

2022-05-16

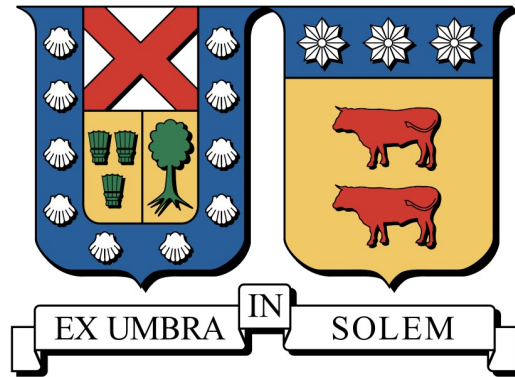
ANÁLISIS DE SEÑALES DE ACELERACIÓN DE LA SUPERFICIE DEL CUELLO PARA LA IDENTIFICACIÓN DE PERSONAS SANAS O PACIENTES CON NÓDULOS VOCALES, INTEGRANDO LA PRESIÓN SUBGLOTAL A UN MODELO DE REGRESIÓN LOGÍSTICA DE 2 PARÁMETROS

CARRERA ALBURQUENQUE, TAMARA

<https://hdl.handle.net/11673/53682>

Repositorio Digital USM, UNIVERSIDAD TECNICA FEDERICO SANTA MARIA

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



“Análisis de señales de aceleración de la superficie del cuello para la identificación de personas sanas o pacientes con nódulos vocales, integrando la presión subglotal a un modelo de regresión logística de 2 parámetros”

Realizado por:

Tamara Carrera Alburquenque

MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL ELECTRÓNICA MENCIÓN
TELECOMUNICACIONES

Profesor Guía: Matías Zañartu

Tutor Correferente: Juan Pablo Cortés

16 de Mayo de 2022

Agradecimientos

Quisiera comenzar agradeciendo a mi familia, a mi madre Sandra y a mi padre Ricardo, por haberme brindado las herramientas para estudiar, por su amor y apoyo incondicional en todas las actividades que se me han ocurrido hacer. Extiendo este agradecimiento a mis hermanos, Loreto y Ricardo, que sin sus consejos y experiencias se me habría hecho más difícil este camino. No puedo dejar de mencionar a mis abuelas, tías y tíos, que siempre han confiado en mis capacidades, permitiendome desarrollarlas con plena confianza y superar cualquier eventualidad con sus palabras de aliento. Me gustaría incluir a mis abuelos, que sé que desde el lugar en el que se encuentren también me cuidan y me apoyan, los siento.

Agradezco también a mis compañeros de Universidad, al grupo de amigos que formé en un comienzo, el cual fue un gran soporte emocional e intelectual. No olvidaré a quienes me han acompañado de alguna u otra forma en las mañanas/tardes/noches de estudio, gracias. Dejando de lado lo académico, agradezco a las grandes mujeres que conocí por el equipo de básquetbol y a mis compañeros de Acción Educa, sin duda todos/as aportaron al aprendizaje propio de la vida universitaria fuera del aula. Incluyo en estos agradecimiento a mis amigas del colegio, con quienes sé que cuento ante cualquier eventualidad y que han estado presente todos estos años. Destaco especialmente a mi mejor amiga Macarena y a mi pololo por su apoyo en la finalización de este trabajo.

El desarrollo de este trabajo no hubiese sido posible sin la confianza del profesor Matías y el apoyo de Juan Pablo, a quien agradezco su disposición en ayudarme y resolver dudas. Agradezco a todos mis profesores y ayudantes por los conocimientos entregados, y a los correspondientes, por la dedicación en sus clases.

Resumen

Las patologías asociadas al comportamiento de la voz ocurren debido a un mal uso o a un uso excesivo de ella. Muchos trabajadores o estudiantes que desarrollan actividades como el canto o la pedagogía, son más propensos a desarrollar estas patologías. Pueden ir desde una disfonía (NPVH) a traumatismos en el tejido de las cuerdas vocales, como la creación de nódulos (PVH), los cuales pueden llegar a ser un gran impedimento para las labores que realizan cotidianamente. El propósito de este estudio fue crear dos modelos de regresión logística a partir de datos ambulatorios, comenzando con un modelo de dos dimensiones obtenido en trabajos anteriores, que funciona como un buen clasificador entre pacientes y controles. El otro modelo propuesto, es de tres dimensiones y considera las 2 dimensiones previas junto con otra obtenida tras estimaciones de presión subglótica. La finalidad es poder realizar una comparación entre los resultados de ambos. Para la estimación de la presión, se utilizó un modelo de redes neuronales y luego se realizó un estudio estadístico para seleccionar qué variable obtenida a partir de su distribución podía diferenciar mejor a los grupos de estudio. De acuerdo a los resultados obtenidos, la exactitud de la clasificación mejoró con el modelo propuesto, obteniendo en un principio 74.62 % con el modelo de dos dimensiones y de 77.95 % en el modelo al cual se le agregó una dimensión. Específicamente, mejoró la clasificación de los sujetos del grupo de pacientes, de un 74.23 % a 80.98 % (sensibilidad), mientras que la clasificación del grupo de control, se mantuvo (especificidad). De las probabilidades de clasificación, también se puede extraer un índice que corresponde al nivel de fonotrauma diario (DPI) que poseen los sujetos de estudio. Este índice también se vio modificado con el nuevo modelo, otorgándole más valores positivos a los pacientes que en el anterior, lo que es útil debido a que valores positivos de DPI están asociados a patológicos y los negativos a un comportamiento normal de la voz. Finalmente, los resultados expuestos en este trabajo logran una mejoría respecto al anterior, sin embargo, no es considerablemente mayor, debido a la correlación que existe entre el parámetro nuevo que se agregó al modelo y uno de los que ya existía (presión subglótica y nivel de presión sonora, respectivamente). No obstante, esta variación en el valor del DPI puede ser utilizada para investigaciones futuras de pacientes que se realicen cirugías o terapias de la voz, para mostrar un cambio significativo al pasar de un valor positivo a uno negativo, o bien, para especificar el grado de atención que necesitan.

Palabras claves: clasificación, hiperfunción vocal, DPI, monitoreo ambulatorio, presión subglótica, red neuronal, regresión logística.

Abstract

Pathologies associated with voice behavior occur due to misuse or overuse of the voice. Many workers or students who develop activities such as singing or pedagogy are more prone to develop these pathologies. They can range from dysphonia (NPVH) to vocal cord tissue trauma, such as the creation of nodules (PVH), which can become a major impediment to the work they perform on a daily basis. The purpose of this study was to create two logistic regression models from ambulatory data, starting with a two-dimensional model obtained in previous work, which works as a good classifier between patients and controls. The other proposed model is three-dimensional and considers the 2 previous dimensions together with another obtained after subglottic pressure estimations. The purpose is to be able to compare the results of both. A neural network model was used to estimate the subglottal pressure and then a statistical study was performed to select which variable obtained from its distribution could best differentiate the study groups. According to the results obtained, the classification accuracy improved with the proposed model, initially obtaining 74.62 % with the two-dimensional model and 77.95 % in the model to which one dimension was added. Specifically, the classification of the subjects in the patient group improved from 74.23 % to 80.98 % (sensitivity), while the classification of the control group was maintained (specificity). From the classification probabilities, an index corresponding to the level of daily phonotrauma (DPI) possessed by the study subjects can also be extracted. This index was also modified with the new model, giving more positive values to the patients than in the previous one, which is useful because positive DPI values are associated with pathological and negative values with normal voice behavior. Finally, the results presented in this work achieve an improvement with respect to the previous one, however, it is not considerably higher, due to the correlation between the new parameter added to the model and one of those that already existed (subglottic pressure and sound pressure level, respectively). Nevertheless, this variation in the DPI value can be used for future investigations of patients undergoing voice surgery or therapy to show a significant change from a positive to a negative value, or to specify the degree of attention they need.

Keywords: ambulatory monitoring, classification, DPI, logistic regression, neural network, subglottal pressure, vocal hyperfunction.

Glosario

ACC: Acelerómetro.

ACFL: Amplitud del flujo de aire glótico inestable.

DPI: Daily phonotrauma index (Índice de fonotrauma diario).

IBIF: Impedance based inverse filtering (Filtrado inverso basado en impedancia).

IF: Inverse filter (Filtrado inverso).

MFDR: Maximum flow declination rate (Tasa máxima de declinación del flujo).

MVHM: Mobile voice health monitoring (Dispositivo móvil de monitoreo de la salud de la voz).

NaN: Not a number (No es un número).

NPVH: Non phonotraumatic vocal hyperfunction (Hiperfunción vocal no fonotraumática).

OQ: Open quotient (Cociente abierto).

PVH : Phonotraumatic vocal hyperfunction (Hiperfunción vocal fonotraumática).

Ps: Presión subglótica.

SPL: Sound pressure level (Nivel de presión sonora).

SQ: Speed quotient (Cociente de velocidad).

VH: Vocal hyperfunction (Hiperfunción vocal).

Índice

1. Introducción	4
1.1. Hiperfunción vocal	4
1.2. Adquisición de datos	5
1.3. Red Neuronal y DPI	6
2. Objetivos	9
2.1. Objetivo General	9
2.2. Objetivos Específicos	9
3. Estado del arte	10
3.1. Monitoreo de la voz	10
3.2. Modelo de filtrado inverso basado en impedancia sugblotal	11
3.3. Parámetros aerodinámicos	13
3.4. Red Neuronal	13
3.5. Regresión logística y DPI	15
4. Propuesta de solución	17
4.1. Base de datos	17
4.2. Procesamiento de los datos	18
4.3. Estimación de la presión subglótica	18
4.4. Regresión logística	19
4.4.1. Modelo de regresión logística base	19
4.4.2. Modelo de regresión logística propuesto	19
5. Implementación de solución	21
5.1. Datos	21
5.2. Estimación presión subglótica y estadísticos	23
5.3. Regresión logística y DPI	25
6. Resultados	27
6.1. Extracción de los parámetros	27
6.2. Estimación presión subglótica	27
6.3. Modelos de regresión logística	29
6.4. DPI	37
7. Conclusiones	40

Índice de figuras

1.	Sistema móvil de monitoreo de la salud de la voz. (a) Acelerómetro y smartphone (b) Acelerómetro ubicado para recolección de datos [11]. . . .	11
2.	Representación del sistema subglótico y modelado. (a) Posición del acelerómetro y partes del modelo. (b) Analogía del sistema subglótico incluyendo cargas por la piel [4].	12
3.	Esquema del modelos de adquisición de datos ambulatorios y modelo inicial de red neuronal de regresión [9].	14
4.	Gráfico de dispersión DPI, antes del tratamiento y después de la cirugía y/o de la terapia [17].	16
5.	Información de los archivos de personas del grupo de control.	21
6.	Información de los archivos de personas del grupo de patológicos.	21
7.	Información de los archivos de personas del grupo de control tras filtrado IBIF.	22
8.	Información de los archivos de personas del grupo de patológicos tras filtrado IBIF.	22
9.	Histograma presión subglótica estimada para el grupo de control.	27
10.	Histograma presión subglótica estimada para el grupo de pacientes.	28
11.	Matriz de confusión modelo de 2 parámetros.	30
12.	Gráfico de dispersión de la desviación estándar H1-H2 versus la asimetría del nivel de presión sonora, para el modelo de 2 dimensiones.	31
13.	Matriz de confusión modelo de tres parámetros.	32
14.	Gráfico de dispersión de la desviación estándar de H1-H2 versus la asimetría de SPL para el modelo de 3 dimensiones con $P_s = 0$	33
15.	Gráfico de dispersión de la asimetría de P_s versus la desviación estándar de H1-H2 para el modelo de 3 dimensiones con $SPL = 0$	34
16.	Gráfico de dispersión de la asimetría de P_s versus la asimetría de SPL para el modelo de 3 dimensiones con $H1-H2 = 0$	34
17.	Gráfico de dispersión de tres dimensiones con hiperplano obtenido del modelo de regresión (asimetría P_s eje Z).	35
18.	Gráfico de dispersión de tres dimensiones con hiperplano obtenido del modelo de regresión (asimetría SPL eje Z).	36
19.	Gráfico de dispersión de tres dimensiones con hiperplano obtenido del modelo de regresión (desv. estándar H1-H2 eje Z).	36
20.	Histograma valores DPI para sujetos del grupo de control (valor diario).	37
21.	Histograma valores DPI para sujetos del grupo de patológicos (valor diario).	38
22.	Gráfico de dispersión valores DPI, comparación resultados DPI para modelo de 3 dimensiones (DPI 2) versus el de 2 dimensiones (DPI 1).	39

Índice de tablas

1.	Formato de los parámetros para entrar a la red neuronal.	23
2.	Valores medios (desviación estándar) de las estadísticas calculadas para la presión subglótica (control y paciente) e índice de Cohen.	28
3.	Valores-p resultantes de las pruebas de significancia para valores estadísticos de presión subglótica.	29

1 Introducción

Los trastornos vocales son patologías asociadas al comportamiento de la voz y pueden ocurrir debido a un mal uso de ella o a utilizarla excesivamente en la cotidianidad. Trabajadores o estudiantes que desempeñan actividades asociadas al canto o a la pedagogía son más susceptibles a recurrir en lesiones vocales, debido al gran uso vocal que conlleva su quehacer. Sin ir más lejos, alrededor del 30 % de la población adulta de Estados Unidos ha experimentado un trastorno de la voz [13]. Aunque la mayoría de estos trastornos de la voz suelen ser de corta duración, son un gran impedimento para desarrollar labores que impliquen gran capacidad de habla cuando se padecen.

El reconocimiento de los comportamientos vocales y el uso de técnicas no invasivas para su detección, es lo que se utilizará en este trabajo para poder detectar personas patológicas y diferenciarlas de un grupo de personas sanas, empleando señales ambulatorias. El uso de estas técnicas para estudiar el comportamiento de la voz es muy importante para lograr un seguimiento de los pacientes. Estas mediciones no se pueden realizar con precisión durante un breve período de tiempo, ya que los comportamientos vocales hiperfuncionales pueden variar en gran medida durante el transcurso del día de un individuo, por lo que puede no demostrarse claramente durante una examinación médica breve.

1.1 Hiperfunción vocal

La hiperfunción vocal (VH por sus siglas en inglés) se refiere a la actividad excesiva de los músculos y articulaciones del mecanismo vocal durante la producción de voz, o más bien, a la “actividad musculoesquelética perilaríngea excesiva” durante la fonación y se considera un componente etiológico en los tipos de trastornos de la voz más frecuentes [8]. Dentro de los trastornos más comunes, están los que incluyen lesiones benignas de las cuerdas vocales, como por ejemplo los nódulos, y también los que ocurren en ausencia de alguna patología diagnosticable, más relacionados con el deterioro vocal, como la disfonía de tensión muscular.

El inicio y la persistencia crónica de VH se han atribuido a una variedad de factores potenciales, pero el papel relativo de estos factores en la etiología y la fisiopatología crónica de VH sigue siendo incierto [8]. Al día de hoy, hay dos tipos principales de VH que son fundamentalmente diferentes en términos de fisiopatología fonatoria, y son llamados: VH fonotraumático (PVH) y VH no fonotraumático (NPVH), los cuales se diferencian por la presencia o ausencia de lesiones vocales, respectivamente. Este trabajo se realizará con datos ambulatorios provenientes de personas que presentan PVH y un grupo de control correspondientes a personas sin comportamientos vocales hiperfuncionales, emparejados por edad y ocupación.

Estudios anteriores han proporcionado evidencia de que los mecanismos fisiopatológicos (biomecánicos) asociados con las dos formas de VH son bastante diferentes. Los estudios de parámetros aerodinámicos glóticos han demostrado que los pacientes con PVH ne-

cesitan producir valores anormalmente altos de presión de aire subglótica, flujo de aire modulado (que refleja una mayor amplitud de la vibración de las cuerdas vocales) y tasa máxima de declinación del flujo (que refleja un aumento en la velocidad de cierre de las cuerdas vocales y colisión) para alcanzar niveles normales de presión sonora [5]. Estos resultados reflejan un mayor potencial para causar trauma en el tejido de las cuerdas vocales, lo que contribuye a la presencia crónica de lesiones en ellas y la disfonía asociada [8]. Por el contrario, los pacientes con NPVH muestran una función fonatoria anormal que es ineficiente. Esto se manifiesta con aumentos en la presión de aire subglótica y cociente abierto (que refleja un cierre glótico incompleto), sin aumento del flujo de aire modulado y tasa máxima de declinación del flujo. Estos cambios pueden producir disfonía, pero no alcanza a causar trauma en el tejido de las cuerdas vocales [5].

Los resultados más recientes, utilizan un monitoreo ambulatorio de la voz y corroboran aún más las descripciones previas de los mecanismos asociados con la hiperfunción vocal fonotraumática. Se ha mostrado que la función vocal diaria de los pacientes se puede diferenciar de los controles (sanos), en función de las características de distribución a largo plazo de dos parámetros (combinados en una regresión logística), que indican una mayor magnitud de las fuerzas fonatorias y la tendencia a emplear con mayor frecuencia un cierre glótico más abrupto [16].

1.2 Adquisición de datos

Al realizar un estudio del comportamiento de la voz de forma no invasiva y ambulatoria, se generan señales de larga duración. En este estudio, las mediciones que se van a utilizar, se obtienen de un dispositivo móvil de monitoreo de la salud de la voz (MVHM por sus siglas en inglés), el cual consiste en un acelerómetro en miniatura ubicado en el cuello para utilizarlo como sensor de fonación. Posteriormente, los datos son transmitidos a un dispositivo smartphone, utilizándolo como plataforma de adquisición. Esto permite obtener la señal de forma no invasiva y durante un largo período de tiempo, con el fin de capturar el comportamiento de un individuo en su estado natural y captando las posibles variaciones que pueda tener, por día, a lo largo de una semana, que es lo que durará la medición.

Este dispositivo es parte de un largo trabajo de investigación, donde anteriormente, distintas señales obtenidas a través del MVHM se han utilizado para extraer parámetros característicos de la voz. En un comienzo, los que sirven para cuantificar el impacto de su uso, como lo son la frecuencia fundamental (f_0) y el nivel de presión sonora (SPL) [7], las medidas de dosis de vibración vocal [14]. Luego, se consideraron medidas espectrales y cepstrales [12] y medidas aerodinámicas ([10] [4]). Las últimas mencionadas corresponden a las más recientes, son obtenidas a través de un procesamiento de filtrado inverso, utilizando un modelado del sistema vocal y realizando un filtrado inverso subglotal basado en impedancias (IBIF por sus siglas en inglés). IBIF es un esquema basado en modelos físicos para estimar el flujo de aire glótico a partir de la aceleración de la superficie del cuello, cuyo objetivo es obtener los parámetros aerodinámicos que

generalmente son obtenidos en laboratorios clínicos, utilizando señales ambulatorias [19].

Los parámetros que se basan en medidas extraídas de la señal acústica, es decir, proveniente del acelerómetro del MVHM, no pueden proporcionar información sobre los mecanismos fisiológicos específicos que producen la voz. Por ejemplo, las características de la fuente de velocidad del volumen glótico, por lo que se han realizado más esfuerzos en monitorización ambulatoria de la voz con características fisiológicamente relevantes que puedan ayudar a identificar la hiperfunción vocal (medidas aerodinámicas) [4].

Para obtener los parámetros IBIF, específicos de cada sujeto, se comparan las estimaciones de la forma de onda del flujo de aire glótico derivadas del IBIF, con las que se utilizan en la actualidad en laboratorios clínicos. Actualmente, se realizan a partir de la estimación de una señal de flujo de aire glótico filtrada inversamente obtenida a partir de grabaciones usando una máscara de neumotacógrafo con ventilación circunferencial [4]. Para el caso del IBIF, la señal se filtra para eliminar la influencia del tracto vocal. Finalmente, se estiman clínicamente los parámetros a partir de la obtención de la forma de onda del flujo de aire glótico, específicamente, la amplitud del flujo de aire glótico inestable (ACFL), cociente abierto (OQ), de velocidad (SQ) y la tasa máxima de declinación del flujo (MFDR).

Los primeros esfuerzos para obtener características aerodinámicas a partir de la aceleración de la superficie del cuello, se limitaron a vocales sostenidas [19]. Mientras que en [4], utilizando datos ambulatorios, se demostró que las medidas de flujo de aire glótico se pueden utilizar con éxito para diferenciar entre la función vocal normal y los mecanismos fisiopatológicos asociados con la PVH. Demostraciones que superaron los resultados anteriores donde han utilizado SPL, f_o o dosis vocales.

1.3 Red Neuronal y DPI

El aprendizaje automático y la inteligencia artificial se están convirtiendo en herramientas relevantes en la ingeniería biomédica, ya que pueden proporcionar predicciones precisas e implementaciones eficientes. Es por esto, que en el presente estudio se busca estudiar e implementar un algoritmo basado en un modelo de redes neuronales, el cual permite la estimación de la presión subglótica (Ps) [9] utilizando parámetros de la señal adquirida del MVHM. Los parámetros a utilizar como entrada a la red, serán 7 y se considerarán los adquiridos en proyectos anteriores, tanto acústicos como aerodinámicos. Específicamente, dentro de los aerodinámicos se tiene ACFL, OQ, MFDR, SQ y la diferencia entre el primer y segundo armónico (H1-H2), por otra parte, los acústicos corresponderán a f_o y SPL.

Desafortunadamente, la medición directa de características vocales importantes, como la presión subglótica, la presión de colisión de las cuerdas vocales y la activación del músculo laríngeo, no es práctica en entornos de laboratorio y ambulatorios [9]. Es por esto que se opta por un método para estimarlas a partir de una señal obtenida con el MVHM,

mediante un marco que integra un modelo fisiológico de producción de voz y herramientas de aprendizaje automático. La señal del acelerómetro de la superficie del cuello se debe procesar primero con el esquema IBIF para producir una estimación del flujo de aire glótico inestable y obtener las medidas aerodinámicas mencionadas previamente.

Se propone estimar la presión subglotal utilizando datos ambulatorios en este trabajo, ya que no se ha ideado ningún método de aprendizaje automático entrenado con un modelo de producción de voz para la señal del acelerómetro con mediciones ambulatorias. En estudios anteriores, se han utilizado señales provenientes de un acelerómetro que solo representaban sílabas específicas para la aplicación de diversos métodos de estimación clínicos, por lo que, analizar señales ambulatorias que representen de mejor forma el comportamiento de la voz de los individuos, resulta ser un método atractivo de estimarla. Importante es destacar que la validación de las estimaciones de P_s , obtenidas en el trabajo [9], pudo realizarse experimentalmente mediante una comparación con estimaciones de referencia de la presión subglótica, derivadas de una técnica estándar de interrupción del flujo de aire en el laboratorio. El detalle de la implementación del modelo de red neuronal se mostrará en la sección Red Neuronal del capítulo 3.

Tras obtener el resultado de los valores de P_s estimados, se procederá a calcular sus estadísticos, con la finalidad de escoger uno de ellos y añadirlo a un modelo de regresión logística, ya construido en un trabajo previo de clasificación entre personas con nódulos en las cuerdas vocales y personas sanas [17]. En el trabajo recién mencionado, se entrenó y probó un modelo de regresión logística *stepwise*, parcialmente impulsado por la teoría, utilizando las características más predictivas seleccionadas a partir de diversas pruebas estadísticas que demostraban un nivel de significancia del 5% ($p\text{-value} < 0.05$) y un tamaño del efecto mediano a grande (*Cohen's d effect size*). Es posible entrenar un mejor modelo utilizando todas las características estadísticamente significativas. En los resultados de este estudio, se concluyó que en general, en comparación con los controles, las únicas diferencias altamente discriminatorias en el uso semanal de la voz y la función vocal de los pacientes PVH, son la asimetría de SPL (skewness) y la variabilidad H1-H2. En otras palabras, los pacientes tienden a pasar más tiempo hablando más alto que el promedio, con un cierre glótico más abrupto en comparación con los controles emparejados [16].

Con los resultados mencionados, se logró obtener un modelo de regresión logística de dos parámetros, el cual es utilizado posteriormente en [17]. Previo a este estudio, el modelo de regresión de dos parámetros solo había clasificado a los pacientes con lesiones fonotraumáticas antes del tratamiento [16], es por esto que fue necesario obtener más información sobre el rol del uso diario de la voz en la etiología y fisiopatología de la PVH. Esto se logró aplicando el DPI, para predecir mejoras basadas en grupos y la variabilidad de sujetos individuales en el uso de la voz ambulatoria de pacientes con PVH, después de cirugía laríngea y/o terapia de voz post quirúrgica [17]. En este modelo, se utilizó la regresión logística obtenida en [16], para utilizar las probabilidades de clasificación, aplicándoles una transformación logit, dando origen a un índice de fonotrauma diario o

DPI. Se utilizó el logit porque los cambios en L (ΔL) se pueden interpretar por igual en toda la escala.

El DPI puede verse como una medida del potencial de trauma de las cuerdas vocales. En [17], se basa en la combinación de las características de distribución a largo plazo (diarias) de dos parámetros, los que representan la magnitud de las fuerzas fonatorias (NSAM o SPL) y la dinámica de cierre de las cuerdas vocales (H1–H2), seleccionados previamente en [16]. En este estudio, se pretende agregar un estadístico de P_s a los parámetros seleccionados, para luego calcular los valores de DPI que se obtendrán modificando el modelo de regresión inicial.

En un inicio, previo a cualquier intervención, el DPI era significativamente alto para los pacientes. Los valores normativos de este índice, es decir, de funcionamiento normal de la voz, corresponden a valores negativos. Cabe destacar que, aunque el estudio de [17] se centró en el uso del DPI para medir los cambios relacionados con el tratamiento en el uso de la voz, también se probaron otras medidas objetivas de la voz, utilizadas tradicionalmente con este conjunto de datos, y ninguna de ellas se asoció de forma significativa con el tratamiento, excepto la variabilidad del f_o .

La disminución significativa del DPI después de la cirugía, antes de la terapia de voz, está probablemente asociada a una reducción de los ajustes compensatorios, que ya no son necesarios porque las lesiones de las cuerdas vocales han sido eliminadas por la extirpación quirúrgica. El hecho de que el DPI de los pacientes, después de la cirugía, siguiera siendo significativamente más alto, que el del grupo de control, apoya la suposición de que había una persistencia de la hiperfunción vocal antes de la terapia de voz, lo que probablemente refleja las condiciones que contribuyeron a la formación (etiología) de las lesiones fonotraumáticas [17].

2 Objetivos

2.1 Objetivo General

El objetivo principal de este proyecto de titulación, es lograr identificar entre 96 personas (48 pares paciente-control), las que cuentan con hiperfunción vocal fonotraumática, específicamente nódulos en las cuerdas vocales, y las personas que están sanas. Esto se realizará considerando señales ya obtenidas en investigaciones previas, las cuales cuentan con sus respectivos parámetros acústicos y aerodinámicos [4].

Para identificar entre pacientes y personas sanas se hará uso de modelos de regresión logística. Se comenzará creando con un modelo, diseñado previamente, de dos parámetros. Luego, se creará otro modelo basado en el primero y agregándole un nuevo parámetro, proveniente de la presión subglótica. Este nuevo parámetro será obtenido a partir de los estadísticos de las estimaciones resultantes de una red neuronal entrenada con datos sintéticos y validada en la investigación de [9].

2.2 Objetivos Específicos

Los objetivos específicos de esta investigación se relacionan con el desarrollo de actividades que permitan cumplir el objetivo general. Se detallan a continuación:

1. Extraer los parámetros de interés a partir de las señales recopiladas desde una gran base de datos relacionada con investigaciones de comportamientos vocales de individuos del Hospital General de Massachusetts.
2. Aplicar modelo de redes neuronales ya entrenado con la finalidad de obtener predicciones de Ps y sus estadísticas para seleccionar una de ellas que sea agregada a un modelo de regresión logística previamente establecido en [16].
3. Comparar el desempeño de clasificación del modelo de regresión creado de 2 parámetros con el modelo propuesto en este trabajo de 3 parámetros, añadiendo una medida estadística de Ps al primer modelo.
4. Analizar los cambios obtenidos en el DPI a raíz de incorporar un nuevo factor al modelo de regresión.

3 Estado del arte

El estudio de comportamiento de la voz y la identificación de hiperfunciones vocales fonatorias, no se pueden realizar con precisión durante un breve período de tiempo, ya que los comportamientos vocales hiperfuncionales pueden variar en gran medida durante el transcurso del día de un individuo, llegando a no demostrarse claramente durante una examinación médica breve. Es por esto, que surge la idea de trabajar con mediciones del comportamiento vocal no invasivas y ambulatorias, con el fin de diferenciar de manera confiable los patrones normales de comportamiento de la voz de los de hiperfunción vocal.

Debido a lo anterior, y a la idea de trabajar con mediciones no invasivas, es que los datos y modelos a utilizar son parte de un proyecto más amplio y continuo de un gran grupo de investigadores. Dicho proyecto, está destinado a desarrollar un monitoreo ambulatorio de la fonación, como una herramienta efectiva de evaluación clínica y biofeedback, por lo que hay varios trabajos desarrollados por el mismo grupo de investigadores que utilizan monitoreo no invasivo y parámetros obtenidos de las señales, ya sean los acústicos o aerodinámicos.

3.1 Monitoreo de la voz

Comenzando con la adquisición de datos, en este proyecto se hace uso de un dispositivo móvil de monitoreo de la salud de la voz, el que cuenta con un acelerómetro adherido a la superficie del cuello, permitiendo obtener una señal de forma no invasiva y durante 7 días, con el fin de capturar el comportamiento de un individuo en su estado natural y con las posibles variaciones que pueda tener a lo largo de una semana. El acelerómetro (ACC) presente en el dispositivo, es un buen sensor de fonación para la monitorización a largo plazo, debido a que puede llevarse discretamente en la base del cuello sobre la clavícula. Tiene características de sensibilidad y rango dinámico que lo hacen muy adecuado para detectar la vibración de las cuerdas vocales. Además, es relativamente inmune a los sonidos ambientales externos y produce una señal relacionada con la voz que no está filtrada por las resonancias del tracto vocal, especificaciones expuestas en [11]. El sistema completo de monitoreo se puede observar en la Figura 1.

El acelerómetro presente en el dispositivo, entrega una señal que mide la aceleración subglotal de la superficie del cuello, la cual es más robusta para detectar la fonación y soluciona problemas asociados a la señal acústica del habla, como por ejemplo, el ruido exterior. De la señal obtenida, se pueden extraer parámetros basados en el flujo de aire glótico, algunos de los cuales ya han sido utilizados en trabajos anteriores para la clasificación de pacientes con hiperfunción vocal y personas sanas. Se ha cuantificado la acumulación del uso de la voz, con parámetros como SPL, f_o , parámetros asociados a la exposición de las cuerdas vocales a la vibración, como el tiempo de fonación, el número de periodos oscilatorios realizado por las cuerdas vocales en el tiempo (cycle dose) y la distancia que recorren las cuerdas vocales durante una vibración (distance dose), cuyos valores medios no han dado buenos resultados en la diferenciación de patológicos y un

grupo de control [15].

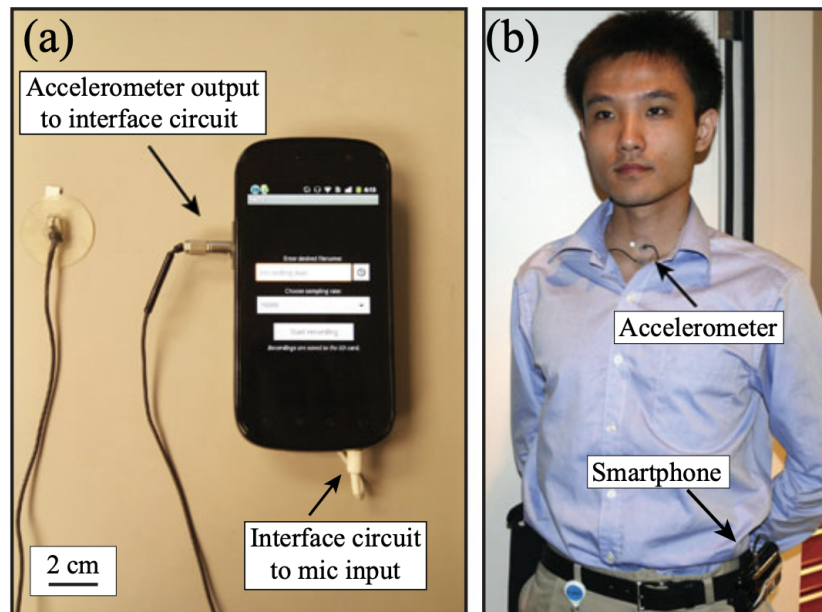


Figura 1: Sistema móvil de monitoreo de la salud de la voz. (a) Acelerómetro y smartphone (b) Acelerómetro ubicado para recolección de datos [11].

Sin embargo, se han desarrollado nuevos esfuerzos para obtener medidas aerodinámicas de la voz. A partir de la misma señal del acelerómetro, se han encontrado mejores resultados al momento de clasificar entre patológicos y grupo de control [16] con estas medidas. La extracción de ellas se realiza a partir de un modelado del sistema vocal, que luego pasa por un filtrado inverso subglotal basado en impedancias (IBIF) [19].

3.2 Modelo de filtrado inverso basado en impedancia suglotal

El objetivo de una evaluación ambulatoria es obtener medidas de la función vocal que sean esencialmente equivalentes a lo que se puede adquirir en el entorno de laboratorio clínico. A la cual se le incluyen una combinación importante de parámetros acústicos y aerodinámicos, obtenidos a medida que los individuos se dedican a sus actividades diarias típicas.

En un principio, los métodos de estudio de señales ambulatorias solo extraían estimaciones de frecuencia fundamental y el nivel de presión sonora, junto con los parámetros asociados al uso de la voz (vocal doses). Si bien, se considera que esas mediciones tienen valor clínico, las capacidades diagnósticas de estas mediciones básicas se consideran inherentemente limitadas porque no incluyen parámetros aerodinámicos [19].

El filtrado inverso (IF) de la voz, se utiliza para estimar la fuente de excitación en la glotis, removiendo los efectos de cargas acústicas en una señal de salida. Esta técnica se realiza

principalmente para el tracto vocal, utilizando grabaciones de vocales sostenidas [19]. El esquema IBIF subglotal, proporciona un método simple para estimar el flujo de aire glotal y los parámetros aerodinámicos, a partir de una señal del ACC ubicado en la piel del cuello. La diferencia con un filtrado inverso tradicional, que generalmente se realiza para eliminar resonancias del tracto vocal, es que en este caso, como se desea utilizar la señal obtenida del ACC del MVHM, se deben eliminar las resonancias subglotales y los efectos de la piel en las mediciones.

El modelo IBIF se representa con parámetros de impedancia en el dominio de la frecuencia, caracterizando la piel del cuello y el ACC, lo cual sirve para modelar de mejor forma la propagación del sonido en el tracto subglotal. Lo que se desea estimar es el flujo de aire glótico, utilizando un modelo basado en analogías mecano-acústicas, a partir de un modelo de líneas de transmisión para crear una función de transferencia entre la aceleración medida en la superficie del cuello y el flujo de aire glotal [19]. En el modelo, el tracto subglotal está compuesto de dos secciones subglotales, donde cada una representa la parte de la tráquea extratorácica por encima y por debajo de la ubicación del acelerómetro.

En la Figura 2 se observa el esquema del modelo IBIF diseñado en [19] y utilizado en [4]. En el modelo equivalente, se observa que el tracto subglotal se compone de dos secciones, sub1 y sub2, las cuales representan la interconexión entre los tractos subglotales por encima y por debajo del acelerómetro, respectivamente. También se observa que hay una impedancia que representa la carga de la piel (Z_{skin}), esta incluye la carga del sensor del acelerómetro.

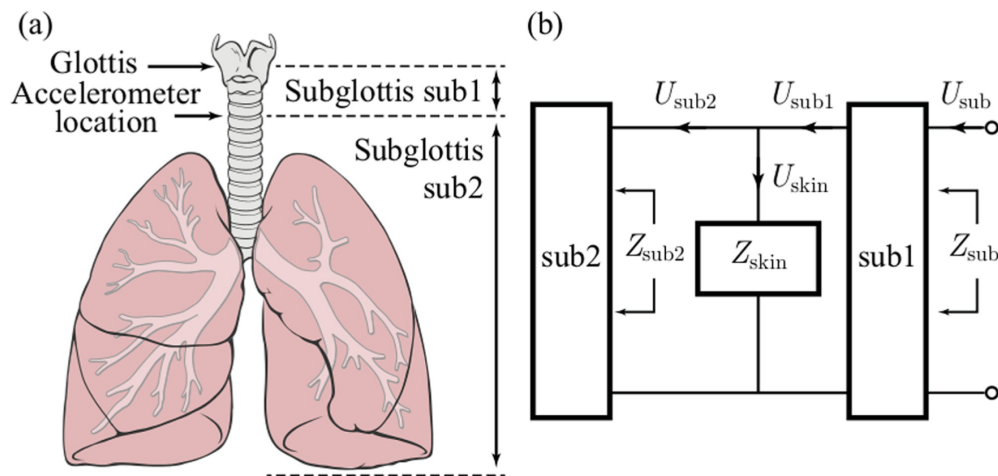


Figura 2: Representación del sistema subglótico y modelado. (a) Posición del acelerómetro y partes del modelo. (b) Analogía del sistema subglótico incluyendo cargas por la piel [4].

Por otra parte, en cuanto a las señales que se observan en la Figura 2, la derivada de la señal U_{skin} corresponde a la señal obtenida del acelerómetro y se considera como entrada.

Se considera como salida el flujo de aire subglotal, U_{sub} . La señal que se desea obtener, corresponde a la velocidad del volumen glotal (GVV) y se define como el negativo del flujo de aire subglotal ($-U_{sub}$).

3.3 Parámetros aerodinámicos

Si bien, en un comienzo se realizaron análisis del funcionamiento de la voz con parámetros acústicos, se verificó que medidas medias de SPL, f_o , CPP y *vocal doses* adquiridas de la vida diaria no fueron significativamente diferentes entre los pacientes con PVH y el grupo de control [16]. La falta de diferencia también sugiere que los pacientes con PVH compensan la presencia de lesiones de pliegue vocal para mantener niveles funcionales de sonoridad vocal, tono y calidad.

Por otro lado, se ha demostrado que las medidas aerodinámicas glotales, de la presión de aire subglotal y el flujo de aire glotal (normalizado por el nivel de presión acústica), se pueden utilizar para identificar mecanismos fonatorios asociados con la hiperfunción vocal, que son claramente diferentes de la función vocal normal [5].

En algunos trabajos posteriores a la elaboración del modelo de filtrado inverso, se hizo uso del IBIF para variadas estimaciones de parámetros aerodinámicos. En [4], se utilizó el esquema IBIF para estimar la forma de onda de flujo de aire glotal a partir de la señal proveniente del ACC de la superficie del cuello, eliminando la influencia del tracto vocal. Su finalidad era estimar los parámetros provenientes de la forma de onda de flujo de aire glotal, ACFL, OQ, MFDR, SQ y H1-H2. Este trabajo ([4]), fue uno de los primeros en incorporar señales ambulatorias para la extracción de parámetros aerodinámicos, utilizando el esquema IBIF, que también incluyó estudios estadísticos de los parámetros obtenidos. Estas medidas se han utilizado para diferenciar el uso diario de la voz de pacientes con hiperfunción vocal de controles emparejados ([7] [4] [18]) y para rastrear cambios relacionados con cirugía y tratamiento de terapia de la voz de los trastornos hiperfuncionales de la voz [17].

3.4 Red Neuronal

Otros esfuerzos recientes que se han implementado, tienen relación con la estimación de parámetros de medidas de función vocal, como la presión subglótica, la presión de colisión de las cuerdas vocales y la activación del músculo laríngeo, estimaciones que no son prácticas en entornos de laboratorio y ambulatorios. En [9], utilizan señales provenientes de un ACC ubicado en la superficie del cuello (no son parte del monitoreo ambulatorio), para entrenamiento y testeo de la red. Estas consistían en grabaciones de datos realizadas en una habitación con sonido aislado, donde el personal del estudio instruyó a cada participante para que repitiera cadenas de sílabas /pæ/ en tres condiciones de volumen (cómodo, alto y bajo). Aunque se instruyó a los sujetos para que mantuvieran un tono y un volumen constantes dentro de cada cadena de sílabas, eran libres de elegir los niveles que les resultaran más naturales sin ningún nivel prescrito de tono absoluto y volumen.

Luego, estas señales fueron procesadas mediante IBIF, estimando el flujo de aire glótico inestable, para luego extraer seis características aerodinámicas, ACFL, MFDR, OQ, SQ, H1-H2 y f_0 .

Se utilizan 7 parámetros como entrada, los extraídos mediante IBIF, junto con SPL (estimado directamente usando un micrófono acústico). Para estimar los parámetros de medidas de función vocal mencionados inicialmente, se emplea un algoritmo de redes neuronales multicapas (Figura 3), basado en un marco de aprendizaje automático supervisado para la regresión. Este modelo posee una capa de entrada de los 7 parámetros, una capa de salida de los 4 parámetros y 2 capas ocultas interconectadas con 10 % de ajuste para evitar sobreajuste [9]. La red fue entrenada a partir de un sintetizador de voz con un modelo de cuerdas vocales.

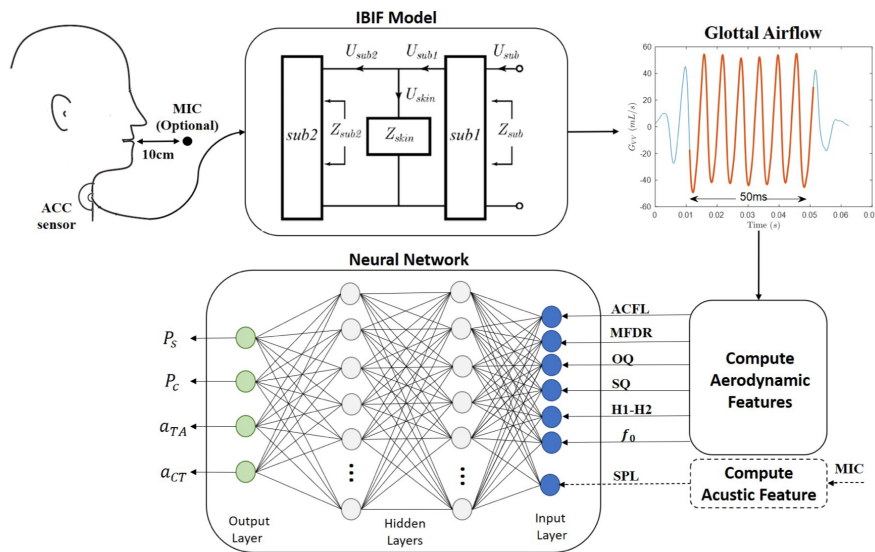


Figura 3: Esquema del modelos de adquisición de datos ambulatorios y modelo inicial de red neuronal de regresión [9].

En la Figura 3, se observa un esquema que representa la adquisición de las señales de voz por medio del ACC y su posterior procesamiento por el esquema IBIF para estimar el flujo de aire glótico. Luego, se realiza el cálculo de los 6 parámetros aerodinámicos, con la finalidad de que ingresen a la capa de entrada de la red neuronal, junto con SPL. Tras pasar por las capas ocultas de neuronas, se obtienen las estimaciones de los parámetros vocales mencionados anteriormente. Es importante mencionar que cada neurona dentro de las capas ocultas tienen parámetros ajustables de peso y sesgo, que se combinan con las salidas de la capa anterior para activar una función de activación de unidad lineal rectificada; luego, la activación resultante sirve como entrada para la siguiente capa [3]. A lo largo de la investigación se prueban distintos modelos de capas ocultas, variando la cantidad de capas y la cantidad de neuronas en cada una. Se encuentra que para la estimación de presión subglótica con datos reales (experimentales) y no sintéticos, el modelo

que entrega mejores resultados es el de 2 capas ocultas con 4 neuronas, porque entrega el menor valor de error medio absoluto (MAE) en comparación a los otros modelos [9].

Las predicciones obtenidas con este esquema, se validan frente a simulaciones numéricas y mediciones de laboratorio. Solo las estimaciones de presión subglótica pudieron ser comparadas con estimaciones de señales experimentales en laboratorio, debido a la dificultad de predecir experimentalmente las otras medidas vocales.

3.5 Regresión logística y DPI

En cuanto al paso final del trabajo a realizar, que corresponde a la regresión encargada de obtener la clasificación entre los pacientes patológicos y el grupo de control, es que se tiene como punto de partida el modelo de regresión logística utilizado en [17], que si bien corresponde a un modelo de dos parámetros, en este trabajo se pretende incorporar un nuevo parámetro para encontrar una mejora al momento de diferenciar los grupos analizados.

Previo al trabajo realizado en [17], se hizo una investigación en [16] para encontrar las características que mejor clasificaban a los grupos en estudio, utilizando un modelo de regresión logística basado en *stepwise*. Esto significa que se debe comenzar probando con un modelo fundamentado en la teoría que considere las características más predictivas, seleccionadas a partir del estudio de diversas pruebas que marcan las diferencias entre los distintos valores estadísticos de los parámetros mencionados, tanto acústicos como aerodinámicos. Una vez encontrada una significancia estadística de los valores menor al 5 %, la diferencia entre ellos es caracterizada por el cálculo del coeficiente de tamaño del efecto o *Cohen's d*. Por lo mencionado, es que en [16] se entrenó y probó un modelo de regresión logística que solo contenía características con tamaño (*Cohen's d*) mediano a grande. Las características que presentaron mayor capacidad de diferenciar ambos grupos resultaron ser dos, la asimetría de SPL (skewness) y la desviación estándar de H1-H2, estas fueron consideradas para el modelo de regresión logística final.

En conocimiento del trabajo previo, en cuanto a la selección de los dos parámetros a utilizar en la regresión logística, es que en [17] utilizan los parámetros mencionados, pero además se realiza una transformación inversa de cada estimación de probabilidad, conocida como logit (L), para crear un índice de fonotrauma diario (DPI). Esta transformación, permite la interpretación significativa de las diferencias o cambios entre los valores en el DPI. El objetivo de ellos era obtener una mayor información sobre el rol que juega el uso diario de la voz en la etiología y la fisiología de la hiperfunción vocal fonotraumática, por lo que estudiaron el DPI para predecir mejoras de los pacientes analizando un antes y después de personas que se realizaron una cirugía laríngea o terapia de la voz.

La hipótesis de [17], era que, en comparación con el tratamiento previo, el DPI disminuiría significativamente hacia la clasificación normal después de la cirugía laríngea y la terapia

de voz, pero los resultados mostraron que solo disminuyó hasta un rango normativo después de la terapia de voz.

En la Figura 4, se observan los resultados obtenidos en [17]. Casi todos los pacientes (25 de 27) mostraron una reducción del DPI tras la cirugía laríngea y/o la terapia vocal, terminando por debajo de la línea diagonal discontinua. 14 de esos pacientes mostraron una reducción tan grande, que sus valores de DPI tras el tratamiento fueron negativos y, por tanto, demostraron la conversión a valores normativos (puntos situados en el cuadrante inferior derecho del gráfico). Nueve de los pacientes con valores DPI positivos mostraron una reducción tras el tratamiento, pero no cruzaron el umbral de normalidad (sus DPI seguían siendo positivos) [17].

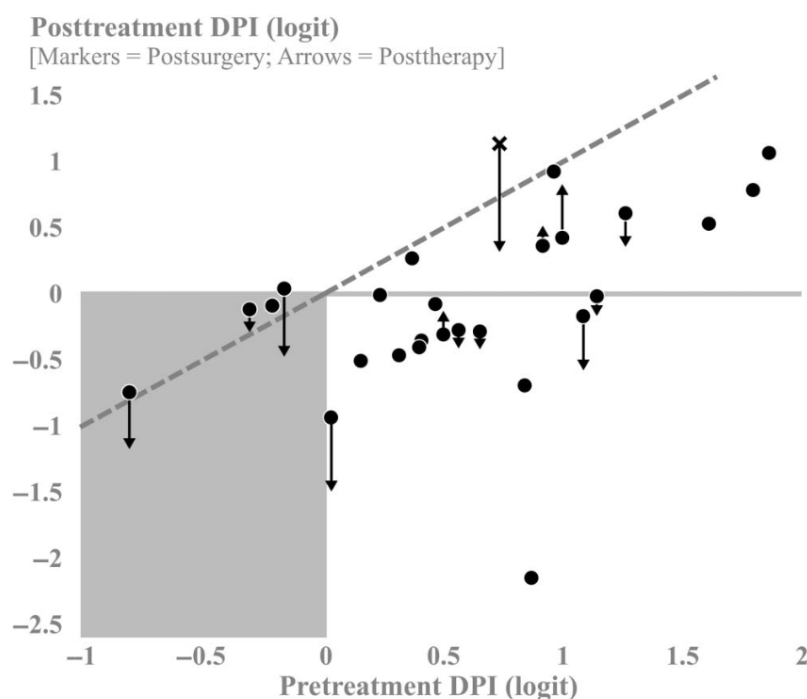


Figura 4: Gráfico de dispersión DPI, antes del tratamiento y después de la cirugía y/o de la terapia [17].

Con los resultados obtenidos en este estudio se podría definir, en futuras investigaciones, por ejemplo, si los valores de DPI más elevados se asocian a pacientes que finalmente deberán someterse a una cirugía laríngea. El valor individual del paciente, también podría ser útil en la discusión de si se debe intentar la cirugía antes de la terapia. Además, los pacientes con valores altos de DPI después de la cirugía, pueden requerir una terapia vocal más prolongada en comparación con los que tienen valores más bajos [17].

4 Propuesta de solución

Tras la investigación teórica de los avances en el uso de datos ambulatorios, relacionada con el estudio del comportamiento de la voz realizado en la sección anterior, es que este trabajo se realizará en base a los trabajos previos mencionados. Se considerarán las señales ambulatorias y los parámetros obtenidos en [4], el modelo de regresión logística y DPI de [16][17] y finalmente el modelo de redes neuronales de [9], cuya aplicación en este trabajo solo se considerará para estimar la presión subglótica. Posterior a esto, tras obtener las estimaciones de P_s , se calcularán medidas estadísticas con la finalidad de escoger uno de esos valores para que sea incorporado a la regresión logística, inicialmente calculada con el valor de asimetría (skewness) de SPL y la desviación estándar de H1-H2, para analizar si cumple el objetivo de mejorar la clasificación entre los grupos en estudio.

4.1 Base de datos

En primer lugar, se debe comenzar con la búsqueda de las señales a utilizar dentro de una extensa base de datos del Hospital General de Massachusetts (MGH), que corresponderán a las utilizadas en [4]. Los individuos que participaron del estudio fueron 48 pares de mujeres adultas (un total de 96 sujetos), donde cada par está compuesto por un paciente con PVH (diagnosticado con nódulos vocales) y un sujeto de control, emparejado con el paciente por edad y ocupación. Los diagnósticos se basaron en una evaluación completa del equipo de laringólogos y patólogos del habla y el lenguaje del Centro de Voz del MGH. Estas señales cuentan con la información de los parámetros acústicos y aerodinámicos ya calculados.

Para los 96 sujetos (48 pares de patológicos emparejados al grupo de control), existen mediciones diarias tras una semana de monitoreo, previa a cualquier intervención quirúrgica o terapia de la voz. Los datos de los sujetos (pacientes y control) se encuentran separados por día (para los días divididos en secciones, se considerará la unión de ellas). Además, para cada día se tienen archivos separados con los parámetros acústicos, obtenidos directamente de la señal del ACC, y los provenientes del IBIF.

Es por esto que se debe realizar un pre-procesamiento de las señales, en primer lugar para considerar solo los segmentos que tienen presencia de voz, y en segundo lugar para extraer los parámetros de interés desde los dos archivos distintos, por cada día de cada individuo. Se deben extraer los parámetros acústicos: f_o , SPL y $H1H2_{ACC}$ desde un archivo, y de otro, los relativos al IBIF: ACFL, OQ, MFDR, SQ, H1H2. El manejo de dos parámetros distintos de H1H2 se debe a que el proveniente del IBIF se calcula con una estimación del flujo de aire glotal del modelo y debe ser utilizado en la red neuronal para la estimación de P_s . Mientras que $H1H2_{ACC}$ se obtiene directamente de la señal del ACC y se debe utilizar en el modelo de regresión junto con SPL y luego, junto SPL y P_s .

Luego de extraer los parámetros de sus respectivas localizaciones, se unirán los ocho en

un solo archivo, que corresponderán a cada día de cada paciente y así poder utilizarlos en las siguientes etapas. Es necesario mencionar que estos archivos no contienen, necesariamente, la misma cantidad de información, ya que esta varía dependiendo del uso de la voz de cada individuo en el día de la medición, por lo que los archivos varían en la cantidad de datos, entre los días de un mismo individuo y en comparación con los otros.

4.2 Procesamiento de los datos

Una vez que se tenga la información de los siete parámetros necesarios para estimar la P_s , se deben realizar ciertos procesamiento para que todos los datos tengan las unidades de medidas correspondientes a las utilizadas para entrenar la red y no existan valores que escapen de los correctos. Para la entrada a la red neuronal, se requiere que el parámetro de ACFL tenga unidades de L/s y en los archivos se encuentra en mL/s, por lo que es necesario realizar una conversión de unidades. Por otra parte, hay algunos datos de los archivos que contienen valores NaN (Not a Number), que por algún motivo (al momento de calcularlos en los trabajos anteriores) fueron descartados como valores apropiados para las mediciones, por ende, se deben descartar. Otro problema que se debe considerar y modificar, es que la medición de SPL en los datos ambulatorios está hecha a 15 cm de la boca y para entrar a la red se necesita la medición a una distancia de 10 cm, por lo que se realiza dicha conversión. Los valores de OQ y f_o también fueron limitados para un rango de valores adecuados, $0 \leq OQ \leq 100\%$ y $f_o < 500$ Hz.

Continuando en la misma línea, con los valores problemáticos, modificados y/o descartados dependiendo del caso, se tendrán los archivos listos para ingresar a la red neuronal y realizar la estimación de la presión subglótica.

4.3 Estimación de la presión subglótica

Tomando como punto de partida los datos obtenidos luego de ser procesados, estos se deben ingresar a la red neuronal, cuyo modelo ya se encuentra entrenado y testeado con datos sintéticos, logrando un modelo que estima valores de P_s con el mínimo error y validado en [9]. Para lograr la estimación de P_s , se cargará el modelo a Python y se deberán leer todos los archivos que contienen los valores de los parámetros con el formato de entrada. Luego, se realizará una normalización para que los datos tengan un buen funcionamiento dentro de la red y finalmente, los valores de P_s resultantes se deberán desnormalizar. Los resultados desnormalizados corresponden a valores de la presión subglótica medidos en Pascales. La medida más común de este parámetro es en cm H_2O , por ende, se efectuará una conversión de unidades. Posterior a eso, se deben analizar los resultados, con la finalidad de eliminar los valores que no sean posibles en la práctica, como por ejemplo, si es que se obtienen valores negativos. Si bien la red está validada con valores reales de P_s , el cambio de utilizar datos obtenidos experimentalmente en laboratorios a usar datos ambulatorios, puede generar cierta diferencia.

Una vez que se obtengan los valores de P_s estimados y modificados, se deberán calcular

algunas medidas estadísticas, para posteriormente encontrar una de ellas que sea considerada en el nuevo modelo regresión propuesto. Luego, se analizará si mejora el resultado de la clasificación que utiliza solo SPL y H1H2. Los estadísticos que se calcularán serán la media, mediana, desviación estándar, asimetría, curtosis, percentiles 5, 95 y el rango. Calculados estos estadísticos, se almacenarán en un solo archivo los 8 valores por día de cada individuo, también se almacenará la información de los valores mínimos y máximos.

4.4 Regresión logística

Del análisis de la clasificación, de los 2 modelos, se verificará si el añadir un nuevo parámetro al modelo previamente utilizado en [16] mejora o no la diferencia entre el grupo de control y de pacientes. Por otra parte, se tiene el DPI, que es un índice utilizado para evaluar el cambio de un mismo individuo tras someterse a alguna variación en su condición de patológico (cirugía o terapia). Para obtener este índice, primero se debe realizar una regresión logística con los parámetros seleccionados y así obtener las probabilidades de clasificación. Posteriormente, se deberá aplicar una transformación inversa a cada una, función conocida como logit. Tras realizar dicha transformación, se crea el DPI. Sin embargo, no será utilizado para clasificación de cada uno de los grupos de estudio en este trabajo, pero se analizará el cambio que genere el agregar un nuevo parámetro al modelo de regresión, en los valores del DPI. Este análisis se hará con la finalidad de que en trabajos futuros se pueda usar el índice obtenido con el modelo de 3 dimensiones para una clasificación, si es que los resultados son favorables hacia la diferenciación.

Como para todo modelo de aprendizaje supervisado que se crea a partir de datos experimentales, la base de datos con la que se cuenta será dividida en dos secciones, correspondientes a un 50 % de entrenamiento y un 50 % de testeo.

4.4.1 Modelo de regresión logística base

Para comenzar, se debe crear un modelo de regresión logística basado en los parámetros que fueron seleccionados en [16] (la asimetría de SPL y la desviación estándar de H1-H2). Con estos parámetros fijos, se obtendrán resultados de clasificación y sus respectivos valores de DPI, los que posteriormente se deberán comparar con los obtenidos por el otro modelo, creado a partir de la incorporación de un nuevo parámetro proveniente de las estimaciones de presión subglótica que se realizarán y sus valores estadísticos. Para realizar este modelo, se utilizará el software MATLAB, apoyando el resultado con un algoritmo que ajusta un modelo de regresión lineal generalizada a los datos.

4.4.2 Modelo de regresión logística propuesto

Para crear el segundo modelo, que consiste en la propuesta de este trabajo, se hará un análisis de distintas pruebas estadísticas comunes para los valores estadísticos de presión subglótica. Específicamente, se aplicará la prueba de Kolmogorov-Smirnov para dos muestras (KStest), la prueba t de dos muestras (t-test), prueba de rango con signo

de Wilcoxon de dos muestras (Wilcoxon sign-rank test), prueba de suma de rangos de Wilcoxon de dos muestras (Wilcoxon rank sum test) y finalmente se calculará el tamaño del efecto de la variable según Cohen (Cohen's d). Este análisis se hará en base a los p-valores (*p-values*) que entregarán las distintas pruebas, escogiendo el mínimo dentro de los valores que cumplan una significancia menor del 5 %. Se busca una característica que rechace la hipótesis de que la misma variable, para pacientes y control, sea igual. Es decir, se busca la variable que permita separar de mejor manera ambos grupos. Por otro lado, el tamaño del efecto de Cohen se calculará para corroborar los resultados anteriores, debiendo tener un mayor tamaño la variable escogida, lo que implicará un mayor grado de diferencia entre ambos grupos analizados.

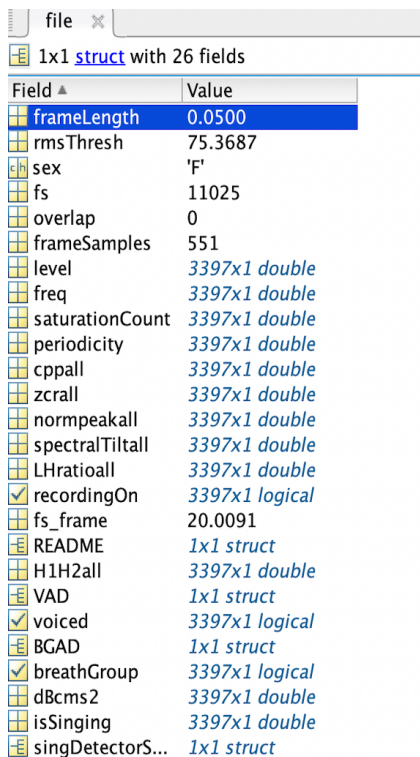
Luego, con el modelo de tres parámetros creado y entrenado, se podrán realizar las predicciones para el grupo de testeo y posteriormente comparar el rendimiento de la clasificación en ambos casos, con los dos modelos creados (2D y 3D). Esta comparación se realizará en base a la exactitud (accuracy) de la clasificación, debido a la importancia de que ambos grupos estén bien clasificados. También, se considerará la sensibilidad (sensitivity) y la especificidad (specificity), para observar los cambios en la clasificación de cada grupo. Finalmente, junto con lo anterior, se observarán los valores del área bajo la curva (AUC), de la curva característica de operación (ROC) para ambos casos.

Por otra parte, se calculará el DPI para ambos resultados de los modelos. Como bien se mencionó anteriormente, el DPI es un índice que muestra el nivel fonotraumático diario de cada individuo, cuya variación (ΔL) es analizada para observar un cambio en el comportamiento diario de la voz, tras una cirugía o terapia. En este caso, al analizar el DPI se prestará atención al cambio que ocurra, ya sea que disminuya o aumente. Esto con el objetivo de introducir una nueva investigación en su análisis, considerando que estos resultados puedan ayudar a identificar de mejor forma a un paciente o a un control, previo a alguna intervención o bien, al grado de intervención que podría necesitar y/o su duración.

5 Implementación de solución

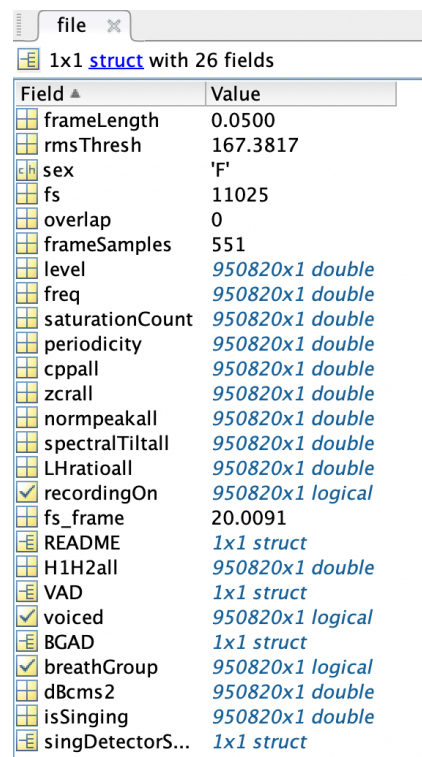
5.1 Datos

En primer lugar, se seleccionan los datos desde el servidor de MGH. Los archivos son los correspondientes al estudio realizado en [4], donde los pacientes están compuestos por individuos con nódulos en las cuerdas vocales (VPH) y cada uno tiene su pareja respectiva dentro del grupo de control, emparejados por edad y ocupación. Como se mencionó en la sección anterior, cada individuo posee una gran cantidad de archivos, correspondientes a las mediciones por día y a los parámetros que fueron calculados. Para esta investigación, se requirió utilizar tanto parámetros acústicos, como provenientes del IBIF. En las Figuras 5 y 6, se puede observar un ejemplo de la información que contienen los archivos, en estas figuras se representa la información recopilada en parte de un día de un individuo del grupo de control (Figura 5) y de un patológico Figura (6). La información que contienen es la misma, pero se debe hacer la diferencia entre ambos, para corroborar los resultados que se obtendrán en la clasificación final. Las variables que se considerarán son *voiced*, *freq*, *level* y *H1H2_{all}* de las Figuras 5 y 6.



Field	Value
frameLength	0.0500
rmsThresh	75.3687
sex	'F'
fs	11025
overlap	0
frameSamples	551
level	3397x1 double
freq	3397x1 double
saturationCount	3397x1 double
periodicity	3397x1 double
cppall	3397x1 double
zcrall	3397x1 double
normpeakall	3397x1 double
spectralTiltall	3397x1 double
LHRatioall	3397x1 double
recordingOn	3397x1 logical
fs_frame	20.0091
README	1x1 struct
H1H2all	3397x1 double
VAD	1x1 struct
voiced	3397x1 logical
BGAD	1x1 struct
breathGroup	3397x1 logical
dBcms2	3397x1 double
isSinging	3397x1 double
singDetectorS...	1x1 struct

Figura 5: Información de los archivos de personas del grupo de control.



Field	Value
frameLength	0.0500
rmsThresh	167.3817
sex	'F'
fs	11025
overlap	0
frameSamples	551
level	950820x1 double
freq	950820x1 double
saturationCount	950820x1 double
periodicity	950820x1 double
cppall	950820x1 double
zcrall	950820x1 double
normpeakall	950820x1 double
spectralTiltall	950820x1 double
LHRatioall	950820x1 double
recordingOn	950820x1 logical
fs_frame	20.0091
README	1x1 struct
H1H2all	950820x1 double
VAD	1x1 struct
voiced	950820x1 logical
BGAD	1x1 struct
breathGroup	950820x1 logical
dBcms2	950820x1 double
isSinging	950820x1 double
singDetectorS...	1x1 struct

Figura 6: Información de los archivos de personas del grupo de patológicos.

Considerar el segmento *voiced* es de gran importancia, debido a que las grabaciones cuentan con información de todo el día de los individuos, por lo que hay momentos del

día donde no se habla y otros donde el uso de la voz va a ir variando de intensidad de acuerdo a la actividad que estén realizando. Es por esto, que las grabaciones se dividieron en fotogramas no superpuestos de 50 ms de duración. Como se hizo en estudios previos ([12] [15]), cada fotograma fue considerado con voz (*voiced*) si superaba los siguientes umbrales: (a) SLP era superior a 45 dB (SPL a 15 cm, como se usa en el MVHM), (b) el primer peak de retardo distinto de cero en la autocorrelación normalizada superaba un umbral de 0,6, (c) f_0 (recíproco del retardo de tiempo del primer peak de autocorrelación distinto de cero) estuvo entre 70 y 1000 Hz, y (d) la relación de energía de baja a alta frecuencia superaba los 20 dB. Estos criterios fueron necesarios para eliminar varios tipos de actividad no fonatoria, como tocar o frotar el sensor, niveles extremadamente altos de ruido ambiental e interferencias/artefactos eléctricos. Esta actividad se obtiene directamente de la señal del ACC y sirve para luego utilizar las medidas del IBIF solo de los segmentos con voz.

Por otra parte, se tiene la información de los archivos que contienen los parámetros IBIF, representados en las Figuras 7 y 8, donde los parámetros a considerar son *acflow* (ACFL), *mfd*, *oq*, *sq* y *h1h2*, que son los necesarios para la estimación de presión subglótica con la red neuronal.

Workspace	
Name ▲	Value
acflow	3397x1 double
boolPhaseCha...	3397x1 double
cq	3397x1 double
file	1x1 struct
h1h2	3397x1 double
hrf	3397x1 double
mfd	3397x1 double
naq	3397x1 double
oq	3397x1 double
sq	3397x1 double
subject	'NF021'
vec	3397x1 double
voicedRMS	3397x1 logical

Figura 7: Información de los archivos de personas del grupo de control tras filtrado IBIF.

Workspace	
Name ▲	Value
acflow	950820x1 double
boolPhaseCha...	950820x1 double
cq	950820x1 double
file	1x1 struct
h1h2	950820x1 double
hrf	950820x1 double
mfd	950820x1 double
naq	950820x1 double
oq	950820x1 double
sq	950820x1 double
subject	'PF021'
vec	950820x1 double
voicedRMS	950820x1 logical

Figura 8: Información de los archivos de personas del grupo de patológicos tras filtrado IBIF.

Para comenzar con la extracción de los parámetros, se tienen todos los archivos correspondientes a cada uno de los días, de todas las parejas control (NF)/paciente (PF). Se procede a extraer los parámetros mencionados al comienzo de este capítulo de los pacientes y del grupo de control, debiendo unir los parámetros acústicos con los aerodinámicos para que ingresen con el formato en que la red neuronal fue creada y entrenada. Estos fueron unidos para cada día de cada NF/PF, de modo tal que las estimaciones resultantes de P_s fueran diarias y sus posteriores medidas estadísticas también.

Previo al ingreso de los parámetros a la red neuronal, se eliminan los datos tipo NaN de todos los archivos, para que en las estimaciones de Ps no estén presentes. Junto con lo anterior, se eliminan los valores de OQ y f_o fuera del rango $0 < OQ < 100\%$ y $f_o < 500$ Hz. Se debe hacer el cambio de SPL de para que su valor sea equivalente a ser medido a 10 cm de la boca ($-20 \cdot \log_{10}(0.1/0.15)$) y el cambio de ACFL de mL/s a L/s, dividiendo en 1.000 el valor extraído de los datos. El formato para la entrada a la red se muestra en la tabla 1, donde se tabulan los primeros 3 datos de cada parámetro de un día de un individuo aleatorio del grupo de control.

ACFL [L/s]	MFDR [L/s ²]	OQ [%]	SQ [%]	f_o [Hz]	H1H2 [dB]	SPL [dB]
0.080876	40.857484	63.983051	102.631579	187.292096	17.909817	62.433280
0.092224	48.274494	78.787878	112.277778	169.863707	28.006628	60.008533
0.107732	57.104335	73.333333	98.611111	169.952057	26.376793	62.614807
...	...	...	...	...	...	...

Tabla 1: Formato de los parámetros para entrar a la red neuronal.

En los archivos donde se almacenaron los datos de entrada a la red, también fue incluida, en otra columna, la medida de H1H2_{ACC}, ya que fue extraída junto con ellos. En la red no fue considerada esta medición, utilizando solo las 7 primeras medidas. La desviación estándar de H1H2_{ACC} se utilizará más tarde para los modelos de regresión logística.

5.2 Estimación presión subglótica y estadísticos

Para la estimación de Ps, se utiliza el modelo de red neuronal creado y entrenado a partir de datos sintéticos [9], donde se escogió el modelo que tuvo la mejor respuesta a la validación con datos reales medidos en laboratorio de Ps. Consistía en un modelo de dos capas ocultas con cuatro neuronas cada una.

El código utilizado para estimar Ps se observa a continuación, donde para efectos del ejemplo, está configurado para los individuos del grupo de control (NF). Cambiando NF a PF (en la lectura/escritura de los archivos .csv), se obtienen las estimaciones para el grupo patológico.

```
import time
import math
import pandas as pd
import numpy as np
import tensorflow as tf
from sklearn.metrics import r2_score
from tensorflow import keras
from tensorflow.keras import layers
from tensorflow.keras.models import load_model
```

```
#Load min-max of data synthetic used in training
train_stats =pd.read_csv('train_stats.csv')
train_stats.rename(index={0: 'ACFL', 1: 'MFDR', 2: 'OQ_mean', 3: 'sQ_mean',
                        4: 'f0', 5: 'H1H2', 6: 'spl'},inplace=True)
labels_train_stats =pd.read_csv('labels_train_stats.csv')
labels_train_stats.rename(index={0: 'Ps', 1: 'Pc', 2: 'a_TA',
                        3: 'a_CT'},inplace=True)

#Define NN architecture
Neurons=4
def build_model():
    model = keras.Sequential([
        layers.Dense(Neurons, activation='relu', input_shape=[7]),
        layers.Dropout(0.1),
        layers.Dense(Neurons, activation='relu'),
        layers.Dropout(0.1),
        layers.Dense(4)
    ])
    optimizer = tf.keras.optimizers.Adam(0.001)
    model.compile(loss='mse',
                  optimizer=optimizer,
                  metrics=['mae', 'mse'])
    return model

# Normalization function
def norm(x):
    return (x - train_stats['min']) / (train_stats['max']- train_stats['min'] )
def norm_labels(x):
    return (x - labels_train_stats['min']) /
    (labels_train_stats['max']- labels_train_stats['min'] )

# Construct and load the trained NN
model = build_model()
model = load_model('model_7F_2HL4N_50.h5',compile=False)
model.summary()

# Obtain PS, Pc, A_TA, A-CT.
for i in range(336):

    j = i+1;
    file= "NF_"+str(j)+"_por_dia.csv";
    test_data = pd.read_csv(file) #Lee el archivo para testear
```

```

ClinicData =
pd.DataFrame(test_data,
columns=['ACFL','MFDR','OQ','SQ', 'f0','H1H2', 'spl'])
ClinicData =
ClinicData.rename(columns = {'OQ': 'OQ_mean', 'SQ': 'sQ_mean'},
inplace=False)
# SPL from 15 cm of the mouth to 10cm of the mouth
ClinicData['spl']= ClinicData['spl']-20*math.log(0.1/0.15,10)
clinic_test_data = ClinicData
normed_clinic_test_data = norm(clinic_test_data)
clinic_test_predictions_1 = model.predict(normed_clinic_test_data)
NN_Prediction_norm = pd.DataFrame(clinic_test_predictions_1,
columns = ['Ps','Pc','a_TA','a_CT'])
NN_Prediction_norm['Ps']=
NN_Prediction_norm['Ps']*(labels_train_stats['max'][0]-
labels_train_stats['min'][0] ) + labels_train_stats['min'][0]
#from Pa to cm h20
NN_Prediction_norm['Ps'] = NN_Prediction_norm['Ps']/98.066
NN_Prediction_norm['Ps'].to_csv('Ps_NF_'+str(j)+'_pordia.csv')

```

Las estimaciones que entrega la red son valores de Ps medidos en Pascales, por lo que previo al proceso de almacenamiento de los datos, se hace la conversión a cm H_2O , medida más común para este parámetro. Sin ir más lejos, los resultados que entrega este predictor incluyen valores negativos, los cuales no corresponden a valores reales de Ps, por consiguiente, se debe hacer un procesamiento de los datos para eliminar los valores negativos para que al momento de calcular los estadísticos no interfieran en ellos.

Ya teniendo las estimaciones de presión subglótica sin valores negativos y en la unidad correspondiente, se procede a calcular sus estadísticos, esto cálculos se realizan con el software MATLAB. Se calculó la media, mediana, desviación estándar, asimetría, curtosis, percentil 5, percentil 95 y el rango, también se almacenó la información de los valores mínimos y máximos. Estas estadísticas fueron obtenidas para cada día de cada paciente/control, almacenándolas en un archivo por paciente, de forma que no se perdiera la información al momento de separar las parejas paciente/control en grupos de entrenamiento y testeo, para las regresiones.

5.3 Regresión logística y DPI

En primera instancia, se comienza creando el modelo de regresión logística con las 2 medidas estadísticas obtenidas en [16], la asimetría de SPL y la desviación estándar de H1-H2, medida directamente de la señal del ACC ($H1H2_{ACC}$). Es importante recalcar que para el entrenamiento del modelo se usa el 50 % de los datos y el otro 50 % para el testeo, además se consideran grupos balanceados de parejas paciente/control.

Tras obtener el primer modelo de regresión, basado en los trabajos anteriores, se comparará su rendimiento en las predicciones realizadas al grupo de testeo, en relación con los resultados del nuevo modelo propuesto en este trabajo. Para seleccionar cuál estadístico de los calculados será añadido a la red, se harán distintas pruebas. Se verificará qué medidas tienen una significancia menor al 5 % y dentro de ellos, seleccionar la de menor valor. Las pruebas aplicadas serán las mencionadas en la sección anterior, además, se caracterizará la diferencia estadística de las medidas, mediante un cálculo del tamaño del efecto ‘d’ de Cohen (Cohen’s d effect size).

Tras seleccionar la medida que indica una capacidad mayor de diferenciar a los grupos de estudio, se crea y entrena el nuevo modelo propuesto, el cual considera las medidas anteriores de asimetría de SPL y desviación estándar de H1H2, agregándole la nueva medida, obteniendo un modelo de 3 parámetros. Ambos modelos de regresión logística se crean y entrenan en MATLAB, las pruebas estadísticas también se realizan con el mismo software.

Con el código implementado y los modelos entrenados, se obtienen los coeficientes de la regresión y luego, se puede realizar una predicción de la salida para los datos del grupo de testeo. La predicción resultante corresponde a las probabilidades de pertenencia a cada uno de los grupos, donde para una $\rho < 0.5$ se clasifica como control y para $\rho \geq 0.5$ como paciente.

Con las probabilidades obtenidas se debe calcular el DPI de la siguiente manera:

$$L = \log \left(\frac{\rho}{1 - \rho} \right) \quad (5.3.1)$$

Para este análisis, donde se estudia el comportamiento vocal de individuos durante una semana, en la que no hay ninguna posibilidad de mejora, ya sea por cirugía o terapia, la comparación de los distintos valores de DPI será con fines de obtener una referencia. Es decir, la variación no corresponderá a un índice de mejoría del paciente, si no más bien, si resulta que hay una variación en el grupo de los pacientes, en la que se refleje un valor de DPI más alto, puede ser beneficioso en el futuro a una mejor diferenciación entre los pacientes y el grupo de control posterior a alguna intervención vocal.

6 Resultados

6.1 Extracción de los parámetros

En este estudio, se utilizaron los valores de cada parámetro (ACFL, MFDR, OQ, SQ, H1-H2, f_o , SPL, H1H2_{ACC}), por día de cada individuo, de los grupos paciente (PF) y control (NF). Fueron obtenidos tras procesar la señal proveniente del ACC, utilizando el esquema IBIF para las mediciones aerodinámicas y la señal pura para las acústicas. Esta información fue recopilada gracias a los trabajos previos, donde utilizaron esta misma base de datos de pacientes/controles [4]. La cantidad de datos por día de cada individuo, pueden variar entre 913 y 299.829, esta diferencia se debe a que solo se consideran los segmentos con voz de los individuos en estudio, por lo que pueden haber días con gran uso vocal y otros con menos.

Se hace una diferencia entre los parámetros correspondientes a H1-H2, porque hay uno que es estimado con IBIF, necesario como entrada a la red neuronal para estimar P_s . Mientras que el H1H2_{ACC}, obtenido directamente de la señal de acelerómetro, se debe utilizar para los modelo de regresión logística.

6.2 Estimación presión subglótica

En primer lugar, se grafican los histogramas de los valores estimados de presión subglótica, para los individuos pertenecientes del grupo de control (NF) y los pacientes (PF) (Figuras 9 y 10). Se puede apreciar que todos los datos son mayores que cero y que se consideran valores hasta los 75 cm H_2O , medidas aceptadas para la presión subglótica estimada a partir de datos ambulatorios.

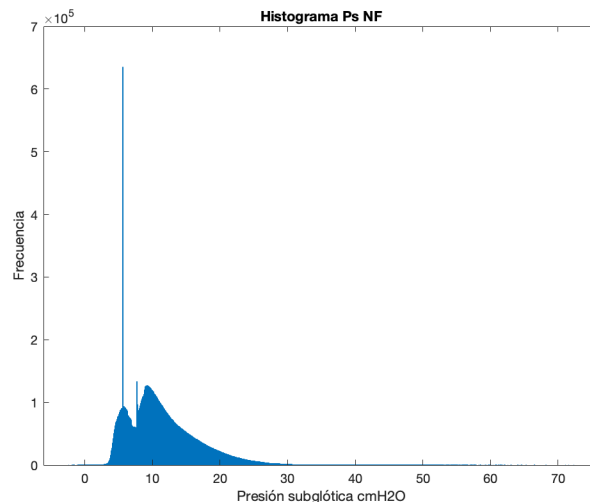


Figura 9: Histograma presión subglótica estimada para el grupo de control.

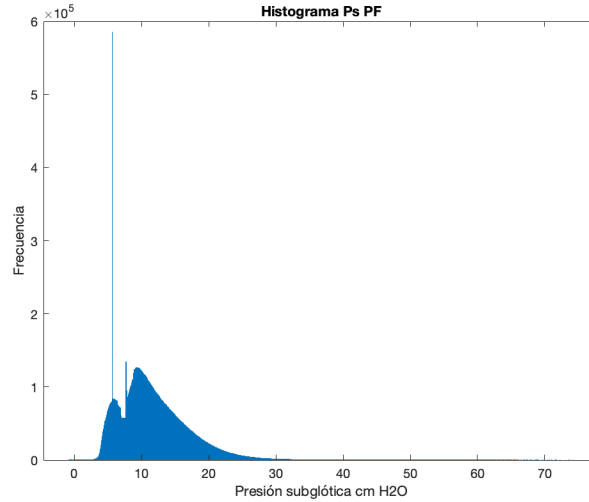


Figura 10: Histograma presión subglótica estimada para el grupo de pacientes.

Los histogramas mostrados son resultados de eliminar algunos valores obtenidos sobre 75 cm H_2O , 2 valores en el caso de los NF y 8 en los casos de los PF.

Teniendo las estimaciones de Ps dentro de un rango de valores tolerables en la práctica, se calcularon las medidas estadísticas de cada día de cada sujeto, las cuales fueron utilizadas para conocer la capacidad de discriminación de cada una entre ambos grupos, aprovechando al máximo el paradigma de control/paciente emparejado (48 pares). En la Tabla 2, se muestran los datos de la media y desviación estándar, para cada una de las medidas calculadas y para ambos grupos, junto a ellas se tabulan los valores obtenidos para el índice ‘d’ de Cohen, el cual clasifica las variables dependiendo del tamaño de la diferencia entre grupos, con lo que discrimina cuál o cuáles tienen más efecto al momento de diferenciar los grupos de estudio.

Estadística	Control	Paciente	‘d’ de Cohen
Mín	1.5698 (1.0114)	1.7226 (0.8850)	0.160814
Máx	37.3223 (10.1985)	38.0397 (11.9288)	0.064640
Media	11.1214 (2.1698)	11.4260 (2.2602)	0.137515
Mediana	10.4125 (2.1656)	10.8454 (2.2067)	0.198033
Desv. est.	4.4505 (1.0754)	4.4059 (1.2673)	-0.037931
Asimetría	0.8700 (0.4020)	0.7611 (0.3478)	-0.289588
Curtosis	4.0776 (1.3661)	3.8725 (1.0082)	-0.170832
Percentil 5	5.2258 (0.8167)	5.3790 (0.7930)	0.190321
Percentil 95	19.4556 (3.9195)	19.5152 (4.3157)	0.014464
Rango	35.7525 (10.1985)	36.3171 (12.1399)	0.050366

Tabla 2: Valores medios (desviación estándar) de las estadísticas calculadas para la presión subglótica (control y paciente) e índice de Cohen.

Se puede observar que, en general, los tamaños ‘d’ de la diferencia de los distintos valores estadísticos son bajos, siendo el que tiene el mayor valor, en magnitud, el de la asimetría, seguido por la mediana y luego el percentil 5. En la misma línea, en la Tabla 3 se tabulan los resultados de los valores p (*p-values*) de las diferentes pruebas estadísticas, donde se puede observar que la medida que tiene el menor valor p es, nuevamente, la asimetría. Siendo esta medida el parámetro escogido para integrar el nuevo modelo de regresión junto a los dos parámetros ya considerados, debido a los resultados de las evaluaciones estadísticas que se realizaron, tanto en la Tabla 2 y 3.

Estadística	Prueba KS	Prueba t	Prueba de rango con signo	Prueba de suma de rangos
Mín	0.120632	0.040153	0.023484	0.057498
Máx	0.758102	0.408801	0.939897	0.986460
Media	0.283741	0.079154	0.100023	0.125129
Mediana	0.054285	0.011569	0.012322	0.028654
Desv. est.	0.205488	0.627830	0.176028	0.274861
Asimetría	0.002716	0.000231	0.000372	0.000375
Curtosis	0.054285	0.029290	0.054779	0.242100
Percentil 5	0.010552	0.015218	0.002693	0.003275
Percentil 95	0.330055	0.853329	0.483428	0.837009
Rango	0.437040	0.519794	0.731528	0.660235

Tabla 3: Valores-p resultantes de las pruebas de significancia para valores estadísticos de presión subglótica.

6.3 Modelos de regresión logística

Para poder comparar el desempeño de los distintos modelos de regresión logística mencionados en este trabajo, se comienza creando y entrenando el primer modelo (en MATLAB), compuesto solo por los parámetros de asimetría de SPL y desviación estándar de H1-H2. Para crear la red, se utilizaron los parámetros calculados por día. El modelo resultante para la regresión logística que considera las medidas de SPL y H1-H2 se expresa en la siguiente ecuación:

$$\rho = \frac{1}{1 + e^{-(1.5674 * S_{skew} - 0.94313 * H_{std} + 5.065)}}$$

Los coeficientes del modelo fueron estimados a partir de los datos de entrenamiento, y ρ representa la probabilidad que entrega el modelo. En la Figura 11, se observa la matriz de confusión resultante de la clasificación del grupo de testeo.

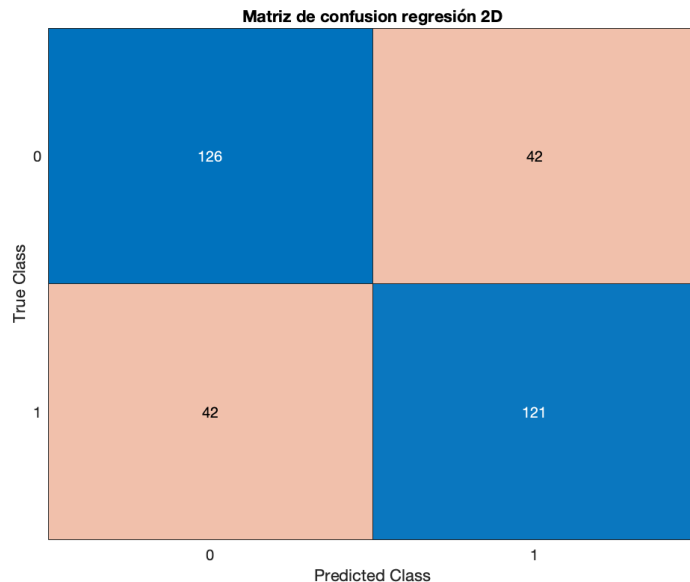


Figura 11: Matriz de confusión modelo de 2 parámetros.

En la matriz se puede observar que la clasificación correcta es mayor (cuadros azules) que los resultados falsos (cuadros rosados), obteniendo una exactitud (accuracy) de 74.62 %, el porcentaje de la clasificación correcta total. Por otra parte, también se obtuvo una sensibilidad de 74.23 %, correspondientes a los que fueron clasificados correctamente como pacientes y una especificidad de 75 %, correspondientes a los clasificados correctamente como control. Finalmente, el área bajo la curva característica (AUC) tiene un valor de 0.8338.

Siguiendo con los resultados de la primera clasificación, en la Figura 12 se grafica la desviación estándar H1-H2 frente a la asimetría de SPL para ilustrar el rendimiento del modelo de dos variables en la clasificación de cada sujeto en función de los datos diarios para el grupo de testeo. En color se representa la magnitud de la presión subglótica para los puntos correspondientes, se representan también con formas distintas los grupos de control/pacientes. La recta representa la regresión logística, la cual muestra la separación de ambos grupos. A simple vista, se observa que los sujetos pertenecientes al grupo de pacientes presentan valores más bajos de asimetría de Ps (tonalidades azules), que los pertenecientes al grupo de control.

La Figura 12 hace referencia a la misma obtenida en [16] (Figure 2), donde se tuvo una exactitud de 76.67 % y un AUC de 0.843, resultados cercanos a los obtenidos en esta clasificación que utiliza los mismos parámetros, pero con distinta base de datos. Entre ambos modelos se obtiene un error absoluto de 0.02 en la exactitud, 0.009 en sensibilidad y 0.05 en especificidad, por lo que se puede decir que el comportamiento de ambos modelos son similares.

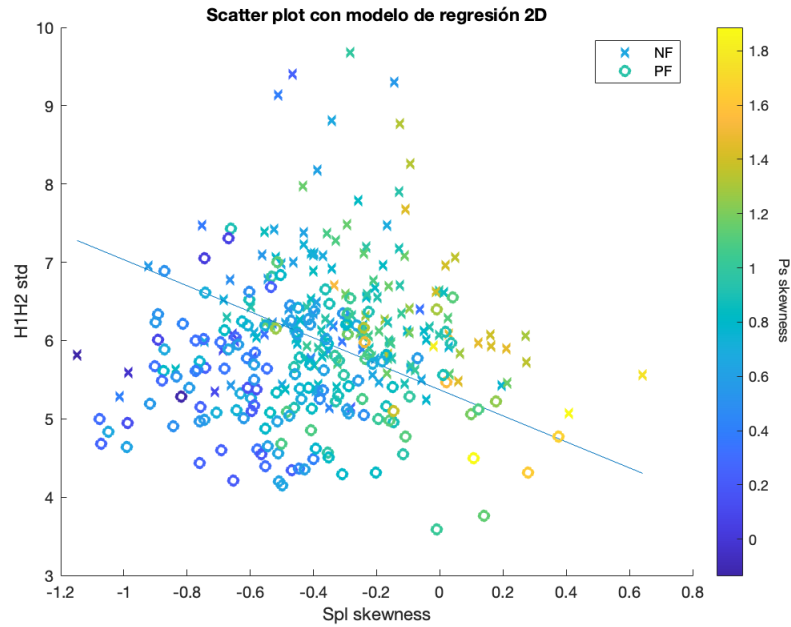


Figura 12: Gráfico de dispersión de la desviación estándar H1-H2 versus la asimetría del nivel de presión sonora, para el modelo de 2 dimensiones.

Por otro lado, el nuevo modelo de regresión logística, creado con la medida estadística de P_s seleccionada previamente, se representa con una ecuación de probabilidad distinta. Esta ecuación contempla los tres parámetros, los dos del modelo anterior y el de P_s , cuyos coeficientes también fueron calculados a partir de los datos de entrenamiento. Esta nueva ecuación queda representada en:

$$\rho = \frac{1}{1 + e^{-(2.2445 * S_{skew} - 0.9311 * H_{std} + 0.831846 * P_{skew} + 4.0933)}}$$

A partir de esta ecuación se obtienen nuevas predicciones para el grupo de testeo. Los resultados de esta nueva clasificación se observan en la Figura 13, donde se puede observar una mejora en la clasificación de los pacientes, mientras que los del grupo de control permanecen igual. Las métricas calculadas para este modelo corresponden a una exactitud de 77.95 %, una sensibilidad de 80.98 % y una especificidad de 75 %. Se observa que en general, con la medición de la exactitud y sensibilidad, la clasificación mejora en comparación al modelo de dos dimensiones. Sin embargo, con la métrica de especificidad no se observa ninguna variación, manteniendo la misma proporción en la clasificación verdadera y falsa. En este caso, el AUC tiene un valor de 0.8460, lo cual también representa una mejora en el rendimiento de la clasificación.

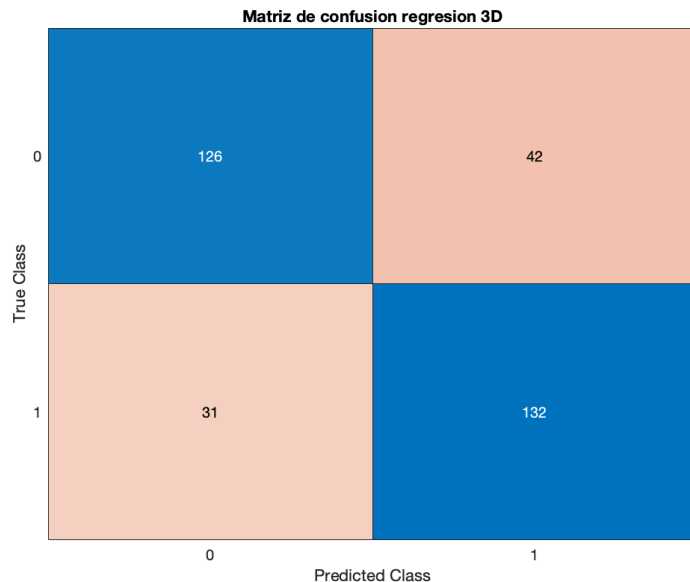


Figura 13: Matriz de confusión modelo de tres parámetros.

Al ser un modelo de 3 dimensiones, cuya clasificación queda condicionada por la relación entre ellas, se complica el graficarlas junto a su hiperplano de corte y analizarlas en un plano tridimensional. Es por esto que para analizar el comportamiento de este modelo, se analizarán dos variables, dejando presente la tercera como un parámetro de color. Este análisis se realizará de tal forma que se logre observar el comportamiento de todas las medidas entre ellas.

En primer lugar, se tiene el gráfico de dispersión de la Figura 14, que representa la relación entre la desviación estándar de H1-H2 y la asimetría de SPL, en color, el valor de la asimetría de presión subglótica. Para graficar esta relación, la recta de regresión es la obtenida para el modelo de tres dimensiones considerando $P_s = 0$. A simple vista se puede observar que se logran diferenciar los grupos de manera similar al primer caso, donde también se consideraron estos mismos parámetros. Con color se puede ver la amplitud de la asimetría de P_s , observándose el mismo efecto de antes, que los patológicos presentan niveles más bajos (tonos azules). La relación entre estas dos variables sí aporta a la clasificación, notándose una separación entre ambos grupos (paciente/control). Asimismo, es posible observar que las tonalidades azules de los pacientes se inclinan a estar en los valores más bajos de asimetría de SPL, vislumbrando cierta relación entre ellas.

Un nuevo análisis, diferente al de los dos parámetros anteriores, es el que se puede observar en la Figura 15, donde se grafican los puntos de dispersión asociados a la asimetría de P_s y la desviación estándar de H1-H2. Se observa en la Figura 15 una recta asociada al modelo de regresión, considerando $SPL = 0$ y expresando su amplitud en

colores.

Se aprecia, nuevamente, que el nivel de asimetría de SPL se relaciona con la asimetría de Ps, debido a que, para este caso, los sujetos pertenecientes al grupo de pacientes también presentan tonalidades más azules que los NF. Casos aislados como algunos pacientes de color amarillo o controles de color azul verdoso, igualmente se observan en el gráfico de la Figura 14. Sin embargo, en este gráfico, la recta de regresión no abarca tantos individuos patológicos al momento de realizar la clasificación, no obstante, si separa parte del grupo. Se puede distinguir un grupo que claramente tiene más presencias de patológicos sobre la recta. Observando la Figura 15, se puede decir que la relación entre estas variables analizadas (asimetría de Ps vs desv. estándar de H1-H2), igualmente aporta a la separación de ambos grupos de estudio en el modelo tridimensional.

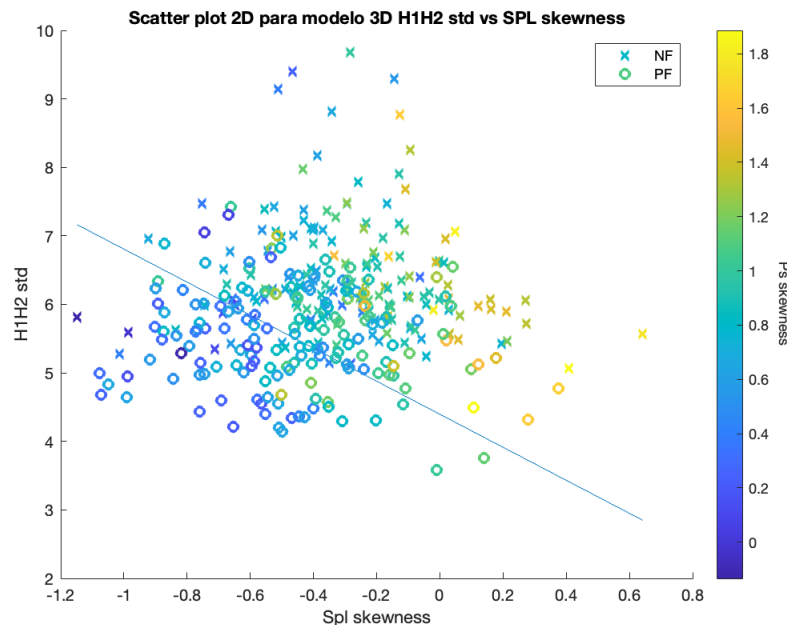


Figura 14: Gráfico de dispersión de la desviación estándar de H1-H2 versus la asimetría de SPL para el modelo de 3 dimensiones con $P_s = 0$.

Por otra parte, y para terminar con el análisis del modelo de tres dimensiones, estudiando la relación de dos parámetros, se tiene el gráfico de la Figura 16, que considera la asimetría de Ps versus la de SPL. Medidas que, como se mencionó anteriormente, y según lo observado en las Figuras 14 y 15, tienen un grado de relación. Esta relación se confirma al observar la Figura 16, donde al graficar la recta de regresión con $H1-H2 = 0$, se aprecia que la recta no corta al grupo de individuos. Por lo demás, estos están muy agrupados, no visualizándose una separación entre ellos. Es debido a esta relación, confirmada en trabajos anteriores (correlación entre SPL y Ps) [5], que no se puede observar una diferencia entre ambos grupos para este caso.

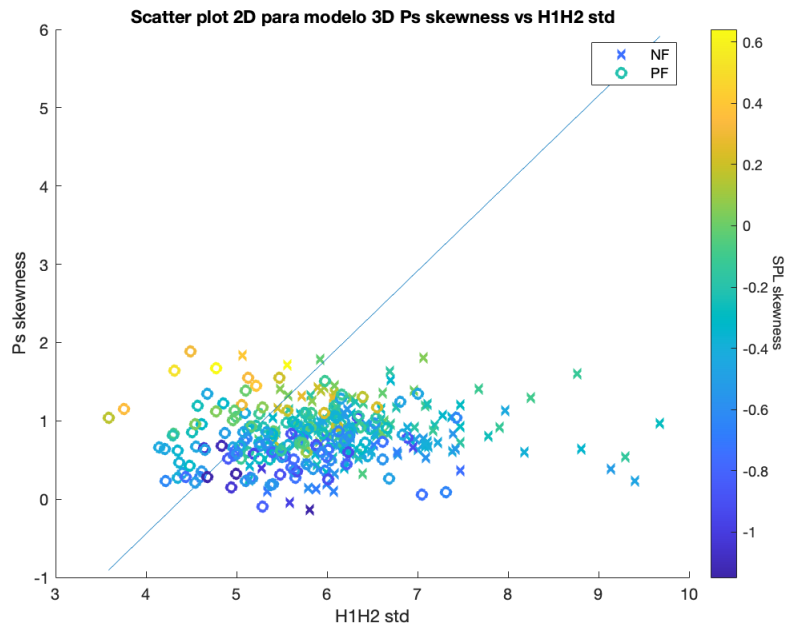


Figura 15: Gráfico de dispersión de la asimetría de Ps versus la desviación estándar de H1-H2 para el modelo de 3 dimensiones con $SPL = 0$.

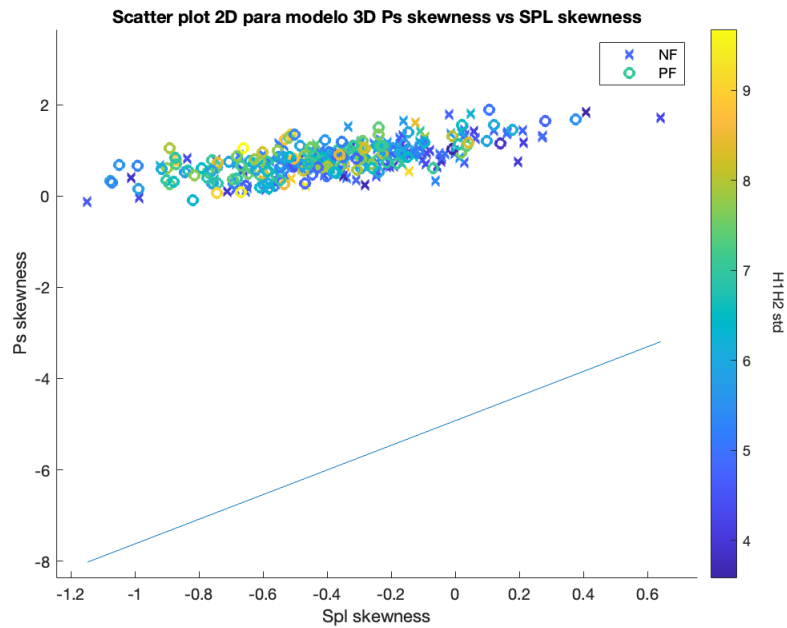


Figura 16: Gráfico de dispersión de la asimetría de Ps versus la asimetría de SPL para el modelo de 3 dimensiones con $H1-H2 = 0$.

El uso de la medida de asimetría de Ps junto con la de SPL, hace que este modelo de 3 dimensiones no tenga un resultado con una mayor diferencia al momento de clasificar. Sin embargo, viendo las métricas analizadas, se observa que el agregar el nuevo parámetro aporta con mejorar la clasificación de individuos patológicos, lo que se refleja al momento de analizar la Figura 15, relación que se incorpora al modelo previamente analizado. Ambos parámetros relacionados entre sí, en conjunto con H1-H2 aportan a la clasificación de forma independiente.

Pese a lo anterior, en la Figura 17 se puede observar un gráfico en tres dimensiones, junto con el hiperplano formado a partir del modelo de regresión logística, el mismo que fue analizado en los gráficos de dos dimensiones haciendo cero una de las variables. La dificultad de analizar este gráfico es que no se puede identificar de buena forma el comportamiento de los grupos, por ende, se decidió realizar los 3 gráficos de dos dimensiones. Sin embargo, se logra observar que los grupos pueden ser efectivamente separados en las 3 dimensiones, de lo contrario el gráfico se vería con los puntos correspondientes a control y pacientes amalgamados y el hiperplano no los separaría, tal como ocurrió en el gráfico de la Figura 16. Este gráfico es para visualizar la capacidad de separar los grupos con el modelo propuesto.

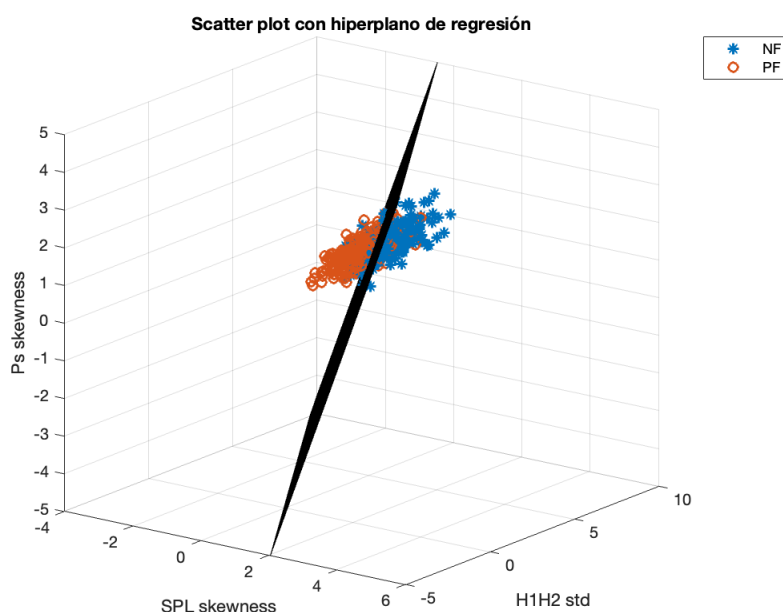


Figura 17: Gráfico de dispersión de tres dimensiones con hiperplano obtenido del modelo de regresión (asimetría Ps eje Z).

En la misma línea, se grafican las Figuras 18 y 19, que muestran otras dos vistas distintas a la Figura 17, con la misma finalidad de visualizar la separación que produce el hiperplano del modelo, cambiando la variable del eje Z.

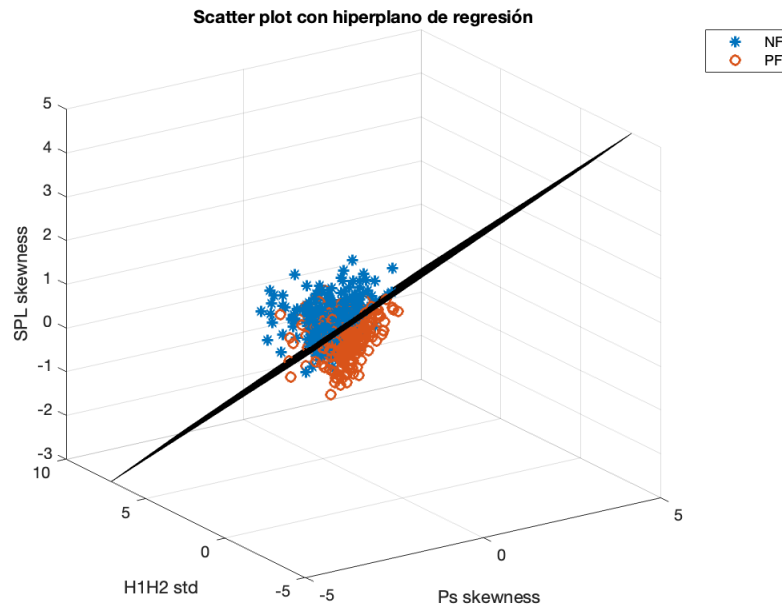


Figura 18: Gráfico de dispersión de tres dimensiones con hiperplano obtenido del modelo de regresión (asimetría SPL eje Z).

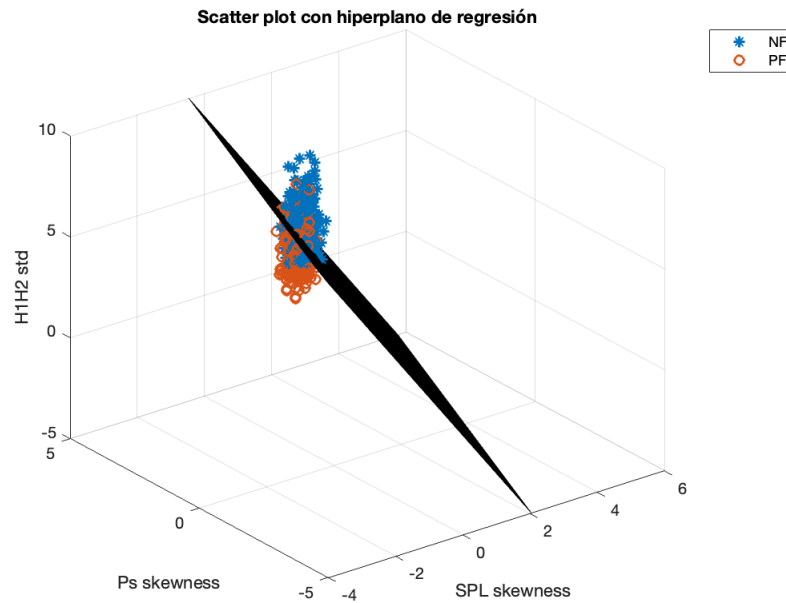


Figura 19: Gráfico de dispersión de tres dimensiones con hiperplano obtenido del modelo de regresión (desv. estándar H1-H2 eje Z).

6.4 DPI

Una vez resuelto el problema de clasificación, se continúa con el desafío siguiente, que es la obtención del DPI. Los valores obtenidos para el DPI corresponden a la transformación logit realizada a las probabilidades calculadas a partir de los modelos creados. Estos valores corresponden a las probabilidades de clasificación del grupo de testeo.

En las Figuras 20 y 21, se observan los histogramas del DPI obtenido con el modelo de 2 dimensiones, junto al DPI obtenido con el modelo de 3 dimensiones, para el grupo de control y de pacientes, respectivamente. Se muestran superpuestos los valores, para poder analizar las diferencias presentes entre ellos y cómo sus valores pueden afectar la clasificación o diferenciación de ambos grupos.

Como se mencionó previamente, lo que se analiza del DPI es su variación tras someterse a algún procedimiento externo que tenga relación con mejorar su condición vocal. Sin embargo, en este trabajo se consideró una semana de estudio de los individuos, donde ninguno asistió a ninguna cita clínica, por lo que los valores obtenidos corresponden a valores iniciales que se obtienen en las investigaciones del DPI donde se estudia su variación. Es por lo mencionado, que los resultados de la clasificación tienen directa relación a los cambios de comportamiento con el DPI, sin ir más lejos, su propio valor depende de las probabilidades de clasificación.

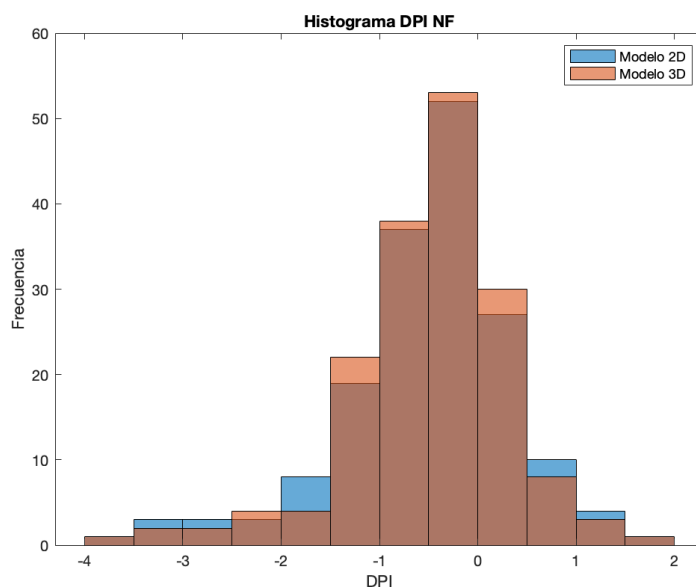


Figura 20: Histograma valores DPI para sujetos del grupo de control (valor diario).

En los estudios previos se asocian los valores negativos de DPI a comportamientos vocales normales y los positivos a la presencia de alguna patología [17]. En la Figura 20 se observa

que hubo algunos cambios en valores negativos y positivos, algunos que estaban en los extremos se movieron al centro, pero manteniendo la proporción en lo relativo al signo. Es por este motivo, que no hubo variación en la clasificación de los sujetos pertenecientes al grupo de control.

Por el contrario, en el caso del DPI para el grupo de pacientes, la subida de los valores a positivos influyó en un aumento en la clasificación de patológicos, lo que se puede verificar observando que las barras positivas aumentaron su valor de frecuencia. El mayor cambio se puede observar en los valores que pasaron del rango $[-0.5 \ 0]$ a $[0 \ 0.5]$, cambio que favorece a la asignación de sujetos patológicos por ser DPI positivo.

Esto indica que, utilizando el modelo de 3 dimensiones para la clasificación, se pueden obtener resultados DPI más elevados, los cuales pueden ser útiles al momento de compararlos con resultados posteriores a tratamientos vocales, donde su resultado debería tender hacia valores negativos.

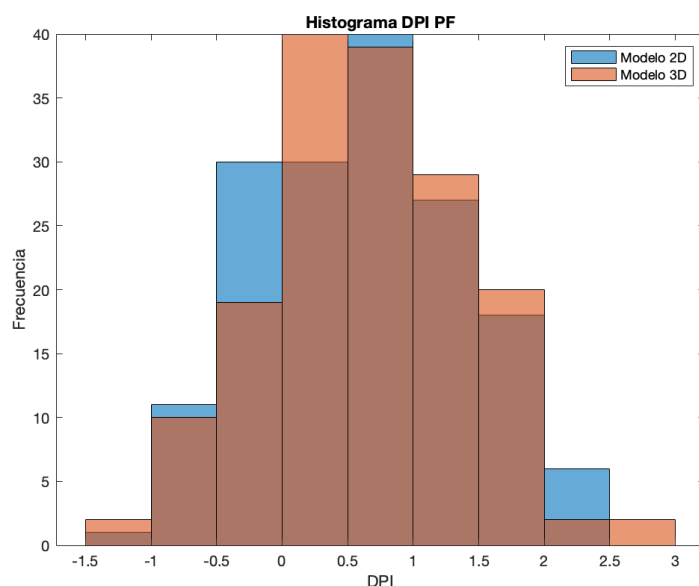


Figura 21: Histograma valores DPI para sujetos del grupo de patológicos (valor diario).

A modo global, en la Figura 22 se representan los valores medios del DPI por sujeto, pertenecientes al grupo de control/paciente. Se observa que hay un aumento mayor por parte de los individuos patológicos en el DPI calculado a partir del modelo de 3 dimensiones, mientras que en el grupo de control los valores permanecen más bien fijos, salvo algunos casos que variaron, pero dentro del rango de comportamiento normal. La recta graficada corresponde a la relación lineal del DPI obtenido del primer modelo analizado, solo con dos dimensiones.

Si bien esta variación no es tan significativa en magnitud, se debe destacar la mejora en el comportamiento del DPI para el caso del grupo de pacientes, donde la variación fue mayor que en el caso del grupo de control. También, es posible notar que hubo un cambio de valores DPI negativos a positivos, mientras que otros positivos aumentaron su valor, dentro del grupo de pacientes.

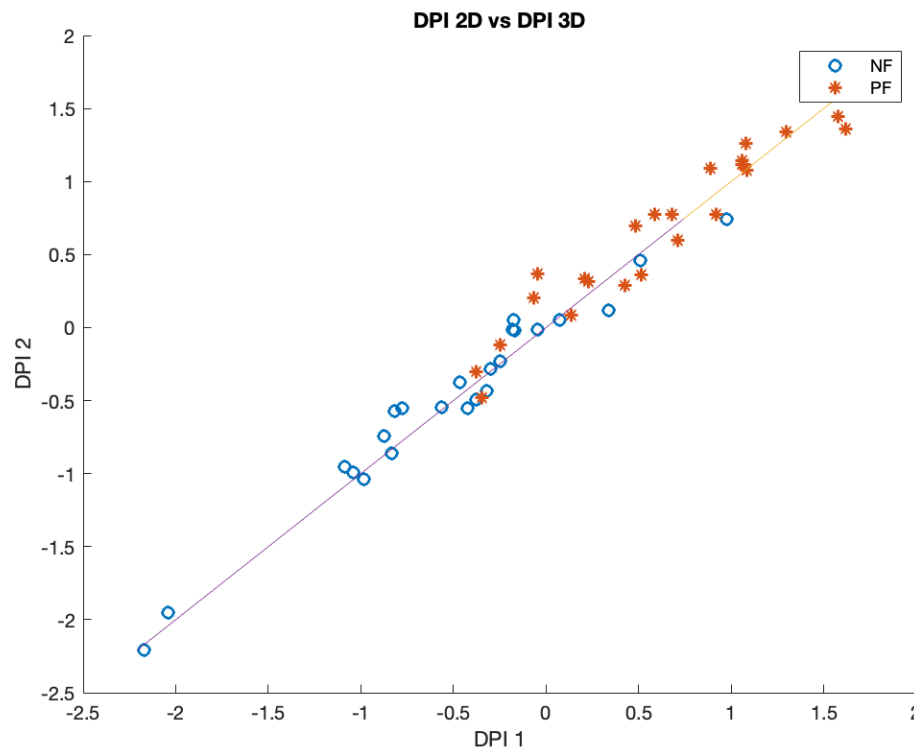


Figura 22: Gráfico de dispersión valores DPI, comparación resultados DPI para modelo de 3 dimensiones (DPI 2) versus el de 2 dimensiones (DPI 1).

7 Conclusiones

Finalizando con este trabajo, es importante destacar que el uso de señales utilizadas en los trabajos previos, facilitó la obtención de los parámetros aerodinámicos y las medidas propias de la señal proveniente del acelerómetro parte del MVHM. Estas fueron de gran importancia al momento de realizar la estimación de la presión subglótica, que se realizó a partir de la red neuronal, y para posteriormente, realizar los modelos de regresión logística que permitieron la clasificación de los individuos de ambos grupos de estudio.

Al comparar los modelos creados en este trabajo, el modelo de regresión de tres dimensiones propuesto, logró una mejor clasificación de los sujetos pertenecientes al grupo de pacientes respecto al modelo base de dos dimensiones, con una sensibilidad de 80.98 %, por sobre a la obtenida en el modelo base que tuvo una sensibilidad de 74.23 %. Sin embargo, se mantuvo igual en el caso de la clasificación de los sujetos del grupo de control, obteniendo una especificidad de 75 % en ambos. No obstante, el resultado que evalúa la clasificación correcta de ambos grupos y el desempeño del modelo corresponde a la exactitud, la cual también obtuvo un valor mayor, de 77.95 %, superando al valor de 0.7462 obtenido en el modelo base. Otra métrica evaluada fue el área bajo la curva, con la cual, de igual manera, se puede ver una mejora en el nuevo modelo, obteniendo un AUC de 0.8460, en comparación con el del modelo base de 0.8338.

En relación con la variación positiva obtenida para el grupo de pacientes, esta no fue mayor debido a la alta correlación existente entre la medición de SPL y de la presión subglótica, razón por la que en la Figura 16, que mostraba el aporte de ambas en la clasificación a partir del nuevo modelo, esta recta no cortaba la agrupación de puntos, y tampoco se podían distinguir de buena forma las dos clases. Si bien, la correlación entre esas mediciones es alta, no es lo suficientemente alta para que al agregar la medida de asimetría de Ps la clasificación no cambie. Esto porque se observa un aumento en la identificación de patológicos, concluyendo que la relación entre la asimetría de Ps con la desviación estándar de H1-H2 sí contribuyó a la mejora en la clasificación.

Por otro lado, paralelamente a lo expuesto en este trabajo, se probaron los parámetros que le seguían a la asimetría de Ps, según los valores p, obteniendo una clasificación peor para ambos casos. Las pruebas establecían, que la asimetría era la medición que mejor separaba ambos grupos, analizando solo estadísticos del parámetro de presión subglótica.

Considerando los resultados del DPI, con una variación de los valores obtenidos con el nuevo modelo, se ve favorecido el reconocimiento de sujetos del grupo de pacientes. Esto puede ser utilizado en una investigación futura, que permita analizar los comportamientos vocales previos o posteriores a cirugías o tratamientos para dicho grupo. Para este caso, se comprobó que hubo un aumento de los valores positivo en ellos, por lo que estos pueden ser utilizados para definir la intervención que necesite el paciente y su duración.

Referencias

- [1] Alzamendi, G., Manríquez, R., Hadwin, P., Deng, J., Peterson, S., Erath, B., et al. (2020). Bayesian estimation of vocal function measures using laryngeal high-speed videoendoscopy and glottal airflow estimates: An in vivo case study. *The Journal of the Acoustical Society of America* 147, EL434–EL439. doi:10.1121/10.0001276
- [2] Alzamendi, G.A., Peterson, S.D., Erath, B.D., Hillman, R.E., and Zañartu, M. (2021). Triangular body-cover model of the vocal folds with coordinated activation of five intrinsic laryngeal muscles with applications to vocal hyperfunction ArXiv preprint arXiv:2108.01115
- [3] Bianco, M.J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M.A., Gannot, S., et al. (2019). Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America* 146, 3590–3628. doi:10.1121/1.5133944
- [4] Cortés J.P., V.M. Espinoza, M. Ghassemi, D.D. Mehta, J.H. Van Stan, R.E. Hillman, J.V. Gutttag, M. Zañartu: Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration. *Plos one*, 13(12):1–22, 2018.
- [5] Espinoza Catalán V., Zañartu M., Van Stan J. H., Mehta D.D., Hillman R. E.: Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research*, 60(8):2159–2169, Aug 2017.
- [6] Espinoza, V.M., Mehta, D.D., Stan, J.H.V., Hillman, R.E., and Zañartu, M. (2020). Glottal aerodynamics estimated from neck-surface vibration in women with phonotraumatic and nonphonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research* 63, 2861–2869. doi:10.1044/2020_JSLHR-20-00189
- [7] Ghassemi, M., Van Stan, J.H., Mehta, D.D., Zañartu, M., Cheyne, H.A. II, Hillman, R.E., et al. (2014). Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: initial results for vocal fold nodules. *IEEE Trans. Biomed. Eng.* 61, 1668–1675. doi: 10.1109/TBME.2013.2297372
- [8] Hillman, R. E., Stepp, C. E., Van Stan, J.H., Zañartu, M., Mehta, D.D. (2020). An Updated Theoretical Framework for Vocal Hyperfunction. *American journal of speech-language pathology*, 29(4), 2254–2260. https://doi.org/10.1044/2020_AJSLP-20-00104
- [9] Ibarra E.J., Parra J., Alzamendi G.A., Cortés J.P., Espinoza V.M., Mehta D.D., Hillman R. E., and Zañartu M.: Estimation of subglottal pressure, vocal fold collision pressure, and intrinsic laryngeal muscle activation from neck-surface vibration using

- a neural network framework and a voice production model. *Frontiers in physiology* vol. 12:732244. 2021.
- [10] Llico, A.F., Zañartu, M., Gonzalez, A.J., Wodicka, G.R., Mehta, D.D., Van Stan, J.H., et al. (2015). Real-time estimation of aerodynamic features for ambulatory voice biofeedback. *J. Acoust. Soc. Am.* 138, EL14–EL19. doi: 10.1121/1.4922364
- [11] Mehta D.D. , M. Zañartu, S.W. Feng H.A. Cheyne R.E. Hillman: Mobile voice health monitoring using wereable accelerometer sensor and a smartphone platform. *IEE Trans. Biomed. Eng.*, 59(11):3090–3096, November 2012.
- [12] Mehta, D.D., Van Stan, J.H., Zañartu, M., Ghassemi, M., Guttag, J.V., Espinoza, V.M., Cortés, J.P., Cheyne, H.A., 2nd, Hillman, R.E. (2015). Using Ambulatory Voice Monitoring to Investigate Common Voice Disorders: Research Update. *Frontiers in bioengineering and biotechnology*, 3, 155. <https://doi.org/10.3389/fbioe.2015.00155>
- [13] Roy N., Merrill R.M., Gray S.D., and Smith E.M., “Voice disorders in the general population: Prevalence, risk factors, and occupational impact,” *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, 2005.
- [14] Titze, I.R., Svec, J.G., and Popolo, P.S. (2003). Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues. *J Speech Lang. Hear. Res.* 46, 919–932. doi: 10.1044/1092-4388(2003/072)
- [15] Van Stan, J.H., Mehta, D.D., Zeitels, S.M., Burns, J.A., Barbu, A.M., Hillman, R.E. (2015). Average Ambulatory Measures of Sound Pressure Level, Fundamental Frequency, and Vocal Dose Do Not Differ Between Adult Females With Phonotraumatic Lesions and Matched Control Subjects. *The Annals of otology, rhinology, and laryngology*, 124(11), 864–874, Nov 2015.
- [16] Van Stan J.H., D.D. Mehta, A.J. Ortiz J.A. Burns K.L. Marks L.E. Toles M. Vangel T. Hron S.M. Zeitels R.E. Hillman: Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions and matched controls. *Journal of speech, Language, and Hearing research*, 63:372–384, February 2020.
- [17] Van Stan J.H., Mehta D. D., Ortiz A.J., Burns J.A., Marks K.L., Toles L.E., Stadelman-Cohen T., Krusemark C. , Muise J., Hron T. , Zeitels S.M., Fox A.B., Hillman R.E.: Changes in a daily phonotrauma index after laryngeal surgery and voice therapy: Implications for the role of daily voice use in the etiology and pathophysiology of phonotraumatic vocal hyperfunction. *Journal of speech, Language, and Hearing research*, 63:3934–3944,2020.
- [18] Van Stan, J.H., Ortiz, A.J., Cortes, J.P., Marks, K.L., Toles, L.E., Mehta, D.D., et al.

- (2021). Differences in daily voice use measures between female patients with nonpho-
notraumatic vocal hyperfunction and matched controls. *J. Speech Lang. Hear. Res.*
64, 1457–1470. doi: 10.1044/2021_JSLHR-20-00538
- [19] Zañartu M., Ho J. C., Mehta D. D., Hillman R. E., and Wodicka G. R., “Sub-
glottal impedance-based inverse filtering of speech sounds using neck surface ac-
celeration,” *IEEE Trans. Audio Speech Lang. Process.* 21(9), 1929–1939, 2013. doi:
10.1109/TASL.2013.2263138