Spontaneous speech production by dysarthric and healthy speakers: Temporal organisation and speaking rate

Laurianne Georgeton, Christine Meunier

Aix-Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France
laurianne.georgeton@lpl-aix.fr, christine.meunier@lpl-aix.fr

ABSTRACT

This study compares speaking rate in spontaneous speech between dysarthric and healthy speakers. Since dysarthria involves heterogeneous pathologies, two types of dysarthria (i.e. Parkinson's disease and amyotrophic lateral sclerosis) have been distinguished. We hypothesize that temporal organisation of speech may be different between healthy and dysarthric speakers, but also between both pathological populations. Four measurements have been explored. Results show that Parkinsonian speakers are characterized by a short Inter-Phrasal Unit (IPU) while words duration is similar to healthy speakers. Amyotrophic Lateral Sclerosis speakers produced long words while IPU duration is similar to healthy speakers. Number of words per IPU distinguishes healthy speakers from dysarthric speakers while number of syllables per second separates Amyotrophic Lateral Sclerosis speakers from Parkinsonian speakers. These results suggest that the boundary between pathological and healthy speech should be examined with multidimensional analyses.

Keywords: dysarthric speakers, spontaneous speech, speaking rate.

1. INTRODUCTION

The production of healthy native speakers is easily intelligible in most of the usual speech situations. Conversely, a characteristic of speech produced by dysarthric speakers (Ds in this paper) is the loss of intelligibility, even in quiet conditions. Nevertheless, the phonetic characteristics of this lack of intelligibility are sometimes hard to identify. Our hypothesis is that speech distortions in Ds are a multidimensional phenomenon that involves several phonetics components. Moreover, the sensitive question of the limits between pathological and healthy speech may not be simply a question of the amount of variation, but also a question of variation organisation.

Dysarthria results from damage to the central or peripheral nervous system impairing the transmission of neural messages to the muscles involved in speech production. The consequence is a deficit in articulation abilities but also at every stage involving motor activities for speech production (respiration, phonation, resonance, prosody, etc.).

Dysarthria includes a set of different pathologies such as Parkinson disease, Amyotrophic Lateral Sclerosis, Cerebellar ataxia, Multiple Sclerosis, etc. Since the pathophysiology is different for each pathology, the consequence, for speech production may also be different. In this paper, we consider two different types of dysarthria: Parkinson Disease and Amyotrophic Lateral Sclerosis (ALS).

Parkinson's disease results from basal ganglia damage that causes stiffness or slowing of movement. Articulation is thus characterized by imprecise consonants, irregular articulatory breakdowns and distorted vowels [7]. Amyotrophic lateral sclerosis is caused by upper and lower motor neuron damage. This disease is typically characterized by muscle weakness and atrophy. ALS speakers have a slower than normal speaking rate, prolonged phonemes, distorted vowels and shortened phrases [3].

Temporal organisation in speech production plays an important role in speech intelligibility. As the speech situation becomes more natural, less prepared, without specific instructions, the forms of variation become more diverse and less predictable. Indeed, for healthy speakers spontaneous speech is characterized by increased articulation rates, decreased frequency and length of silent pauses [10, 13], a phonetic reduction leading to a variation in syllabic timing [4]. For dysarthrics speakers, the motor constraints due to the pathology forces speakers to a re-organization of speech temporal parameters. This reorganisation induces also variability, leading to an irregular interval between syllables [1], a shortened speech segments [5], an increased silent pauses [8, 9] and an unexpected location of pauses [14]. These impairments can contribute to reduce intelligibility and altered naturalness of dysarthric speech [2, 12].

The aim of this study is to compare both populations in spontaneous speech in order to quantify the temporal aspects of production. To answer this question, the objective is two-fold: (1) to investigate speaking rate variations within two groups of dysarthric populations: Parkinson speakers (Ps) and Amyotrophic Lateral Sclerosis speakers (ALSs); (2)

to compare and evaluate the similarities and differences in speaking rates between both dysarthric and healthy speakers (Hs) in spontaneous speech.

2. MATERIAL AND METHODS

2.1. Corpus and speakers

Recordings are extracted from databases developed within two different projects. 8 speakers with Parkinson's Disease and 12 speakers with Amyotrophic Lateral Sclerosis disease were extracted from the DesPhoAPaDy project (ANR-08-BLAN-0125). 6 healthy speakers were extracted from the TYPALOC project (ANR-12-BSHS2-0003). Ps and ALSs were chosen according to their age and severity in the disease (0 corresponds to normality and 3 corresponds to high severity). The criterion was to avoid non-intelligible speech as well as non-affected speech (similar to normal speech) (Table 1).

Table 1: Mean, minimal and maximal values of age and severity disease for each population (ALSs, Ps, Hs).

Populations		Age	Severity disease
Ds	ALSs	66.2 <81-50>	2.02 <1.2-2.7>
	Ps	63.9 <81-48>	0.99 <0.4-1.6>
Hs		69.8 <82-63>	-

For Ds, instructions are to tell their everyday or a typical day in the hospital. For healthy speakers, instructions are to tell their professional career or personal events. Then, for both populations speech production context is a narrative situation. The main difference between healthy and dysarthric corpora is the duration of the recordings. Narration is quite long for healthy speakers (mean: 10.78 min), while Ds speak less (mean: .93 min for Ps, 1.44 min for ALSs). Most of Ds avoid speaking situations. As a consequence they don't speak very long time.

2.2. Measurements

Each audio file was manually transcribed, while word and phonetic alignment were performed by an automatic system (developed by the Laboratoire d'Informatique d'Avignon, France; see also the DesPho-APaDy Project for details of segmentation procedure, [6]). In this study, four measurements have been examined:

- Duration of Inter Pausal Units (IPU)
- Duration of words
- Number of syllables per second
- Number of words per IPU

The Inter Pausal Units (IPU) are defined as speech sequences separated by pauses larger than 250ms. All measurements are extracted from these IPU. Consequently, pauses larger than 250ms are not taken into account in the measurements. The durations of IPU are extracted from the boundary (manually labelled). The durations of words are extracted from the word boundaries (automatically labelled). Number of syllables per second is deducted from the number of vowels in the IPU. The number of words per IPU is calculated by summing all the IPU durations for each speaker's file and dividing it by the sum of words in the same file. These measurements have been extracted for each speaker.

2.2. Statistical analysis

First, for IPU durations and word durations, a mixed-effect general linear model test was performed. The dependent variable was one of the two duration measurements (i.e. duration of IPU or duration of words). "Subject" was the random effect and "population" (3 levels: Hs, Ps, ALSs) was the fixed-factor. P-values were obtained by a likelihood ratio test with Satterthwaite approximations. As there were two comparisons made within each group, a Bonferroni correction was used to adjust the alpha level (i.e. .025). Second, for other measures (number of syllables per second, number of words per IPU), a one-way ANOVA was conducted in order to test the effect of "population" (3 levels: Hs, Ps, ALSs). All post-hoc pairwise comparisons were done with Tukey post-hoc tests. Critical significance was set at $p<.05$. The data analysis was carried on using the language R ([11]).

3. RESULTS

3.1. Durations

There are significant effects of "population", one for IPU durations with χ^2 (2)= 19, p=.001 and the other for word durations with χ^2 (2)= 11, p=.005. The means for each of the four measurements are presented in Table 2.

Table 2: Means of IPU durations (sec), word durations (sec), syllables/second, words/IPU for each population (Hs, Ps, ALSs).

Mean	Hs	Ps	ALSs
Duration of IPU (sec)	2.78	1.39	2.55
Duration of words (sec)	0.27	0.23	0.41
Syllables/second	4.66	5.27	3.57
Words/IPU	9.98	5.95	6.53

As shown in Figure 1 and 2, Ps show significantly lower IPU durations than Hs by 1.33 sec (s.e=±.26) (t=5, p<.001) and than ALSs by 1.03 sec (s.e=±.25), (t=4.06, p<.0002). No significant distinction is observed between Hs and ALSs (t=-1.3, p=.2). For word durations, ALSs show significantly higher duration of words than Ps by 0.16 sec (s.e.=±.05), (t=3.4, p<.0021) and than Hs by 0.13 sec (s.e.=±.05), (t=2.5, p<.02). No significant distinction is observed between Hs and Ps (0.6, p=.53).

Figure 1: Box plot of IPU durations for each population (Hs, Ps and ALSs) (in sec). (ns: non significant, *: p<.025)

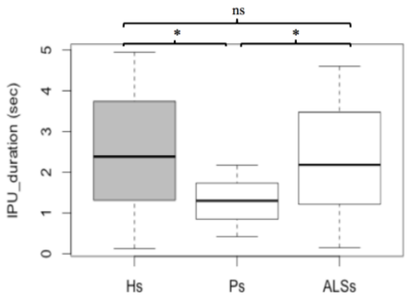
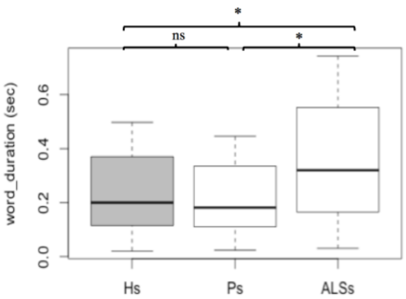


Figure 2: Box plot of words duration for each population (Hs, Ps and ALSs) (in sec). (ns: non significant, *: p<.025)



3.2. Speaking rate

Speaking rate between populations is cued by variation in number of syllables per second (F(2,23)=7.5, p=.003), and in number of words per IPU (F(2,23)=11, p<.0004). As shown in Figure 3, post-hoc comparisons show that Ps produces an higher number of syllables per second than ALSs (Ps>*ALSs, p=.003). Results for words per IPU show a significant distinction between Hs and each of dysarthric populations (figure 4): Hs>*Ps with p<.0006 and Hs>*ALSs with p<.001.

Figure 3: Box plot of syllables per second for each population (Hs, Ps and ALSs) (ns: non significant, *: p<.005)

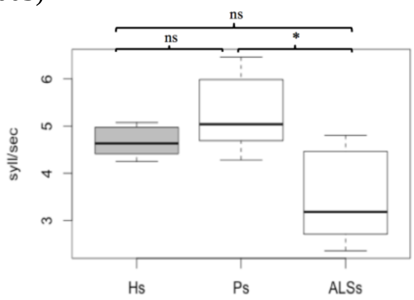
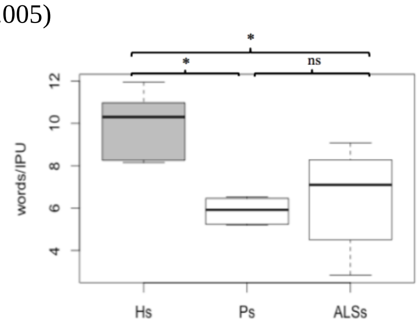


Figure 4: Box plot of words per IPU for each population (Hs, Ps and ALSs) (ns: non significant, *: p<.005)



4. DISCUSSION

As predicted in our hypothesis, the distinction between healthy and pathological speech is complex and cannot be simply represented through a unique parameter. Our results show that speech sequences are differently organized according to each population (Ps, ALSs and Hs). Indeed, the main difference between Hs and Ds is the number of words they are able to produce within an IPU. Hs clearly produces more words in IPU than Ds do. Consequently, the ability to produce a great amount of words in a sequence of speech seems to be a difficult task for Ds. Nevertheless, if this disability is common for Ps and ALSs, these two groups show clear differences concerning the other parameters. Parkinsonian productions are often characterized by a rapid and stuttered speech. The physical manifestation of this characterization is the highest number of syllables per second, while ALSs produce the lowest (Hs being intermediate). This implies a high speech rate, even if word durations for Ps do not differ from Hs ones. In fact, Ps speak faster than other populations, but they produce short IPU whereas word durations similar to Hs ones. At the opposite, ALSs speak slower, with longer words whereas IPU durations are similar to Hs ones.
In sum, the specificity of Ds is to produce very few words in an IPU, but both populations realize this in

two different ways: through a rapid speech flow for Ps and a slow speech flow for ALSs.
According to the disability of producing a lot of words in the IPU, we may hypothesize that is can be due either to: 1/ the deficit: the motor and breathing effort necessary to produce a long sequence of words requires an unreachable target for Ds; 2/ a strategy: in order to maintain sufficient intelligibility, Ds prioritize sequences with few words as to limit the production of reduced phonemes.
Finally, we have to mention that another characteristic of Ds is speaker's heterogeneity. Especially for ALSs, we observed that two speakers present opposite profiles: one with an extreme ALSs profile (very low speaking rate) and another one with an extreme Ps profile (very high speaking rate). This heterogeneity obviously blurs the characteristics of each population.

5. CONCLUSION

Our study shows that several parameters are necessary to reflect speech distortions for Ds and to quantify the distinction between healthy and dysarthric speakers. Temporal organisation of Parkinson speech is structured by small IPU, containing few words, and a high speech rate. ALSs are characterized by a slow speech, long words, and IPU durations similar to healthy speakers. The specificity of Hs, compared to Ds, is their ability to produce a lot of words within a speech sequence (IPU).
These results suggest that boundary between pathological and healthy speech is blurred and should be examined with multiparametric analyses. The specificity of dysarthric speech may not be simply represented by a "higher" variation degree but rather by a different variation structure.

Acknowledgments: this paper was funded by two grants from *The French National Research Agency* (TYPALOC: ANR-12-BSHS2-0003 and DesPhoAPaDy: ANR-08-BLAN-0125).

7. REFERENCES

[1] Boutsen, F.R., Bakker, K., Duffy, J.R. 1997. Subgroups in ataxic dysarthria. *Journal of Medical Speech-Language Pathology*, 5, 27–36

[2] Bunton, K., Kent, R. D., Kent, J. F. & Duffy, J. R. 2001. The effects of flattening fundamental frequency contours on sentence intelligibility in speakers with dysarthria. *Clinical Linguistics and Phonetics*, 15(3), 181-93.

[3] Darley, F. L., Aronson, A. E., Brown, J. R. 1975. *Motor Speech Disorders*. Philadelphia, Saunders.

[4] Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. 2006. Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, 120(2), 1040-1051.

[5] Forrest, K., Weismer, G., Turner, G.S. 1989. Kinematic, acoustic and perceptual analyses of connected speech produced by parkinsonian and normal geriatric adults. *Journal of the Acoustical Society of America,* 85, 2608-2622.

[6] Fougeron C., Crevier-Buchman L., Fredouille C., Ghio A., Meunier C., Chevrie-Muller C., Audibert N., Bonastre J.F., Simon A., Delooze C., Duez D., Gendrot C., Legou T., Lévêque N., Pillot C., Pinto S., Pouchoulin G., Robert D., Vaissiere J., Viallet F., Vincent C. 2010. Developing an acoustiphonetic characterization of dysarthric speech in French, *Proc. 7th international conference on Language Resources and Evaluation*, Malta, p. 2831-2838.

[7] Goberman, A., Elmer, L. 2005. Acoustic Analysis of Clear versus Conversational Speech in Individuals with Parkinson Disease. *Journal of Communication Disorders*, 38, 215-230.

[8] Hammen V, Yorkston K, Beukelman D. 1989. Pausal and speech duration characteristics as a function of speaking rate in normal and dysarthric individuals. In: Yorkston KM, Beukelman DR, editors. *Recent advances in clinical dysarthria*. Austin, TX: Pro-Ed, pp. 213–224

[9] Illes J, Metter EJ, Hanson WR, Iritani S. 1988. Language production in Parkinson's disease: Acoustic and linguistic considerations. *Brain and Language*, 33:146–160.

[10] Krause, J. C, Braida, L. D. 2004. Acoustic properties of naturally produced Clear speech at normal speaking rates. *Journal of the Acoustical Society of America*, 115, 362-378.

[11] R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

[12] Ramig, L. (1992). The role of phonation in speech intelligibility: a review and preliminary data from patients with Parkinson's disease. In R. D. Kent (Ed.) *Intelligibility in Speech Disorders: Theory, Measurement, and Management Amsterdam*: John Benjamin, 119-156.

[13] Uchanski, R. M., Choi, S. S., Braida, L. D., Durlach, N. I. 1996. Speaking clearly for the hard of hearing. IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research*, 39, 494–509.

[14] Yorkston, K. Beukelman, D. (1981). Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *Journal of Speech and Hearing Disorders*, 46. 296 □ 301