

Marco Teorico

Logopedia

La logopedia es la disciplina que aborda los trastornos del habla, lenguaje y comunicación. Tradicionalmente, el proceso terapéutico se basa en la observación directa del logopeda y la repetición de ejercicios por parte del paciente. Sin embargo, los avances tecnológicos han permitido la aparición de la logopedia digital, un enfoque que utiliza sistemas computacionales para registrar, analizar y retroalimentar el desempeño vocal del usuario.

De acuerdo con Ladefoged (2001), el estudio de la fonética acústica se centra en las propiedades físicas del sonido, incluyendo parámetros como la frecuencia fundamental (F0), jitter (variación de frecuencia), shimmer (variación de amplitud) y la relación armónico-ruido (HNR). Estos indicadores cuantitativos permiten evaluar de manera objetiva la calidad de la voz y su estabilidad.

Fonética Acústica

La fonética acústica estudia las propiedades físicas del sonido del habla, analizando su frecuencia, intensidad y duración. Según Ladefoged (2001), estos parámetros permiten caracterizar la voz humana y evaluar su estabilidad. En el proyecto, los sonidos emitidos por los usuarios son transformados en señales digitales que se analizan mediante algoritmos especializados. Este análisis proporciona información cuantitativa sobre la calidad vocal y permite identificar patrones de pronunciación anormales, esenciales para la corrección terapéutica.

Frecuencia Fundamental (F0)

La frecuencia fundamental (F0) representa el número de vibraciones por segundo de las cuerdas vocales y determina el tono de la voz. Es uno de los parámetros más relevantes en la evaluación fonética, ya que revela el control respiratorio y muscular del paciente. Alteraciones significativas en la F0 pueden indicar tensiones, problemas de articulación o variaciones emocionales. Su análisis se realiza mediante transformadas de Fourier y algoritmos de detección de picos en Python, lo que permite obtener valores precisos del tono vocal en cada grabación.

Jitter

El jitter mide las variaciones de frecuencia entre ciclos consecutivos de vibración de las cuerdas vocales. Un valor elevado de jitter indica inestabilidad en la frecuencia, lo que puede reflejar un control insuficiente del tono o irregularidades en la fonación (Boersma & Weenink, 2023). En este proyecto, el análisis del jitter es esencial para detectar fluctuaciones anormales que afectan la claridad de la pronunciación. Su registro continuo permite observar mejoras progresivas en la estabilidad vocal conforme el paciente avanza en la terapia.

Shimmer

El shimmer cuantifica las variaciones en la amplitud de la señal vocal entre ciclos sucesivos. Es un indicador de irregularidad en la intensidad o fuerza de la voz. Una voz saludable presenta valores bajos de shimmer, mientras que niveles altos pueden asociarse a fatiga vocal, disfonía o falta de control respiratorio. La plataforma medirá el shimmer en cada sesión para identificar alteraciones y ofrecer retroalimentación sobre la uniformidad de la emisión sonora del paciente.

Relación Armónico-Ruido (HNR)

La relación armónico-ruido (HNR) evalúa la proporción entre los componentes armónicos y el ruido presente en una señal vocal. Un valor bajo de HNR indica la presencia de turbulencias o distorsiones, lo que afecta la claridad del sonido. Este parámetro resulta fundamental en la detección de problemas laríngeos y en la valoración de la calidad general de la voz. El sistema calculará la HNR junto a otras métricas acústicas para determinar un índice global de calidad fonética.

Big Data

El término Big Data hace referencia al manejo y análisis de grandes volúmenes de información que superan la capacidad de las herramientas tradicionales de gestión de datos. Según Gartner (2012), Big Data “comprende volúmenes de información de gran tamaño, velocidad y variedad que demandan tecnologías innovadoras para capturar, gestionar y analizar datos con el fin de generar valor”. En el ámbito del presente proyecto, Big Data permite almacenar y procesar múltiples grabaciones de voz obtenidas de diferentes pacientes, sesiones y ejercicios, generando una base de conocimiento sobre patrones fonéticos comunes.

El uso de Big Data en logopedia permite crear modelos predictivos que identifican errores sistemáticos en la pronunciación y evalúan la evolución del paciente a lo largo del tiempo. Esto posibilita pasar de terapias reactivas a terapias predictivas y adaptativas, basadas en evidencia cuantitativa. En este sentido, el proyecto se alinea con la tendencia mundial de la salud digital, donde los datos masivos se convierten en una fuente de mejora continua para la atención personalizada.

Relación con las 5 V de Big Data

V	Aplicación en el Proyecto
Volumen	Se espera recolectar 70 grabaciones de voz en la fase inicial, expandiéndose a miles conforme se integren nuevos pacientes. Cada archivo contiene múltiples métricas fonéticas (F0, jitter, shimmer, HNR, MFCCs), generando una base de datos de gran tamaño que requiere almacenamiento escalable (por ejemplo, en la nube o en sistemas distribuidos).
Velocidad	El procesamiento de grabaciones se realizará por lotes, pero con potencial de evolución hacia análisis en tiempo real mediante streaming de audio. Python permitirá la automatización del preprocesamiento y Power BI actualizará los dashboards dinámicamente para terapeutas.
Variedad	Los datos incluirán audios WAV, metadatos estructurados (edad, género, tipo de ejercicio), y análisis fonéticos (json con métricas acústicas). Se integrarán tanto datos estructurados como no estructurados, gestionados en una arquitectura híbrida.
Veracidad	Se aplicarán procedimientos de limpieza y normalización de audio para eliminar ruido, junto con filtros de calidad ($\text{SNR} \geq 20 \text{ dB}$). Además, los datos se anonimizarán y validarán por terapeutas certificados.
Valor	Los resultados se traducirán en informes visuales interactivos en Power BI que mostrarán la evolución de cada paciente y las recomendaciones personalizadas, optimizando el proceso terapéutico y la toma de decisiones clínicas.

Justificación Técnica

El enfoque de análisis adoptado es descriptivo y predictivo, combinando exploración inicial de datos acústicos con modelos de Machine Learning capaces de predecir la mejora fonética a lo largo del tratamiento. El análisis descriptivo permitirá

comprender las distribuciones, correlaciones y patrones dentro del conjunto de datos, mientras que el análisis predictivo ofrecerá estimaciones sobre el progreso esperado de cada paciente.

Se ha optado por Python debido a su robustez en análisis de datos, facilidad para implementar pipelines y compatibilidad con bibliotecas científicas. Herramientas como pandas y numpy manejarán los datos tabulares, librosa realizará el análisis de señales acústicas y scikit-learn permitirá entrenar y validar modelos predictivos. Por otro lado, Power BI se integrará para la visualización interactiva y generación de reportes clínicos accesibles a terapeutas y familias.

La arquitectura propuesta se basa en un pipeline modular controlado desde un archivo principal (main.py) que coordina los procesos de carga, limpieza, análisis y almacenamiento. Esta estructura garantiza la reproducibilidad y escalabilidad del proyecto, permitiendo ampliar el flujo de trabajo conforme se agreguen nuevos módulos o fuentes de datos.

Finalmente, la claridad del flujo de trabajo se asegurará mediante documentación técnica, control de versiones y visualizaciones estandarizadas. De esta manera, el sistema no solo proporcionará resultados precisos, sino que también servirá como una herramienta educativa para otros investigadores interesados en la aplicación de Big Data a la logopedia y la rehabilitación del habla.

Bibliografía

- Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer program]. University of Amsterdam.
- Gartner IT Glossary. (2012). *Definition of Big Data*. Gartner Inc.

- Ladefoged, P. (2001). *A Course in Phonetics* (4th ed.). Harcourt College Publishers.
- *(DTW) Techniques*. *Journal of Computing*, 2(3), 138–143.
- Wickham, H. (2014). *Tidy Data*. *Journal of Statistical Software*, 59(10), 1–23.