A preliminary approach to the acoustic-perceptual characterization of dysarthria

Eugenia San Segundo[1], Jonathan Delgado[2]

[1]Department of Spanish Language & General Linguistics, UNED, Madrid, Spain,
[2]Department of Developmental and Educational Psychology, La Laguna University, Tenerife, Spain,
[1]*esansegundo@flog.uned.es*, [2]*extjdelgado@ull.edu.es*

Abstract

We have conducted an acoustic-perceptual evaluation of 15 dysarthric and 15 neurologically-healthy speakers. On the one hand, speech samples were analysed with *Praat* (13 acoustic parameters were extracted, related to F0, frequency and amplitude variation, as well as other source and vocal tract measures). On the other hand, two raters evaluated all the voices perceptually using the Simplified Vocal Profile Analysis (SVPA), which implements a visual analog scale for each voice quality setting in an online interface.

The results show, for the perceptual analyses, that (1) intra- and interrater agreement is overall very good; and that (2) the perceptual settings 'vocal tract tension' and 'laryngeal tension' are the most useful to characterize dysarthria. In terms of statistical modelling, most linear models were significant using only 4-5 acoustic parameters, but the specific parameters in each model depend on the VPA setting under consideration.

All in all, acoustic-perceptual assessment through the SVPA seems to be an important complement to traditional assessment in dysarthria, as it provides information on the functioning of the supraglottic structures commonly affected in this type of motor speech disorder, in which the muscles used to produce speech are damaged, paralyzed, or weakened.

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

1. 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 [2-4], 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 [5, 6].

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 [7]. In this investigation, the Vocal Profile Analysis (VPA) [8] 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 [9], some previous studies exist which have explored the vocal profile characteristics of specific groups. For instance, in Down syndrome speakers, Beck [10] 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.

Preprint version
To appear in Proceedings of the 3rd International Symposium on Applied Phonetics (ISAPh 2021)

2. Method

2.1. Participants

30 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, 10 participants present ataxic dysarthria, 2 spastic dysarthria and 3 mixed dysarthria.

2.2. 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 [11] (SNR>30dB). The speech material consisted in reading aloud four phonetically balanced sentences of the Spanish Matrix Sentences Test [12].

2.3. 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 [13]. 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 [14]. Vowel Space Area (VSA) was calculated using the equation outlined by Sapir et al. [15]. 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 1: Extracted acoustic measures. CPPs: smoothed Cepstral Peak Prominence; VSA: Vowel Space Area.

F0	Freq. variation	Amplitude variation	Voice source features	Vocal tract features
Mean F0	Jitter absolute	Shimmer absolute	CPPs	LTAS slope
F0 st dev	Jitter %	Shimmer%	HNR	VSA
	Jitter RAP	Shimmer APQ3		F2 range

2.4. 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 [16]. 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 setting.

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.

3. Results

3.1. Intra- and interrater agreement

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.

3.2. Linear models

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.

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 most VPA settings, the best model fitted to the data based on model comparison was a simpler model.

Table 2: Summary of results for linear models

VPA setting	Predictors	Model significance
Laryngeal tension	CPPs + F0 + Jitter% + Shimmer% + HNR + Group*** + Sex	(F(7,22)= 7.6, p<0.001 ***
Vocal tract tension	LTAS slope + VSA + F2 range + F0 + Group*** + Sex	(F(6,23)= 22.4, p<0.001 ***
Pharyngeal setting	CPPs* + LTAS slope + VSA + F2 range + Group + Sex	(F(6,23)= 22.4, p<0.05 *
Dorsal setting	CPPs* + LTAS slope + VSA + F2 range + Group + Sex	(F(6,23)= 9.5, p<0.001 ***
Apical setting	CPPs + LTAS slope + VSA + F2 range + Group + Sex	(F(6,23)= 7.1, p<0.001 ***

Table 2 shows the five VPA setting which had better intra- and interrater agreement (see section 3.1). Because reliability in the other settings was not so good (in some cases only slightly worse), particularly in inter-rater terms, these five were the only settings selected for the linear model fitting.

3.2.1. Laryngeal tension

For laryngeal tension, the model was very significant with five acoustic variables, whether the speakers are dysarthric or not, and whether they are male or female. However, the predictor ‘group’ (i.e. whether the speakers are dysarthric or not) weighs more in this model to explain the ratings (see Table 2). Figure 1 shows that raters are more likely to rate a voice in the range 65-75 along the scale (for laryngeal tension) if the speaker has dysarthria and in the range 45-55 if the speaker belongs to the control group.

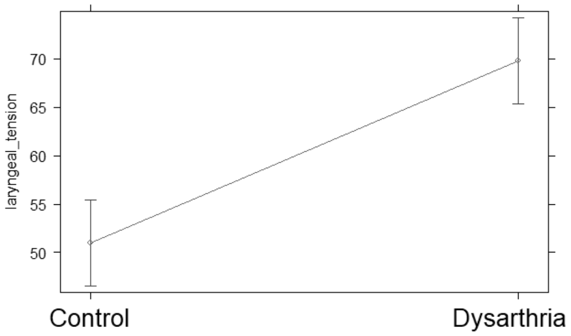

Figure 1: Group effect plot for laryngeal tension

3.2.2. Vocal tract tension

For vocal tract tension (VTT), we constructed a simpler 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), as shown in Table 2. Results also show that raters are more likely to rate a voice in the range 70-85 along the VAS if the speaker has dysarthria and in the range 45-60 if the speaker belongs to the control group (Figure 2).

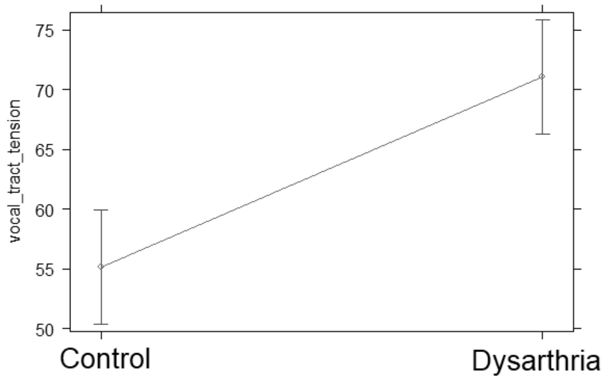

Figure 2: Group effect plot for vocal tract tension

4. Discussion

Intra- and interrater agreement was considerably good for most settings, so we conducted the linear model analyses with the averaged ratings of the two raters. Most linear models were significant using only 4-5 acoustic parameters, which suggests that simpler models are preferred over more complicated ones. The precise acoustic parameters used in each model depend on the VPA setting under consideration. For instance, it seems logical that jitter and shimmer appear in the laryngeal tension model because it is the only setting in Table 2 related to the voice source. The others are supralaryngeal settings and different types of acoustic parameters are important to explain the interaction between acoustics and perception, mostly a combination of CPPs, LTAS slope, VSA and F2 range.

It is worth mentioning that CPPs appears as a significant acoustic parameter in a couple of models. This might imply that dysarthric speech evaluation could be a new clinical field where CPP has great value for pathological speaker characterization. Indeed, previous studies have already shown the potential of CPP for determining VQ in other pathologies. For example, in 2018 the American Speech Language Hearing Association recommended CPP as a general measure of dysphonia. Other studies show that CPP is a cue to detect specific voice disorders, such as laryngeal disorders, vocal fold nodules, etc. [17, 18]. Some researchers have even proposed its application to other purposes such as the detection of cognitive load or the evaluation of the sexual appeal of a voice. [19, 20; in 21].

5. Conclusions

VPA settings related to tension are key to characterize dysarthria perceptually. In VPA terms, tension can occur both in the larynx and in the vocal tract. Dysarthric voices are perceived as more tense than non-dysarthric voices both at the laryngeal and supralaryngeal level. This seems to be a salient perceptual feature, as there is no overlap between both groups of speakers for these two aspects. However, a cluster of other VPA settings (related to the pharynx and the tongue: tip and blade) also play a role in distinguished dysarthric patients from the control group. These results can be explained by the fact that dysarthria is a motor speech disorder in which the muscles used to produce speech are damaged, paralyzed, or weakened. This weakness seems to be compensated with a certain overexertion or overstraining affecting muscular movements; in other words, a kind of tension.

All in all, we have shown that the VPA protocol allows for clinical applications besides its most common use in sociolinguistic and forensic studies. We can summarize the main implications of our study for clinical practice as follows. The combination of acoustic and perceptual assessment through the VPA scheme is an important complement to traditional assessment methods in dysarthria (typically unguided auditory evaluations, i.e. without following a specific protocol). The VPA scheme provides relevant information of the functioning of the supraglottic structures commonly affected in this type of motor speech disorders. Therefore, the information obtained through the VPA makes it possible to assess the evolution of dysarthria and to establish therapeutic objectives.

As directions for future work, we plan to add at least one more rater, to measure intelligibility of dysarthric speakers, which could be another variable to add to our models and might give us an idea of the degree of severity of dysarthria. Likewise, it would be useful to analyse prosodic, time-based measures, such as articulation rate, to further characterize this disorder. Eventually, if enough patients can be recruited, we will propose a VPA-based tool specific for dysarthria which can distinguish also between different types of dysarthria.

6. References

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

[2] Pinto, S., Chan, A., Guimarães, I., Rothe-Neves, R., Sadat, J. 2017. A cross-linguistic perspective to the study of dysarthria in Parkinson’s disease. *J Phonetics* 64, 156-167.

[3] Kreiman, J., Gerratt, B. 2019. Acoustic Analysis and Voice Quality in Parkinson Disease. *Automatic Assessment of Parkinsonian Speech Workshop*. Cham: Springer, 1-23.

[4] Gómez, A., Tsanas, A., Gómez-Vilda, P., Palacios-Alonso, D., Rodellar, V., Álvarez, A. 2021. Acoustic to kinematic projection in Parkinson's disease dysarthria. *Biomed. Signal Process. Control* 66, 102-422.

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

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

[7] Ma, E.P.M., Yu, E.M.L. 2005. Multiparametric evaluation of dysphonic severity. *J. Voice* 20, 380–390.

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

[9] San Segundo, E. 2021. International survey on voice quality: Forensic practitioners versus voice therapists. *Estudios de Fonética Experimental* 30, 8-33.

[10] Beck, J.M. 1997. Organic variation of the vocal apparatus. In: Gibbon, F., Hardcastle, W. J., Laver, J (eds), *The Handbook of Phonetic Sciences*. NJ: Wiley, 256-297.

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

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

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

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

[15] Sapir, S., Ramig, L., Spielman, J., Fox, C. 2010. Formant centralization ratio (FCR) as an acoustic index of dysarthric vowel articulation: comparison with vowel space area in Parkinson disease and healthy aging. *J Speech Lang Hear Res* 53, 114-125.

[16] San Segundo, E., Skarnitzl, R. 2021. A computer-based tool for the assessment of voice quality through visual analogue scales: VAS-Simplified Vocal Profile Analysis. *J Voice* 35(3), 497.e9-497.e21.

[17] Rosa, M.O., Pereira, J.C., Greller, M., Carvalho, A. 1999. Signal processing and statistical procedures to identify laryngeal pathologies. *Proc. IEEE Internal Conf. Electronics, Circuits, and Systems, ICECS'99*, vol. 1, Pafos, 423-426.

[18] Kumar, B.R., Bhat, J.S., Prasad, N. 2010. Cepstral analysis of voice in persons with vocal nodules. *J Voice* 24, 651-653.

[19] Yap, T.F, Epps, J., Ambikairajah, E., Choi, E.HC. 2011. Voice source features for cognitive load classification. *Proc. IEEE Internat. Conf. Acoustics, Speech, and Signal Processing - ICASSP 2011*, Prague, 5700-5703.

[20] Balasubramanium, R.K., Bhat, J.S., Srivastava, M., Eldose, A. 2012. Cepstral analysis of sexually appealing voice. *J Voice* 26, 412-415.

[21] Fraile, R., Godino-Llorente, J.I. 2014. Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control* 14, 42-54.