



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO Y EDUCACIÓN CONTINUA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**PREDICCIÓN DEL ENTENDIMIENTO DEL HABLA EN RUIDO EN
PACIENTES CON IMPLANTES COCLEARES USANDO SEÑALES DE
ELECTROCOCLEOGRAFÍAS**

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

GABRIELA ISIDORA MORA MOYA

PROFESOR GUÍA:
SEBASTIÁN RÍOS PÉREZ

PROFESOR CO-GUÍA:
PAUL DÉLANO REYES

COMISIÓN:
JORGE SILVA SÁNCHEZ

SANTIAGO DE CHILE

2024

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS Y MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL
POR: GABRIELA ISIDORA MORA MOYA
FECHA: 2024
PROF. GUÍA: SEBASTIÁN RÍOS PÉREZ
PROF. CO-GUÍA: PAUL DÉLANO REYES

PREDICCIÓN DEL ENTENDIMIENTO DEL HABLA EN RUIDO EN PACIENTES CON IMPLANTES COCLEARES USANDO SEÑALES DE ELECTROCOCLEOGRAFÍAS

Si bien los implantes cocleares permiten restaurar la audición de pacientes con sordera, se ha observado una notable variabilidad en la capacidad de comprensión del habla, sobre todo con ruido de fondo. Esta disparidad representa un desafío particular para aquellos que experimentan resultados deficientes, ya que puede afectar diversos aspectos de su vida. Por ende, resulta crucial identificar los factores que inciden en esta variabilidad para poder predecir de mejor forma el desempeño auditivo y comprender su relación con la variable en cuestión. En este contexto, abordar este problema desde la perspectiva de la ingeniería, con un enfoque específico en el uso de herramientas de ciencia de datos, emerge como una opción lógica. La ciencia de datos ofrece una gama de técnicas para desarrollar modelos predictivos, lo cual resulta especialmente pertinente para comprender y anticipar la variabilidad en la comprensión del habla dentro de este grupo de pacientes.

De esta forma el siguiente trabajo de tesis aborda el diseño de una herramienta que permite definir niveles de comprensión del habla mediante modelos que puedan contribuir al desarrollo de mejores estrategias para abordar los desafíos en el tratamiento de pacientes que tienen malos desempeños en la comprensión del habla en entornos ruidosos después de la implantación.

La herramienta se ha desarrollado utilizando un conjunto de datos de señales de electrocoocleografías registrados de pacientes implantados en el Hospital Clínico de la Universidad de Chile. Este proyecto ha sido realizado considerando el conocimiento en procesamiento de señales, así como el expertise en el campo de la ciencia de datos, con el objetivo de implementar un enfoque de modelado integral que incluye tanto la regresión como la clasificación.

En base a las herramientas desarrolladas, se evalúan diferentes representaciones de la señal como también diferentes modelos según el enfoque de clasificación y de regresión, donde la representación y modelo seleccionado para la predicción del entendimiento del habla en ruido, resultó ser un modelo de red neuronal en el cual todas las representaciones estudiadas lograron el mismo desempeño con un MAPE¹ de un 79 % en contraste con el mejor resultado del estado del arte de un MAPE de 167 %.

¹ El MAPE (Mean Absolute Percentage Error) es una métrica utilizada para evaluar el desempeño de un modelo de predicción o pronóstico. Mide el promedio de los errores porcentuales absolutos entre las predicciones y los valores reales.

A mi familia, los amo mucho.

Agradecimientos

Quiero expresar mi más sincero agradecimiento a quienes han sido pilares fundamentales en mi camino hacia la culminación de esta etapa académica.

En primer lugar, a mi familia, cuyo apoyo incondicional ha sido esencial en mi trayectoria. A mis padres, les agradezco profundamente por brindarme la tranquilidad necesaria para concentrarme en mis estudios, sin preocupaciones adicionales. Cada acción que han emprendido en mi beneficio refleja su amor incondicional. Papá, agradezco enormemente tus esfuerzos incansables para asegurarte de que nunca nos falte nada. Recuerdo con gratitud una vida sin carencias gracias a ti. Mamá, con tu dedicación absoluta en nuestra crianza, has dejado una huella imborrable de ternura y entrega hacia quienes amamos. Los llevo en mi corazón con profundo amor y gratitud.

A mis hermanos, quienes han sido mi constante apoyo a lo largo de este viaje. A Dani, por su calidez reconfortante que siempre ilumina mi espíritu cuando estamos juntas. A Coni, por su compañerismo y orientación que han estado presentes desde mi infancia. Y a mi querido Dieguito, por su alegría contagiosa y risas que siempre llenan nuestro hogar de felicidad. Los amo a todos profundamente.

También quiero expresar mi gratitud a aquellos que ya no están físicamente entre nosotros. A mi tata, quien siempre estuvo a mi lado desde mi infancia, y aunque ahora solo me acompañe su recuerdo, este me brinda ánimo en momentos de fatiga. Asimismo con mi tío Rodrigo, quien sé que habría sentido una inmensa alegría al saber que su sobrina, con quien compartió tantas experiencias y lazos afectivos, ha culminado su carrera universitaria.

Por último, deseo expresar mi gratitud hacia todas las personas que me acompañaron intelectualmente durante este proceso. Mi más profundo agradecimiento a mi profesor guía, Sebastián Ríos, cuya orientación siempre fue ejemplar, guiándome con respeto y compartiendo sus vastos conocimientos. Asimismo, deseo expresar mi agradecimiento a los profesores Jorge Silva y Paul Délano por su invaluable ayuda en las áreas que son de su especialización, en las cuales carecía de experiencia. También quiero reconocer a Javiera Herrada y Vicente Medel, quienes en las primeras etapas del proceso demostraron disposición para resolver mis dudas sin esperar nada a cambio. Finalmente, agradezco al Centro AC3E por su apoyo en campos que no dominaba, permitiéndome dar coherencia y profundidad a mi investigación.

Tabla de Contenido

1. Introducción	1
1.1. Motivación e hipótesis	1
1.2. Fundamentos del problema	4
1.3. Objetivos del trabajo	5
2. Metodología, conceptos y teoría	7
2.1. Metodología	7
2.2. Conceptos y teoría	8
2.3. El oído	9
2.3.1. Partes de oído	9
2.3.2. Trasmisión del sonido al cerebro	10
2.4. Implante coclear	10
2.4.1. Contexto e historia	11
2.4.2. Transmisión del sonido con implante coclear	13
2.4.3. Electrocoqueografía (ECoG)	13
2.5. Procesamiento de señales	14
2.5.1. Qué es una señal	14
2.5.2. Trasformada de fourier de tiempo discreto	15
2.5.3. Identidad de Parseval	16
2.5.4. Señal evocada	17
2.5.5. Señal inducida	17
2.6. Conceptos de las electrocoqueografías	17
2.6.1. Potenciales de electrocoqueografíos (ECoG) evocados	17
2.6.2. Actividad espontánea del nervio auditivo	19
2.7. Estado del arte	20
2.8. Aprendizaje de máquinas	22
2.8.1. Contexto	22
2.8.2. Modelos discriminativos	23
2.8.3. Modelos generativos	35
2.8.4. Métricas de desempeño	36
3. Resultados	40
3.1. Comprensión de los datos	40
3.2. Análisis y selección de datos	42
3.2.1. Análisis de los datos	42
3.2.2. Selección de los datos	43
3.3. Preprocesamiento de los datos	43

3.3.1.	Representación de la señal	43
3.3.2.	Feature engineering	43
3.4.	Modelamiento	49
3.4.1.	Modelos discriminativos	49
3.4.2.	Modelo generativo	52
3.5.	Evaluación	53
3.5.1.	Evaluación de los modelos de regresión	53
3.5.2.	Evaluación de los modelos de clasificación	59
3.5.3.	Evaluación de modelo generativo	64
4.	Conclusiones	66
	Bibliografía	68
	Anexos	71
A.	Señal de electrocolceografía evocada a 500 Hz de todos los pacientes	71
B.	Señal de electrocolceografía evocada a 2000 Hz de todos los pacientes	72
C.	Microfonico coclear a 500 Hz	73
D.	Potencial de acción compuesto a 500 Hz	74
E.	Microfónico coclear a 2000 Hz	75
F.	Potencial de acción compuesto a 2000 Hz	76

Índice de Tablas

1.1.	Presentación de objetivos.	6
2.1.	Criterios de los niveles de sordera según decibeles	11
2.2.	Criterios de implante coclear en la población adulta según FDA	12
2.3.	Benchmark de las investigaciones sobre el entendimiento del habla	20
3.1.	Cantidad de dataframes para probar modelos	50
3.2.	Resultados de MAPE en silencio a 500 Hz	54
3.3.	Resultados de MAPE en silencio a 2000 Hz	55
3.4.	Resultados de MAPE en silencio sumando ambas respuestas	56
3.5.	Resultados de MAPE en ruido a 500 Hz	57
3.6.	Resultados de MAPE en ruido a 2000 Hz	58
3.7.	Resultados de MAPE en ruido sumandos ambas respuestas	59
3.8.	Resultados de Recall Macro en silencio a 500 Hz	60
3.9.	Resultados de Recall Macro en silencio a 2000 Hz	61
3.10.	Resultados de Recall Macro en silencio sumandos ambas respuestas	62
3.11.	Resultados de Recall Macro en ruido a 500 Hz	63
3.12.	Resultados de Recall Macro en ruido a 2000 Hz	63
3.13.	Resultados de Recall Macro en ruido sumandos ambas respuestas	64

Índice de Ilustraciones

1.	Esquema del sistema auditivo en humanos. Se muestran ilustraciones de las células ciliadas de la cóclea (órgano receptor auditivo) y como estas se van dañando a lo largo de la vida. Dr Pathak, S. Age Related hearing Loss: Presbycusis. Reproducido de: https://www.drshrehapathakent.com	2
2.	Fases CRISP-DM. Reproducido de: https://healthdataminer.com	8
3.	Imagen de partes del oído. Reproducido de : https://www.childrensmn.org	10
4.	Partes de un implante coclear. Reproducido de: https://www.advancedbionics.com	13
5.	Implante coclear desde la perspectiva quirúrgica. Reproducido de https://www.advancedbionics.com	13
6.	Señal de electrocolegografía de paciente que fue transplantado de implante coclear al que fue sometido a prueba audiológica de tono puro a 500 Hz a 6 meses de la cirugía	14
7.	Efecto de <i>aliasing</i> . Reproducido de: https://www.physik.uzh.ch	16
8.	Procedimiento para separar los potenciales electricos de las células ciliadas y del nervio auditivo. Reproducido de: Katz, J et al. (2015). Handbook of clinical audiology. Hearing with a cochlear implant. Página: 226	18
9.	Respuesta obtenida a 750 Hz a una respuesta de diferencia y de suma. Reproducido de doi: 10.1097/AUD.0000000000000385	19
10.	Ejemplo de árbol de decisión. Reproducido de https://www.geo.fu-berlin.de	24
11.	Ejemplo de problema de regresión con árbol de desición. Reproducido de https://saedsayad.com	25
11.	Ejemplo de cálculo de la desciación estándar por clases en una variable. Reproducido de https://saedsayad.com	25
12.	Resultado de la desviación estándar por cad atributo. Reproducido de https://saedsayad.com	26
14.	Ejemplo de criterio de terminación de una rama de decisión. Reproducido de: https://saedsayad.com	27
15.	Cálculo de los criterios de terminación de una rama. Reproducido de: https://saedsayad.com	27
16.	Ejemplo de las hojas del árbol. Reproducido de: https://saedsayad.com	28
17.	Ejemplo final de cómo queda el árbol de decisión. Reproducido de: https://saedsayad.com	28
18.	Ejemplo de red neuronal con 5 capas. Reproducido de: https://afit-r.github.io	31
19.	Representación de mínimos locales y mínimos globales cuando la dimensionalidad de las variables es de 3. Reproducido de: https://medium.com	33
20.	Estructura de un autoencoder. Reproducido de: https://towardsdatascience.com/	36
21.	Matriz de Confusión.	37
21.	Protocolo de obtención de datos. Reproducido de doi: 10.1371/journal.pone.0299911	41

22.	Voltajes a través del tiempo del paciente 1	41
23.	Gráfico de cajas sobre la prueba Matrix test en los escenarios de silencio y ruido	42
24.	Intuición del <i>feature</i> de amplitud absoluta	45
25.	Intuición del <i>feature</i> de amplitud ponderada	46
26.	Intuición del <i>feature</i> de área bajo la curva discreta	46
27.	Intuición del <i>feature</i> de área bajo la curva continua	47
28.	Análisis inducido con claro peak en 3100 Hz	48
29.	Análisis inducido con comportamiento oscilatorio	48
30.	Ilustración de la resta de señales oscilatorias en el paciente 1	49
31.	Criterio de asignación de clase	51
32.	Perturbación del espacio latente	52
33.	Representaciones del autoencoder	65

Capítulo 1

Introducción

1.1. Motivación e hipótesis

Acorde a la WHO, la hipoacusia se define como la pérdida parcial o total de la audición mayor a 25 dB en uno o ambos oídos. Puede manifestarse durante diferentes etapas de la vida humana, ya sea en el período prenatal debido a factores genéticos e infecciones intrauterinas, o en la tercera edad debido a degeneraciones neurosensoriales. Se estima que la prevalencia de esta enfermedad corresponde a más del 5% en la población global, lo que corresponde a alrededor de 430 millones de personas[1], donde 396 millones son adultos y 34 millones son niños. La magnitud de esta cifra revela la importancia de comprender y abordar los diversos aspectos relacionados a esto.

Según el nivel de pérdida auditiva, las hipoacusias se pueden clasificar en grado leve (25-40 dB), moderada (40-70 dB), severa (70 a 90 dB), y profunda (>90 dB de pérdida)[2]. Estos niveles dan cuenta del nivel necesario en decibeles para responder a estímulos auditivos en distintas frecuencias. Estas pérdidas auditivas tienen impacto significativos en la calidad de vida de los pacientes, extendiéndose más allá de la pérdida de un sentido, ya que esto afecta la comunicación, el habla, la capacidad cognitiva, e incluso puede tener efectos importantes en la educación y el empleo. Adicionalmente, la hipoacusia no tratada puede generar aislamiento social, estigma y soledad[1].

Afortunadamente, existen diversas formas de tratamiento y apoyo para las personas con hipoacusia, cuyo enfoque puede variar según la gravedad y las necesidades individuales. Una opción común para el tratamiento es usar dispositivos tecnológicos para el apoyo de la audición de los paciente como: (i) audífonos, (ii) implantes de oído medio e (iii) implantes cocleares[3]. En particular, para los casos más graves, donde los audífonos no son opción por la ausencia o daño de células ciliadas (Figura 1), los implantes cocleares son la mejor opción. Estos han revolucionado el campo de la rehabilitación auditiva en los últimos 30 años [4] [5][6] [7] al proporcionar a las personas con pérdida de audición severa o profunda la capacidad de percibir y entender el habla.

Si bien los implantes cocleares han demostrado éxito al restaurar los umbrales auditivos y la capacidad de detectar estímulos auditivos simples, existe una considerable variabilidad entre los pacientes en cuanto a la habilidad para interpretar dichos sonidos como lenguaje hablado en un contexto comunicativo. Este aspecto adquiere una importancia crucial, dado que

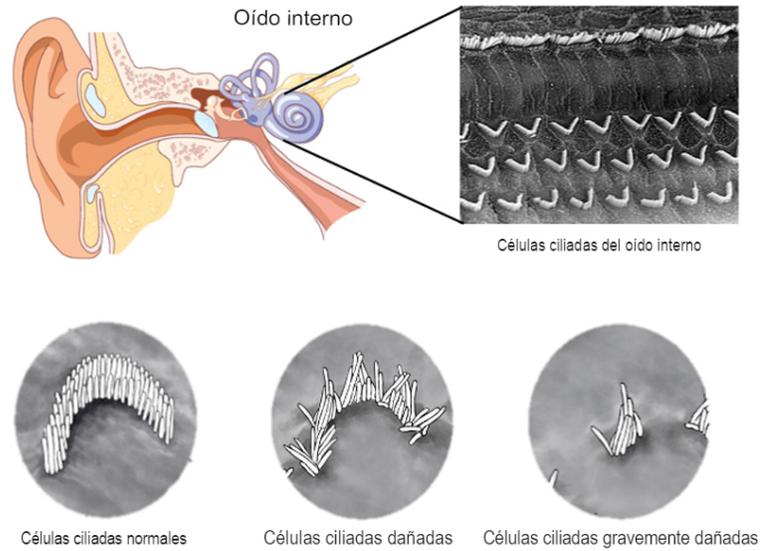


Figura 1: Esquema del sistema auditivo en humanos. Se muestran ilustraciones de las células ciliadas de la cóclea (órgano receptor auditivo) y como estas se van dañando a lo largo de la vida. Dr Pathak, S. Age Related hearing Loss: Presbycusis. Reproducido de: <https://www.drshrehapathakent.com>

en individuos con discapacidad auditiva, la interpretación del sonido de un mensaje hablado, especialmente en diversas condiciones ambientales, puede representar un desafío significativo. La capacidad de interpretación se convierte en un factor determinante para que el paciente logre identificar el mensaje transmitido de manera oral. [8].

Como se menciona en el párrafo anterior, es crucial separar los procesos fisiológicos de la percepción de sonidos y del habla. El primer proceso implica que los sonidos se transformen en señales dentro del oído para luego ser transportados al cerebro. El segundo proceso consiste en que el cerebro pueda traducir esa señal y lograr interpretarla. La complejidad de este se manifiesta en que las señales no siempre se presentan en formas fácilmente comprensibles para el cerebro. En el caso de personas con sordera que utilizan implantes cocleares, la interpretación de señales auditivas degradadas se convierte en un desafío significativo, ya que el cerebro no está naturalmente preparado para este tipo de procesamiento, dado que partes específicas encargadas de esta función tienden a degradarse por desuso [8].

El desuso afecta directamente a la capacidad del cerebro para aprender sobre señales sonoras que vienen del entorno. El cerebro organiza los sonidos conocidos en objetos auditivos, permitiéndole asociar señales degradadas a algún objeto auditivo familiar. En el caso de hablantes nativos, el aprendizaje mediado por plasticidad cerebral facilita el acceso a estas representaciones auditivas, posibilitando la comprensión de señales degradadas, ya que el cerebro completa la señal. Sin embargo, esta capacidad se ve limitada en pacientes con implantes cocleares debido a la privación de este aprendizaje por la sordera. En consecuencia, al recibir estímulos auditivos, la comprensión del habla **puede** resultar confusa tanto en situaciones de silencio como en entornos ruidosos. Se destaca la palabra "puede" debido a la

variabilidad el entendimiento del habla en pacientes con implantes cocleares [9].

Lo anteriormente mencionado representa un desafío significativo para los pacientes. Muchos de ellos optan por no someterse a la operación debido a tres razones: la incertidumbre sobre el funcionamiento del implante en su caso particular, la falta de herramientas no invasivas para detectar en qué falló el implante en los casos donde no hay un buen desempeño postoperatorio, y por último, la posible pérdida de la audición residual luego de la intervención[10] [11].

La carencia de herramientas para pacientes con resultados postoperatorios desfavorables se atribuye a la complejidad que implica evaluar la interacción del implante con el cuerpo, especialmente cuando se encuentra ubicado en una región tan reducida como la cóclea la cual se encuentra dentro de la cabeza. Los resultados adversos después de la operación pueden originarse por la formación de fibrosis y/o inflamación alrededor de los electrodos, lo cual afecta directamente la transmisión de señales al cerebro a través del implante [12].

Estos factores también inciden en la eficacia de la electrococleografía, un método de evaluación electrofisiológico que utiliza el mismo implante coclear para registrar la funcionalidad de la cóclea del paciente la cual permite medir las respuestas del nervio auditivo y de las células ciliadas cocleares y de esta manera evaluar la calidad de la señal. Esto se debe a que las interacciones previamente mencionadas con el cuerpo también afectan a los electrodos encargados de llevar a cabo la electrococleografía. En otras palabras, esta interferencia repercute en la calidad de la señal recopilada por la electrococleografía, ya que los electrodos se ven afectados por la inflamación o fibrosis, lo cual puede afectar a la señal electrococleografica.

La identificación de estas alteraciones podría facilitar la manipulación de variables ajustables, como la ubicación de los electrodos y la rehabilitación auditiva postauricular, con el objetivo de optimizar el desempeño individual de cada paciente. Sin embargo, es esencial establecer de manera precisa la relación entre las señales de las electrococleografías y la comprensión del habla para detectar posibles problemas de fibrosis o inflamación a lo largo del tiempo. Por ende, para mejorar los resultados de los pacientes, es imperativo desarrollar una herramienta que permita trazar la correlación entre las electrococleografías y la capacidad de comprensión del habla antes de abordar las variables modificables.

Por otro lado, La distinción entre el entendimiento del habla en silencio y en entornos ruidosos añade una capa adicional de complejidad. En el escenario de silencio, donde dos personas conversan sin interferencias sonoras externas, los pacientes con implantes cocleares tienden a mostrar un mejor desempeño en comparación con situaciones ruidosas. Sin embargo, surge un dilema al medir esta métrica en silencio, ya que la comunicación entre individuos rara vez ocurre en ambientes sin ruido. Por lo tanto, resulta esencial abordar este fenómeno en presencia de condiciones ruidosas para obtener una comprensión más precisa y aplicable a la vida cotidiana y así poder mejorar la calidad de vida de estos pacientes.

En la actualidad, los estudios más recientes han incorporado las características de las señales evocadas¹ extraídas de las electrococleografías posoperatorias en modelos de regresión lineal. Las características mencionadas se originan de la exposición de los pacientes a tonos de frecuencias específicas, tales como 250, 500, 750, 1000, 2000 y 4000 Hz. Estos tonos provocan

¹ Es un método que promedia las señales con el fin de reducir o mitigar el ruido presente en ellas.

respuestas en los oídos de los individuos, las cuales pueden ser representadas en forma de señales. A partir de estas respuestas, se derivan señales de fase de condensación y rarefacción. Posteriormente, estas dos fases se promedian y se identifican los peaks generados en el primer y segundo armónico para luego ser sumados, resultando en una métrica que engloba la totalidad de estos datos la cual se llama electrocoqueografía de respuesta total. Sin embargo, a pesar de estos esfuerzos, estos modelos solo logran explicar un 34 % de la variabilidad en la comprensión del habla en entornos ruidosos [9].

Los resultados modestos sugieren que la relación subyacente entre estas dos variables podría no ser lineal. En consecuencia, se plantea la idea de que la creación de modelos de aprendizaje automático, utilizando señales evocadas a partir de los registros de electrocoqueografías de prueba única, podría proporcionar predicciones más precisas para la comprensión del habla en condiciones de ruido, superando así el estándar actual representado por las regresiones lineales.

1.2. Fundamentos del problema

A lo largo de los años, se han estudiado múltiples variables que se creían que afectaban el entendimiento del habla como variables de la historia del paciente, registros audiológicos, psicológicos y antecedentes médicos del paciente. En la década de los 90 se hacían estudios con variables temporales del paciente como edad de implantación del implante coclear, duración de la sordera profunda y la causa de la sordera. Los resultados de esa época señalaban que la duración de la sordera era el mejor predictor individual y preoperatorio para predecir el desempeño del habla en ruido [4]. Sin embargo, a medida que pasaron los años, esa medida empezó a cobrar menos importancia, explicando menos del 20 % de la varianza, debido a que hay mejores tratamientos para personas que empiezan a quedar sordas y además, ya no se esperan tantos años para los implantes[6][7]. Luego se empezaron a estudiar variables electrofisiológicas pues permiten registrar cómo está la salud coclear y del nervio auditivo[13].

Una de las señales electrofisiológicas que se han estudiado son las electrocoqueografías. Estas consisten en una técnica de diagnóstico utilizada en audiología para evaluar la función coclear en pacientes con implantes cocleares, es decir, el funcionamiento del oído interno y, específicamente, de la cóclea y del nervio auditivo. Esta prueba registra las respuestas eléctricas generadas por las células ciliadas del oído interno y del nervio auditivo en respuesta a estímulos sonoros presentados con una polaridad alterna de condensación y rarefacción [3]. Son muy útiles a la hora de ver la salud coclear, de hecho estas presentan una alta correlación con las pruebas de audiometría de tonos puros, las cuales son una prueba en que el paciente participa mencionando cuáles umbrales puede oír[14].

Las formas de descomponer las señales de las electrocoqueografías en los últimos años han dado resultado a 4 potenciales eléctricos que provienen de diferentes partes del oído. Los que provienen de las células ciliadas son el microfónico coclear, y el potencial de sumación[15] y los que corresponden a potenciales neuronales son el neurofónico del nervio auditivo[16] y el potencial de acción compuesto. Todos esos potenciales, se miden utilizando técnicas de promediación de la señal en el tiempo, y se han utilizado para intentar predecir el entendimiento del habla en pacientes, y a lo largo de los años han cobrado mayor relevancia a la hora de explicar esa variable [9].

Por otra parte, existe un potencial eléctrico poco explorado en estudios humanos: la actividad espontánea del nervio auditivo. Investigaciones recientes sugieren su posible presencia a una frecuencia de 3100 Hz, y su medición podría añadir otra dimensión para evaluar el estado del nervio auditivo [17]. Es importante señalar que la mayoría de los estudios han centrado su atención en la medición de las electrococleografías después de la intervención quirúrgica para implantes cocleares. Esto se debe a que el propio dispositivo implantado permite la captura de señales eléctricas en su lugar de origen, a diferencia de por ejemplo, realizar la medición en el tímpano, donde la señal obtenida tiende a debilitarse. Esta particularidad es crucial, ya que los pacientes con implantes cocleares son particularmente idóneos para este tipo de mediciones en comparación con individuos sanos. Esto se debe a que, en el caso de personas sin discapacidad auditiva, la obtención de señales de alta calidad sería notablemente invasivo.

La representación de la electrococleografía que mejores resultados ha dado es la electrococleografía de respuesta total, la cual también utiliza promedios de series de tiempo a diferentes frecuencias (250, 500, 750, 1000, 2000 y 4000 Hz) intentando abarcar todo el espectro del habla [16], la forma de medirlo es que se obtiene la magnitud del estímulo en cada una de las frecuencias para luego surmarlas y tener una única medida. Dicha información ha dado resultados de la explicabilidad del 59 % de la variación del entendimiento del habla en silencio, y un 34 % en el habla en ruido en condiciones postoperatorias [9].

Con el fin de abordar la problemática asociada a la creación de una herramienta diagnóstica para comprender el bajo rendimiento postoperatorio en pacientes con implantes cocleares, se emplearán los registros postoperatorios, para que en una primera instancia se pueda estudiar si hay una relación fuerte entre esta y la variable del entendimiento del habla en ruido y en un futuro poder relacionarla con el deterioro de la señal a lo largo del tiempo provocada por la inflamación y fibrosis del tejido que esta en contacto con los electrodos. En consecuencia, este enfoque permitiría ajustar variables modificables, tales como la colocación de electrodos, y definir una rehabilitación auditiva individual adecuada, con el propósito de optimizar el rendimiento de cada paciente.[4][5][18][6][7][16][19][9].

Una vía efectiva para desarrollar esta herramienta es a través de la ciencia de datos, un campo que integra conocimientos de estadística, aprendizaje automático y análisis de datos para extraer información significativa de conjuntos de datos complejos. La aplicación de técnicas de ciencia de datos en el ámbito de la predicción del entendimiento del habla en pacientes con implantes cocleares tiene el potencial de generar hallazgos valiosos que contribuyan a mejorar los resultados clínicos y la calidad de vida de estos individuos. Es fundamental destacar que, hasta el momento, el estado del arte en esta problemática se limita al uso de modelos de regresión lineal multivariable para explicar la variable del entendimiento del habla.

1.3. Objetivos del trabajo

En la siguiente tabla se presenta el objetivo general definido en este trabajo, junto con sus objetivos específicos correspondientes.

Tabla 1.1: Presentación de objetivos.

Objetivo General		
Investigar y desarrollar un modelo de predicción no lineal del entendimiento del habla en condiciones de ruido para pacientes con implantes cocleares. Para ello se propone utilizar señales evocadas de los ensayos clínicos de las electrocoeleografías y además, técnicas de ciencia de datos (modelos de aprendizaje supervisado y análisis de señales) con el fin de mejorar los resultados clínicos y la calidad de vida de estos individuos.		
Objetivos Específicos		
1. Desarrollar y establecer un diseño experimental a partir del estado del arte del problema.	2. Desarrollo de los modelos propuestos con las señales de las electrocoeleografías.	3. Evaluación del desempeño de los modelos.
Resultados Esperados		
<ul style="list-style-type: none"> • Estudiar y entender la problemática de los pacientes • Realizar un bechmark relacionado con el estado del arte del problema. • Definir los modelos del benchmark como línea base para el trabajo. 	<ul style="list-style-type: none"> • Un script de Python que incluya todas las configuraciones necesarias para el procesamiento de señales electrofisiológicas y el entrenamiento de modelos 	<ul style="list-style-type: none"> • Definición de métricas para evaluar los modelos. • Evaluar el desempeño de los modelos • Seleccionar los mejores modelos según el caso

Capítulo 2

Metodología, conceptos y teoría

2.1. Metodología

La metodología empleada en esta tesis de magíster en Ciencia de Datos para la predicción del entendimiento del habla en pacientes con implantes cocleares a través de señales de electrococleografía se fundamenta en la estructura CRISP-DM (Cross Industry Standard Process for Data Mining)[20]. No obstante, se realizaron ajustes a dicho enfoque, ya que el problema abordado se centra en ensayos clínicos que, por un lado, no están vinculados a una industria, sino que se enfocan en el paciente, y por otro lado, hasta la redacción de este trabajo, la fase de despliegue no se ha ejecutado debido a la dificultad para reclutar pacientes, quienes deben someterse previamente a la aprobación de un comité ético de la Universidad de Chile.

En este contexto, la metodología CRISP-DM se ha modificado en su primera etapa, la comprensión del negocio, y se ha eliminado la última etapa, correspondiente al despliegue, debido a las razones mencionadas anteriormente. La estructura resultante se configura de la siguiente manera:

La primera fase consiste en comprender el problema de investigación y establecer los objetivos específicos de la tesis. Para lograr esto, se lleva a cabo una revisión exhaustiva de la literatura existente relacionada con implantes cocleares, electrococleografías y técnicas de predicción del entendimiento del habla, además de revisar conceptos médicos para entender cómo se transmite el sonido en el oído de los humanos. Esta revisión proporciona una base sólida de conocimientos y antecedentes para la investigación.

En la segunda fase, se pone énfasis en comprender los datos disponibles. Esto implica tener acceso a los datos de electrococleografías de pacientes con implantes cocleares y realizar una exploración inicial para comprender su estructura, calidad y características. Durante esta etapa, también se pueden identificar posibles desafíos o problemas relacionados con los datos, como la presencia de datos faltantes o ruidos en las señales. Esta comprensión profunda de los datos es crucial para garantizar la validez y la calidad de los resultados obtenidos.

La tercera fase se dedica a la preparación de los datos para su posterior análisis. Esto implica realizar tareas como la limpieza de los datos para eliminar ruidos o valores atípicos, la normalización de los datos para que estén en un rango común y la selección de las variables relevantes para el análisis. Esta etapa requiere un cuidadoso procesamiento de los datos, con

el objetivo de asegurar que estén en un formato adecuado y listos para su uso en el modelado.

En la cuarta fase, se lleva a cabo el modelado de los datos. Aquí, se seleccionan los modelos de predicción más adecuados para predecir el entendimiento del habla en pacientes con implantes cocleares utilizando las señales de electrococleografías, además de replicar los modelos utilizados en el estándar de referencia actual. Los modelos seleccionados pueden ser de regresión y clasificación. Estos modelos se entrenan y ajustan utilizando los datos preparados en la fase anterior.

La quinta y última fase está dedicada a la evaluación de los modelos. Se utilizan diversas métricas y técnicas para evaluar el rendimiento de los modelos de predicción. Esto permite determinar qué modelos son los más efectivos en la predicción del entendimiento del habla en pacientes con implantes cocleares comparándolos también, con los modelos de referencia que hay actualmente.

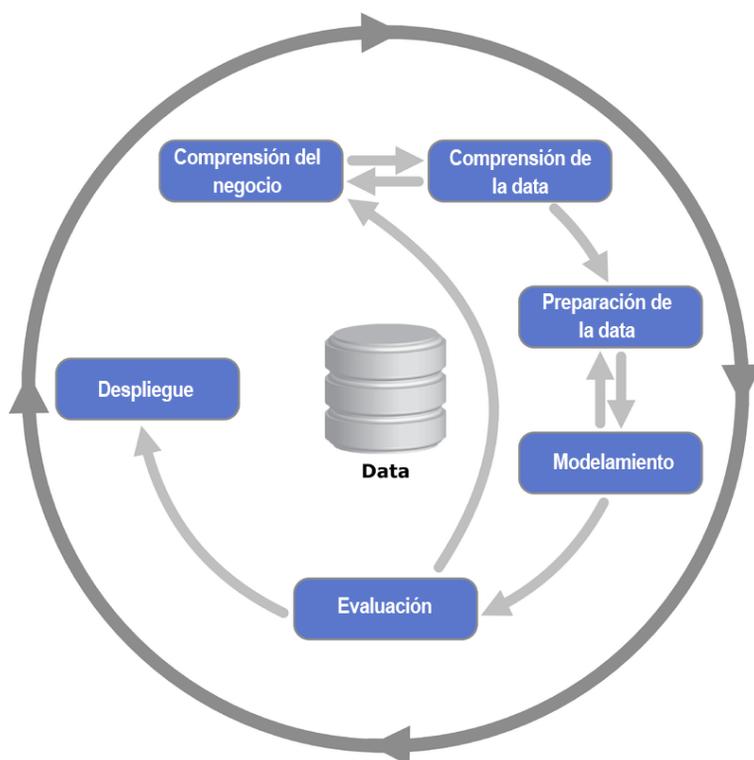


Figura 2: Fases CRISP-DM. Reproducido de: <https://healthdataminer.com>

2.2. Conceptos y teoría

Para el correcto análisis y modelamiento del problema es crucial entender conceptos de médicos del funcionamiento del oído e implantes cocleares, procesamiento de señal, electrococleografías y de modelos de aprendizaje de máquinas. A continuación se presenta cada uno separado por sección:

2.3. El oído

Para comprender el desafío de predecir la comprensión del habla en pacientes con implantes cocleares, es fundamental entender la anatomía auditiva, la transmisión del sonido en el oído y la composición y funcionamiento del implante coclear. A continuación, se detallarán cada uno de estos aspectos.

2.3.1. Partes de oído

El oído se compone de tres partes, estas son el oído externo, el oído medio y el oído interno. Cada uno de ellos se compone de las siguientes partes:

Oído externo: El oído externo es la parte del oído que se encuentra en el exterior del cráneo y está compuesto por el pabellón auricular (o la oreja) y el canal auditivo externo.

- **Pabellón auricular:** Es la parte visible del oído que se encuentra en el exterior del cráneo. Está formado por cartílago y piel, y tiene una forma característica que ayuda a dirigir el sonido hacia el canal auditivo.
- **Canal auditivo externo:** Es un tubo estrecho y curvo que conecta el pabellón auricular con el tímpano. El canal está cubierto de piel y tiene glándulas que producen cerumen para proteger el oído de la suciedad y las bacterias.

Oído medio: El oído medio es la sección del oído que se encuentra entre el tímpano y el oído interno. Está compuesto por la caja del tímpano y tres huesos llamados martillo, yunque y estribo.

- **Tímpano:** Es una membrana delgada que separa el oído externo del oído medio. Vibra en respuesta a las ondas sonoras y transmite las vibraciones al oído interno.
- **Caja del tímpano:** Es una pequeña cámara de aire detrás del tímpano. Contiene los tres huesos del oído medio y ayuda a amplificar las vibraciones del tímpano.
- **Martillo, yunque y estribo:** Son los tres huesos más pequeños del cuerpo humano. Conectan el tímpano con la cóclea (el órgano del oído interno).

Oído interno: El oído interno es la sección del oído que se encuentra dentro del cráneo, y es responsable de la transducción de las vibraciones sonoras en señales eléctricas que son enviadas al cerebro. Está compuesto por la cóclea, el vestíbulo y los canales semicirculares.

- **Cóclea:** Es un órgano en forma de caracol que se encuentra en el oído interno. Contiene células ciliadas que convierten las vibraciones sonoras en señales eléctricas que son enviadas al cerebro.
- **Vestíbulo:** Es una pequeña estructura en el oído interno que ayuda a mantener el equilibrio y la orientación espacial.
- **Canales semicirculares:** Son tres tubos curvos que también se encuentran en el oído interno y están llenos de líquido. Ayudan a detectar los movimientos de la cabeza y el cuerpo, y son importantes para mantener el equilibrio.[21]

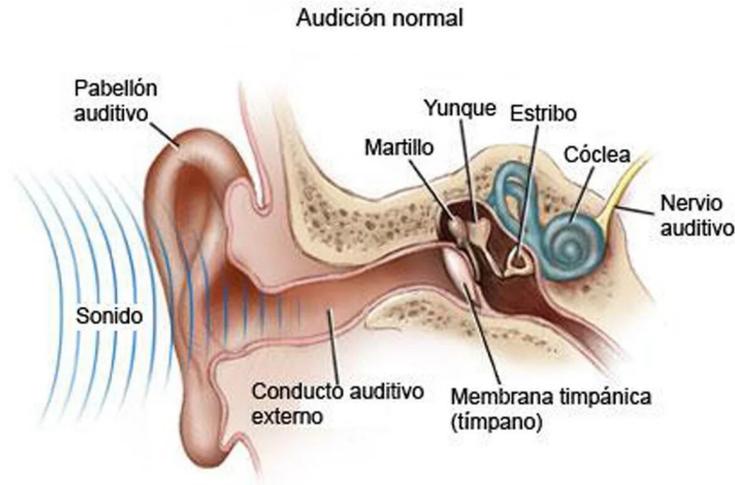


Figura 3: Imagen de partes del oído. Reproducido de : <https://www.childrensmn.org>

2.3.2. Trasmisión del sonido al cerebro

Cuando una persona está expuesta a un sonido, el oído externo captura las vibraciones y las dirige a través del conducto auditivo externo hacia el oído medio, compuesto por el tímpano y la cadena de huesecillos. Estas vibraciones sonoras provocan el movimiento de los huesecillos, generando a su vez el desplazamiento del fluido dentro de la cóclea. Este movimiento del fluido estimula las células ciliadas, pequeños receptores auditivos que recubren la interioridad de la cóclea. Al inclinarse hacia adelante y hacia atrás, los cilios transmiten señales eléctricas al nervio auditivo, que a su vez transmite estas señales al cerebro para su interpretación.

Sin embargo, debido al proceso de envejecimiento, factores genéticos, ciertas enfermedades, infecciones o la exposición repetida a ruidos intensos, los cilios pueden resultar dañados o incluso destruidos. En casos donde los cilios no funcionan correctamente, el nervio auditivo no recibe estimulación y, por ende, no puede enviar información al cerebro, resultando en la pérdida de la capacidad auditiva.

Para quienes presentan daño en las células del oído interno y sufren de sordera severa a profunda, la recomendación principal es el implante coclear. Este dispositivo actúa como un sustituto de la cóclea, específicamente de la función de las células ciliadas donde el implante genera señales eléctricas que estimulan el nervio auditivo, facilitando la transmisión de información al cerebro. En esencia, este aparato desempeña el papel crucial de generar las señales necesarias para compensar la pérdida de la función de las células ciliadas, permitiendo así mejorar la capacidad auditiva en personas con este tipo de afectaciones.[22]

2.4. Implante coclear

2.4.1. Contexto e historia

Una de las alternativas para abordar la pérdida de audición consiste en el empleo de diversos dispositivos implantables que posibilitan la sustitución de distintas partes de las estructuras que facilitan la audición. Entre ellos, se encuentran dispositivos anclados al hueso, implantes de oído medio, *implantes cocleares* e implantes auditivos de tronco encefálico. En el contexto de esta tesis, resulta crucial comprender el funcionamiento de los implantes cocleares y su evolución histórica para analizar las implicaciones de investigaciones previas y evaluar las limitaciones de los datos que se pueden obtener mediante su utilización.

Los implantes cocleares fueron inicialmente concebidos para personas con deterioro profundo de la función del oído interno. Esta elección de pacientes se restringía a aquellos con sordera profunda, y más específicamente, con anacusia, es decir, la incapacidad total de percibir las frecuencias que conforman los sonidos cotidianos[23]. Una clasificación de los niveles de sordera se presenta a continuación:

Tabla 2.1: Criterios de los niveles de sordera según decibeles

Nivel de sordera	Nivel necesario en decibeles para responder a estímulos auditivos en distintas frecuencias
Leve	(25-40 dB)
Moderado	(40-70 dB)
Severa	(70 a 90 dB)
Profunda	(>90 dB de pérdida)

Con el transcurso del tiempo, el criterio de selección de pacientes para implantes cocleares ha experimentado una flexibilización. A medida que avanzan las décadas, se siguen llevando a cabo estudios clínicos con el fin de evaluar la posibilidad de ampliar la elegibilidad de los pacientes. Esto ha permitido la inclusión de individuos con pérdidas auditivas de menor grado, impulsado en parte por los avances tecnológicos que han mejorado las técnicas quirúrgicas. Como resultado, se ha reducido la probabilidad de pérdida de audición residual durante la inserción del implante [24]. A continuación, se presenta una tabla que detalla los diversos criterios que han regido a lo largo de los años:

Tabla 2.2: Criterios de implante coclear en la población adulta según FDA

Año	Cambios en los criterios para implante coclear en población adulta
1985	Aprobado para adultos con sordera postlocutiva de 18 años o más con pérdida auditiva bilateral profunda (mayor o igual a 90 dB HL) y sin habilidades de reconocimiento de voz.
1990	Criterio edad: sin cambios. Criterio grado severidad: sin cambios. Criterio percepción del habla: sin cambios.
1998	Criterio edad: se adicionan adultos con hipoacusia prelocutiva. Criterio grado severidad: se amplía y considera hipoacusias severas y profundas (PTP mayor o igual a 70 dB HL). Criterio percepción del habla: considera a personas cuya percepción del habla es menor o igual al 40 % en el mejor oído con adaptación auditiva.
2000	Criterio edad: sin cambios. Criterio grado severidad: sin cambios. Criterio percepción del habla: menor o igual al 50 % correcto en el oído candidato a ser implantado. Menor o igual al 60 % con adaptación auditiva.
2014	Criterio edad: sin cambios. Criterio grado de severidad: en las bajas frecuencias audición normal o con hipoacusia leve – moderada (< 60 dB HL) con pendiente pronunciada (mayor o igual a 70 dB HL) para dispositivos híbridos, para hipoacusias moderadas a profundas considerar dispositivos tradicionales. Criterio percepción del habla: rendimiento entre 10 % a 60 % en el oído a ser implantado y < a 80 % en el oído contralateral.
2019	Criterio edad: sin cambios. Criterio grado de severidad: se adiciona hipoacusia unilateral y pérdida auditiva asimétrica con menos de 10 años de inicio. Criterio percepción del habla: sin cambios.
2020	Criterio edad: sin cambios. Criterio grado de severidad: sin cambios. Criterio percepción del habla: sin cambios.

En términos de su funcionamiento, el implante coclear opera de dos maneras distintas. La primera funciona para que el paciente pueda oír y la segunda funciona como una forma de registro de las señales que produce el propio oído de la persona. En otras palabras el primero es la transmisión de sonido con el implante coclear y la segunda en la forma de electrococleografía que es la forma de registro.

En lo que respecta al financiamiento en Chile, el costo de cada unidad de este dispositivo se sitúa en torno a los 20 mil dólares. Para los recién nacidos, el financiamiento se realiza a través del GES, mientras que para los adultos se cubre bajo la ley Ricarte Soto (Ley 20.850, específicamente para la patología de Hipoacusia Sensorineural Bilateral severa a profunda postlocutiva)[25].

2.4.2. Transmisión del sonido con implante coclear

El implante coclear consta de cinco partes las cuales contemplan funciones completamente distintas en la trasmisión de sonido. Como se puede observar en la imagen, el punto 1 es el micrófono del aparato el cual capta las ondas sonoras del ambiente, luego estas son pasadas por el punto 2 que es el procesador de sonido el cual transforma la señal análoga en una señal digital y luego esta es traspasada al punto 3 el cual es una antena receptora que envía las señales digitales al implante coclear y este genera señales eléctricas que son transmitidas a los electrodos del punto 4 y a su vez estos electrodos estimulan el nervio auditivo en el punto 5.



Figura 4: Partes de un implante coclear. Reproducido de: <https://www.advancedbionics.com>

Desde otra perspectiva ya implantada el aparato se ve de la siguiente manera:

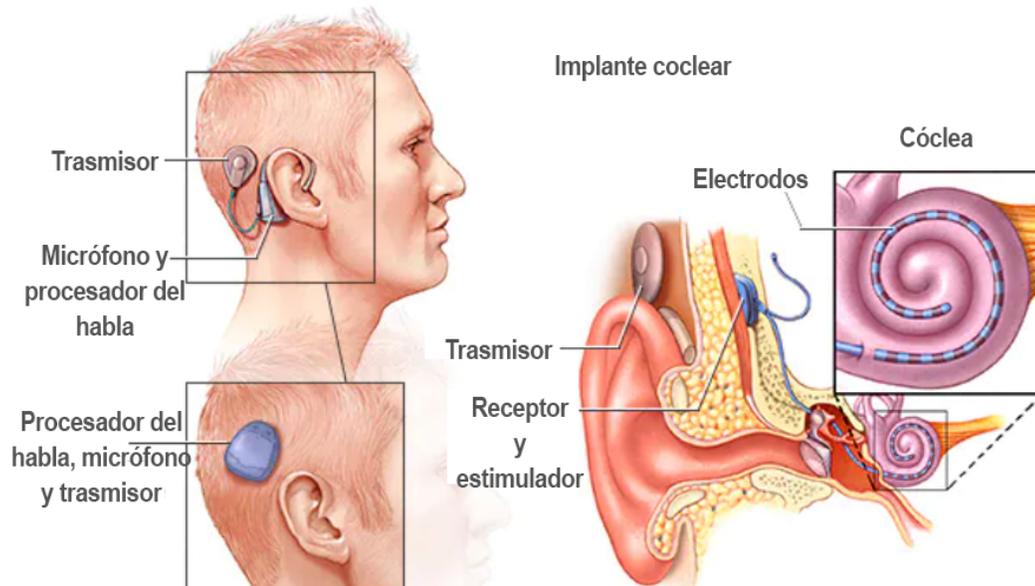


Figura 5: Implante coclear desde la perspectiva quirúrgica. Reproducido de <https://www.advancedbionics.com>

2.4.3. Electrocolegrafía (ECoG)

Este examen evalúa la actividad eléctrica de las células ciliadas y del nervio auditivo, permitiendo la medición de los potenciales eléctricos asociados a cada una de esas regiones. Ambas áreas son fundamentales para la audición: las células ciliadas transmiten la información

procesada al nervio auditivo y el nervio auditivo se encarga de hacer llegar esa información al cerebro.

2.5. Procesamiento de señales

2.5.1. Qué es una señal

Cuando nos referimos a señales, hablamos de cualquier magnitud física que experimenta cambios a lo largo del tiempo, el espacio u otras variables independientes. Un ejemplo de esto es una señal electrofisiológica del cuerpo humano la cual podría variar en voltaje a través del tiempo. Ejemplos de estas señales electrofisiológicas son los electrocardiogramas, encefalogramas y electrococcleografías las cuales miden la variación de voltaje en el corazón, encéfalo y coclea. Las señales anteriormente mencionadas podrían ser descritas como una función que varía el voltaje a través del tiempo y a modo de ilustración se presentará una de electrococcleografía:

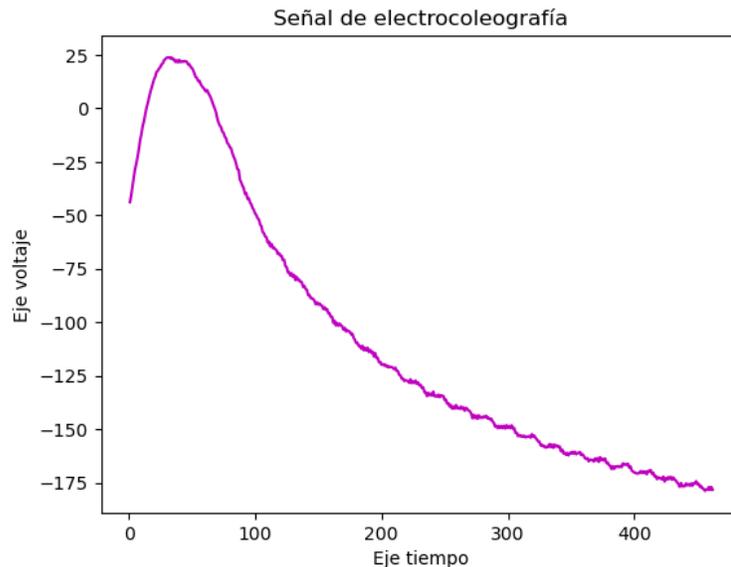


Figura 6: Señal de electrococcleografía de paciente que fue transplantado de implante coclear al que fue sometido a prueba audiológica de tono puro a 500 Hz a 6 meses de la cirugía

2.5.1.1. Contexto

Hablar sobre el procesamiento de señales implica comprender el concepto fundamental de sistema en este ámbito. En términos sencillos, un sistema puede ser un dispositivo físico o un conjunto de software y hardware que ejecuta operaciones sobre una señal. Cuando dirigimos una señal a través de un sistema, como en el caso del filtrado, podemos afirmar que la señal ha sido procesada. En esta situación, el procesamiento de la señal implica la eliminación de ruido e interferencias de la señal deseada.

En términos generales, un sistema se define por el tipo de operación que realiza sobre la señal. Por ejemplo, si la operación es lineal, se le denomina sistema lineal. Si la operación es

no lineal, el sistema se clasifica como no lineal, y así sucesivamente. Estas operaciones suelen agruparse bajo el término "procesamiento de señales"[26].

En el caso específico que nos ocupa, las señales digitales que recibimos fueron las provenientes de ensayos únicos de electrocoleografías en pacientes con implantes cocleares. El sistema que procesó estas señales fue un software, en concreto Python, que llevó a cabo un análisis espectral. Este enfoque se seleccionó debido a que la representación más natural implicaba analizar las frecuencias capturadas por los implantes. Por lo tanto, se aplicó una Transformada de Fourier Discreta a las señales para obtener información significativa sobre su contenido espectral.

2.5.1.2. Transformación análogo digital

La transformación analógica a digital es el proceso mediante el cual una señal analógica, que varía de manera continua en el tiempo, se convierte en una representación digital, es decir, una secuencia de valores discretos en el tiempo. Este proceso es esencial en una amplia variedad de aplicaciones, como la obtención de datos electrofisiológicos.

En el contexto de la tesis, la frecuencia de muestreo es un factor importante para futuras aplicaciones de la transformada de fourier. En específico, la frecuencia de muestreo es el número de muestras de una señal analógica que se toman en un intervalo de tiempo específico. En otras palabras, es la cantidad de veces por segundo que se registran o discretizan los valores de una señal analógica para convertirla en una señal digital.

Cuando se adquiere o procesa una señal analógica para convertirla en una forma digital (digitalización), la frecuencia de muestreo es un parámetro clave que determina cuántas mediciones se realizan por unidad de tiempo. Se expresa en unidades de Hertz (Hz) y representa el número de muestras por segundo.

El teorema de muestreo de Nyquist-Shannon establece que la frecuencia de muestreo debe ser al menos el doble de la frecuencia máxima presente en la señal analógica para evitar el fenómeno de aliasing y permitir una reconstrucción fiel de la señal original. Esta frecuencia máxima en la señal analógica se conoce como la frecuencia de Nyquist.

El aliasing es un fenómeno no deseado que puede ocurrir al muestrear una señal analógica a una frecuencia de muestreo insuficiente. Se manifiesta cuando frecuencias en la señal original que están por encima de la mitad de la frecuencia de muestreo (frecuencia de Nyquist) se "doblan." se reflejan incorrectamente en el rango de frecuencias más bajo, dando lugar a frecuencias espurias que no estaban presentes en la señal original [27]. Un ejemplo de esto se puede observar en la Figura 7:

2.5.2. Transformada de fourier de tiempo discreto

La transformada de Fourier y la serie de Fourier son herramientas matemáticas fundamentales en el análisis y diseño de representaciones de señales las cuales pasan del dominio del tiempo al dominio de la frecuencia y a través de esto permiten saber cuales son las frecuencias que están presentes en ellas. En el caso de esta investigación, debido al carácter digital de las señales, la herramienta que se usa es la de la **Transformada de Fourier Discreta** debido a que se tiene una muestra de la señal es un tiempo limitado. La formula es la siguiente:

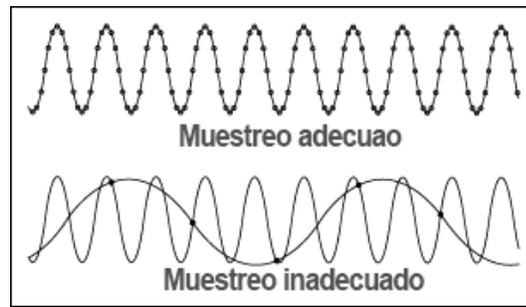


Figura 7: Efecto de *aliasing*. Reproducido de: <https://www.physik.uzh.ch>

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$$

Donde $X(\omega)$ representa la suma de todas las frecuencias que contiene la señal $x(n)$, n representa el tiempo por el que se mueve la señal, la j es un número imaginario y la ω representa la frecuencia angular que tiene por fórmula $\omega = \frac{2\pi f}{N}$ donde N es la **frecuencia de muestreo** y f es la frecuencia. Cabe destacar que al ser tratada como una transformada de fourier de tiempo discreto, todas las frecuencias que contiene la señal quedan descritas por el intervalo de frecuencias entre $(-\pi, \pi)$ o equivalentemente en $(0, 2\pi)$. Por lo tanto, para obtener la frecuencia máxima contenida en la señal se tiene que despejar lo siguiente:

$$\omega = \frac{2\pi f}{N}$$

Como ω debe tomar el valor de π para ser límite, se reemplaza por ese valor y se despeja f :

$$\pi = \frac{2\pi f}{N}$$

$$f = \frac{\pi N}{2\pi}$$

Dando así el resultado de la frecuencia más alta que descompone la transformada de fourier para la señal respectiva [28]

2.5.3. Identidad de Parseval

La Identidad de Parseval es una fórmula que permite calcular la energía total de una señal ya sea mediante la representación en el dominio del tiempo o la representación en el dominio de las frecuencias, las cuales son equivalentes. La fórmula descrita es la siguiente:

$$N \sum_{n=0}^{N-1} |x[n]|^2 = \sum_{k=0}^{N-1} |X[k]|^2$$

Donde el lado izquierdo de la igualdad representa la suma al cuadrado de toda la señal en el dominio dominio del tiempo y en el lado derecho de la igualdad está la suma al cuadrado de toda la señal en el dominio de la frecuencia dividido por la máxima frecuencia que está contenida en la señal. Cabe mencionar que el $X[k]$ es el resultado de la transformada de

fourier [28].

2.5.4. Señal evocada

La señal evocada consiste en promediar las respuestas individuales de señales en el dominio del tiempo a estímulos para luego pasarla por una transformada de fourier, permitiendo así la eliminación del ruido presente en la señal. No obstante, este proceso de promediado también implica la posibilidad de pérdida de información en la señal.

2.5.5. Señal inducida

La señal inducida se obtiene al someter las respuestas individuales de señales en el dominio del tiempo a una transformada de Fourier, seguida por la realización de un promedio de todas estas transformadas. Se estima que este enfoque evita la pérdida de información inherente a las señales individuales.

2.6. Conceptos de las electrocoeleografías

Cuando nos referimos a electrocoeleografías, buscamos medir las respuestas de los potenciales eléctricos del oído y su correlación con la percepción del habla. Entre estos potenciales se incluyen el microfónico coclear, el potencial de acción compuesto y la actividad espontánea del nervio auditivo.

2.6.1. Potenciales de electrocoeleografías (ECoG) evocados

Los potenciales ECoG evocados son el resultado de señales evocadas de electrocoeleografías de las cuales se les extraen estas representaciones. Estos estímulos acústicos resultan de la superposición de tres componentes, dos de ellos provienen de las células ciliadas de la cóclea, siendo el microfónico coclear (CM) y el potencial de sumación (SP); mientras que, el tercero, el potencial de acción compuesto (CAP), surge de las fibras nerviosas del nervio auditivo [29]. En los registros obtenidos en respuesta a estímulos de una polaridad específica, estos componentes se entrelazan, y dependiendo del tipo y la intensidad de la estimulación acústica, no es fácil distinguirlos entre sí. La observación de lo anterior cuando se presenta un tono se refleja en la Figura 8. En la parte A, se distinguen las distintas polaridades: la polaridad de condensación se muestra mediante la línea gris, mientras que la rarefacción se indica con la línea negra. Cada una de estas líneas representa el resultado de promediar las señales separadas por polaridades. En B, se aprecia la suma de las 2 respuestas representadas en A, que da lugar al CAP y al SP, siendo el CAP la curva más prominente. Finalmente, en C se muestra la diferencia entre la respuesta de rarefacción y la de condensación de la figura A, lo que resulta en el MC.

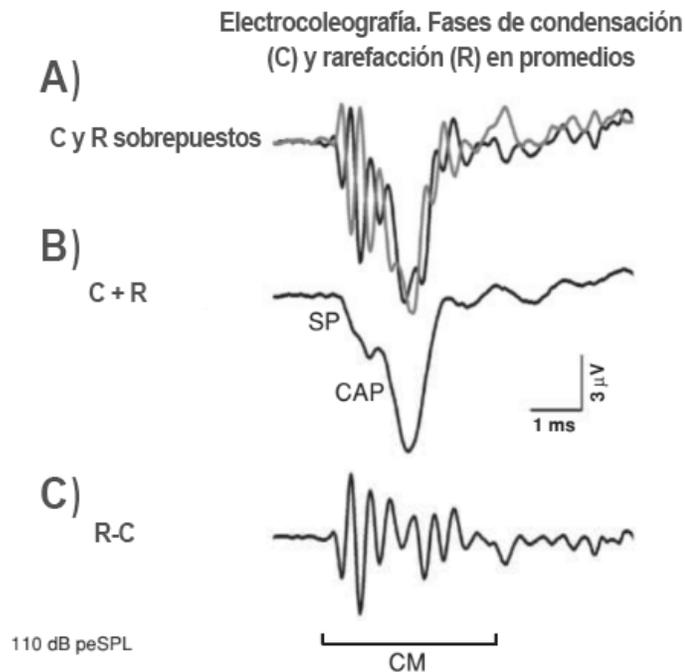


Figura 8: Procedimiento para separar los potenciales electricos de las células ciliadas y del nervio auditivo. Reproducido de: Katz, J et al. (2015). Handbook of clinical audiology. Hearing with a cochlear implant. Página: 226

Por otra parte, una alternativa para representar la señal es someterla a un análisis mediante la transformada de fourier discreta, como se ilustra en la Figura 9. En la sección A de la figura se presentan las señales evocadas de rarefacción y condensación, las cuales, al ser transformadas mediante esta técnica, producen las señales que se muestran en la sección D. Asimismo, al aplicar la transformada de fourier discreta a la respuesta de diferencia (MC), se obtiene la representación mostrada en la sección E. Finalmente, al procesar la respuesta de suma (CAP) con esta transformada, se obtiene la representación que se visualiza en la sección F.

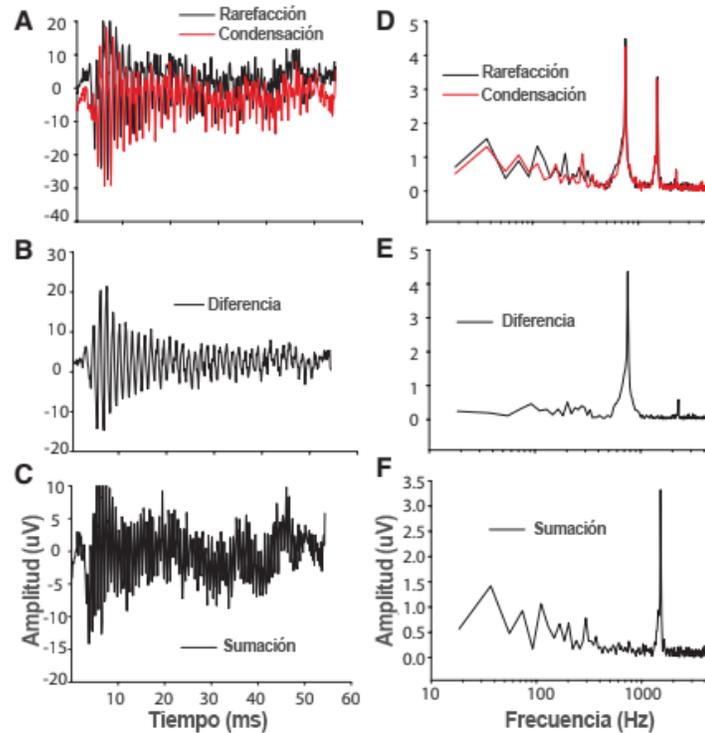


Figura 9: Respuesta obtenida a 750 Hz a una respuesta de diferencia y de suma. Reproducido de doi: 10.1097/AUD.0000000000000385

Al analizar la imagen, los resultados esperados para cualquier análisis de este tipo deberían ser similares, como es el caso particular aquí examinado, que corresponde a un tono de 750 Hz. Se observa que la respuesta de diferencia, que representa el microfónico coclear, presenta un claro peak en el primer armónico y un leve peak en el tercer armónico. Por otro lado, en la respuesta de suma, que indica el potencial de acción compuesto, se aprecia únicamente un peak en el segundo armónico. Este patrón de comportamiento es el que se espera para todas las frecuencias analizadas.

2.6.2. Actividad espontánea del nervio auditivo

La actividad espontánea del nervio auditivo se refiere a la actividad eléctrica que puede ser detectada en la ausencia de estímulos externos [30]. La importancia de esta actividad radica en que un nervio auditivo funcional es crucial para la transmisión de la información aferente al sistema auditivo central. Por lo tanto, la medición de la actividad espontánea del nervio auditivo se convierte en otro indicador valioso para evaluar la funcionalidad auditiva en una persona y garantizar una óptima percepción del sonido [31].

El estudio más reciente sobre este tema en humanos reveló un posible pick en 3100 Hz al aplicar un análisis inducido en las señales de los pacientes. Sin embargo, la actividad espontánea a esta frecuencia no puede ser confirmada de manera definitiva. Esto se debe a que solo la mitad de los pacientes de la muestra mostraron este pick en 3100 Hz, mientras que la otra mitad exhibió un comportamiento oscilatorio en el que también se observó un peak en esa misma frecuencia. Es importante destacar que este comportamiento oscilatorio se considera posiblemente artefactual debido a que todos los pacientes con este patrón mostraron la mis-

ma oscilación en las mismas frecuencias, sugiriendo una posible influencia de los dispositivos utilizados en la medición. A pesar de estas consideraciones, este estudio preliminar podría indicar que la frecuencia emitida por el nervio auditivo en humanos se encuentra alrededor de los 3100 Hz [17].

2.7. Estado del arte

A lo largo de los estudios orientados a descubrir la relación entre el entandimiento del habla y otras variables, se han estudiado multiples aristas que podrían afectar. En la siguiente tabla se puede observar en más detalle todos estos avances:

Tabla 2.3: Benchmark de las investigaciones sobre el entendimiento del habla

Nombre original	Año	Medición de las variables	Variables independientes más importantes	Variable dependiente	Modelo
Multivariate predictors of audiological success with multichannel cochlear implants	1993	Preoperatorias y postoperatorias	Duración de la sordera profunda	Entendimiento de habla, no hay una distinción entre hacerlo en ruido o en silencio	Regresión lineal
Predictors of audiological outcome following cochlear implantation in adults	2007	Preoperatorias y postoperatorias	Duración de la sordera profunda	Entendimiento de habla, no hay una distinción entre hacerlo en ruido o en silencio	Regresión lineal
Pre-, Peri- and Postoperative Factors Affecting Performance of Postlinguistically Deaf Adults Using Cochlear Implants: A New Conceptual Model over Time	2012	Preoperatorias, perioperatorias y postoperatorias	Umbral promedio de tonos puros del mejor oído, la marca del dispositivo, el porcentaje de electrodos activos, el uso de audífonos durante el período de sordera profunda y la duración de sordera moderada.	Entendimiento del habla en ruido y silencio	Regresión lineal generalizada

Round Window Electrocochleography Just Before Cochlear Implantation Relationship to Word Recognition Outcomes in Adults	2014	Perioperatoria	Electrococleografía de respuesta total	Entendimiento del habla en silencio	Regresión lineal
Round Window Electrocochleography and Speech Perception Outcomes in Adult Cochlear Implant Subjects: Comparison With Audiometric and Biographical Information	2014	Preoperatorias y perioperatorias	Electrococleografía de respuesta total	Entendimiento del habla en silencio	Regresión lineal
Further evidence of the relationship between cochlear implant electrode positioning and hearing outcomes	2019	Postoperatorio	Posicionamiento de los electrodos	Entendimiento del habla en ruido y silencio	Regresión lineal generalizada
Electrococleography and cognition are important predictors of speech perception outcomes in noise for cochlear implant recipients	2022	Postoperatoria	Electrococleografía de respuesta total y MoCa	Entendimiento del habla en ruido y silencio	Regresión lineal

El estudio más reciente, que también refleja el estado actual del problema, revela que el mejor modelo seleccionado es una regresión lineal que incorpora variables de electrocografía de respuesta total y una medida de nivel cognitivo conocida como MoCa². Esta información se encuentra resumida en la última fila de la tabla mencionada anteriormente. La aplicación de este modelo se realiza mediante la siguiente ecuación:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \varepsilon$$

donde:

- y : variable del entendimiento del habla (en silencio o en ruido).

² La medida de inteligencia MoCA, o Montreal Cognitive Assessment, es una herramienta de evaluación breve utilizada en entornos clínicos para detectar y monitorear el deterioro cognitivo leve.

- x_1 : electrocolegrafía de respuesta total (ECog-T).
- x_2 : MoCa.
- β_0 : intercepto.
- β_1 y β_2 : coeficientes de regresión que representan las pendientes de las variables independientes.
- ε : error aleatorio.

La idea de este modelo de regresión lineal es encontrar los valores de los coeficientes β_0 , β_1 y β_2 que minimizan la función de pérdida $J(\beta_0, \beta_1, \beta_2)$, dada por:

$$J(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

donde n es el número de observaciones, y_i es el valor observado, y \hat{y}_i es el valor predicho por el modelo. Una vez despejada la anterior ecuación, los valores de los coeficientes son:

$$\beta_1 = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)(y_i - \bar{y})}{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}$$

$$\beta_2 = \frac{\sum_{i=1}^n (x_{2i} - \bar{x}_2)(y_i - \bar{y})}{\sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}$$

$$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x}_1 - \beta_2 \cdot \bar{x}_2$$

donde \bar{x}_1 y \bar{x}_2 son las medias de las variables independientes x_1 y x_2 respectivamente, y \bar{y} es la media de y . Cabe mencionar que en el estudio descrito, cada paciente tenía solo una tupla del estilo (ECog-T, MoCa)

En relación a la implementación del estado del arte utilizando los datos disponibles en esta tesis, se emplearán exclusivamente los datos de electrocolegrafías para construir la variable asociada a las señales, dado que no se dispone de una variable como MoCa en el conjunto de datos. Por otro lado, es importante señalar que solo se cuenta con frecuencias de 500 Hz y 2000 Hz, lo que impide obtener una electrocolegrafía de respuesta total con los datos actualmente disponibles por lo que más adelante se describirán las estrategias implementadas para poder crear variables a pesar de las limitaciones existentes.

2.8. Aprendizaje de máquinas

2.8.1. Contexto

El aprendizaje de máquinas, un componente clave de la inteligencia artificial, se enfoca en el desarrollo de modelos que permiten a las computadoras aprender patrones y realizar tareas sin necesidad de programación explícita. En lugar de depender de instrucciones específicas, las máquinas utilizan datos para mejorar su rendimiento en una tarea específica a lo largo del tiempo.

Este proceso de aprendizaje de máquinas guarda similitudes con el aprendizaje humano en varios aspectos. La máquina se expone a un conjunto de datos de entrada y utiliza algoritmos para identificar patrones y relaciones en esos datos. A medida que más datos están disponibles, la máquina ajusta sus modelos, mejorando así su capacidad para hacer predicciones o tomar decisiones. Este enfoque resulta particularmente valioso en situaciones donde las reglas o patrones no son fácilmente programables de manera manual.

Dentro del ámbito del aprendizaje de máquinas, se destacan dos enfoques fundamentales: los modelos discriminativos y los modelos generativos. Mientras que los modelos discriminativos se centran específicamente en aprender la frontera de decisión entre clases o categorías para lograr clasificaciones o regresiones precisas, los modelos generativos van más allá. Estos últimos buscan comprender la distribución completa de los datos y tienen la capacidad de generar nuevas instancias que se asemejan a las observaciones de entrenamiento. La elección entre un modelo discriminativo y uno generativo depende de la naturaleza específica del problema y de los objetivos de la tarea en cuestión.

En esta situación particular, los modelos discriminativos y generativos desempeñan roles distintos a lo largo de diferentes fases. Inicialmente, se desarrollan modelos discriminativos con el propósito de hacer predicciones y establecer una referencia para evaluar la efectividad de los modelos. Es importante señalar que, para ampliar el análisis y mejorar la precisión de las predicciones, sería valioso explorar el uso de modelos más sofisticados que trasciendan las simples regresiones lineales empleadas hasta el momento. Alternativas como random forest, xgboost y redes neuronales podrían proporcionar beneficios significativos al capturar relaciones no lineales y manejar conjuntos de datos más complejos. Estos modelos tienen la capacidad de discernir patrones complejos en los datos y adaptarse a diferentes tipos de relaciones entre las variables, lo que potencialmente podría mejorar la capacidad predictiva del sistema en general.

Posteriormente, se introduce la perspectiva de modelos generativos, impulsada por los desafíos inherentes a los datos, los cuales se caracterizan por la disponibilidad limitada de información, en este caso debido al reducido número de pacientes reclutados. Este nuevo enfoque propone la generación de nuevas señales a partir de la información existente, superando así las limitaciones presentes en el conjunto de datos.

2.8.2. Modelos discriminativos

2.8.2.1. Modelos de regresión

Los modelos de regresión son componentes fundamentales en el ámbito del aprendizaje de máquinas, especialmente concebidos para resolver desafíos relacionados con la regresión. En términos sencillos, estos modelos proporcionan valores numéricos continuos dentro del rango establecido por los datos de entrenamiento. A modo de ejemplo se señalarán algunos modelos de regresión sin orden de importancia:

1. Regresión Lineal Simple y Múltiple
2. Regresión Regularizadas
3. árbol de decisión

4. Regresión de Bosques Aleatorios (Random Forest Regression)
5. XGBoost regresor
6. Regresión de Máquinas de Soporte Vectorial (Support Vector Regression)
7. Regresión de Vecinos más Cercanos (K-Nearest Neighbors Regression)
8. Redes Neuronales (Neural Network Regression)

En esta tesis, se emplean diversos modelos, incluyendo random forest, xgboost y redes neuronales feedforward, los cuales serán explicados en detalle más adelante para dar un contexto de cómo se aplican. Estos modelos se han seleccionado con el propósito de contrastar su desempeño con respecto al de modelo de regresión lineal mencionado previamente en la revisión del estado del arte del problema.

2.8.2.1.1. Árbol de decisión

Un árbol de decisión es una estructura jerárquica donde cada nodo representa una característica, cada rama representa una decisión basada en esa característica, y cada hoja representa el resultado de esa decisión. Un ejemplo de esto se puede observar en la Figura 10.

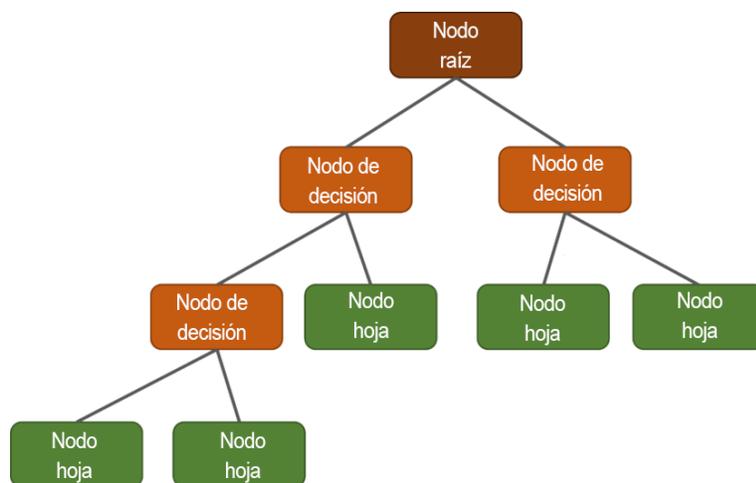


Figura 10: Ejemplo de árbol de decisión. Reproducido de <https://www.geo.fu-berlin.de>

En el caso de un árbol de decisión de regresión el resultado final de las hojas varía levemente a cómo se calcula en un árbol de decisión de clasificación, en donde las hojas es la clasificación deseada para el caso de la clasificación. En un árbol de decisión para regresión se calcula mediante el siguiente algoritmo el cual será ilustrado con el siguiente ejemplo en donde se quiere predecir la cantidad de horas jugadas:

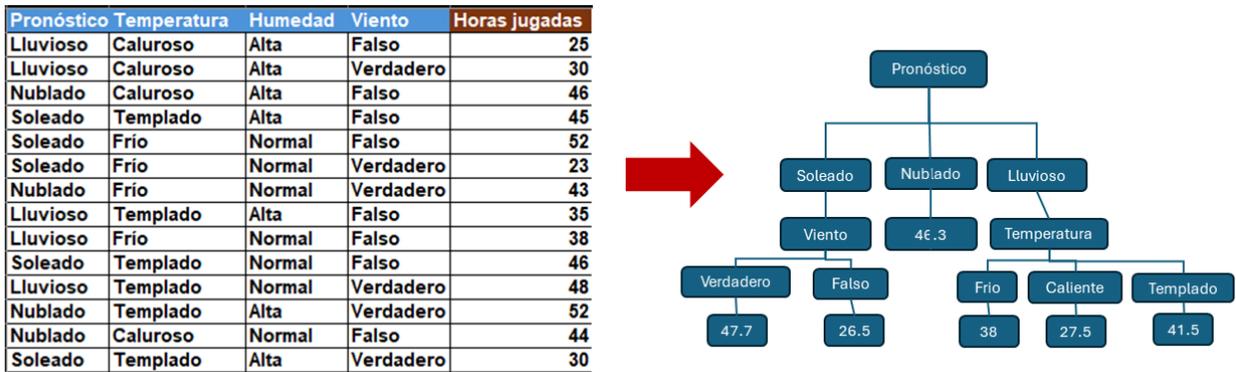


Figura 11: Ejemplo de problema de regresión con árbol de decisión. Reproducido de <https://saedsayad.com>

Algoritmo

- Un árbol de decisión se construye de arriba hacia abajo a partir de un nodo raíz e implica dividir los datos en subconjuntos que contienen instancias con valores iguales. Para dividir los datos por primera vez se calcula la desviación estándar de cada atributo de manera separada para así elegir el mejor candidato. La manera de hacerlo por atributo es calcular la desviación estándar de las horas jugadas por opciones de cada atributo. En el caso de pronóstico opción nublado según la muestra 46, 43, 52, 44 el resultado de la desviación estándar es igual a 3.49
- Luego se calcula la desviación estándar de cada una de las otras opciones y luego para calcular la desviación estándar del atributo en total se pondera por la cantidad de instancias que hay en la muestra, cómo se puede observar a continuación:

$$S(T, X) = \sum_{c \in X} P(c)S(c)$$

		Horas jugadas	Cantidad
Pronóstico	Nublado	3.49	4
	Lluvioso	7.78	5
	Soleado	10.87	5
			14

Figura 11: Ejemplo de cálculo de la desviación estándar por clases en una variable. Reproducido de <https://saedsayad.com>

$$S(horas, pronostico) = P(Soleado) * S(Soleado) + P(Nublado) * S(Nublado) + P(Lluvioso) * S(Lluvioso)$$

Reemplazando por valores:

$$= (4/14) * 3.49 + (5/14) * 7.78 + (5/14) * 10.87 = 7.66$$

- Los siguientes resultados son posteriores a haber calculado la desviación estándar por cada atributo:

		Horas jugadas			Horas jugadas	
Pronóstico	Nublado	3.49		Temp	Frio	10.51
	Lluvioso	7.78			Caluroso	8.95
	Soleado	10.87			Templado	7.65
		SDR=1.66		SDR=0.48		
		Horas jugadas			Horas jugadas	
Humedad	Alta	9.36		Viento	Falso	7.87
	Normal	8.37			Verdadero	10.59
		SDR=0.28		SDR=0.29		

Figura 12: Resultado de la desviación estándar por cada atributo. Reproducido de <https://saedsayad.com>

- Luego lo que se realiza es lo que se llama reducción de la desviación estándar que es en palabras simples, es restar la desviación estándar del atributo a predecir con la desviación estándar calculada por cada atributo. El que termina teniendo un valor mayor es el atributo que se elige, como se puede observar con el atributo pronóstico.

$$SDR(T, X) = S(T) - S(T, X)$$

$$SDR(Horas, Pronostico) = S(Horas) - S(Horas, Pronostico)$$

$$= 9.32 - 7.66 = 1.66$$

- Este proceso se itera por los atributos a menos que en alguna de las ramas de los nodos de decisión cumpla con algunos criterios de terminación como el coeficiente de desviación (CV) el cual se calcula como $\frac{DesvEs}{\hat{x}}$ y como la cantidad de ejemplos que se tiene por atributo. Si por ejemplo este fuera de un CV de 10% y una cantidad de 3 o menos ejemplos por atributos, se tendría lo siguiente:



Figura 14: Ejemplo de criterio de terminación de una rama de decisión. Reproducido de: <https://saedsayad.com>

La rama de nublado tendría un CV de 8% y por tanto, las decisiones por esa rama se cortan y se calcula como resultado de su hoja la media de los atributos que cayeron en esa categoría, los cuales son 46, 43 , 52 y 44:

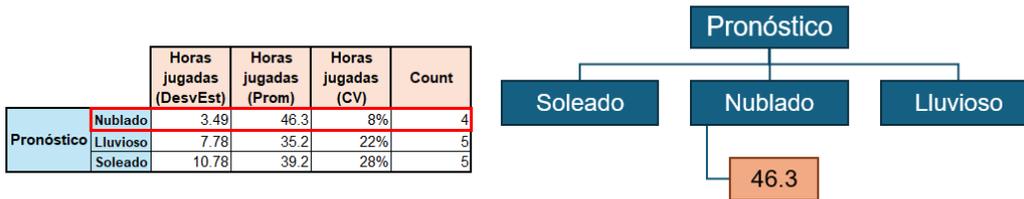


Figura 15: Cálculo de los criterios de terminación de una rama. Reproducido de: <https://saedsayad.com>

- Como las otras ramas de ese atributos tienen CV mayores a un 10% se sigue el proceso por cada atributo, pero como los siguientes atributos tienen 3 o menos ejemplos, se separa por las clases del atributo y se calcula el promedio de los que quedan como nodo hoja.

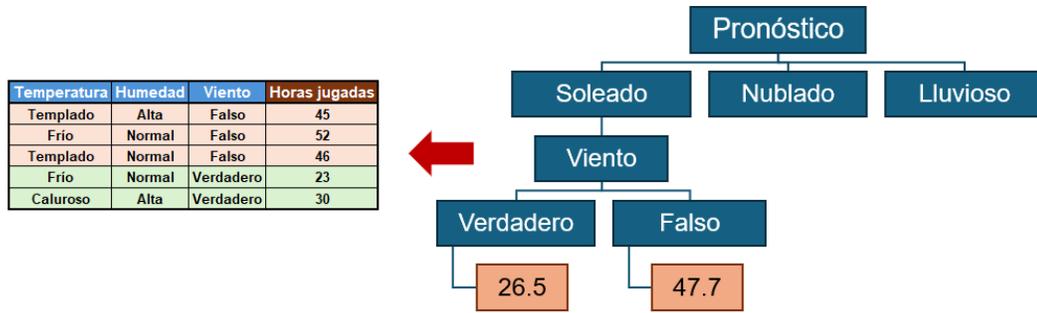


Figura 16: Ejemplo de las hojas del árbol. Reproducido de: <https://saedsayad.com>

- Por último, el árbol queda de la siguiente manera:

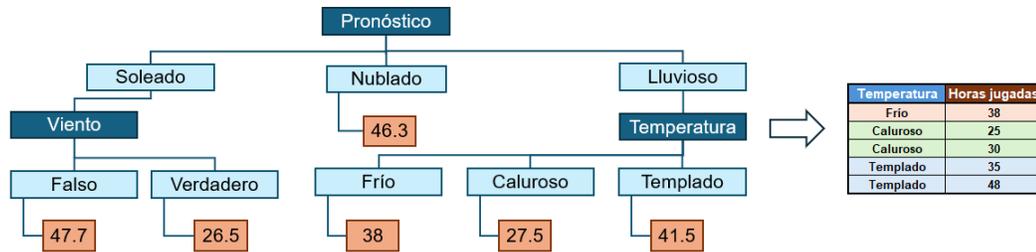


Figura 17: Ejemplo final de cómo queda el árbol de decisión. Reproducido de: <https://saedsayad.com>

2.8.2.1.2. Random forest regression

El algoritmo Random Forest Regression es una técnica de aprendizaje supervisado utilizada para problemas de regresión. Se basa en la construcción de múltiples árboles de decisión durante el entrenamiento y la combinación de sus predicciones para obtener una predicción más robusta y generalizable. Referente a este modelo, hay dos conceptos importantes para tener en cuenta, los cuales son el bootstrapping y el ensamblaje.

El bootstrapping implica crear múltiples muestras (llamadas muestras bootstrap) mediante el muestreo aleatorio con reemplazo de los datos disponibles. Cada muestra bootstrap tiene el mismo tamaño que la muestra original, pero puede contener duplicados de observaciones y perder otras observaciones en el proceso de reemplazo.

Por otro lado, el ensamblaje combinan las predicciones de varios modelos individuales para mejorar la precisión y el rendimiento general del sistema. Estos modelos aprovechan la diversidad y la independencia de múltiples modelos para mitigar las limitaciones y debilidades de cada uno por separado.

Algoritmo

- Se selecciona aleatoriamente un conjunto de características (bootstrapping) del conjunto de datos de entrenamiento. Se construye un árbol de decisión utilizando solo ese subconjunto de características. Este árbol se construye de manera recursiva dividiendo los

datos en nodos basándose en la característica que proporciona la mejor separación según algún criterio (como la reducción de desviación estándar).

- Se repite el paso anterior para construir múltiples árboles, cada uno con un conjunto diferente de características seleccionadas aleatoriamente.
- Con todos los árboles de decisión creados, se pasa la instancia que se quiere predecir y se obtiene la predicción de la instancia x del árbol i , $f(x)_i$.
- Al tener todas las predicciones de la instancia x en todos los árboles se calcula un promedio y ese es el resultado de esa instancia

$$\hat{y}(x) = \frac{1}{N} \sum_{i=1}^N f_i(x)$$

2.8.2.1.3. XGBoost regresor

El modelo XGBoost (Extreme Gradient Boosting)[32] es un algoritmo de aprendizaje automático que utiliza árboles de decisión para realizar tareas de regresión. El modelo se construye como una combinación ponderada de árboles de decisión. Cada árbol se agrega secuencialmente al modelo, y cada árbol se entrena para corregir los errores del modelo existente. La predicción final se obtiene sumando las predicciones de todos los árboles en el conjunto:

$$\hat{y} = \sum_{k=1}^K f_k(x_i)$$

donde:

- K : representa la cantidad de árboles.
- $f_k(x_i)$: representa el resultado del árbol k a la instancia x_i .

Cabe destacar que el modelo de XGBoost está basado en Gradient Tree Boosting que a su vez está basado en un modelo de Tree Boosting que en palabras simples es un modelo secuencial que también tiene como resultado la suma de todos los árboles creados. Donde en cada iteración en la cual se construye un árbol se intenta minimizar la función objetivo que cuenta con una función de pérdida y con términos de regularización. La función objetivo es:

$$\mathcal{L}(\phi) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

donde:

- n es el número de observaciones en el conjunto de entrenamiento.
- y_i es el valor real de la variable objetivo para la observación i .
- \hat{y}_i es la predicción del modelo para la observación i .
- K es el número total de árboles en el modelo.
- L es la función de pérdida que mide la diferencia entre y_i y \hat{y}_i .

- $\Omega(f_k)$ son términos de penalización que controlan la complejidad del árbol k .

Debido a que la función anteriormente descrita no puede ser optimizada mediante métodos de optimización tradicionales en el espacio euclidiano, Gradient Tree Boosting se crea como una versión mejorada para resolver esto en donde se redefine la predicción de \hat{y}^t donde lo predicho es la suma de los árboles anteriores más el valor de la hoja respectiva. Esto queda de la siguiente manera:

$$\mathcal{L}(\phi)^{(t)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

La variación de XGBoost con Gradient Tree Boosting es que aproxima el resultado anterior a una aproximación de Taylor de segundo orden, provocando que el computo de los resultado sea mucho más rápido, el cual queda de la siguiente manera:

$$\mathcal{L}(\phi)^{(t)} \simeq \tilde{\mathcal{L}}(\phi)^{(t)} = \sum_{i=1}^n \left[L(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

donde $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ es el gradiente de primer orden y $h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)})$ es el gradiente del segundo orden en la función de pérdida. Por otro lado, $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$.

Después de denotar $I_j = \{i \mid q(x_i) = j\}$ como el conjunto de instancias de la hoja j , se simplifica la ecuación anterior y quedade la siguiente manera:

$$\tilde{O}^{(t)} = \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T$$

Así, la solución para el peso óptimo ω_j^* de una hoja j para una estructura de árbol fija $q(x)$ se puede obtener aplicando la siguiente fórmula:

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

Donde, sustituyendo ω_j^* en la ecuación anterior se obtiene:

$$\tilde{O}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T$$

La ecuación anterior representa una métrica de puntuación que se utiliza para evaluar la configuración del árbol $q(x)$ con el fin de identificar las configuraciones óptimas para la clasificación. No obstante, en la realidad, resulta inviable explorar todas las posibles configuraciones del árbol q . En el trabajo de investigación de [33], se propone un algoritmo "codicioso" que comienza con una única hoja y , de manera iterativa, agrega ramificaciones para expandir la estructura del árbol. La decisión de agregar una división a la estructura del árbol existente puede determinarse mediante la siguiente función:

$$O_{split} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma$$

Donde I_{izq} y I_{der} representan los conjuntos de instancias de los nodos izquierdo y derecho, respectivamente, tras la división, y I denota la unión de I_{izq} y I_{der} .

2.8.2.1.4. Red neuronal feedforward fully connected

Un modelo de regresión utilizando una red neuronal feedforward fully connected, se puede ver como una composición de funciones que se van realizando por capas. Para explicar bien este problema se presentará una imagen de su estructura básica, la cual consiste de capa de entrada, capas ocultas, capa de salida, neuronas, pesos, bias y función de activación:

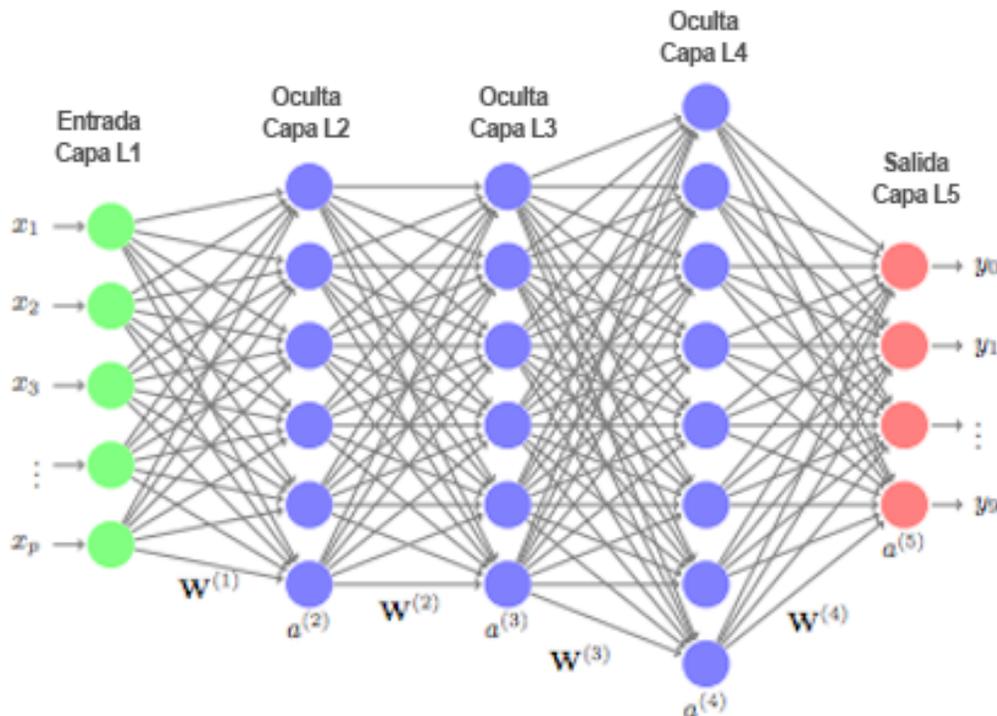


Figura 18: Ejemplo de red neuronal con 5 capas. Reproducido de: <https://afit-r.github.io>

En la imagen se pueden observar diferentes letras y colores que consisten en diferentes partes del modelo. Las pelotas de color verde al inicio forman la capa de entrada y cada una de las pelotas conforman las neuronas de la capa de entrada, luego las pelotas de color morado son las neuronas de las capas ocultas, donde en este caso en particular son 3 capas ocultas y por último, las neuronas de color rojo son las neuronas de la capa de salida. Por otro lado vemos letras W que consisten en los pesos del modelo, además faltan los bias que consisten en un escalar que se suma al resultado de cada neurona. La función que modela el valor de cada neurona según la capa l y la función de activación de la capa l viene dada por lo siguiente:

$$x^{(l)} = f(\mathbf{W}^{(l)} \mathbf{x}^{(l-1)} + b^{(l)})^{(l)}$$

Es importante resaltar que el proceso de asignar valores a cada una de las neuronas se denomina *feedforward*. Al realizar este proceso por primera vez, los valores de $W^{(l)}$ y $b^{(l)}$ son

asignados de manera aleatoria. Como es característico en todos los modelos de aprendizaje de máquinas, nuestro objetivo radica en ajustar los parámetros del modelo de manera que este pueda aprender de los datos. La variación de estos parámetros se logra al reducir el error entre la variable predicha y la etiqueta real de los datos. En términos matemáticos, la función que buscamos minimizar es la siguiente:

$$Loss(\hat{y}, y) = \frac{1}{N} \sum_0^N error(\hat{y}, y)$$

La correcta adaptación de los parámetros en una red neuronal se logra a través del método de descenso de gradiente. De manera sencilla, este proceso implica situarse en un punto $W^{(l)}$ y determinar en qué dirección moverlo para reducir el término $Loss(\hat{y}, y)$. El descenso de gradiente busca llevar a cabo este ajuste para todos los parámetros del modelo. Sin embargo, ejecutar este procedimiento no es trivial; se necesita identificar la dirección adecuada para ajustar los parámetros con el propósito de minimizar la función $Loss(\hat{y}, y)$. La obtención de todas las derivadas de los parámetros con respecto a la función de pérdida se realiza mediante un algoritmo conocido como **backpropagation**, el cual calcula todas las derivadas parciales de cada parámetro.

Es importante destacar que el descenso de gradiente identifica mínimos locales, ya que la cantidad de parámetros asociados a una red es considerable, y, dadas las limitaciones de tiempo de entrenamiento y capacidad computacional, se asegura al menos la convergencia a un mínimo local en la función de $Loss(\hat{y}, y)$. La complejidad de este proceso, se puede observar en la Figura 19 donde se puede apreciar cómo los mínimos locales pueden confundirse con mínimos globales en el algoritmo. La fórmula que define el descenso de gradiente es la siguiente:

$$W_{new} = W_{old} - \alpha \frac{dW}{dJ}$$

Después de abordar en detalle los procedimientos fundamentales en la red neuronal, que incluyen el feedforward y el backpropagation, se desvela el proceso esencial del entrenamiento en esta estructura. La secuencia comienza con una iteración de feedforward, utilizando valores aleatorios para los parámetros. Posteriormente, se lleva a cabo el descenso de gradiente empleando las derivadas parciales calculadas a través del backpropagation. Esta sucesión de feedforward y descenso de gradiente se repite según la necesidad de procesar todo el conjunto de datos a través de la red, con el objetivo de lograr una minimización óptima del error de la red.

2.8.2.2. Modelos de clasificación

Los modelos de clasificación desempeñan un papel fundamental en el campo del aprendizaje de máquinas, siendo diseñados de manera específica para resolver problemas de clasificación. En contraste con los modelos de regresión, cuyo enfoque radica en la predicción de valores continuos, los modelos de clasificación están destinados a asignar etiquetas o categorías a las observaciones en función de atributos específicos. A modo de ejemplo se señalarán algunos modelos sin orden de importancia:

1. Regresión Logística
2. Bosques Aleatorios (Random Forest)

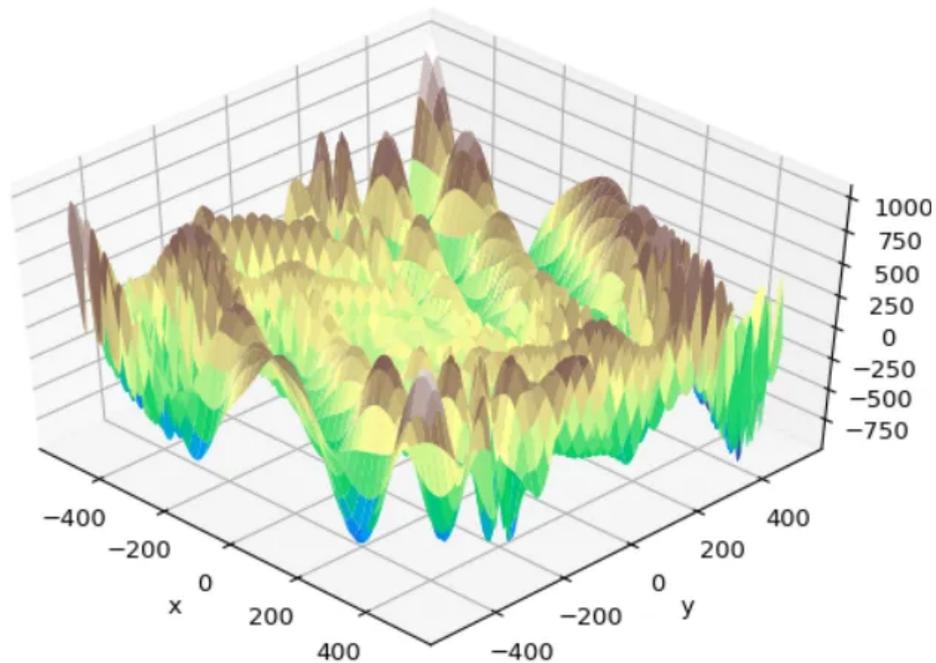


Figura 19: Representación de mínimos locales y mínimos globales cuando la dimensionalidad de las variables es de 3. Reproducido de: <https://medium.com>

3. Vecinos más Cercanos (K-Nearest Neighbors)
4. Redes Neuronales
5. Máquinas de Vectores de Soporte (SVM) Multiclase
6. Naive Bayes
7. XGBoost clasificador

En particular, en esta tesis se utilizaron los modelos de Random Forest, XGBoost y Redes Neuronales.

2.8.2.2.1. Random forest classifier

El algoritmo Random Forest Classifier es una técnica de aprendizaje supervisado utilizada para problemas de clasificación. Se basa en la construcción de múltiples árboles de decisión durante el entrenamiento y la combinación de sus predicciones para obtener una clasificación más robusta y generalizable. Al igual que en su versión regresora, se siguen ocupando los métodos de bootstrapping y de ensamblaje para la construcción de árboles.

Algoritmo

- Se selecciona aleatoriamente un conjunto de características (bootstrapping) del conjunto de datos de entrenamiento. Se construye un árbol de decisión utilizando solo ese subconjunto de características. Este árbol se construye de manera recursiva dividiendo los

datos en nodos basándose en la característica que proporciona la mejor separación según algún criterio (como la ganancia de información o la impureza Gini).

- Se repite el paso anterior para construir múltiples árboles, cada uno con un conjunto diferente de características seleccionadas aleatoriamente.
- Con todos los árboles de decisión creados, se pasa la instancia que se quiere clasificar y se obtiene la clasificación de la instancia x del árbol i , $c(x)_i$.
- Al tener todas las clasificaciones de la instancia x en todos los árboles, se realiza una votación (mayoría) para determinar la clase final asignada a la instancia

$$\hat{y}(x) = \text{votación}\{c_i(x)\}$$

2.8.2.2.2. XGBoost classifier

El algoritmo de XGBoost para clasificación es similar en a su versión regresora con la diferencia de que la función de pérdida cambia. Para un problema de clasificación multinomial con z clases, se utiliza la pérdida de entropía cruzada categórica, también conocida como la pérdida softmax. Esta función de pérdida se utiliza para calcular la penalización asociada con las predicciones incorrectas en problemas de clasificación multiclase.

La fórmula de la pérdida de entropía cruzada categórica para un problema de z clases se expresa como:

$$\text{Pérdida} = - \sum_{i=1}^n \sum_{k=1}^K y_{i,k} \cdot \log(p_{i,k})$$

Donde:

- n es el número de instancias de datos.
- K es el número de clases (en este caso, $K = z$).
- $y_{i,k}$ es un valor binario que indica si la instancia i pertenece o no a la clase k . Por lo tanto, $y_{i,k}$ es 1 si la instancia i pertenece a la clase k y 0 de lo contrario.
- $p_{i,k}$ es la probabilidad predicha de que la instancia i pertenezca a la clase k .

Esta función de pérdida penaliza las predicciones incorrectas con un valor proporcional a la diferencia entre la predicción y la etiqueta verdadera. Al igual que en el caso binario, cuando la predicción se acerca más a la etiqueta verdadera, la pérdida tiende a cero, y cuando la predicción es incorrecta y está lejos de la etiqueta verdadera, la pérdida es alta.

XGBoost utilizará esta función de pérdida durante el proceso de entrenamiento para ajustar los parámetros del modelo, minimizando la pérdida total sobre el conjunto de entrenamiento.

2.8.2.2.3. Red neuronal feedforward fully connected classifier

La única diferencia entre su versión regresora y la versión clasificadora radica en la configuración específica de la última capa de la red neuronal. En el contexto de la clasificación, la función de activación utilizada en la última capa se adapta para abordar las necesidades

particulares del problema. Por ejemplo, podría emplearse la función softmax, cuya fórmula está dada por:

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}},$$

donde z_i es la entrada para la clase i y K es el número total de clases. La función softmax asigna probabilidades normalizadas a cada clase, garantizando que la suma de estas probabilidades sea igual a 1. Esto es esencial en problemas de clasificación multiclase, donde se busca determinar la probabilidad de pertenencia a cada categoría.

Esta elección en la función de activación tiene una consecuencia directa en la función de pérdida utilizada durante el entrenamiento de la red neuronal para clasificación. En lugar de las funciones de pérdida típicas de regresión, como la pérdida cuadrática o la pérdida absoluta, se emplea comúnmente la entropía cruzada categórica. La fórmula de la entropía cruzada categórica se ajusta a la naturaleza categórica de las predicciones y los objetivos en problemas de clasificación, siendo crucial para evaluar la discrepancia entre las distribuciones de probabilidad predichas y las distribuciones reales de las clases. La función de la entropía cruzada con K clases es la siguiente:

$$\text{Pérdida} = - \sum_{i=1}^n \sum_{k=1}^K y_{i,k} \cdot \log(p_{i,k})$$

Donde:

- n es el número de instancias de datos.
- K es el número de clases.
- $y_{i,k}$ es un valor binario que indica si la instancia i pertenece o no a la clase k . Por lo tanto, $y_{i,k}$ es 1 si la instancia i pertenece a la clase k y 0 de lo contrario.
- $p_{i,k}$ es la probabilidad predicha de que la instancia i pertenezca a la clase k .

2.8.3. Modelos generativos

Los modelos generativos son herramientas en el campo del aprendizaje de máquinas que se enfocan en entender y modelar la estructura subyacente de un conjunto de datos con el objetivo de generar nuevas instancias que compartan características similares con los datos originales. Estos modelos son fundamentales en problemas donde se busca comprender la distribución de los datos y, a partir de esa comprensión, crear nuevas muestras que se asemejen a las observaciones existentes. Algunos ejemplos de estos modelos son los siguientes:

- Autoencoder
- VAE (Variational Autoencoder)
- GPT (Generative Pre-trained Transformer)
- GAN (Generative Adversarial Network)

2.8.3.1. Autoencoder

Los autoencoders son arquitecturas de redes neuronales que se emplean con frecuencia para aprender representaciones eficientes y compactas de conjuntos de datos de alta dimensión. Estos modelos funcionan a través de una estructura que consiste en capas de codificación (encoder), donde la información se reduce a su forma más esencial, seguida de capas de decodificación (decoder) que buscan reconstruir la entrada original. La singularidad de los autoencoders radica en su capacidad para capturar patrones significativos y resumir la información esencial del conjunto de datos.

Específicamente, el autoencoder generativo va más allá al incorporar la capacidad de generar nuevos ejemplos que comparten características con los datos originales. Esto implica que el autoencoder no solo aprende a comprimir la información, sino que también puede crear instancias adicionales que son consistentes con las características aprendidas. En resumen, el autoencoder no solo es una herramienta para representar datos de manera eficiente, sino que también actúa como un generador capaz de producir muestras nuevas basadas en la estructura subyacente del conjunto de datos original. Un ejemplo de cómo puede ser la red es la siguiente:

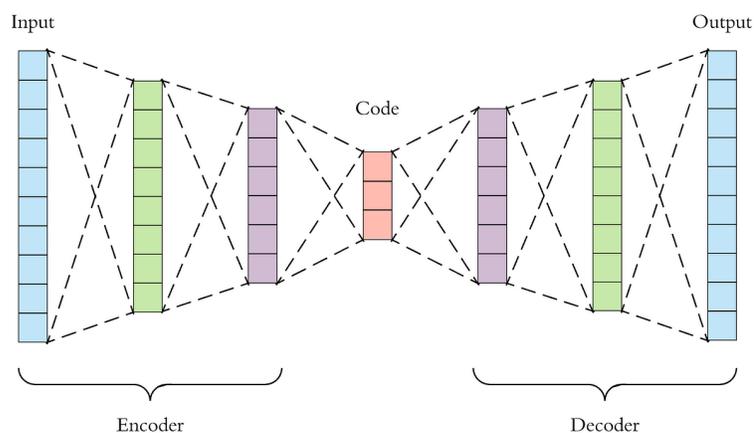


Figura 20: Estructura de un autoencoder. Reproducido de: <https://towardsdatascience.com/>

2.8.4. Métricas de desempeño

Las métricas de desempeño son herramientas críticas que permiten evaluar y cuantificar la calidad y eficacia de los modelos predictivos. Estas métricas proporcionan una medida objetiva de qué tan bien un modelo se ajusta a los datos y cómo realiza sus predicciones en comparación con la realidad. La elección de métricas apropiadas es esencial para comprender el rendimiento del modelo y tomar decisiones informadas sobre su utilidad en un contexto específico.

En el caso de modelos de regresión, algunas de las métricas que se suelen utilizar son el **MSE**, **RMSE**, **MAE** y **MAPE**, las cuales se calculan como:

MSE:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

RMSE:

$$\sqrt{MSE}$$

MAE:

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

MAPE:

$$\frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Donde y_i representa el valor real de la etiqueta en el índice i , y la variable \hat{y}_i representa el valor predicho por el modelo para el índice i , por último, n representa el número total de la muestra.

Por otra parte, los modelos de clasificación tienen otro tipo de métricas que están basadas en cálculos hechos en base a la matriz de confusión de los modelos. El ejemplo de matriz de confusión ilustrado en la Figura 21 es para ejemplos donde la clasificación es binaria, consiste en filas que representan las clases predichas por el modelo, mientras que las columnas representan las clases reales del modelo. La matriz se divide en cuatro secciones:

	Clase Positiva	Clase Negativa
Predicción Positiva	TP	FP
Predicción Negativa	FN	TN

Figura 21: Matriz de Confusión.

Donde cada sección de la matriz se define de la siguiente manera:

1. Verdaderos Positivos (TP): Representa la cantidad de instancias que fueron correctamente clasificadas como positivas por el modelo.
2. Verdaderos Negativos (TN): Representa la cantidad de instancias que fueron correctamente clasificadas como negativas por el modelo.
3. Falsos Positivos (FP): Representa la cantidad de instancias que fueron incorrectamente clasificadas como positivas por el modelo.
4. Falsos Negativos (FN): Representa la cantidad de instancias que fueron incorrectamente clasificadas como negativas por el modelo.

Algunas de las métricas calculadas a partir de la matriz de confusión son las de **Accuracy**, **Recall**, **Precision**, **F1-Score**, donde cada una se calcula de la siguiente manera:

Accuracy:

$$\frac{TP + TN}{TP + FP + FN + TN}$$

Recall (Sensibilidad):

$$\frac{TP}{TP + FN}$$

Precision:

$$\frac{TP}{TP + FP}$$

F1-Score:

$$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Cuando el número de clases de los modelos es superior a 2, las métricas se calculan individualmente para cada clase, generando múltiples resultados correspondientes a cada una. Para consolidar esta información, se utilizan técnicas conocidas como micro average y macro average, las cuales proporcionan un único valor para las métricas de Recall, Precision y F1-Score, ofreciendo así una síntesis representativa de los resultados globales. Lo anteriormente mencionado se calcula de la siguiente manera:

Micro Average: Para calcular el micro average de recall, precision y F1-score, primero sumas los verdaderos positivos (TP), falsos positivos (FP) y falsos negativos (FN) de todas las clases. Luego, aplicas las fórmulas correspondientes:

$$\text{Micro Average Recall} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)}$$

$$\text{Micro Average Precision} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FP_i)}$$

$$\text{Micro Average F1-Score} = \frac{2 \times \text{Micro Average Precision} \times \text{Micro Average Recall}}{\text{Micro Average Precision} + \text{Micro Average Recall}}$$

Donde n es el número de clases.

Macro Average: Para el macro average, calculas las métricas por clase y luego tomas el promedio aritmético.

$$\text{Macro Average Recall} = \frac{1}{n} \sum_{i=1}^n \text{Recall}_i$$

$$\text{Macro Average Precision} = \frac{1}{n} \sum_{i=1}^n \text{Precision}_i$$

$$\text{Macro Average F1-Score} = \frac{1}{n} \sum_{i=1}^n \text{F1-Score}_i$$

Donde n es el número de clases.

Capítulo 3

Resultados

3.1. Comprensión de los datos

La información utilizada en esta tesis proviene de datos recopilados para una investigación previa que analizaba las señales mediante un enfoque inducido [17]. Es fundamental destacar que la recopilación de datos no fue óptima para este estudio, ya que estos no se obtuvieron con el propósito de predecir el habla, sino de estudiar las señales en busca de información adicional, como la actividad espontánea del nervio auditivo.

La ineficiencia en la recopilación de datos se atribuye a la omisión de frecuencias clave en comparación con estudios más recientes que han demostrado mejores resultados en el cálculo de la electrocoqueografía de respuesta total. Estos estudios consideran frecuencias de 250 Hz, 500 Hz, 750 Hz, 1000 Hz, 2000 Hz y 4000 Hz [9] [16]. Sin embargo, en este caso específico, únicamente se registraron datos correspondientes a 500 Hz y 2000 Hz.

El protocolo de adquisición de datos utilizado en este estudio involucró a un grupo de 20 pacientes. Durante el procedimiento, cada paciente fue expuesto de manera individual a tonos puros de 500 Hz y 2000 Hz, mientras se registraba la respuesta interna de la cóclea a estos estímulos sonoros. Este proceso dio como resultado la obtención de las señales electrocoqueográficas.

La prueba tuvo una duración de 3 minutos para cada tono, durante los cuales se obtuvieron aproximadamente 1800 muestras por fase de condensación y rarefacción. Cada señal se representa mediante una tupla de voltaje frente al tiempo, con un total de 466 puntos en cada señal, es decir, 466 mediciones de voltaje a lo largo de los 3 minutos.

En el diagrama que se presenta a continuación, se destaca en verde el momento específico en el cual se recopilaron los datos que se usaron para esta investigación.

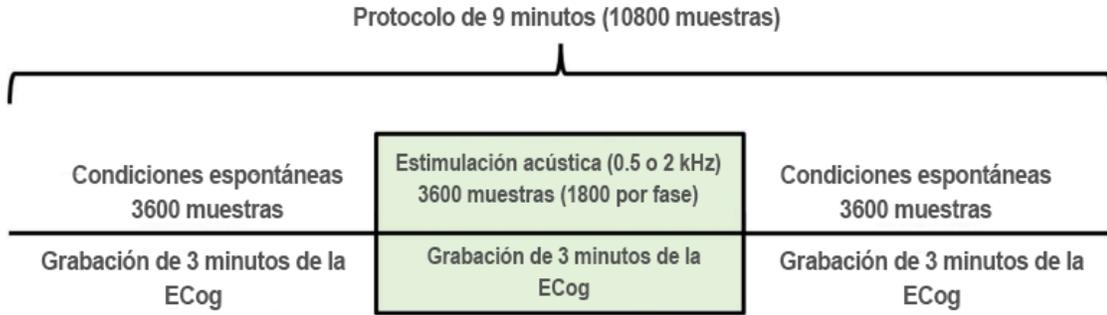


Figura 21: Protocolo de obtención de datos. Reproducido de doi: 10.1371/journal.pone.0299911

Las curvas de voltaje versus tiempo de cada paciente a cada frecuencia se pueden observar en el Anexo A y B resumidas en un promedio por fase. Un ejemplo de ellas es la del paciente 1 donde la fase azul es la de condensación y la naranja es la de rarefacción:

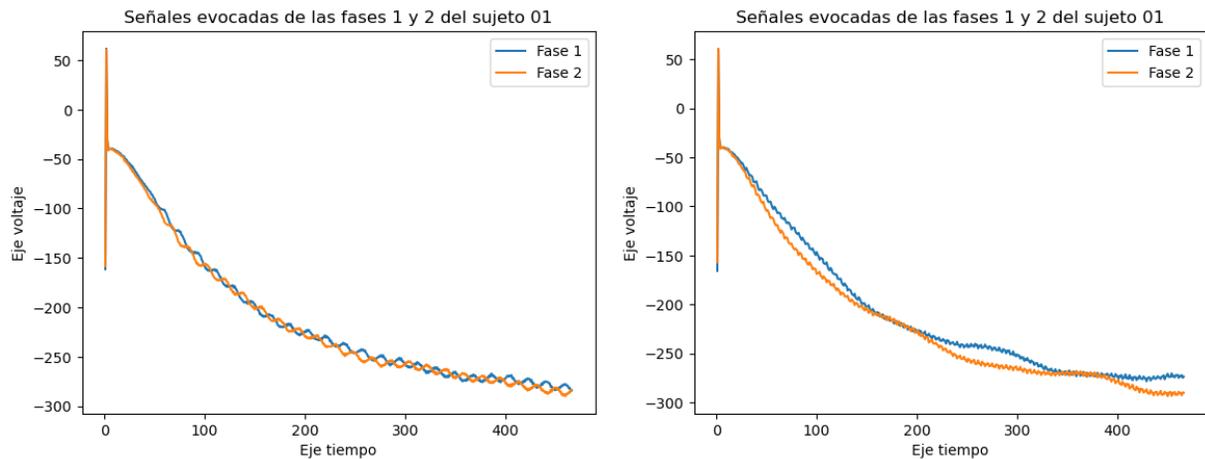


Figura 22: Voltajes a través del tiempo del paciente 1

Por otro lado, se dispone de etiquetas que reflejan el 'entendimiento del habla en silencio' y el 'entendimiento del habla en ruido', representadas en una escala del 0% al 100%. Estas etiquetas fueron generadas a través de la prueba "Matrix Test", la cual se llevó a cabo simultáneamente con la adquisición de las señales de electrococcleografía. Este test consiste en la repetición de palabras en entornos tanto de silencio como de ruido, calculando posteriormente el porcentaje de respuestas correctas por parte de los pacientes. Para obtener una visión detallada de los resultados, se remite a la Figura 25.

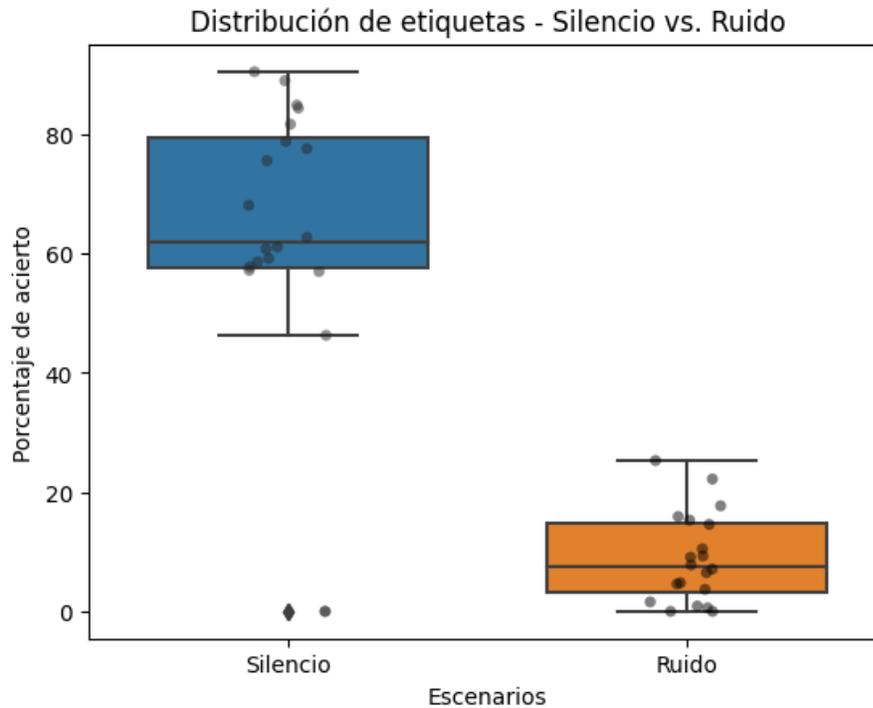


Figura 23: Gráfico de cajas sobre la prueba Matrix test en los escenarios de silencio y ruido

3.2. Análisis y selección de datos

3.2.1. Análisis de los datos

Al analizar los datos de las electrococcleografías es importante destacar la presencia de un artefacto en estos datos, manifestado por un inicio de la señal con un voltaje más elevado que su presentación a lo largo del tiempo. Aunque este artefacto no es una medición deseada, forma parte inherente del experimento.

En cuanto a las etiquetas, se destaca la presencia de pacientes cuyos resultados en las pruebas son nulos (0%) tanto en condiciones de silencio como en entornos ruidosos, a pesar de mostrar respuestas en las electrococcleografías. Esto es particularmente evidente en los pacientes 3 y 17, lo que sugiere la posible presencia de neuropatía auditiva en estos individuos. Esta condición se caracteriza por un funcionamiento adecuado en la cóclea, pero con el nervio auditivo incapaz de transmitir la información al cerebro. Esta discrepancia se refleja en el gráfico de las electrococcleografías, donde se observa la estimulación pero las pruebas de habla no registran aciertos.

Adicionalmente, se observa una limitación en la cantidad de datos disponibles. No obstante, esta circunstancia es común en estudios médicos debido al elevado costo de algunos exámenes o, en este caso, la dificultad para encontrar pacientes que cumplan con las características específicas necesarias, como aquellos con un implante coclear.

3.2.2. Selección de los datos

En cuanto a las electrocolegrafías, debido al efecto artefactual, se decide eliminar los primeros veinte puntos registrados en ambas fases durante el procesamiento para así eliminar el efecto que se producía al inicio de la señal. Y en cuanto a las etiquetas, los pacientes que presentaban este posible comportamiento de neuropatía, se decide eliminarlos de la muestra porque se escapan de lo que se quiere estudiar en esta investigación, que es finalmente el estado de la cóclea y cómo esta afecta el entendimiento del habla. Estos pacientes al tener problemas en el nervio auditivo no pueden entender aunque la cóclea funcione en perfecto estado.

3.3. Preprocesamiento de los datos

Debido a la naturaleza del problema, es necesario hacerse cargo de dos tareas en el preprocesamiento de los datos. La primera consiste en generar una representación de la señal que tenga que ver con el objetivo estudiado y por otro lado, generar variables relevantes para que un modelo aprenda.

3.3.1. Representación de la señal

La correcta representación de una señal es esencial para facilitar análisis futuros relacionados con la misma. Por lo tanto, resulta de vital importancia que la señal que pretendemos analizar esté directamente vinculada a la magnitud que deseamos medir, en este caso, la energía generada por el oído interno al ser estimulado a diferentes frecuencias. Afortunadamente, en el ámbito del procesamiento de señales detallado en la sección de *Procesamiento de señales*, contamos con herramientas específicas que desempeñan un papel fundamental en este proceso.

La Transformada de Fourier, por un lado, realiza de manera precisa la conversión de una señal desde el dominio temporal al de las frecuencias. Por otro lado, la Identidad de Parseval desempeña una función crucial al calcular la energía asociada a cada componente de frecuencia presente en la señal. Estas herramientas se integran de manera coherente con el propósito de lograr una representación integral de la señal, abarcando tanto sus características frecuenciales como energéticas.

En consecuencia, cualquier representación de la cual se extraigan características para entrenar modelos se deriva de este resultado de señal que proporciona la energía asociada a cada componente de frecuencia presente.

3.3.2. Feature engineering

Debido a la naturaleza de las variables destinadas a la predicción, es decir, las señales, nos enfrentamos a la presencia de numerosos componentes de señal que podrían contribuir únicamente con ruido. Para abordar este desafío, una estrategia consiste en construir variables en torno a la magnitud que deseamos medir, en este caso, los potenciales eléctricos generados en el oído interno. En la literatura, se identifican dos enfoques principales para el análisis de estos potenciales eléctricos: análisis evocado y análisis inducido. A través de dichos análisis, es posible obtener los potenciales eléctricos en cuestión.

3.3.2.1. Análisis evocado

A partir del análisis evocado detallado en la sección de *Procesamiento de señales*, se generan los potenciales eléctricos evocados del micrófono coclear y el potencial de acción compuesto, tal como se describe en la sección de *Conceptos de las electrocolegrafías*, para las frecuencias específicas de 500 Hz y 2000 Hz presentes en los datos disponibles. Y luego de eso, se lleva a cabo el cálculo de la energía correspondiente a cada componente de señal previamente mencionado en la sección de *Procesamiento de señales*, utilizando las señales resultantes. Este procedimiento asegura la precisa representación mencionada anteriormente.

En cuanto a los resultados generados por el microfónico coclear y el potencial de acción compuesto, se espera observar un peak en el microfónico coclear que coincida con la frecuencia del tono, mientras que en el potencial de acción compuesto se anticipa un peak en el segundo armónico. La anticipación se fundamenta en la información expuesta en la sección de *Conceptos de las electrocolegrafías*. Por lo tanto, para observar si efectivamente esas respuestas se encuentran en los datos de la investigación se observan las respuestas evocadas de las figuras de los Anexos C, D, E y F

Al examinar los potenciales eléctricos de las dos frecuencias, se evidencia que la respuesta del microfónico coclear es consistentemente perceptible en la mayoría de los casos. En contraste, la respuesta del potencial de acción compuesto muestra una presencia limitada a la frecuencia de 500 Hz en algunos casos, y no es discernible en ningún paciente a la frecuencia de 2000 Hz. A partir de este análisis, se concluye que no se procederá a la extracción de características del potencial de acción compuesto. Esta decisión se fundamenta en la observación de que la respuesta en estos análisis no es claramente identificable, y aunque existen diversas posibles explicaciones, la razón principal parece ser la baja respuesta del nervio auditivo en comparación con el ruido de la señal. Al enfocarse exclusivamente en el microfónico coclear, los rasgos generados estarán directamente vinculados con la respuesta de las células ciliadas de la cóclea.

3.3.2.1.1. Creación de la tabla

Un problema asociado al análisis evocado radica en la determinación del número óptimo de señales a promediar para obtener la señal que posteriormente será utilizada en la creación de características. Es importante señalar que en la sección previa no se abordó este tema, centrándonos únicamente en observar las señales resultantes de promediar todas las señales por paciente, las cuales, por consecuencia, exhiben la mejor relación pico versus ruido.

En este contexto, la cantidad de variables generadas para construir la tabla final de datos está directamente vinculada a este aspecto. Por ejemplo, consideremos un escenario donde un paciente presenta 1300 señales, y se requiere promediar 50 señales para generar la señal evocada. En este caso, se obtendrían 26 señales evocadas para dicho paciente, permitiendo la extracción de 26 características. Si, por otro lado, se emplean 65 señales para el promedio, la cantidad de señales evocadas se reduciría a 20.

La resolución de este problema no es trivial. Por un lado, aumentar la cantidad de señales promediadas disminuye el ruido en la señal, mejorando así la relación de amplitud de las características en comparación con el ruido. Por otro lado, al incrementar el tamaño del conjunto promediado, se reduce la cantidad de datos disponibles por paciente, lo que resulta

en una base de datos más limitada para el entrenamiento de los modelos.

La propuesta para abordar esta cuestión consiste en transformar estos parámetros en hiperparámetros de los modelos donde se investigará su desempeño con 40, 50 y 60 señales a promediar, permitiendo así la selección de la mejor configuración de conjuntos de señales a través de agregar este hiperparámetro a la búsqueda de los mejores hiperparámetros de los modelos.

3.3.2.1.2. Features

Los atributos derivados del micrófono coclear se han diseñado con el propósito de evaluar la respuesta en frecuencia del tono presentado a los pacientes. Por ende, estos atributos están conceptualizados en relación con el punto máximo de la señal en el micrófono coclear. Se han generado atributos tanto normalizando la señal por la energía como sin llevar a cabo esta normalización. La normalización de la energía implica la creación de una señal de micrófono coclear, seguida del cálculo del área bajo la curva total. El resultado se divide posteriormente por cada componente de la señal, siguiendo la fórmula:

$$señal_i^{nuevo} = \frac{señal_i^{viejo}}{AreaTotal}$$

En cada instancia de separación, ya sea mediante normalización o sin ella, se generan los siguientes atributos:

- Amplitud absoluta: se calcula a partir de medir la amplitud en la frecuencia de la señal generada al introducir un tono (en 500 Hz o en 2000 Hz). En este caso sería la magnitud de la línea roja.

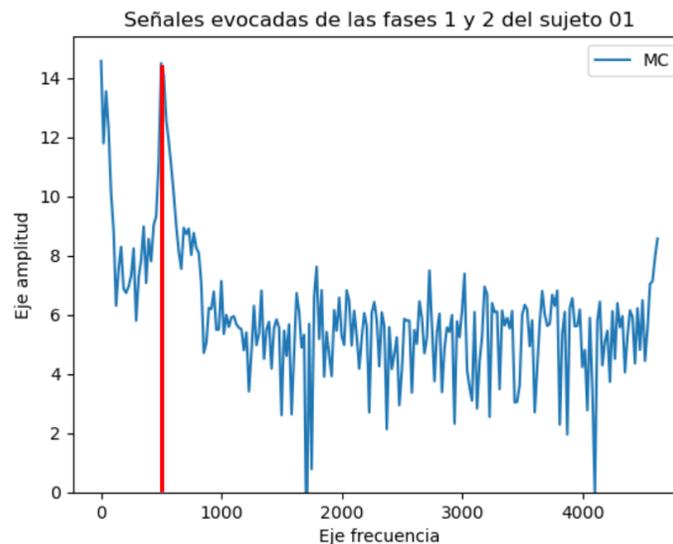


Figura 24: Intuición del *feature* de amplitud absoluta

- Amplitud ponderada: se calcula a partir de medir la amplitud en la frecuencia de la señal generada al introducir un tono (en 500 Hz o en 2000 Hz) dividido por la media del piso de la señal. En este caso sería la magnitud de la línea roja dividida por el valor de la amplitud de la línea verde.

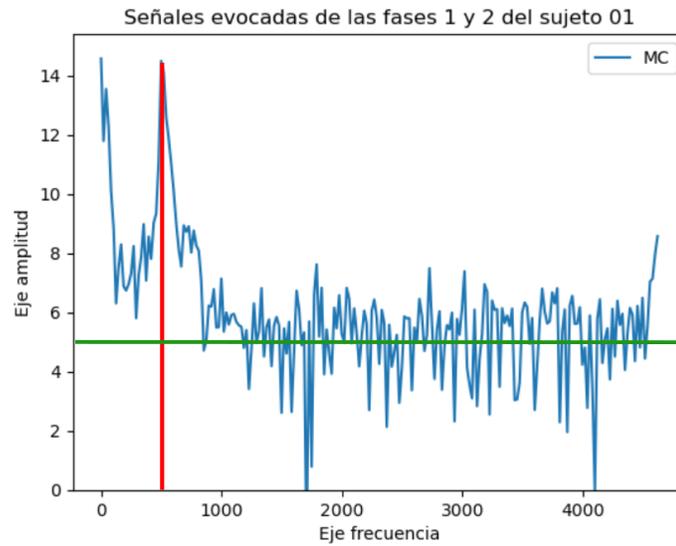


Figura 25: Intuición del *feature* de amplitud ponderada

- Área bajo la curva discreta: se calcula a partir de la suma de las líneas rojas y verdes por separado. Donde el total de la suma de las líneas rojas es la energía en el peak y se divide por el total de las líneas verdes más el total de líneas rojas. Es similar a un cálculo de la razón que hay bajo el peak versus toda la energía de la señal. Cabe destacar que en la imagen de ejemplo no están señaladas todas las líneas verdes que deberían haber, pero la idea es que se trazara en todos los puntos del eje x.

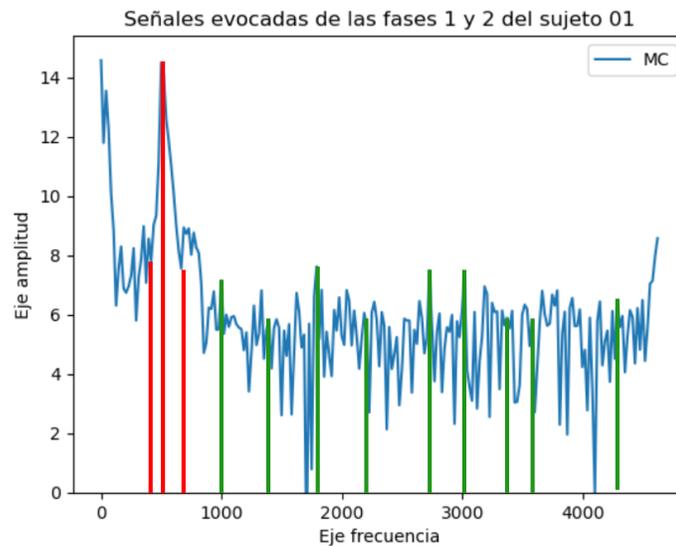


Figura 26: Intuición del *feature* de área bajo la curva discreta

- Área bajo la curva continua: se calcula a partir de el área bajo la curva que se representa en rojo dividido por toda el área bajo la curva, que en este caso sería todo lo que está en rojo y verde.

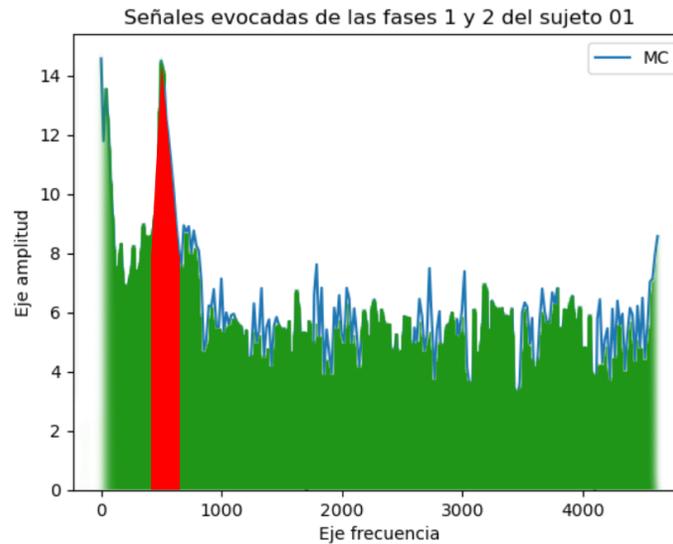


Figura 27: Intuición del *feature* de área bajo la curva continua

También se generan otros 4 atributos, pero sumando las respuestas de 500 Hz y 2000 Hz que dan por separado en cada *feature* mencionado anteriormente en cada uno de los escenarios, ya sea tipo de condición y normalización de energía.

3.3.2.2. Análisis inducido

Del análisis evocado del microfónico coclear, podemos extraer una medida de las células ciliadas, pero no así del nervio auditivo debido a que las señales no eran suficientemente buenas para capturar el potencial de acción compuesto. Con el análisis inducido podríamos lograr esto, debido a que se estima que mediante este análisis se podría obtener la actividad espontánea del nervio auditivo tal como se menciona en la sección de *Conceptos de las electrocoleografías*, y con esto tener una variable que estime el funcionamiento del nervio auditivo. Al realizar el análisis inducido en los datos se pueden observar dos comportamientos en los pacientes, un comportamiento oscilatorio y un claro peak en 3100 Hz. Los cuales se ven de las siguientes maneras:

- Claro peak en 3100:

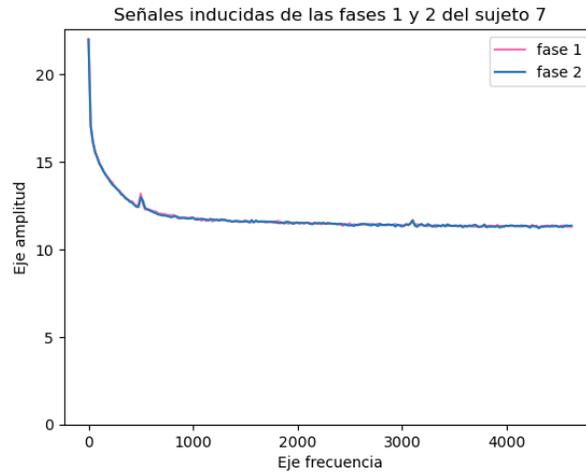


Figura 28: Análisis inducido con claro peak en 3100 Hz

- Comportamiento oscilatorio con peak en 3100 Hz:

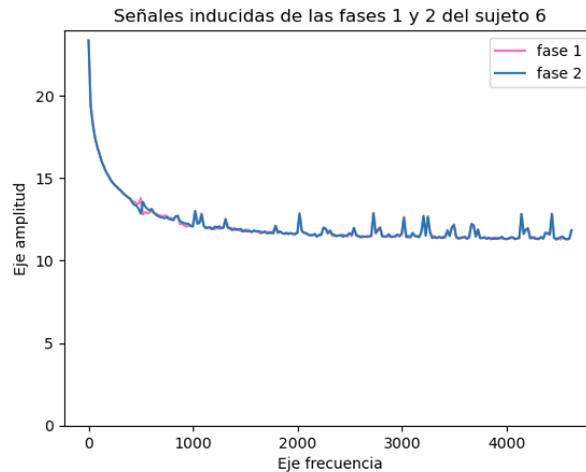


Figura 29: Análisis inducido con comportamiento oscilatorio

Este comportamiento oscilatorio imposibilita el uso directo de este feature debido a que no se puede descartar que ese comportamiento sea instrumental. De todas formas en el estudio del que se habla en *Conceptos de las electrocoleografías* se menciona que el comportamiento oscilatorio se presenta solo con el promedio de 10 señales y cuando se trata de las señales con un claro peak en 3100 Hz, este aparece luego del promedio de 100 señales aproximadamente.

Lo anteriormente mencionado se usa para crear una estrategia que permita eliminar el comportamiento oscilatorio en los pacientes que tienen esa conducta, el cual consiste en que se promedien 100 señales, esto con el objetivo de poder aparecer el peak en 3100 Hz que se observa en los otros paciente y esa señal resultante restarla con un conjunto de 10 señales las cuales tendrían ese comportamiento oscilatorio. Lo anteriormente mencionado queda descrito de la siguiente manera:

$$señal_{\hat{x}=100} - señal_{\hat{j}=10} \quad (3.1)$$

Donde el primer término es el promedio de 100 señales y el segundo término es el promedio de 10 señales que están contenidas en el conjunto de 100 señales del primer término. El resultado de esto no fue el esperado, se repitió el experimento para ambas fases en cada una de las frecuencias y no se pudo aislar el efecto de 3100 Hz, de hecho las señales quedaban de la siguiente manera:

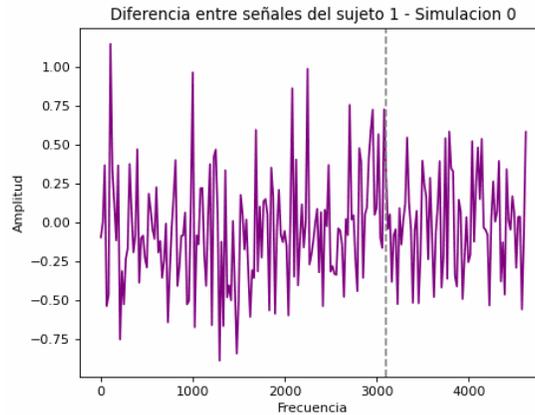


Figura 30: Ilustración de la resta de señales oscilatorias en el paciente 1

Por lo que este feature se descarta y sólo se usarán los creados en el análisis evocado.

3.4. Modelamiento

En el modelamiento el objetivo será diseñar modelos de predicción del entendimiento de habla en silencio y en ruido, y además, crear un modelo generativo que genere datos de señales con sus etiquetas. Lo anteriormente mencionado sugiere dos enfoques:

Modelos discriminativos

- Predicción del habla en silencio
- Predicción del habla en ruido

Modelos generativos

- Modelo generativo para generar señales y sus etiquetas

a continuación se detalla el proceso de modelamiento de cada enfoque.

3.4.1. Modelos discriminativos

Este enfoque considera dos formas de ver el problema de predicción del entendimiento del habla. Los enfoques son los de regresión y de clasificación. El primer enfoque tiene por objetivo crear un modelo de regresión que permita predecir en una escala del 0% al 100% el entendimiento del habla en ruido y en silencio y por otro lado, el problema de clasificación en el cual se separaron los resultados en tres clases según el nivel del entendimiento del habla.

3.4.1.1. Modelos de regresión

Estructura de datos

Se utilizan los datos de 18 pacientes que cuentan con etiquetas del entendimiento del habla en ruido y silencio, además de las electrococleografías.

Creación de la tabla de entrenamiento

Para cada escenario en silencio y en ruido, se crean conjuntos considerando los siguientes features: frecuencia, normalización de la energía y tipo de feature, dando como resultado un total de 48 dataframes que cuentan con una variable independiente y una variable dependiente, que en este caso sería la predicción del habla en ruido o en silencio, según corresponda:

Tabla 3.1: Cantidad de dataframes para probar modelos

Condición	Frecuencia	Energía	Tipo de feature	Cantidad	
silencio	500 Hz	No normalizado por energía	Amplitud absoluta	1	
			Amplitud ponderada	1	
			Área bajo la curva discreta	1	
		No normalizado por energía	Área bajo la curva continua	1	
			Amplitud absoluta	1	
			Amplitud ponderada	1	
	2000 Hz	No normalizado por energía	Área bajo la curva discreta	1	
			Área bajo la curva continua	1	
			Amplitud absoluta	1	
		No normalizado por energía	Amplitud ponderada	1	
			Área bajo la curva discreta	1	
			Área bajo la curva continua	1	
	Ambos	No normalizado por energía	Amplitud absoluta	1	
			Amplitud ponderada	1	
			Área bajo la curva discreta	1	
		No normalizado por energía	Área bajo la curva continua	1	
			Amplitud absoluta	1	
			Amplitud ponderada	1	
	Ruido	500 Hz	No normalizado por energía	Área bajo la curva discreta	1
				Área bajo la curva continua	1
				Amplitud absoluta	1
			No normalizado por energía	Amplitud ponderada	1
				Área bajo la curva discreta	1
				Área bajo la curva continua	1
2000 Hz		No normalizado por energía	Amplitud absoluta	1	
			Amplitud ponderada	1	
			Área bajo la curva discreta	1	
		No normalizado por energía	Área bajo la curva continua	1	
			Amplitud absoluta	1	
			Amplitud ponderada	1	
Ambos		No normalizado por energía	Área bajo la curva discreta	1	
			Área bajo la curva continua	1	
			Amplitud absoluta	1	
		No normalizado por energía	Amplitud ponderada	1	
			Área bajo la curva discreta	1	
			Área bajo la curva continua	1	
Total de conjuntos creados				48	

Selección de variables

Todos los modelos que se prueban usan sólo una variable, la cual está descrita en la tabla mencionada anteriormente, la cual tiene los features relacionado a tipo de condición, frecuen-

cia y energía.

División del conjunto de datos

La forma de dividir los datos consistió en elegir aleatoriamente el 80 % de los datos para entrenar y el 20 % restante para testear y este proceso repetirlo 25 veces. Por lo tanto, se tienen 25 conjuntos distintos tanto de entrenamiento como de testeo.

Modelos de regresión

Los modelos que se ocupan para cada conjunto de datos son los de Random Forest, Red Neuronal Feed Forward, XGBoost y regresión lineal univariada, esta última es para comparar los resultados de los otros modelos con el modelo que se usa actualmente en este problema.

3.4.1.2. Modelos de clasificación

Estructura de datos

Se utilizan los datos de 18 pacientes que cuentan con etiquetas del entendimiento del habla en ruido y silencio, además de las electrococleografías.

Creación de clases

La forma en que se crearon las clases de las etiquetas fue a partir del siguiente criterio:

Silencio			Ruido		
Etiqueta en silencio	Etiqueta relajada	Recuento	Etiqueta en ruido	Etiqueta relajada	Recuento
<65	60	9	<5	1	6
65 < x < 75	70	3	5 < x < 14	10	6
75 <	80	6	14 <	15	6

Figura 31: Criterio de asignación de clase

Creación de la tabla de entrenamiento

Las tablas de entrenamiento son las mismas descritas en la Tabla 4.1

Selección de variables

Todos los modelos que se prueban usan sólo una variable, la cual está descrita en la tabla mencionada anteriormente, la cual tiene los features relacionado a tipo de condición, frecuencia y energía.

División del conjunto de datos

La forma de dividir los datos consistió en elegir aleatoriamente el 80 % de los datos para entrenar y el 20 % restante para testear y este proceso repetirlo 25 veces. Por lo tanto, se tienen 25 conjuntos distintos tanto de entrenamiento como de testeo.

Modelos de clasificación

Los modelos que se ocupan para cada conjunto de datos son los de Random Forest, Red Neuronal Feed Forward y XGBoost.

3.4.2. Modelo generativo

Selección de variables

Las variables seleccionadas para este enfoque son las representaciones de microfónico coclear de un promedio de 50 señales pasadas por una transformada de fourier y luego por el teorema de parseval sin la promediación de estas. O sea, si un paciente tiene 1300 señales de condensación y rarefacción, tendrá 1300 representaciones del micrófono coclear representado por la energía que hay en cada frecuencia. Además a esta señal de microfónico coclear se le añade la variable de la etiqueta del paciente al final.

División del conjunto de datos

Se usan todos los datos para entrenar

Modelo utilizado

Se utilizó un modelo de autoencoder para entrenar, al cual se le pasaron todos los microfónicos cocleares. Luego para regenerar señales se modificaba el espacio latente de la siguiente manera:

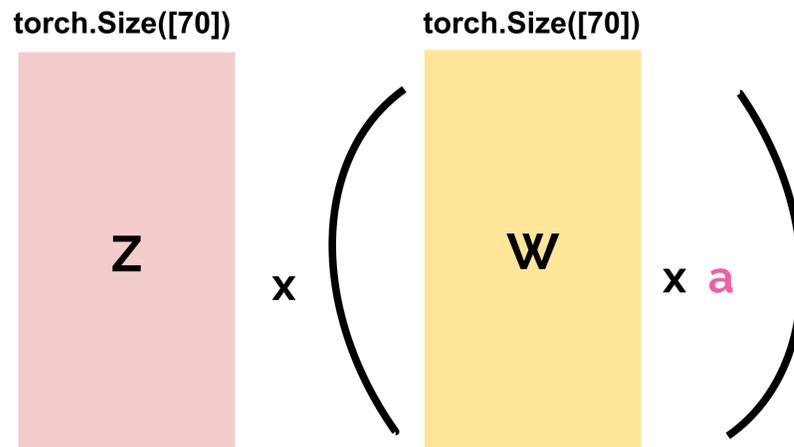


Figura 32: Perturbación del espacio latente

Donde Z es el espacio latente generado por el entrenamiento, W es un vector del mismo tamaño del espacio latente con valores aleatorios de una distribución normal con media 0 y varianza 1 y la variable a variaba en 99 valores, donde 33 tenían valores aleatorios entre $[0.5, 1[$, 33 tenían valores aleatorios entre $[1, 1.5[$ y 33 tenían valores aleatorios entre $[1.5, 2]$. Esto con el objetivo de generar 99 señales nuevas.

3.5. Evaluación

La evaluación de los modelos se lleva a cabo a través de 3 etapas, donde la primera etapa consiste en la evaluación de los modelos de regresión, la segunda la evaluación de los modelos de clasificación y por último, la evaluación del modelo generativo.

3.5.1. Evaluación de los modelos de regresión

Se lleva a cabo una búsqueda exhaustiva de los hiperparámetros para cada uno de los algoritmos mediante una grilla de búsqueda. Esto incluye la variación de los valores asociados al tamaño del conjunto utilizado para la promediación de señales, previamente especificado en la sección de Creación de tabla. El objetivo es identificar los hiperparámetros óptimos que conduzcan a las métricas más destacadas. Este proceso se realiza empleando el método de validación cruzada, implementado en este contexto mediante la creación de 25 instancias independientes de entrenamiento y prueba que fueron mencionadas anteriormente, seleccionadas de manera aleatoria. Cada instancia de validación cruzada se compone de un 80 % de los datos destinados al entrenamiento y un 20 % para la fase de prueba. La evaluación de los hiperparámetros se realiza considerando la métrica de error absoluto porcentual medio (MAPE), y se determinan como los mejores aquellos que resultan en un valor de MAPE más bajo.

Como se menciona en el párrafo anterior, la métrica de desempeño elegida para evaluar los modelos es el MAPE, o Error Porcentual Absoluto Medio. Esta elección ha sido cuidadosamente considerada debido a las características inherentes que hacen que el MAPE sea altamente adecuado para nuestro contexto específico.

En primer lugar, la naturaleza porcentual del MAPE proporciona una medida intuitiva y fácilmente comprensible del rendimiento de los modelos. Esta cualidad es esencial para garantizar una comunicación efectiva de los resultados, dirigida tanto a expertos en análisis de datos como a aquellos con menos experiencia en el campo, facilitando así una comprensión general del desempeño del modelo.

En segundo lugar, el MAPE se destaca por su capacidad para manejar conjuntos de datos con diferentes magnitudes y escalas. Dado que la variabilidad en la escala de las variables es común en diversos escenarios, la elección del MAPE permite una evaluación equitativa del rendimiento del modelo independientemente de estas diferencias de escala.

Asimismo, el MAPE penaliza de manera equitativa los errores en las predicciones, ya sean sobreestimaciones o subestimaciones. Esta característica es crucial para asegurar una evaluación imparcial del modelo en todos los rangos de valores posibles, contribuyendo así a una medición equilibrada del desempeño.

Por último, la interpretación clara del MAPE como el porcentaje promedio de error en las predicciones ofrece una visión transparente de la magnitud de los errores relativos. Esta claridad interpretativa facilita la toma de decisiones informadas y la implementación de mejoras iterativas en el modelo.

A continuación, se presenta una tabla con los valores MAPE separado por escenario, de cada uno de los modelos y sus características:

Tabla 3.2: Resultados de MAPE en silencio a 500 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Silencio	500 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	0.102
				Regresión lineal	0.161
				Red neuronal	0.109
				XGBoost	0.106
			Amplitud ponderada	Random Forest	0.13
				Regresión lineal	0.164
				Red neuronal	0.134
				XGBoost	0.135
			Área bajo la curva discreta	Random Forest	0.157
				Regresión lineal	0.173
				Red neuronal	0.155
				XGBoost	0.155
		Área bajo la curva continua	Random Forest	0.154	
			Regresión lineal	0.173	
			Red neuronal	0.158	
			XGBoost	0.154	
		Normalizado por energía	Amplitud absoluta	Random Forest	0.137
				Regresión lineal	0.161
				Red neuronal	0.144
				XGBoost	0.138
			Amplitud ponderada	Random Forest	0.13
				Regresión lineal	0.163
				Red neuronal	0.133
				XGBoost	0.135
Área bajo la curva discreta	Random Forest		0.157		
	Regresión lineal		0.173		
	Red neuronal		0.155		
	XGBoost		0.155		
Área bajo la curva continua	Random Forest	0.154			
	Regresión lineal	0.173			
	Red neuronal	0.158			
	XGBoost	0.154			

Tabla 3.3: Resultados de MAPE en silencio a 2000 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Silencio	2000 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	0.141
				Regresión lineal	0.161
				Red neuronal	0.142
				XGBoost	0.143
			Amplitud ponderada	Random Forest	0.155
				Regresión lineal	0.162
				Red neuronal	0.152
				XGBoost	0.156
			Área bajo la curva discreta	Random Forest	0.159
				Regresión lineal	0.163
				Red neuronal	0.161
				XGBoost	0.16
		Área bajo la curva continua	Random Forest	0.159	
			Regresión lineal	0.164	
			Red neuronal	0.159	
			XGBoost	0.159	
		Normalizado por energía	Amplitud absoluta	Random Forest	0.152
				Regresión lineal	0.161
				Red neuronal	0.158
				XGBoost	0.153
			Amplitud ponderada	Random Forest	0.155
				Regresión lineal	0.162
				Red neuronal	0.152
				XGBoost	0.156
Área bajo la curva discreta	Random Forest		0.159		
	Regresión lineal		0.163		
	Red neuronal		0.161		
	XGBoost		0.16		
Área bajo la curva continua	Random Forest	0.159			
	Regresión lineal	0.164			
	Red neuronal	0.16			
	XGBoost	0.159			

Tabla 3.4: Resultados de MAPE en silencio sumando ambas respuestas

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Silencio	Ambos	No normalizado por energía	Amplitud absoluta	Random Forest	0.132
				Regresión lineal	0.16
				Red neuronal	0.138
				XGBoost	0.135
			Amplitud ponderada	Random Forest	0.149
				Regresión lineal	0.161
				Red neuronal	0.149
				XGBoost	0.15
			Área bajo la curva discreta	Random Forest	0.162
				Regresión lineal	0.168
				Red neuronal	0.155
				XGBoost	0.166
		Área bajo la curva continua	Random Forest	0.161	
			Regresión lineal	0.168	
			Red neuronal	0.155	
			XGBoost	0.162	
		Normalizado por energía	Amplitud absoluta	Random Forest	0.152
				Regresión lineal	0.16
				Red neuronal	0.157
				XGBoost	0.152
			Amplitud ponderada	Random Forest	0.149
				Regresión lineal	0.16
				Red neuronal	0.149
				XGBoost	0.15
Área bajo la curva discreta	Random Forest		0.162		
	Regresión lineal		0.168		
	Red neuronal		0.155		
	XGBoost		0.166		
Área bajo la curva continua	Random Forest	0.161			
	Regresión lineal	0.168			
	Red neuronal	0.155			
	XGBoost	0.162			

Tabla 3.5: Resultados de MAPE en ruido a 500 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Ruido	500 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	1.25
				Regresión lineal	1.86
				Red neuronal	0.801
				XGBoost	1.25
			Amplitud ponderada	Random Forest	1.32
				Regresión lineal	1.79
				Red neuronal	0.801
				XGBoost	1.31
			Área bajo la curva discreta	Random Forest	1.74
				Regresión lineal	1.94
				Red neuronal	0.801
				XGBoost	1.72
		Área bajo la curva continua	Random Forest	1.70	
			Regresión lineal	1.95	
			Red neuronal	0.801	
			XGBoost	1.7	
		Normalizado por energía	Amplitud absoluta	Random Forest	1.37
				Regresión lineal	1.86
				Red neuronal	0.801
				XGBoost	1.39
			Amplitud ponderada	Random Forest	1.32
				Regresión lineal	1.79
				Red neuronal	0.801
				XGBoost	0.131
Área bajo la curva discreta	Random Forest		1.74		
	Regresión lineal		1.94		
	Red neuronal		0.801		
	XGBoost		1.72		
Área bajo la curva continua	Random Forest	1.7			
	Regresión lineal	1.95			
	Red neuronal	0.801			
	XGBoost	1.7			

Tabla 3.6: Resultados de MAPE en ruido a 2000 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Ruido	2000 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	1.06
				Regresión lineal	1.76
				Red neuronal	0.797
				XGBoost	1.08
			Amplitud ponderada	Random Forest	0.92
				Regresión lineal	1.67
				Red neuronal	0.797
				XGBoost	0.963
			Área bajo la curva discreta	Random Forest	1.87
				Regresión lineal	1.87
				Red neuronal	0.797
				XGBoost	1.90
		Área bajo la curva continua	Random Forest	1.67	
			Regresión lineal	1.89	
			Red neuronal	0.797	
			XGBoost	1.73	
		Normalizado por energía	Amplitud absoluta	Random Forest	1.18
				Regresión lineal	1.76
				Red neuronal	0.797
				XGBoost	1.21
			Amplitud ponderada	Random Forest	0.92
				Regresión lineal	1.67
				Red neuronal	0.79
				XGBoost	0.96
Área bajo la curva discreta	Random Forest		1.87		
	Regresión lineal		1.87		
	Red neuronal		0.797		
	XGBoost		1.9		
Área bajo la curva continua	Random Forest	1.67			
	Regresión lineal	1.89			
	Red neuronal	0.797			
	XGBoost	1.73			

Tabla 3.7: Resultados de MAPE en ruido sumandos ambas respuestas

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Ruido	Ambos	No normalizado por energía	Amplitud absoluta	Random Forest	1.53
				Regresión lineal	1.73
				Red neuronal	0.80
				XGBoost	1.52
			Amplitud ponderada	Random Forest	1.17
				Regresión lineal	1.59
				Red neuronal	0.8
				XGBoost	1.24
			Área bajo la curva discreta	Random Forest	1.82
				Regresión lineal	1.84
				Red neuronal	0.8
				XGBoost	1.85
		Área bajo la curva continua	Random Forest	1.73	
			Regresión lineal	1.85	
			Red neuronal	0.8	
			XGBoost	1.76	
		Normalizado por energía	Amplitud absoluta	Random Forest	1.22
				Regresión lineal	1.73
				Red neuronal	0.8
				XGBoost	1.23
			Amplitud ponderada	Random Forest	1.17
				Regresión lineal	1.73
				Red neuronal	0.8
				XGBoost	1.24
Área bajo la curva discreta	Random Forest		1.82		
	Regresión lineal		1.84		
	Red neuronal		0.8		
	XGBoost		1.85		
Área bajo la curva continua	Random Forest	1.73			
	Regresión lineal	1.85			
	Red neuronal	0.8			
	XGBoost	1.76			

De los resultados previamente expuestos, se destaca que en ambos escenarios, tanto en presencia de ruido como en silencio, todos los modelos superan el rendimiento de la regresión lineal. Es importante señalar que en el escenario de silencio, el modelo de amplitud absoluta a 500 Hz sin normalización de energía exhibe un MAPE del 10 %, en contraste con el mejor modelo de regresión, que alcanza un 16 %.

Por otro lado, en el escenario con presencia de ruido, las configuraciones más destacadas corresponden a todos los modelos de red neuronal a 2000 Hz, los cuales arrojan un MAPE del 79 %, en contraste con el mejor desempeño obtenido por la regresión lineal, que presenta un MAPE del 167 %.

3.5.2. Evaluación de los modelos de clasificación

Al igual que en la evaluación de los modelos de regresión, se hace una grilla de búsqueda con una validación cruzada con la misma configuración para buscar el desempeño del mejor modelo. En este caso se busca optimizar la métrica de recall macro average.

El recall macro average en modelos de clasificación es valioso por varias razones. Calcula el recall por clase y luego promedia estos valores, lo que proporciona un equilibrio entre clases y una visión general del rendimiento del modelo. Es sensible a clases minoritarias, lo que es crucial en problemas con clases desbalanceadas. Además, su interpretación clara y su capacidad para comunicar resultados lo hacen útil en diferentes contextos. También es más robusto ante clases mal etiquetadas, lo que garantiza una evaluación más precisa del modelo.

Dicho lo anterior, se presentan los resultados que obtuvieron los modelos:

Tabla 3.8: Resultados de Recall Macro en silencio a 500 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	Recall Macro
Silencio	500 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	0.62
				Red neuronal	0.58
				XGBoost	0.63
			Amplitud ponderada	Random Forest	0.52
				Red neuronal	0.54
				XGBoost	0.53
			Área bajo la curva discreta	Random Forest	0.44
				Red neuronal	0.42
				XGBoost	0.43
		Área bajo la curva continua	Random Forest	0.47	
			Red neuronal	0.41	
			XGBoost	0.47	
		Normalizado por energía	Amplitud absoluta	Random Forest	0.51
				Red neuronal	0.44
				XGBoost	0.52
			Amplitud ponderada	Random Forest	0.52
				Red neuronal	0.54
				XGBoost	0.53
Área bajo la curva discreta	Random Forest		0.44		
	Red neuronal		0.42		
	XGBoost		0.43		
Área bajo la curva continua	Random Forest	0.47			
	Red neuronal	0.42			
	XGBoost	0.47			

Tabla 3.9: Resultados de Recall Macro en silencio a 2000 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	Recall Macro
Silencio	2000 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	0.58
				Red neuronal	0.52
				XGBoost	0.58
			Amplitud ponderada	Random Forest	0.49
				Red neuronal	0.49
				XGBoost	0.50
			Área bajo la curva discreta	Random Forest	0.40
				Red neuronal	0.41
				XGBoost	0.41
			Área bajo la curva continua	Random Forest	0.41
				Red neuronal	0.40
				XGBoost	0.42
		Normalizado por energía	Amplitud absoluta	Random Forest	0.53
				Red neuronal	0.57
				XGBoost	0.57
			Amplitud ponderada	Random Forest	0.49
				Red neuronal	0.50
				XGBoost	0.50
			Área bajo la curva discreta	Random Forest	0.40
				Red neuronal	0.41
				XGBoost	0.41
			Área bajo la curva continua	Random Forest	0.41
				Red neuronal	0.40
				XGBoost	0.42

Tabla 3.10: Resultados de Recall Macro en silencio sumandos ambas respuestas

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Silencio	Ambos	No normalizado por energía	Amplitud absoluta	Random Forest	0.71
				Red neuronal	0.53
				XGBoost	0.7
			Amplitud ponderada	Random Forest	0.51
				Red neuronal	0.51
				XGBoost	0.52
			Área bajo la curva discreta	Random Forest	0.42
				Red neuronal	0.43
				XGBoost	0.43
		Área bajo la curva continua	Random Forest	0.43	
			Red neuronal	0.42	
			XGBoost	0.43	
		Normalizado por energía	Amplitud absoluta	Random Forest	0.53
				Red neuronal	0.49
				XGBoost	0.53
			Amplitud ponderada	Random Forest	0.51
				Red neuronal	0.51
				XGBoost	0.52
Área bajo la curva discreta	Random Forest		0.42		
	Red neuronal		0.43		
	XGBoost		0.43		
Área bajo la curva continua	Random Forest	0.43			
	Red neuronal	0.42			
			XGBoost	0.43	

Tabla 3.11: Resultados de Recall Macro en ruido a 500 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	Recall Macro
Ruido	500 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	0.62
				Red neuronal	0.60
				XGBoost	0.62
			Amplitud ponderada	Random Forest	0.56
				Red neuronal	0.57
				XGBoost	0.56
			Área bajo la curva discreta	Random Forest	0.46
				Red neuronal	0.44
				XGBoost	0.46
			Área bajo la curva continua	Random Forest	0.49
				Red neuronal	0.44
				XGBoost	0.48
		Normalizado por energía	Amplitud absoluta	Random Forest	0.6
				Red neuronal	0.44
				XGBoost	0.6
			Amplitud ponderada	Random Forest	0.56
				Red neuronal	0.57
				XGBoost	0.56
			Área bajo la curva discreta	Random Forest	0.44
				Red neuronal	0.42
				XGBoost	0.43
			Área bajo la curva continua	Random Forest	0.49
				Red neuronal	0.44
				XGBoost	0.48

Tabla 3.12: Resultados de Recall Macro en ruido a 2000 Hz

Condición	Frecuencia	Energía	Tipo de feature	Modelos	Recall Macro
Ruido	2000 Hz	No normalizado por energía	Amplitud absoluta	Random Forest	0.57
				Red neuronal	0.57
				XGBoost	0.57
			Amplitud ponderada	Random Forest	0.52
				Red neuronal	0.52
				XGBoost	0.53
			Área bajo la curva discreta	Random Forest	0.42
				Red neuronal	0.44
				XGBoost	0.43
			Área bajo la curva continua	Random Forest	0.42
				Red neuronal	0.43
				XGBoost	0.45
		Normalizado por energía	Amplitud absoluta	Random Forest	0.53
				Red neuronal	0.45
				XGBoost	0.53
			Amplitud ponderada	Random Forest	0.52
				Red neuronal	0.52
				XGBoost	0.53
			Área bajo la curva discreta	Random Forest	0.42
				Red neuronal	0.44
				XGBoost	0.43
			Área bajo la curva continua	Random Forest	0.43
				Red neuronal	0.43
				XGBoost	0.45

Tabla 3.13: Resultados de Recall Macro en ruido sumandos ambas respuestas

Condición	Frecuencia	Energía	Tipo de feature	Modelos	MAPE
Ruido	Ambos	No normalizado por energía	Amplitud absoluta	Random Forest	0.69
				Red neuronal	0.62
				XGBoost	0.69
			Amplitud ponderada	Random Forest	0.59
				Red neuronal	0.58
				XGBoost	0.58
			Área bajo la curva discreta	Random Forest	0.45
				Red neuronal	0.46
				XGBoost	0.44
		Área bajo la curva continua	Random Forest	0.45	
			Red neuronal	0.45	
			XGBoost	0.44	
		Normalizado por energía	Amplitud absoluta	Random Forest	0.55
				Red neuronal	0.53
				XGBoost	0.55
			Amplitud ponderada	Random Forest	0.59
				Red neuronal	0.57
				XGBoost	0.58
Área bajo la curva discreta	Random Forest		0.45		
	Red neuronal		0.46		
	XGBoost		0.44		
Área bajo la curva continua	Random Forest	0.45			
	Red neuronal	0.45			
	XGBoost	0.44			

De acuerdo a los resultados presentados anteriormente, se puede apreciar que en el entorno de silencio, el clasificador más destacado corresponde a un modelo Random Forest que emplea la característica de amplitud absoluta a con ambas frecuencias sin normalización de energía, obteniendo un recall macro del 71%. Por otro lado, en el escenario con ruido, el mismo modelo sobresale como el mejor, logrando un recall macro del 69%.

3.5.3. Evaluación de modelo generativo

La evaluación del modelo generativo se llevó a cabo mediante la inspección y comparación de sus resultados en contraste con los datos reales. Dado que los atributos de amplitud absoluta y amplitud ponderada demostraron poseer el mayor poder predictivo, se centró la comparación en estos aspectos. A modo de ejemplo se presentarán los resultados a 500 Hz de ambas representaciones:

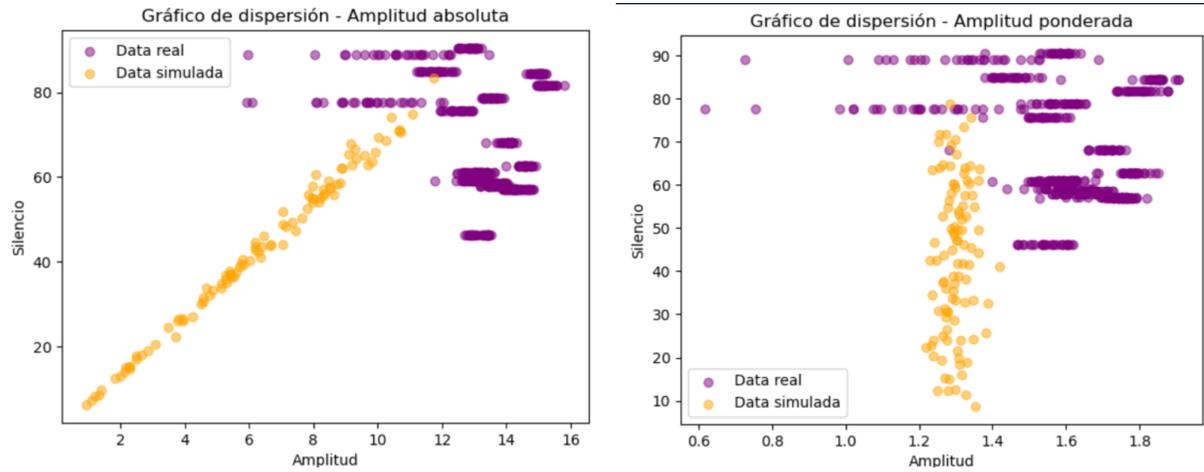


Figura 33: Representaciones del autoencoder

Al analizar detenidamente los gráficos, se evidencia que el modelo generativo no logró replicar de manera precisa los datos reales. En el caso de la amplitud absoluta, se observa un comportamiento lineal que no se corresponde con la naturaleza de los datos reales, los cuales presentan un patrón de agrupamiento por niveles de silencio. En cuanto a la amplitud ponderada, se aprecia una persistencia del nivel de amplitud, independientemente del nivel de silencio. Este comportamiento se repetía en los datos a 2000 Hz

Debido a la falta de fidelidad en la representación de los datos generados, se desestima la utilización de esta información para el entrenamiento del modelo, ya que no refleja adecuadamente el comportamiento real de los datos.

Capítulo 4

Conclusiones

En este trabajo de tesis se presenta una herramienta para definir el nivel de comprensión del habla que podría tener un paciente con implantes cocleares a partir de las señales de las electrococleografías que se obtienen del mismo aparato. En particular, mediante el uso de modelos de aprendizaje de máquinas se generaron predicciones tanto para ambientes de silencio como en ruido de tal forma de poder medir mediante un modelo regresor este nivel o bajo una categoría que engloba umbrales de audición en cada segmento. Al desarrollar y evaluar diferentes modelos se seleccionaron diferentes modelos que satisfacen un mayor poder predictivo tanto para el problema de regresión como el de clasificación.

En cuanto al problema de regresión en el escenario de silencio el modelo seleccionado es un random forest con un feature de amplitud absoluta a 500 Hz sin normalización de la energía el cual da un MAPE de un 10 % y en el escenario de en silencio el modelo seleccionado fue una red neuronal con cualquiera de los features entregados a 2000 Hz sin normalización de la energía el cual da un MAPE de un 79 %. Los resultados de estos modelos confirman la hipótesis principal de esta tesis, donde los modelos anteriormente mencionados superan los resultados de las regresiones lineales y en particular, podemos ver que en el escenario de ruido, los otros modelos se ajustan mucho mejor que las regresiones lineales logrando la mitad de error que presentan las regresiones lineales. De esta forma, se cumple con el objetivo general, y con los objetivos específicos 1, 2 y 3 que fueron presentados al inicio de este texto.

Por otro lado, desde la perspectiva de clasificación del nivel del umbral del habla del paciente, se obtuvieron resultados similares en cuanto a la métrica de recall macro con un valor de un 71 %-69 % en los escenarios de silencio y en ruido. Si bien los resultados no son malos aún hay mucho espacio de mejora en este enfoque. Para seguir desarrollándolo hay que primero crear algún criterio en el que se puedan comparar estos resultados son los de regresión y eventualmente evaluar según la aplicación, cuál de las dos herramientas es la más correcta. Otro espacio de mejora en este enfoque es evaluar los umbrales de cada categoría, de tal forma de averiguar otras configuraciones que logran capturar mejor el efecto que hay por categoría, ya sea aumentando o disminuyendo las clases o definir de otras maneras los límites de audición por cada una de las categorías.

Es importante resaltar que tanto los resultados del enfoque regresor como los de clasificación pueden mejorar significativamente a medida que se aumente el tamaño de la muestra de pacientes. En el momento de la recolección de datos, solo se contaba con datos de seña-

les de 20 pacientes, de los cuales solo se utilizaron 18. Este hecho claramente introduce un sesgo en los resultados. Aunque las hipótesis se cumplan, todavía existe un amplio margen para mejorar. Una de las áreas de mejora es precisamente esta, y otras están relacionadas con los tipos de datos obtenidos de los pacientes y la cantidad de señales recolectadas de ellos.

Inicialmente, se tuvo acceso únicamente a las señales generadas a 500 Hz y 2000 Hz de los pacientes. Sin embargo, los estudios más recientes, que han demostrado los mejores resultados en la predicción del entendimiento del habla, sugieren la inclusión de una medida de Electrocolegografía de Respuesta Total. Esta medida considera muestras de señales a frecuencias adicionales, específicamente a 250, 500, 750, 1000, 2000 y 4000 Hz. Por lo tanto, para futuros análisis con el objetivo de mejorar estas métricas, es necesario recopilar datos de estas frecuencias adicionales.

Por otro lado, también es crucial obtener más muestras por paciente. La toma de ventanas de señales promediadas conlleva a una disminución en la cantidad de datos disponibles por paciente. Por lo tanto, se recomienda adquirir una mayor cantidad de muestras por paciente para enriquecer el conjunto de datos y mejorar así la robustez y precisión de los análisis.

En relación con la perspectiva de generar nuevos datos para abordar el problema de la escasez de datos, se evaluó un modelo de autoencoder. Sin embargo, según los resultados obtenidos, se estima que este modelo no logró capturar adecuadamente la naturaleza de las señales. Aunque las señales generadas parecían similares a las reales superficialmente, al comparar los features generados con las etiquetas también producidas por el autoencoder, se pudo observar que los datos no seguían la distribución real observada en los datos auténticos.

A pesar de que esta implementación condujo a resultados incorrectos, este enfoque merece ser investigado más a fondo. Por ejemplo, podría explorarse la generación de señales basada en el nivel de comprensión del habla. Esto se debe a que en los datos reales se observa que los features varían dentro de rangos definidos según el nivel de comprensión del habla. Con este enfoque se tendrían diferentes autoencoders por nivel de entendimiento del habla.

Bibliografía

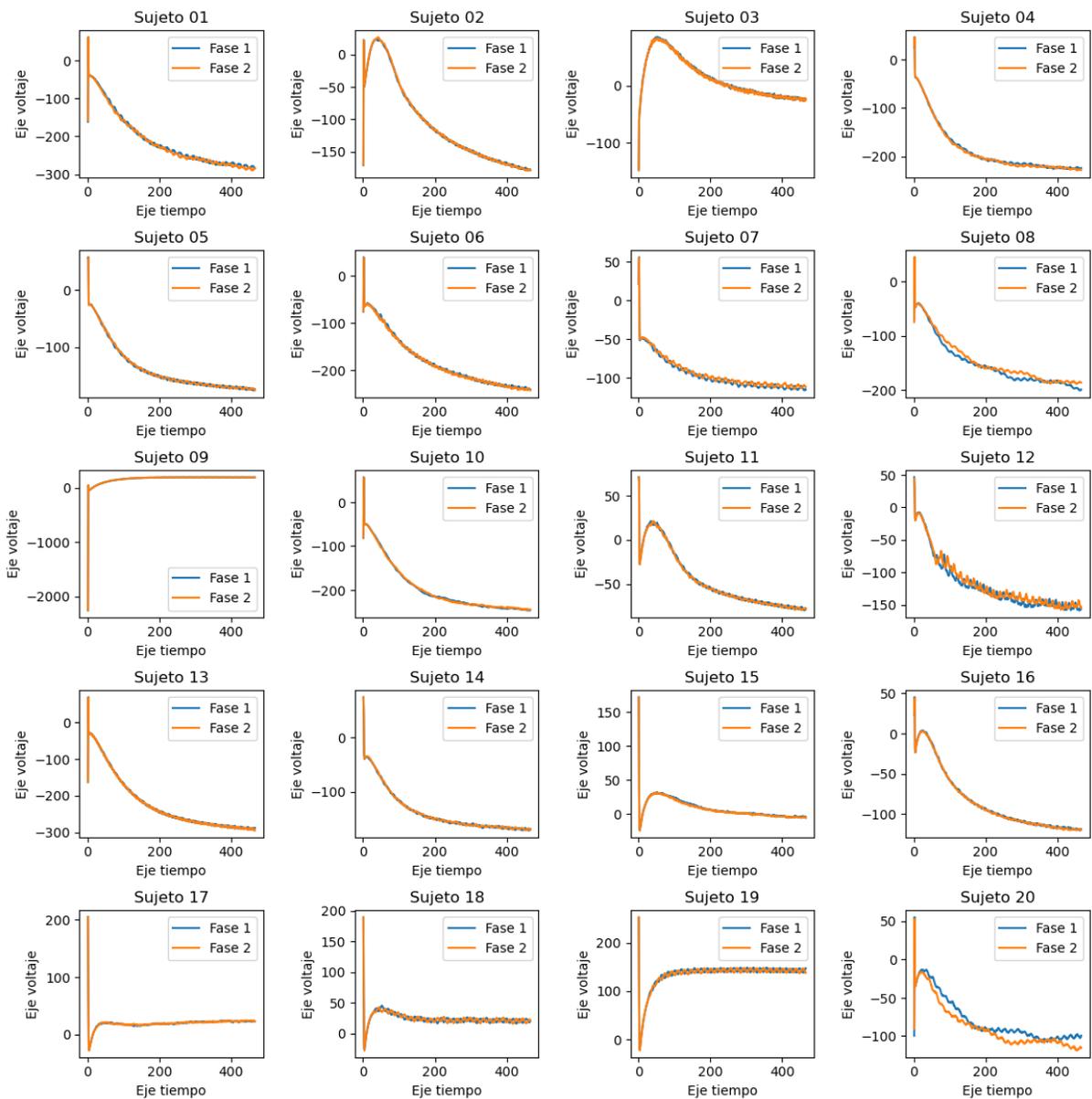
- [1] “Sordera y pérdida de la audición.”, <https://www.who.int/es/news-room/fact-sheets/detail/deafness-and-hearing-loss>.
- [2] “Nederlandstalige versie - aanbeveling 02/1 bis,” 2016, <https://www.biap.org/es/archivo/65-ct-2-classification-des-surdites>.
- [3] Katz, J., English, K., English, K., Chasin, M., Chasin, M., Tillery, K., y Hood, L., “Handbook of clinical audiology,” vol. Seventh Edition, pp. 891–921, 2015.
- [4] Gantz, B. J., Woodworth, G. G., Knutson, J. F., Abbas, P. J., y Tyler, R. S., “Multivariate predictors of audiological success with multichannel cochlear implants,” *Annals of Otology, Rhinology amp;amp; Laryngology*, vol. 102, no. 12, p. 909–916, 1993, [doi:10.1177/000348949310201201](https://doi.org/10.1177/000348949310201201).
- [5] Blarney, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., Larky, J., Lindström, B., Nedzelski, J., Peterson, A., y et al., “Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants,” *Audiology and Neurotology*, vol. 1, no. 5, p. 293–306, 1996, [doi:10.1159/000259212](https://doi.org/10.1159/000259212).
- [6] Lazard, D. S., Vincent, C., Venail, F., Van de Heyning, P., Truy, E., Sterkers, O., Skarzynski, P. H., Skarzynski, H., Schauwers, K., O’Leary, S., y et al., “Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time,” *PLoS ONE*, vol. 7, no. 11, 2012, [doi:10.1371/journal.pone.0048739](https://doi.org/10.1371/journal.pone.0048739).
- [7] Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., y et al., “Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients,” *Audiology and Neurotology*, vol. 18, no. 1, p. 36–47, 2012, [doi:10.1159/000343189](https://doi.org/10.1159/000343189).
- [8] Kral, A., Dorman, M. F., y Wilson, B. S., “Neuronal development of hearing and language: Cochlear implants and critical periods,” *Annual Review of Neuroscience*, vol. 42, no. 1, p. 47–65, 2019, [doi:10.1146/annurev-neuro-080317-061513](https://doi.org/10.1146/annurev-neuro-080317-061513).
- [9] Walia, A., Shew, M. A., Kallogjeri, D., Wick, C. C., Durakovic, N., Lefler, S. M., Ortman, A. J., Herzog, J. A., y Buchman, C. A., “Electrocochleography and cognition are important predictors of speech perception outcomes in noise for cochlear implant recipients,” *Scientific Reports*, vol. 12, 2022, [doi:10.1038/s41598-022-07175-7](https://doi.org/10.1038/s41598-022-07175-7).
- [10] Radeloff, A., Shehata-Dieler, W., Scherzed, A., Rak, K., Harnisch, W., Hagen, R., y Mlynski, R., “Intraoperative monitoring using cochlear microphonics in cochlear implant patients with residual hearing,” *Otology amp;amp; Neurotology*, vol. 33, no. 3, p. 348–354, 2012, [doi:10.1097/mao.0b013e318248ea86](https://doi.org/10.1097/mao.0b013e318248ea86).

- [11] Wijewickrema, S., Bester, C., Gerard, J.-M., Collins, A., y O’Leary, S., “Automatic analysis of cochlear response using electrocochleography signals during cochlear implant surgery,” *PLOS ONE*, vol. 17, no. 7, 2022, doi:10.1371/journal.pone.0269187.
- [12] Rahman, M. T., Chari, D. A., Ishiyama, G., Lopez, I., Quesnel, A. M., Ishiyama, A., Nadol, J. B., y Hansen, M. R., “Cochlear implants: Causes, effects and mitigation strategies for the foreign body response and inflammation,” *Hearing Research*, vol. 422, p. 108536, 2022, doi:10.1016/j.heares.2022.108536.
- [13] Katz, J., English, K., English, K., Chasin, M., Chasin, M., Tillery, K., y Hood, L., “Handbook of clinical audiology,” vol. Seventh Edition, pp. 207–228, 2015.
- [14] Koka, K., Saoji, A. A., y Litvak, L. M., “Electrocochleography in cochlear implant recipients with residual hearing,” *Ear and Hearing*, vol. 38, no. 3, 2017, doi:10.1097/aud.000000000000385.
- [15] Dallos, P., *The auditory periphery = biophysics and physiology*. Academic Press, 1973.
- [16] Fitzpatrick, D. C., Campbell, A. T., Choudhury, B., Dillon, M. P., Forgues, M., Buchman, C. A., y Adunka, O. F., “Round window electrocochleography just before cochlear implantation,” *Otology amp; Neurotology*, vol. 35, no. 1, p. 64–71, 2014, doi:10.1097/mao.000000000000219.
- [17] Herrada, J., Medel, V., Dragicevic, C., Maass, J. C., Stott, C. E., y Delano, P. H., “A frequency peak at 3.1 khz obtained from the spectral analysis of the cochlear implant electrocochleography noise,” 2023, doi:10.1101/2023.09.09.556985.
- [18] Green, K., Bhatt, Y., Mawman, D., O’driscoll, M., Saeed, S., Ramsden, R., y Green, M., “Predictors of audiological outcome following cochlear implantation in adults,” *Cochlear Implants International*, vol. 8, no. 1, p. 1–11, 2007, doi:10.1179/cim.2007.8.1.1.
- [19] Chakravorti, S., Noble, J. H., Gifford, R. H., Dawant, B. M., O’Connell, B. P., Wang, J., y Labadie, R. F., “Further evidence of the relationship between cochlear implant electrode positioning and hearing outcomes,” *Otology amp; Neurotology*, vol. 40, no. 5, p. 617–624, 2019, doi:10.1097/mao.0000000000002204.
- [20] Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez Orallo, J., Kull, M., Lachiche, N., Ramirez Quintana, M. J., y Flach, P. A., “Crisp-dm twenty years later: From data mining processes to data science trajectories,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2020, doi:10.1109/tkde.2019.2962680.
- [21] “Anatomía y fisiología del oído.”, <https://www.stanfordchildrens.org/es/topic/default?id=anatomaylafisiologadelodo-90-P05132>.
- [22] “Los implantes de cóclea.”, <https://www.childrensmn.org/educationmaterials/parents/article/10549/los-implantes-de-coclea/>.
- [23] Veronique J.O. Verhaegen, Ad F.M. Snik, A. J. B. A. R. L. E. A. M., “Preservation of low-frequency residual hearing after cochlear implantation. is soft surgery effective?,” *The Journal of International Advanced Otology*, vol. 6, no. 2, pp. 125–130, 2010.
- [24] Gautschi-Mills, K., Khoza-Shangase, K., y Pillay, D., “Preservation of residual hearing after cochlear implant surgery: an exploration of residual hearing function in a group of recipients at cochlear implant units,” *Brazilian Journal of Otorhinolaryngology*, vol. 85, no. 3, p. 310–318, 2019, doi:10.1016/j.bjorl.2018.02.006.

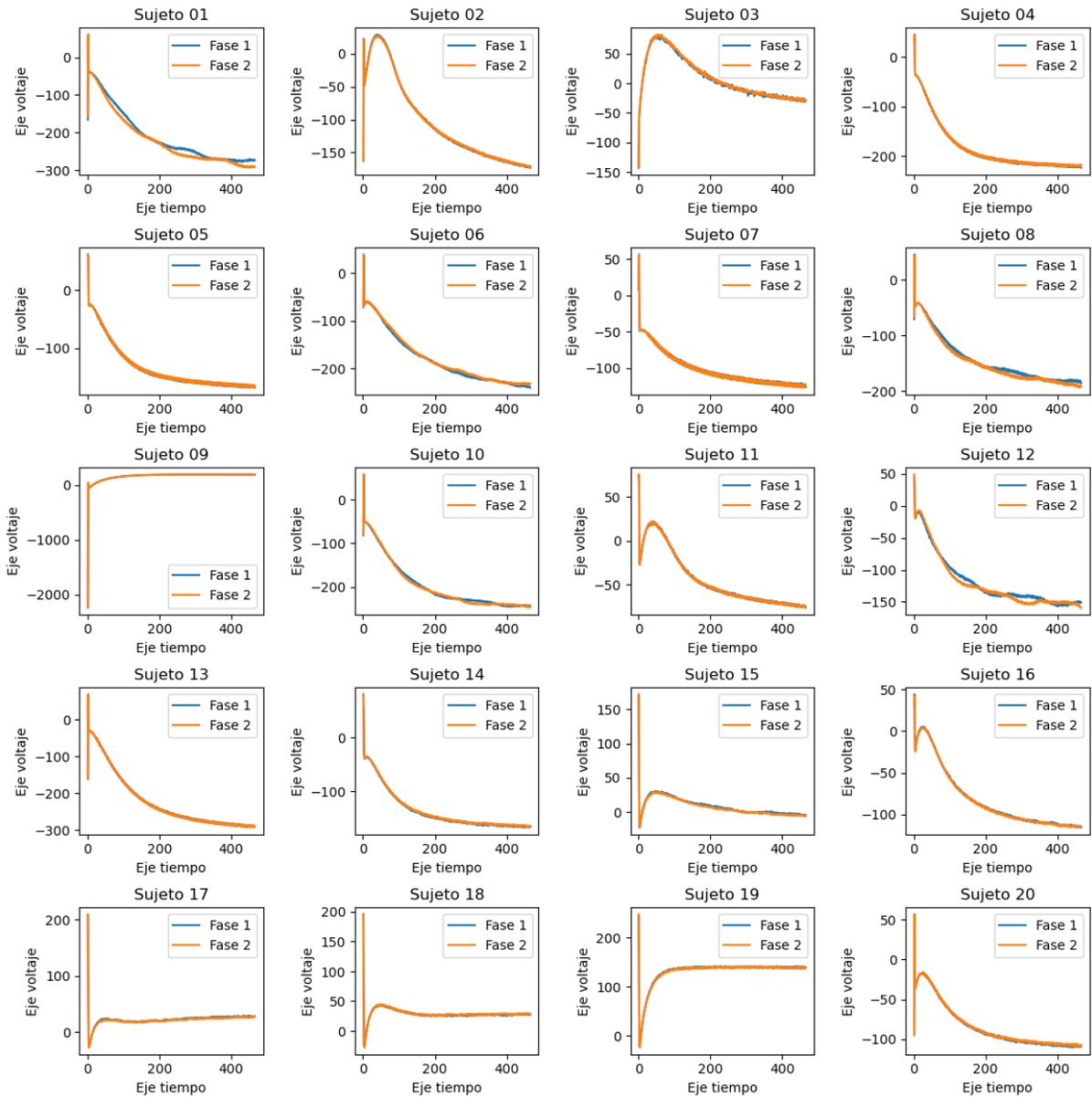
- [25] “Dispositivo de implante coclear unilateral para personas con hipoacusia sensorioneural bilateral severa o profunda postlocutiva,” 2019, https://www.minsal.cl/wp-content/uploads/2019/07/22.-01072019-Protocolo-Hipoacusia_FINAL.pdf.
- [26] Proakis, J. G. y Manolakis, D. G., “Digital signal processing,” vol. Third Edition, pp. 2–5, 1995.
- [27] Proakis, J. G. y Manolakis, D. G., “Digital signal processing,” vol. Third Edition, pp. 21–39, 1995.
- [28] Proakis, J. G. y Manolakis, D. G., “Digital signal processing,” vol. Third Edition, pp. 230–288, 1995.
- [29] Eggermont, J. J., “Electrocochleography,” vol. Fifth Edition, pp. 625–705, 1976.
- [30] Kiang NYS, Watanabe T, T. E. C. T., “Discharge patterns of single fibers in the cat’s auditory nerve,” 1965.
- [31] Louage, D. H. G., van der Heijden, M., y Joris, P. X., “Temporal properties of responses to broadband noise in the auditory nerve,” *Journal of Neurophysiology*, vol. 91, no. 5, p. 2051–2065, 2004, [doi:10.1152/jn.00816.2003](https://doi.org/10.1152/jn.00816.2003).
- [32] Chen, T. y Guestrin, C., “Xgboost: A scalable tree boosting system,” en *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, ACM, 2016, [doi:10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [33] Chen, T. y Guestrin, C., “Xgboost: A scalable tree boosting system,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16*, 2016, [doi:10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).

Anexos

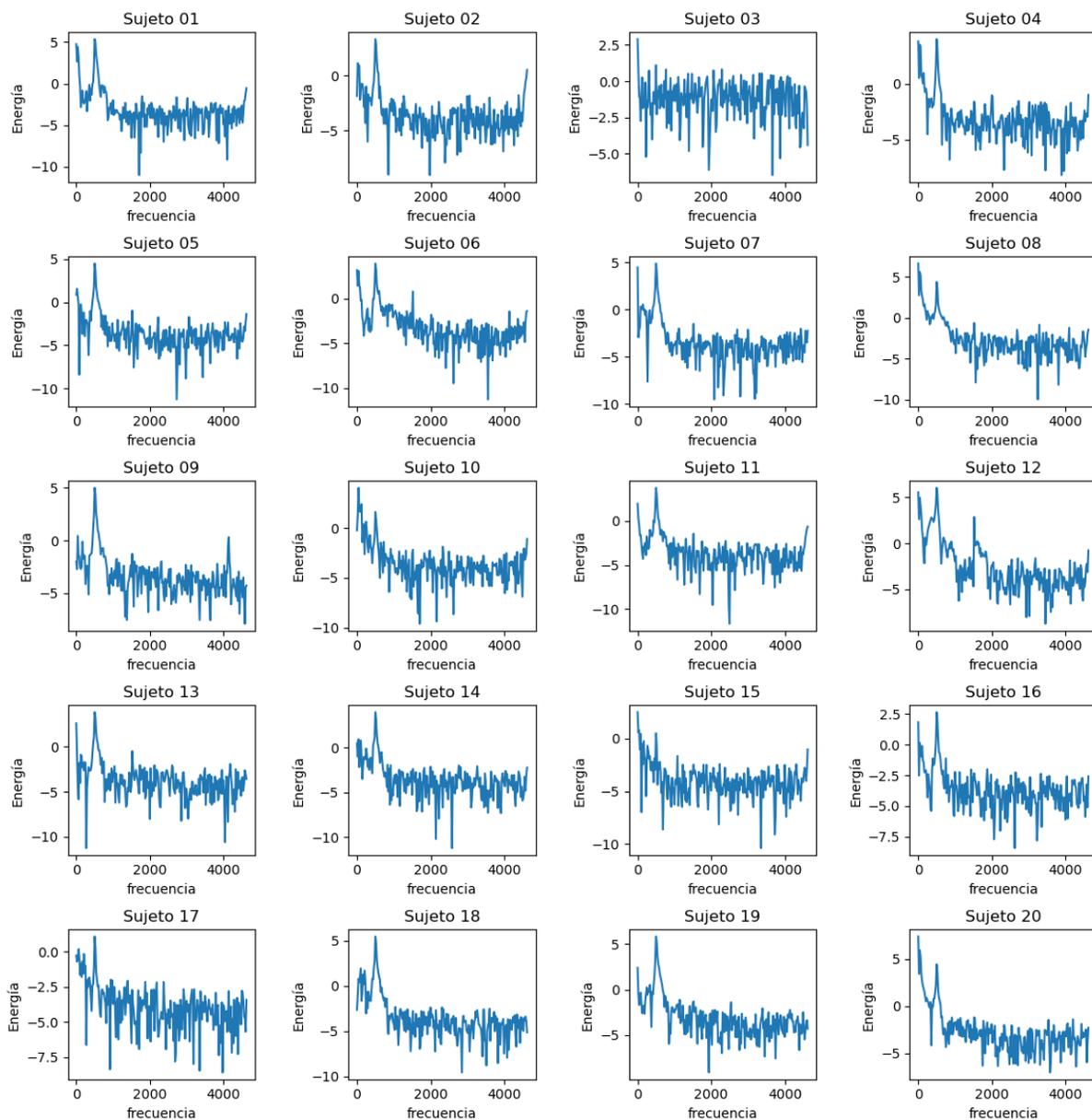
Anexos A. Señal de electrocolceografía evocada a 500 Hz de todos los pacientes



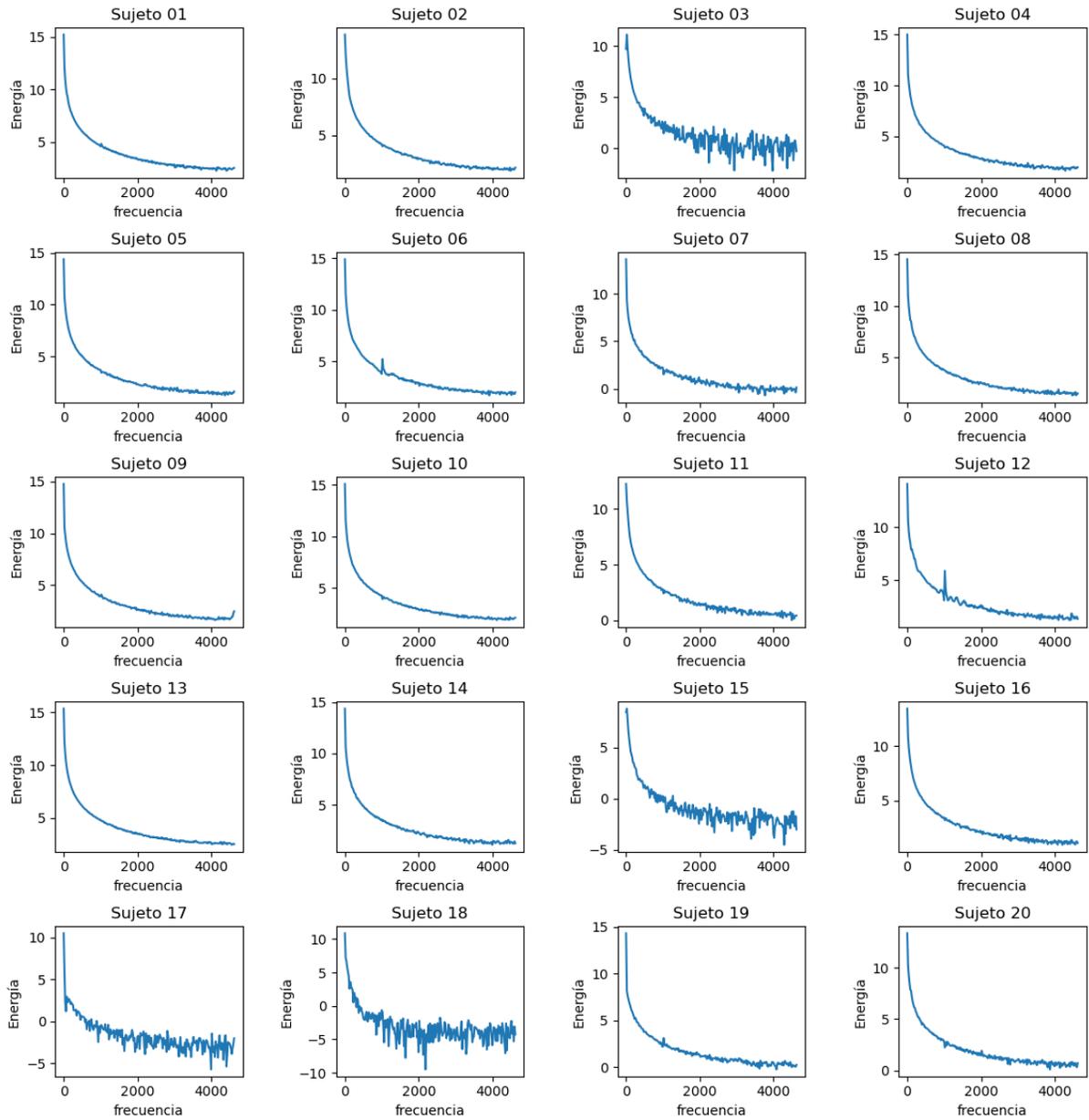
Anexos B. Señal de electrocolceografía evocada a 2000 Hz de todos los pacientes



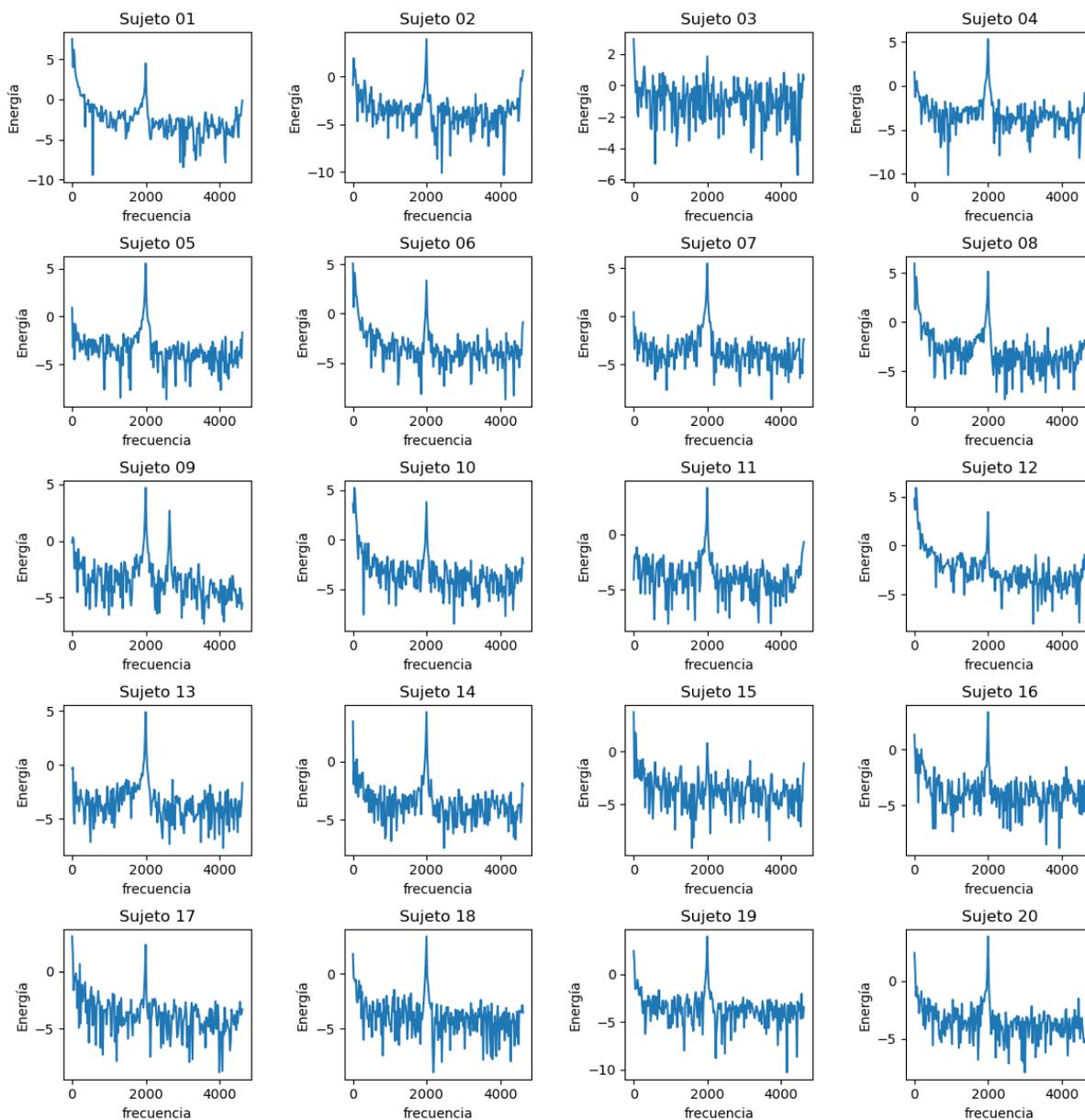
Anexos C. Microfonico coclear a 500 Hz



Anexos D. Potencial de acción compuesto a 500 Hz



Anexos E. Microfónico coclear a 2000 Hz



Anexos F. Potencial de acción compuesto a 2000 Hz

