Biomedical applications of voice and speech processing

Pedro Gómez Vilda

Neuromorphic Processing Laboratory (NeuVox Lab)
Center for Biomedical Technology, Universidad Politécnica de Madrid
Campus de Montegancedo, s/n, 28223, Pozuelo de Alarcón, Madrid, Spain
pedro@fi.upm.es ORCID: http://orcid.org/0000-0003-3283-378X

ABSTRACT: Neurological deterioration presents different variants depending on their classification criterion, which may be their anatomic localization or their disease clinical features, although there is not a clear cut between both. Anatomically this ample group of disorders may affect the central nervous system (brain and spinal cord), or the peripheral nervous system. Clinically, the neurodegenerative disorders are classified as affecting cognitive functions or neuromotor capabilities. In the group of neurodegenerative diseases of the central nervous system, Alzheimer's disease (AD) or Fronto-Temporal Dementia (FTD) are to be found, whereas in the second group certain pathologies as Parkinson's Disease (PD), Amyotrophic Lateral Sclerosis (ALS), Huntington's Disease (HD) or *myasthenia gravis* (MG) are among the most frequent ones, although "the number of neurodegenerative diseases is currently estimated to be a few hundred" (Przedborski et al., 2003). All these pathologies produce correlates in speech at different levels: in fluency, in prosody, in articulation or in phonation. Speech technologies offer computer solutions to evaluate objectively detected anomalies in each level, adding statistical robustness, which makes them suitable for their clinical and rehabilitative application. The present issue is devoted to briefly review the characteristics of the diseases mentioned before, defining the foundations of the correlate features present in each one. Some computer solutions available in detecting and monitoring illness progress are reviewed in the contributions of different research groups working in this field.

Keywords: speech processing; neurologic diseases; computer-aided diagnosis; Alzheimer Disease; Parkinson Disease; Amyotrophic Lateral Sclerosis; Myasthenia Gravis.

RESUMEN: *Aplicaciones biomédicas del procesamiento de la voz y el habla.* – El deterioro neurológico presenta diferentes variantes, dependiendo del criterio de clasificación que se emplee, que puede ser o bien su localización anatómica, o bien sus rasgos clínicos, aunque no existe una frontera clara entre ambos. Desde el punto de vista anatómico, los trastornos de este tipo pueden afectar ya sea al sistema nervioso central (cerebro y médula espinal), ya sea al sistema nervioso periférico. Desde el punto de vista clínico, se clasifican como de tipo cognitivo o de tipo neuromotor. Se adscriben al primer grupo las patologías neurodegenerativas del sistema nervioso cortical, como la enfermedad de Alzheimer (EA) o la Demencia Frontotemporal (DFT), mientras que en el segundo grupo se enmarcan patologías como la enfermedad de Parkinson (EP), la Esclerosis Lateral Amiotrófica (ELA), la enfermedad de Huntington (EH) o la miastenia grave (MG), aunque "el número de trastornos neurodegenerativos estimado actualmente puede alcanzar unos pocos cientos" (Przedborski *et al.*, 2003). Todas estas patologías producen correlatos en el habla en diferentes niveles: en la fluidez, en la prosodia, en la articulación o en la fonación. Las Tecnologías del Habla ofrecen aplicaciones capaces de evaluar objetivamente las anomalías detectadas en cada nivel, aportando además una robustez estadística que las hace adecuadas para su uso clínico y rehabilitador. El presente número revisa brevemente las características de los tipos de enfermedades citados, define la base de los correlatos detectables en cada una de ellas, y aporta información sobre algunas aplicaciones capaces de detectar y monitorizar su evolución, a través de las diferentes contribuciones de grupos de investigación que trabajan en dichos temas.

Palabras clave: procesado del habla; trastornos neurológicos; diagnóstico asistido por computador; Alzheimer; Parkinson; Esclerosis Lateral Amiotrófica; miastenia grave.

1. INTRODUCTION

Computer science and artificial intelligence have experienced great advancements during the last two decades. New applications in the field of speech technologies have appeared as advanced signal processing and pattern recognition methods have provided notorious progress in the fields of automatic speech and speaker recognition. The applications of these methods to speech production and perception are less known, but no less relevant. Recently large companies in the field of Computer Science have realized the relevance of speech technologies in the biomedical field, and are seeking the development of applications to the field of health care and well being (Cecchi, 2017).

Speech is a very useful behavioral manifestation of the neurological, emotional and psychological state of the speaker, as it can be easily recorded, stored, transmitted and processed. Important semantic correlates may be derived from speech fluency, prosody, articulation and phonation, these being the principal domains where cognitive and neuromotor information is hidden. Speech involves the joint action of the respiratory, phonatory and articulatory systems under a strict coordination of the neuromotor system, ultimately driven by the cognitive functions associated to speech elaboration and comprehension. The message is constructed in the speech primary area (Broadman), involving also Broca's and visual areas. Once the discourse is defined, neuromotor actions are instantiated from the premotor cortex, with the concourse of the hippocampal region. Prelearned motor maps are recovered from the hippocampus (Moser et al., 2014) and sent to the sub-thalamic centers to be transformed in specific motor action sequences instructed to muscles. The proprioceptive feedback encoded in temporal and auditory sensation is used in modeling aspects as intonation contours, articulation positions or loudness control through frontal and cerebellar centers. The main muscle sets involved in the biomechanical production of speech are the diaphragm, the larynx muscles (cricoarytenoid, thyroarytenoid, and transverse and oblique laryngeal), the lingual muscles (geniohyoglossus, styloglossus, intrinsic), the velopharyngeal muscles, and the facial muscles (masseter, orbicularis, zygomatic, risorius, etc.).

Consequently, the main types of neurologic pathologies of interest affecting speech can be classified in the following main groups:

- Neuromotor. They are produced by the malfunction of neuromotor units, these comprising the secondary neurons activating the muscular structures, or by the degeneration of the muscle in itself. Typically, Parkinson Disease (PD) is produced by a lack of the neurotransmitter dopamine activating the secondary neurons. Degeneration of the secondary neuron body or axon is behind amyotrophic lateral or multiple sclerosis. Different asthenias are produced by the degeneration of the muscular unit.

- Cognitive. They are based mainly in the degeneration of primary neurons in the cortex or other subcortical regions. Alzheimer's Disease (AD) is one of the most widespread and with largest social impact. Although its ultimate cause is still a matter of study, it seems that progressive poisoning of primary neurons in the cortex lead to a deterioration of short-term (working) memory, and to a loss of communication abilities of the patient.
- Psychologic. These pathologies are less based on a physiological reason, they being associated to systemic malfunction, i.e., they do not seem to be due to neuron unit deterioration, but to the improper social behavior of many neuron subsets in their mutual interaction. Autism, depression, or psychotic diseases may be included in this large group.

The groupings within this coarse taxonomy may overlap, especially when illness progresses. For instance, PD will affect neuromuscular activity and speech, depression being another associated symptom present in many neurological diseases, and cognitive deterioration will appear with time. AD will impair the patient's communication abilities at a higher level, but will also end with neuromotor incompetence as well. The main correlates of speech, which may be affected by these pathologies are the following:

- Fluency, to be measured by the count of silent and phonated pauses between articulated segments, as well as by syllable counts. These correlates may be given as counts in a given interval (typically from seconds to hours), or by modeling the energy and fundamental tone profiles, either by dynamic nonlinear theory, or by histograms and associated probability density functions.
- Prosody is typically measured by the timely evolution of the fundamental tone in phonated segments, and by the average energy envelope of the speech stream. These correlates also admit a modeling in terms of nonlinear theory, or by probability density functions.
- Articulation may be characterized in different ways. The position of the articulation organs is directly inferred either from devices such as the Articulograph (Savariaux et al., 2017), which depend on fixtures attached to the articulation organs, or from indirect estimations derived from formant kinematics (Carmona-Duarte et al., 2016), or from the electromyographic activity recorded on the skin surface (Gómez-Vilda et al., 2017). The usual correlates estimated are the positions of the articulators in time, or their time derivatives (Yunusova et al., 2008). The Vowel Space Area or Speech Centralization Ratio (Sapir et al., 2010) are examples of static articulation correlates.
- Phonation is perhaps the best studied process, in part by virtue of the advances in organic pathology evaluation from voice. In fact, there is a myriad of

different estimates, including stability measurements as jitter, shimmer and their multiple variants, the ratios of harmonic energy to noise and vice versa, the ratios between glottal and turbulent excitation, the estimates of tremor and biomechanical unbalance, chaos derived coefficients, etc. The interested reader may check Mekyska et al. (2015), for a comprehensive review.

In general, it may be observed that these correlates can be divided into long-term and short-term ones, ranging from the characterization of fluency (long-term, depending on discourse duration, up to hours) to phonation (short term, lasting the duration of a short syllable, typically 30-50 ms).

The purpose of the present preamble is to guide the interested reader to Speech Dysfunction Modeling by Speech Technologies, introducing a group of papers in this same issue under a common framework. The authors come from a community very active in different areas of Speech Technologies during the last years, which has developed a strong cooperation on Neurologic Disease Characterization by Speech.

The set of papers included in this issue are briefly commented in relation to the use of speech technologies in healthcare and voice education.

2. PAPERS IN THE ISSUE

The first paper is entitled "Monitoring Parkinson Disease from Speech Articulation Kinematics", by the Center for Biomedical Technology, Universidad Politécnica de Madrid. Basically, it presents the Neuromechanical Model of the main articulation joint (lower jaw, tongue and lip) as a biomechanical system acting in response to different muscular forces activated by cranial nerves, which have an important influence in the first formants of phonated speech. A statistical distribution of the kinematic variables associated to this system is defined and used to estimate distances between utterances from different subjects in terms of Information Theory principles. This kinematic statistical distribution is used to evaluate the correlation between information theory based-distances and neurological test scores on a PD database of vowel utterances.

The next paper in the issue is entitled "*ALZUMERIC*: a decision support system for diagnosis and monitoring of cognitive impairment", by the Systems Engineering and Automation Department, University of the Basque Country. The purpose of this paper is twofold: on the one hand, it is intended to describe a working platform to allow the clinician to obtain speech samples and transmit them to a server for their processing in real time, offering a diagnose help to neurologists working in AD and MCI. On the other hand it discusses the features and methods to produce scoring results under statistical and security guarantees.

This work is followed by another contribution entitled "Estudio de diferentes parámetros para la detección de la Esclerosis Lateral Amiotrófica a partir del movimiento articulatorio" by the Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones, Universidad de Las Palmas de Gran Canaria, the Center for Biomedical Technology, Universidad Politécnica de Madrid, Laboratoire Scribens, Département de Génie Électrique, École Polytechnique de Montréal, and the Instituto de Medicina Molecular, Universidade de Lisboa. It is devoted to describing how the kinematic variables derived from speech articulation acoustics can be used to monitor the progress of Amyotrophic Lateral Sclerosis (ALS). This is a specific neurodegenerative disease, which presents its first symptoms as a progressive dysarthia. A method to generate the articulation patterns from formants is presented, and two features based on articulation kinematics are evaluated with respect to illness progress.

The next paper included in this issue is entitled "Lo que la voz nos cuenta acerca de los síndromes genéticos: el caso del Síndrome de Williams", by the Department of Linguistics of the Universidad Autónoma de Madrid and the Center for Biomedical Technology of Universidad Politécnica de Madrid. Its aim is to describe the phonation of children with Williams' Syndrome (WS), a rare disease of genetic origin producing different behavioral, cognitive and physiological problems in the patients who are suffering it. The research is based on the study of twelve cases of children with WS by contrasting the biomechanical features of their phonation with a large database of normative children in the same age range. The results corroborate former subjective observations by other researchers and open interesting hypotheses on the relationship between genetics and behavior.

Finally, a fifth paper is included under the title "La reeducación vocal y su evaluación a través del Procesado Digital de Señales: un estudio de caso", by the Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones, the Departamento de Didácticas Especiales of Universidad de Las Palmas de Gran Canaria, and the Conservatorio Profesional de Música de Canarias. It is devoted to the description of a study case of phonation rehabilitation based on a new method called "The Cellophane Screen". Using this simple device and under the guidance of a speech therapist, the rehabilitation of a patient suffering from *myasthenia gravis* is reported. The rehabilitation process is assessed by speech processing techniques, extracting the spectrogram, the short-time spectrum of the glottal source, and its cepstrum. It may be seen that the rehabilitation enhances the harmonic spectrum of the patient and her phonation stability, leading to a substantial improvement in her voice quality.

REFERENCES

Carmona-Duarte, C., Alonso, J. B., Díaz, M., Ferrer, M. A., Gómez-Vilda, P., & Plamondon, R. (2016). Kinematic modelling of diphthong articulation. In A. Esposito et al. (Eds.), *Recent advances in nonlinear speech processing*, vol. 48, 53-60. Cham: Springer. doi.org/10.1007/978-3-319-28109-4_6

Cecchi, G. (2017). With AI, our words will be a window into our mental health. Retrieved from https://www.ibm.com/blogs/research/2017/1/ibm-5-in-5-our-words-will-be-the-windows-to-our-mental-health/

Gómez-Vilda, P., Palacios-Alonso, D., Gómez-Rodellar, A., Ferrández-Vicente, J. M., Álvarez-Marquina, A., Martínez-Olalla, R., & Nieto-Lluis, V. (2017). Relating facial myoelectric activity to speech formants. In J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, J. Toledo Moreo & H. Adeli (Eds.), *Proceedings of IWINAC 2017 LNCS 10338, Vol. 2.* (pp. 520-530). Cham: Springer. doi.org/10.1007/978-3-319-59773-7

Mekyska, J., Janousova, E., Gomez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., … López-de-Ipiña, K. (2015). Robust and complex approach of pathological speech signal analysis. *Neurocomputing, 167,* 94-111. doi.org/10.1016/j.neucom.2015.02.085

Moser, E. I., Roudi, Y., Witter, M. P., Kentros, C., Bonhoeffer, T., & Moser, M. B. (2014). Grid cells and cortical representation. *Nature Reviews Neuroscience*, *15*(7), 466-481. doi.org/10.1038/nrn3766

Przedborski, S., Vila, M., & Jackson-Lewis, V. (2003). Neurodegeneration: What is it and where are we? *The Journal of Clinical Investigation, 111,* 3-10. doi.org/10.1172/JCI200317522

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language and Hearing Research*, *53,* 114-125. doi.org/10.1044/1092-4388(2009/08-0184)

Savariaux, C., Badin, P., Samson, A., & Gerber, S. (2017). A comparative study of the precision of Carstens and Northern Digital Instruments electromagnetic articulographs. *Journal of Speech, Language and Hearing Research, 60*, 322-340. doi.org/10.1044/2016_JSLHR-S-15-0223

Yunusova, Y., Weismer, G., Westbury, J. R., & Lindstrom, M. J. (2008). Articulatory movements during vowels in speakers with dysarthria and healthy controls. *Journal of Speech, Language, and Hearing Research*, *51*(3), 596-611. doi.org/10.1044/1092-4388(2008/043)