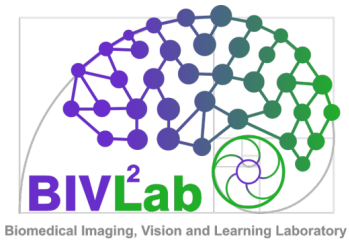

A DIGITAL BIOMARKER TO QUANTIFY PARKINSONIAN PATTERNS USING AUDIO-VISUAL DATA

BRAYAN CAMILO VALENZUELA RINCÓN

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMATICA
BUCARAMANGA

2023

A DIGITAL BIOMARKER TO QUANTIFY PARKINSONIAN PATTERNS USING AUDIO-VISUAL DATA

BRAYAN CAMILO VALENZUELA RINCÓN

Research work in partial fulfillment of the requirements for the degree of:
Magíster en Ingeniería de sistemas e informática

Advisor:
Fabio Martínez Carrillo
Ph.D in Systems and Computer Engineering
Co-Advisor:
John Edilson Arévalo Ovalle
Ph.D in Systems and Computer Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍA FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMATICA
BUCARAMANGA
2022

ACKNOWLEDGEMENTS

The author expresses his acknowledgement:

I would like to express my sincere gratitude to everyone who has significantly contributed to the completion of my master's thesis. First and foremost, I would like to thank my academic advisors Fabio Martinez and John Arevalo for their expert guidance, constant support, and valuable feedback throughout the research process. Their knowledge and dedication have been instrumental in the success of this project.

I cannot fail to mention my colleagues of the $Bivl^2ab$ research group, whose support and collaboration have been invaluable. Through our enriching discussions and debates, we have tackled complex challenges and gained new perspectives. Their contributions have been essential to the development of my work.

Furthermore, I would like to extend my gratitude to the *Escuela de ingeniería de sistemas e informática* from the *Universidad Industrial de Santander* who provided access to data, literature, and resources necessary to conduct my research. Their generosity and availability were crucial to the quality and depth of the present work.

Last but not least, I want to thank my family and friends for their unconditional love, understanding, and support throughout this academic journey. Their constant encouragement and words of affirmation gave me the strength to persevere during challenging times.

To all of you, my heartfelt thanks. Without your contribution and support, this achievement would not have been possible. Your positive influence on my academic and personal life is invaluable, and I am deeply grateful for your presence in my life.

CONTENTS

page

LIST OF FIGURES

page

LIST OF TABLES

page

LIST OF APPENDICES

page

ABSTRACT

TITLE: A DIGITAL BIOMARKER TO QUANTIFY PARKINSONIAN PATTERNS USING AUDIO-VISUAL DATA [*]

AUTHOR: BRAYAN CAMILO VALENZUELA RINCON [**]

KEYWORDS: FACIAL BRADYKINESIA, PARKINSON'S DISEASE (PD), HYPOMIMIA, SPEECH DISORDERS, MULTIMODAL LEARNING.

DESCRIPTION: Parkinson's disease is a neurodegenerative disorder that affects a large number of people worldwide. Voice and facial alterations are representative symptoms of the disease, studied manually by expert neurologists. In this context, specialized neural networks have been developed for the analysis of audio (voice) and video (face), in order to support the diagnosis of the disease. In this work, it has been proposed the integration of audio-visual deep representations, learned by two independent neural networks specialized in the analysis of facial and voice impairments. For this, a capture protocol was defined to acquire fully synchronized audiovisual sequences in a population of patients diagnosed with Parkinson's disease and control subjects. The results obtained demonstrated that the integration of neurologically synchronized information sources plays a fundamental factor in the detection of Parkinson's disease-related patterns, achieving an improvement in the diagnosis of up to 17.67% of the Area under the ROC Curve (AUC). In particular, the results suggest that the information learned by the audio network acts in a complementary manner to the video data, suggesting that simple linear integration from different sensory modalities, is enough to improve the detection and diagnosis of Parkinson's disease. This work represents a preliminary effort toward multimodal analysis of these symptoms, with the objective of enhancing both the comprehension and diagnosis of the disease.

---

[*] Research work

[**] Facultad de ingeniería fisicomecánicas
Escuela de ingeniería de sistemas e informatica. Advisor: Fabio Martínez Carrillo, Ph.D.

RESUMEN

TÍTULO: BIOMARCADOR DIGITAL PARA LA CUANTIFICACIÓN DE PATRONES PARKINSONIANOS USANDO INFORMACIÓN AUDIOVISUAL. [*]

AUTOR: BRAYAN CAMILO VALENZUELA RINCON [**]

PALABRAS CLAVE: BRADICINESIA FACIAL, ENFERMEDAD DE PARKINSON (PD), HIPOMIMIA, TRASTORNOS DEL HABLA, APRENDIZAJE MULTIMODAL.

DESCRIPCIÓN: La enfermedad de Parkinson es un trastorno neurodegenerativo que afecta a un gran número de personas en todo el mundo. Desordenes del habla y las alteraciones del movimiento facial son síntomas representativos de la enfermedad, estudiados manualmente por neurólogos expertos. En este contexto, se han desarrollado redes neuronales especializadas para el análisis de audio (voz) y vídeo (rostro), con el fin de apoyar el diagnóstico de la enfermedad. En este trabajo, proponemos la integración de representaciones profundas audiovisuales, aprendidas por dos redes neuronales independientes especializadas en el análisis de alteraciones faciales y auditivas. Para ello, se definió un protocolo de captura para adquirir secuencias audiovisuales totalmente sincronizadas en una población de pacientes diagnosticados de enfermedad de Parkinson y sujetos control. Los resultados obtenidos demostraron que la integración de fuentes de información neurológicamente sincronizadas, juega un factor fundamental en la detección de patrones relacionados con la enfermedad de Parkinson, consiguiendo una mejora en el diagnóstico de hasta el 10,22% del Área bajo la curva ROC (AUC). En particular, los resultados sugieren que la información aprendida por la red de audio actúa de forma complementaria a la información procedente del vídeo, sugiriendo que la simple integración lineal a partir de diferentes modalidades sensoriales, es suficiente para mejorar la detección y diagnóstico de la enfermedad de Parkinson. Este trabajo representa un esfuerzo preliminar hacia el análisis multimodal de estos síntomas, con el objetivo de mejorar tanto la comprensión como el diagnóstico de la enfermedad.

---

[*] Trabajo de investigación

[**] Facultad de ingeniería fisicomecánicas
Escuela de ingeniería de sistemas e informatica. Advisor: Fabio Martínez Carrillo, Ph.D.

INTRODUCTION

Parkinson's disease (PD) is the second most important neurodegenerative disorder worldwide, affecting between 2 and 3% of the population, over the age of 65 years old. Parkinson patients develop communication disorders (around 75%) related to changes in amplitude, speed and rigidity of orofacial movements [1][2]. In fact, there are reported impaired expressions related to speech and face alterations [3][4]. Regarding the face, Hypomimia, or facial Amimia, is one of the most distinctive PD signs, reporting alterations at the face's upper zone like a greater sinking of the eyelids and a reduced blinking rate. While at lower face: involuntary opening of the mouth and involuntary and spontaneous anger/smiling expressions are reported [5]. Also, regarding speech-related disorders are reported: reduction in voice amplitude over time (Hyphonia), poor voice quality (dysphonia), and reduced pitch inflection (Hypoprosody) [6]. In the literature, these PD affectations are related to neurological areas such as the motor cortex, cerebellum, basal ganglia, thalamus, and auditory pathways. Being responsible to coordinate oral production and

---

[1] Jeri A Logemann et al. "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients". In: *Journal of Speech and hearing Disorders* (1978).

[2] Cynthia M Fox and Lorraine Olson Ramig. "Vocal sound pressure level and self-perception of speech and voice in men and women with idiopathic Parkinson disease". In: *American Journal of Speech-Language Pathology* (1997).

[3] Marcia C Smith, Melissa K Smith, and Heiner Ellgring. "Spontaneous and posed facial expression in Parkinson's disease". In: *Journal of the International Neuropsychological Society* 2.5 (1996), pp. 383–391.

[4] P Madeley, AW Ellis, and RHS Mindham. "Facial expressions and Parkinson's disease". In: *Behavioural Neurology* 8.2 (1995), pp. 115–119.

[5] Jennifer L Spielman, Joan C Borod, and Lorraine O Ramig. "The effects of intensive voice treatment on facial expressiveness in Parkinson disease: preliminary data". In: *Cognitive and behavioral neurology* 16.3 (2003), pp. 177–188.

[6] Shimon Sapir, Lorraine Ramig, and Cynthia Fox. "Speech and swallowing disorders in Parkinson disease". In: *Current opinion in otolaryngology & head and neck surgery* 16.3 (2008), pp. 205–210.

facial expressions. [7,8].

Several studies have quantified orofacial PD alterations, to provide an index related to the diagnosis, but also to follow the progression of the disease [5,7,9]. These studies are mainly based on an observational test and try to coarsely stratify prosody tasks [10] and facial expression, during emotional expression tasks [11]. Despite of neurological relationship between mechanisms dedicated to voice production and facial movements, some clinical protocols and scales as unified Parkinson's disease rating scale (UPDRS) [12], analyze independently such symptoms. Besides, much of these scales remain expert dependent, introducing an inter and intra subjectivity, reported in some cases an error diagnosis above 24% [13].

Multiple computational approaches have emerged to support the quantification of these patterns. Regarding facial movements, some approaches recorded imitation or evoked facial expressions, and further captured kinematic patterns from facial tracking models. The output of such modeling is projected into classifiers to carry out a discrimination of parkinsonism patterns. Other strategies include sophisticated devices to capture depth information and construct descriptors

---

[7] Michael D McClean and Stephen M Tasko. "Association of orofacial with laryngeal and respiratory motor output during speech". In: *Experimental brain research* 146.4 (2002), pp. 481–489.

[8] Joan C Borod et al. "Parameters of emotional processing in neuropsychiatric disorders: Conceptual issues and a battery of tests". In: *Journal of communication disorders* 23.4-5 (1990), pp. 247–271.

[9] L Ricciardi et al. "Hypomimia in Parkinson's disease: an axial sign responsive to levodopa". In: *European Journal of Neurology* 27.12 (2020), pp. 2422–2429.

[10] Elliott D Ross, Robin D Thompson, and Joseph Yenkosky. "Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions". In: *Brain and language* 56.1 (1997), pp. 27–54.

[11] Paul Ekman. "Pictures of facial affect". In: *Consulting Psychologists Press* (1976).

[12] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. "The unified Parkinson's disease rating scale (UPDRS): status and recommendations". In: *Movement Disorders* 18.7 (2003), pp. 738–750.

[13] Werner Poewe et al. "Parkinson disease". In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–21.

to characterize Parkinsonism-related patterns [14][15][16]. On the other hand, voice impairments have been characterized by audio features such as vowel quality, coordination of laryngeal and supra-laryngeal activity, consonant articulation accuracy, and Weighted-Mel Frequency Cepstrum Coefficients (MFCC), using automatic classification models such as Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) [17][18][19]. These works, nonetheless, require sophisticated setup protocols and are restricted in general to handcrafted features, *i.e.,* specialized descriptors designed to recover a very particular pattern. These descriptors may be also sensitive, noisy and limited to capture large variability in recordings that enhance patterns of interest. Also, these strategies underlie in the naive hypothesis of uncorrelated information among audio and video sequences.

This work presents a retrospective study that captures paired facial and speech patterns, through deep descriptors regarding the capability to discriminate between Parkinson's and the control population. A main contribution of this study is the evaluation of the contribution to correlated orofacial patterns, and how such integration may have a major correlation with the disease. The

---

[14] Nomi Vinokurov et al. "Quantifying hypomimia in parkinson patients using a depth camera". In: *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer. 2015, pp. 63–71.

[15] Andrea Bandini et al. "Analysis of facial expressions in parkinson's disease through video-based automatic methods". In: *Journal of neuroscience methods* 281 (2017), pp. 7–20.

[16] Athina Grammatikopoulou et al. "Detecting hypomimia symptoms by selfie photo analysis: for early Parkinson disease detection". In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 2019, pp. 517–522.

[17] Michal Novotnỳ et al. "Automatic evaluation of articulatory disorders in Parkinson's disease". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.9 (2014), pp. 1366–1378.

[18] Juan Rafael Orozco-Arroyave et al. "Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson's disease". In: *Sixteenth annual conference of the international speech communication association*. 2015.

[19] Marek Wodzinski et al. "Deep learning approach to Parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification". In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 717–720.

contributions reported in this work are:

- A deep volumetric architecture that captures spatiotemporal patterns to represent facial expressions. Regarding audio patterns, a convolutional net was adjusted to recover patterns described as Mel Spectrogram inputs.
- A correlation study among paired audio-video sequences to measure the contribution of complementary and synchronized Parkinson patterns from both modalities.
- A dataset of synchronized gesture auditory sequences, recorded from a control and a Parkinson's population. Each participant was requested to perform different pronunciation tasks such as sustained vowels, phonemes, words.

1. FUNDAMENTALS

1.1. Facial bradykinesia and speech disorders

Parkinson's is a multifactorial disease expressed by different motor and non-motor patterns. Today, the literature is reported a notable correlation between Facial bradykinesia and speech disorders with the disease [9,20]. These patterns are nonetheless measured independently at different stages of the disease. We hypothesize that the study of the multimodal source of information may be crucial to better understand the development of the disease and may be promising as a digital biomarker alternative to support early diagnosis and achieve an effective following of the disease. These patterns are described as follows:

- **Facial bradykinesia**, commonly known as Hypomimia, is one of the most distinctive clinical features of Parkinson's disease, producing reduction or loss of spontaneous facial movements and emotional expression [21]. The amimia is rarely asymmetrical [20] and may be present at very early stages of the disease, revealing observable patterns such as palpebral fissures are wider (staring expression), nasolabial folds are flattened, orbicularis oculi (eyelid muscles) wrinkles are reduced and the mouth opens involuntarily [21]. Thus, in the upper face, hypomimia usually manifests as a reduction of blink rate and an alteration in the closing and opening phase of the eyelid during voluntary blinking. Regarding the lower face, usually, it is manifested problems in smiling spontaneously as well as impairment of voluntary orofacial movements (organized and coordinated movements of the different

[20] Teresa Maycas-Cepeda et al. "Hypomimia in Parkinson's Disease: What Is It Telling Us?" In: *Frontiers in Neurology* 11 (2021), p. 1775.

[21] Joseph Jankovic. "Parkinson's disease: clinical features and diagnosis". In: *Journal of neurology, neurosurgery & psychiatry* 79.4 (2008), pp. 368–376.

parts of the face and mouth) [22]. Additionally, in most patients, the mentioned symptoms derive the incapacity to identify emotions due to their reduced capacity to imitate them [9, 23].

- **Speech disorders** in PD is related to reduced loudness or tendency to reduce loudness over time (hypophonia), poor voice quality (dysphonia), little change in pitch (hypoprosody), reduced range of articulation (hypokinetic articulation), a tendency to accelerate speech articulation (rush), and hesitant or slurred speech. Collectively referred to Hypokinetic dysarthria (HKD) [6]. In addition, it is known that about 90% of patients with PD develop speech and swallowing disorders during the disease [24, 25].

1.2. Multimodal learning

The main interest in this work is to fuse speech and face gesture information into a multimodal model that allows exploiting the multi-factorial nature of PD. The multimodal characterization is referred then to as the harmonious integration of information from textual, visual, auditory, and other sources [26]. Particularly, multimodal learning aims to build models with the ability to process and relate information from a variety of domains, capturing correspondences between modalities of different forms and dimensions, projecting them in a common representation space

---

[22] Matteo Bologna et al. "Facial bradykinesia". In: *Journal of Neurology, Neurosurgery & Psychiatry* 84.6 (2013), pp. 681–685.

[23] M Katsikitis and I Pilowsky. "A controlled quantitative study of facial expression in Parkinson's disease and depression." In: *Journal of Nervous and Mental Disease* (1991).

[24] Lorraine O Ramig, Cynthia Fox, and Shimon Sapir. "Speech treatment for Parkinson's disease". In: *Expert Review of Neurotherapeutics* 8.2 (2008), pp. 297–309.

[25] Shimon Sapir, Lorraine Olson Ramig, and Cynthia Fox. "Voice, speech, and swallowing disorders". In: *Handbook of Parkinson's Disease*. CRC Press, 2007, pp. 469–492.

[26] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. "Multimodal machine learning: A survey and taxonomy". In: *IEEE transactions on pattern analysis and machine intelligence* 41.2 (2018), pp. 423–443.

[26,27].

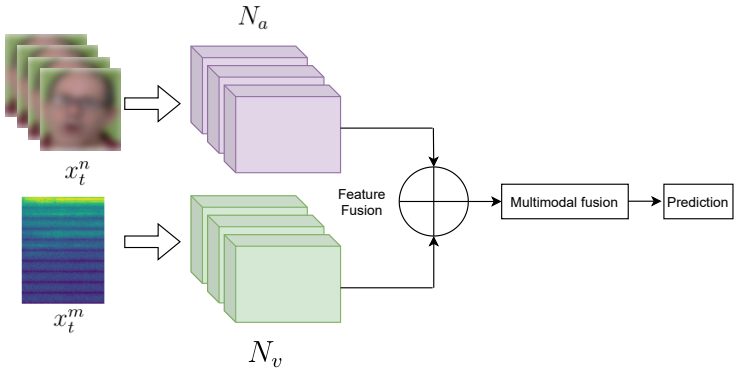

Figure 1. Illustration of a multimodal model to fuse audiovisual feature maps into a common representation scheme, using the feature fusion operation. To finally use this joint representation in a classification task

Following the formulation proposed by Bayoudh in [27], input sequences are defined for different modalities such as $X_a = \{x_1^n, , , , , x_T^n\}$ and $X_v = \{x_1^m, , , , , x_T^m\}$, where x_t^n and x_t^m are features vectors $n-$ and $m-$dimensional for the modalities a and v, respectively. Given a set of labels $Z = \{Z^1, , , , , Z^T\}$, an M multimodal learning model is trained to map X_a , X_v in the same categorical set Z. To perform the analysis of gestural auditory sequences, two unimodal architectures are constructed for X_a and X_v, denoted as N_a y N_v, respectively, where $N_a : X_a \rightarrow Y$, $N_v : X_v \rightarrow Y$ y $M = N_a \oplus N_v$, being Y the label of the class predicted as the output of the constructed M model and $\oplus$ indicates the feature fusion operation. The approach described can be seen in the Figure 1. This multimodal generalization may be realized using a variety of fusion strategies, each of which is defined by how the information learnt by models N_a and N_v is combined. The most common techniques are those that integrate the learned representations for the X_a and X_v models, as well as those that integrate the predictions of models N_a and N_v given the X_a and X_v inputs, illustrated in Figures 2, 3.

---

[27] Khaled Bayoudh et al. "A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets". In: *The Visual Computer* (2021), pp. 1–32.

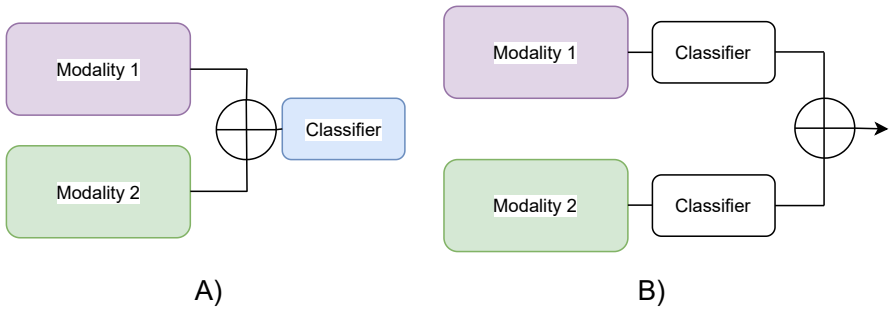

Figure 2. A) Early fusion: This technique refers to the integration of low-level features extracted from N_a and N_v models. These are merged from operators such as concatenation or correlation before the classification stage is performed. B) Late fusion: It consists of performing the classification of each one of the features extracted from N_a and N_v independently, to finally integrate these classification results to get the final classification result.

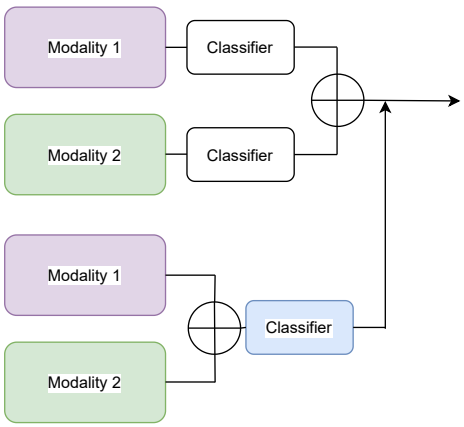

Figure 3. Hybrid fusion: Hybrid fusion method consists in combining the two methods of early and late fusion before performing the final classification.

2. PREVIOUS WORKS

2.1. Parkinson disease from audio analysis

Some of the initial approaches to speech analysis focused on the study of articulatory deficits rather than on the assessment of dysphonia. This fact was established in order to model the difficulty that some PD patients have evidenced in starting and stopping fold movements. Thus, Michal Novotný et al [17] designed an automatic approach to estimate articulatory speech deficits in Parkinson's disease, based on rapid repetition of the syllables /pa/-/ta/-/ta/-/ka/. Multiple features were analyzed to describe articulatory aspects of speech and PD patients showed poor articulatory performance in the investigated speech dimensions. In a later study, Orozco-Arroyave et al [18] analyzed voice recordings from written texts and monologues in 3 languages. Such approach models the audio signals during voiced and unvoiced segments, evidencing the limitations of patients to stop/start vocal fold movement during the production of a continuous conversation (monologue).

Subsequent works have focused on the use of CNN representations, adjusted on datasets that include phonation and articulation-related vocal features from tasks such as word pronunciation, phonemes, and sustained vowels [28,19,29,30]. These approaches have evidenced a remarked gain of CNNs and transfer learning strategies on the analysis and automatic classification of audio sequences related to Parkinson's disease. Also, one of the main representations used to encode

---

[28] Hakan Gunduz. "Deep learning-based Parkinson's disease classification using vocal feature sets". In: *IEEE Access* 7 (2019), pp. 115540–115551.

[29] Juan Camilo Vásquez-Correa et al. "Multimodal assessment of Parkinson's disease: a deep learning approach". In: *IEEE journal of biomedical and health informatics* 23.4 (2018), pp. 1618–1630.

[30] Juan Camilo Vásquez-Correa et al. "Convolutional neural networks and a transfer learning strategy to classify Parkinson's disease from speech in three different languages". In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 24th Iberoamerican Congress, CIARP 2019, Havana, Cuba, October 28-31, 2019, Proceedings 24*. Springer. 2019, pp. 697–706.

information from audio sequences are frequency spectrograms and Mel Spectrograms. These representations capture the information in the frequency domain as a function of the temporal information, being a two-dimensional representation that can be interpreted by a CNN.
Similarly, in the study conducted by Vazques Correa et al. [29], speech signal characterization was complemented with unsynchronized information from handwriting and gait. Such approach includes a similar approximation that proposed in [18]. In such case the speech is modeled to capture discriminatory patterns during the transition between vocal and non-vocal segments. The characterization of such transitions is also complemented with handwriting start-stop transitions, and even also with start-stop gait transitions. Despite the results of the aforementioned approaches, disease analysis is currently limited to the analysis of non-synchronized and isolated information sources. This experimental setup may impact in the representation of speech, the discriminatory character and may impact in the generalization of the proposed approach.

2.2. Parkinson disease from video analysis

The level of face expressiveness, a main Parkinsonism indicator, has been supported by computational approaches that take as input representations of the facial movement to identify the capability to express emotions such as sadness, anger, and disgust, as well their ability to mimic them. A seminal approach was introduced by Katsikitis et al. [31], developing a model to highlight important manual points, marked over facial expression while the patients smiling. Founding statistical difference between Parkinson, and a control groups. Also, Wu et al. [32] introduced a study to recover amusement, sadness, anger, disgust, surprise, and fear stimuli, recorded using facial electromyography and electrocardiogram signals. These signals were used

---

[31] Mary Katsikitis and I Pilowsky. "A study of facial expression in Parkinson's disease using a novel microcomputer-based method." In: *Journal of Neurology, Neurosurgery & Psychiatry* 51.3 (1988), pp. 362–366.

[32] Peng Wu et al. "Objectifying facial expressivity assessment of Parkinson's patients: preliminary study". In: *Computational and mathematical methods in medicine* 2014 (2014).

as input to train an SVM classifier. Capturing the differences in facial expressivity of patients with PD and control subjects. In same line, Bandini *et al.* [15] conducted a study over different facial expressions (happiness, anger, disgust and sadness) to automatically compute landmarks as facial displacements regarding the neutral expression (serious). In that work was validated and quantified the hypothesis that facial movements performed by control subjects have a greater displacement than those performed by Parkinson's patients.

Recently, computational approaches have been addressed to improve the characterization of hypomimia related patterns. Using convolutional neural networks (CNN) adjusted from facial movement approximations such as points of interest, displacement quantification, among other [33,34]. One of the most recent works proposes a multimodal analysis of facial symptoms, based on the integration of static (frame) and dynamic (video) features, estimated from deep representations and face landmarks [35,36]. A main limitation of these approaches is the use of face landmarks, which result sensitive to fast movements and environmental variables such as lighting. In addition, several of these methods make use of facial expressions (natural or simulated) to identify alterations in facial movement, without include any linguistic exercise that could provide complementary information.

---

[33] Avner Abrami et al. "Automated Computer Vision Assessment of Hypomimia in Parkinson Disease: Proof-of-Principle Pilot Study". In: *Journal of Medical Internet Research* 23.2 (2021), e21037.

[34] Bo Jin et al. "Diagnosing Parkinson disease through facial expression recognition: video analysis". In: *Journal of medical Internet research* 22.7 (2020), e18697.

[35] Luis F Gomez et al. "Improving parkinson detection using dynamic features from evoked expressions in video". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 1562–1570.

[36] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

3. RESEARCH PROBLEM

Parkinson's disease causes motor impairments, whose symptoms are associated with rigidity, tremor, and postural instability. The main limitation with diagnosis and treatment planning is that motor symptoms only appear when the majority of dopaminergic neurons have already been lost. Moreover, the analysis of disease is principally carried out by observational protocols, which induce an intrinsic expert variability.

The computational strategies may be key to support diagnosis at early stages, allowing, among others, to plan treatment and properly follow the progression of each patient. Today, there exists evidence that correlates hypomimia with orofacial and axial symptoms with PD. Nonetheless, the description of such PD patterns remains limited to isolated quantification of observational sources (speech and facial expressions). The study of multimodal approaches that integrates multiple motor disabilities recognize PD as a multifactorial disease. These approaches may be key to complement current standards, allowing to find hidden relationships of the disease with motor disabilities, captured in audio-visual data.

Research Question

How to design a computational strategy to integrate audiovisual information, related to the facial and auditory symptoms of Parkinson's disease?

4. OBJECTIVES

General Objective

- To develop a multimodal deep representation to integrate gesture-auditory sequences to describe patterns related with Parkinson's disease.

Specific Objectives

- To capture a set of synchronized audio (speech) and video (face gestural) sequences of Parkinson's patients and control subjects.
- To build a deep representation for each audio and video modality that project sources in a common representation space.
- To develop a multimodal strategy that integrate audio-visual data to find motor correlations with Parkinson's disease.
- To validate multimodal approach, as potential digital biomarker, with respect to the capability to discriminate PD regarding a control population.

5. DATASET

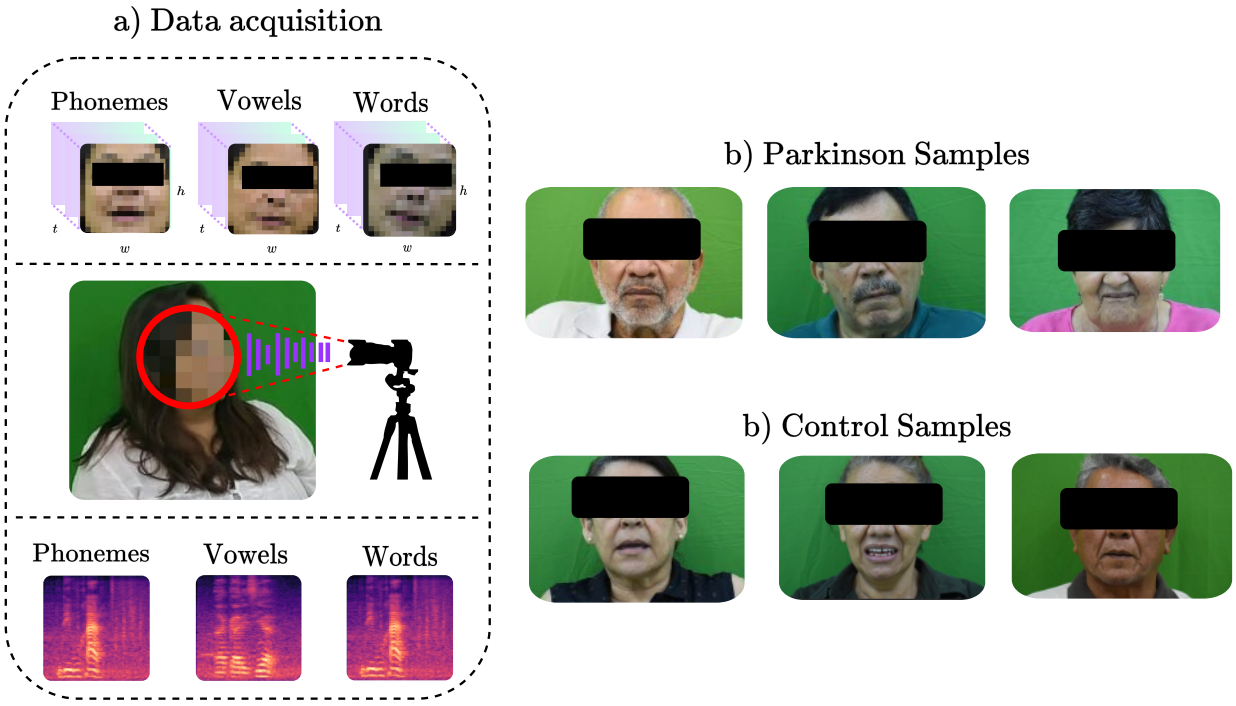

Figure 4. Graphical representation of the data capture process, as well as the population captured for the present work.

A retrospective study was conducted to validate the capability of audio and video descriptors to discriminate and model Parkinsonism disorders. During the protocol, a total of 14 participants were invited to develop gestural and speaking exercises. In general, our study population included 7 patients diagnosed with PD (average age of 65 ± 4) and 7 control subjects (average age of 61 ± 3), with a gender distribution of 4 males and 3 females for the Parkinson's population and 2 males and 5 females for the control population. In addition, it is known that 2 of 7 patients were off medication at the moment of the data acquisition, the remaining 5 patients were medicated with Levodopa. Each one gave informed consent before participating in the study, which was approved by the ethics committee of the Universidad Industrial de Santander. To the

best of our knowledge, this dataset represents the first effort to recover synpatients diagnosed with PDchronized audio and video modalities from a Parkinson and control population. Details of recorded patterns are described as follows.

Recording settings For each participant was recorded a video sequence with the synchronized audio signal, while the participants performed the oral production of multiple exercises, including sustained vowel pronunciation, phonemes, and words, like in [18]. All participants considered in this study are Colombian native Spanish speakers.
Audio and video recordings were conducted in the same environment, for Parkinson and control populations, to preserve similar background noise and lighting variables. For that a Nikon D3500 digital camera with integrated monaural microphone was used. The video acquisition was carried out at a resolution of 1080p at 60 fps, focused on the face region. While audio captures were performed at a sampling rate of 48000 kHz. All patients included in the study were diagnosed by an expert neurologist and partially labeled according to the HY scale [37].

Orofacial recorded exercises This study includes different phonation and articulation tasks, which are summarized as follows:

- To record associated patterns with phonation, three repetitions of the 5 vowels were carried out in the study.
- To articulation ability was accessed from the pronunciation of phonemes that force the movement of muscles affected for PD. The phonemes included in the study are: *(pa, pe, ta, ka): pa-ta-ka, pa-ka-ta, pe-ta-ka.*
- The repetition of three groups of words were also included to enrich phonation and articulatory analysis. The first group include the words of the phonological Spanish: *petaca,*

[37] Roongroj Bhidayasiri et al. "Parkinson's disease: Hoehn and Yahr scale". In: *Movement Disorders: A Video Atlas: A Video Atlas* (2012), pp. 4–5.

bodega, pato, apto, campana, presa y plato. The second group include motor verbs, such as: *acariciar, aplaudir, agarrar y dibujar*. The third group nouns of well kwnon objects, such as *barco, bosque, ciudad, establo, hospital, luna y montaña.*

6. PROPOSED APPROACH

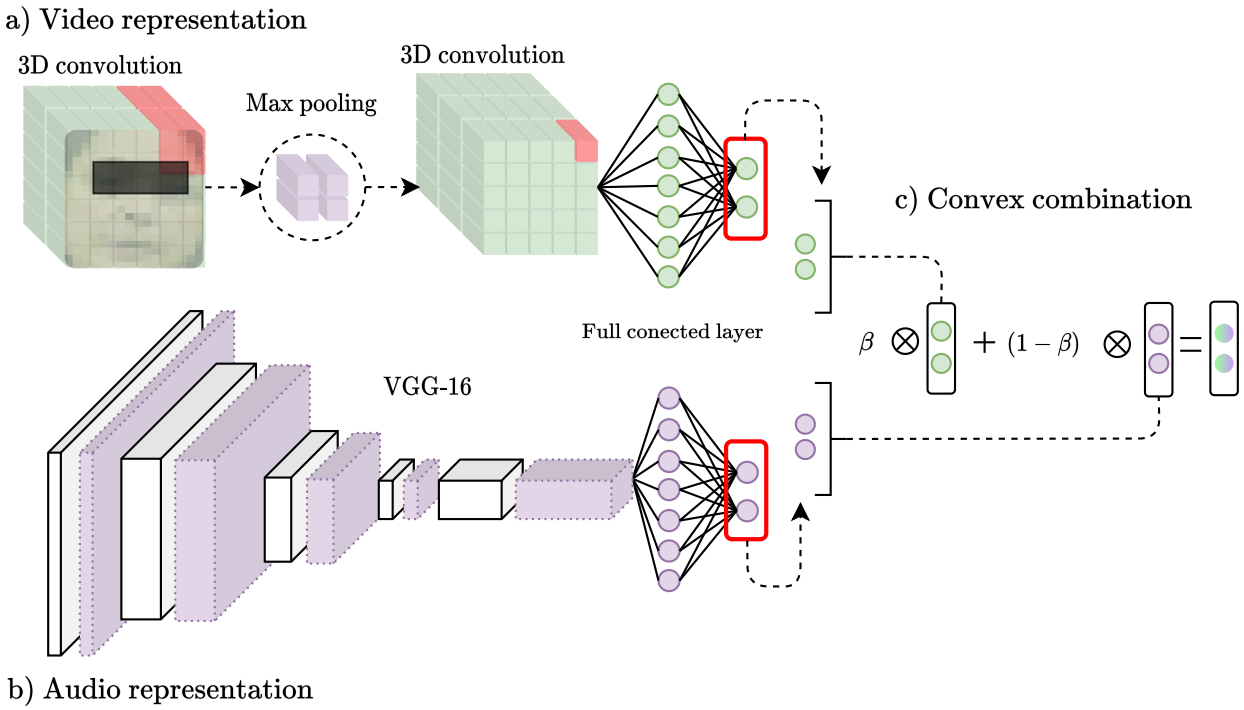

Figure 5. Pipeline of the proposed approach. a) and b) represent the convolutional architectures Resnet-50 and I3D used to encode the facial and auditory motion patterns in a low dimensional representation. c) Audiovisual integration process to improve the representation of auditory gestural motion.

We propose to codify voice and face movement signals coded as deep representations, to find correspondences with PD patterns related with speech impairments and hypomimia patterns, respectively. The resultant embedded descriptors are then integrated to analyze the correlation of audiovisual patterns to improve the diagnosis of PD. The general pipeline is illustrated in Figure 5. The complete content of this section has been accepted in the 44th Annual International

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [38], and a journal manuscript is under construction for the Computer Methods and Programs in Biomedicine.

6.1. Video representation

Modeling and measuring spatial and temporal patterns related to facial movement are crucial to characterize Hypomimia. Here, a comprehensive representation of facial motion was achieved by using a custom implementation of a 3D convolutional neuronal network (3D ConvNet). As previous studies have shown, hypomimia has a significant impact on the facial expression of PD patients, causing their facial movements to become small in amplitude and slow in speed [22]. For this reason, the proposed video representation is focused on analyzing the motion present in a facial video sequence. As has been quantified in a variety of works [39][40][41], the use of 3D convNets have demonstrated better performance in video classification tasks compared to 2D convolutional networks, due to their ability to model complex spatiotemporal relationships, extract relevant features more effectively, and leverage temporal information for greater generalization. Therefore, this proposal is focused on modeling altered movement patterns in facial video sequences. For this, we make use of 3D convolutions that allow modeling facial movements over time, without losing the information of the patient's gestures, present in the spatial domain of the video.

The proposed 3D ConvNet receives as input video clips of patients while developing some lin-

---

[38] Brayan Valenzuela et al. "A Spatio-Temporal Hypomimic Deep Descriptor to Discriminate Parkinsonian Patients". In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2022, pp. 4192–4195.

[39] Christian Szegedy et al. "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

[40] Du Tran et al. "A closer look at spatiotemporal convolutions for action recognition". In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2018, pp. 6450–6459.

[41] Du Tran et al. "Learning spatiotemporal features with 3d convolutional networks". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4489–4497.

guistic exercises, *i.e.,* pronunciation of vowels, phonemes, and words. This volumetric input is decomposed in 3D local convolutional filters, that learn not only the geometric structure of the image but also the local kinematic components of facial gestures. Thus, in a progressive hierarchical representation, the spatiotemporal primitives are adjusted to recover the main volumetric patterns involved in language exercises, which maximize differences between PD and control patients.

One of the main advantages is the fact that our representation does not require invasive methods such as facial electromyography, manually defined features such as facial landmarks, or the measurement of distances to compare displacements. This contributes to the generalization of our approach, facilitating its use in new data sets, in which multiple variations in facial motion are contemplated, resulting from tasks such as the production of emotional facial expressions and the pronunciation of linguistic exercises.

6.2. Audio representation

One of the main strategies used in literature to process and enhance patterns in audio sequences are Mel Spectrograms [42]. This audio representation has demonstrated high performance in automatic learning tasks designed to identify representative aspects of the human voice related to a particular task. In such a case, we project the audio signal (represented as Mel Spectrograms) to a convolutional net to analyze abnormal audio patterns in PD patients, following articulatory and phonation tasks [30, 19].

Specifically, the audio signals are commonly acquired as one-dimensional vectors, denoted as $A_{\{t,Hz\}}$, where t represents the duration of the audio and Hz is the sampling frequency. A more robust analysis of such signals is carried out on the frequency domain, and therefore the original input is divided into ordered and overlapping audio segments, denoted as

[42] Zhijing Xu et al. "Voiceprint recognition of Parkinson patients based on deep learning". In: *arXiv preprint arXiv:1812.06613* (2018).

$\{A_0, A_1, ..., A_{n-1}, A_n\}$ and then recovered the respective frequency spectrogram, as the respective maps: $A_{Spec} = Concat(\{A_0, A_1, ..., A_{n-1}, A_n\})$. Each spectrogram uses the Mel scale, to enhance the small frequency variations in the human voice. The result of such processing is a 2D matrix representation denoted as $MSpec = Mscale(Spec)$. This frequency representation $MSpec$ is used to adjust deep convolutional architectures.

In this work, we implemented three different convolutional architectures, two of them inspired on standard baseline approaches for classification. Basically, these nets decompose spectrograms regarding to a set of learned filters, that try to enhance local patterns according to a discrimination rule. In this case, net was adjusted according to the separation between Parkinson's and control patients. The set of convolutional filters operates among layers outputs, progressively, allowing embedded information into a single hidden vector, which in turn is projected to a classification task. This modeling has been widely validated in the literature, and this work, was dedicated to selecting the best net to represent audio information.

6.3. Audio-Visual Convex combination

In the clinical field, the nature and level of coupling between the orofacial, respiratory and laryngeal systems is an important topic of interest, associated with the understanding of disease [7, 8, 9]. Thus, previous work suggests increased connectivity between neural systems linking the jaw area (among other facial regions) with the laryngeal and respiratory systems. This proposes that the neural circuits involved in motor control of the face for speech, have relatively strong links to systems related to oral production.

The main purpose of this work is the study how voice audio and face video descriptors can be integrated to form orofacial representation that can robustly discriminate Parkinsonism patterns. Also, from this study, we are interested in evaluating the correlation correspondence and integrating such information sources. In such a sense, we implemented a dedicated deep representation that exploits raw signals and project information into an embedding space. Hence, each representation by itself each able to score a PD probability according to the learned rep-

resentation.

Hence, following a multimodal fusion scheme [26], a late convex fusion is herein introduced to integrate the audio and video information. In this work we are particularly interested in studying the contribution of hypomimia and voice impairments. For such reason, we decide to adopt a late fusion scheme along with a linear combination of characteristics. The deep video representation $V_{n,2}$ and the deep audio descriptor $A_{n,2}$ are independent Parkinson probabilities coded using CNN, for a batch of n samples. These probabilities are then integrated as the linear combination: $D_{integrated} = V_{n,2} \cdot \beta + A_{n,2} \cdot (1-\beta)$, where $\beta \in \mathbb{R}^+$. This linear combination serves as a refinement process for the estimated predictions in each modality independently, improving the overall behavior of the model. In such cases, the β value measures the importance of each modality, allowing bring importance to each exercise, according to validated orofacial findings.

7. EXPERIMENTAL SETUP

The audio and video patterns were here coded from independent convolutional architectures. For audio recognition, the recordings were coded as Mel Spectograms using 40 Mel-frequency cepstral coefficients, sampling rate of 48000 kHz and a Discrete cosine transform of type 2 retrieving spatial maps of (73×40). Then, a deep representation was built from a Resnet-50 [43] a VGG-16 [36] and a custom convolutional neuronal network (2D CNN). These architectures were trained and adjusted to discriminate between control and Parkinson population from scratch. For the Resnet-50 and VGG-16 architectures, the final classification layer was modified to solve binary classification. The custom convolutional net consists of 3 blocks composed of a 2D convolutional layer (kernel size = (3,3), filters = 64, activation=Relu), followed by a 2D max pooling operation with a filter size of (3,3). The three architectures were adjusted following a classical binary cross-entropy, Adam optimizer, early stopping and 25 epochs using a learning rate of 1×10^{-5}.

To recover Hypomimia patterns from video recordings, we implemented a compact 3D ConvNet, this consists of 2 blocks of a 3D convolutional layer (kernel size = (3,3,3), filters = 64, activation=Relu), followed by a 3D max pooling operation with a filter size of (2,2,2). The rest of the architecture is composed of dense layers to perform binary classification. We also implement a volumetric representation from an inflated 3D (I3D) architecture that take advantage of 3D convolutional kernels to compute local volumetric patterns over input video-sequences. This net take initially 2D weights from an image classification problem. Then, such weights are inflated to temporal dimension by concatenating the 2d learned filters. The architecture consist of inception layers that fully exploit spatiotemporal representation at different scales.

As a baseline comparison, we implemented a 2D convolutional net, using as reference the Resnet-

---

[43] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

50 architecture. This capture face information from isolated frames, taking into account our interest to correlate spatiotemporal face motion patterns with PD. All network configurations mentioned were adjusted using binary cross-entropy, Adam optimizer, early-stopping and 25 epochs with a learning of 1×10^{-5}.

The validation of whole deep schemes (for audio, video and the convex fusion) following the leave-one-out cross-validation scheme. In such case, at each experiment, the audio and video samples for one patient were left out, while the rest of samples were used for training both respective nets. For each experiment, the capability of the proposed approach to classify Parkinson associated patterns were accessed from the metrics of precision, recall, f1-score, accuracy and Area under the ROC Curve (AUC). As reference, each prediction was compared with the diagnosis of an expert neurologist. In this research we conduct several experiments to establish the capability of audio-visual representation. Each experiment was split according to exercises that carry out the patients, *i.e.*,, vowels, phonemes, and words. The results of these experiments are subdivided into exercises that are analyzed independently for each modality. During validation, video and audio experiments were conducted independently, which thereafter are fused to provide a better support for disease diagnosis.

A confounding variables analysis was conducted to comprehensively assess and know the impact of external variables. Patient age and gender were of particular interest among the variables examined, as they had the potential to introduce distortion in the relationship between the independent variables (Parkinson's and Control) and the dependent variables of interest (Audio or video signals). Two logistic regression fits were carried out, denoted as $Y_{pred} = B_0 + B_{1-Adjusted}Y_{true} + B_2C_1 + B_3C_2$ and $Y_{pred} = B_0 + B_{1-Crude}Y_{true}$, with the inclusion of the relevant confounding variables (denoted as C_1 and C_2). By computing the absolute percentage error between the values $B_{1-Adjusted}$ and $B_{1-Crude}$ as $err\% = |B_{1-Crude} - B_{1-Adjusted}|/B_{1-Crude}$. The achieved results ensure the validity and reliability of our findings.

8. EVALUATION AND RESULTS

In this work was conducted a study to explore audio and video descriptors with respect to the capability to discriminate Parkinson disease. In such sense, deep dedicated representations were adjusted to code audio and video signals, independently. Once the best configurations (architectures) in each modality were identified, the respective output probabilities in each modality were merged following our convex fusion approach. The validation of each stage is described as follows.

8.1. Parkinson classification from audio deep representations

To represent speech-related patterns, Table 1 summarizes the results achieved for the three different convolutional architectures, and regarding each orofacial exercise. As observed in results, the VGG-16 backbone properly recovers speech patterns from frequency Mel Spectrograms, achieving coherent result across different exercises. It should be noted a better capability to discriminate between control and Parkinson patients (AUC = 73.55) when patients pronounce vowels. Such case may be associated to long and repetitive audios that allows to capture essential frequency features from Mel Spectograms. Thus, both architectures, the baseline convolutional architecture and Resnet-50 achieving an average AUC of only 58.81% and 55.06%, resulting in a limited audio representation to discriminate between both labeled patterns.

The VGG-16 network is the best architecture for our audio representation, achieving an average AUC of 69.72%. For Phonemes and words that recover more structured and complex language, the VGG-16 remains competitive in the discrimination task. Hence, the following multimodal integration are carried out from audio representation coded from VGG-16.

Network	*Exercise*	*Metrics*				
		Prec	*Rec*	*F1*	*Acc*	*AUC*
2D-conv net	*Phonemes*	53,85	55,56	54,69	53,97	$56,03 \pm 9,94$
	Vowels	63,64	53,33	58,03	61,43	$63,00 \pm 7,4$
	Words	54,55	36,51	43,74	53,04	$57,42 \pm 3,93$
Resnet-50	*Phonemes*	62,26	52,38	56,90	60,32	$57,07 \pm 10,15$
	Vowels	45,63	44,76	45,19	45,71	$46,80 \pm 8$
	Words	66,14	43,92	52,78	60,71	$61,31 \pm 4,1$
VGG-16	*Phonemes*	62,32	68,25	65,15	63,49	$\mathbf{67,72 \pm 9,2}$
	Vowels	72,34	64,76	68,34	70,00	$\mathbf{73,55 \pm 7,03}$
	Words	62,98	69,31	65,99	64,29	$\mathbf{67,91 \pm 3,74}$

Table 1. Ablation study from audio patterns. The use of the VGG-16 and Resnet-50 networks is considered when these are trained from scratch.

8.2. Parkinson classification from a video representation

Regarding the capability to recover spatiotemporal patterns from patients, observed during the develop of language exercises, three different architectures were here included in the validation. Firstly, we included a typical 2D convolutional architecture, from a Resnet-50 architecture, that try to code gestural patterns from isolated frame observations. In such case, baseline architectures argue that gestural parkinsonian patterns may discriminative from isolated and static observations, during the develop of some linguistic and emotional exercises.

Regarding volumetric representations, we conduct experiments using two deep representations, a custom 3D-Conv net, and an Inflated-3D (I3D). We expect not only capture gestural abnormalities in Parkinson patients but also recover movement disorders during language exercises. In such case, these architectures should be more informative about abnormalities related with PD that going to improve the characterization of video-sequences.

Table 2 summarizes the results achieved for the three architectures in the three different exercises and under the discrimination task between Parkinson and control patients. It should be noted that 2D convolutional net is able to recover gestural patterns, during phonemes (AUC = 74.25%) and vowels (AUC = 63.82%) exercises, achieving coherent discrimination between

Network	*Exercise*	*Metrics*				
		Prec	*Rec*	*F1*	*Acc*	*AUC*
2D-baseline	*Phonemes*	87,50	44,44	58,95	69,05	74,25
	Vowels	57,73	53,33	55,45	57,14	63,82
	Words	54,55	36,51	43,74	53,04	57,42
Inflated-3D	*Phonemes*	62,26	52,38	56,90	60,32	57,07
	Vowels	51,72	42,86	46,88	51,43	45,23
	Words	56,70	53,48	55,04	56,32	60,05
3D-conv net	*Phonemes*	80,00	50,79	62,14	69,05	**82,59**
	Vowels	66,67	51,43	58,06	62,86	**70,46**
	Words	72,00	52,38	60,64	66,01	**79,53**

Table 2. Results of the 3 architectures proposed in the video modality, for the exercises of vowels, phonemes and words. Where the best result is estimated using the 3D conv-net network.

two populations. In such case, both phonemes and vowels, are repetitive that requires few gesture configurations, resulting in a sufficient representation of the 2D information. However, the pronunciation of words, with more complex linguistic structures is limited for this configuration. In the case of I3D model, the results are limited in the whole configuration. This fact may be associated to the parameters to be adjusted (around 27 million) and regarding the total dataset considered in this work. Contrary, the 3D convolutional architecture achieves a superior performance to discriminate Parkinson from the control population in whole three considered exercises. In such a case, the 3D-CNN recover gestural but also temporal information, coded into a compact representation, to discriminate abnormal patterns. Overall, our findings suggest that the choice of architecture plays a crucial role in determining the performance of facial characterization, and 3D architectures should be considered as a viable option for achieving improved results. For the next multimodal validations, we consider visual patterns only coded using 3D-CNN.

8.3. Multimodal convex fusion

One of the main goals of this work was to explore audio-visual integrations to validate the capability of discrimination between Parkinson´s and control populations. For doing so, probabilities responses are taken. For audio was selected the responses brought by the VGG-16, were while for video were used the responses from a 3D CNN convolutional net.

The integration of both modalities follows a convex rule, where modality importance was validated at different β values. Figure 6 shows the achieved results for independent modalities (red dotted line), and the results achieved for the linear integration for each β values. Experiments were carried out for each considered exercises, *i.e.*, phonemes, words and vowels.

To integrate modalities into a Leave one patient out cross-validation scheme, for each fold, we run the experiment with the same patients splits for both: audio and visual representations. Once adjusted each modality model, we integrate output probabilities for each patient, taking into account the beta weight. The average results are then reported for each fold and in general for whole considered experiments.

As expected, better integration is a direct response to better audio/visual representations. Thus, we estimated the best features integration in the phonemes exercises with an AUC of 85,39% for a β of 0.5. However, as shown in Figure 6, the findings indicated that for any β value, the convex fusion achieves better results than the analysis for independent modalities. This observation may corroborate the clinical hypothesis that patterns of facial movement and speech disorders act in a complementary way [45], further highlighting the utility of the proposed approach.

Figure 8 reports a more detailed analysis of patient prediction for each modality using phonemes, for the best convex combination ($\beta = 0.5$). In general, it is observed that convex integration better support Parkinson´s disease classification but also resolves with major confidence the characterization of each patient. For instance, video representation of the patients $\{P0, P1, P3\}$ result in the wrong classification, while for audio there exists high confidence with the disease. In such cases, the multi-modal integration achieves a more robust prediction, being only $P0$ a false negative, being this result, a consequence associated with the patient's medication, since

in our data set all patients except for $P0$ and $P2$ were off medication.

Table 4 reports the complete analysis for different exercises considering the independent and fused deep representations. As expected, is evident the gain of combining audio and visual patterns for different thresholds of relevance. Larger differences are observed in phonemes excercise, where the multi-modal approach has a gain of 17,67% and 2,8% regarding the audio and visual information, treated independently.

Additionally, we applied DeLong's statistical test to determine differences from the independent audio and video representations. The results, shown in Table 3, revealed a *p-value* less than the 5% significance level, thereby rejecting the null hypothesis that our auditory gestural representation is the same as the audio and video representations.

Exercises	Phonemes		Vowels		Words	
Models	Video vs Convex	Audio vs Convex	Video vs Convex	Audio vs Convex	Video vs Convex	Audio vs Convex
p-values	0.0478	$9,256 \times 10^{-6}$	0.0063	0.0012	0.039	3.638×10^{-15}

Table 3. *p-values* estimated between the independent modalities and our joint representation using the DeLong statistical test.

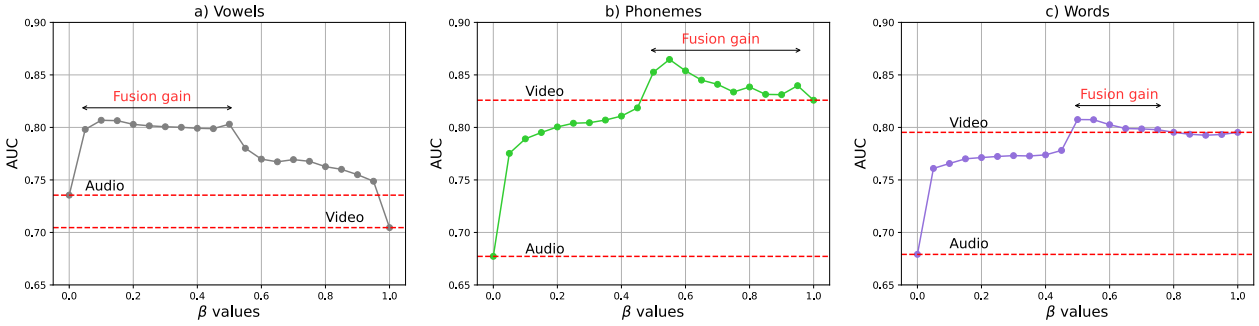

Figure 6. Iterative search for the optimal β parameters for the vowel, phoneme, and word exercises, with respect to the AUC metric. This process was performed on the best-estimated results for each modality.

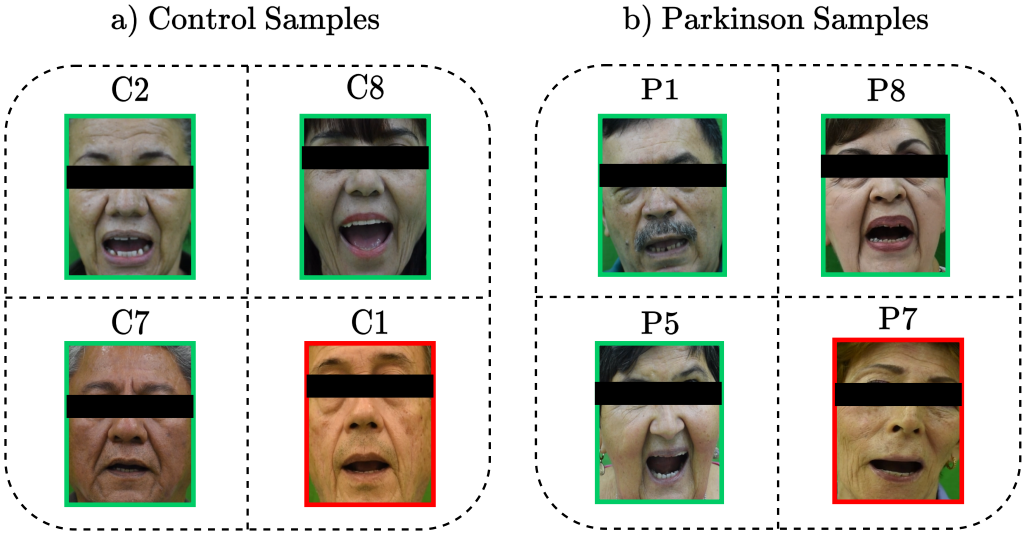

Figure 7. Examples of Parkinson's (b) and control (a) samples. In green is marked the contrast of the patients correctly identified in their class, while in red identified those that were wrongly predicted.

Modality	*Phonemes*					*Vowels*					*Words*				
	Prec	*Rec*	*F1*	*Acc*	*AUC*	*Prec*	*Rec*	*F1*	*Acc*	*AUC*	*Prec*	*Rec*	*F1*	*Acc*	*AUC*
Video	80,00	50,79	62,14	69,05	$82,59 \pm 6,91$	66,67	51,43	58,06	62,86	$70,46 \pm 7,09$	72,00	52,38	60,64	66,01	$79,53 \pm 3,07$
Audio	62,32	68,25	65,15	63,49	$67,72 \pm 9,1$	72,34	64,76	68,34	70,00	$73,55 \pm 7,03$	62,98	69,31	65,99	64,29	$67,91 \pm 3,74$
Fusion	88,89	63,49	74,07	77,78	$85,39 \pm 6,27$	72,63	65,71	69,00	70,48	$80,68 \pm 5,87$	77,48	76,46	76,96	77,12	$80,75 \pm 3$

Table 4. Experimental results for three experiments considered: video modality using 3D convolutions, audio modality using VGG-16, and the proposed approach using convex integration. For each of the experiments, the performance of the vowel, phoneme, and word exercises was evaluated independently. Determining the best result, the use of phonemes reached an AUC value of 85,39% when applying the proposed feature fusion approach.

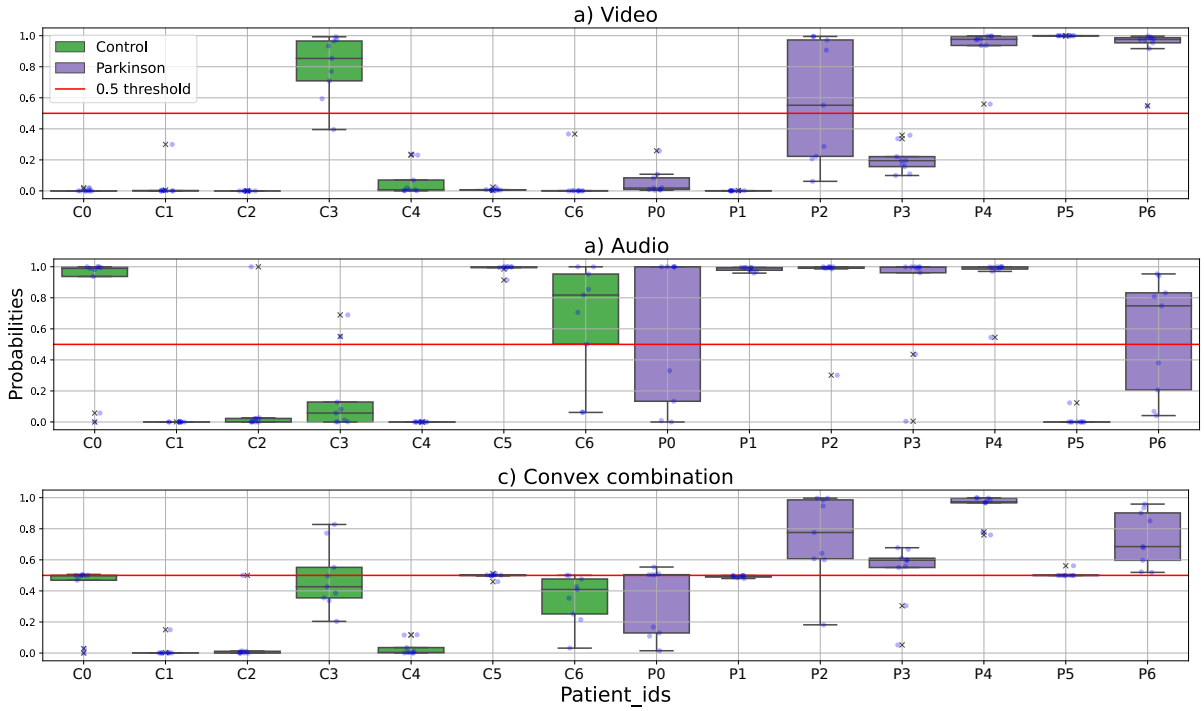

Figure 8. Probability distributions for gesture-auditory modalities and their integration for each patient included in the study, using box plots, scatter plots and phonemes exercises. Figure a) shows the estimated probability distributions for the Video modality, while Figure b) shows the probability distributions for the Audio modality. Finally, Figure c) shows the integrated probability distribution. A horizontal red line is used as a reference value (0.5) to determine whether a sample was predicted with Parkinson's or a control sample

9. DISCUSSION

This work reported the potential correlation between hypomimia and speech disorders, which may impact Parkinson's characterization and the consequent discrimination with a control population. Input video and audio signals were projected to deep convolutions representations to retrieve embedding descriptors that recover principal components associated with the alterations. The study was carried out as a retrospective study with 14 participants (7 Parkinson's and 7 control subjects), who develop phonation and linguistic exercises, while the face and audio were recorded. In this work was found that a spatiotemporal convolutional representation favors the face expressiveness characterization, achieving an 82.59% AUC performance, during the pronunciation of phonemes, for a Parkinson classification task. The adjusted 3D convolutional representation has a gain of around 7% regarding the typical approximation of face modeling from isolated frames of a particular sequence. In the same sense, the speech-related disorders were coded from input audio signals of patients while developing exercises of communication. In our study, the transformation of such raw signal to Mel Spectrograms and the respective projection to a convolutional net was sufficient to achieve a representation with the capability to discriminate Parkinson patients with an AUC of 73%, while pronouncing the vowels.

Remarkably, the study herein carried out evidence of the importance of multimodal analysis of the disease, demonstrating better and consistent characterization when probability outputs from audio-deep representation were integrated with the respective probabilities of facial deep representation. In fact, during this study the patients were invited to develop three different pronunciation exercises (vowels, phonemes, and words), being the convex integration consistently better in whole experiments. As evidence, the proposed convex integration approach yields promising improvements, estimating a gain in results of a maximum of 17.67% for phonemes and a minimum of 1.22% for words. These results may be key to studies with complex manifestations of the disease, impacting the decrease of false positive diagnoses. Also, the processing of multiple modalities may impact in an appropriate following of disease progression, and as a

tool to measure the impact of a particular therapy.

In the literature, the orofacial disorders for Parkinson's disease have principally modeled from independent audio and video sources. Regarding face-gestural representations, some works have reported the capability of discrimination between Pakinson and control patients, from isolated face images (a study with 54 PD patients, achieving a 71% of AUC) [33]. The approaches related to isolated image expressions may be sensitive to the geometrical properties of the images and they have to lead to high face variability. Also, in such works is lost much of the neuromotor impairments only are visible in dynamic sequences. In fact, this study evidenced how a volumetric deep architecture better discriminates Parkinson's patients approaching the kinematic information during the development of communication exercises. Some approaches have included specialized equipment to measure facial electromyography and electrocardiogram signals, as input to supply the quantification. In such a study was validated the discrimination of electrical facial signals with 7 PD patients, observing significant differences between his group of Parkinson's patients and control subjects [32]. The approaches based on external and invasive instruments may alter the natural facial gestures and restrict the analysis of expressions. Also, the modeling of facial landmarks has been proposed as a descriptor to discriminate 30 patients diagnosed with Parkinson, while developing emotional communication exercises [35]. In such work was achieved 88.46% of AUC using landmarks descriptors in a support vector machine. Along the same line, the landmarks-based approaches approximate expressions form a reduced set of key points which may lose some tremor and rigidity profiles during the language exercises. Contrarily, this work introduced markerless strategies that only use video recordings without additional artifacts to capture gesture dynamics. At the same time, the proposed approach models video clips that allow to capture of structural face information but also describe localized dynamics during exercises. In this sense, the proposed approach achieves an AUC of 85,39 and a recall of 76,46, for PD patients diagnosed by an expert Neurologist.

Some complementary works have been dedicated to analyzing Parkinson's language affectations but using the response from audio information. These studies focused to measure audio repre-

sentations as feature vectors. For instance, discrimination with an accuracy of 88% was achieved from phoneme pronunciations, revealing that the onset and offset of audio segments are highly correlated with the presence of Parkinson's disease [17]. Also, a study that included monologues and phrases in three distinct languages showed disease markers in voiced and unvoiced sounds [18]. More recent works have coded sounds in frequency-based representations to capture information from deep representation, achieving an effective classification performance (up to 90% in a cohort with 50 Parkinson patients) [19][30].

Despite remarked classification from isolated sources, there exists neurological evidence that orofacial respiratory and laryngeal systems are correlated and may have a major impact on Parkinson's disease characterization [7][8]. Also, this work constructed a paired dataset with synchronized audio and face gestures that include exercises of phonation, articulation, and prosody (vowels, phonemes, and words). This dataset also constitutes a contribution for the scientific community to explore alternative integration of both sources of information to achieve a better characterization of disease. For instance, some relationships have consistently been identified between the jaw muscles and the different respiratory systems (involving mechanical stimulation of peripheral speech structures).

Taking into account such premises, this study integrated the speech and gestural representations, obtained from audio and video signals. A simple linear integration demonstrated a high capability to discriminate Parkinsonian patterns in different speech exercises. In fact, an AUC of 85,39 was achieved in the exercise of phonemes, when β has a weight of 0.5 for video. Contrary, for the exercise of vowels, with a β of 0.1, was evidence of a remarked contribution to audio representation.

To quantify the impact of external variables in our study, an analysis of Confounding variables on the age and gender of the patients was performed for all the exercises studied. When studying the impact of the mentioned variables on the Video modality, an average absolute percentage error of $err\% = 0.1666$ was estimated. On the other hand, in the Audio modality $err\% = 0.1866$. To the gestural-auditory representation, we found a $err\% = 0.1934$. Thus, we

were able to identify how both the independent modalities and their combination, the impact of age and gender does not exceed an error of 20%, so they do not represent a significant bias in the results presented in this work.

10. CONCLUSIONS AND FUTURE WORK

This work introduced a novel PD digital descriptor that codifies speech and faces gesture representations to classify Parkinson's patients regarding a control population. For this purpose, we designed and recover a dataset with synchronized audio (voice) and video (face) sequences while patients perform multiple linguistic exercises designed for PD analysis. Hence, this work designed deep representations to analyze video patterns from a 3D convolutional net, while the speech was coded in Mel Spectrograms and analyzed with a standard convolutional approach. Both nets have the capability to discriminate between Parkinson´s and control subjects. Hence, both probability outputs were integrated using a simple linear combination, showing a positive impact on the classification task. The coded representation and following the integration rule, it was showing major support in PD characterization, with a significative correlation between oral and gestural patterns. Also, the deep representation was robust to capture disease patterns, leading to face variations and speech variability from patients included in this study.

Future works include an extended study in the design and modeling of fusion strategies to approach orofacial integration at different scales of representation. For instance, the use of multimodal nets that approach information and fusion at different scales, which in turn learn better fusion among modalities to categorize the disease. Also, from the findings of this work, there exist other validation scenarios that can be validated for analysis of the patient stage. For instance in open dialogues of patients, where orofacial impairments can be detected to support expert analysis.

BIBLIOGRAPHY

Abrami, Avner et al. "Automated Computer Vision Assessment of Hypomimia in Parkinson Disease: Proof-of-Principle Pilot Study". In: *Journal of Medical Internet Research* 23.2 (2021), e21037 (cit. on pp. 21, 42).

Baltrušaitis, Tadas, Chaitanya Ahuja, and Louis-Philippe Morency. "Multimodal machine learning: A survey and taxonomy". In: *IEEE transactions on pattern analysis and machine intelligence* 41.2 (2018), pp. 423–443 (cit. on pp. 16, 17, 31).

Bandini, Andrea et al. "Analysis of facial expressions in parkinson's disease through video-based automatic methods". In: *Journal of neuroscience methods* 281 (2017), pp. 7–20 (cit. on pp. 13, 21).

Bayoudh, Khaled et al. "A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets". In: *The Visual Computer* (2021), pp. 1–32 (cit. on p. 17).

Bhidayasiri, Roongroj et al. "Parkinson's disease: Hoehn and Yahr scale". In: *Movement Disorders: A Video Atlas: A Video Atlas* (2012), pp. 4–5 (cit. on p. 25).

Bologna, Matteo et al. "Facial bradykinesia". In: *Journal of Neurology, Neurosurgery & Psychiatry* 84.6 (2013), pp. 681–685 (cit. on pp. 16, 28).

Borod, Joan C et al. "Parameters of emotional processing in neuropsychiatric disorders: Conceptual issues and a battery of tests". In: *Journal of communication disorders* 23.4-5 (1990), pp. 247–271 (cit. on pp. 12, 43).

Ekman, Paul. "Pictures of facial affect". In: *Consulting Psychologists Press* (1976) (cit. on p. 12).

Fox, Cynthia M and Lorraine Olson Ramig. "Vocal sound pressure level and self-perception of speech and voice in men and women with idiopathic Parkinson disease". In: *American Journal of Speech-Language Pathology* (1997) (cit. on p. 11).

Gomez, Luis F et al. "Improving parkinson detection using dynamic features from evoked expressions in video". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 1562–1570 (cit. on pp. 21, 42).

Grammatikopoulou, Athina et al. "Detecting hypomimia symptoms by selfie photo analysis: for early Parkinson disease detection". In: *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 2019, pp. 517–522 (cit. on p. 13).

Gunduz, Hakan. "Deep learning-based Parkinson's disease classification using vocal feature sets". In: *IEEE Access* 7 (2019), pp. 115540–115551 (cit. on p. 19).

He, Kaiming et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778 (cit. on p. 32).

Jankovic, Joseph. "Parkinson's disease: clinical features and diagnosis". In: *Journal of neurology, neurosurgery & psychiatry* 79.4 (2008), pp. 368–376 (cit. on p. 15).

Jin, Bo et al. "Diagnosing Parkinson disease through facial expression recognition: video analysis". In: *Journal of medical Internet research* 22.7 (2020), e18697 (cit. on p. 21).

Katsikitis, M and I Pilowsky. "A controlled quantitative study of facial expression in Parkinson's disease and depression." In: *Journal of Nervous and Mental Disease* (1991) (cit. on p. 16).

Katsikitis, Mary and I Pilowsky. "A study of facial expression in Parkinson's disease using a novel microcomputer-based method." In: *Journal of Neurology, Neurosurgery & Psychiatry* 51.3 (1988), pp. 362–366 (cit. on p. 20).

Logemann, Jeri A et al. "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients". In: *Journal of Speech and hearing Disorders* (1978) (cit. on p. 11).

Madeley, P, AW Ellis, and RHS Mindham. "Facial expressions and Parkinson's disease". In: *Behavioural Neurology* 8.2 (1995), pp. 115–119 (cit. on p. 11).

Maycas-Cepeda, Teresa et al. "Hypomimia in Parkinson's Disease: What Is It Telling Us?" In: *Frontiers in Neurology* 11 (2021), p. 1775 (cit. on p. 15).

McClean, Michael D and Stephen M Tasko. "Association of orofacial with laryngeal and respiratory motor output during speech". In: *Experimental brain research* 146.4 (2002), pp. 481–489 (cit. on pp. 12, 30, 37, 43).

Novotnỳ, Michal et al. "Automatic evaluation of articulatory disorders in Parkinson's disease". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.9 (2014), pp. 1366–1378 (cit. on pp. 13, 19, 43).

Orozco-Arroyave, Juan Rafael et al. "Voiced/unvoiced transitions in speech as a potential biomarker to detect Parkinson's disease". In: *Sixteenth annual conference of the international speech communication association.* 2015 (cit. on pp. 13, 19, 20, 25, 43).

Parkinson's Disease, Movement Disorder Society Task Force on Rating Scales for. "The unified Parkinson's disease rating scale (UPDRS): status and recommendations". In: *Movement Disorders* 18.7 (2003), pp. 738–750 (cit. on p. 12).

Poewe, Werner et al. “Parkinson disease”. In: *Nature reviews Disease primers* 3.1 (2017), pp. 1–21 (cit. on p. 12).

Ramig, Lorraine O, Cynthia Fox, and Shimon Sapir. “Speech treatment for Parkinson’s disease”. In: *Expert Review of Neurotherapeutics* 8.2 (2008), pp. 297–309 (cit. on p. 16).

Ricciardi, L et al. “Hypomimia in Parkinson’s disease: an axial sign responsive to levodopa”. In: *European Journal of Neurology* 27.12 (2020), pp. 2422–2429 (cit. on pp. 12, 15, 16, 30).

Ross, Elliott D, Robin D Thompson, and Joseph Yenkosky. “Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions”. In: *Brain and language* 56.1 (1997), pp. 27–54 (cit. on p. 12).

Sapir, Shimon, Lorraine Ramig, and Cynthia Fox. “Speech and swallowing disorders in Parkinson disease”. In: *Current opinion in otolaryngology & head and neck surgery* 16.3 (2008), pp. 205–210 (cit. on pp. 11, 16).

Sapir, Shimon, Lorraine Olson Ramig, and Cynthia Fox. “Voice, speech, and swallowing disorders”. In: *Handbook of Parkinson’s Disease*. CRC Press, 2007, pp. 469–492 (cit. on p. 16).

Simonyan, Karen and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014) (cit. on pp. 21, 32).

Smith, Marcia C, Melissa K Smith, and Heiner Ellgring. “Spontaneous and posed facial expression in Parkinson’s disease”. In: *Journal of the International Neuropsychological Society* 2.5 (1996), pp. 383–391 (cit. on p. 11).

Spielman, Jennifer L, Joan C Borod, and Lorraine O Ramig. “The effects of intensive voice treatment on facial expressiveness in Parkinson disease: preliminary data”. In: *Cognitive and behavioral neurology* 16.3 (2003), pp. 177–188 (cit. on pp. 11, 12, 30, 37).

Szegedy, Christian et al. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826 (cit. on p. 28).

Tran, Du et al. “A closer look at spatiotemporal convolutions for action recognition”. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2018, pp. 6450–6459 (cit. on p. 28).

Tran, Du et al. “Learning spatiotemporal features with 3d convolutional networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4489–4497 (cit. on p. 28).

Valenzuela, Brayan et al. “A Spatio-Temporal Hypomimic Deep Descriptor to Discriminate Parkinsonian Patients”. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2022, pp. 4192–4195 (cit. on p. 28).

Vásquez-Correa, Juan Camilo et al. “Convolutional neural networks and a transfer learning strategy to classify Parkinson’s disease from speech in three different languages”. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 24th Iberoamerican Congress, CIARP 2019, Havana, Cuba, October 28-31, 2019, Proceedings 24*. Springer. 2019, pp. 697–706 (cit. on pp. 19, 29, 43).

Vásquez-Correa, Juan Camilo et al. “Multimodal assessment of Parkinson’s disease: a deep learning approach”. In: *IEEE journal of biomedical and health informatics* 23.4 (2018), pp. 1618–1630 (cit. on pp. 19, 20).

Vinokurov, Nomi et al. “Quantifying hypomimia in parkinson patients using a depth camera”. In: *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer. 2015, pp. 63–71 (cit. on p. 13).

Wodzinski, Marek et al. “Deep learning approach to Parkinson’s disease detection using voice recordings and convolutional neural network dedicated to image classification”. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 717–720 (cit. on pp. 13, 19, 29, 43).

Wu, Peng et al. “Objectifying facial expressivity assessment of Parkinson’s patients: preliminary study”. In: *Computational and mathematical methods in medicine* 2014 (2014) (cit. on pp. 20, 42).

Xu, Zhijing et al. “Voiceprint recognition of Parkinson patients based on deep learning”. In: *arXiv preprint arXiv:1812.06613* (2018) (cit. on p. 29).

APPENDICES

Anexo A. Academic Products

Journals

- B. Valenzuela, J. Arevalo, W. Contreras, F. Martinez, Orofacial Parkinson discrimination throughout deep dedicated representations, Submitted to Computer Methods and Programs in Biomedicine.

Conference papers

- Valenzuela, B., Arevalo, J., Contreras, W., Martinez, F. (2022, July). A Spatio-Temporal Hypomimic Deep Descriptor to Discriminate Parkinsonian Patients. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC) (pp. 4192-4195). IEEE.

Collaborations

- C.C. Viáfara, B. Valenzuela, F. Martínez, J.J. Penagos, A method to analyze wear mechanisms on worn chute lining surfaces using computer vision tools, Tribology International, Volume 186, 2023.

Anexo B. Informed Consent

Versión 0.1
Código: _____

ESCUELA DE INGENIERÍA DE SISTEMAS
UNIVERSIDAD INDUSTRIAL DE SANTANDER - Laboratorios Vive Digital

CONSENTIMIENTO INFORMADO

Proyecto: Cuantificación de patrones locomotores para el diagnóstico y seguimiento remoto en zonas de difícil acceso
Responsables: Fabio Martínez Carrillo, Gabriel Rodrigo Pedraza Ferreira.

9/10/2020

Con base en los reglamentos establecidos en la Resolución Nº 008430 del 4 de octubre de 1993 por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud en Colombia y según el artículo 15 relacionado con el Consentimiento Informado usted deberá conocer de forma completa y clara los aspectos de la investigación que se desarrollará. Usted ha sido convocado para este proyecto por cumplir con los requisitos de inclusión para la grabación de un conjunto de datos en vídeo de movimientos en marchas anormales y patológicas de manera natural. Por tal motivo se le invita formalmente a que participe del estudio teniendo en cuenta los siguientes criterios de inclusión:

- Ser mayor de edad.
- Poder realizar 4 ciclos de marcha (8 pasos) de manera seguida

De acuerdo con lo anterior y en cumplimiento de estos criterios, por favor indique con una X en una de las siguientes opciones qué tipo de participante es usted:

___ **Persona control:** Es aquella persona que no presenta ninguna dificultad motora, implicando que no ha sido diagnosticada de ninguna enfermedad que afecte su movimiento natural.

___ **Persona con afectación en la marcha:** Es aquella persona que tiene algún tipo de patología diagnosticada la cual pueda afectar su marcha.

Tenga en cuenta que su participación en este proyecto es ***absolutamente voluntaria***. Por favor lea con cuidado el documento y haga todas las preguntas que desee hasta su total comprensión.

JUSTIFICACIÓN

Usted está invitado a participar en este proyecto para crear un conjunto de datos usados en el análisis de patrones de movimiento en marchas normales y patológicas registradas en vídeo. Los movimientos se esperan analizar en vídeo, registrados de manera natural mediante una cámara convencional buscando capturar el movimiento natural de mínimo 3 ciclos de marcha (6 pasos) en línea recta. Para investigaciones futuras los datos recolectados serán usados únicamente para fines de investigación y sus respectivas publicaciones.

OBJETIVO

Desarrollar una estrategia computacional para el registro, cuantificación y representación de patrones de la marcha en ambientes semi-controlados como soporte de procesos de asistencia a la telerehabilitación.

DESCRIPCIÓN

El proceso de captura será el mismo para personas control como para personas con algún tipo de afectación en sus patrones de marcha. Esto debido a que el estudio considera que los datos han sido capturados bajo las mismas condiciones en ambos tipos de población, en miras de una adecuada evaluación estadística. Esta convocatoria se limita a población con edad mayor a 18 años.

A cada participante, en presencia de sus acompañantes, familiar o un testigo, se le entregará para su lectura. Si decide participar, podrá proceder a firmarlo. Seguidamente se registran sus datos personales. La grabación de vídeos tomará un tiempo aproximado de 10 minutos. Ante la cámara, usted deberá hacer caminata de 3 ciclos de marcha (6 pasos) equivalente a 8 metros aproximadamente, esta caminata deberá ser repetida 3 veces con el fin de poder capturar cualquier tipo de variación en los patrones de marcha.

9/10/2020

Para la sesión de grabación es necesario que usted como los investigadores cumplan con unas medidas de bioseguridad para evitar el contagio del COVID-19, es por esto que a continuación presentamos a usted unas recomendaciones dadas por el Comité de Protocolos de Bioseguridad de la Universidad Industrial de Santander para el desarrollo de la sesión:

1. Mantener el distanciamiento mínimo de dos metros entre personas
2. Todas las personas usan los elementos de protección personal (EPPs) y el uso de careta las personas que participan.
3. El uso de gel antibacterial o alcohol al ingreso y salida del sitio de prueba de cada persona que participa en el proyecto, si la persona tiene contacto con alguna superficie deberá igualmente desinfectar sus manos.
4. Verificar que las personas que participan en el proyecto tengan una afiliación al Sistema de Seguridad Social.

Al participar en este estudio, usted no recibirá ningún tipo de subvención económica o material ni deberá aportar herramienta alguna para la intervención. Al finalizar la investigación, usted podrá recibir los resultados obtenidos de la captura. Este material será presentado a usted por los investigadores cuando culmine la actividad.

Las inquietudes adicionales que surjan en relación con el desarrollo e implicaciones del proyecto podrán ser aclaradas por Fabio Martínez Carrillo, Profesor de la Escuela de Ingeniería de Sistemas e Informática, a quien puede contactar en el teléfono 3103054041, o mediante correo electrónico dirigido a famarcar@saber.uis.edu.co; o directamente en su oficina en la Universidad Industrial de Santander (sede principal) ubicada en la Cra. 27 #9 Ciudad Universitaria, Edificio de Laboratorios Pesados, oficina 231; o al teléfono 577- 6344000 extensión 2110.

RIESGOS

De acuerdo con el Artículo 11 de la Resolución No. 8430 de 4 de octubre de 1993, esta investigación se considera sin riesgo para el participante dado que el estudio únicamente emplea el registro de datos a través de un procedimiento común de captura de vídeos por medio de cámaras ordinarias. De tal forma, ninguno de los métodos utilizados es invasivo o penetra la piel. si durante la captura de los vídeos usted experimenta cualquier tipo de malestar, la grabación será suspendida de inmediato y se le ubicará en estado de reposo. De presentarse cualquier incidente relacionado con esta jornada donde el participante requiere valoración médica inmediata será remitido al servicio de urgencias del Hospital Universitario de Santander o si es su decisión al servicio de urgencias de la entidad donde se encuentre afiliado al sistema de seguridad social. Durante este proceso será acompañado por el investigador principal.

DERECHO A RETIRARSE

Su participación en este estudio es autónoma y voluntaria, en donde podrá actuar acorde a sus principios personales. Si usted decide no participar, no implicará sanción alguna. Además, usted cuenta con el derecho a negarse a responder a preguntas concretas si así lo desea. También puede optar por retirarse en cualquier momento y toda su información será descartada y eliminada.

CONFIDENCIALIDAD

Los resultados de las pruebas y la información que usted nos ha dado son de carácter absolutamente confidencial, de manera que solamente usted y el investigador principal tendrán acceso a estos datos.

Una copia de los registros con la información de cada participante será archivada por el investigador principal y a cada registro se le asignará un número con el cual se identificará y codificará para su ingreso a la base de datos durante la sistematización de la información. Por lo anterior, los nombres de los participantes no serán divulgados en forma alguna; y cuando los resultados de este estudio sean publicados en revistas o congresos científicos, la información personal de los participantes será debidamente anonimizada previamente.

A menos que Usted dé una autorización específica cuando la ley lo permita, sus resultados personales no estarán disponibles para terceras personas como empleadores, organizaciones gubernamentales, compañías de seguros o instituciones educativas. Esto también aplica a su cónyuge, a otros miembros de su familia. Sin embargo, con el objetivo de realizar un manejo adecuado de los datos, un miembro del Comité de Ética de la Universidad Industrial de Santander podrá consultar sus datos y su registro. Por lo anterior, atentamente se le invita a participar en el estudio y si está de acuerdo, se le solicita su nombre y la firma en las casillas abajo descritas.

9/10/2020

AUTORIZACIÓN PARA EL USO DE LA INFORMACIÓN EN ESTUDIOS FUTUROS

Dentro del equipo de investigación al que pertenecen los investigadores responsables (Grupo de Investigación BIVL^{2}ab - *Biomedical Imaging, Vision and Learning Laboratory*) de la Universidad Industrial de Santander, se espera seguir utilizando la información registrada en este estudio para el desarrollo de estudios futuros y derivados. Por lo tanto, al firmar este consentimiento usted puede autorizar al investigador principal a ceder su información a otros investigadores de su equipo de investigación, con previa aprobación del Comité de Ética de la Universidad Industrial de Santander para realizar los estudios mencionados. Por favor marcar con una X si autoriza o no autoriza, y firmar en caso de si autorizar.

9/10/2020

___ Si autorizo
___ No autorizo

_____

Firma Participante
Nombre:
C.C.

Huella digital
(En caso que se amerite)

_____

Firma Investigador Principal
Nombre:

Huella digital
(En caso que se amerite)

Yo_____,
identificado con _____ N°_____ de _____ al firmar este consentimiento el día ___ de _____ del _____, acepto participar de manera voluntaria en el presente estudio y autorizo la grabación de mis vídeos y el uso de mis datos individuales para los análisis requeridos. He leído y entendido la información registrada en este documento y mis dudas fueron aclaradas. Entiendo que soy libre de retirarme del estudio. Por otro lado, se me ha garantizado justicia, equidad, autonomía en la participación y la confidencialidad en el manejo de toda la información recolectada, teniendo en cuenta que los resultados del procesamiento de dicha información podrán ser divulgados con fines científicos, mediante presentaciones en congresos o publicaciones en revistas científicas, con la debida protección de mi identidad

_____ Huella digital
Firma Participante (En caso que se amerite)
Nombre:

_____ Huella digital
Firma Profesional Salud (Testigo 1) (En caso que se amerite)
Nombre:

_____ Huella digital
Firma Acompañante Captura de datos (Testigo 2) (En caso que se amerite)
Nombre:

_____ Huella digital
Firma Investigador Principal (En caso que se amerite)
Nombre:

9/10/2020

Contacto Comité de Ética: Para preguntas o aclaraciones acerca de los aspectos éticos de ésta investigación pueden comunicarse con el Comité de Ética en Investigación Científica de la Universidad Industrial de Santander (CEINCI-UIS), o con cualquiera de los miembros del Comité, al teléfono 6344000 Extensión 3808 ó al correo comitedetica@uis.edu.co.