A preliminary approach to the acoustic-perceptual characterization of dysarthria

E. San Segundo[1] and J. Delgado[2]
[1]Dept. Spanish Language & General Linguistics, UNED, Madrid Spain
[2] Dept. Developmental and Educational Psychology, La Laguna University, Tenerife, Spain

Keywords — dysarthria, acoustic analyses, perceptual analysis, Vocal Profile Analysis, linear models

I. Introduction

Dysarthria is a speech disorder derived from a neurological damage of the central nervous and/or peripheral system that produces difficulties in motor programming or execution, resulting in the presence of alterations in the muscular pathway, strength, tone, speed and precision of the movements performed by the muscles involved in speech production, i.e. breathing, phonation, articulation and resonance [1]. While most recent studies place the focus on dysarthric individuals of specific neurodegenerative disorders, such as Parkinson's disease, in this investigation we have evaluated dysarthric patients in general (ataxic, spastic and mixed dysarthria) by analysing their speech production both acoustically and perceptually, and comparing the same acoustic and perceptual measures in a control population of neurologically healthy speakers. Our aim is to be able to characterize dysarthria by exploring a large number of acoustic predictors as well as different perceptual features relating to voice quality (VQ), understood in a broad sense as comprising both laryngeal and supralaryngeal long term configurations, or settings.

As for the acoustic features that have been investigated before for dysarthria, we find a range of spectral and cepstral measures, as well as measures of F0 variability, frequency and amplitude variation from cycle to cycle, features covering loudness, voicing, articulation and also temporal measures such as speech rate, syllable rate, syllable duration, number and length of pauses – to mention but a few [2, 3].

Less is known about the perceptual characterization of dysarthria, even though auditory assessment in clinical studies is still regarded as the gold standard with which acoustic measures are compared [4]. In this investigation, the Vocal Profile Analysis (VPA) [5] is used for the first time –to the best of our knowledge– to analyse dysarthria perceptually. Although the use of this protocol has been more widespread among forensic phoneticians [6], some previous studies exist which have explored the vocal profile characteristics of specific groups. For instance, in Down syndrome speakers Beck [7] found several vocal characteristics which were significantly different from an age-matched control group, many of them related to particular configurations of the vocal tract (e.g. protruded jaw or open jaw, fronted tongue body, pharyngeal constriction). Because the VPA protocol includes several supralaryngeal settings, it seemed a very appropriate scheme to analyse the speech of dysarthric subjects, whose characteristic muscle weakness results in difficult or unclear articulation of speech.

II. Method

A. *Participants*

Thirty native Spanish-speaking subjects voluntarily participated in this study, 15 with dysarthria (mean age 42.93, SD 10.31) and 15 neurologically healthy (mean age 41.86, SD 13.62). The two experimental groups (dysarthria and control) were sex-matched. Within the dysarthria group, ten participants present ataxic dysarthria, two spastic dysarthria and three mixed dysarthria.

B. *Recording setup and speech samples*

All recordings were conducted in a soundproof booth with an AKG C544L head-mounted condenser microphone digitized at a sampling rate of 44.1 kHz, and 16 bits of resolution using the audio interface *Alesis io2 express* and a personal computer with *Praat*. The signal-to-noise ratio (SNR) was measured post hoc to check the level of environmental noise of the voice recordings. All samples were consistent with the recommended threshold proposed by [8] (SNR>30dB). The speech material consisted in reading aloud four phonetically balanced sentences of the Spanish Matrix Sentences Test [9].

C. *Acoustic analysis*

All the acoustic measures (Table I) were extracted and analysed with *Praat*. Temporal acoustic analyses were conducted only on voiced segments of continuous speech through an automated detection *Praat* script [10]. The Cepstral and spectral acoustic analysis were obtained for every complete sample. The CPPs and LTAS slope values were calculated with the same Acoustic Voice Quality Index 03.01 script configuration [11]. Vowel Space Area (VSA) was calculated using the equation outlined by Sapir et al. [12]. The first and second formants of the stressed vowels /i/, /a/ and /u/ from the words "Carmen", "libros" and "azules", respectively, were extracted. Then, the means of each vowel formant were used to calculate the VSA. F2 range was calculated as the difference between the maximum and minimum F2 frequencies in the diphthong [je] of the word "tiene".

TABLE I.

Extracted acoustic measures				
F0	***Frequency disturbance***	***Amplitude disturbance***	***Other voice source features***	***Vocal tract features***
mean F0, standard deviation of the F0	jitter absolute, jitter (%), jitter (RAP)	shimmer absolute, shimmer (%), shimmer APQ3	smoothed cepstral peak prominence (CPPS), HNR	LTAS slope, VSA, F2 range

D. Perceptual analysis

The protocol used for the perceptual analysis was the Spanish version of the Simplified Vocal Profile Analysis (SVPA) that implements a Visual Analogue Scale per VQ setting in a computer-based interface, described in [13]. This protocol includes the following 10 settings, which describe the VQ of a speaker in terms of laryngeal and supralaryngeal long-term configurations: (1) *voice type*, (2) *laryngeal tension*, (3) *vocal tract tension*, (4) *laryngeal height*, (5) *pharyngeal*, (6) *velopharyngeal*, (7) *dorsal*, (8) *apical*, (9) *mandibular* and (10) *labial s*etting.

Two raters (one phonetician and one speech-language pathologist) listened to 36 speech samples and rated them (blindly; i.e. they did not know whether they were dysarthric or control), in the same order and with a short break every 12 stimuli. The stimuli belonged to the 30 participants described above, with 6 repeated voices, randomly selected with the aim of calculating intra-rater reliability. None of these raters had previous experience in using this specific protocol to judge voice or speech disorders.

III. RESULTS

Intraclass correlations coefficients (ICCs) were calculated to determine intra- and inter-rater reliability. In both cases, a two-way, consistency, single-measure model (ICC (2,1)) was used. The results show excellent inter-rater agreement in *vocal tract tension* (0.88) and *dorsal setting* (0.79); good agreement in *apical* (0.73) and *pharyngeal* (0.62); and fair agreement in *laryngeal tension* (0.53). Intra-rater agreement was excellent (> 0.75) in both raters for these five settings as well. These five settings were therefore selected for the linear model fitting.

For each of the five VPA settings indicated above we constructed a linear model as function of the acoustic variables, whether speakers are dysarthric or not, and whether they are male or female. The ratings of both raters were averaged for the purpose of statistical modelling. The goal was to find the relationship –and test for the interaction– between the perceptual ratings and the acoustic predictors. Statistical modelling was done using the linear model function of *R*. We tested collinearity of independent variables in order reduce the potential number of predictors down to a small number of non-correlated dimensions. The 8 acoustic predictors from Table 1 were finally used.

All the linear models were significant, meaning that the perceptual ratings given to the five VQ settings can be explained by a model covering all those predictors. For some VPA settings, the best model fitted to the data based on model comparison was a simpler model. This is the case of *vocal tract tension (VTT)*. We constructed a linear model of *VTT* ratings as a function of LTAS slope, VSA, F2 range, F0, whether speakers are dysarthric or not, and whether they are females or males. This model was significant ($F(6,23)= 22.4$, $p<0.001$). Results also show that raters are more likely to rate a voice in the range 70-80 along the VAS if the speaker has dysarthria and in the range 50-60 if the speaker belongs to the control group.

REFERENCES

[1] N. Melle, Guía de intervención logopédica en la disartria. Madrid: Editorial Síntesis, 2007.

[2] J. Delgado-Hernández, "Estudio piloto sobre los valores acústicos de las vocales en español como indicadores de la gravedad de la disartria," Rev Neurol 2017; 64: 105-11.

[3] E.J. Metter and W.R. Hanson, "Clinical and acoustical variability in hypokinetic dysarthria," Journal of communication disorders, 1986, 19(5), 347-366.

[4] E.P.M. Ma and E.M.L. Yu, "Multiparametric evaluation of dysphonic severity," J. Voice. 2005; 20, pp. 380–390.

[5] J. Laver, The phonetic description of voice quality. London: Cambridge Studies in Linguistics London, 1980.

[6] E. San Segundo, "International survey on voice quality: Forensic practitioners versus voice therapists," Est. de Fonética Exper, 2021; 30, pp. 8-33.

[7] J.M. Beck, "Organic variation of the vocal apparatus". In .- Gibbon, F., Hardcastle, W. J., & Laver, J (Eds), The handbook of phonetic sciences, pp. 256-297. Wiley. 1997.

[8] D.D. Deliyski, H.S. Shaw, M.K. Evans, "Adverse effects of environmental noise on acoustic voice quality measurements," J Voice, 2005; 19, pp. 15-28.

[9] S. Hochmuth, T. Brand, M.A. Zokoll, F. Zenker, N. Wardenga, B. Kollmeier B, "A Spanish matrix sentence test for assessing speech reception thresholds in noise," Int J Audiol, 2012; 51, pp. 536-544.

[10] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt, "Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels.," J Voice, 2010; 24, pp. 540-555.

[11] B. Barsties, Y. Maryn, "External validation of the Acoustic Voice Quality Index Version 03.01 with extended representativity," Ann Otol Rhinol Laryngol. 2016;125, pp. 571-583.

[12] E. San Segundo and R. Skarnitzl, "A computer-based tool for the assessment of voice quality through visual analogue scales: VAS-Simplified Vocal Profile Analysis," J Voice 2019.