226	**0.240**
F0 & Int	**0.420**	**0.436**	**0.241**	0.175
In comparison with the preceding syllable				
dF0min	−0.175	−0.193	−0.200	−0.158
dF0max	**−0.367**	**−0.337**	**−0.271**	**−0.230**
dInt	0.029	−0.046	−0.007	−0.133
dDrange	−0.080	0.157	0.057	0.054
In comparison with the sentence				
F0maxM	0.239	0.146	0.113	−0.015
DM	−0.059	−0.263	−0.186	−0.076
IntM	0.212	0.168	**0.435**	**0.568**

The values of these coefficients indicated the features predominantly used to produce accent in healthy and dysarthric speech. Variations were observed between the severity levels. For example, the HCP tended to use changes in F0 and intensity (F0 & Int) within the target syllable supported by an increase in F0 in relation to the preceding syllable (dF0max) to produce a detectable accent. The group of speakers with mild dysarthria showed a similar tendency, meaning that they retained control over F0 to highlight important information in the sentence. Then, as the severity progressed, the pattern increasingly deviated from the HCP pattern. The group with moderate dysarthria used the same two main features, although they were not as prominent as in the HCP. In addition, the target syllables were highlighted by means of intensity contrast to the rest of the sentence (IntM). On the other hand, the participants with severe dysarthria only used one of the main features applied by HCP, (dF0max) and supplemented this strategy by manipulating intensity more prominently (Int and IntM). This group of speakers appeared to have less control over F0 but managed to compensate with intensity changes.

4. Discussion

4.1. Cross-Population Validation of Acoustic Features

This study validated an automatic system that extracts ten specific acoustic features derived from F0, intensity, and duration, used for sentence accent identification across different languages and speaker populations with atypical prosody. The acoustic features were divided into three categories (Table 5), the syllable's inherent parameters, the parameters of the syllable in contrast with the preceding syllable and the parameters of the syllable in contrast with the entire sentence. This set of features was used in a discriminant function to classify between accented and unaccented syllables, achieving 91.9% of correct classification of accented syllables and 92.2% of correct classification of unaccented syllables for the new population of English speakers affected with ataxic dysarthria. The classification accuracy results are comparable with the results of our previous study for native Dutch speakers (healthy and dysarthric speech) and with other studies of accent detection in healthy speech [30–35]. The results suggest that combining the ten acoustic parameters developed by Mendoza et al. [41] has a good capacity to discriminate between accented and unaccented syllables in healthy and speech-impaired speakers of Germanic languages with comparable accentuation patterns, such as English and Dutch.

In clinical practice, this automatic accent detection system could significantly reduce the time required to analyse speech data and provide quantitative information of prosodic parameters that could be useful as diagnostic and outcome measures. This could help clinicians define and implement more precise therapeutic approaches based on the identification of specific compensatory strategies of accent production. In addition, the current system's focus on within utterance variables may, in the future, allow a move away from structured sentence accent tasks toward more naturalistic speech samples as the basis for analysis, thus providing greater face validity to the information gained from the investigation of both healthy and disordered speech.

This study did not investigate the erroneous classifications in further detail. However, a preliminary inspection of the misclassified syllables showed some utterances where the system detected two accents and the listeners only one. Such cases could indicate specific dysarthric speech deficits such as excess stress where several syllables in an utterance received similar levels of accent as often reported for ataxic dysarthria or the reduced stress characteristic of hypokinetic dysarthria where no syllable, in particular, is highlighted from the rest. In such cases, listeners might have felt compelled to identify a single accent target, leading to the mismatch between perceptual and acoustic analysis results. Further investigation of such utterances and perceptual studies of what prompts a listener to identify a particular word in an utterance as accented may shed more light on these cases in the future. In the meantime, it is important to keep in mind that the so-called errors made by the automatic analysis might not reflect analysis mistakes but additional features of dysarthric speech performance.

4.2. Impact of Severity on Accent Production

As the current data show, the ability to accentuate or highlight information within an utterance is not related to the overall severity of the dysarthria, as even the severely affected participants (SAP) managed to place the accent in their sentences successfully. However, limited information is available to date on whether severity influences the acoustic patterns used to signal sentence accent. This could guide more effective intervention approaches for the speakers. Previous research, such as a study by Lowit et al. [9], did not find a strong correlation between dysarthria severity and accent production patterns; however, this study was based on a much-reduced set of features than that applied by Mendoza et al. [41].

The present study shows clear differences in how acoustic prosodic features were manipulated by the different speaker groups. The speakers with mild dysarthria tended to use similar strategies to the control group, i.e., they conveyed accent by making changes to F0 within the target syllable, with a simultaneous increase in the intensity (F0 & Int) and contrast in the frequency between the target syllable and the preceding syllable (dF0max). This result is in line with previous studies that reported F0 as the primary marker for accent perception [50–52] and found that speakers with acquired motor speech disorders could use the same pitch patterns as HCP [10,53,54].

The group of moderately affected participants (MAP) appeared to still have control over F0 in the target syllable and in contrast with the preceding syllable, although these features were not used as prominently as in the HCP. As a result, the MAP used additional compensatory strategies to produce accent, i.e., they also applied changes to intensity relative to the rest of the sentence (IntM).

The SAP demonstrated a reduced ability to control F0. However, they compensated for this by using mainly intensity (both within syllables (Int) and compared with the median intensity of the entire sentence (IntM)). This is an interesting observation given that Lowit et al. [38] demonstrated in a perceptual experiment that intensity could be a powerful signal for listeners and used in compensation for pitch. Thus, the speakers with ataxia seem to naturally employ the most effective compensatory feature to counteract their deficit in F0 manipulation.

This objective analysis of accent patterns within different levels of severity contributes to a better understanding of the nature of dysarthric speech and its deviant characteristics. It could also help clinicians determine the remaining acoustical cues for accent production in order to select optimal treatment strategies.

5. Limitations and Further Directions

Although the results of this study are promising for the automatic detection of sentence accent in dysarthria within different aetiologies and all severity levels, several topics require further research. The potential of this analysis can go beyond the traditional focus tasks and analyse more natural language production, which is highly important in clinical practice; therefore, this will be considered in future studies. The different patterns of accent production between groups of dysarthria severity levels observed in this study also deserve further investigation. More detailed investigations are required to further validate the generality of these results and to expand their scope. It would be useful to carry out a replication study on a larger population, which would allow for a more robust (statistical) analysis of the results.

As discussed above, there was a mismatch between the un/accented syllables classified by the automatic system and the listeners. Further analysis into listener strategies to identify accented words would be helpful to clarify the extent to which this phenomenon was due to speaker characteristics or measurement errors, which would require further refinement of the acoustic features in order to improve system performance. In addition, the quality of the recordings was not optimal in some samples, further analysis of the degree to which this might have impacted the accuracy of the results would be significant. Future studies could evaluate the algorithm performance in non-Germanic languages. Additionally, an investigation of accentuation patterns within different types of dysarthria

might be useful to better understand underlying problems and compensatory strategies. Despite these limitations, the presented method is suitable for future research investigating larger samples of dysarthric speech. It can provide insight into the patients' motor control processes and support patient-tailored therapeutic interventions—what the problem is and how to compensate for it.

6. Conclusions

This cross-population study validated a detailed set of acoustic descriptors related to F0, intensity, and duration (calculated within utterances) used for the automatic detection of sentence accent in dysarthric speech. The discrimination between accented and unaccented syllables using the automatic algorithm was accurate for both populations (Dutch and English speakers with dysarthria).

In addition, the validated acoustic features could adequately describe the strategies used for accent production across different severities. They provided a detailed objective description and a deeper understanding of the strategies or compensatory mechanisms used by speakers with dysarthria to highlight important information in a spoken message.

The clinical significance of this study is threefold: faster automatic detection of accent production, an objective analysis useful as an outcome measure, and support in determining therapeutic strategies.

Author Contributions: V.M.R. and A.L. conceptualised the design and methodology of the study, conducted the formal data analysis, and wrote the manuscript. A.L. collected the data of English speakers. A.L., L.V.d.S., M.D.B. and G.N performed perceptual evaluations of the data. L.V.d.S. participated in the data analysis. H.A.K.H.-D., M.E.H.-D.H. and V.M.R. designed and developed the automatic acoustic analysis. M.E.H.-D.H. participated in the data processing. M.D.B. and G.V.N. acquired the funding, supervised the study, and reviewed and edited the manuscript. All authors contributed to the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union's Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie (grant agreement number 766287).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, Ethical approval granted by NHS research ethics board Glasgow West, REC Ref 06/S0709/12 and by the Ethics Committee of the Antwerp University Hospital (Ref 12/35/273 and date of approval 17/09/2012).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The English speech corpus was collected as part of a grant provided by Ataxia UK to the second author. We thank all participants for their involvement in providing the speech samples used for this analysis. We also thank Heidi Martens (SLP), who collected the Dutch corpus for prosody research, funded by the Flemish Agency for Innovation by Science and Technology (IWT-TBM 080662).

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Darley, F.L.; Aronson, A.E.; Brown, J.R. Differential Diagnostic Patterns of Dysarthria. *J. Speech Hear. Res.* **1969**, *12*, 246–269. [CrossRef]
2. Darley, F.L.; Aronson, A.E.; Brown, J.R. *Motor Speech Disorders*; W. B. Saunders: Philadelphia, PA, USA, 1975.
3. Duffy, J.R. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, 4th ed.; Elsevier: St. Louis, MO, USA, 2020.
4. Peppé, S.J.E. Why is Prosody in Speech-Language Pathology So Difficult? *Int. J. Speech-Lang. Pathol.* **2009**, *11*, 258–271. [CrossRef]
5. Darley, F.L.; Aronson, A.E.; Brown, J.R. Clusters of Deviant Speech Dimensions in the Dysarthrias. *J. Speech Hear. Res.* **1969**, *12*, 462–496. [CrossRef] [PubMed]
6. Kent, R.; Rosenbek, J.C. Prosodic Disturbance and Neurologic Lesion. *Brain Lang.* **1982**, *15*, 259–291. [CrossRef]

7. Rosenbek, J.C.; LaPointe, L.L. The Dysarthrias: Description, Diagnosis, and Treatment. In *Clinical Management of Neurogenic Communicative Disorders*, 2nd ed.; Johns, D.F., Ed.; Little Brown: Boston, MA, USA, 1985; pp. 97–152.
8. Yorkston, K.M.; Beukelman, D.R.; Minifie, F.; Sapir, S. Assessment of stress patterning. In *The Dysarthria: Physiology, Acoustics, Perception, Management*; McNeil, M., Rosenbek, J., Aronson, A., Eds.; Pro-Ed: Austin, TX, USA, 1984; pp. 131–162.
9. Lowit, A.; Kuschmann, A.; MacLeod, J.M.; Schaeffler, F.; Mennen, I. Sentence Stress in Ataxic Dysarthria: A Perceptual and Acoustic Study. *J. Med Speech-Lang. Pathol.* **2010**, *18*, 77–82.
10. Lowit, A.; Kuschmann, A.; Kavanagh, K. Phonological Markers of Sentence Stress in Ataxic Dysarthria and their Relationship to Perceptual Cues. *J. Commun. Disord.* **2014**, *50*, 8–18. [CrossRef] [PubMed]
11. Sluijter, A.M.; Van Heuven, V.J. Spectral Balance as an Acoustic Correlate of Linguistic Stress. *J. Acoust. Soc. Am.* **1996**, *100*, 2471–2485. [CrossRef]
12. Sluijter, A.M.; van Heuven, V.J. Acoustic Correlates of Linguistic Stress and Accent in Dutch and American English. In Proceedings of the Fourth International Conference on Spoken Language (ICSLP 96), Philadelphia, PA, USA, 3–6 October 1996; Volume 2, pp. 630–633.
13. Cutler, A. Stress and Accent in Language Production and Understanding. In *Intonation, Accent and Rhythm: Studies in Discourse Phonology*; Gibbon, D., Richter, H., Eds.; De Gruyter: Berlin, Germany, 1984; pp. 77–90.
14. Rietveld, A.C.M.; van Heuven, V.J. *Algemene Fonetiek [General Phonetics]*; Coutinho: Bussum, The Netherlands, 2009.
15. Thies, T.; Mücke, D.; Lowit, A.; Kalbe, E.; Steffen, J.; Barbe, M.T. Prominence Marking in Parkinsonian Speech and its Correlation with Motor Performance and Cognitive Abilities. *Neuropsychologia* **2020**, *137*, 107306. [CrossRef]
16. Patel, R. Assessment of Prosody. In *Assessment of Motor Speech Disorders*; Lowit, A., Kent, R., Eds.; Plural Publishing: San Diego, CA, USA, 2011; pp. 75–95.
17. Cutler, A.; Dahan, D.; Van Donselaar, W. Prosody in the Comprehension of Spoken Language: A Literature Review. *Lang. Speech* **1997**, *40*, 141–201. [CrossRef]
18. De Bodt, M.S.; Huici, M.E.H.-D.; Van de Heyning, P. Intelligibility as a Linear Combination of Dimensions in Dysarthric Speech. *J. Commun. Disord.* **2002**, *35*, 283–292. [CrossRef]
19. Fry, D.B. Duration and Intensity as Physical Correlates of Linguistic Stress. *J. Acoust. Soc. Am.* **1955**, *27*, 765–768. [CrossRef]
20. Fry, D.B. Experiments in the Perception of Stress. *Lang. Speech* **1958**, *1*, 126–152. [CrossRef]
21. Lieberman, P. Some Acoustic Correlates of Word Stress in American English. *J. Acoust. Soc. Am.* **1960**, *32*, 451–454. [CrossRef]
22. Lehiste, I. *Suprasegmentals*; MIT Press: Cambridge, MA, USA, 1970.
23. Lehiste, I. Suprasegmental Features of Speech. In *Contemporary Issues in Experimental Phonetics*; Lass, N.J., Ed.; Academic Press: New York, NY, USA, 1976; pp. 225–239.
24. Patel, R.; Campellone, P. Acoustic and Perceptual Cues to Contrastive Stress in Dysarthria. *J. Speech. Lang. Hear. Res.* **2009**, *52*, 206–222. [CrossRef]
25. Pierrehumbert, J. *The Phonology and Phonetics of English Intonation*; MIT Press: Cambridge, MA, USA, 1980.
26. Wang, D.; Narayanan, S. An Acoustic Measure for Word Prominence in Spontaneous Speech. *IEEE Trans. Audio Speech Lang. Process.* **2007**, *15*, 690–701. [CrossRef] [PubMed]
27. Tamburini, F.; Wagner, P. On Automatic Prominence Detection for German. In Proceedings of the Eighth Annual Conference of the International Speech Communication Association, Antwerp, Belgium, 27–31 August 2007.
28. Al Moubayed, S.; Beskow, J. Prominence Detection in Swedish Using Syllable Correlates. In Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010, Makuhari, Chiba, Japan, 26–30 September 2010; pp. 1784–1787.
29. Streefkerk, B.M. Acoustical Correlates of Prominence: A Design for Research. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.38.2075&rep=rep1&type=pdf (accessed on 3 October 2021).
30. Streefkerk, B.M.; Pols, L.C.; Bosch, L.F.T. Acoustical Features as Predictors for Prominence in Read Aloud Dutch Sentences Used in ANN's. In Proceedings of the Sixth European Conference on Speech Communication and Technology, Budapest, Hungary, 5–9 September 1999.
31. Ananthakrishnan, S.; Narayanan, S.S. Automatic Prosodic Event Detection Using Acoustic, Lexical, and Syntactic Evidence. *IEEE Trans. Audio. Speech Lang. Process.* **2007**, *16*, 216–228. [CrossRef] [PubMed]
32. Johnson, D.O.; Kang, O. Automatic Prominent Syllable Detection with Machine Learning Classifiers. *Int. J. Speech Technol.* **2015**, *18*, 583–592. [CrossRef]
33. Nielsen, E.; Steedman, M.; Goldwater, S. The Role of Context in Neural Pitch Accent Detection in English. *arXiv* **2020**, arXiv:2004.14846.
34. Scharenborg, O.; Kakouros, S.; Post, B.; Meunier, F. Cross-Linguistic Influences on Sentence Accent Detection in Background Noise. *Lang. Speech* **2020**, *63*, 3–30. [CrossRef]
35. Li, K.; Mao, S.; Li, X.; Wu, Z.; Meng, H. Automatic Lexical Stress and Pitch Accent Detection for L2 English Speech Using Multi-Distribution Deep Neural Networks. *Speech Commun.* **2018**, *96*, 28–36. [CrossRef]
36. Li, K.; Liu, J. English Sentence Accent Detection Based on Auditory Features. *J. Tsinghua Univ. Sci. Technol.* **2010**, *50*, 613–617.
37. Li, K.; Zhang, S.; Li, M.; Lo, W.-K.; Meng, H. Prominence Model for Prosodic Features in Automatic Lexical Stress and Pitch Accent Detection. In Proceedings of the Twelfth Annual Conference of the International Speech Communication Association, Florence, Italy, 27–31 August 2011.

38. Lowit, A.; Ijitona, T.B.; Kuschmann, A.; Corson, S.; Soraghan, J. What Does it Take to Stress a Word? Digital Manipulation of Stress Markers in Ataxic Dysarthria. *Int. J. Lang. Commun. Disord.* **2018**, *53*, 875–887. [CrossRef]
39. Kuschmann, A.; Lowit, A. Sentence Stress in Children with Dysarthria and Cerebral Palsy. *Int. J. Speech-Lang. Pathol.* **2018**, *21*, 336–346. [CrossRef]
40. Tamburini, F. Automatic Prosodic Prominence Detection in Speech Using Acoustic Features: An Unsupervised System. In Proceedings of the Eighth European Conference on Speech Communication and Technology, Geneva, Switzerland, 1–4 September 2003.
41. Mendoza Ramos, V.; Hernandez-Diaz, H.A.K.; Huici, M.E.H.D.; Martens, H.; Van Nuffelen, G.; De Bodt, M. Acoustic Features to Characterise Sentence Accent Production in Dysarthric Speech. *Biomed. Signal Process. Control.* **2020**, *57*, 101750. [CrossRef]
42. Kochanski, G.; Grabe, E.; Coleman, J.; Rosner, B. Loudness Predicts Prominence: Fundamental Frequency Lends Little. *J. Acoust. Soc. Am.* **2005**, *118*, 1038–1054. [CrossRef]
43. Martens, H.; Dekens, T.; Latacz, L.; Van Nuffelen, G.; Verhelst, W.; De Bodt, M. Automated Assessment and Treatment of Speech Rate and Intonation in Dysarthria. In Proceedings of the ICTs for Improving Patients Rehabilitation Research Techniques, Venice, Italy, 5 May 2013.
44. Martens, H.; Van Nuffelen, G.; Cras, P.; Pickut, B.; De Letter, M.; De Bodt, M. Assessment of Prosodic Communicative Efficiency in Parkinson's Disease as Judged by Professional Listeners. *Parkinson's Dis.* **2011**, *2011*, 129310. [CrossRef] [PubMed]
45. Martens, H.; Van Nuffelen, G.; Wouters, K.; De Bodt, M. Reception of Communicative Functions of Prosody in Hypokinetic Dysarthria due to Parkinson's Disease. *J. Park. Dis.* **2016**, *6*, 219–229. [CrossRef]
46. Kuschmann, A.; Lowit, A. Pausing and Sentence Stress in Children with Dysarthria due to Cerebral Palsy. *Folia Phoniatr. Logop.* **2020**, 1–10. [CrossRef]
47. Gussenhoven, C.; Hart, J.; Collier, R.; Cohen, A. A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Speech Melody. *Language* **1990**, *68*, 610. [CrossRef]
48. Klecka, W.R. *Discriminant Analysis*; Sage: Beverly Hills, CA, USA, 1980.
49. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A.E. Linear Discriminant Analysis: A Detailed Tutorial. *AI Commun.* **2017**, *30*, 169–190. [CrossRef]
50. Bolinger, D.L. A Theory of Pitch Accent in English. *WORD* **1958**, *14*, 109–149. [CrossRef]
51. Van Katwijk, A. *Accentuation in Dutch: An Experimental Linguistic Study*; Van Gorcum: Amsterdam, The Netherlands; Assen, The Netherlands, 1974.
52. Hasegawa, Y.; Hata, K. Fundamental Frequency as an Acoustic Cue to Accent Perception. *Lang. Speech* **1992**, *35*, 87–98. [CrossRef] [PubMed]
53. Mennen, I.; Schaeffler, F.; Watt, N.; Miller, N. An Autosegmental-Metrical Investigation of Intonation in People with Parkinson's Disease. *Asia Pac. J. Speech Lang. Hear.* **2008**, *11*, 205–219. [CrossRef]
54. Verhoeven, J.; Mariën, P. Neurogenic Foreign Accent Syndrome: Articulatory Setting, Segments and Prosody in a Dutch Speaker. *J. Neurolinguistics* **2010**, *23*, 599–614. [CrossRef]