



FACULTAD DE MEDICINA

DISEÑO Y DESARROLLO DE DISPOSITIVO DE DOSIMETRÍA DE VOZ

Tesis Doctoral

MARCOS LLORENTE ORTEGA

Pamplona 2019

La presente tesis, titulada

"Diseño y desarrollo de dispositivo de dosimetría de voz",

Presentada por **D. Marcos Llorente Ortega**, realizada bajo nuestra dirección en el
Programa de Doctorado en Investigación Médica Aplicada,
en la Universidad de Navarra.

El resultado de sus investigaciones queda expuesto en esta Memoria que, una vez
revisada reúne los requisitos de calidad y originalidad requeridos por la Ley.

Revisado el texto, estoy conforme con su presentación para ser defendida.

Dr. Secundino Fernández González

Director de la tesis

Dr. Adam Podhorski

Co-Director de la tesis

Pamplona, 17 de junio de 2019

*Sin desafíos la vida es una rutina,
una lenta agonía.*

-Albert Einstein

AGRADECIMIENTOS.

A la Universidad de Navarra, por confiar en este proyecto, aceptando su realización y defensa como mi tesis doctoral.

A la Asociación de Amigos de la Universidad de Navarra, por su apoyo fundamental para la realización de esta tesis.

Al Dr. Secundino Fernández, por el apoyo y confianza desde incluso antes de acabar el grado hace ya seis años. Un largo camino que juntos hemos recorrido: trabajo fin de grado, trabajo fin de máster, tesis doctoral y que continúa con el Laboratorio de Ingeniería Médica (LabIM).

Al Dr. Adam Podhorski, por su tiempo y formación en un campo de sistemas embebidos y procesado de señal del que mi conocimiento era muy reducido.

A Beatriz, por prestar su voz y ser mi sujeto de pruebas con infinita paciencia y máxima disponibilidad.

A Xabi, por el apoyo y comprensión durante la tesis y por el gran trabajo realizado que ha hecho que juntos levantemos el LabIM.

Al equipo (Nekane, David y Rubén), por los ánimos recibidos y los buenos momentos vividos, dentro y fuera de la Universidad.

A mis amigos, por seguir creyendo en mí a pesar de estar un año diciendo “en un mes deposito”.

A Ana, porque durante todos estos años ha estado en los buenos y malos momentos, tirando de mí para que no me relajase y apoyándose siempre que lo he necesitado.

Y por supuesto, agradecer a mi familia su preocupación y estima, haciéndome saber que ellos siempre están y estarán para lo que necesite.



FACULTAD DE MEDICINA

DISEÑO Y DESARROLLO DE DISPOSITIVO DE DOSIMETRÍA DE VOZ

Trabajo presentado por Marcos Llorente Ortega, para optar al grado de Doctor en
Investigación Médica Aplicada.

Pamplona 2019

I. ÍNDICE

I. ÍNDICE	11
II. ABREVIATURAS	17
III. PRÓLOGO	21
IV. INTRODUCCIÓN	25
1. MOTIVACIONES.....	28
2. HIPÓTESIS	29
3. OBJETIVOS	29
3.1. General	29
3.2. Específicos	30
V. ESTADO DEL ARTE	31
1. FUNDAMENTOS DE LA VOZ	33
2. PARÁMETROS DE LA VOZ	33
2.1. Frecuencia fundamental (F_0).....	35
2.2. Intensidad (I)	35
2.3. Tiempo de fonación (t).....	36
3. DISPOSITIVOS DE DOSIMETRÍA DE LA VOZ	37
4. ESTUDIOS BASADOS EN DOSÍMETROS DE VOZ	42
VI. MATERIAL Y MÉTODO	45
1. DESARROLLO DEL SISTEMA.....	47
1.1. Hardware.....	47
1.1.1. Acelerómetro.....	47
1.1.2. Microcontrolador	49
1.1.3. Colgante.....	51
1.1.4. Circuito.....	52
1.2. Algoritmo.....	53
1.2.1. Registro de los valores del sensor	54
1.2.2. Procesado de los datos.....	58
1.2.2.1. Cálculo de la intensidad relativa.....	59
1.2.2.2. Cálculo de la FFT	60

1.2.2.3. Cálculo de la magnitud de la FFT	64
1.2.2.4. Filtrado de la magnitud de la FFT	67
1.2.2.5. Cálculo de la autocorrelación	69
1.2.2.6. Cálculo de la frecuencia fundamental de la voz.....	73
1.2.2.7. Grabación de datos.....	75
1.3. Gestión de la memoria	77
2. VALIDACIÓN.....	78
2.1. Validación del microcontrolador y del algoritmo	78
2.2. Validación de los parámetros de interés en la voz humana	79
2.2.1. Validación de la frecuencia fundamental.....	80
2.2.2. Validación de la intensidad relativa	80
2.2.3. Validación del tiempo de fonación.....	82
3. REGISTRO DE VOZ DURANTE LARGOS PERIODOS DE TIEMPO	84
VII. RESULTADOS	85
1. DESARROLLO DEL SISTEMA.....	87
1.1. Hardware.....	87
1.2. Algoritmo.....	89
1.2.1. Tiempo empleado para ejecutar cada proceso.....	89
1.2.2. Consumo de energía.....	94
2. VALIDACIÓN.....	95
2.1. Validación del microcontrolador y del algoritmo	95
2.2. Validación de los parámetros de interés en voz humana.....	109
2.2.1. Validación de la frecuencia fundamental.....	109
2.2.2. Validación de la intensidad relativa	112
2.2.3. Validación del tiempo de fonación.....	115
3. REGISTRO DE VOZ DURANTE LARGOS PERIODOS DE TIEMPO	116
3.1. Sujeto 1.....	117
3.1.1. Análisis día normal del sujeto 1.....	120
3.1.2. Análisis día hiperfuncional del sujeto 1.....	129
3.2. Sujeto 2.....	138
3.2.1. Análisis de 3 días.....	141
VIII. DISCUSIÓN	147
1. INTRODUCCIÓN.....	149

2. HARDWARE.....	150
3. VALIDACIÓN.....	151
4. REGISTROS DE VOZ.....	154
5. LIMITACIONES Y LÍNEAS FUTURAS	156
IX. CONCLUSIONES	157
X. BIBLIOGRAFÍA.....	161
XI. GLOSARIO	167
XII. ANEXOS.....	175
ANEXO I. Documento técnico VoxLog	177
ANEXO II. Vocal Holter – Prototype Med 1.1.1	179
ANEXO III. Vocal Holter – Prototype Med 2.0.0	181
ANEXO IV. Datasheet acelerómetro	183
ANEXO V. Diagrama de bloque del microcontrolador msp430fr5994	185
ANEXO VI. Dibujo técnico de prototipo colgante	187
ANEXO VII. Dibujo técnico de soporte para anclar circuito a tarjeta de desarrollo	189
ANEXO VIII. Fonetograma registrado con el dispositivo y con LingVAWES.....	191
ANEXO IX. Datasheet lector de tarjeta microSD.....	193
ANEXO X. Datasheet LiPo batería 400 mAh.....	195
ANEXO XI. Dibujo técnico de dispositivo final	197
ANEXO XII. Boxplot de los 14 días de frecuencia fundamental e intensidad relativa de sujeto 1	199
ANEXO XIII. Boxplot de los 11 días de frecuencia fundamental e intensidad relativa de sujeto 2.....	201
ANEXO XIV. Comunicaciones	203

II. ABREVIATURAS

ADC	Analog to Digital Converter (Conversor analógico digital)
APM	Ambulatory Phonation Monitor (Monitor de fonación ambulatorio)
CPU	Central Processing Unit (Unidad central de procesamiento)
CUN	Clínica Universidad de Navarra
DMA	Direct Memory Access (Acceso directo a memoria)
EGG	Electroglotografía
FFT	Fast Fourier Transform (Transformada rápida de Fourier)
FRAM	Ferroelectric Random Access Memory (Memoria de acceso aleatorio ferroeléctica)
LEA	Low Energy Accelerator (Acelerador de baja energía)
ME	Modo de Energía
PIO	Presión subglótica
RAM	Random Access Memory (Memoria de acceso aleatorio)
RL	Resistencia laríngea
RMS	Root Mean Square (Error medio cuadrático)
RTC	Real Time Clock (Reloj de tiempo real)
SPI	Serial Peripheral Interface (Interfaz Periférica Serial)
SPL	Sound Pressure Level (Nivel de presión sonora)
SRAM	Static Random Access Memory (Memoria de acceso aleatorio estática)
TI	Texas Instruments
V	Voltio

III. PRÓLOGO

El trabajo que se describe a continuación se enmarca dentro de un entorno particular. Se trata de la primera tesis desarrollada por un ingeniero en el marco del programa de Doctorado de Investigación Médica Aplicada de la Facultad de Medicina de la Universidad de Navarra con un enfoque clínico y cuyo cuerpo es principalmente técnico.

En esta posición, la comunicación interdisciplinar cobra especial importancia para posibilitar el entendimiento entre médico e ingeniero. Parte fundamental para el desarrollo de esta habilidad de comunicación se encuentra en el entorno de trabajo. El Laboratorio de Ingeniería Médica (LabIM), lugar donde se desarrolló el proyecto, ofrece el clima idóneo debido a que todo el trabajo que se realiza en él es colaboración entre médicos de la Clínica Universidad de Navarra o de la Facultad de Medicina de la Universidad de Navarra e ingenieros.

Con el objetivo de que toda explicación sea entendida por el lector, proceda del entorno clínico o del ingenieril, el lenguaje se ha adaptado a esta interdisciplinariedad. Se han explicado detalles que podrían darse por supuestos para unos, para que los otros entiendan la totalidad. En el glosario aparecerán las definiciones de aquellos términos que sea necesario definir puesto que, dependiendo del ámbito, pueden usarse diferentes acepciones. También se han adaptado cuánto se profundiza en las explicaciones aportando la información necesaria y relevante para el proyecto. Además, se procurará implicar el menor número de detalles que requirieran explicación que no aporten valor al documento.

IV. INTRODUCCIÓN

La objetivación y cuantificación de los principales parámetros acústicos de la voz es un método de análisis muy relevante para los especialistas en laringología e investigadores de la voz, a la hora de valorar los trastornos fonatorios. La caracterización acústica precisa de la voz supone una herramienta muy útil tanto desde el punto de vista diagnóstico de las diferentes entidades patológicas como para la constatación de resultados de los diferentes tratamientos. Además, la caracterización acústica permite la valoración de patrones de riesgo a la hora de prevenir trastornos en el futuro. El análisis instrumental de los trastornos de la voz conlleva el estudio acústico, aerodinámico y electroglotográfico de los elementos respiratorios, valvulares, vibratorios y resonadores que contribuyen a la génesis de la voz. Estas pruebas de análisis se llevan a cabo con equipamiento especializado y disponible habitualmente en los laboratorios de voz adscritos a consultas de laringología y centros de investigación de la voz donde valoran: la frecuencia fundamental (F_0 en Hz), intensidad (I en dB_{SPL}), el espectrograma acústico, el contacto de la superficie de las cuerdas vocales (EGG), el flujo transglótico (Volumen en l/s), la presión subglótica (PIO en mmH_2O), la resistencia laríngea (RL en $\text{mmH}_2\text{O/l/s}$) y otras variables de interés. Estos estudios se realizan en un momento concreto, en unas instalaciones y con unos medios y condiciones determinadas. Los resultados, aunque son muy informativos, son específicos de ese momento y se realizan en unas condiciones concretas de laboratorio. El estudio y seguimiento continuo en el medio habitual del paciente de los parámetros mencionados y otros como el tiempo de uso de voz, durante largos periodos de tiempo, queda restringido a casos concretos para investigación.

Además del análisis de la señal acústica de la voz, se puede emplear otro método de medición para valorar la actividad fonatoria de las cuerdas vocales durante la fonación, y que no se utiliza habitualmente en el ámbito clínico, que se basa en la cuantificación de las vibraciones de la piel a nivel de la laringe mediante acelerómetros.^{1,2} Este método de medición tiene varias ventajas entre las que se encuentra la posibilidad de cuantificación y monitorización de diferentes parámetros fisiológicos que suponen una fuente de información diagnóstica y pronóstica muy relevante que respeta la intimidad del paciente en lo referente al contenido de su locución. La tecnología permite

implementar técnicas y medidas de parámetros fisiológicos cada vez más fiables, precisas y que no supongan ningún tipo de incomodidad ni molestia para el paciente.

1. MOTIVACIONES

Para dar una respuesta adecuada a los trastornos y enfermedades de los pacientes es cada vez más importante poder realizar tanto el diagnóstico como el tratamiento de forma individualizada, que es lo que viene a denominarse como “medicina personalizada”. Esta manera de enfocar los problemas, aunque comúnmente puede asociarse solo al estudio genérico de los problemas, se refiere al análisis y control de los diferentes parámetros que definen los problemas en los ámbitos de actuación reales de manera individualizada.

En el entorno sanitario, el estudio diagnóstico y el tratamiento personalizado, así como el seguimiento del paciente fuera del entorno hospitalario, tienen un impacto positivo en los resultados del paciente en términos de precisión diagnóstica, eficiencia terapéutica y calidad de vida.

En la actualidad, la medición de la actividad, las calorías consumidas, el registro de la calidad del sueño, la medición de la temperatura corporal, la frecuencia cardíaca, la saturación de oxígeno en un determinado momento, a lo largo de una actividad concreta o durante varios días son registros habituales y posibles con dispositivos asequibles para cualquiera. El progreso de la tecnología es la clave de este avance.

La tecnología permite disponer de sensores más pequeños y precisos, procesadores más potentes y eficientes, baterías más duraderas y comunicaciones más accesibles de manera que cada vez es más común que los dispositivos se conecten entre ellos y los sistemas de comunicación (desde el hardware hasta los protocolos) sean más fáciles de implementar.

Son muchos los parámetros fisiológicos que se están registrando por dispositivos cada vez más asequibles, precisos y exactos; sin embargo, es llamativo el escaso desarrollo relacionado con el registro de parámetros de la voz desde el punto de vista fisiológico o médico, teniendo en cuenta que es el método de comunicación habitual de las personas y que para muchas es su herramienta fundamental de trabajo. Por ejemplo, en Estados

Unidos se estima que el 6.6 % de la población en edad de trabajar sufre o sufrirá algún desorden de la voz.³

En el ámbito de este trabajo es posible encontrar cuatro dispositivos comerciales que han intentado cumplir los requisitos: “Ambulatory Phonation Monitor” (KayPENTAX, EE.UU.), VoxLog (Sonvox AB, Suecia), VocaLog (Griffin Laboratories, EE.UU.) y Voice-Care (PR.O. VOICE, Italia). En el momento actual solo VoxLog, Vocalog y Voice-Care siguen comercializándose y son muy difíciles de conseguir. Quizá el elevado coste, entre los 500 y 5000 dólares y problemas de índole práctica (tamaño, peso, sistema de fijación del sensor...) hayan contribuido a que su utilización esté muy dirigida hacia la investigación y menos hacia el interés clínico que sin duda tienen estos dispositivos.

2. HIPÓTESIS

Nos preguntamos si es posible desarrollar un dispositivo vestible que permita la monitorización durante largos periodos de tiempo de los principales parámetros de la voz.

Si es posible diferenciar patrones fonatorios.

3. OBJETIVOS

3.1. General

Desarrollar un dispositivo vestible que registre y mida los principales parámetros (frecuencia fundamental e intensidad relativa) de la voz, así como el tiempo de uso de la voz realizando todo el procesado de manera interna. Con ello se posibilitará el seguimiento de pacientes fuera del entorno hospitalario y disponer de datos objetivos del uso y características de la voz con fines diagnósticos y terapéuticos por parte de los profesionales sanitarios y pacientes.

3.2. Específicos

El dispositivo debería:

- Permitir la medición de la frecuencia fundamental.
- Permitir la medición de la intensidad relativa de la voz.
- Permitir la medición del tiempo de fonación.
- Tener un tamaño reducido para permitir que sea fácilmente vestible y no interfiera en la emisión normal de la voz y vida cotidiana del sujeto que porte el dispositivo.
- Tener un consumo energético contenido para posibilitar la monitorización durante uno o varios días de forma ininterrumpida.
- Almacenar los parámetros medidos en memoria extraíble.
- Tener un precio asequible para su comercialización a profesionales, investigadores y usuarios domésticos.

V. ESTADO DEL ARTE

1. FUNDAMENTOS DE LA VOZ

La voz es el producto de la modificación en las cavidades supraglóticas del sonido generado en las cuerdas vocales. La acción de generar este sonido se denomina fonación. Los pliegues vocales, por la acción de los músculos intrínsecos laríngeos, se ponen en contacto cerrando la glotis. El cierre de la glotis unido a la liberación de aire de los pulmones eleva la presión subglótica. De acuerdo con la teoría muco-ondulatoria enunciada por Jorge Perelló en 1962⁴, en el momento en que la presión subglótica supera un umbral, se produce ondulación de la mucosa de las cuerdas vocales de abajo hacia arriba dejando pasar pulsos de aire; una vez liberado el aire de un pulso la mucosa de las cuerdas vocales vuelve a contactar cerrándose. Estos ciclos de apertura y cierre de las cuerdas vocales, son los que generan las vibraciones en la laringe y el sonido que, en las cavidades resonadoras, es amplificado, filtrado y liberado a través de la boca y nariz para generar la voz. Estas vibraciones se transmiten por el tejido circundante y llegan hasta la superficie de la piel. La frecuencia de estas vibraciones es la frecuencia fundamental de la voz y la presión de aire liberada determina la intensidad de la misma. Técnicamente se denomina al tejido pliegue vocal, aunque está admitido el coloquialismo cuerda vocal. De igual manera el término coloquial vibración de la cuerda vocal hace referencia al movimiento ondulatorio del pliegue vocal y también está admitido. Por ello, de aquí en adelante se emplearán ambos términos.

2. PARÁMETROS DE LA VOZ

En la voz existen tres parámetros fundamentales: frecuencia fundamental, intensidad y tiempo de fonación. La frecuencia y el tiempo de fonación se pueden extraer directamente de las mediciones del sensor, mientras que la intensidad de la voz debe estimarse a partir de la amplitud de las vibraciones registradas por el dispositivo y un proceso de calibración.⁵⁻⁷

Como se muestra en la Tabla 1, tres de los dispositivos que han sido comercializados para la dosimetría de voz miden las tres variables básicas. El APM (Ambulatory Phonation Monitor) además, partiendo de la información de tiempo de fonación y de la frecuencia fundamental, calcula el número de ciclos ondulatorios, y con ellos

(conociendo el recorrido medio de las cuerdas vocales) calcula la distancia recorrida por las cuerdas vocales.⁸

Variables	APM	VoxLog	VocaLog	Voice-care
Intensidad (dB _{SPL})	X	X	X	X
Respuesta a la intensidad	X	X	X	
Frecuencia fundamental (Hz)	X	X		X
Respuesta a la frecuencia	X	X		
Tiempo de fonación (hh:mm:ss)	X	X	X	X
Ruido ambiente (dB _{SPL})		X		
Número de ciclos	X			
Distancia recorrida por las cuerdas (metros)	X			
Calibración de intensidad	Diaria	No necesaria	Una vez	Una vez

Tabla 1. Diferencias entre los dispositivos comercializados.^{8,9}

En cuanto a la respuesta a la frecuencia y a la intensidad emitida por los dispositivos, éstas hacen referencia a que los dispositivos emiten avisos cuando el parámetro al que se asocia la respuesta supera unos límites preestablecidos. La intención de esta respuesta es que el usuario no sobrepase ciertos límites de una determinada variable de la voz.

Por último, destacar que VoxLog es capaz de medir el ruido ambiente gracias a que incorpora un micrófono. Al tener micrófono integrado, el dispositivo VoxLog también es capaz de hacer una autocalibración de la intensidad.

2.1. Frecuencia fundamental (F_0)

La frecuencia fundamental de la voz, número de pulsos de aire que atraviesan la glotis por segundo, en una persona adulta se encuentra entre los 180-250 Hz en mujeres y los 100-150 Hz en hombres.¹⁰

La modificación de este parámetro a lo largo del tiempo está relacionada con el aumento de la carga vocal.¹¹⁻¹⁷ Se define como carga vocal el esfuerzo realizado por el órgano fonador debido al uso de la voz.¹⁸ Por ejemplo, aumenta la frecuencia fundamental como motivo del esfuerzo para hacerse oír.^{17,19}

Aumentos de la frecuencia fundamental por encima de la normal para cada individuo, sin la técnica apropiada, puede conllevar lesiones orgánicas.

También se menciona en la bibliografía que la frecuencia fundamental varía con el simple hecho de modificar las condiciones de entorno donde se mide, por ejemplo estando en el laboratorio de voz o fuera de él.¹⁹ Por ello es importante un seguimiento continuo y a lo largo de un amplio periodo de tiempo para poder obtener datos concluyentes.

La medición de la frecuencia fundamental a través de la señal del acelerómetro puede realizarse de diferentes modos,²⁰ siendo los más comunes el uso de la transformada rápida de Fourier (FFT)^{21,22} y la autocorrelación²³.

2.2. Intensidad (I)

La intensidad es la medida de presión del sonido expresada en decibelios (dB_{SPL}). Siendo esos dB_{SPL} la relación en escala logarítmica entre la presión sonora eficaz P_1 y la presión de referencia P_0 (20 μPa).

$$SPL = 20 \times \log \left(\frac{P_1}{P_0} \right) \quad (1)$$

Para medir la intensidad habitualmente se emplea un sonómetro. Sin embargo, la presión puede correlacionarse con la amplitud de la vibración del acelerómetro y

de este modo es posible calcular la intensidad.²⁴ Para esto, es preciso una calibración específica del sistema.^{25,26}

Son varios los estudios que destacan el aumento de la intensidad debida a una carga vocal prolongada durante el día.^{11,12,16,17}

2.3. Tiempo de fonación (t)

Esta medida hace referencia al tiempo fonado durante la medición o como lo describen T. Masuda et al. (1993) el periodo durante el cual las cuerdas vocales están vibrando.²⁷ Anteriormente en 1989 A. Ohlsson et al. también definieron este tiempo como aquel durante el cual ocurren periodos fundamentales detectables. Posteriormente, en 2001, A. Szabo et al. aclaran la diferencia entre tiempo de fonación y tiempo de locución. El primero es aquel en el que las cuerdas vocales vibran y el segundo incluye los segmentos sordos y sonoros además de los tiempos de silencio. El tiempo de fonación, al igual que la frecuencia fundamental, se puede medir a partir de la señal recibida por el acelerómetro.²⁰

Los trastornos de la voz más comunes son crónicos o recurrentes como resultado de un abuso o mal uso de la voz.²⁸

La diferencia en el tiempo de uso de la voz puede observarse entre personas con voz sana y personas con problemas en la voz.²⁹

Partiendo de estos tres parámetros (F_0 , I, t), podemos calcular otros parámetros de la voz que puedan ser de interés, como el número de ciclos vocales o la distancia recorrida por las cuerdas vocales como calcula el APM.³⁰ Aunque probablemente los más interesantes sean el *shimmer* y *jitter*. Estos parámetros miden la variabilidad de la amplitud (*shimmer*) y de la frecuencia (*jitter*). En la bibliografía se pueden encontrar diferentes fórmulas para calcular cada uno de estos parámetros.³¹

Por otro lado, M. Zanartu et al. presentó en 2014 el modelo para la estimación de la presión subglótica a través de mediciones realizadas con acelerómetro.³² Y usando este modelo se ha comprobado recientemente que es más precisa la medición de la presión

subglótica que la intensidad de la voz, por lo que las medidas obtenidas a partir de acelerometría aportan información valiosa.³³

3. DISPOSITIVOS DE DOSIMETRÍA DE LA VOZ

Los dispositivos de dosimetría o monitorización de la voz se pueden dividir en dispositivos de monitoreo o de respuesta, aunque también los hay que integran ambas categorías. Los dispositivos de monitoreo tienen como objetivo registrar los parámetros medidos para un posterior análisis. Los de respuesta son aquellos que, según la medida realizada en tiempo real y unos límites establecidos en el dispositivo, responden con algún tipo de estímulo (acústico o vibratorio) para avisar de que el límite preestablecido ha sido superado.

La mayoría de estos dispositivos usan un micrófono para el registro de la señal, pero a partir de 2003 con el trabajo de HA. Cheyne et al.⁶ se comenzaron a emplear acelerómetros para realizar estos registros.

Del mismo modo que se diferencian en el sensor utilizado, también se diferencian en la localización del sensor. Los micrófonos pueden localizarse al aire libre enfrente de la boca, o en contacto con la piel en la región supraesternal bajo la laringe. Los acelerómetros se colocan todos ellos en la zona supraesternal por debajo del cartílago tiroides.

Otro factor importante que diferencia los dispositivos de monitoreo son los parámetros que analizan. Frecuencia, intensidad y tiempo de fonación son los parámetros más comúnmente analizados, pero cada dispositivo tiene una combinación entre ellos distinta, pudiendo medir uno de ellos, dos o tres. Algún dispositivo calcula otros parámetros a partir de las medidas.³²

En la siguiente tabla se analizan todos los dispositivos de dosimetría o monitoreo según las características mencionadas y según otros parámetros.

Autor	Año	Nombre	Monitoreo o respuesta	Sensor	Parámetros	Tiempo de funcionamiento	Frecuencia de muestreo (Hz)	Intervalo de análisis (subintervalo)	Tamaño (cm)	Peso (gr)	Coste
Holbrook, Rolnick, & Bailey ³⁴	1974	VIC	Respuesta	Micrófono de contacto	Amplitud y tiempo de fonación				12x6x4		
Zicker, Tompkins, Rubov/, & Abbs ³⁵	1980		Respuesta	Micrófono	Amplitud		200	0,5 s	15x10x8	1000	
Ryu, Komiyama, Kannae, & Wotanabe ³⁶	1983		Monitoreo	Micrófono de contacto	Amplitud y tiempo de fonación	12 horas		60 s ⁶	9x6x3	150	
Ohlsson, Brink, & Löfqvist ³⁷	1989	RA-LOF01	Monitoreo	Micrófono de contacto	Frecuencia y tiempo de fonación	12 horas	1200-1800	6 s (200 ms)	19x11x5	475	
Masuda, Ikeda, Manako, & Komiyamo ²⁷	1993		Monitoreo	Micrófono de contacto	Amplitud y tiempo de fonación	12 horas	0,1 segundos				

McGillivray, Proctor-Williams, & McLister ³⁸	1994					Amplitud													
Rantala, Haataja, Viikman, & K�rkk� ³⁹	1994				1,5 horas	Frecuencia y amplitud		5000		70-250 ms (2,5-10 ms)									
Buekers, Bierens, Kingma, & Marres ⁴⁰	1995				12 horas	Amplitud y tiempo de fonaci�n		1		1 s		15x9x4		600					
Airo, Olkinuora, & Sala ⁴¹	2000				8 horas	Amplitud y tiempo de fonaci�n		100		2 s (10 ms)				120					
Szabo, Hammarberg, Hakansson, & Sodersten ⁴²	2001		VACLF1		Aprox. 12 horas	Frecuencia y tiempo de fonaci�n				1-10 s (50, 100, 200, 500 ms)		14x7x3		250					
Cheyne, Hanson, Genereux, Stevens, & Hillman ⁶	2003		PVA o APM		12 horas	Frecuencia, amplitud y tiempo de fonaci�n		11025		125 ms (25 ms)		12x8,5x2		200					5000\$

Autor	Año	Nombre	Monitoreo o respuesta	Sensor	Parámetros	Tiempo de funcionamiento	Frecuencia de muestreo (Hz)	Intervalo de análisis (subintervalo)	Tamaño (cm)	Peso (gr)	Coste
Popolo, Svec, & Titze ⁵	2005		Monitoreo	Acelerómetro	Frecuencia, amplitud y tiempo de fonación	2-3 horas		30 ms	13x8x1,6	180	
Marcus Wirebrand ²⁰	2011	VoxLog	Monitoreo y respuesta	Acelerómetro y micrófono	Frecuencia, amplitud y tiempo de fonación	2-14 horas ^(a)		5-60 s (100-1000 ms) ^(a)	12x2,5x2,5 ₈		5000
Lindstrom; Wayne, Södersten; McAllister, Ternström ⁴⁴	2011		Monitoreo	Acelerómetro y micrófono	Frecuencia, amplitud y tiempo de fonación		48000	30 ms			
Mehta, Zafartu, Feng, Cheyne, & Hillman ⁴⁵	2012	VHIM	Monitoreo y respuesta ⁴³	Acelerómetro	Frecuencia, amplitud y tiempo de fonación	18 horas		50 ms (25 ms)		175 ^(b)	
Carullo, Vallan & Astolfi ²³	2013	Voice-Care	Monitoreo	Micrófono de contacto	Frecuencia, amplitud y tiempo de fonación	11 horas ⁹ 5 horas ^(c) 6 horas ^(d)	38460 (calibración) 19230 (monitoreo)	30 ms	15x10x5 ^(b)		400€ ⁹ 900€ ^(c) 1800€ ^(d)

Griffin Laboratories, Temecula, CA, EE.UU.	2014 ⁴⁹	Vocalog	Monitoreo y respuesta ⁸	Micrófono de contacto ⁸	Amplitud y tiempo de fonación ⁸	3 semanas ^{(e), 8}	1 s ⁸	7x5,5x1,5	999\$
	2016 ⁵⁰	Vocalog2	Monitoreo y respuesta ⁴⁷	Micrófono de contacto ⁴⁷	Amplitud y tiempo de fonación ⁴⁷	3 días ⁴⁸	1s ⁴⁷		394\$ ⁸
	2017		Monitoreo	Acelerómetro y micrófono	Frecuencia, amplitud y tiempo de fonación	8 horas	50ms	68x14,5x5 ^(f)	12,5 ^(e)

Tabla 2: Recopilatorio dispositivos de dosimetría de voz.

^(a) Es el tiempo máximo por grabación, puede almacenar 100 grabaciones dispone de memoria para 1400 horas. No se menciona duración de la batería (ANEXO I).

^(b) Calculado a partir de los componentes citados en el artículo. Medidas Nexus S, Peso 139 g (Nexus s) – 9 (batería original Nexus S) +45 (batería ampliada Mugen 3900mAh XL extended battery).

^(c) ANEXO II

^(d) ANEXO III

^(e) Tiempo que es capaz de almacenar en memoria, no se comenta la duración de la batería.

^(f) Esas son las medidas del emisor colocado sobre la piel. El receptor mide 59x25x10 mm.

^(g) Es el peso del emisor colocado sobre la piel. El receptor pesa 47,5 g sin incluir la batería que no mencionan.⁴⁶

Algunos de los dispositivos mencionados se muestran en las siguientes figuras.



Figura 1: Dispositivos de monitorización de la voz comercializados.⁵⁰



Figura 2: Dispositivos en desarrollo de monitorización de la voz basados en teléfonos móviles.⁵⁰

4. ESTUDIOS BASADOS EN DOSÍMETROS DE VOZ

A continuación, se presenta una revisión de estudios en los que se han utilizado dispositivos de monitorización de la voz. Con esto se quiere mostrar la relevancia de los datos obtenidos y el abanico de entornos susceptibles de su implementación.

Gran parte de las publicaciones tiene relación con el colectivo de profesores debido a que la voz es una de sus herramientas de trabajo más relevantes. L. Rantala y E. Vilkman (1999) relacionan los problemas de la voz del profesorado con la carga vocal a la que exponen a su aparato fonador.¹⁴ En ese mismo artículo se menciona cómo las diferencias entre los dos grupos estudiados son mayores al final de cada día y al final de la semana. Años más tarde en un estudio similar llegan a las mismas conclusiones.¹⁵

E. Hunter y I. Titze (2010) documentaron el uso de la voz de 57 profesores durante dos semanas y con la información recopilada constataron que el tiempo de fonación en horario laboral (29,9 % de cada hora) era de más del doble con respecto al periodo no laboral, la intensidad era 2.5 dB_{SPL} mayor y la frecuencia natural entre 1 y 1.5 tonos superior.⁵² Algún estudio ha analizado el uso de la voz por el profesorado en diferentes circunstancias, como puede ser el empleo de amplificadores. Los dosímetros de voz descritos en estos artículos, concretamente el APM, han servido para demostrar que el uso de amplificadores disminuye la intensidad de la voz y por lo tanto reduce la carga sobre la misma.^{16,53,54} Registros parecidos fueron tomados con otro dispositivo, Voice-Care, y mostraron que una mala acústica del aula conlleva un aumento de la intensidad empleada por los profesores.⁵⁵ Otra faceta estudiada es la diferencia del uso de la voz entre las horas laborables (23 %) y el descanso (12 %).⁵⁶ Incluso dentro de las horas laborables pueden llegar a identificar estilos de enseñanza^{39,57} o nivel de enseñanza.³⁰ El ruido ambiente al que se enfrenta el locutor también es una variable relacionada con el aumento de la intensidad.⁵⁸

Aunque el profesorado es uno de los grupos más afectado por los trastornos de la voz y por lo tanto el más estudiado, hay otras profesiones que tienen una carga vocal elevada como los logopedas, entrenadores, enfermeras, teleoperadores o recepcionistas.⁴⁰ Toda profesión en la que la voz es instrumento importante, es susceptible de ser estudiada con estos dispositivos.

A nivel clínico, los parámetros medidos con estos dispositivos tienen una gran relevancia diagnóstica y en la valoración del pronóstico. La relación entre la disfonía y el tiempo de fonación es la conclusión principal a la que llegan muchos trabajos.^{13,29,59}

El uso de un sistema de registro de voz con respuesta desarrollado por R. McGillivray et al. en 1993 con población infantil, demuestra cómo en apenas 6 sesiones de entre 20 y 30 minutos el niño consiguió el objetivo de mantener una intensidad menor a 65 dBa.³⁸ En 2015 Van Stan et al. describen el uso de respuesta en días alternativos. Los resultados muestran cómo los días en los que el dispositivo de respuesta está activo la intensidad disminuye, sin embargo, esta disminución no es asimilada pues en los días sin respuesta por parte del dispositivo se registran valores similares a los iniciales.⁶⁰ El campo de estudio con el uso de este tipo de dispositivos no queda restringido únicamente a la medición de los parámetros de la voz, ya que pueden estudiarse las pausas.^{56,61} Con estas medidas, A. Holbrook et al. describieron en 1974 el primer caso de uso de un sistema de dosimetría de voz con respuesta, demostrando cómo el uso de este tipo de dispositivo puede incluso evitar la cirugía. En el estudio se menciona que 11 de los 32 pacientes que llevaron el dispositivo experimentaron una completa recuperación (pólipos, nódulos o úlceras de contacto) y que otros 8 vieron reducido el tamaño de sus lesiones pudiendo evitar la cirugía.³⁴

El trabajo realizado por Y. Horii y B. Fuller en 1990 constató que, en pacientes sometidos a intubación orotraqueal relativamente corta (1.5-23.5 horas) para ser sometidos a una cirugía, los valores de shimmer y jitter eran mayores en vocales sostenidas y aumentó la frecuencia fundamental media en la lectura.¹ Más recientemente se ha estudiado la voz antes y después de una cirugía laríngea. A los pacientes sometidos a cirugías se les recomienda reposo, y estos dispositivos de dosimetría de la voz permiten el seguimiento de ese reposo. Así, Misono et al., comprobaron cómo los pacientes del estudio redujeron tras la cirugía del 29 % al 12 % el tiempo de fonación y de 66.9 a 64.5 dB_{SPL} la intensidad.⁶²

VI. MATERIAL Y MÉTODO

1. DESARROLLO DEL SISTEMA

1.1. Hardware

El hardware que integra el dispositivo desarrollado consta de dos componentes principales. Un acelerómetro, que transforma las vibraciones de las cuerdas vocales transmitidas a través de la piel en una señal eléctrica analógica; y el microcontrolador, que se encarga de convertir esta señal analógica a digital, procesarla, y guardar la señal procesada en un dispositivo de almacenamiento para que pueda ser analizada posteriormente. Además, es necesario un pequeño circuito para adaptar la alimentación suministrada por el microcontrolador al acelerómetro y una pequeña pieza donde posicionar el sensor y poder colgarlo del cuello en la posición óptima.

Para lograr un dispositivo que podamos denominar como vestible, éste tiene que cumplir un requisito fundamental, y es que su tamaño y peso permita llevarlo cómodamente encima. Para ello todos los componentes que lo forman tienen que estar dirigidos a cumplir su función con el menor tamaño y peso posible.

En la Figura 1 y Figura 2, se muestran los dispositivos comerciales actuales y dos en desarrollo que usan dispositivos móviles. De todos ellos el único que podríamos considerar como dispositivo vestible es el Vocalog2. Todos los demás, además de llevar cables, suelen tener un volumen y peso considerable ya comentado en la Tabla 2.

1.1.1. Acelerómetro

El acelerómetro usado es el BU-27135-0000 de Knowles Electronics, EE.UU. Sus dimensiones son reducidas (7.92 mm × 5.59 mm × 2.24 mm), la sensibilidad es unidireccional (solo mide la aceleración en un eje) y la respuesta en frecuencia es constante entre los 20 y 2000 Hz (Figura 3). El acelerómetro tiene tres pines que corresponden a la alimentación, tierra y señal, respectivamente. Es alimentado con 1,5 V (ANEXO IV).

El tamaño reducido, clave para conseguir un dispositivo vestible, permite un correcto posicionamiento entre la parte superior del esternón y la parte inferior de la laringe. Es importante que sea esta zona porque es la más cercana a las cuerdas vocales sin estar directamente sobre la laringe, lo que se traduce en una buena recepción de las vibraciones generadas en las cuerdas vocales. Este reducido tamaño también aportará mayor comodidad.

Para la medición de la intensidad relativa es fundamental que la respuesta del dispositivo sea lineal en el rango de frecuencias medidas. Esto permite correlacionar la misma amplitud con la misma intensidad relativa a pesar de existir una variación en la frecuencia. Significa que la señal del acelerómetro tendrá una proporcionalidad lineal a la intensidad de la oscilación que mide.

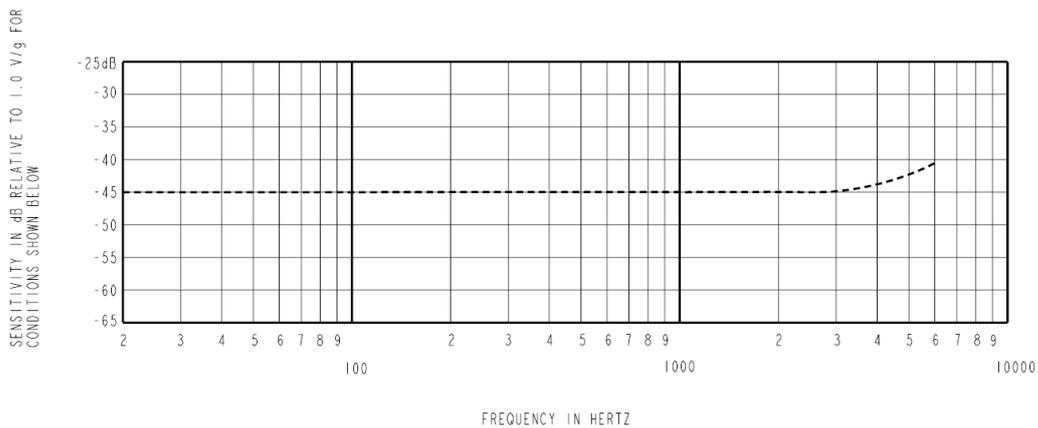


Figura 3: Sensibilidad del acelerómetro BU-27135-0000 con respecto a la frecuencia.

La elección de este acelerómetro frente a un micrófono ha sido por dos motivos fundamentales. El primero, es la relación señal ruido. El acelerómetro no se ve afectado por otras señales más que las vibraciones de la piel con la que está en contacto, sin embargo, el micrófono recoge señales de su alrededor que no son el objeto de medición.⁶³ Además, por este mismo motivo podría registrar incluso voces de otras personas y dar como resultados mediciones incorrectas. El segundo motivo es la privacidad: es menos probable que una persona esté dispuesta a llevar un micrófono durante todo el día registrando su voz y la de los demás. De igual modo, las personas que se encuentren alrededor o tengan algún

tipo de conversación con el usuario del dispositivo pueden no estar dispuestas a que se recoja, aunque el registro de su voz no sea el objetivo del dispositivo.

1.1.2. Microcontrolador

El microcontrolador usado es el MSP430FR5994 (ANEXO V) de Texas Instruments (TI), EE.UU. Las características principales que reúne este modelo por las cuales se ha decidido usarlo para este trabajo son las siguientes:

- Posee un coprocesador de bajo consumo llamado LEA (Low Energy Accelerator) para operaciones matemáticas. Entre estas operaciones se encuentran la transformada rápida de Fourier y el filtrado lineal. Estas dos operaciones, como se explica más adelante, son fundamentales para el cálculo de la frecuencia fundamental de la voz. Este acelerador permite realizar estas operaciones de una manera más rápida y más eficiente energéticamente que si se hubiesen realizado en el procesador principal del microcontrolador (llamado CPU o unidad central de procesamiento, del inglés Central Processing Unit).
- Este microcontrolador se encuentra dentro de la gama de ultra bajo consumo de TI. Se caracteriza por un consumo base reducido y además ofrece modos de bajo consumo adicionales en que distintas partes del microcontrolador se apagan cuando no están en uso en un momento dado y que permiten reducir aún más el consumo de la energía para prolongar la vida de la batería. Esta característica permite reducir el tamaño del dispositivo dado que tiene un impacto sobre el tamaño de la batería necesaria para mantener en funcionamiento el equipo durante un mismo periodo de tiempo.
- La tercera característica importante del dispositivo es la posibilidad de registrar la entrada del conversor analógico-digital al mismo tiempo que se ejecutan otras instrucciones.
- Por último, TI ofrece una tarjeta de desarrollo con este microcontrolador denominada MSP-EXP430FR5994 (Figura 4) y varios ejemplos de uso. Esta tarjeta fue la que se utilizó durante todo el trabajo.

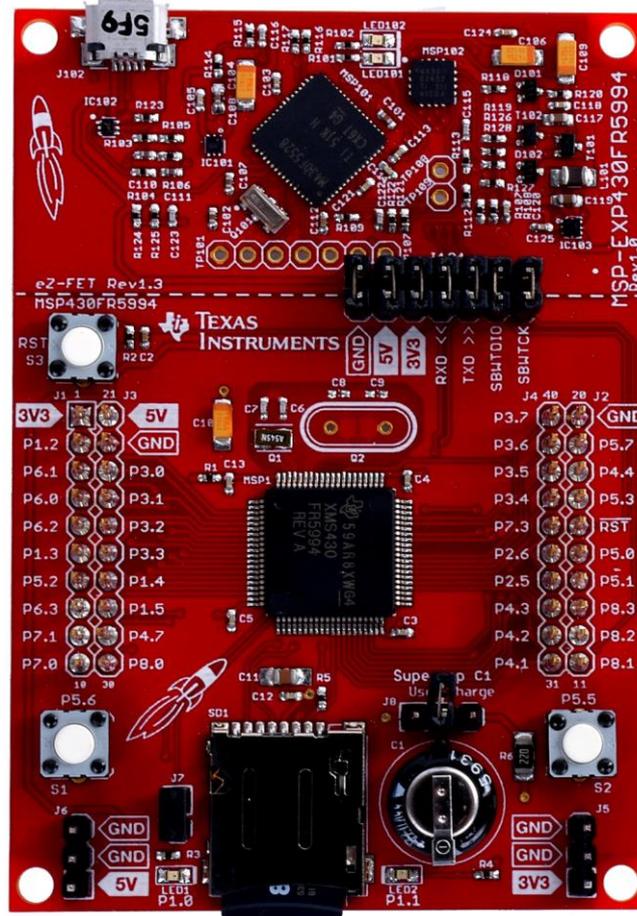


Figura 4: Tarjeta de desarrollo MSP-EXP430FR5994.

Otras características relevantes, pero menos importantes, son:

- Dispone de un reloj interno (RTC) que permitiría incluir información temporal de los registros.
- Dispone de la comunicación SPI para traspasar la información almacenada en la memoria del microcontrolador a una memoria microSD.
- Dispone de acceso directo a memoria (DMA) que permite trasladar datos de un punto a otro de la memoria sin necesidad de ocupar ciclos del procesador.

La tarjeta de desarrollo, posee varias características que fueron útiles para el desarrollo del sistema de monitorización de voz.

- Ranura para tarjetas microSD.
- Circuito integrado para el análisis del consumo del dispositivo.
- Circuito integrado para realizar la depuración del programa o *debugging*.
- Botones y LEDs para interactuar con el sistema y señalar eventos.

1.1.3. Colgante

Para las primeras pruebas del dispositivo, el sensor se aguantaba con la mano y se posicionaba entre la laringe y el esternón. Este método era suficiente para la medición de la frecuencia fundamental durante menos de un minuto. Este método no era válido para la medición de la intensidad puesto que la presión ejercida variaba y por lo tanto la medición de la amplitud de la vibración registrada. El sensor tiene que estar apoyado sobre la piel con la misma presión durante toda la medición para que los datos recogidos puedan ser comparables. Otros trabajos utilizan cintas adhesivas de doble cara para pegar el sensor que utilizan sobre la piel, pero estas molestan y/o acaban despegándose afectando a la medida del sensor^{5,64}. En este trabajo se diseñó un soporte sobre el que se posiciona el sensor y que tiene un sistema de anclaje para una goma elástica que pasa alrededor del cuello y mantiene el sensor en una posición fija (Figura 5). Para ajustar la tensión también se diseñó otra pieza. Esta disposición utilizada para las pruebas de validación realizadas, se acerca mucho a la disposición final que tendría el dispositivo que se mostrará en los resultados. Estas dos piezas fueron diseñadas en el programa de diseño 3D Fusion 360 y fabricados mediante impresión 3D con la impresora Form2 de Formlabs en resina Clear.

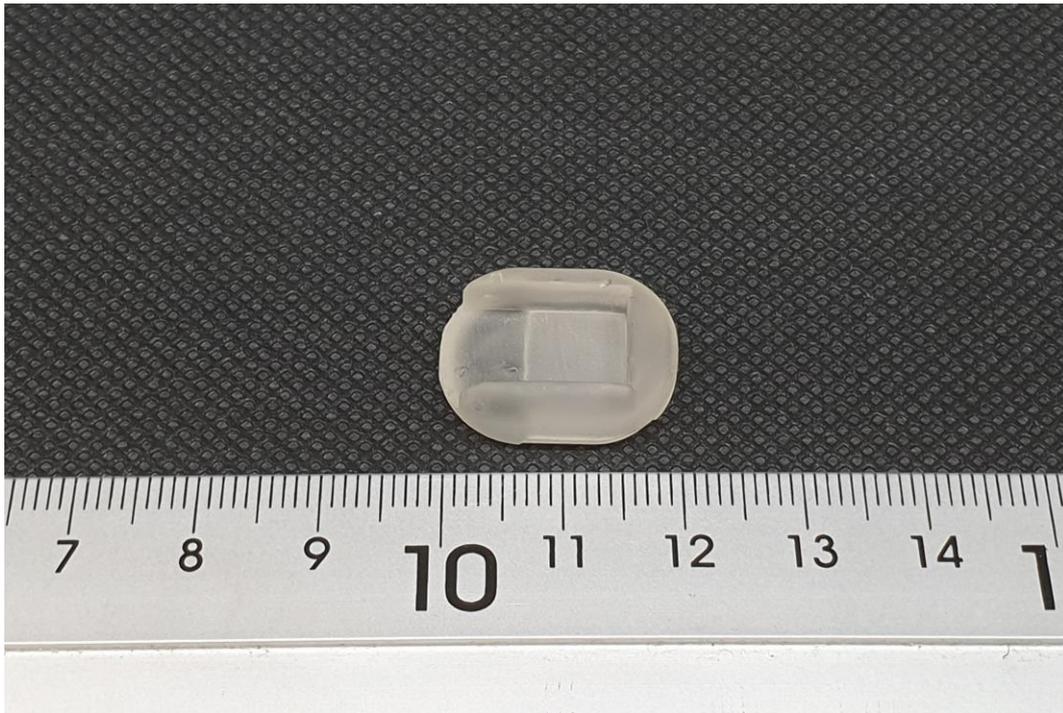


Figura 5: Pieza para posicionar el acelerómetro sobre la región supraesternal (ANEXO VI).

1.1.4. Circuito

Para conectar el acelerómetro a la tarjeta de desarrollo es necesario un pequeño circuito para ajustar la tensión de alimentación que proporciona el microcontrolador a los valores que marca la hoja de datos del sensor.

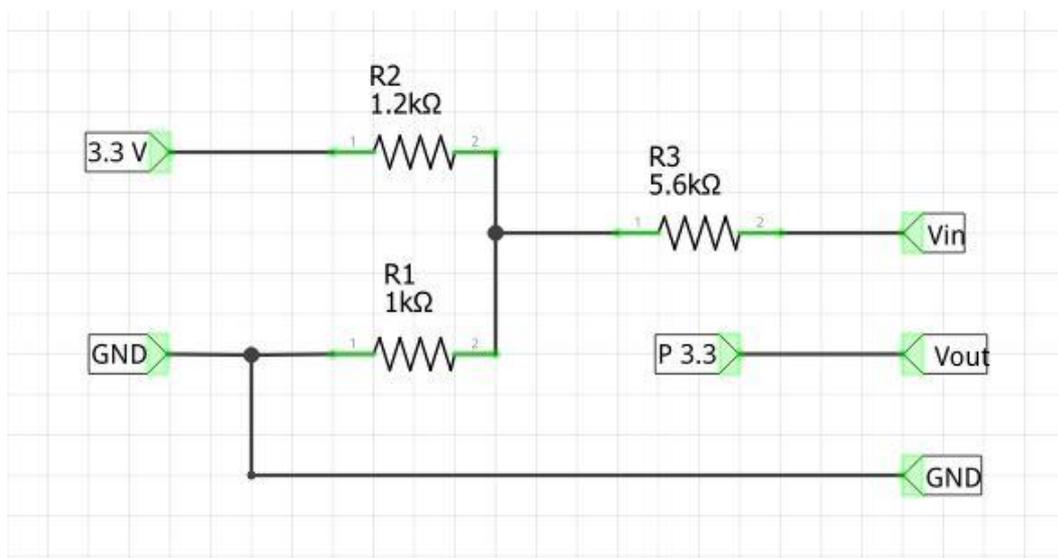


Figura 6: Circuito de alimentación del acelerómetro BU-27135-0000. Vin, Vout y GND son los pines del acelerómetro. P 3.3 es el pin del microcontrolador donde se conecta la salida del acelerómetro.

Para que todo el conjunto (Figura 7) fuera lo más compacto posible se diseñó y fabricó mediante impresión 3D un soporte para el circuito que se pudiera anclar a la placa de desarrollo (ANEXO VII).

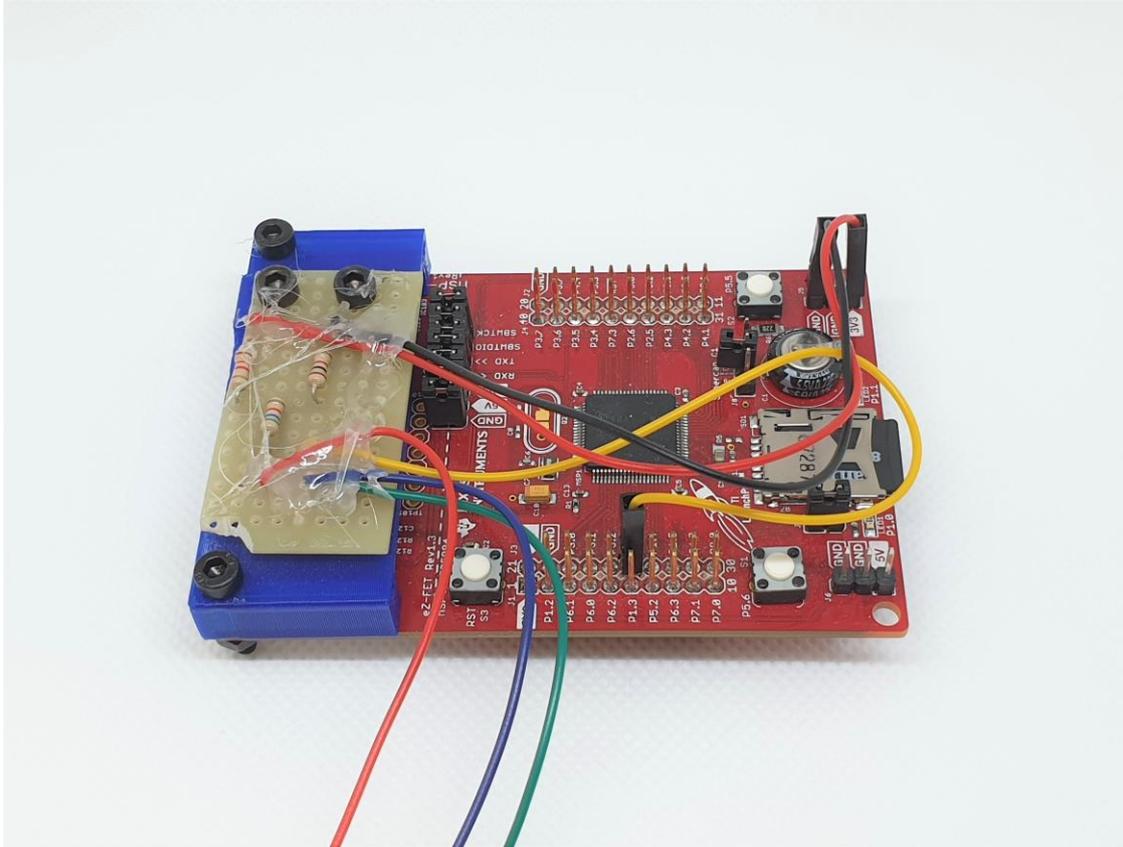


Figura 7: Tarjeta de desarrollo con el circuito para el acelerómetro (Figura 6) acoplado usando la pieza fabricada mediante impresión 3D (color azul).

1.2. Algoritmo

El microcontrolador se programó en lenguaje C++ usando el programa de TI CCS_7.2.0.00013. Para este dispositivo en concreto se usó también la librería DSPLib_1_30_00_02 que contiene las funciones matemáticas que se van a necesitar optimizadas para su ejecución en LEA.

El algoritmo tiene tres partes generales (Figura 8):

- El registro de los valores del sensor. Esto es la lectura de los valores registrados por el sensor, su transformación a formato digital y su almacenamiento en la memoria del microcontrolador.

- El análisis de esos valores. Con el objetivo de calcular los valores de intensidad y frecuencia, deben realizarse operaciones como la transformada de Fourier, filtrado o correlación.
- La grabación de los datos calculados en la memoria microSD.

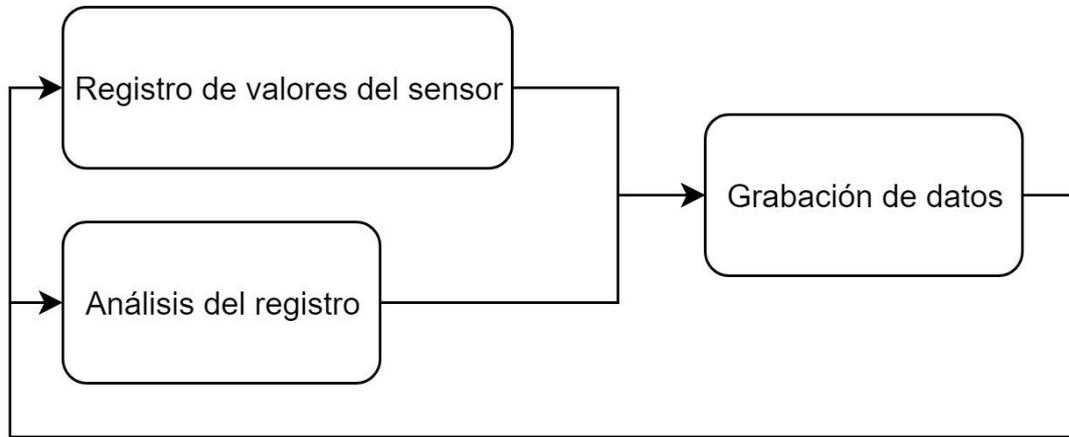


Figura 8: Diagrama de flujo general. Los procesos de registro y análisis se realizan en paralelo.

1.2.1. Registro de los valores del sensor

Las vibraciones generadas por las cuerdas vocales son transmitidas al tejido circundante hasta el acelerómetro. Al tratarse de un sensor analógico, este produce una señal de salida continua que varía en el tiempo y es proporcional al efecto que está midiendo, en este caso las vibraciones de la piel con la que está en contacto.

Esta señal llega al conversor analógico digital (ADC) del microcontrolador. El conversor transforma la señal continua producida por el sensor en una señal digital con la que pueda trabajar el microcontrolador. En el caso del MSP430FR5994, el ADC tiene la resolución de 12 bits. Esto quiere decir que el rango comprometido en los dos valores (máximo y mínimo) se divide en 4096 valores (para una variable sin signo, el valor máximo representado por 12 bits en binario es 1111 1111 1111, que en decimal es $2^{12}-1 = 4095$). Como el rango de medición está entre 0 V y 1.5 V, cada valor discreto tiene una resolución en voltaje de 366 μ V.

El rango normal de frecuencias de la voz hablada se encuentra entre los 100 Hz y los 250 Hz.¹⁰ Según la teoría de Nyquist-Shannon, para digitalizar una señal

análoga la frecuencia de muestreo debe ser como mínimo el doble de la frecuencia máxima que se está midiendo. En nuestro caso si la frecuencia máxima de la señal a medir es de 250 Hz deberíamos utilizar una frecuencia de muestreo de al menos 500 Hz. Sin embargo, para poder asegurar que se cubra todo el rango frecuencial de la voz hablada y posibles desviaciones, se decide ampliar el rango de medición hasta los 1000 Hz. Por consiguiente, la frecuencia mínima de muestro debe ser de 2000 Hz.

Una vez determinada la frecuencia de muestreo, debemos determinar el intervalo de análisis. Este intervalo es el tiempo durante el cual se estarán registrando los valores del acelerómetro antes de ejecutar el análisis para obtener los parámetros buscados (F_0 , I y t). Este intervalo en dispositivos existentes, como se muestra en la Tabla 2, varía desde los 30 ms hasta los 60 segundos. Teniendo este último caso como ejemplo, el dispositivo obtendría un valor de frecuencia y amplitud de la señal registrada cada 60 segundos. De este modo también la unidad mínima para distinguir si hay voz o no, sería de 60 segundos. Si se desea aumentar la resolución temporal debe disminuirse el tamaño del intervalo. Esto tiene una consecuencia directa que es el aumento de la cantidad de los datos recolectados y por lo tanto de la memoria necesaria para almacenarlos.

En nuestro caso el intervalo de análisis es de 250 ms. Este tiempo viene determinado por la frecuencia de muestreo y el número de muestras dentro del intervalo de análisis. El número escogido de muestras por intervalo es de 512 y la frecuencia de muestreo es de 2048 Hz (pasamos de los 2000 Hz mencionados anteriormente a esta frecuencia para que los resultados al calcular la frecuencia sean números enteros puesto que la gran mayoría de parámetros del algoritmo serán potencias de dos).

En este trabajo se escogieron 512 muestras para analizar porque es el valor máximo para que todas las variables necesarias en el programa entren dentro de la capacidad de memoria que tiene LEA para operar. Las variables corresponden con espacios reservados de memoria en los que se guardan y modifican datos. Estas variables, dependiendo de su tipo, ocupan un espacio determinado.

El microcontrolador dispone de una memoria interna denominada SRAM de 8 kB. Este tipo de memoria RAM pierde la información almacenada en ella cuando deja de estar alimentada. Esta memoria RAM se divide en dos bloques de 4 kB (Figura 9), uno destinado a la aplicación y el otro destinado a LEA cuando LEA se encuentra activa, mientras que si no se usara LEA los 8 kB estarían disponibles para la aplicación. Ambos bloques tienen una parte de esos 4 kB destinados a uso interno, por lo que realmente el usuario no dispone de toda esta memoria.

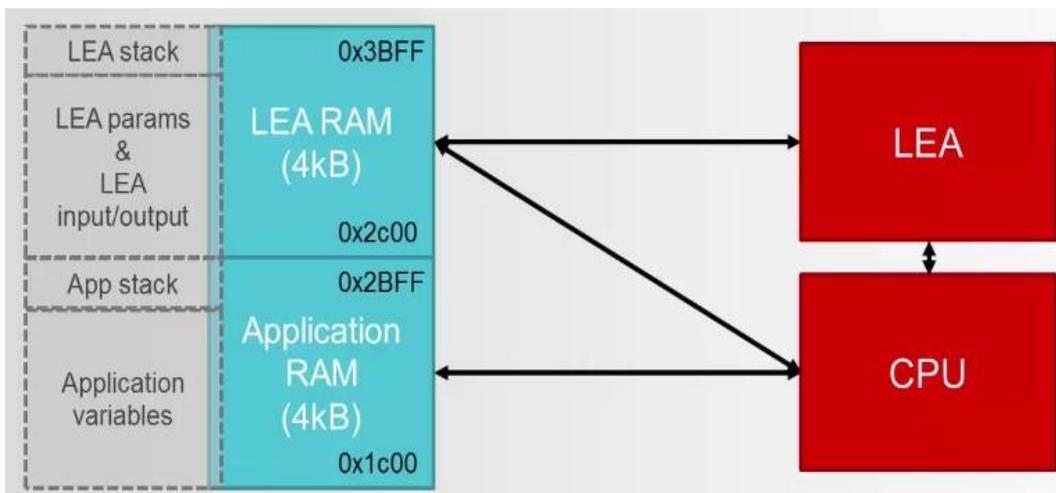


Figura 9: Esquema de memoria RAM del microcontrolador MSP430FR5994.

Con menos de 4 kB disponibles, la gestión de las variables y su posicionamiento dentro de la memoria será un aspecto crítico.

La variable en la que se almacenan los valores registrados por el acelerómetro para ser analizados, a la cual llamaremos `Input` de aquí en adelante, será un vector de tipo `_q15` de 512 valores. Si tenemos 512 valores `_q15` que ocupan 2 bytes cada uno, estamos ocupando un total de 1024 bytes. Esta variable, como posteriormente será utilizada para el cálculo de la transformada de Fourier, es necesario que se encuentre en la región de RAM compartida con LEA.

Ahora bien, si se desea que el registro y el procesamiento de los datos se realicen en paralelo para posibilitar el procesamiento en tiempo real, no es posible guardar directamente los datos recogidos del sensor en la variable `Input` porque esta se está utilizando para grabar los resultados de operaciones

realizadas sobre la variable. Para solventar este problema, se guardan los datos de la señal recibida por el sensor en otro vector. Los datos los se guardan en la parte de la memoria RAM destinada a la CPU. En esta zona de memoria se generan dos variables. Estas variables, a las que a partir de ahora se denominarán `Buffer1` y `Buffer2`, servirán de almacenamiento temporal de las muestras de la señal de acelerómetro. Los valores del acelerómetro comenzarán a guardarse en el `Buffer1`, cuando se completen los 512 valores establecidos, empezarán a guardarse en el `Buffer2`. Mientras se guardan en el `Buffer2`, el contenido del `Buffer1` se copia en la posición de memoria correspondiente a la variable `Input`. Esta técnica se denomina *double buffering* o *ping-pong buffering*. Se realiza esta copia puesto que las variables `Buffer1` y `Buffer2` no se encuentran en una zona de memoria compartida con LEA y no se posicionan las variables `Buffer1` y `Buffer2` dentro de la región de memoria compartida con LEA porque con la poca memoria que se puede compartir con LEA nos quedaríamos sin espacio para el resto de variables que sí deben estar situadas en esta región de memoria. Una vez el `Buffer2` se ha llenado y el `Buffer1` ha sido copiado en la variable `Input`, volverán a guardarse los valores en el `Buffer1`. En el tiempo que tarda en llenarse el *buffer* correspondiente (250 ms) se copian y procesan los datos registrados en el otro. Este proceso cíclico no se detiene hasta que el dispositivo se apaga o se encuentra grabando en la memoria microSD los datos ya procesados.

Para que el almacenamiento de los datos recibidos por el acelerómetro y digitalizados por el ADC sea más rápido y eficiente, se utiliza una de las características de este microcontrolador, el acceso directo a memoria (DMA, del inglés Direct Memory Access). Este controlador interno permite la transferencia de paquetes de datos de un punto a otro de la memoria sin necesidad de intervención de la CPU. De este modo, el guardado de los datos en el *buffer* no consume tiempo de CPU. Esto implica un menor uso de energía y que la CPU esté disponible para otras tareas.

Por lo tanto y a modo de resumen, las vibraciones generadas por las cuerdas vocales se transmiten a través de la piel hasta el acelerómetro. Éste transforma

estas vibraciones en una variación de potencial eléctrico que el conversor analógico digital transforma en una señal discreta. La señal se guarda en uno de los *buffers*. Cuando este se completa, comienza a guardarse en el otro. Mientras se está guardando en el otro *buffer*, los datos del que acaba de llenarse se transfieren a la variable `Input` que se encuentra dentro de la región de memoria de la RAM que tiene disponible LEA para su procesado.

1.2.2. Procesado de los datos

Una vez tenemos los datos registrados por el acelerómetro copiados en la variable `Input`, que se encuentra dentro de la memoria a la que tiene acceso LEA, podemos comenzar a procesar los datos para obtener los valores de intensidad relativa y frecuencia fundamental correspondientes en el intervalo analizado.

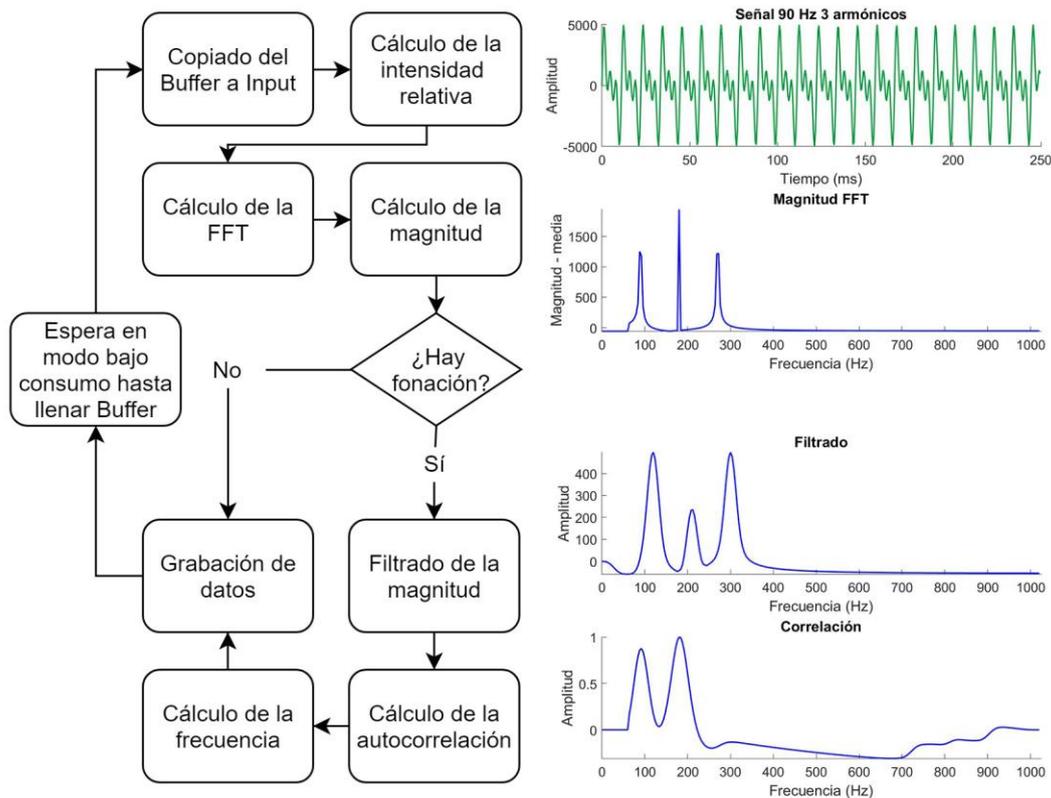


Figura 10: Diagrama de flujo del procesamiento de los datos a la izquierda. A la derecha, representación de la señal y los resultados de los pasos cálculo de la magnitud, filtrado de la magnitud y cálculo de la autocorrelación.

En la Figura 10 se muestra el diagrama de flujo del procesado. Una vez que se hayan copiado los datos del *buffer* a la variable `Input` el primer paso es el cálculo de la intensidad. Esto se calcula directamente a partir de la señal del acelerómetro (en esta figura una señal de 90 Hz y tres armónicos). Posteriormente se calcula la FFT y su magnitud a la que se le resta la media. En este punto se observan los armónicos de la señal, y se determina si la señal corresponde a voz o no. En el caso de no corresponder a voz, directamente se guarda intensidad relativa y frecuencia fundamental 0. En caso de determinarse que corresponde a voz, la señal de la magnitud de FFT es filtrada y de ésta se calcula la correlación (autocorrelación dado que se correlaciona consigo misma). Tras este procesado, se calcula la frecuencia fundamental y se guarda junto a la intensidad relativa calculada anteriormente. Por último, se espera al llenado del *buffer* en modo de bajo consumo para disminuir el consumo energético.

1.2.2.1. Cálculo de la intensidad relativa

Con los valores de la señal obtenida del acelerómetro almacenados en la variable `Input`, el primer paso del procesado es el cálculo de la intensidad relativa. Cálculos posteriores se realizarán utilizando la variable `Input` como argumento de entrada de funciones cuyo resultado será guardado en esta misma variable `Input`. Esta sobrescritura ocasiona la pérdida de los datos de la señal obtenida del acelerómetro, pero es necesario para no ocupar más espacio en la memoria.

El cálculo de la intensidad relativa se realiza mediante la media cuadrática (RMS por las siglas en inglés de Root Mean Square)

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}, \quad (2)$$

siendo N el número de valores del intervalo analizado.

Este valor es calculado en subintervalos dentro del intervalo de medición. El intervalo de análisis tiene la duración de 250 ms y comprende 512 muestras.

Este intervalo se divide en 8 subintervalos en los que se realiza el cálculo de la media cuadrática, cada uno de 64 muestras y 31.25 ms de duración. Esta división en subintervalos se realiza por dos motivos:

- El primer motivo es el aumento de la resolución temporal en el cálculo de la intensidad relativa. Debido a que la duración del intervalo se ha elegido teniendo en cuenta el cálculo de la frecuencia fundamental, éste no es un intervalo óptimo para estimar la intensidad relativa que varía dentro del mismo. Por lo tanto, tomar un valor de intensidad relativa cada 250 ms no es representativo y se opta por el uso de subintervalos para reducir el error cometido.
- El segundo motivo es referente al tiempo de fonación. Como durante los 250 ms de registro la intensidad relativa puede variar, esta variación también puede conllevar que no todo el registro de 250 ms sea voz. Con los subintervalos pasamos de una resolución temporal de 250 ms a una resolución de 31.25 ms a la hora de discriminar entre voz y no voz.

Por lo tanto, por cada intervalo de 250 ms obtenemos 8 valores de intensidad relativa correspondientes al cálculo de la media cuadrática de cada uno de los 8 subintervalos.

1.2.2.2. Cálculo de la FFT

Las vibraciones producidas por las cuerdas vocales generan una señal oscilatoria. Para analizar una señal oscilatoria, una de las herramientas matemáticas más útiles es el cálculo de la transformada discreta de Fourier (DFT). Para el cálculo de la DFT se utiliza un conjunto de operaciones reunidas en un operador que se denomina transformada rápida de Fourier (FFT) cuyo resultado es la DFT pero el tiempo de computación es mucho menor. Esta operación transforma la señal del dominio del tiempo al dominio de la frecuencia obteniendo la descomposición de la señal en senos armónicamente relacionados y cada uno de ellos contribuye con una

frecuencia fundamental. El resultado de la FFT se denota en números complejos. De aquí en adelante se hará referencia a la DFT como FFT sin distinguir entre operación y resultado de la operación para facilitar la lectura.

Cuando se calcula la magnitud de FFT en un intervalo que contiene fonación, obtendremos como resultado una gráfica con una serie de picos. Cada uno de estos picos es un armónico, y el primer pico es la frecuencia fundamental.

En la Figura 11 se muestra una señal compuesta por tres señales sinusoidales de distintas frecuencias. Debajo de todas ellas está representada la magnitud de la FFT de la señal compuesta inicial y se observan los 3 picos correspondientes a las 3 señales que la componen.

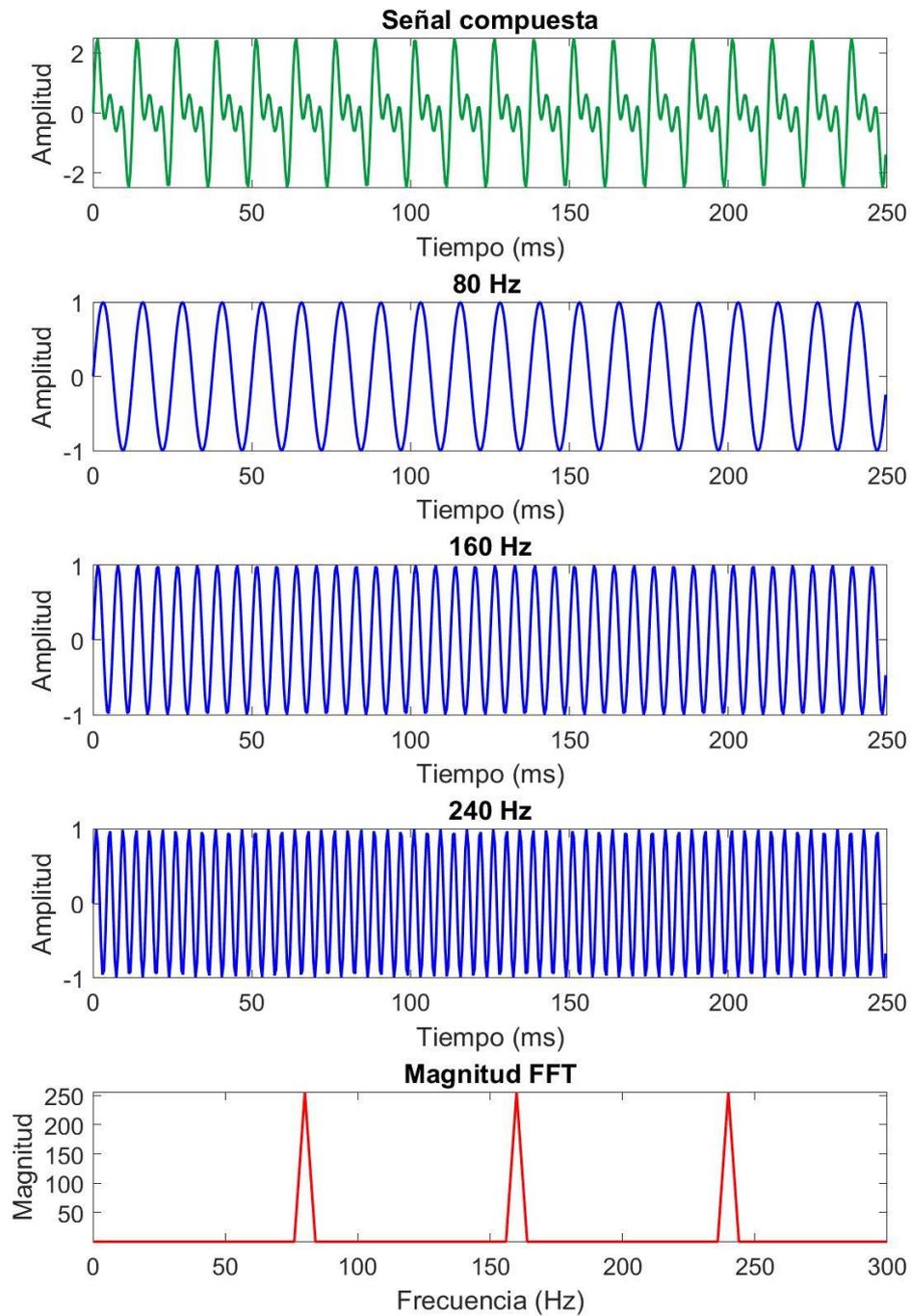


Figura 11: Descomposición de una señal compuesta en las tres señales que la forman y la magnitud de la FFT de la señal compuesta entre 0 y 300 Hz.

Uno de los motivos por los que se escogió el microcontrolador MSP430FR5994 fue porque integraba un coprocesador de bajo consumo, llamado LEA, para realizar el cálculo de la FFT de manera muy rápida y eficiente energéticamente. Pero estas ventajas conllevan ciertas limitaciones. Con LEA solo se pueden realizar cálculos de la FFT de hasta 1024 puntos. Esto quiere decir que el vector que contiene la señal y que va a ser analizado usando la FFT debe tener como máximo 1024 valores. Para calcular la resolución frecuencial se divide la frecuencia de muestreo entre el número de muestras. Con la frecuencia de muestreo escogida de 2048 Hz y el máximo de 1024 puntos admitidos para que la FFT pueda ser calculada por LEA, obtendríamos una resolución frecuencial de 2 Hz. Sin embargo, como hemos mencionado antes, el número de muestras que escogemos tomar es de 512 para no exceder el límite de memoria disponible para LEA. Por ello la resolución frecuencial será de 4 Hz.

Durante la programación del microcontrolador para el cálculo de la FFT surgió un problema importante. Las variables de entrada y salida de la función `msp_fft_fixed_q15` para calcular la FFT, tienen que tener un alineamiento determinado. Esta función para el cálculo de la DFT se encuentra en la librería DSPLib. El alineamiento quiere decir que no pueden estar en cualquier posición de la memoria, sino que (en el caso de la FFT) la posición de comienzo de la variable debe ser múltiplo del doble de la longitud de la variable en cuestión. Esto quiere decir que, si la variable tiene una longitud de 0x200, la posición de memoria debe ser divisible por 0x400. El fabricante del microcontrolador proporciona la librería DSPLib que contiene una función específica `DSPLIB_DATA(var, align)`, siendo `var` la variable que se desea posicionar en memoria y `align` el alineamiento que debe tener. Esta función posiciona `var` con el alineamiento `align` en la posición de memoria que cumple la norma descrita antes. Resulta que esta función no funciona correctamente, desaprovechando gran parte de la memoria. En nuestro caso, en el que necesitamos aprovechar al máximo la memoria disponible para LEA, es un

grave problema. El fabricante admitió un problema en su software sin proporcionar una solución. Se consiguió resolver el problema indicando manualmente cuál debía ser la posición en memoria de todas las variables situadas en la sección de memoria compartida con LEA.

Una vez posicionada en memoria la variable `Input` en el lugar correspondiente, se utiliza la función `msp_fft_fixed_q15` para calcular la FFT. El resultado de la función se guarda de nuevo en la variable `Input`. Este resultado está compuesto por 256 números complejos. En las posiciones pares de la variable se almacena la parte real y en la impar la parte imaginaria.

1.2.2.3. Cálculo de la magnitud de la FFT

Para analizar el resultado de la FFT debemos calcular la magnitud de ésta. Como se menciona en el apartado anterior, el resultado de la FFT es un conjunto de números complejos. Primero debemos declarar una nueva variable para almacenar el resultado del cálculo de la magnitud. Esta debe ser del tipo `int16_t` con una longitud de 256 valores (la mitad de la longitud de la variable `Input`). Es la mitad de la longitud de la variable `Input` porque por cada número complejo que está compuesto de dos partes (real e imaginaria) obtenemos un número real. Esta variable en la que se guardarán los resultados de las distintas operaciones la denominaremos `RdO`.

La operación para el cálculo de la magnitud de la FFT viene determinada por la fórmula:

$$m = \sqrt{re^2 + im^2}. \quad (3)$$

Siendo *re* la parte real, e *im* la parte imaginaria del número complejo y *m* el resultado.

Realizar esta operación para calcular los 256 valores de la variable `RdO` tiene un tiempo de computación elevado, concretamente 300 ms. Si tenemos en

cuenta que el tiempo del intervalo son 250 ms, no es posible realizar esta operación en tiempo real.

Del resultado de esta operación obtendremos una gráfica en la que los picos representan los armónicos de la señal y en el que el primero de ellos es la frecuencia fundamental de la señal Figura 11. Por lo tanto, lo que nos interesa es la posición de dichos picos y no su magnitud. Por ello proponemos el uso de la siguiente ecuación para el cálculo de la magnitud:

$$m = |re| + |im|. \quad (4)$$

Para comparar cuál es la diferencia entre ambas operamos elevando al cuadrado:

$$m^2 = (\sqrt{re^2 + im^2})^2 = re^2 + im^2, \quad (5)$$

$$m^2 = (|re| + |im|)^2 = |re|^2 + |im|^2 + 2|re||im| = re^2 + im^2 + 2|re||im|. \quad (6)$$

Observando ambas ecuaciones la diferencia es $2|re||im|$. Este valor solo puede ser positivo, por lo tanto, la diferencia de usar la ecuación (3) o la (4) es la amplitud del valor y siempre una sería mayor a la calculada con la ecuación (4) que con la (3).

Esta otra manera de calcular la magnitud, a pesar de incurrir en un error de escala y no ser propiamente la magnitud, mantiene la posición de los picos de la magnitud de FFT. De aquí en adelante consideraremos esta aproximación como la magnitud. De este modo, el tiempo necesario para calcular la magnitud se reduce a 30 ms, una décima parte. Esta reducción del tiempo de computación permite operar el algoritmo en tiempo real.

En este paso aprovechamos para eliminar el ruido de frecuencias bajas. Si consideramos que la frecuencia mínima de la voz es de 72 Hz, podemos decir que las magnitudes que representan el intervalo 0-72 Hz deben ser 0 puesto que no puede haber señal correspondiente a esas frecuencias. Como la resolución frecuencial con la que estamos trabajando es de 4 Hz, las

posiciones de la variable RdO que representan el intervalo 0-72 Hz son 0-17 (el valor de 72 no lo incluimos). Por lo tanto, cambiamos los valores que tienen estas posiciones de la variable RdO por 0.

Por último, y dado que vamos a seguir utilizando la variable RdO para otros cálculos sobrescribiendo y perdiendo los datos que en este paso se almacenan, buscamos un valor de posición que sirva de referencia para calcular la frecuencia fundamental. Intuitivamente, diríamos que la posición correspondiente al valor máximo de la variable RdO correspondería con la frecuencia fundamental, asumiendo que el primer pico es el de mayor magnitud. Por lo que guardamos esta posición en la variable Pos_Fft y el valor de RdO en esa posición en la variable Val_Fft . Aunque en muchos casos es así, no siempre el primer armónico es el que contiene más energía y por lo tanto el que tiene un mayor valor. Para salvar este inconveniente, y sabiendo que usaremos este valor únicamente en casos en los que la autocorrelación (véase el apartado 1.2.2.5 de este capítulo) no es precisa, analizaremos si el máximo de RdO es o no es el primer armónico. Experimentalmente se ha observado que, si hay algún armónico anterior al que tiene el valor máximo, siempre tiene un valor mayor a $1/3$ del máximo. Analizamos dos supuestos, el primero en el que no sea el primer armónico que contenga más energía sino el segundo, es decir que en la posición $Pos_Fft/2$ debe encontrarse otro pico correspondiente a un armónico. Como las posiciones son números enteros y al estar trabajando con variables enteras, puede que el pico no se encuentre justo en posición resultado de la operación $Pos_Fft/2$, ampliamos la búsqueda a $Pos_Fft/2 \pm 1$. Entre los tres valores, $Pos_Fft/2 - 1$, $Pos_Fft/2$ y $Pos_Fft/2 + 1$ se selecciona el máximo de los tres, y si el valor de RdO en esa posición es mayor a $1/3$ del valor máximo de RdO , se almacena dicha posición en la variable Pos_Fft y el valor de la magnitud en esa posición en la variable Val_Fft . En el caso que Pos_Fft no sea divisible por 2 ó 3, al tratarse de una variable entera y el resultado ser también entero se redondea el resultado de la división. El segundo supuesto sería que no se tratase ni del primer armónico ni del

segundo, sino del tercero. Para comprobar si esto es así y corregirlo, se sigue el mismo procedimiento que en el primer supuesto, pero esta vez se divide la posición `Pos_Fft` entre tres.

En este punto, experimentalmente y analizando varios conjuntos de datos, se puede determinar que el valor umbral para distinguir si un intervalo contiene voz o no es 45 unidades (Figura 12). De este modo, si se supera este umbral, se considera que en el intervalo de análisis se ha registrado voz y por lo tanto se continúa con el procesamiento de los datos. En caso contrario, se considera que en el intervalo no se ha registrado voz y por lo tanto se asigna una intensidad relativa y frecuencia fundamental de 0 al intervalo y, por consiguiente, a cada uno de los subintervalos que lo forman. En esta circunstancia, al no continuar con el resto de cálculos, permitimos que el procesador entre en un modo de ahorro de energía hasta que se termine de llenar el *buffer* y se envíen los datos del nuevo intervalo a la variable `Input` para su procesado. En modo de ahorro de energía el consumo del dispositivo se reduce. Teniendo en cuenta que la gran parte del tiempo que el dispositivo va a estar en funcionamiento es tiempo en el que no se va a registrar voz, esto tiene un impacto importante sobre el consumo del dispositivo y por lo tanto en la autonomía. Entre el 20-50 % del tiempo de una lectura o conversación se corresponde a pausas.⁶⁵

1.2.2.4. Filtrado de la magnitud de la FFT

El resultado del cálculo de la magnitud de la FFT sobre la señal obtenida con los parámetros comentados nos devuelve un resultado abrupto, con muchos picos como se muestra en la Figura 12. Esto se debe a dos factores: la señal analizada está compuesta de muchas otras sinusoidales con diferentes frecuencias fundamentales; y el segundo factor está relacionado con los parámetros de medida y procesado. El cálculo de la frecuencia fundamental se realizará mediante el método de correlación para localizar la distancia entre armónicos. Para simplificar la búsqueda de picos es preferible una señal con pocos picos. Como se muestra en la Figura 13, la correlación

calculada a partir de la magnitud de FFT es más abrupta que la calculada usando la magnitud FFT filtrada.

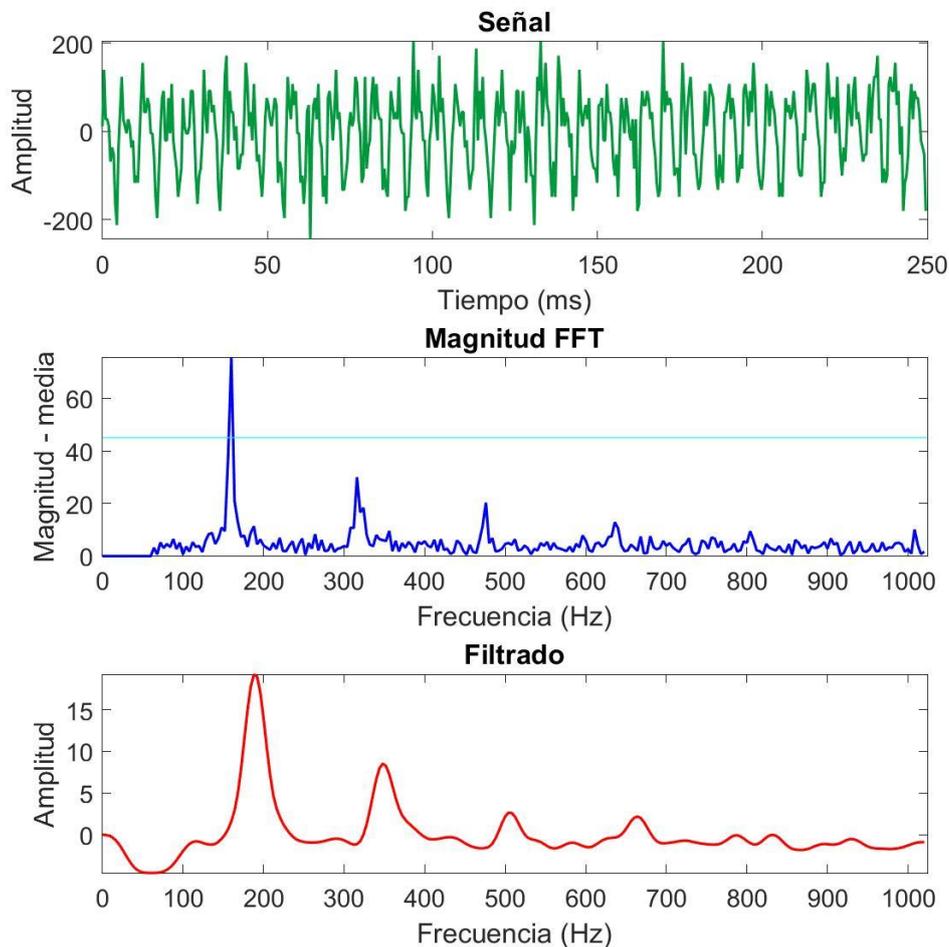


Figura 12: Señal del acelerómetro, cálculo de la magnitud de la FFT y filtrado. En la magnitud de FFT se muestra en cyan el límite para considerar si hay fonación o no.

Por ello se opta por filtrar el resultado obtenido de la magnitud de la FFT usando como filtro una ventana de Hamming de orden 16. Esto suaviza la magnitud y se pueden observar los picos referentes a los armónicos. Además, mejora el resultado del cálculo de la correlación puesto que se elimina la posibilidad de encontrar correlación entre picos que se no corresponden con los armónicos de la señal. El orden elegido para la ventana debe ser potencia de 2 según los parámetros admitidos por la función `msp_fir_q15` de la librería DSPLib. Los coeficientes del filtro fueron calculados en Matlab usando la función `fir1(16, 0.1)` y se tomaron los

primeros 5 decimales que es la resolución de la representación `_q15`. El primer argumento de `fir1` es el tamaño del filtro. Como se ha mencionado, debemos trabajar con potencias de 2. El tamaño escogido es 16 puesto que es el mínimo tamaño que elimina los picos no correspondientes a armónicos.

Al realizar cálculos de multiplicaciones y sumas, que son las que forman la operación de filtrado, si los multiplicandos tienen valores elevados, es posible que se supere el valor máximo que se puede representar con el tipo de variable usado para representar esos multiplicandos. Para evitar esto, la propia función de la librería `DSP_Lib msp_fir_q15` obliga a introducir los valores del filtro en el formato `_q15`. Los valores del filtro se convierten a `_q15` usando una macro destinada a ello `_Q15()`. Por lo tanto, lo que hacemos es guardar los valores obtenidos en Matlab transformados por la función `_Q15()` en la variable `Fil_Coef` que es de tipo `_q15`. Además de ello, para reducir el valor de los resultados de las operaciones, restamos la media de la señal a toda ella antes de ejecutar la función de filtrado. Por último, la función de filtrado necesita una variable reservada para sus cálculos. Esta variable, de tipo `_q15`, se genera en la región de memoria de RAM destinada a LEA con un tamaño del doble de la variable `Fil_Coef` y que denominaremos `Buff_Fil_Coef`.

Los argumentos necesarios por la función (`RdO`, `Fil_Coef` y `Buff_Fil_Coef`) deben estar también dentro de la región de memoria destinada a LEA para que ésta pueda acelerar el cálculo. Para reducir el uso de la memoria se utiliza la variable `RdO` como variable de salida (además de ser la variable de entrada) sobrescribiendo los valores de entrada.

1.2.2.5. Cálculo de la autocorrelación

Calcular la frecuencia usando los resultados de la magnitud de la FFT de una señal sinusoidal pura periódica puede parecer fácil si identificamos el primer pico. La señal registrada por el acelerómetro es una señal formada por un conjunto de frecuencias que a su vez pueden variar dentro del propio

intervalo. Si a esto le sumamos las limitaciones del dispositivo a la hora de registrar y calcular la FFT, resulta que el cálculo de la frecuencia fundamental a partir de la magnitud de la FFT no es sencillo. Podría caerse en el error de decir que la posición del valor máximo de la FFT corresponde al primer pico de la misma y por lo tanto a la frecuencia fundamental. Pero esto es erróneo puesto que puede darse que el máximo de la magnitud corresponda a otro armónico.

Los armónicos están equidistantes entre sí, y esa distancia es igual a la frecuencia fundamental. Esto ocurre en caso de tener una señal ideal, en nuestro caso los fonemas sonoros son cuasi-periódicos y el análisis, aunque aproximado, sigue siendo válido. Por lo tanto, si se calcula la autocorrelación del resultado de la magnitud de la FFT obtenemos una señal cuyo primer pico corresponde con la distancia entre los armónicos. Como se ha mencionado antes, la FFT calculada es abrupta, por lo que la autocorrelación puede encontrar equidistancia entre otros picos que no corresponden con los armónicos y dar un resultado erróneo. Esto no pasa con la señal filtrada.

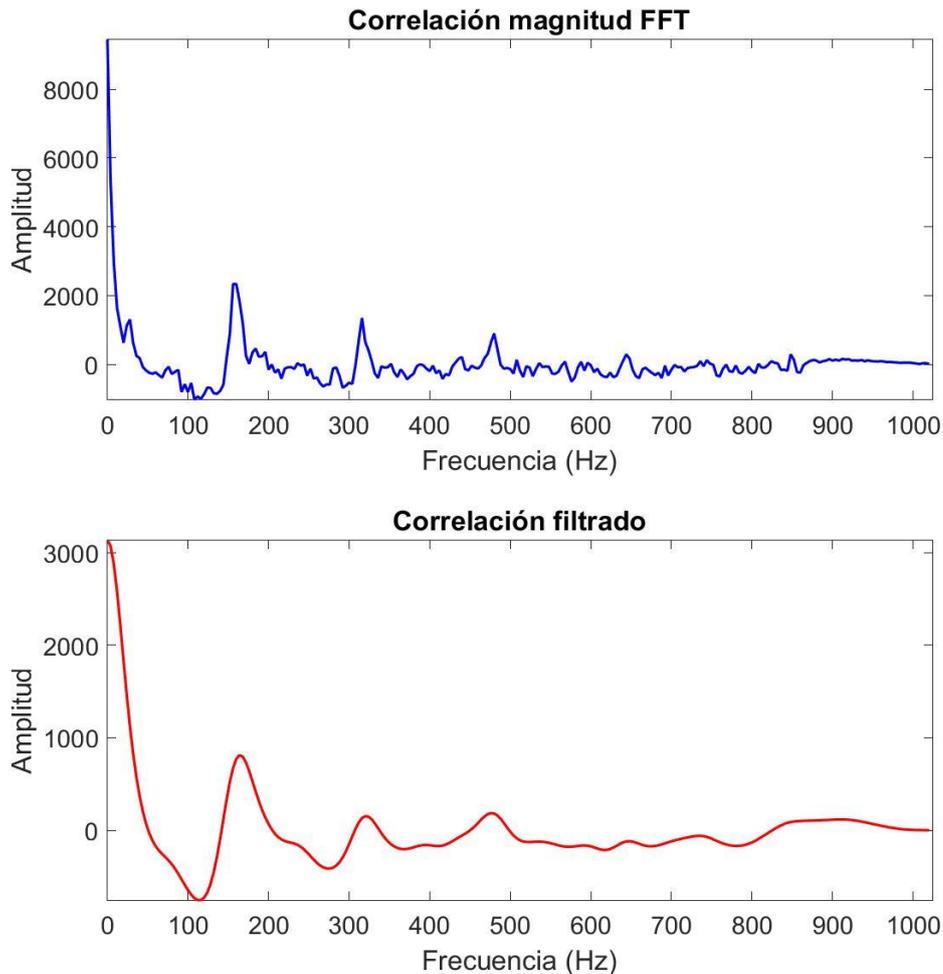


Figura 13: Comparación entre la señal de la correlación a partir de la magnitud de FFT y la correlación calculada a partir del filtrado.

La autocorrelación se puede calcular filtrando el vector de la señal con el mismo vector invertido. Este es el procedimiento que hemos seguido puesto que podemos usar la misma función de filtrado que hemos comentado antes (`msp_fir_q15`) y así, realizar el cálculo utilizando LEA aprovechando su velocidad y bajo consumo. Dentro de los argumentos de la función `msp_fir_q15`, el filtro es la señal invertida en el tiempo que guardamos en la variable `Fil_Coef_Corr` de tipo `_q15`. La variable reservada para cálculos internos será `Buff_Fil_Coef_Corr` de tipo `_q15` con tamaño del doble de `Fil_Coef_Corr`. Sin embargo, el tiempo necesario para convertir los valores de `Buff_Fil_Coef_Corr` a formato `_q15` usando la macro `Q15()` excede los 250ms de los que disponemos para realizar todo el

procesado de la señal. La función disponible en la librería está preparada para utilizar este tipo `_q15` para la ventana y si introducimos una variable tipo `int16_t` el resultado no está correctamente escalado. Para conseguir un resultado que aproveche el rango ofrecido por la variable `int16_t` (- 32768, 32767) debemos escalar la variable `Buff_Fil_Coef_Corr` por el factor

$$E = \sqrt{\frac{3276800}{RdO(255)}} \quad (7)$$

obtenido empíricamente. Por lo tanto, para realizar el cálculo de la autocorrelación el proceso consta de los siguientes pasos:

- Se calcula la autocorrelación sin el escalado.
- Se calcula el factor E de escalado utilizando los valores obtenidos de la correlación.
- Como hemos sobrescrito los valores de `RdO`, escalamos los valores de los `Fil_Coef_Corr`.
- Se copia la variable `Fil_Coef_Corr` invertida en `RdO`.
- Se ejecuta de nuevo la función `mSP_fir_q15` para calcular la autocorrelación. Es el mismo paso que la primera autocorrelación solo que los valores ahora están escalados.

De este modo se obtiene como resultado de la autocorrelación una señal cuyo máximo es el valor máximo de la variable `int16_t` y donde podemos observar cómo el primer pico se corresponde con la distancia entre los armónicos de la señal medida.

Por último, se cambian los primeros 17 valores que contiene la variable `RdO` donde se ha guardado el resultado de la autocorrelación por 0. Estos valores hacen referencia a valores de frecuencia entre 0 y 72 Hz, valores que no se consideran voz.

1.2.2.6. Cálculo de la frecuencia fundamental de la voz

Los dispositivos existentes obtienen la frecuencia fundamental de la voz usando o la transformada de Fourier o la autocorrelación⁴⁷. Sin embargo, nosotros no podemos usar solamente uno de estos sistemas. En nuestro caso, un algoritmo de búsqueda de picos basado en la FFT falla cuando nos encontramos en intensidades de la señal de la voz bajas y frecuencias bajas, donde la intensidad de alguno de los armónicos se puede confundir con los valores correspondientes al ruido. Para obtener resultados correctos mediante la autocorrelación, es necesario que los armónicos estén dentro del rango de la FFT. En nuestro caso la FFT abarca desde los 0 Hz hasta los 1024 Hz, por lo que a partir de los 340 Hz solo tendremos dos armónicos de la frecuencia fundamental y a partir de 512 Hz solo tendremos uno que es cuando el método empezará a fallar.

Afortunadamente, en el rango de frecuencias en que un método no funciona, el otro sí lo hace y viceversa. Por ello, nuestra propuesta es una combinación de ambos. Primero nos fijamos en la autocorrelación, pues abarca el rango de frecuencias más habituales de la voz y es más robusta en frecuencias bajas (<480 Hz) e intervalos que contienen señal con múltiples frecuencias. Para calcular la frecuencia a partir de la autocorrelación trabajamos con dos supuestos: que el primer valor de R_{dO} distinto de cero sea positivo o negativo (Figura 14). Si es positivo, analizamos si los valores aumentan o disminuyen a partir de este valor. En el caso en que los valores aumenten, buscamos el primer máximo y guardamos su posición en la variable `Pos_Max_Corr`. Si por lo contrario decrece, buscamos la posición del primer valor que sea menor a cero. Conocida esta posición seguimos buscando hasta encontrar la posición en la que el valor vuelve a ser mayor de cero, y a partir de ese punto buscamos el primer máximo local y guardamos su posición en la variable `Pos_Max_Corr`. Si por el contrario el primer valor distinto de R_{dO} es negativo, recorreremos la variable en busca de la posición en la que el valor correspondiente es mayor de cero y desde esta

posición buscamos el primer máximo local para guardar su posición en la variable `Pos_Max_Corr`.

Para el cálculo de la F_0 se comprueban dos condiciones. La primera es si la posición del máximo y el valor guardado en la variable `Pos_Max_Corr` coinciden y `Pos_Max_Corr` sea menor que `Pos_Fft+10`, y la segunda condición es que el valor de `Pos_Fft` es menor de 120 (correspondiente a la frecuencia 480 Hz) y el valor contenido en la variable `Val_Fft` es menor de 100. Cumpliéndose la primera y no la segunda obtenemos la frecuencia fundamental multiplicando el valor de la variable `Pos_Max_Corr+1` por 4 Hz (la resolución frecuencial). El valor de corte seleccionado para la variable `Val_Fft` se eligió de manera experimental tras analizar cientos de intervalos en todo el rango de frecuencias e intensidades. Como se mostrará más adelante, es necesario sumar 1 debido al retraso de una muestra en el vector de correlación que ocasiona el cálculo de la correlación utilizando la función de filtrado. En el resto de las circunstancias, es probable que no existan suficientes armónicos para el cálculo por autocorrelación o que estemos ante una señal cuya frecuencia fundamental supere los 480 Hz. En este caso se toma la variable `Pos_Fft` que, de igual modo, multiplicamos por 4 Hz para calcular la frecuencia fundamental.

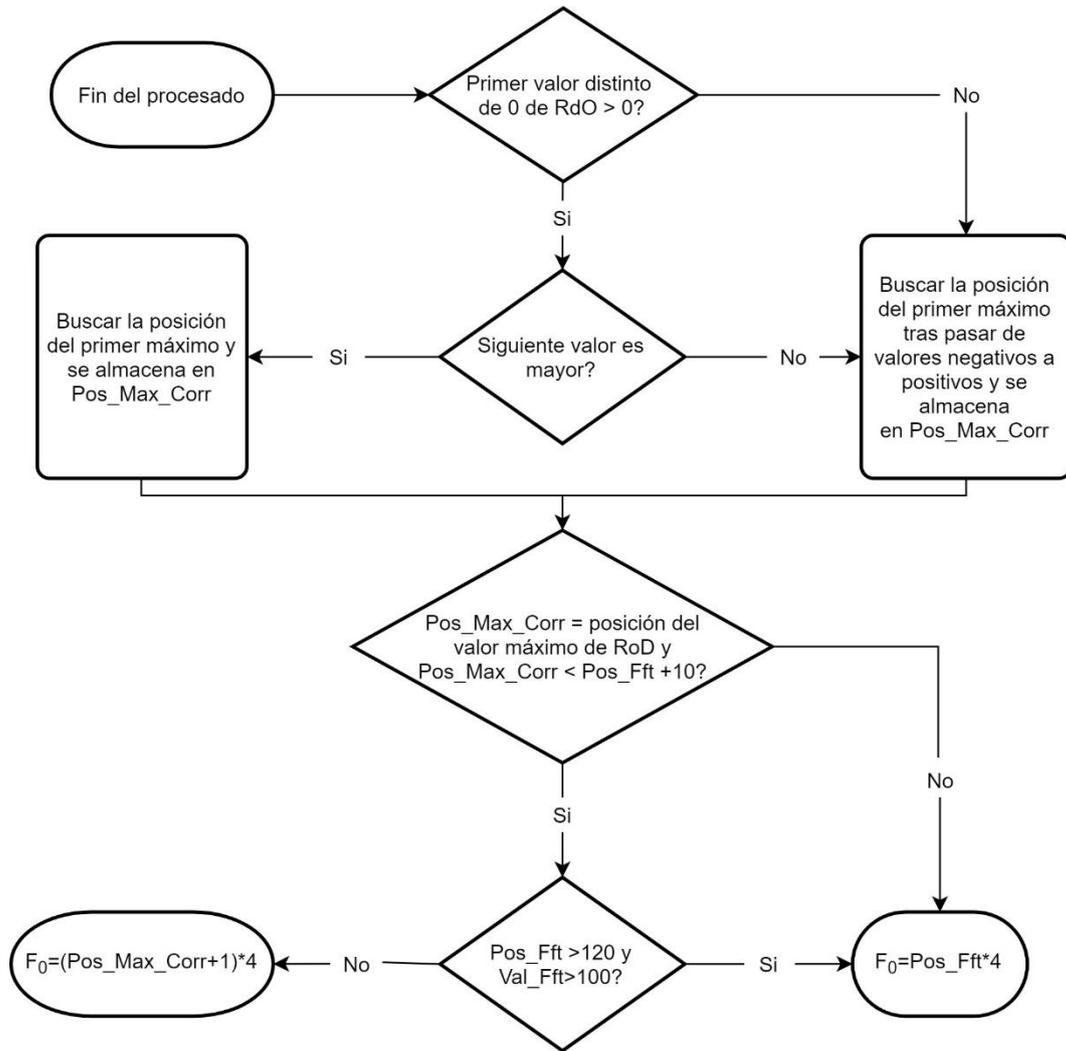


Figura 14: Diagrama del algoritmo para el cálculo de la frecuencia fundamental.

1.2.2.7. Grabación de datos

Cada intervalo de medición de 250 ms contiene 8 subintervalos de 31.25 ms, por lo que se generan 16 valores (8 correspondientes a frecuencia y 8 correspondientes a intensidad relativa). Los valores de la frecuencia se almacenan en un variable que llamaremos DF y los referentes a la intensidad relativa en la variable DI . Estas dos variables ocupan 15 kB por lo que se almacenan en la región de memoria FRAM y almacenan 8 datos cada 250 ms. Posteriormente, cada 2 minutos (la elección de este periodo se explica más en adelante), se copian los valores de estas dos variables (DF y DI) a la memoria microSD para poder leer la información en un ordenador. Como

resultado se obtiene un fichero de texto (extensión .txt) que contiene de manera secuencial frecuencia e intensidad relativa de cada uno de los subintervalos medidos.

El almacenamiento en la memoria microSD ha sido uno de los puntos que más tiempo de desarrollo ha consumido. Y el motivo principal es que existe una interferencia entre el registro del acelerómetro y la comunicación SPI con la microSD. Esta interferencia impide que se produzca una transferencia de datos a la microSD mientras se está registrando. La solución propuesta consiste en que en el momento en el que se desea realizar la copia de las variables DF y DI en la microSD se detiene el registro, se transfieren los datos, y se reemprende el registro. Esto conlleva no registrar durante la totalidad del tiempo en que el dispositivo esté en funcionamiento. Como se ha mencionado antes se opta transferir cada 2 minutos todos los datos registrados durante los 480 intervalos que contiene ese periodo de 2 minutos, en total 7680 valores. Comenzar y terminar la comunicación con la microSD consume tiempo y energía, por lo que es más eficiente (en tiempo y consumo) transferir conjuntos de datos que hacerlo de manera individual. Sin embargo, cuantos más datos se envían en cada paquete más tiempo está sin registrar el dispositivo de manera continua. Por esto se escogieron 2 minutos de registro con una pausa para transferencia de 5 segundos aproximadamente como un equilibrio entre tiempo de registro y tiempo de pausa.

De manera suplementaria se funcionalizó uno de los botones de la tarjeta de desarrollo para almacenar manualmente la información. Así, al pulsarlo, se guarda la información de las variables DF y DI en la micro SD y se genera un nuevo archivo. Esto fue utilizado en las pruebas de corta duración (menos de dos minutos) realizadas con el dispositivo.

1.3. Gestión de la memoria

En todo momento, durante la explicación de cada uno de los pasos que se han realizado para la obtención de los parámetros de la voz elegidos se ha hecho hincapié en las variables, su tamaño y posicionamiento. En este apartado vamos a detallar más estos aspectos, fundamentales para el funcionamiento del dispositivo.

El microcontrolador utilizado, el MSP430FR5994 dispone de 256 kB de memoria FRAM y 8 kB de memoria SRAM. Como ya se ha mencionado anteriormente, de estos 8 kB de memoria SRAM sólo 4 KB están destinados para LEA, de los cuales además 312 bytes pertenecen a LEASTACK. LEASTACK es la memoria reservada para el sistema para realizar operaciones internas de LEA y en la que el usuario no debe posicionar variables para la correcta ejecución de las funciones que usan esta memoria. Por lo tanto, quedan 3784 bytes de memoria SRAM disponible para las variables que LEA va a utilizar.

Las variables que vamos a posicionar en esta sección de memoria son las de la Tabla 3:

Nombre	Tipo	Posición en memoria (hexadecimal)	Alineamiento	Longitud	Tamaño (bytes)
Input	_q15	0x00002c00	1024	512	1024
RdO	_q15	0x00003400	1024	256	512
Fil_Coef	_q15	0x00003800		16	32
Buff_Fil_Coef	_q15	0x00003840		32	64
Fil_Coef_Corr	_q15	0x00003600		256	512
Buff_Fil_Coef_Corr	_q15	0x00003000		512	1024

Tabla 3: Distribución de las variables en memoria RAM dentro de la sección correspondiente a LEA.

La suma total de los tamaños de las variables es de 3168 bytes de memoria ocupados de los 3784 disponibles. Con una ocupación del más de 83 %, teniendo libres solo 616 bytes. Esto hace que sea imposible aumentar la resolución frecuencial, ya que requeriría por lo menos duplicar el tamaño de las variables `Input`, `RdO`, `Fil_Coef_Corr` y `Buff_Fil_Coef_Corr`. Las funciones `msh_fft_fixed_q15` y `msh_fir_q15` requieren que los argumentos de entrada sean potencias de dos. Y aun sobreescribiendo la variable `Buff_Fil_Coef_Corr` en la posición de memoria de la variable `Input` se necesitaría 4192 bytes, lo que excede los 3784 bytes disponibles.

2. VALIDACIÓN

Todo el proceso anteriormente descrito debe ser validado para cerciorarse que el resultado obtenido por el dispositivo corresponde con la realidad. Para ello se les sometió a diferentes pruebas que serán descritas en los siguiente subapartados.

2.1. Validación del microcontrolador y del algoritmo

El primer aspecto a validar es el propio algoritmo implementado en el microcontrolador. Es preciso comprobar que los resultados obtenidos de las operaciones ejecutadas por el microcontrolador se corresponden con los resultados esperados. Con este objetivo se generaron varias señales de prueba con características conocidas que fueron procesadas por el microcontrolador y en el software Matlab (Matlab R2018b, Mathworks). Se escogió Matlab como método de verificación dado que consideramos que sus resultados son válidos.

Las señales se generaron en Matlab y fueron las siguientes (Figura 28):

- Seno de 90 Hz y 3 armónicos. Amplitud relativa entre los armónicos: 1, 1, 1.
- Seno de 120 Hz y 3 armónicos. Amplitud relativa entre los armónicos: 1, 1, 1.
- Seno de 200 Hz y 3 armónicos. Amplitud relativa entre los armónicos: 0.5, 0.7, 1.
- Seno de 400 Hz y 2 armónicos. Amplitud relativa entre los armónicos: 0.5, 1.
- Seno de 600 Hz y 1 armónico.

Todas estas se generaron con una frecuencia de muestreo de 2048 Hz y un número de 512 muestras para hacerlas coincidir con el tipo de señal que recopila el dispositivo por parte del acelerómetro. Estos datos generados en Matlab se almacenaron en el dispositivo de manera manual para asegurar que los valores correspondientes a todos los puntos de la señal sean exactamente los mismos tanto en Matlab como en el dispositivo y así poder realizar la comparación.

Estas señales se eligen con las características descritas para mostrar cómo el algoritmo trabaja en diferentes situaciones. La primera ejemplifica el motivo por el que la correlación no es suficiente y cómo el algoritmo utiliza la información de la magnitud de la FFT para calcular correctamente la F_0 . La segunda, tercera y cuarta representan las opciones contempladas para encontrar el armónico correspondiente a la frecuencia fundamental en la magnitud de la FFT. A su vez, en estos tres casos la correlación será suficiente para el cálculo de F_0 . Por último, la quinta señal representa la actuación del algoritmo al tener un solo armónico.

Una vez almacenados los valores se ejecutó el algoritmo en ambos sistemas. Para poder hacer una mejor comparación se extraen del dispositivo los datos de cada uno de los pasos del algoritmo.

2.2. Validación de los parámetros de interés en la voz humana

El siguiente paso realizado es la comprobación de los resultados obtenidos por el dispositivo en cuanto a los parámetros de la voz que estamos estudiando (frecuencia fundamental, intensidad relativa y tiempo de fonación). Para ello se colocó el dispositivo en tres sujetos (dos hombres y una mujer) y se realizaron grabaciones con el dispositivo y, simultáneamente, con el equipo usado por el laboratorio de voz de la Clínica Universidad de Navarra (CUN), Pamplona, para la medición de cada una de estos parámetros. Todos los registros fueron tomados en el laboratorio de voz de la Clínica Universidad de Navarra.

2.2.1. Validación de la frecuencia fundamental

Para la validación de la medición de la frecuencia fundamental se colocó el dispositivo en los tres sujetos. Estos sujetos generaron una escala tonal, cada uno de ellos dentro de su rango confortable y sin predeterminedar la frecuencia de cada uno de los escalones tonales. Cada una de las notas, realizada con una vocal, se mantuvo entre 1 y 3 segundos de manera constante y con breve silencio entre cada una de ellas.

La medición de la frecuencia fundamental se realizó usando dos métodos. El primero, es el dispositivo que se presenta en este documento. El segundo fue a través de un micrófono Logitech 980186-0403. La señal es digitalizada en la tarjeta A/DSoundScope/16, GW instruments y estudiada en el programa SoundScope 1.2 del mismo fabricante de la tarjeta instalado en un Power Macintosh 9600/233, Apple Inc. Este equipo es el usado por el laboratorio de voz de la CUN para realizar los análisis de voz. El resultado obtenido de las mediciones del SoundScope es un valor de frecuencia por cada escalón tonal que se corresponde con la frecuencia fundamental media medida por el programa durante dicho escalón.

Posteriormente, ambos registros se analizaron en Matlab usando el estadístico de correlación y calculando el error absoluto del dispositivo respecto al SoundScope.

2.2.2. Validación de la intensidad relativa

De un modo similar se comparan las mediciones de intensidad relativa registradas por el dispositivo con las medidas de intensidad registradas por el equipo de elección del laboratorio de voz de la CUN.

De nuevo, los tres sujetos realizan sendas escalas de intensidad dentro del rango que cada uno de ellos encuentra confortable. Cada una de las intensidades emitidas se mantuvo entre 1 y 3 segundos de manera constante produciendo una vocal /a/ con un breve silencio entre ellas.

La medición de la intensidad se registró usando dos equipos. El primero, es el dispositivo que se presenta en este documento. El segundo fueron dos sonómetros (Figura 15), equipo usado en el laboratorio de voz de la CUN. Los valores de uno de los sonómetros fueron anotados a mano, un valor por cada intensidad emitida. El otro registraba las mediciones en un ordenador con el software lingWAVES, Wevosys para obtener el fonetograma (ANEXO VIII).



Figura 15: Dispositivos para la medición en la validación de la intensidad relativa. A la izquierda la placa de desarrollo con el microcontrolador elegido para este trabajo, el acelerómetro y la batería, y a la derecha los dos sonómetros.

Posteriormente se realizó la comparación con los valores del sonómetro a través del estadístico correlación calculado en Matlab.

2.2.3. Validación del tiempo de fonación

Por último, se comparó el tiempo de fonación registrado por el dispositivo y el registrado a través de micrófono YETI PRO, Blue Microphones conectado a un ordenador con Windows XP y el software Sound Forge 10.0, Sony. Para ello, cada uno de los tres sujetos realizó un monólogo durante 2-3 minutos que fue registrado por los dos medios mencionados. De manera adicional, se grabó un pequeño fragmento únicamente con vocales y pausas con uno de los sujetos.

Para el análisis de la grabación usando micrófono se utilizó el software Audacity 2.3.1, Audacity Team. Dado que el dispositivo almacena los datos en la memoria microSD cada 120 segundos, se tomaron los primeros 100 s del monólogo. Se eliminaron todas las pausas de la grabación de audio para contabilizar el tiempo que corresponde a voz dentro de la grabación. Para ello se importa el archivo de audio en el programa. Se selecciona toda la pista y se aplica el efecto “truncado de silencio”.

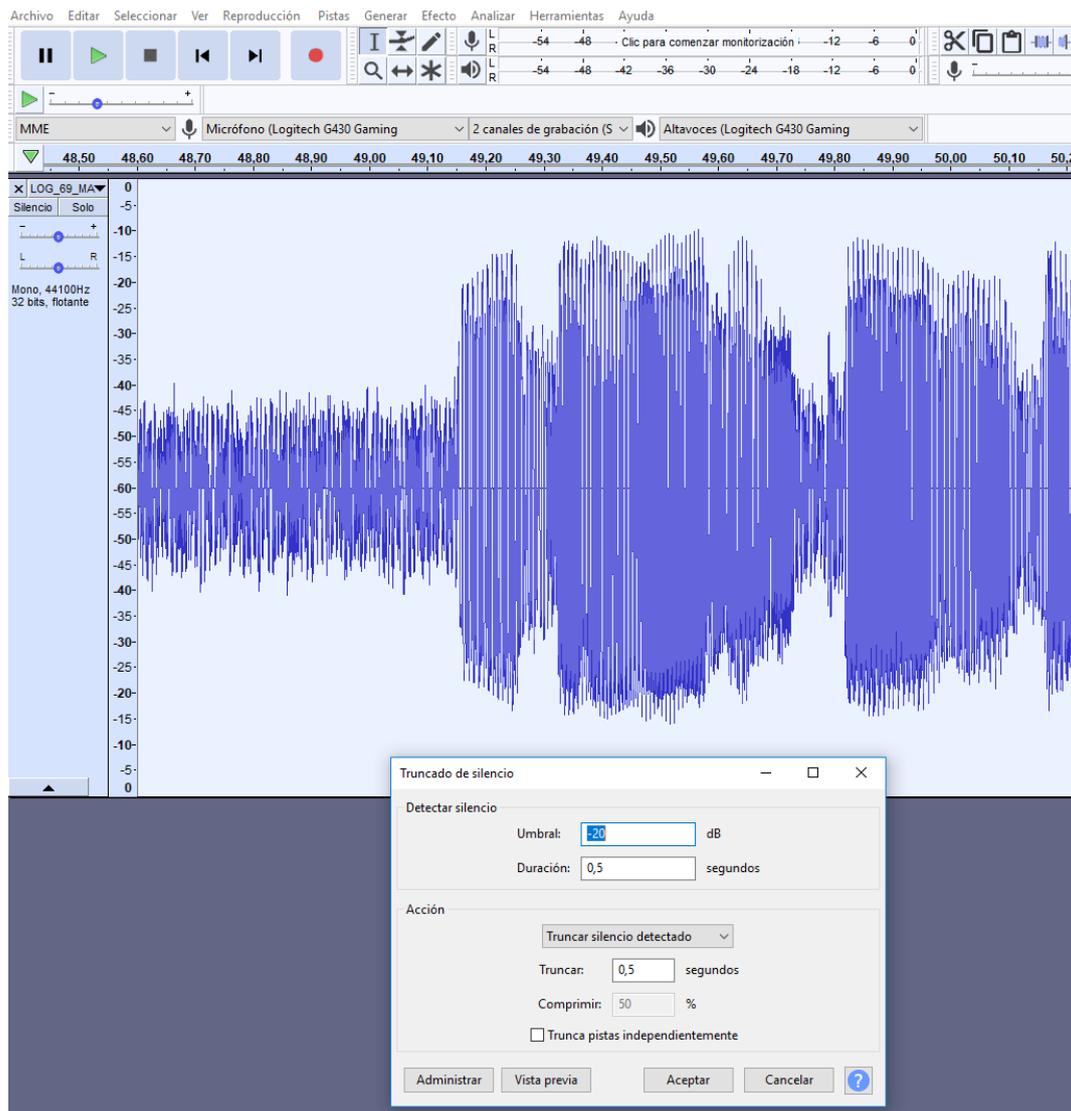


Figura 16: Captura de pantalla del software Audacity. Desplegado el menú del efecto truncado de silencio.

Como se muestra en la Figura 16, el efecto “truncado de silencio” tiene varios parámetros que se pueden ajustar. Dentro de los parámetros de detección de silencio, se ajusta el umbral siendo el valor máximo en dB de la señal en un fragmento identificado como silencio. La duración se determinó como 0.1 segundos para que fuera muy estricto, ya que el dispositivo tiene una resolución temporal de 31.25 ms. En los parámetros de acción, se selecciona la opción “truncar silencio detectado” y se determina como 0 segundos la opción Truncar. Esto hará que cuando se detecte un periodo de silencio que corresponda con las condiciones determinadas se sustituye por un fragmento de 0 segundos de silencio, o lo que es lo mismo, se elimina el silencio detectado.

Con esto se obtiene una señal de audio que concatena todos los fragmentos de voz registrados por el micrófono.

Para contabilizar el tiempo de fonación medido por el dispositivo se usó Matlab. Se lee el archivo guardado en la memoria micro SD y se contabiliza la cantidad de valores correspondientes a la frecuencia fundamental que son distintos de 0 en los 100 primeros segundos de registro. El resultado es el número de aquellos fragmentos de 31.25 ms que se han considerado corresponden a voz. De esta manera, se multiplica este número por los 31.25 ms a los que representa y se obtiene el tiempo total de fonación.

3. REGISTRO DE VOZ DURANTE LARGOS PERIODOS DE TIEMPO

Para finalizar el estudio del dispositivo se realizaron grabaciones durante largos periodos de tiempo con 2 sujetos. El sujeto 1 registró 14 días, y el sujeto 2 registró 11 días.

Para ello se preparó una caja donde se almacenó la placa del microcontrolador, el circuito de adaptación para el sensor y una batería externa de 2000 mAh. De la caja salía un metro de cable que conectaba el acelerómetro con el microcontrolador. Esta caja se cerraba y se entregaba durante la mañana a los dos sujetos. Estos se la guardaban en un bolsillo de su bata y se colocaban el colgante como se muestra más adelante en la Figura 18. Finalizada la jornada laboral, el dispositivo se devolvían para almacenar los datos y recargar la batería. Ambas acciones no eran estrictamente necesarias dado que tanto la batería como la memoria podrían aguantar más días, pero se decidió así para asegurar que el funcionamiento era el correcto y que no se perdía la información registrada.

A ambos sujetos se les pidió que hicieran un uso normal de la voz salvo uno de los días que deberían de simular un patrón hiperfuncional. Este patrón hiperfuncional consistía en usar frecuencias y/o intensidades fuera del rango habitual en un entorno fuera de la jornada laboral.

VII. RESULTADOS

1. DESARROLLO DEL SISTEMA

1.1. Hardware

El prototipo del dispositivo, objeto de la tesis, se muestra en la Figura 17. La placa se alimentó con una batería externa con capacidad de 2000 mAh y con salida de 5 V que se conecta a través del puerto micro USB.

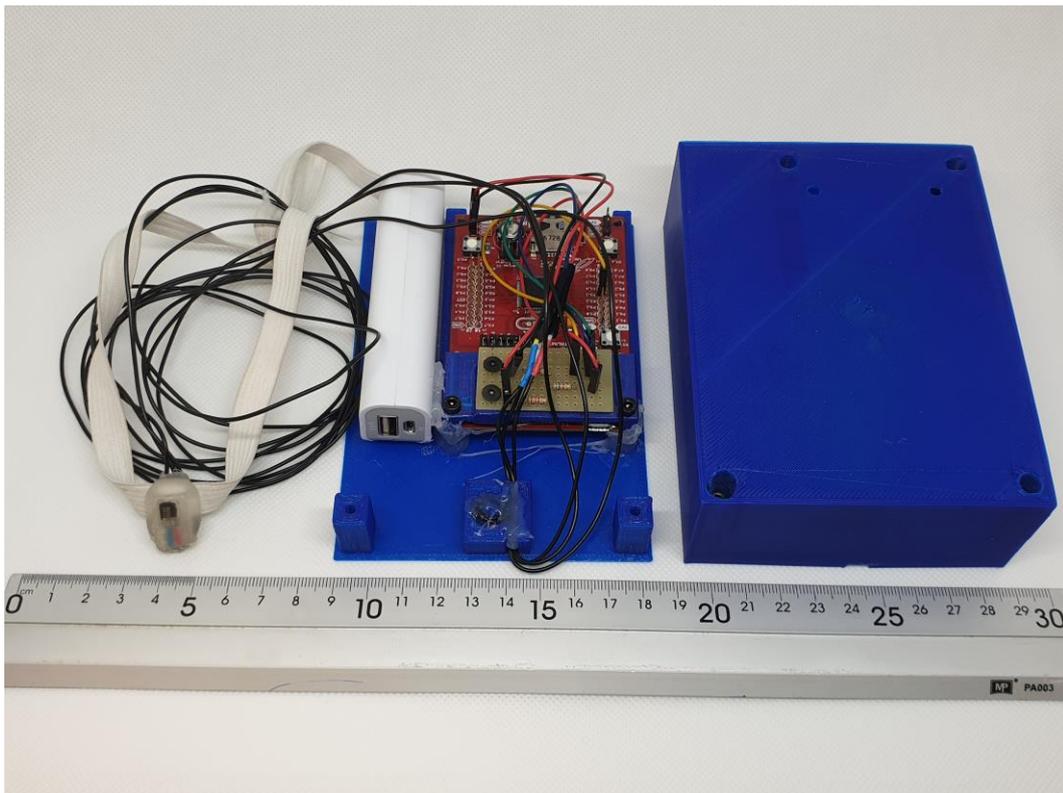


Figura 17: Prototipo para los ensayos realizados. Acelerómetro en la pieza para usarlo de colgante, tarjeta de desarrollo con circuito de la Figura 6 acoplado y caja donde almacenarlo junto a la batería.

El acelerómetro se posiciona en la pieza mostrada en la Figura 5, se adhiere con cola caliente y se coloca una cinta elástica. El resultado de su posicionamiento se muestra en la Figura 18. Los tres cables que se observan van conectados a la placa donde se encuentra el microcontrolador.



Figura 18: Disposición del sensor sobre la laringe utilizando la pieza detallada anteriormente.

Esta es la configuración que se utilizó para tomar los registros. De manera paralela se diseñó lo que podría ser el dispositivo final. Se tuvieron en cuenta los siguientes componentes para generar el volumen del dispositivo y calcular peso:

Componente	Medidas (mm)	Peso (g)	Precio (€)
Microcontrolador MSP430FR5994 IPN	16x16x1.5	0.462	7.18
BU-27135-000	7.92x5.59x2.28	0.28	41.25
Ranura micro SD Molex 504077- 1891 (ANEXO IX)	11.32x15.4x2.3	0.188	1.77
Placa base + resistencias smd	20x16x2	<1	<2
Batería 400 mAh (ANEXO X)	37x25x5.2	13	4.44
Circuito para cargar la batería	37x10x2.5	<5	<10

Tabla 4: Medidas, peso y coste de los componentes necesarios más relevantes para el montaje del dispositivo definitivo.

El peso y precio de la placa base, resistencias y circuito de carga para la batería se han estimado en base a experiencias anteriores con este tipo de dispositivos. Teniendo estos datos en cuenta, el peso es menor de 20 g y el coste de menos de 67 €. A estos componentes habría que sumar el coste de carcasa de plástico, de la correa y del montaje. En el peor de los casos el dispositivo pesaría menos de 40 g y tendría un coste menor a 100 €. La apariencia sería la mostrada en la Figura 19.

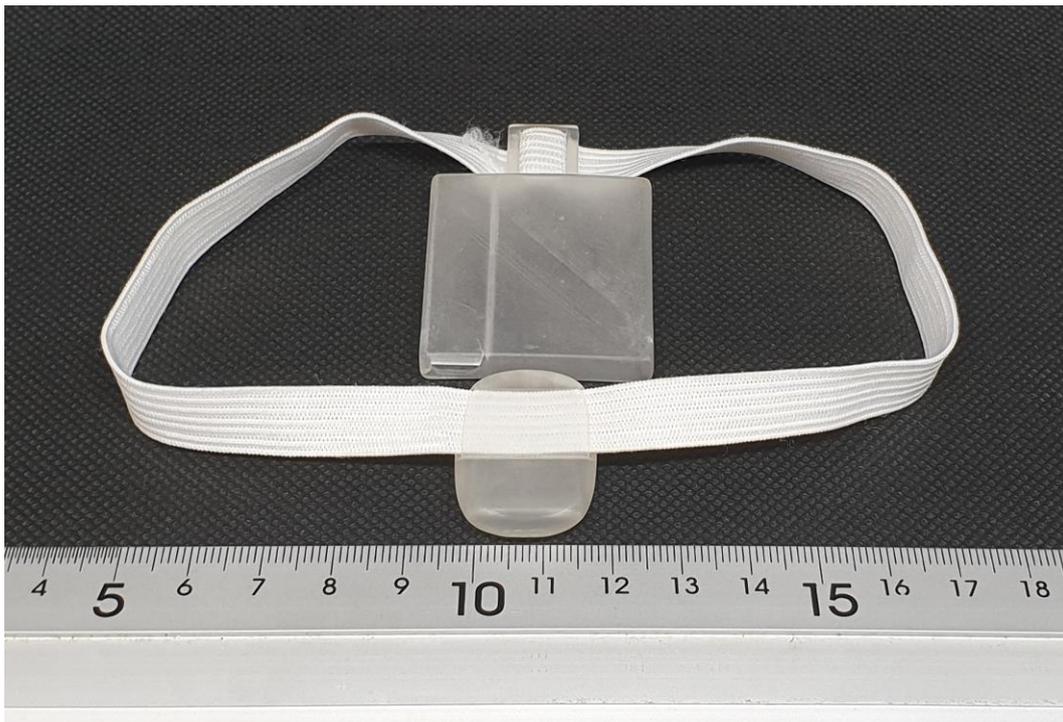


Figura 19: Prototipo del dispositivo final (ANEXO XI).

La batería seleccionada es de 400 mAh, lo que supondría una duración de 7 días de registro ininterrumpido.

1.2. Algoritmo

1.2.1. Tiempo empleado para ejecutar cada proceso

En este trabajo el tiempo es muy importante, no solo porque es una de las características de la voz que vamos a medir (tiempo de fonación), sino porque queremos que todo el procesamiento de la señal se ejecute en tiempo real. Para poderlo ejecutar en tiempo real, la duración de todas las operaciones tiene que

ser menor al tiempo del intervalo de medición que es de 250 ms. Además, cuanto menor sea el tiempo necesario para las operaciones más tiempo se encontrará el dispositivo en modo de ahorro de energía y mayor será el tiempo máximo de funcionamiento (a igual batería).

Operación	Tiempo (ms)
Copia valores acelerómetro	1.4
Cálculo de la intensidad relativa	13.4
Cálculo de la FFT	0.91
Cálculo de la magnitud	3.61
Resta de la media	1.27
Cálculo frecuencia en Fft	0.40
Filtrado	2.77
Correlación 1	10.56
Cálculo del multiplicador	1.70
Escalado	7.44
Correlación 2	8.55
Cálculo de frecuencia fundamental	0.77
Grabación en microSD	3500-5200

Tabla 5: Tiempo necesario para la realización de cada operación.

En el caso en que identifique que no hay voz en el intervalo de medición, el tiempo de procesado es de 21 ms aproximadamente. En caso en que sí se detecta voz, el procesado tarda 53 ms. Con lo que se disminuye un 60 % el tiempo de cálculo y por lo tanto se reduce el consumo en caso de no detectarse voz. Así pues, de los 250 ms del intervalo, como mínimo 197 ms (el 79 %) es tiempo en el que el microcontrolador se encuentra en modo de bajo consumo. Este porcentaje aumenta al 91 % en caso de no reconocer voz.

En las siguientes imágenes podemos ver el tiempo de actividad de bajo consumo del procesador en cada ciclo de registro, sin fonación detectada y con ella.

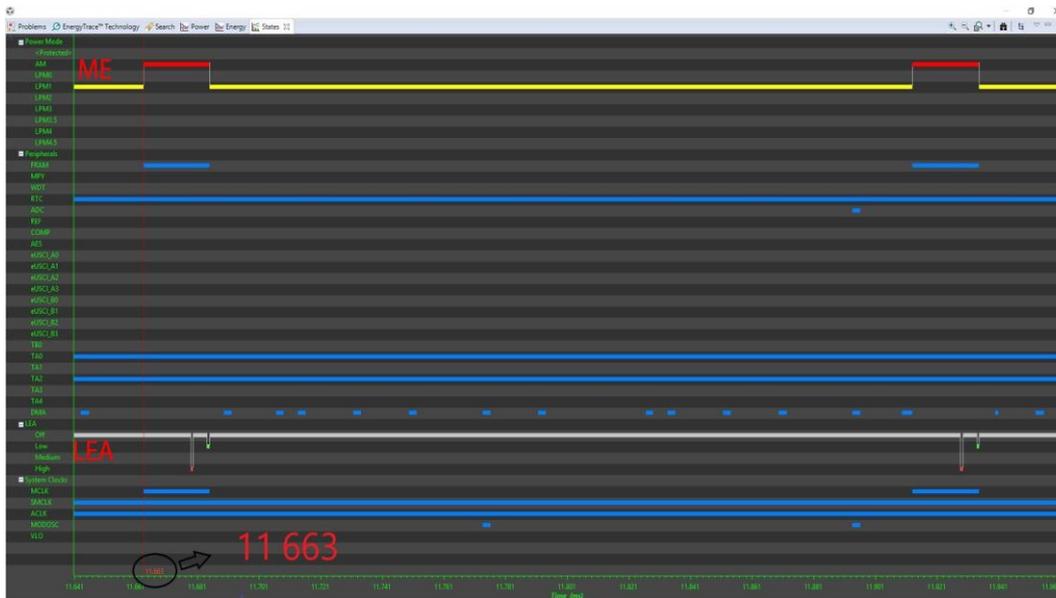


Figura 20: Comienzo de registro y procesado, 11663 ms. Fonación no detectada. La gráfica titulada ME (Modo de Energía) representa en amarillo el tiempo en que el microcontrolador se encuentra en modo de bajo consumo y en rojo cuando se encuentra a pleno rendimiento. La gráfica titulada LEA muestra la actividad de LEA, en gris cuando no está activa y en rojo o verde cuando está activa.

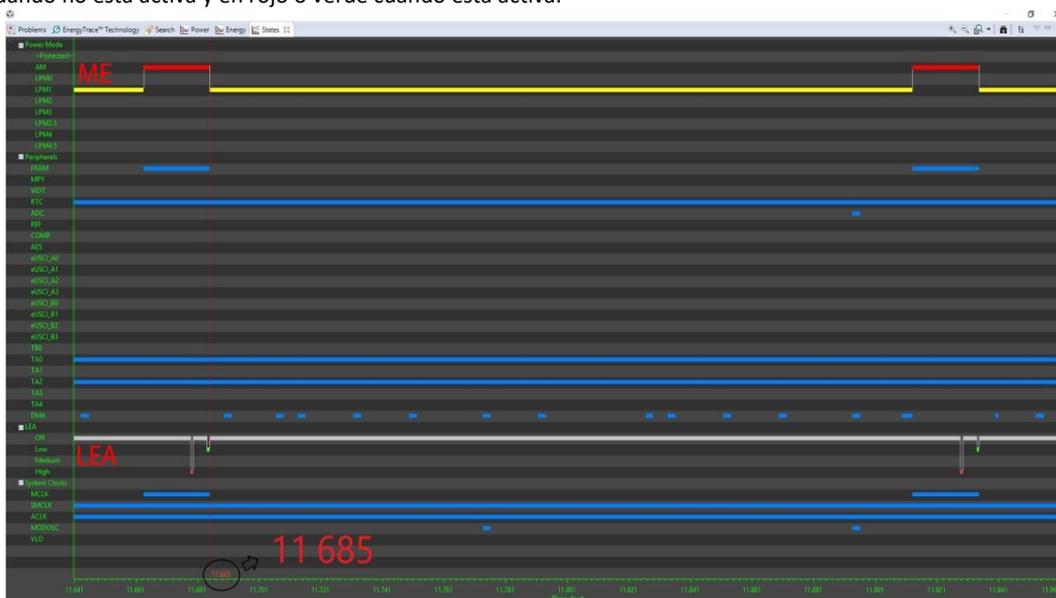


Figura 21: Final del procesado, 11685 ms. Fonación no detectada. El significado de las marcas ME y LEA es el mismo que en la Figura 20.

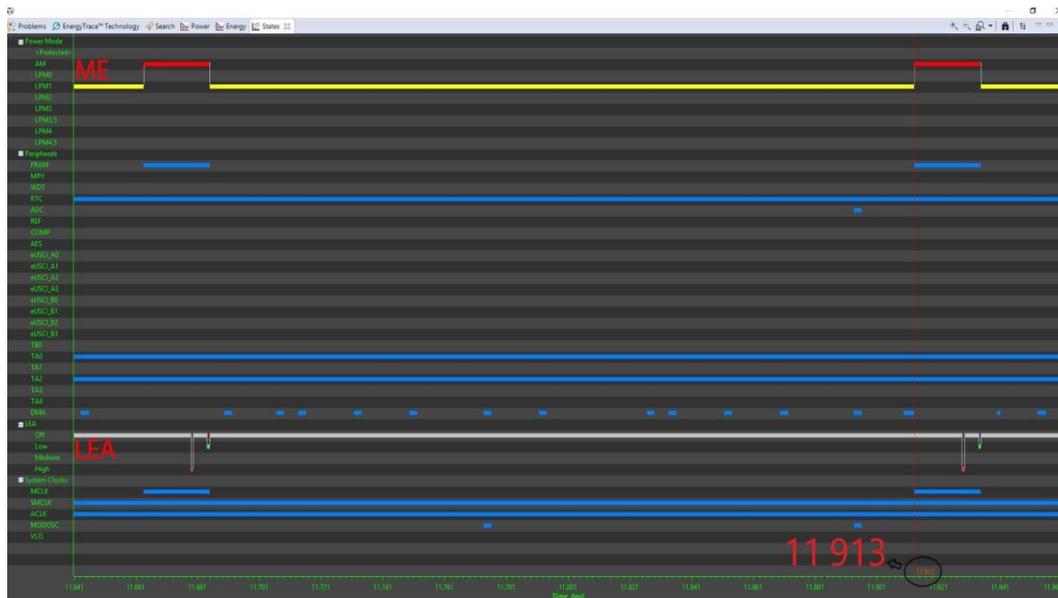


Figura 22: Fin del registro, 11913 ms. Fonación no detectada. El significado de las marcas ME y LEA es el mismo que en la Figura 20.

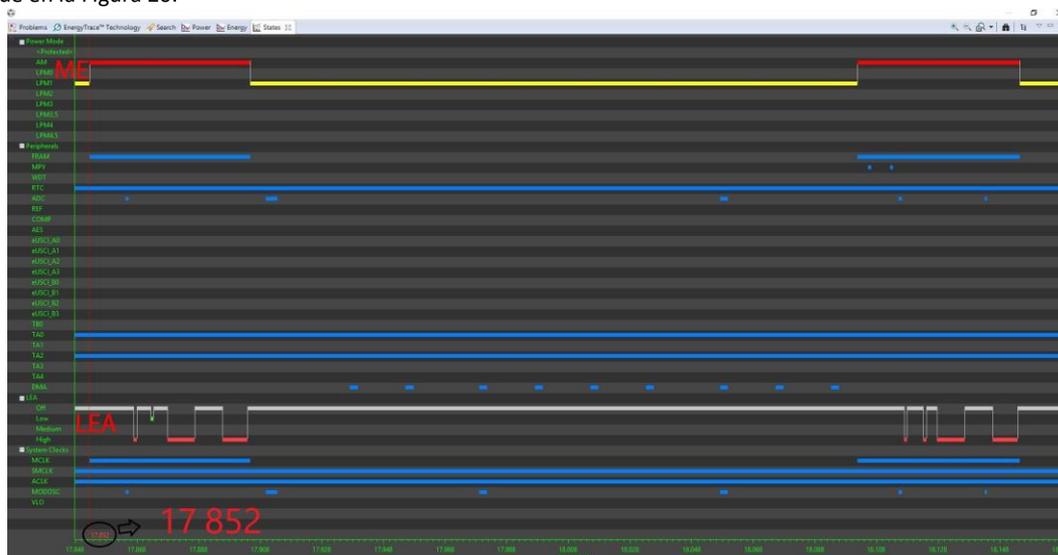


Figura 23: Comienzo de registro y procesado, 17852 ms. Fonación detectada. El significado de las marcas ME y LEA es el mismo que en la Figura 20.

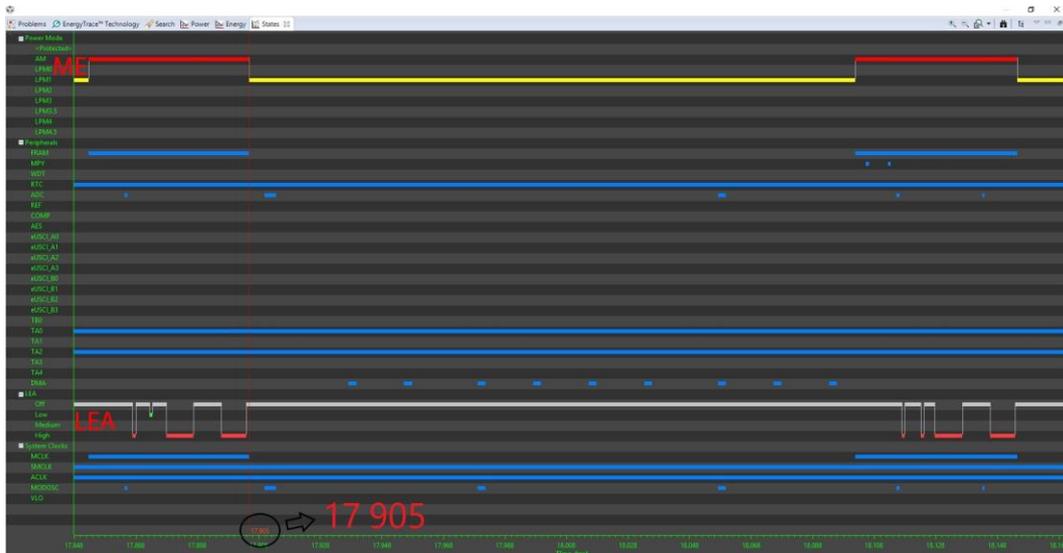


Figura 24: Fin del procesado, 17905 ms. Fonación detectada. El significado de las marcas ME y LEA es el mismo que en la Figura 20.

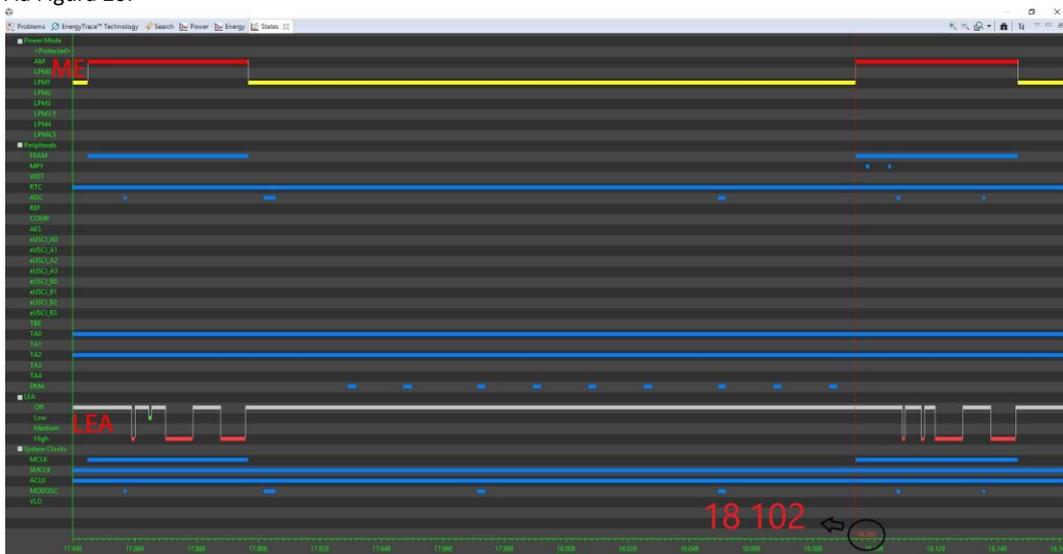


Figura 25: Fin del registro, 18102 ms. Fonación detectada. El significado de las marcas ME y LEA es el mismo que en la Figura 20.

La línea ME de la Figura 20 a la Figura 25 representa el modo en el que se encuentra el microcontrolador. El color rojo significa que el microcontrolador está activo a pleno rendimiento y el amarillo significa que el procesador se encuentra en modo de bajo consumo. Si nos fijamos en la línea temporal (eje horizontal de la Figura 20 a la Figura 25), expresada en milisegundos, comprobamos que los tiempos corresponden con los cálculos de la Tabla 5. En ambos casos son 250 ms el tiempo entre el inicio y el fin del registro. El tiempo entre el comienzo del procesado y el fin del procesado en el caso de no detectar

fonación es de 22 ms. En el caso de detectar fonación el tiempo de procesado es de 53 ms.

También en estas imágenes se puede observar la activación de LEA. Para ello, debemos fijarnos en la línea blanca de las figuras. Esta línea se ve interrumpida y aparece en rojo o verde cada vez que LEA se activa. En la Figura 20 se pueden observar dos activaciones durante el periodo de procesado. Estas corresponden al cálculo de la FFT y al cálculo de la magnitud. Sin embargo, en la Figura 23 podemos observar que el número de activaciones dentro de cada ciclo de procesado asciende a cuatro. Además del cálculo de las FFT y la magnitud, se realiza filtrado y dos correlaciones. El motivo por el que aparecen cuatro puntos de activación en vez de cinco, como cabría esperar, es debido a que el filtrado y la primera correlación se hacen consecutivamente, por lo que LEA no llega a desactivarse.

Si se tienen en cuenta también los periodos de grabación de datos, el porcentaje de tiempo en bajo consumo quedaría en torno al 73 %. En la Figura 26 se muestra el resultado obtenido en el programa CCS después de 600 segundos de grabación. Este fragmento puede tomarse como representativo de cuál sería el porcentaje de tiempo en bajo consumo del dispositivo. A partir del segundo 400 la variación en los porcentajes no superó el $\pm 0,5$ %.

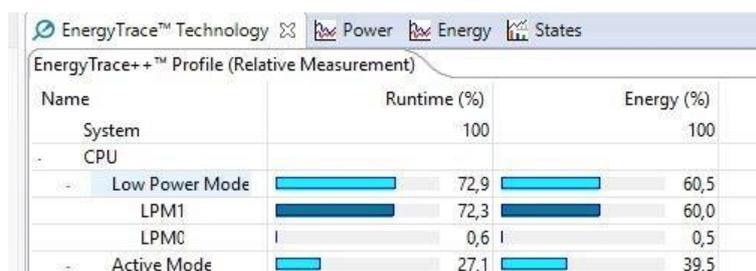


Figura 26: Datos de actividad del microcontrolador durante grabación de 600 segundos mostrados en el programa CCS.

1.2.2. Consumo de energía

La reducción de consumo del dispositivo es un punto muy importante en el objetivo de conseguir que éste sea vestible. La tarjeta de desarrollo usada junto

con el software CCS permite evaluar cuál es el consumo de energía en la opción de EnergyTrace. Además, realiza un cálculo estimado de cuál sería la duración de la batería que el usuario indique.

EnergyTrace™ Profile (Relative Measurement)	
Name	Live
System	
Time	600 sec
Energy	4866,939 mJ
Power	
Mean	8,0118 mW
Min	0,0000 mW
Max	172,3725 mW
Voltage	
Mean	3,2873 V
Current	
Mean	2,4371 mA
Min	0,0000 mA
Max	52,4726 mA
Battery Life	2xAAA: 17,7 day (est.)

Figura 27: Consumo energético del dispositivo propuesto.

Como se muestra en la Figura 27, el consumo energético medio es de 2.44 mA. Esto significa que alimentando el dispositivo con 2 pilas AAA (1150 mAh), el tiempo de funcionamiento sería de 17.7 días. Esto es posible gracias a las características del microcontrolador y en gran parte a la estructura del algoritmo.

2. VALIDACIÓN

2.1. Validación del microcontrolador y del algoritmo

Las cinco señales generadas para comprobar el desempeño del dispositivo y validar el algoritmo se generaron en Matlab y son las mostradas a continuación.

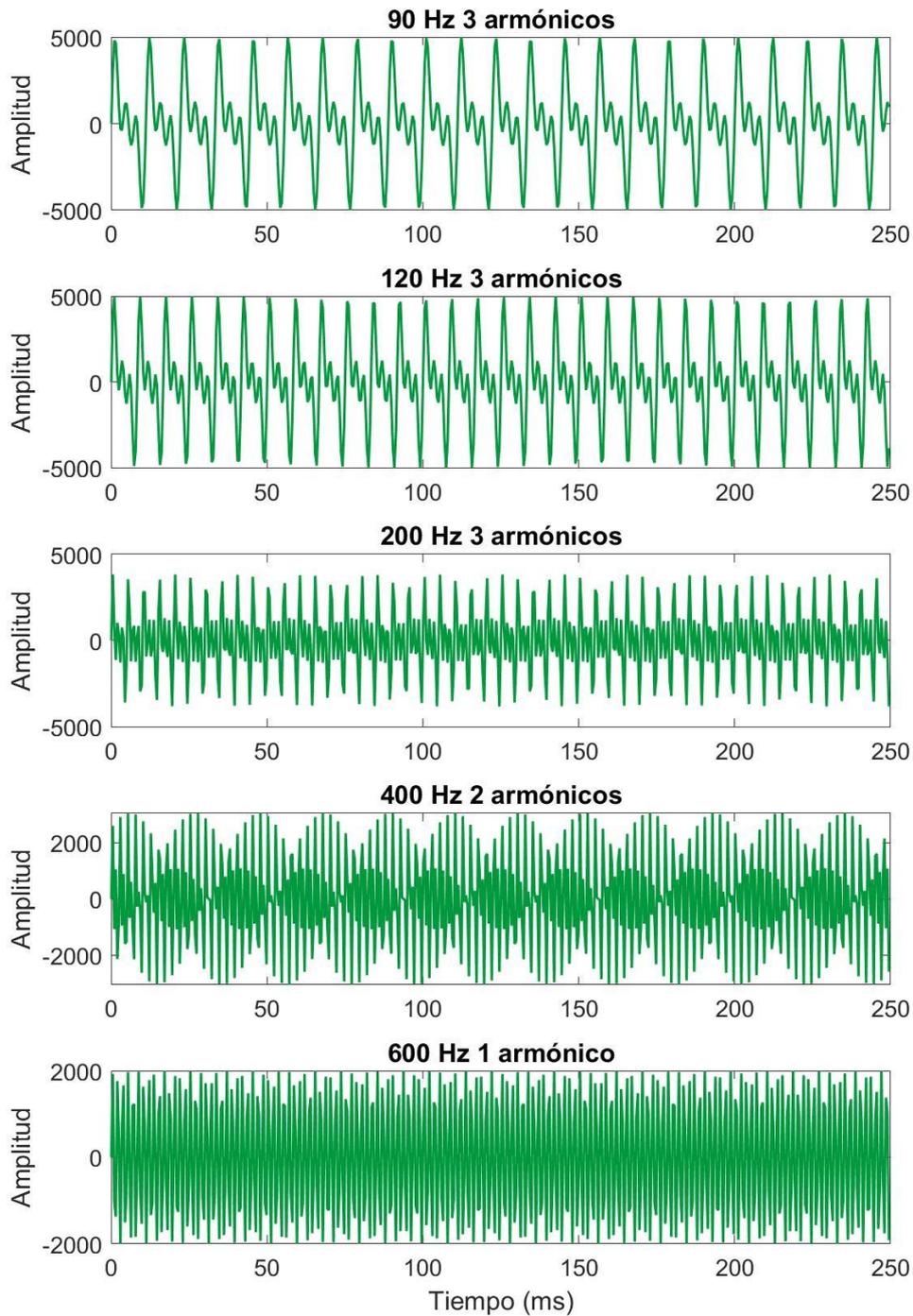


Figura 28: Las cinco señales de prueba generadas en Matlab. La frecuencia de muestreo utilizada para generarlas fue de 2048 Hz y se recogieron 512 muestras.

Estas 5 señales corresponden a tres casos distintos en el recorrido del algoritmo para el cálculo de la frecuencia fundamental.

Caso 1:

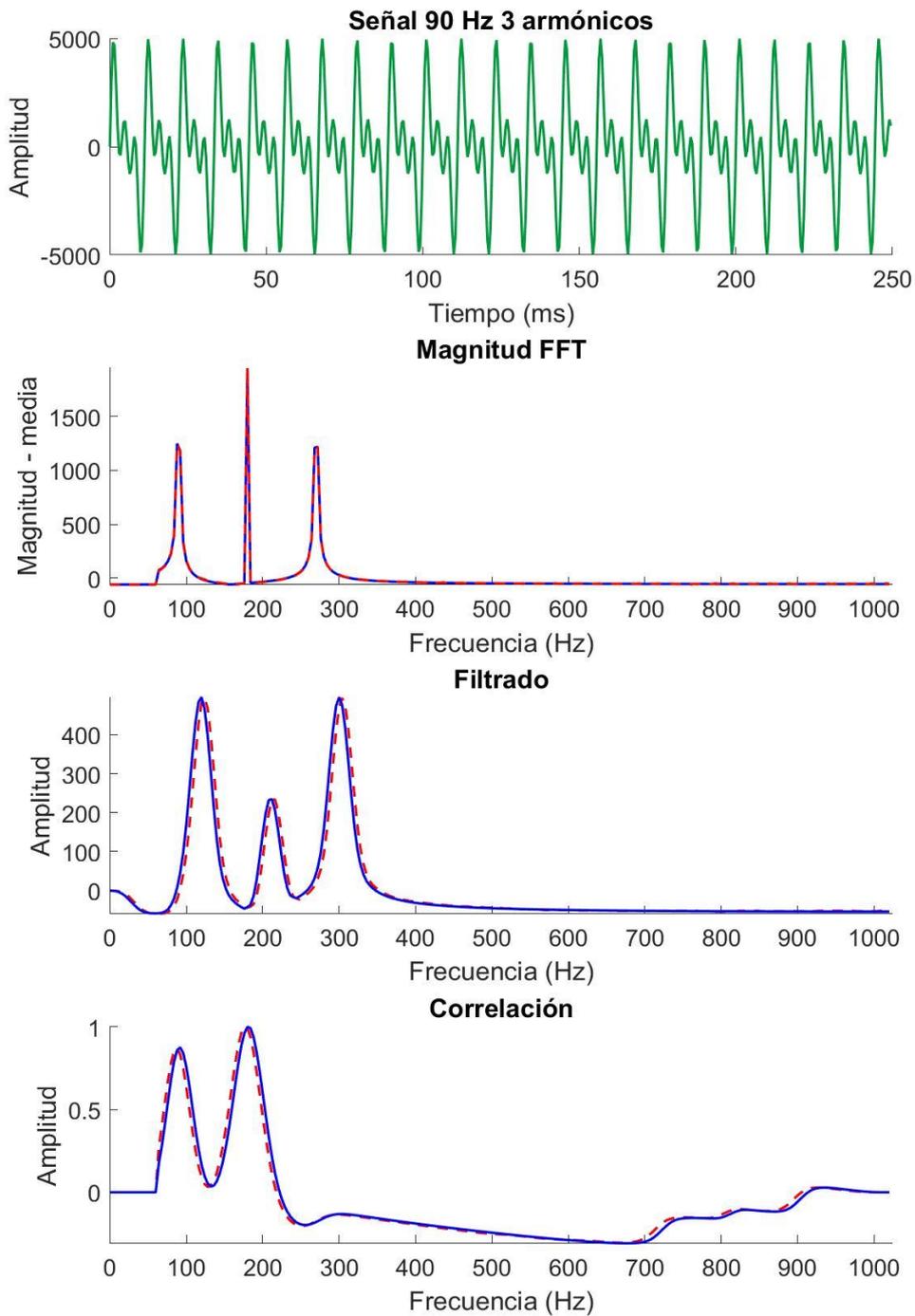


Figura 29: Análisis de la señal de 90 Hz y 3 armónicos por parte del algoritmo usando el dispositivo y Matlab. La línea discontinua de color rojo representa los datos procesados por el dispositivo y la línea continua azul representa los datos procesados en Matlab.

En la Figura 29 se muestra el tipo de análisis empleado en el procesado de cada una de estas señales para observar cómo es el comportamiento del algoritmo.

En este análisis de la señal, mostrada en la Figura 29, se destacan dos aspectos que se trasladarán al resto de los análisis, sea cual sea la señal introducida.

El primero es que, en el cálculo de la magnitud de la FFT, se observa que el cálculo realizado en Matlab y con el dispositivo es prácticamente idéntico, la posición de los armónicos es la misma y varía muy ligeramente la magnitud (error absoluto medio de 1.74 %). Esta diferencia se debe a que las funciones FFT no son exactamente las mismas y los resultados en Matlab tiene decimales mientras que los del dispositivo son números enteros. Esto confirma que el cálculo de la FFT realizado en el dispositivo se pueden considerar como correcto puesto que solo se tomará en cuenta la posición del armónico y no la magnitud.

El segundo aspecto a destacar es el retraso introducido en el dispositivo en el cálculo del filtrado. En la Figura 30 se observa la misma gráfica de correlación que en la Figura 29, pero con un zoom sobre el máximo. Este mismo retraso se observa en la correlación. Se debe tener en cuenta que los valores mostrados en la gráfica de correlación tienen un orden invertido al que se obtiene tras el filtrado. Por ello en la gráfica de correlación la señal del dispositivo está retrasada cuando en la de filtrado está adelantada. Como se comenta en el apartado 1.2.2.6 del capítulo VI, en el cálculo de la frecuencia se tiene en cuenta este retraso y por ello se suma una posición a la posición obtenida de la correlación. O lo que es lo mismo, se suman 4 Hz a los resultados.

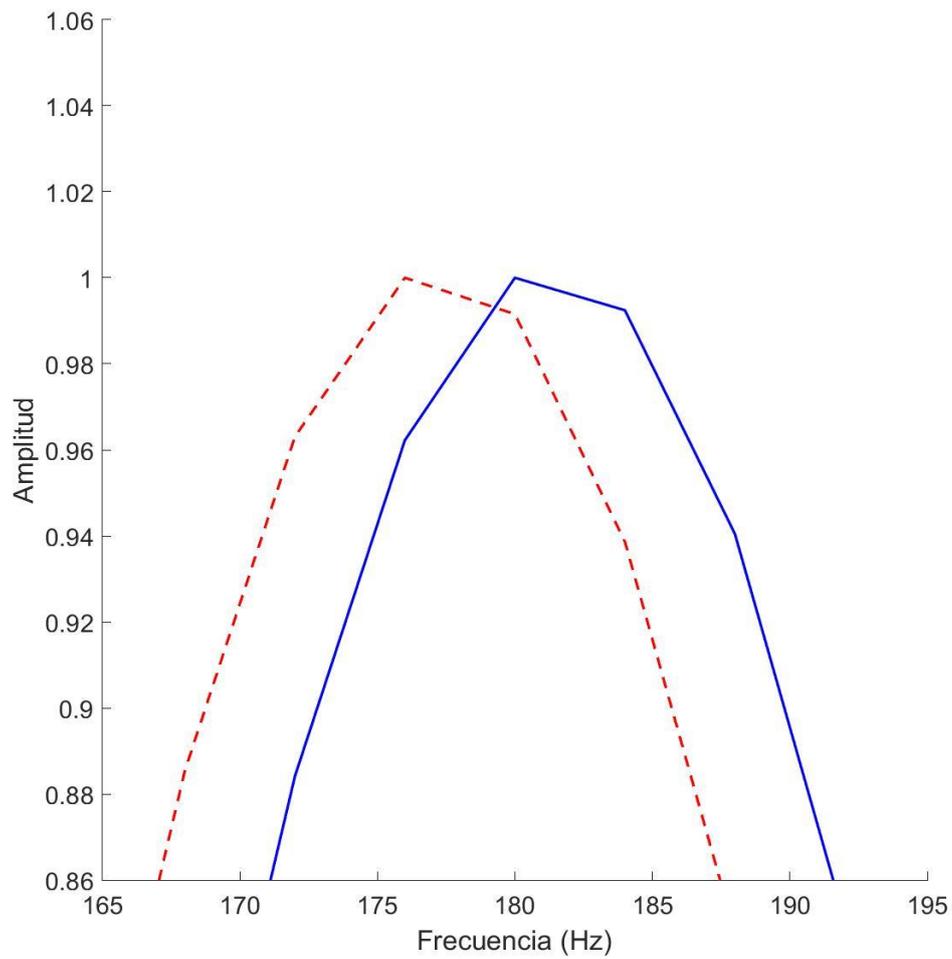


Figura 30: Zoom realizado sobre el máximo de la gráfica de correlación correspondiente a la señal de 90 Hz y 3 armónicos (Figura 29). La línea discontinua de color rojo representa los datos procesados por el dispositivo y la línea continua azul representa los datos procesados por Matlab.

Según lo expuesto anteriormente se comprueba que tanto Matlab como el dispositivo obtienen como resultado una frecuencia de 88 Hz cuando la esperada es de 90 Hz.

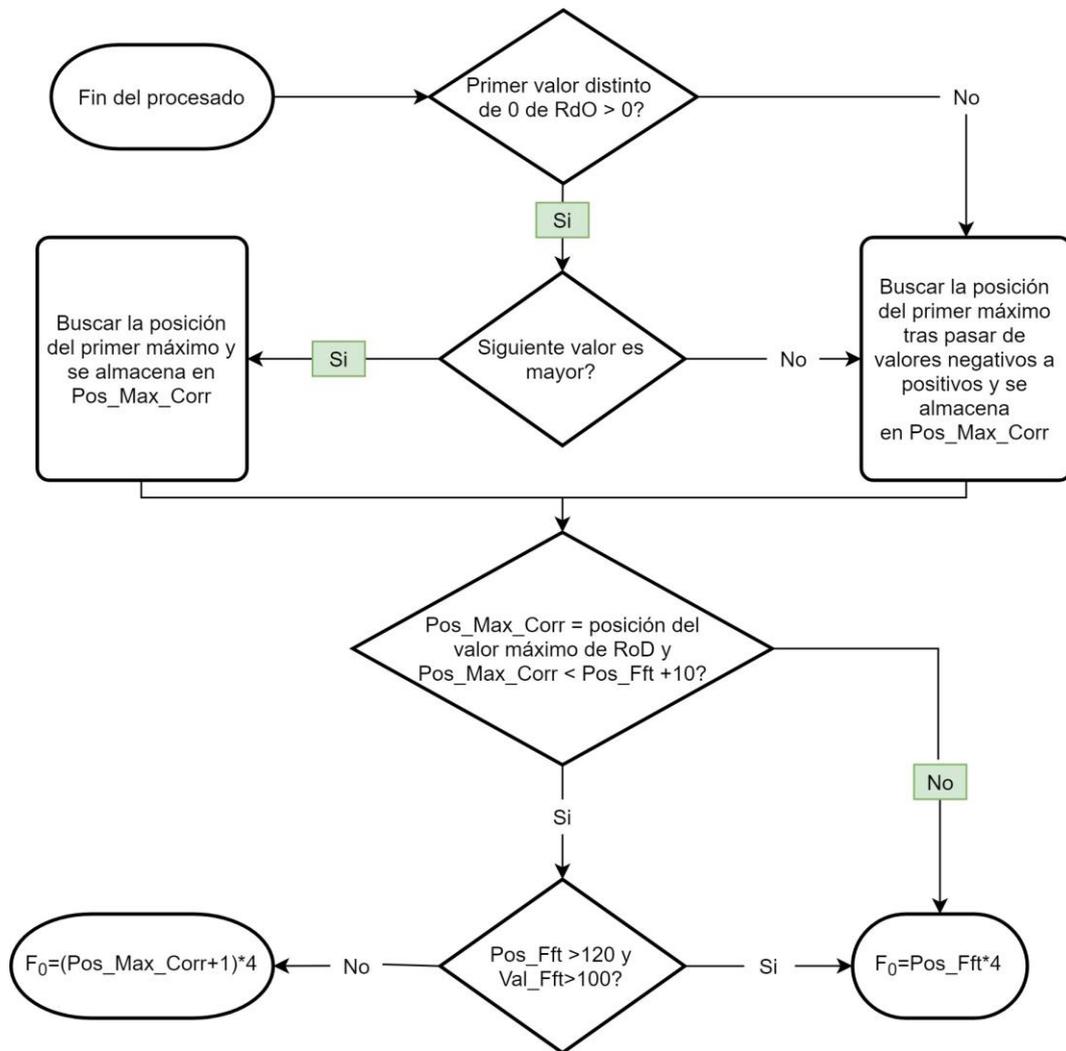


Figura 31: Diagrama del algoritmo para el cálculo de la frecuencia fundamental. En verde se marca el camino que sigue para la señal del caso 1.

Este caso ilustra cómo trabaja el algoritmo en el supuesto en el que el primer armónico no es el que más energía aporta y en donde el máximo de la correlación no coincide con el primer pico. Por ello, para el cálculo de frecuencia fundamental el algoritmo se fija en la magnitud de la FFT. Como se explica en el apartado 1.2.2.3 del capítulo VI, se busca el máximo y se comprueba si es el primer armónico o no. En este caso no lo es, y se encuentra un armónico antes y se queda con la posición de este (en la variable `Pos_Fft`) para realizar el cálculo de la frecuencia fundamental (Figura 31).

Caso 2:

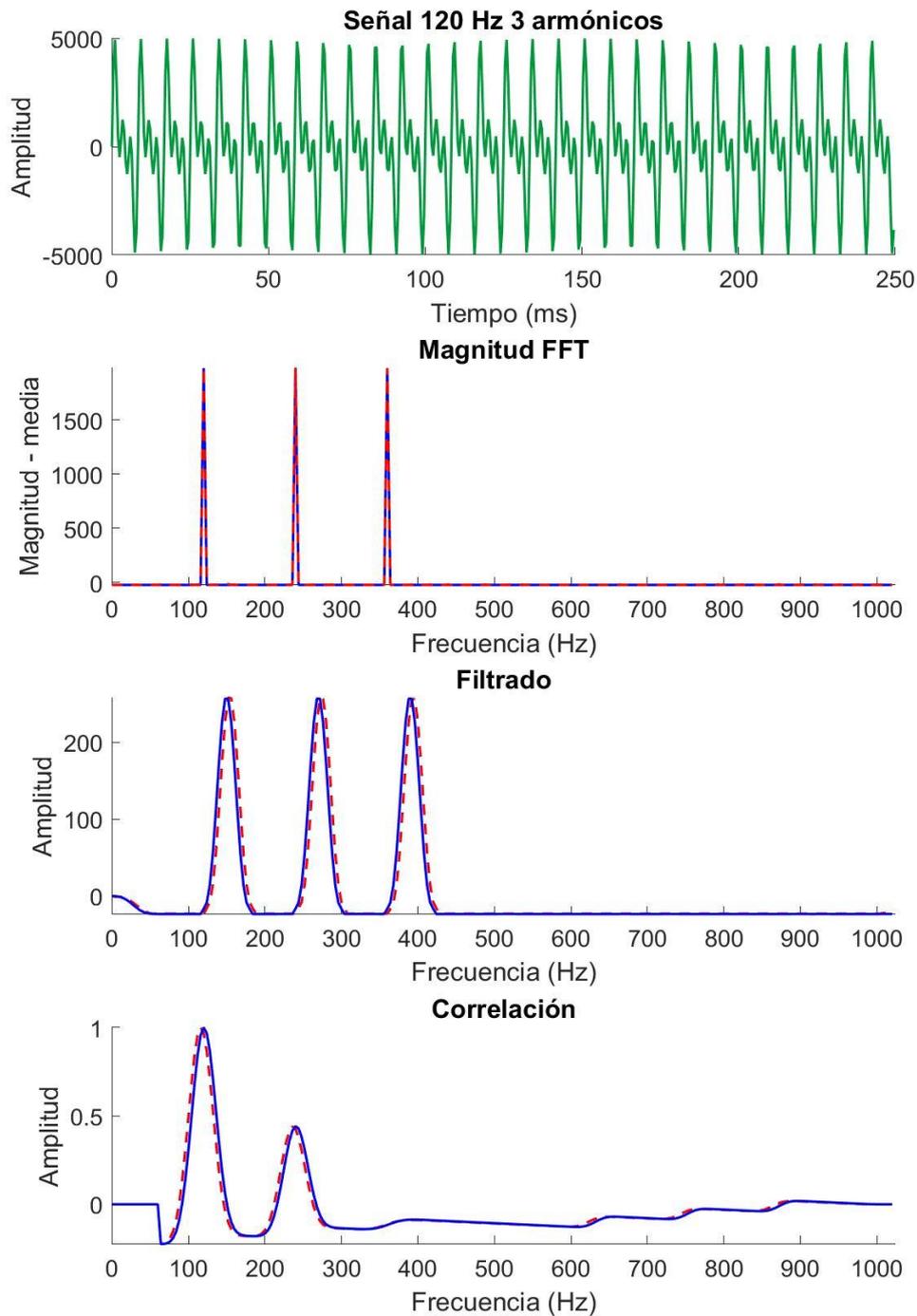


Figura 32: Análisis de la señal de 120 Hz y 3 armónicos por parte del algoritmo usando el dispositivo y Matlab. La línea discontinua de color rojo representa los datos procesados por el dispositivo y la línea continua azul representa los datos procesados en Matlab.

Las siguientes tres señales siguen el mismo recorrido para el cálculo de la F_0 pero con diferencias en el cálculo de Pos_Fft durante el procesado.

El procesado de la señal de 120 Hz con 3 armónicos (Figura 32) muestra el caso en el cual el máximo de la magnitud de la FFT se encuentra en al primer armónico y el máximo de la correlación coincide con el primer pico.

Como el máximo de la correlación coincide con el primer pico de la magnitud de la FFT, para el cálculo de la frecuencia fundamental se toma la posición de este pico de la correlación teniendo en cuenta el retardo de la correlación. El resultado obtenido en Matlab y en el dispositivo es de 120 Hz, que corresponde con el valor esperado.

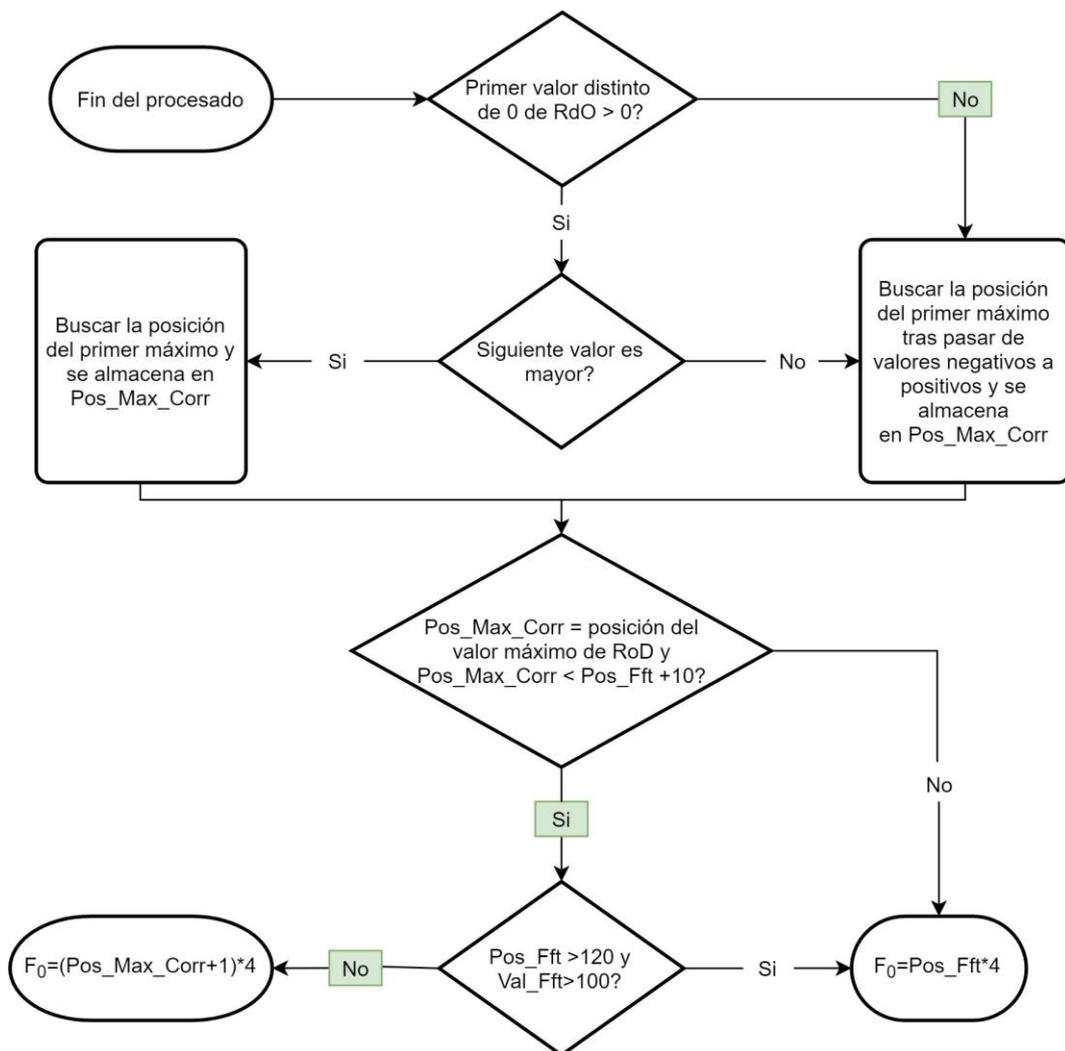


Figura 33: Diagrama del algoritmo para el cálculo de la frecuencia fundamental. En verde se marca el camino que sigue para las señales del caso 2.

La tercera señal (Figura 34) correspondiente a una frecuencia fundamental de 200 Hz y 3 armónicos con distinta relación de amplitudes entre ellos (0.5, 0.7 y 1). En este caso, el máximo de la magnitud de FFT se encuentra en el tercer armónico. Como se explica en el punto 1.2.2.3 del material y métodos en la página 64, durante el procesado se analiza la magnitud de FFT para comprobar si el máximo corresponde con el primer armónico. En esta señal se encuentra otro pico a 1/3 de la posición del máximo y es la posición de éste la que se almacena en `Pos_Fft`. Como el primer pico de la correlación coincide con el máximo de la correlación, el algoritmo se fija en la correlación para el cálculo de la frecuencia (Figura 33). Su procesado es el mismo que el de la señal de 120 Hz, y del mismo modo, tanto Matlab como el dispositivo obtienen la frecuencia esperada de 200 Hz como frecuencia fundamental.

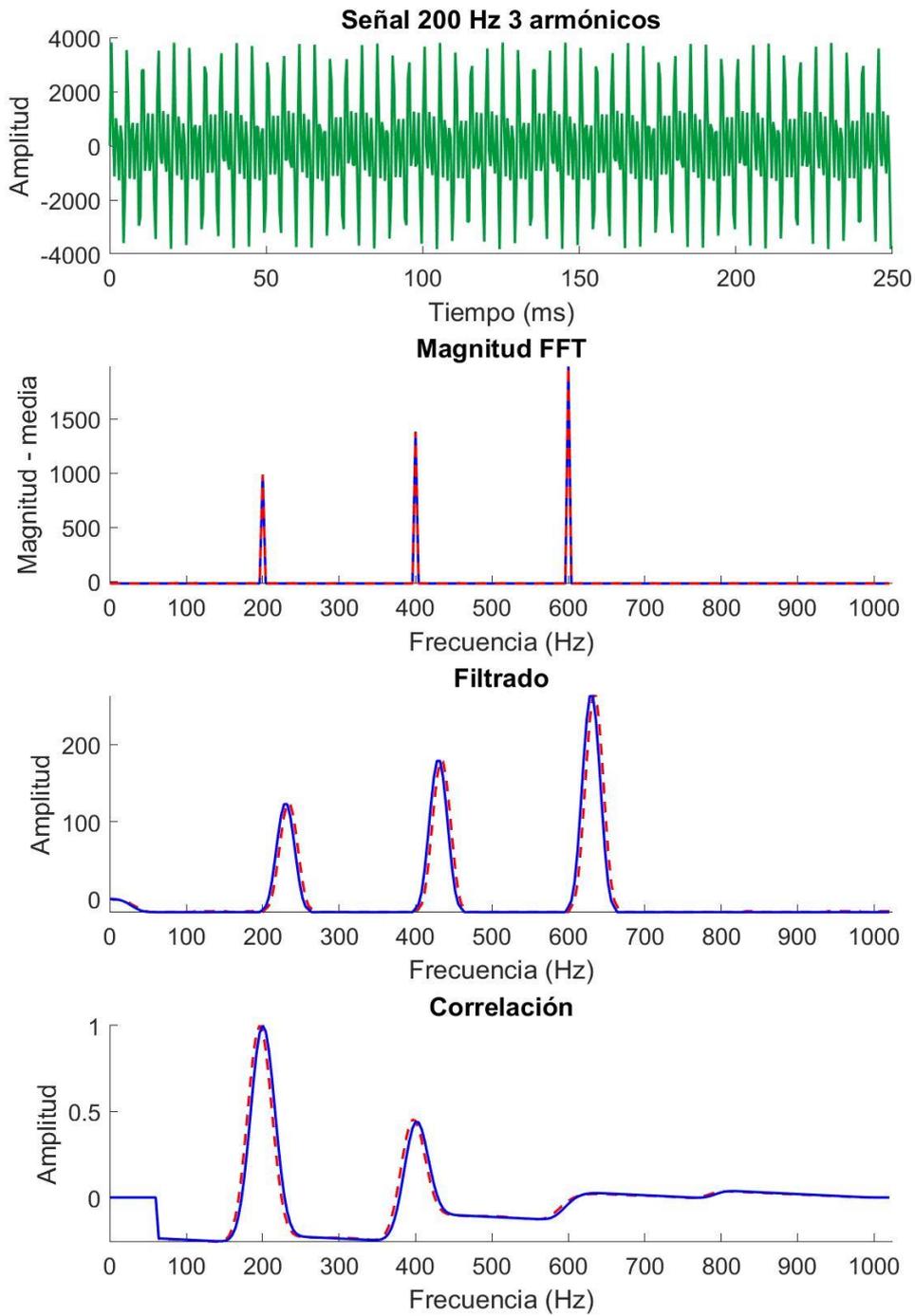


Figura 34: Análisis de la señal de 200 Hz y 3 armónicos por parte del algoritmo usando el dispositivo y Matlab. La línea discontinua de color rojo representa los datos procesados por el dispositivo y la línea continua azul representa los datos procesados en Matlab.

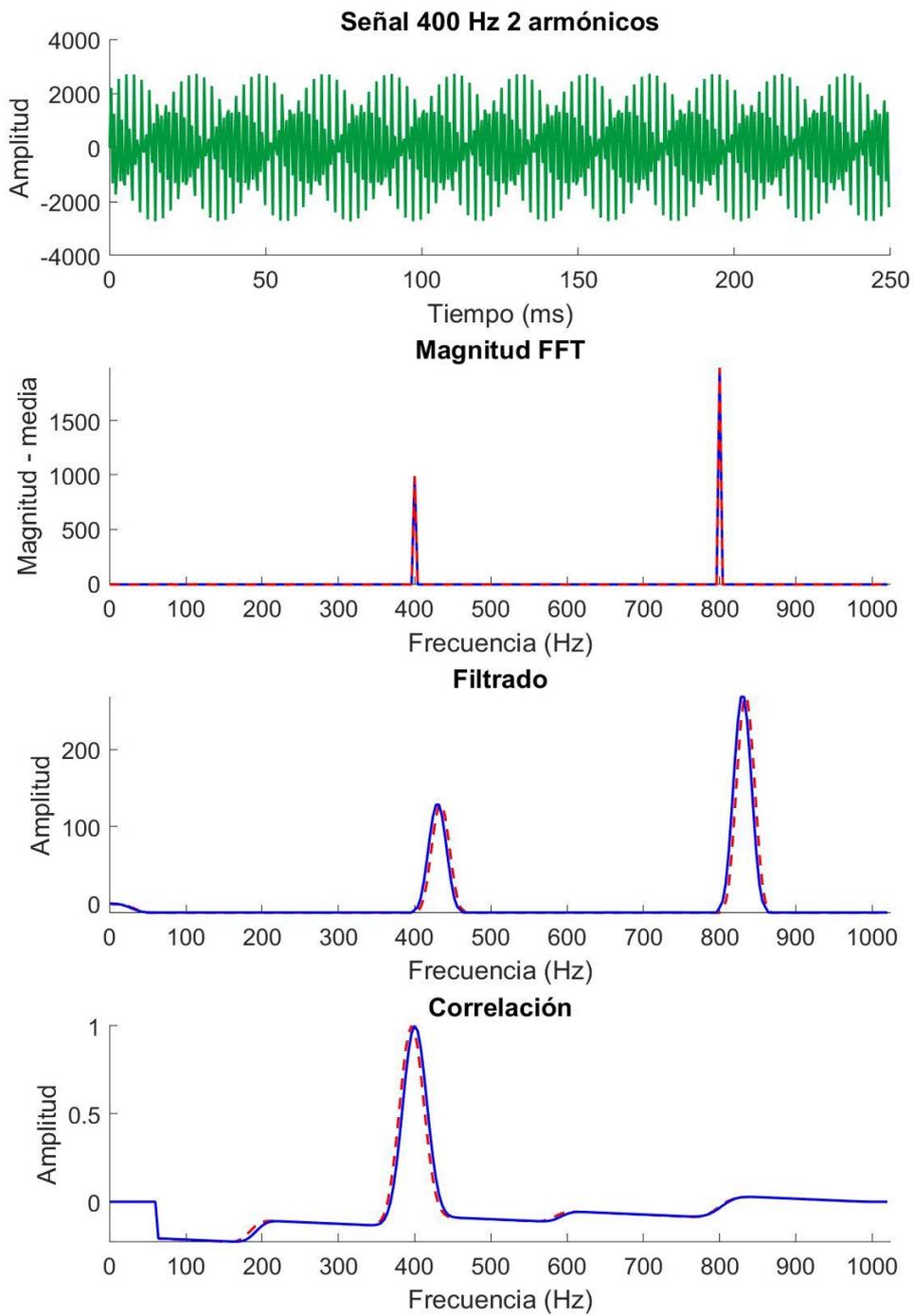


Figura 35: Análisis de la señal de 400 Hz y 3 armónicos por parte del algoritmo usando el dispositivo y Matlab. La línea discontinua de color rojo representa los datos procesados por el dispositivo y la línea continua azul representa los datos procesados en Matlab.

Para la señal de 400 Hz y dos armónicos con relación de amplitud 0,5 y 1 que se muestra en la Figura 35 durante el procesado se encuentra un armónico a $1/2$ de la distancia del máximo de la magnitud de la FFT y es la posición de armónico la que se almacena en `Pos_Fft`. En la correlación se detecta que el primer pico con valor de amplitud positiva corresponde con el máximo de la correlación y se toma como frecuencia fundamental, teniendo en cuenta como siempre, el retraso de la señal de la correlación. Tanto Matlab como el dispositivo calculan correctamente la frecuencia fundamental esperada, 400 Hz.

Caso 3:

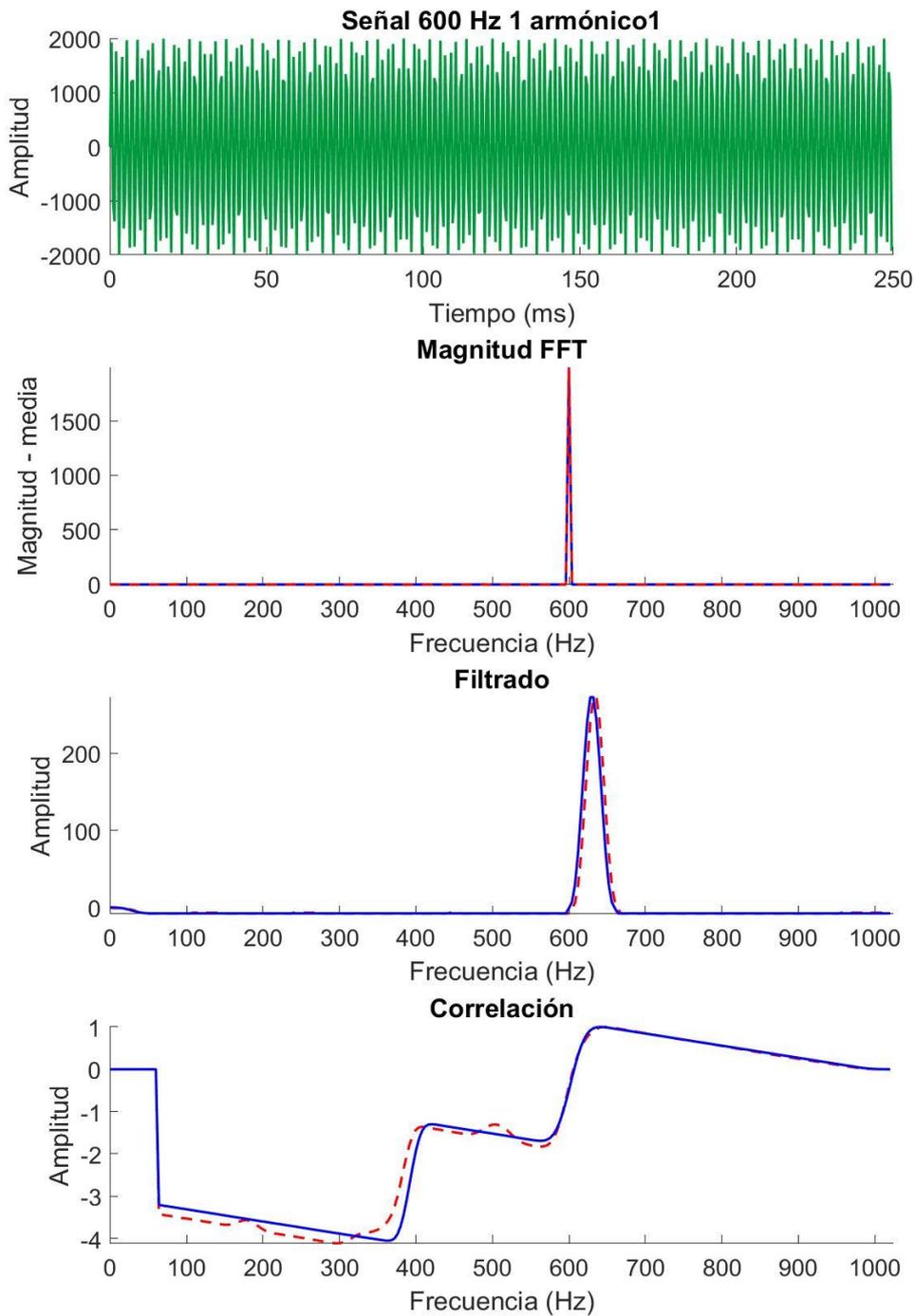


Figura 36: Análisis de la señal de 600 Hz y 3 armónicos por parte del algoritmo usando el dispositivo y Matlab. La línea discontinua de color rojo representa los datos procesados por el dispositivo y la línea continua azul representa los datos procesados en Matlab.

En el caso tres (señal de 600 Hz y un armónico) el algoritmo no utiliza los datos de la correlación dado que el máximo de la magnitud de FFT es mayor a los 480 Hz. Como no se encuentra ningún pico que cumpla las condiciones descritas en el apartado 1.2.2.3, la posición del máximo de la magnitud de la FFT define la frecuencia fundamental (Figura 37). El cálculo en Matlab y en el dispositivo devuelve 600 Hz como resultado, valor que corresponde con el esperado.

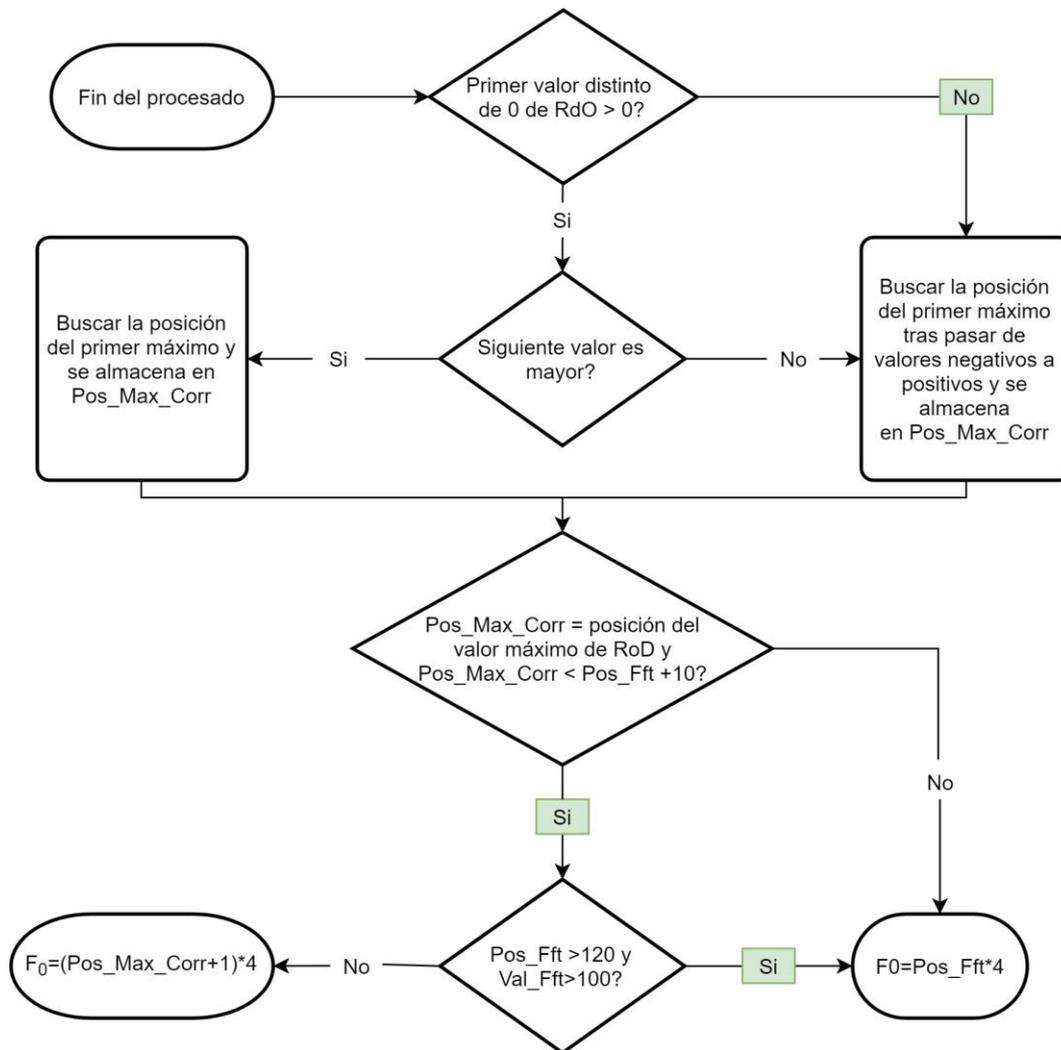


Figura 37: Diagrama del algoritmo para el cálculo de la frecuencia fundamental. En verde se marca el camino que sigue para la señal del caso 3.

Estas cinco señales se analizaron en Matlab y en el dispositivo. Los resultados obtenidos en el cálculo de la frecuencia son los mostrados en la siguiente tabla.

Valor real	Valor Matlab	Valor Dispositivo
90	88	88
120	120	120
200	200	200
400	400	400
600	600	600

Tabla 6: Comparación de frecuencias fundamentales (Hz) esperadas con las obtenidas en el dispositivo y en Matlab usando el algoritmo descrito en esta Tesis.

Como se observa, tanto Matlab y el dispositivo calculan correctamente la frecuencia fundamental de la señal introducida. En el caso de la señal de 90 Hz, ambos métodos obtienen un resultado de 88 Hz. Esto tiene dos lecturas. La primera es que tanto Matlab como el dispositivo coinciden, por lo que se comprueba que los cálculos realizados en el dispositivo son correctos. La segunda lectura tiene relación con la resolución frecuencial derivada de la frecuencia de muestreo y el número de muestras. 90 Hz no es divisible entre 4 Hz, que es la resolución frecuencia que tenemos. Por lo tanto, el algoritmo devuelve el valor más cercano divisible, en este caso 88.

2.2. Validación de los parámetros de interés en voz humana

2.2.1. Validación de la frecuencia fundamental

A continuación, se muestran los resultados obtenidos durante los registros realizados con el software SoundScope y el dispositivo para la medición de la frecuencia fundamental.

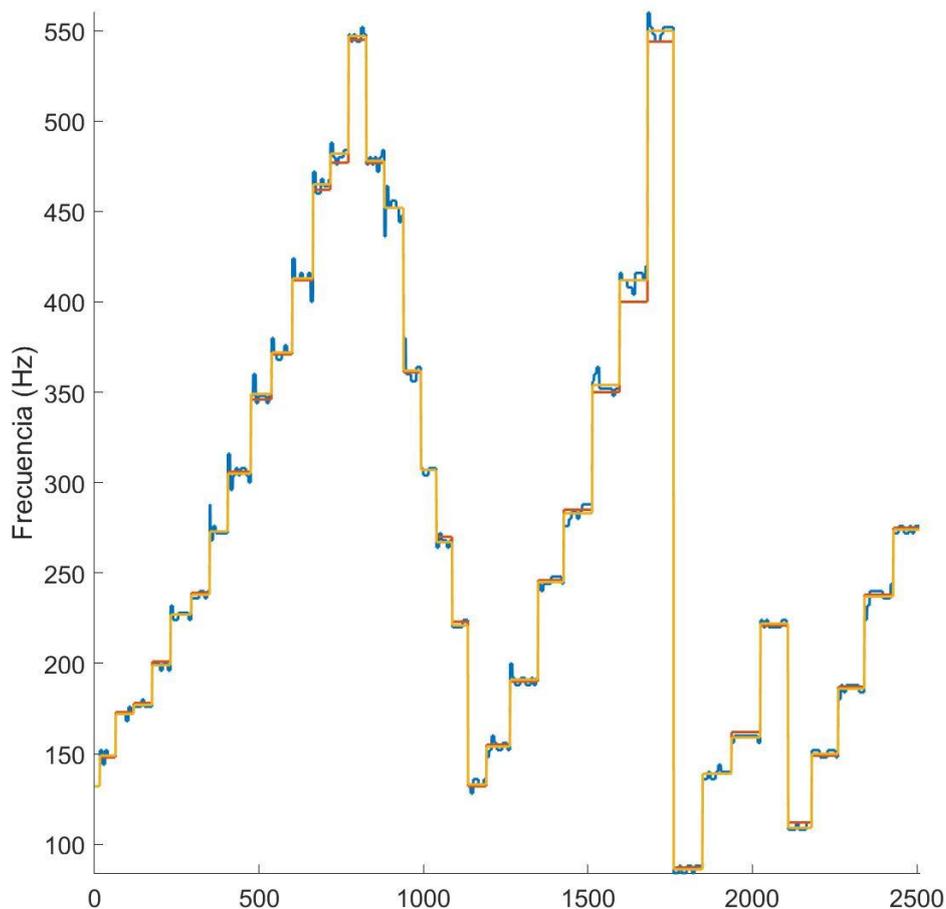


Figura 38: Gráfica de 38 escalones tonales registrados por el dispositivo y el SoundScope. En azul se muestran los registros del dispositivo, en naranja los del SoundScope y en amarillo los valores del dispositivo mediados por escalón tonal.

En la Figura 38 se muestran los 2508 datos recogidos por el dispositivo en los 38 escalones tonales. Los primeros 29 escalones corresponden al sujeto 1 tomados en dos mediciones, la primera de 21 escalones tonales y la segunda de 8. Entre cada una de ellas se quitó y colocó el dispositivo de su posición sobre el espacio supraesternal. Los siguientes cuatro escalones tonales corresponden al sujeto 2 y los últimos cinco al sujeto 3. Se observa que los valores devueltos por el SoundScope se mantienen estables durante todo el escalón tonal. Esto es debido a que de SoundScope solo se toma un valor de frecuencia por cada escalón tonal que corresponde con la frecuencia media del escalón tonal. Los resultados proporcionados por el dispositivo concuerdan con los proporcionados

por SoundScope. Además, debido a la resolución temporal del dispositivo desarrollado, permite observar cómo en el ataque (al iniciar un tono) la frecuencia fundamental es ligeramente mayor y posteriormente se estabiliza.

Para que la comparación fuese más objetiva, de los datos obtenidos por el dispositivo se calculó la frecuencia fundamental media obtenida por cada escalón tonal.

En la Figura 39 se muestra la correlación entre los valores de frecuencia fundamental obtenidos usando el SoundScope y los valores de frecuencia fundamental obtenidos con el dispositivo.

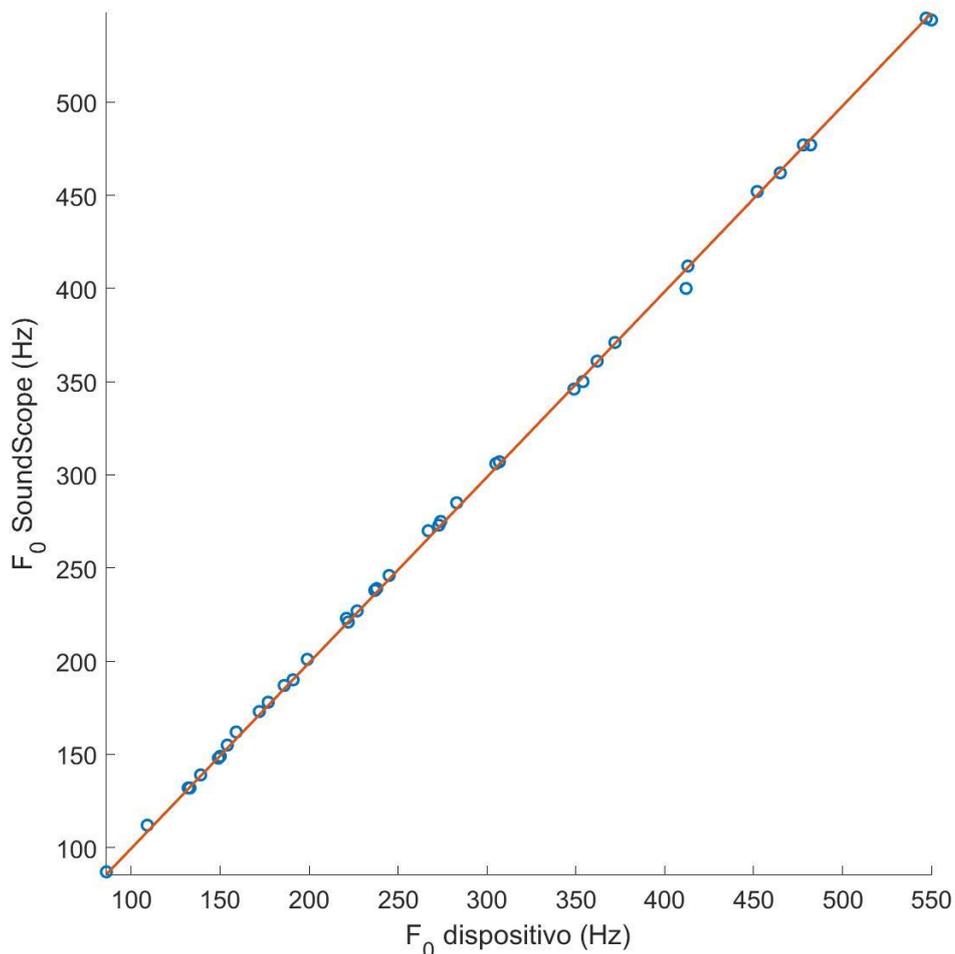


Figura 39: Diagrama de dispersión que correlaciona los valores de la frecuencia fundamental registrados por el SoundScope y la frecuencia fundamental media por escalón tonal calculada por el dispositivo.

La correlación de estos valores es $R = 0.9998$.

Por último, se calcula el error absoluto del dispositivo tomando como referencia el valor del SoundScope (Figura 40).

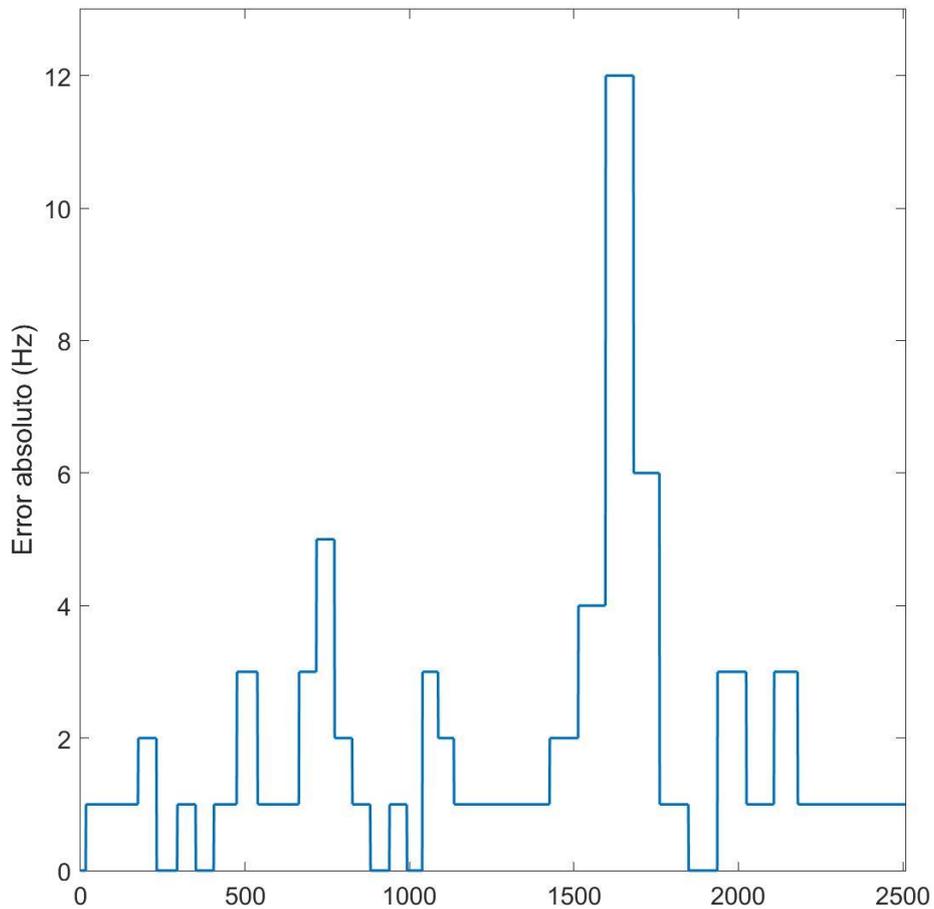


Figura 40: Cálculo del error absoluto entre lo registrado con el SoundScope y la media por escalón tonal registrada por el dispositivo.

El error absoluto medio es de 1.94 Hz con una desviación estándar de 2.32 Hz.

2.2.2. Validación de la intensidad relativa

Como resultado de las grabaciones realizadas para la validación de la intensidad relativa se tomaron 77 mediciones de intensidad constante del sujeto 1. De cada uno de estos valores se anotó el valor del sonómetro (77 valores), mientras que el dispositivo arrojó 5190 valores correspondientes al mismo periodo.

Del sujeto 2 se tomaron 8 mediciones de intensidad que corresponden a 646 datos en el dispositivo.

Las mediciones del sujeto 3 se descartaron debido a que no realizó el ejercicio de la forma establecida en el método. No mantuvo una vocal a intensidad constante.

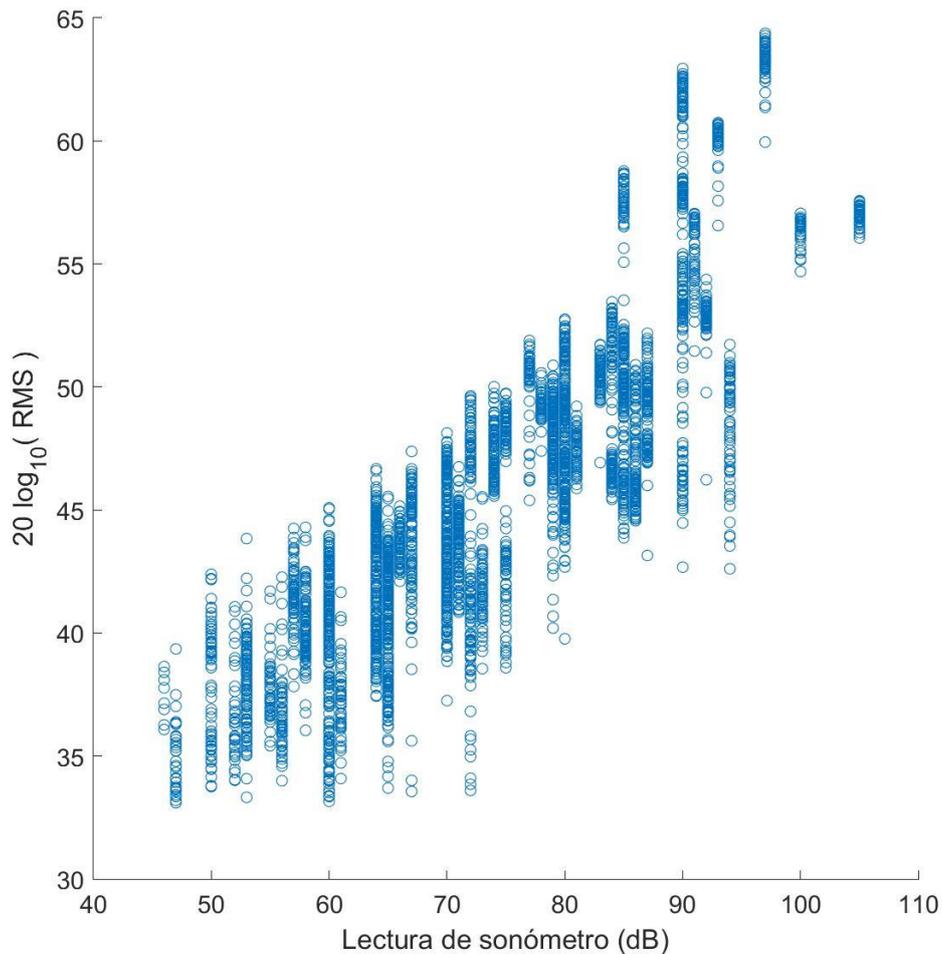


Figura 41: Diagrama de dispersión del logaritmo de los valores de intensidad relativa obtenidos por el dispositivo (RMS) y la intensidad registrada por el sonómetro. Registro tomado con el sujeto 1.

En la Figura 41, tras el cálculo del logaritmo, se observa una correlación lineal cuyo coeficiente de correlación es de $R = 0.86$. Este resultado tiene más relevancia cuando se tiene en cuenta que con el sujeto 1 las mediciones se realizaron días distintos. Por lo que el hecho de quitarse y ponerse el dispositivo,

con la posible pequeña variación en la tensión del colgante, no afecta a la correlación de los datos.

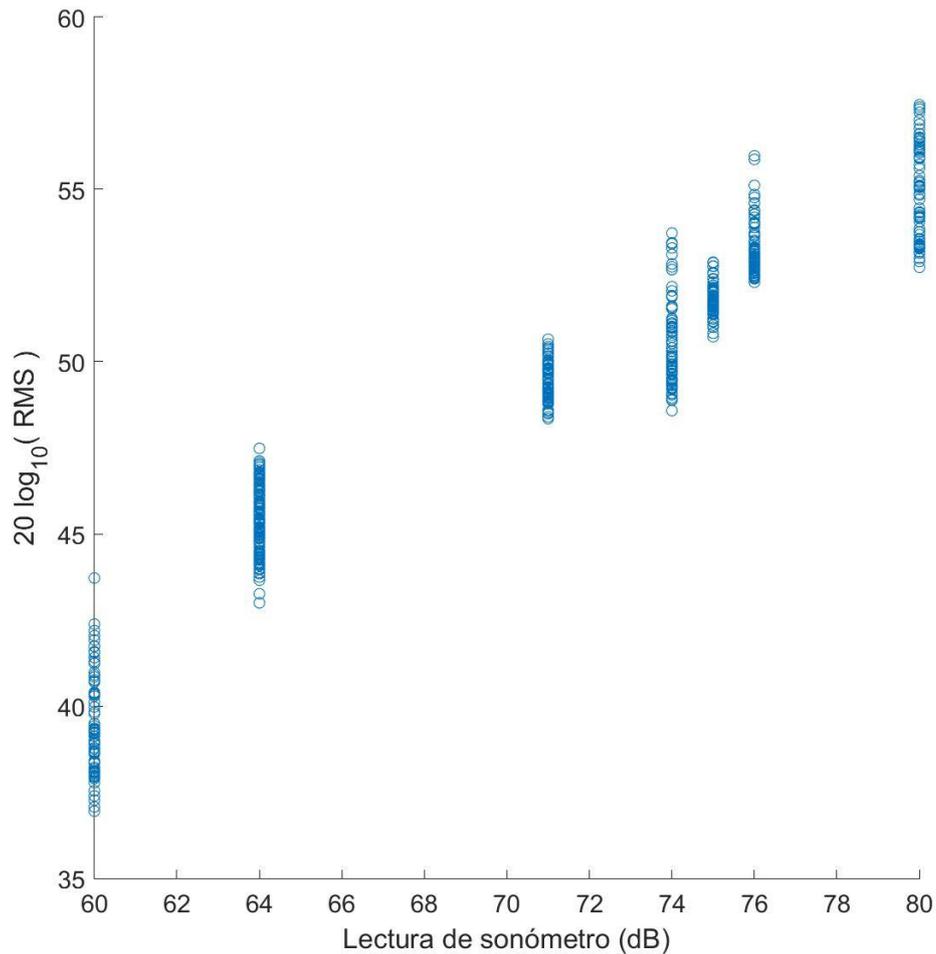


Figura 42: Diagrama de dispersión del logaritmo de los valores de intensidad relativa obtenidos por el dispositivo (RMS) y la intensidad registrada por el sonómetro. Registro tomado con el sujeto 2.

Por último, se muestra la gráfica de dispersión de la grabación realizada con el sujeto 2. La correlación lineal de los datos del dispositivo con los del sonómetro es $R = 0.96$ (Figura 42).

2.2.3. Validación del tiempo de fonación

Los resultados de las grabaciones se muestran a continuación en la Tabla 7. Como se observa, sólo en el caso de la grabación del sujeto 1 el tiempo difiere más del 80 % usando micrófono y Audacity, y el tiempo registrado por el dispositivo.

Grabación	Audacity (s)	Dispositivo (s)	Relación (%)
Sujeto 1	81	35	43
Sujeto 2	74	60	81
Sujeto 3	75	69	92
Vocales	7.7	7.9	103

Tabla 7: Comparación del cálculo o medición de los tiempos de fonación con diferentes sistemas: Audacity y el dispositivo.

Los resultados de la Tabla 7 son coherentes con el fundamento del dispositivo. El dispositivo está diseñado para cuantificar el tiempo de fonación. Como se ha definido en el capítulo V punto 2.3, el tiempo de fonación solo incluye los momentos en los que las cuerdas vocales vibran. Sin embargo, la grabación de la señal acústica mediante micrófono no permite, usando Audacity, la distinción entre sonidos sonoros y sordos (con vibración de las cuerdas vocales o sin vibración). Por lo tanto, la medición realizada usando micrófono contabiliza sonidos sordos y sonoros, obteniéndose el tiempo total mayor. La discrepancia de la relación entre dispositivo y Audacity en el Sujeto 1 se estudió minuciosamente y se determinó que el dispositivo dejó de hacer contacto con la piel por una incorrecta colocación.

En la grabación en la que únicamente se registra voz generada al emitir vocales y pausas, todo es voz vocálica. En ese caso coinciden los valores calculados por el dispositivo con los obtenidos a partir de la grabación acústica (Figura 43).

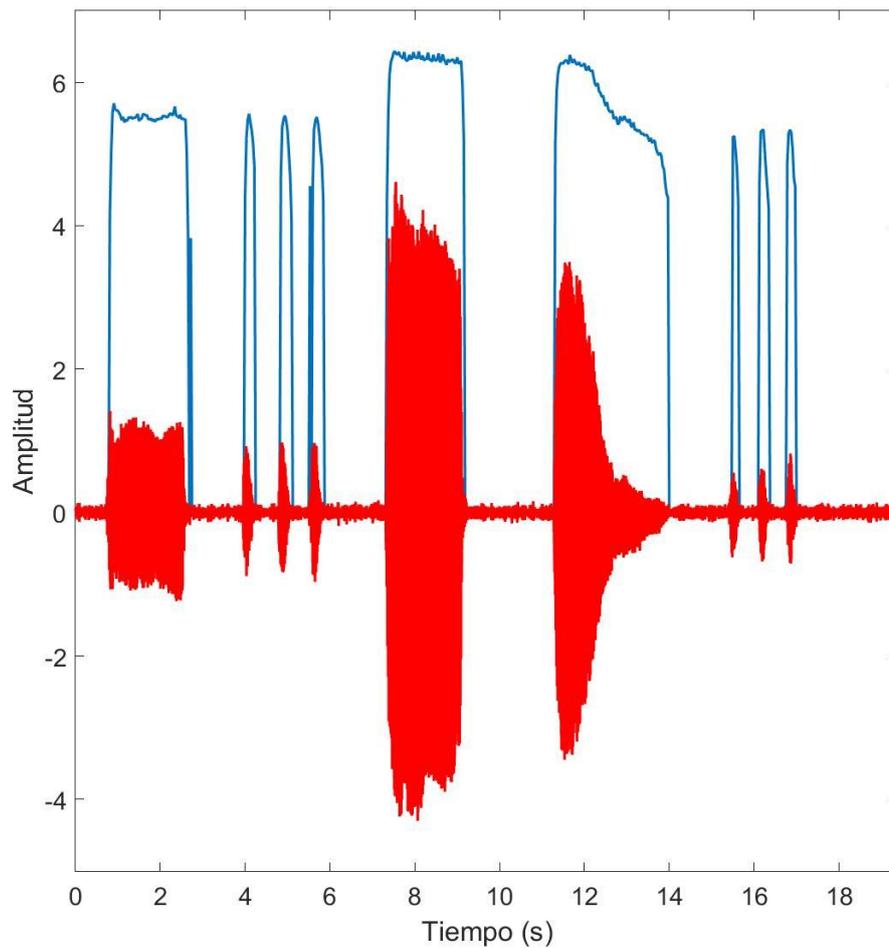


Figura 43: Comparación señal acústica (rojo) y medición de intensidades por el dispositivo (azul) del registro de vocales.

3. REGISTRO DE VOZ DURANTE LARGOS PERIODOS DE TIEMPO

Para organizar los resultados y poder explicarlos de una manera más ordenada y clara, se ha optado por separar los resultados obtenidos del sujeto 1 de los obtenidos del sujeto 2.

3.1. Sujeto 1

Los registros del sujeto 1 se llevaron a cabo durante 14 días. La duración de estas grabaciones se muestra en la Figura 44.

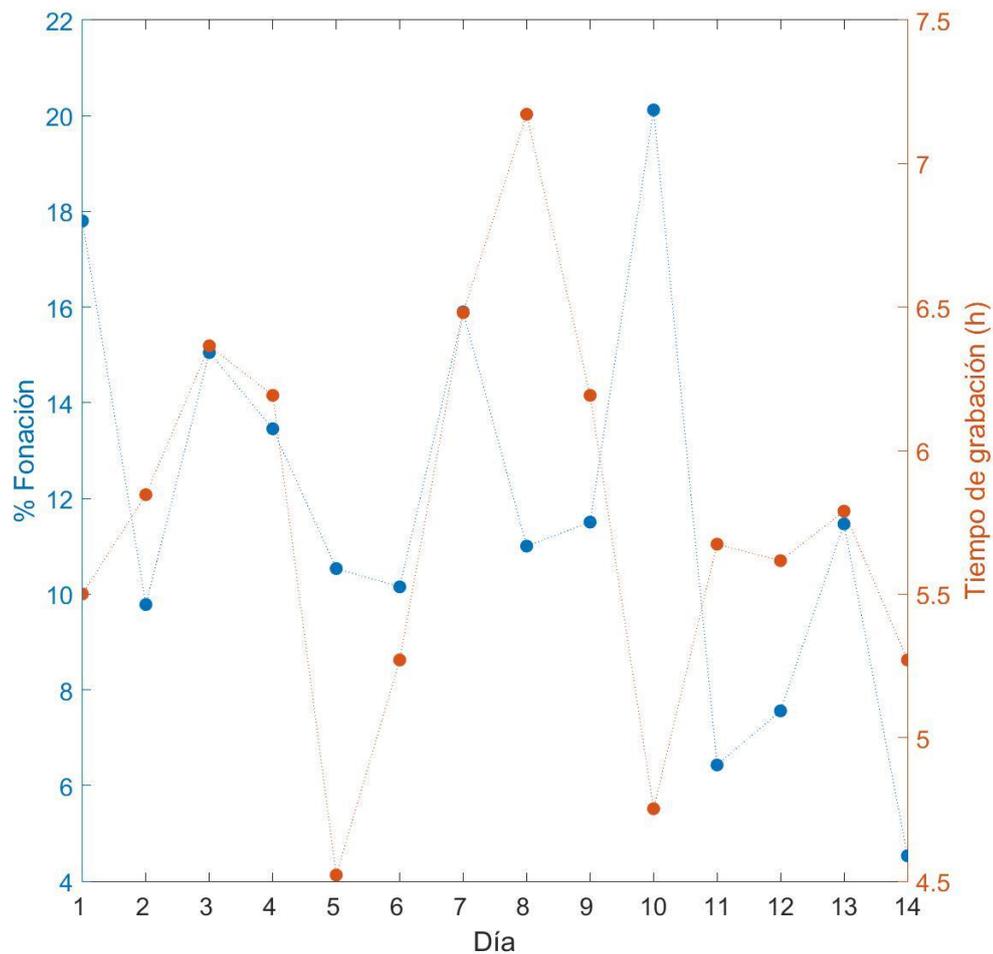


Figura 44: Registros de larga duración del sujeto 1. En rojo: se muestra el tiempo de grabación y en azul el porcentaje de ese tiempo que el dispositivo ha contabilizado como fonación.

El tiempo de registro oscila entre 4.5 y 7.1 horas diarias y el porcentaje de tiempo diario de fonación entre 4.5 y 20 %.

Los datos registrados durante los 14 días se han condensado en las dos siguientes figuras (Figura 45 y Figura 46) donde se muestran los histogramas de frecuencias fundamentales e intensidades relativas.

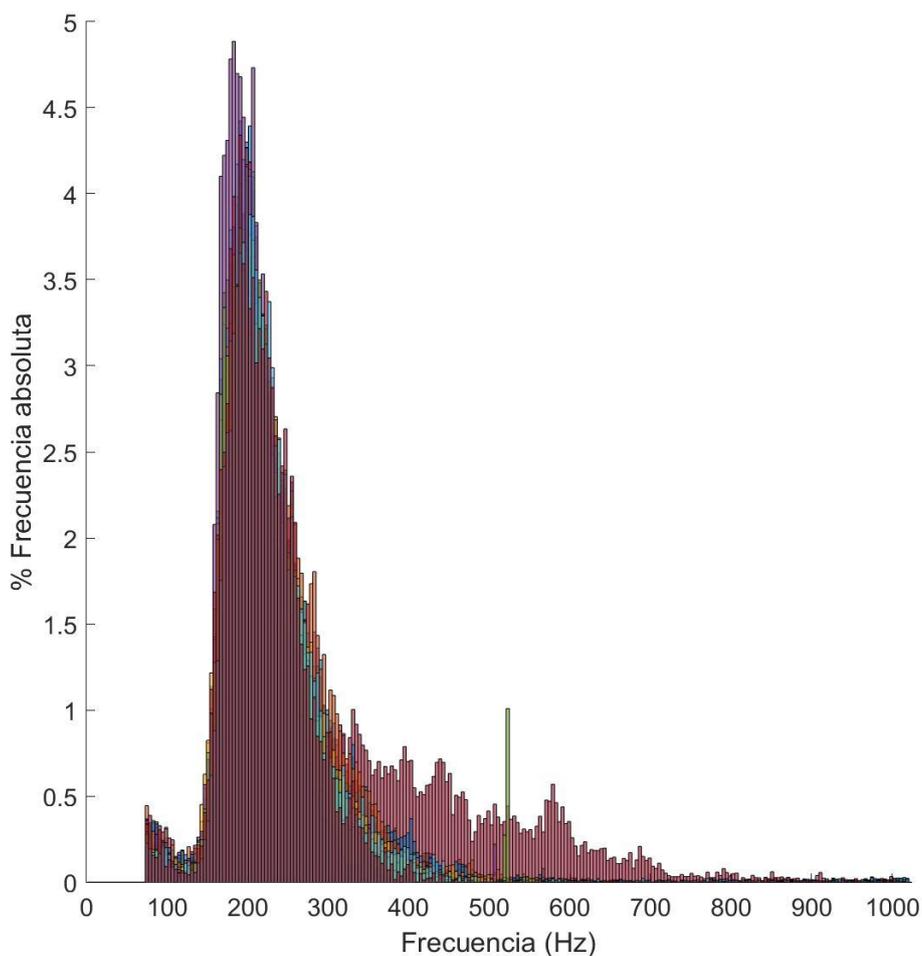


Figura 45: Histogramas de frecuencias de los 14 días de registro del sujeto 1.

En la Figura 45 se muestran los histogramas de frecuencias de los 14 días de registro del sujeto 1. La superposición de unos sobre otros impide verlos todos, pero este hecho hace referencia a la poca variabilidad de las frecuencias fundamentales registradas a lo largo de los distintos días.

Es necesario destacar que uno de los histogramas muestra frecuencias más elevadas que los demás. Este histograma coincide con el día que el sujeto 1 llevó a cabo un patrón anormal, en concreto una sesión de canto. Más adelante se analizará más en detalle este día.

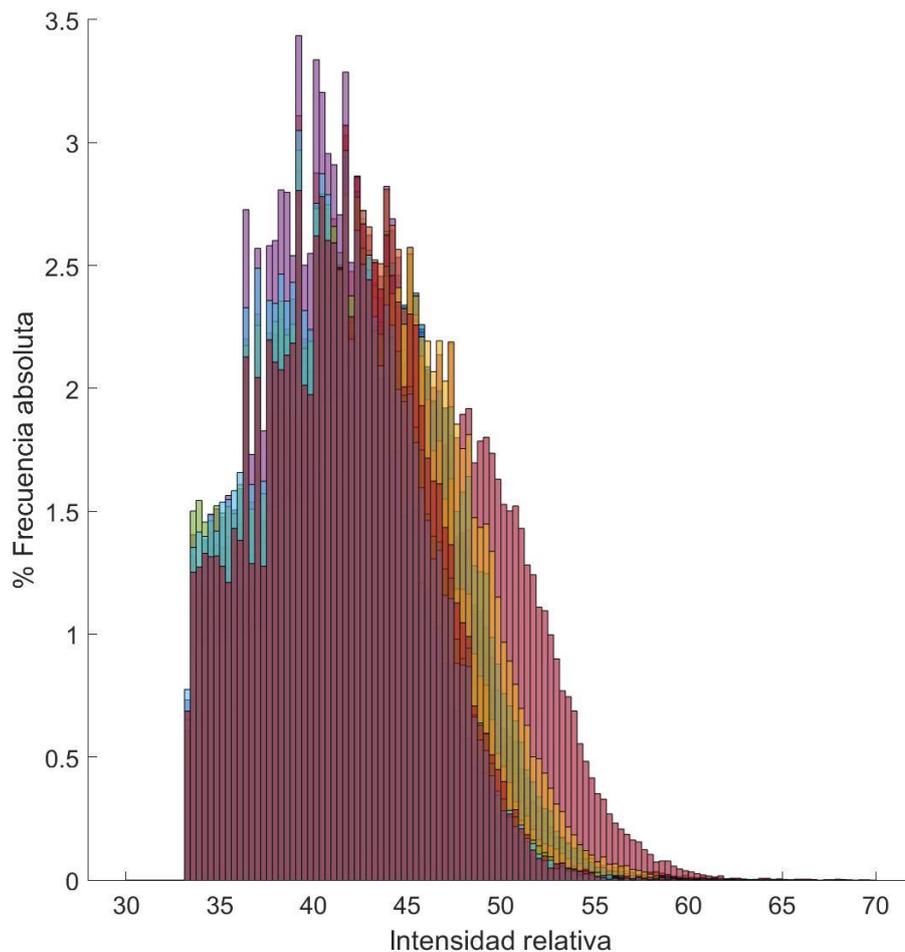


Figura 46: Histogramas de intensidad relativa de los 14 días de registro del sujeto 1.

Al igual que en la Figura 45, en la Figura 46 la superposición de los histogramas impide tener una visión clara de los días por separado. Aun así, se puede observar de nuevo que uno de los días registró intensidades más elevadas que los demás. Al igual que se ha mencionado anteriormente, este día coincide con el día en el que el sujeto 1 llevó a cabo un patrón anormal.

En estos histogramas de intensidades relativas se ve mayor variabilidad entre días en comparación con el histograma de frecuencias.

A continuación, se analizan más en detalle dos días: un día normal y un día con patrón anormal. La comparación entre todos ellos se encuentra en el ANEXO XII.

3.1.1. Análisis día normal del sujeto 1

A continuación, se analiza con detenimiento el día 1, cuyo patrón de voz fue normal.

Como se ha mostrado con anterioridad en el conjunto de los días, se puede visualizar el histograma de frecuencias fundamentales registradas durante el día.

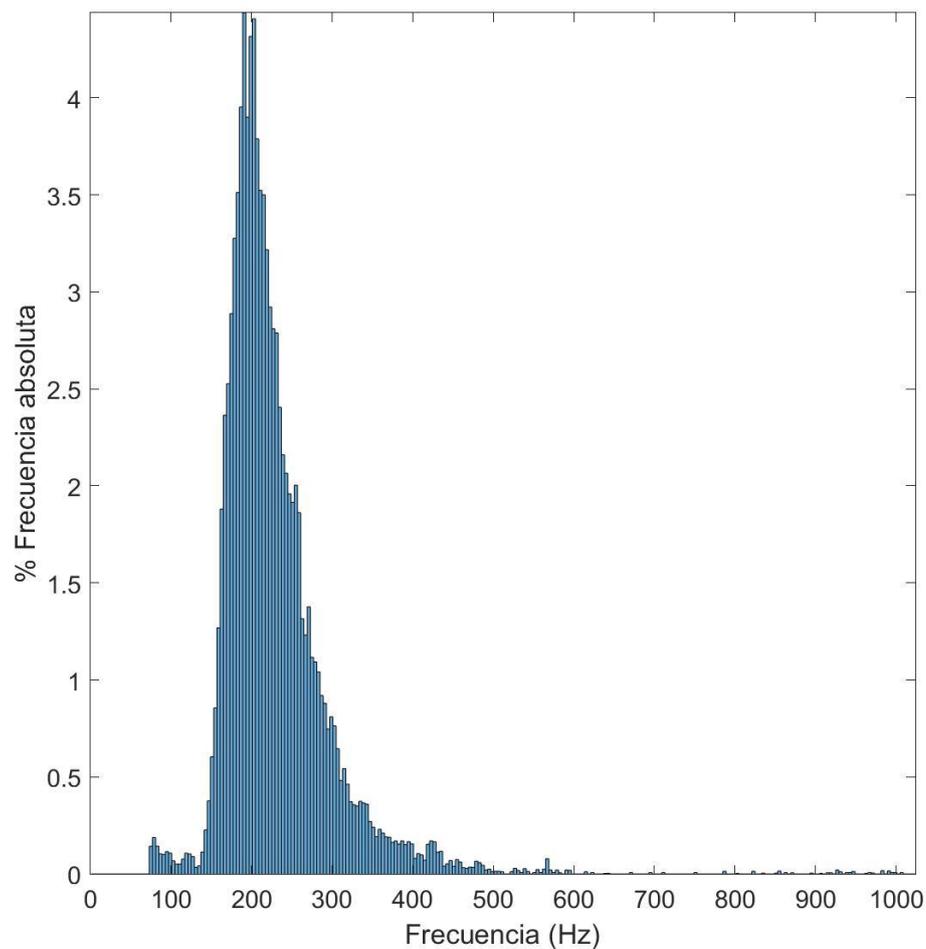


Figura 47: Histograma de frecuencias fundamentales del día 1 del sujeto 1.

La Figura 47 muestra el histograma de las frecuencias fundamentales a lo largo de la jornada laboral del sujeto 1. En el boxplot (Figura 48) se muestra la mediana (216 Hz) y el primer y tercer cuartil (188 Hz y 256 Hz respectivamente). Los valores representados en rojo, aunque visualmente aparenten ser muchos, corresponden a menos del 5 % de los datos.

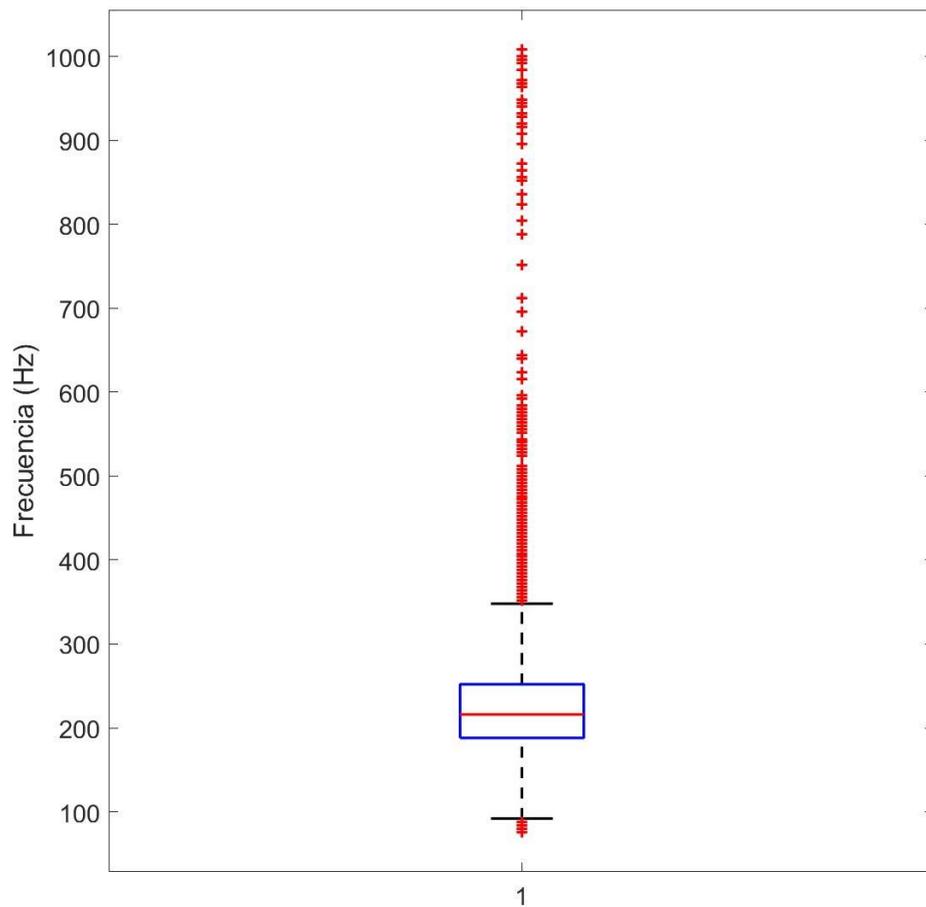


Figura 48: Boxplot de frecuencias fundamentales del día 1 del sujeto 1.

Del mismo modo se puede visualizar el histograma de intensidades relativas (Figura 49). En el boxplot (Figura 50) de intensidades relativas se muestra la mediana (42.61) y el primer y tercer cuartil (39.18 y 45.76 respectivamente).

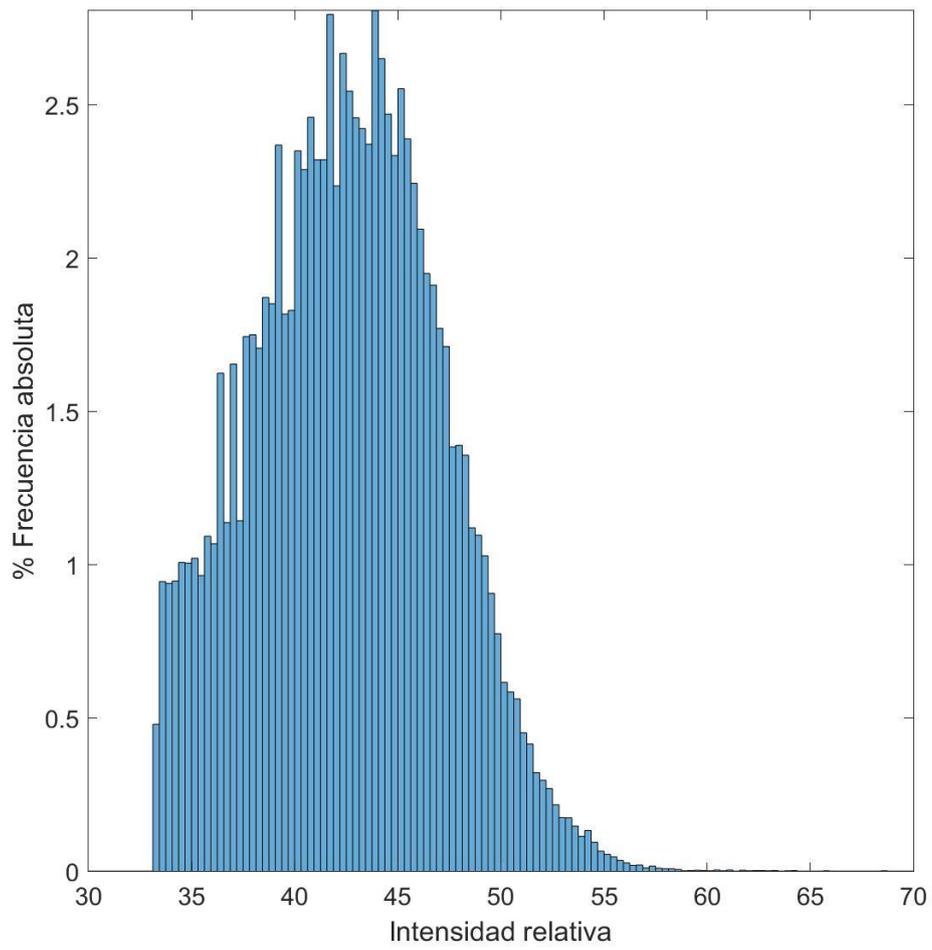


Figura 49: Histograma de intensidades relativas del día 1 del sujeto 1.

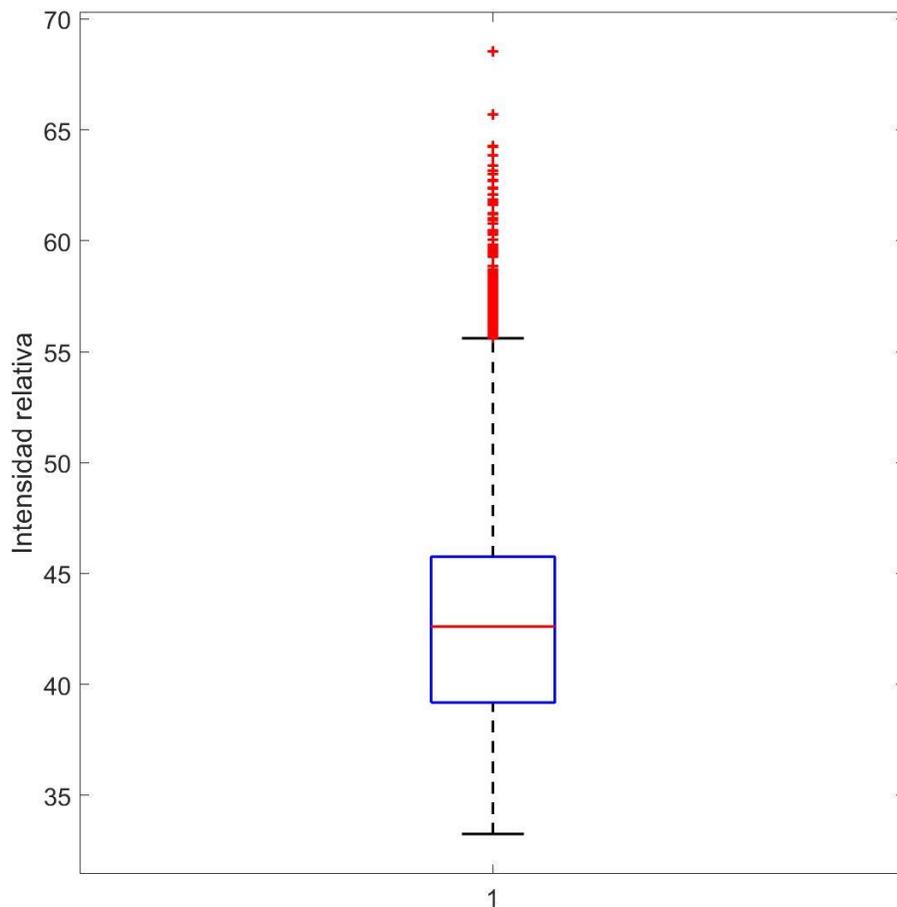


Figura 50: Boxplot de intensidades relativas del día 1 del sujeto 1.

La mejor manera de combinar la información de frecuencias fundamentales e intensidades relativas es el fonetograma (Figura 51). En el fonetograma se muestra en el eje de abscisas las frecuencias fundamentales y en el de ordenadas las intensidades relativas. De este modo se representa mejor cuál es el rango vocal y en qué frecuencia fundamental e intensidad relativa se concentra el tiempo de fonación.

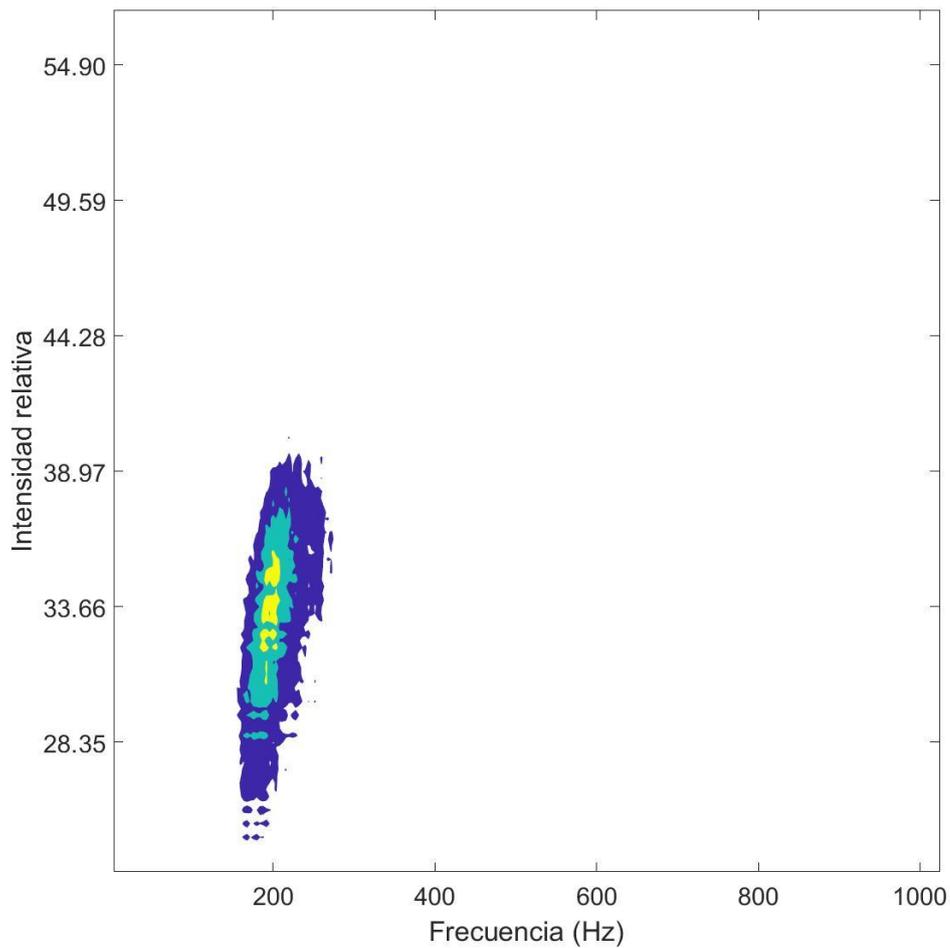


Figura 51: Fonetrograma del día 1 del sujeto 1. Los colores dividen las regiones según el tiempo acumulado con esa frecuencia fundamental e intensidad relativa. La región amarilla representa los puntos en que la proporción entre el tiempo en ese punto y el máximo está entre el 75 y el 100 %, en verde entre el 50 y el 75 % y en azul entre el 25 y el 50 %.

El dispositivo posibilita el análisis de los dos parámetros a lo largo del tiempo. Así se pueden mostrar la evolución de la frecuencia fundamental y de la intensidad relativa durante el registro.

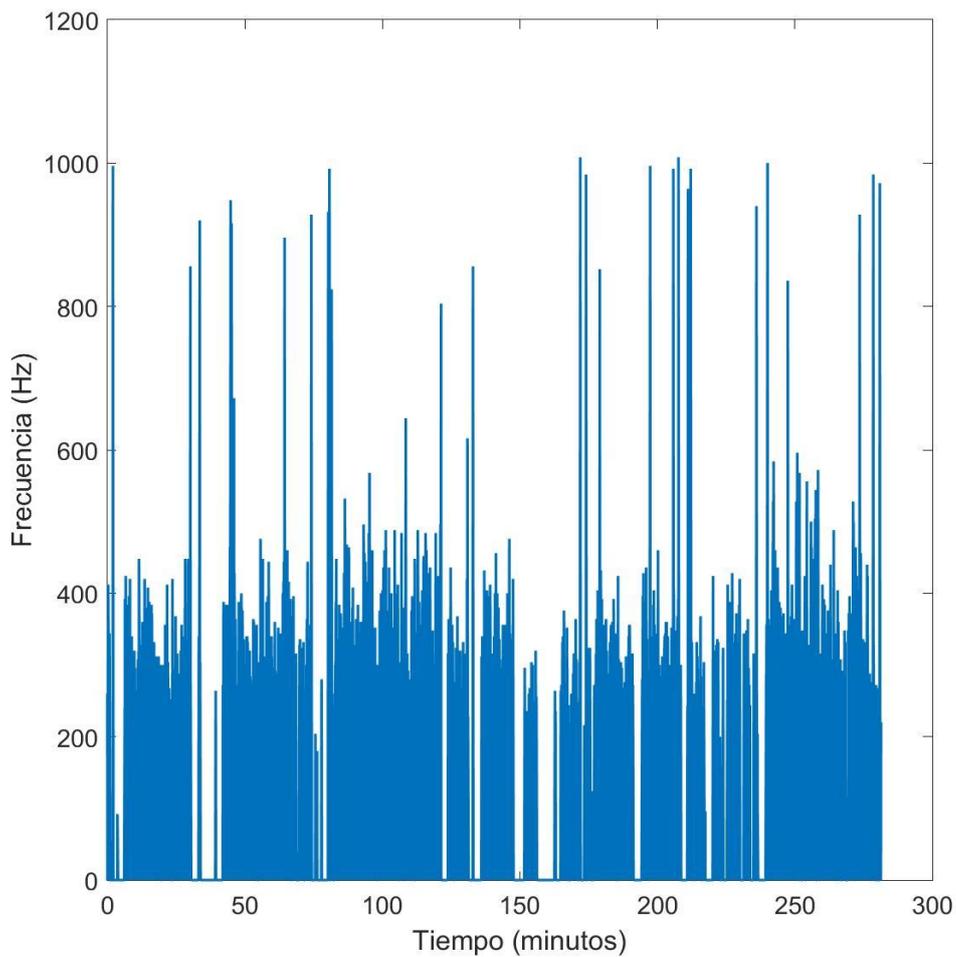


Figura 52: Gráfica de la evolución temporal de la frecuencia fundamental a lo largo del registro del día 1 del sujeto 1.

En la Figura 52 se puede observar la evolución de la frecuencia fundamental a lo largo del tiempo de registro. Se observa que la frecuencia fundamental se concentra entre los 200 y los 300 Hz. También destacan los periodos de registro donde la frecuencia fundamental es 0 Hz. En estos periodos la fonación es nula y por lo tanto son los periodos de pausa. Además, se observan picos elevados correspondientes a registros de frecuencia superiores a 600 Hz. Probablemente estos serán errores de cálculo de la frecuencia dado que la fonación a 1000 Hz de forma espontánea es extremadamente excepcional en la voz hablada. Al adoptar la representación gráfica de estos valores se destacan mucho estas magnitudes aisladas y por ello, se opta por calcular la frecuencia mediana por segmentos temporales de 5 minutos.

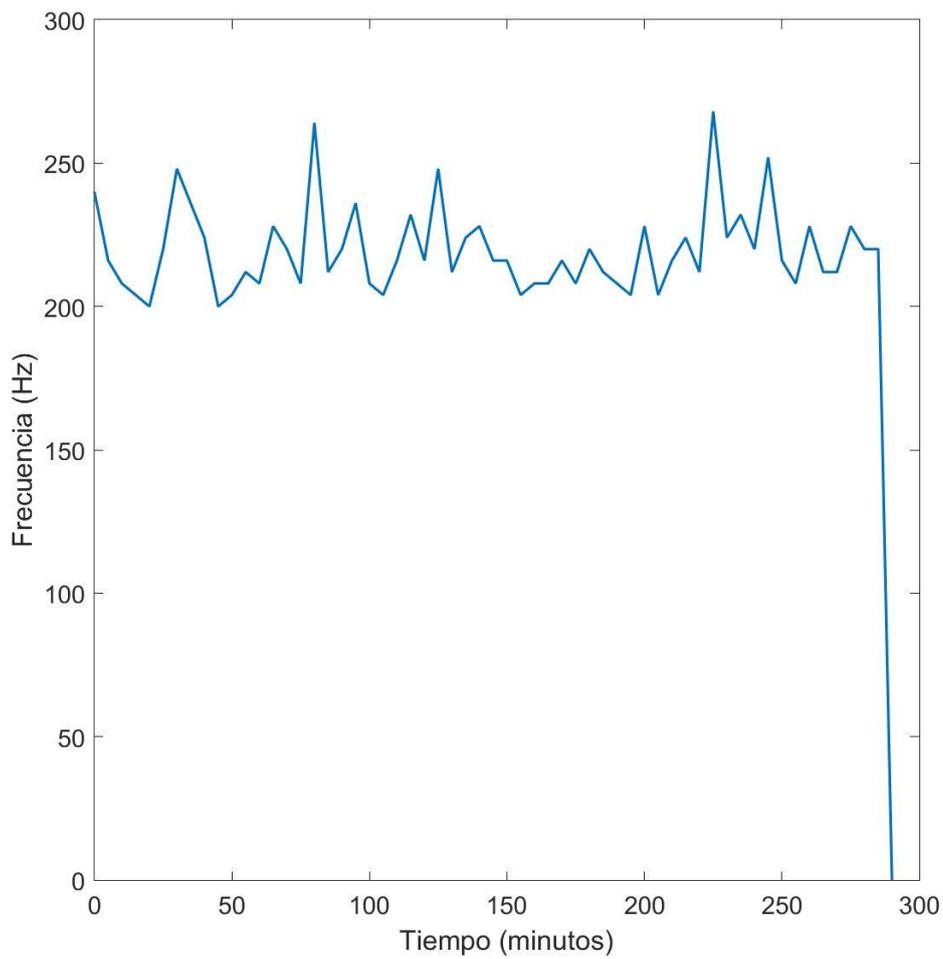


Figura 53: Gráfica de la evolución temporal de la mediana de las frecuencias fundamentales en segmentos de 5 minutos a lo largo del registro del día 1 del sujeto 1.

En la Figura 53 se observa mejor la evolución de la frecuencia fundamental a lo largo del tiempo de registro.

De la misma manera se presenta la intensidad relativa en la Figura 54 y Figura 55.

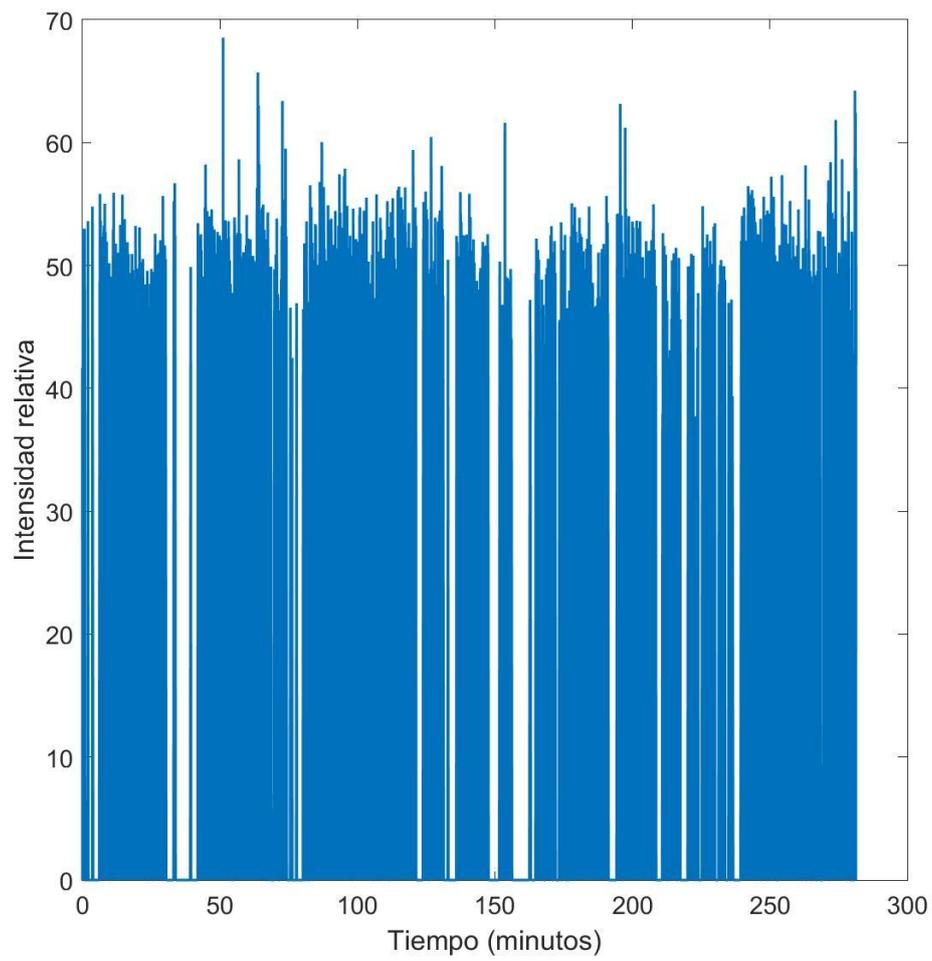


Figura 54: Gráfica de la evolución temporal de la intensidad relativa a lo largo del registro del día 1 del sujeto 1.

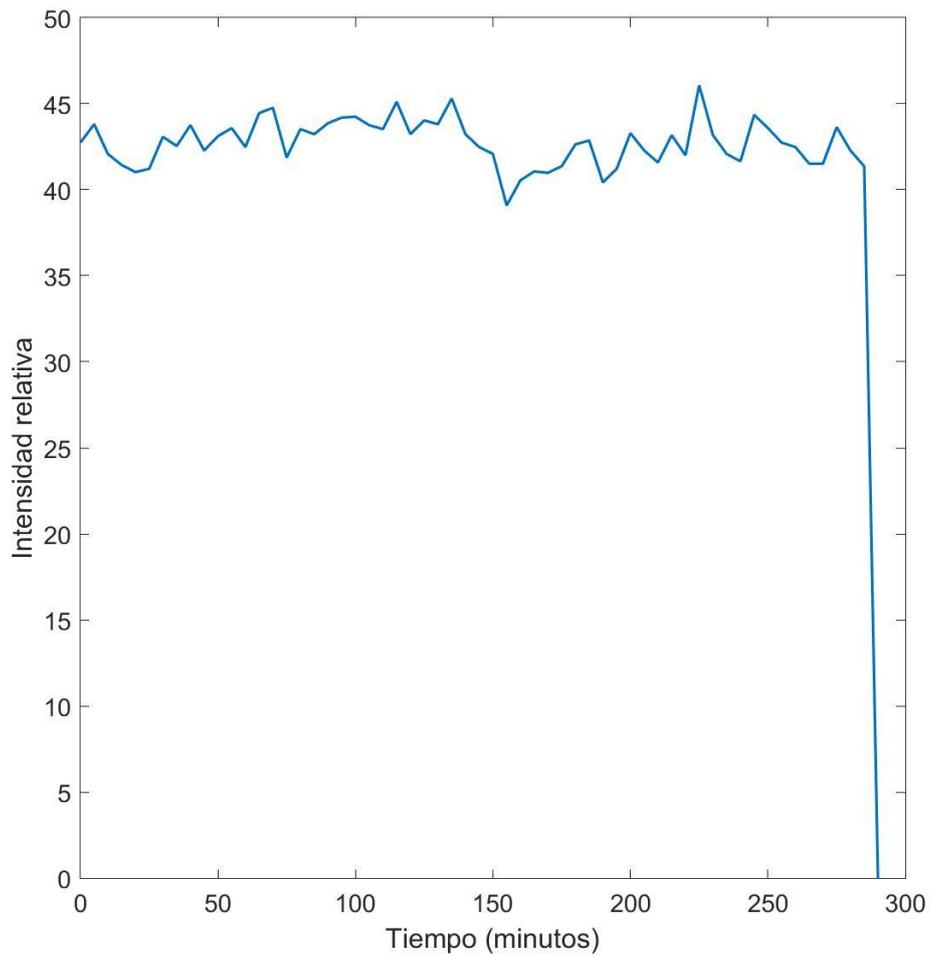


Figura 55: Gráfica de la evolución temporal de la mediana de las intensidades relativas en segmentos de 5 minutos a lo largo del registro del día 1 del sujeto 1.

Utilizando los mismos segmentos de 5 minutos, se calcula el porcentaje de tiempo de fonación en dichos segmentos y se representa a lo largo del tiempo (Figura 56).

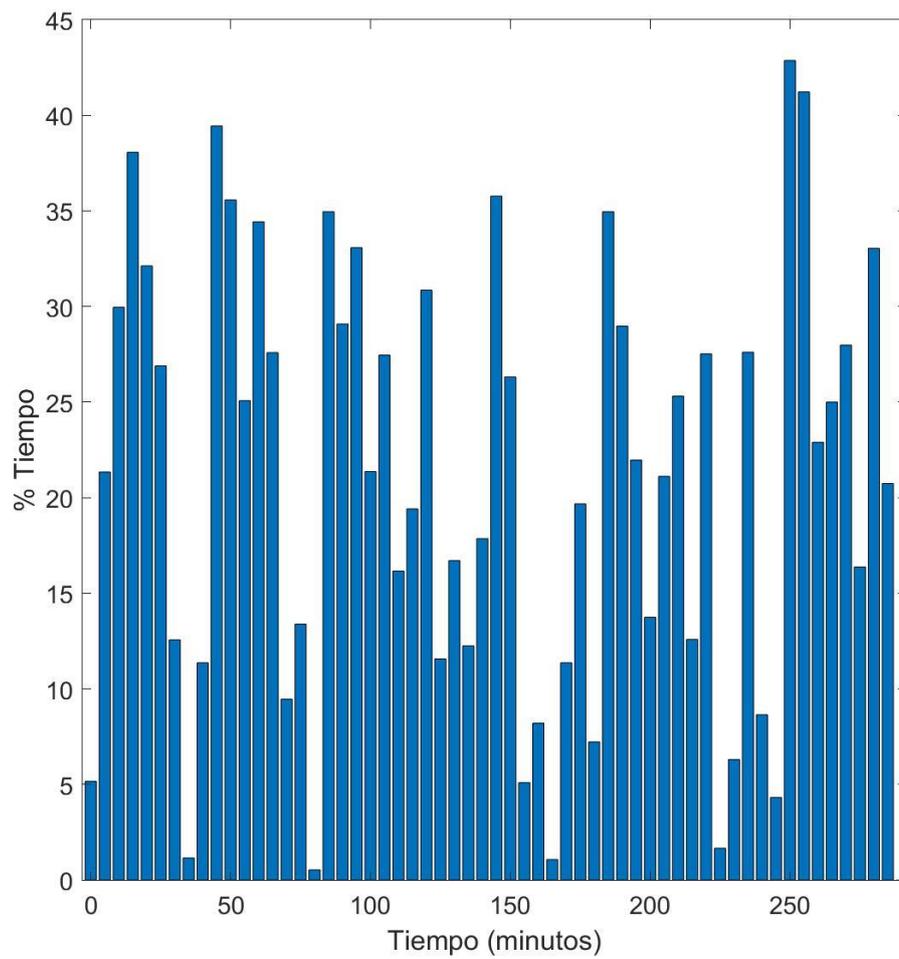


Figura 56: Gráfica de la evolución del porcentaje de fonación en segmentos de 5 minutos a lo largo del registro del día 1 del sujeto 1.

3.1.2. Análisis día hiperfuncional del sujeto 1

El mismo trabajo realizado con el día de fonación normal se lleva a cabo con el día de patrón hiperfuncional correspondiente al día 7.

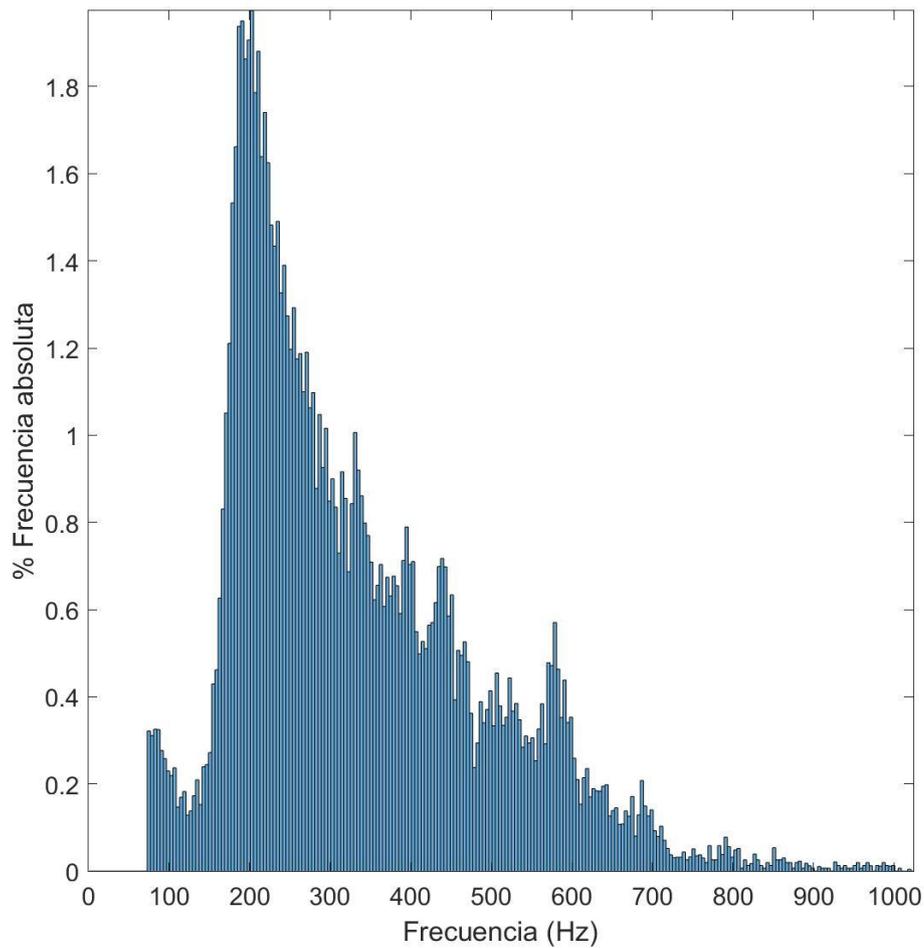


Figura 57: Histograma de frecuencias fundamentales del día 7 del sujeto 1.

En el histograma de frecuencias fundamentales (Figura 57) ya se observa un aumento del porcentaje de frecuencia absoluta en el rango de 400 a 600 Hz en comparación frente a la del registro de la Figura 45.

En el histograma de intensidades relativas del día 7 se observa un ligero desplazamiento del histograma hacia intensidades relativas mayores.

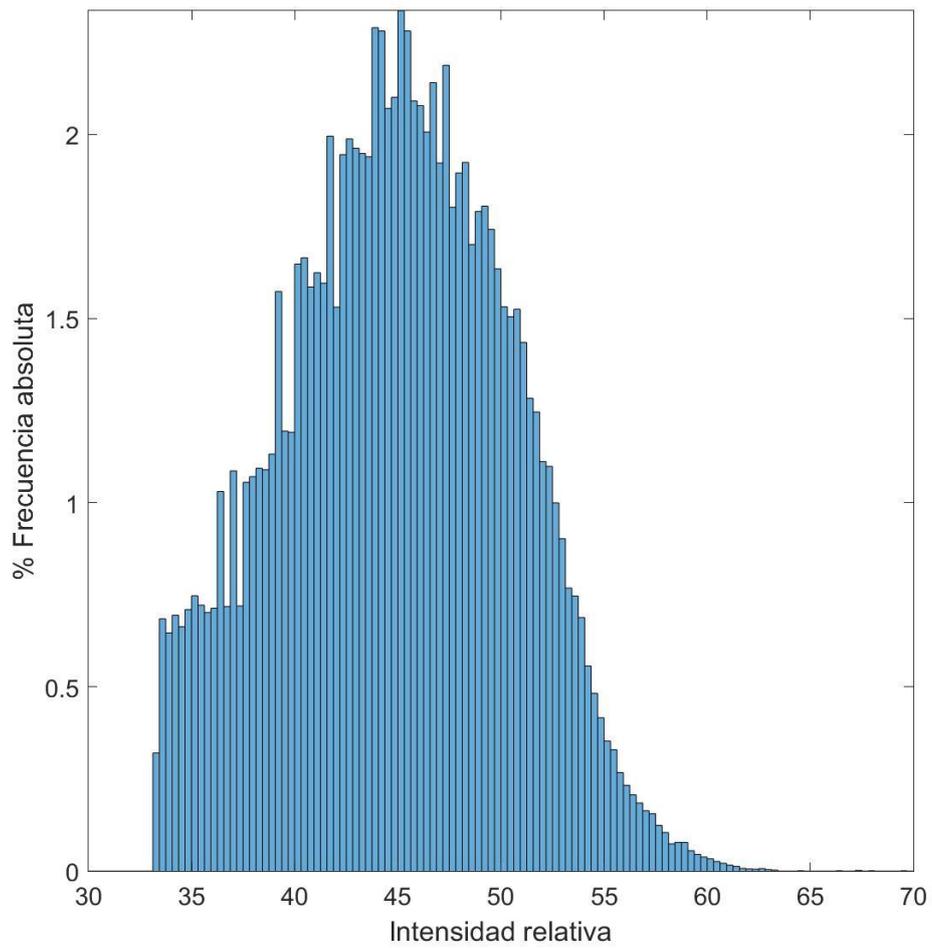


Figura 58: Histograma de intensidades relativas del día 7 del sujeto 1.

En la Figura 59 se observa cómo se expande el fonetograma presentando el mayor rango de frecuencias fundamentales empleado.

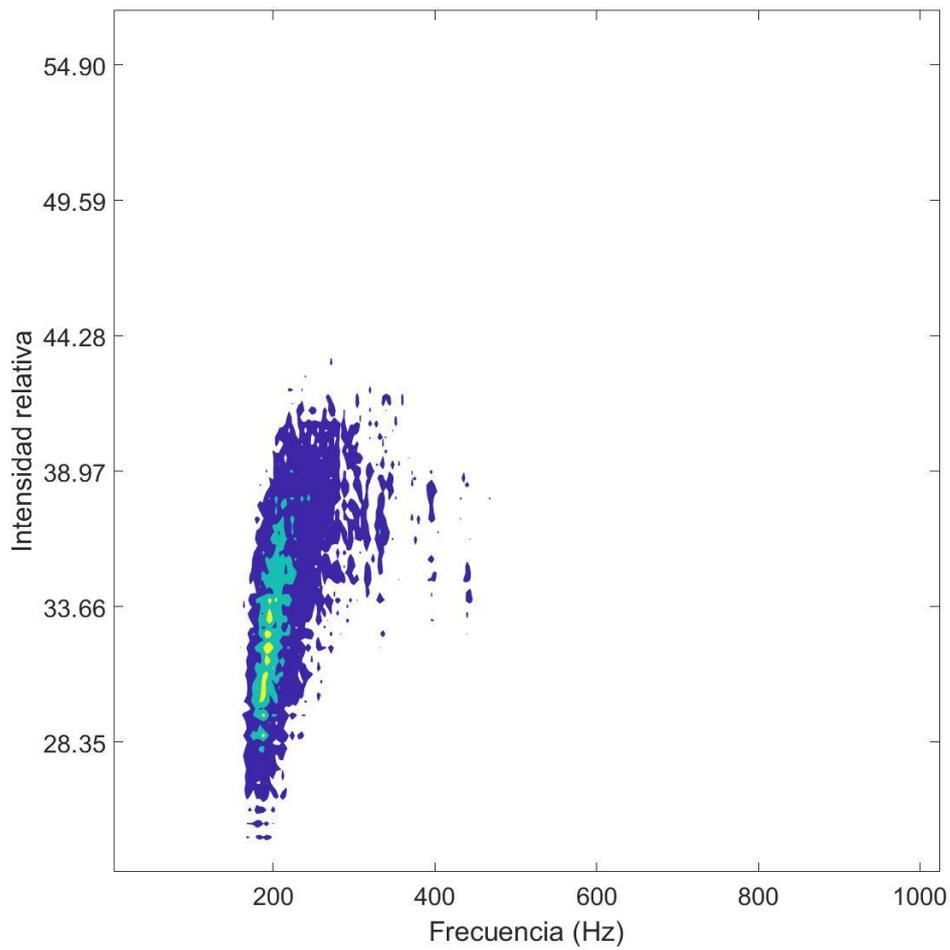


Figura 59: Fonetrograma del día 7 del sujeto 1. Los colores dividen las regiones según el tiempo acumulado con esa frecuencia fundamental e intensidad relativa. La región amarilla representa los puntos en que la proporción entre el tiempo en ese punto.

De nuevo, puede analizarse la evolución de la frecuencia fundamental a lo largo del tiempo de registro.

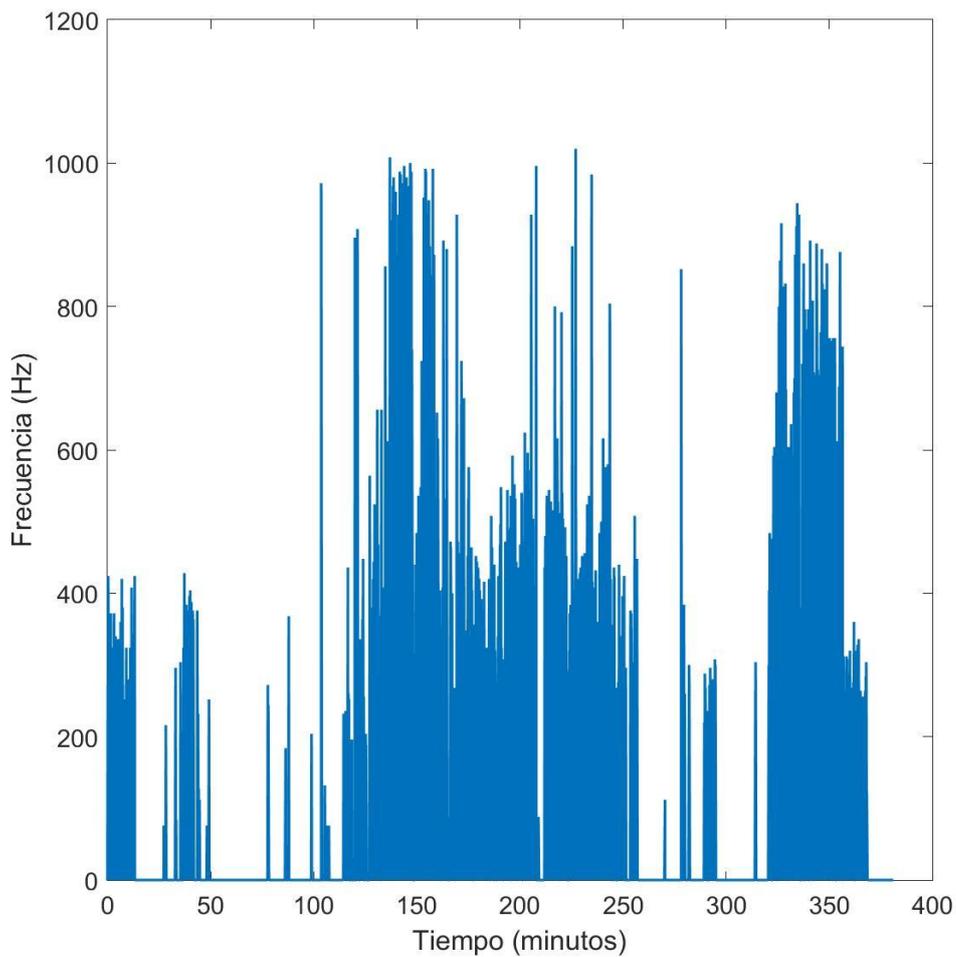


Figura 60: Gráfica de la evolución temporal de la frecuencia fundamental a lo largo del registro del día 1 del sujeto 1.

En la Figura 60 se observa la evolución de la frecuencia fundamental a lo largo del registro realizado el día 7. Durante este registro el sujeto 1 realizó un uso normal de la voz hasta el minuto 320. A continuación, tuvo una sesión de canto seguida, a partir del minuto 355, de un pequeño periodo de conversación. Durante el periodo de uso normal se pueden distinguir momentos en los que la frecuencia fundamental asciende (entre el minuto 100 y el 250) debido a que el sujeto 1 se encontraba realizando escalas tonales. También se distinguen entre el minuto 0 al 50 y del 250 al 300, periodos en los que la frecuencia fundamental no aumenta tanto.

La distribución del tiempo de fonación durante este registro se observa en la Figura 61.

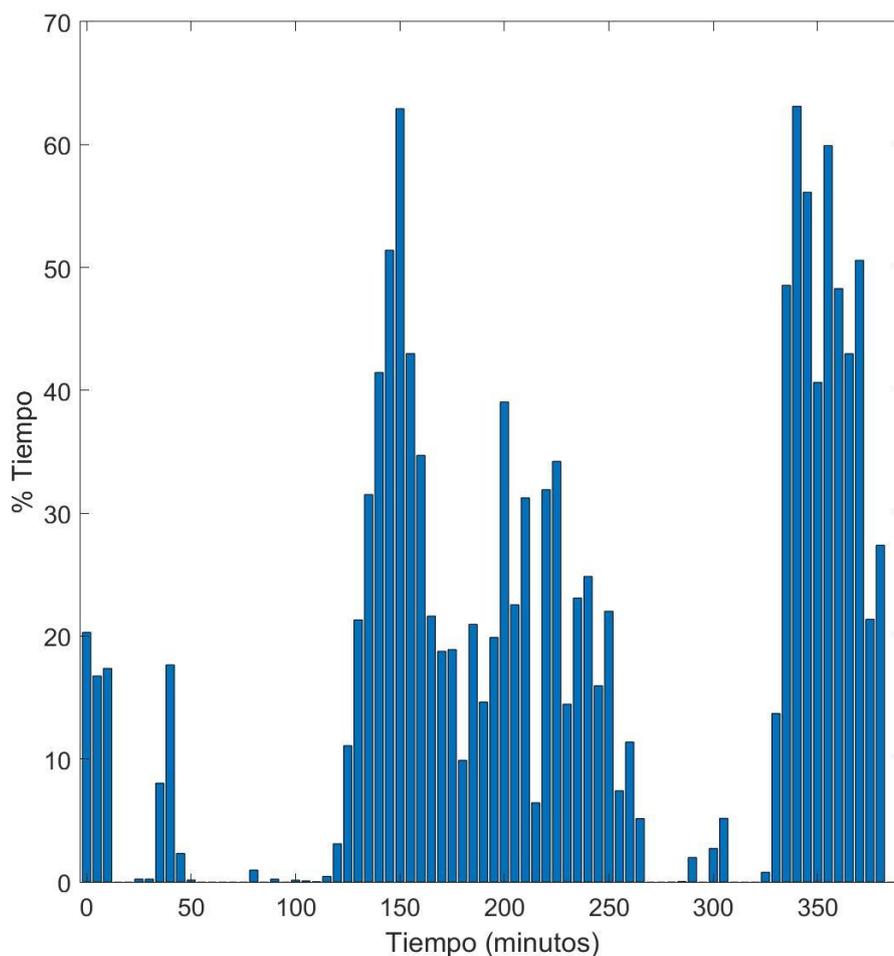


Figura 61: Gráfica de la evolución del porcentaje de fonación en segmentos de 5 minutos a lo largo del registro del día 7 del sujeto 1.

Los mismos intervalos temporales que hemos destacado anteriormente en el análisis de la Figura 60 se pueden observar en la Figura 61. En el periodo de uso normal de la voz (hasta el minuto 320) se observan distintas variaciones en el porcentaje de tiempo de fonación. Y a partir del minuto 320 se alcanza un tiempo en el que se mantiene un porcentaje de tiempo de fonación superior al 50 % para luego reducirse en la última parte que corresponde con conversación.

Estas tres partes en las que puede dividirse el registro del día 7 son susceptibles de un análisis más profundo. Por ello se analiza la frecuencia fundamental y la intensidad relativa en tres segmentos. El primero que abarca desde el minuto 0 al 320 que corresponde con el uso normal de la voz dentro de la jornada laboral del sujeto 1. El segundo segmento del minuto 320 al 355 que corresponde con la

sesión de canto. Y el último segmento que corresponde con el periodo que comprende del minuto 355 al 400, en el cual el sujeto 1 tiene una conversación.

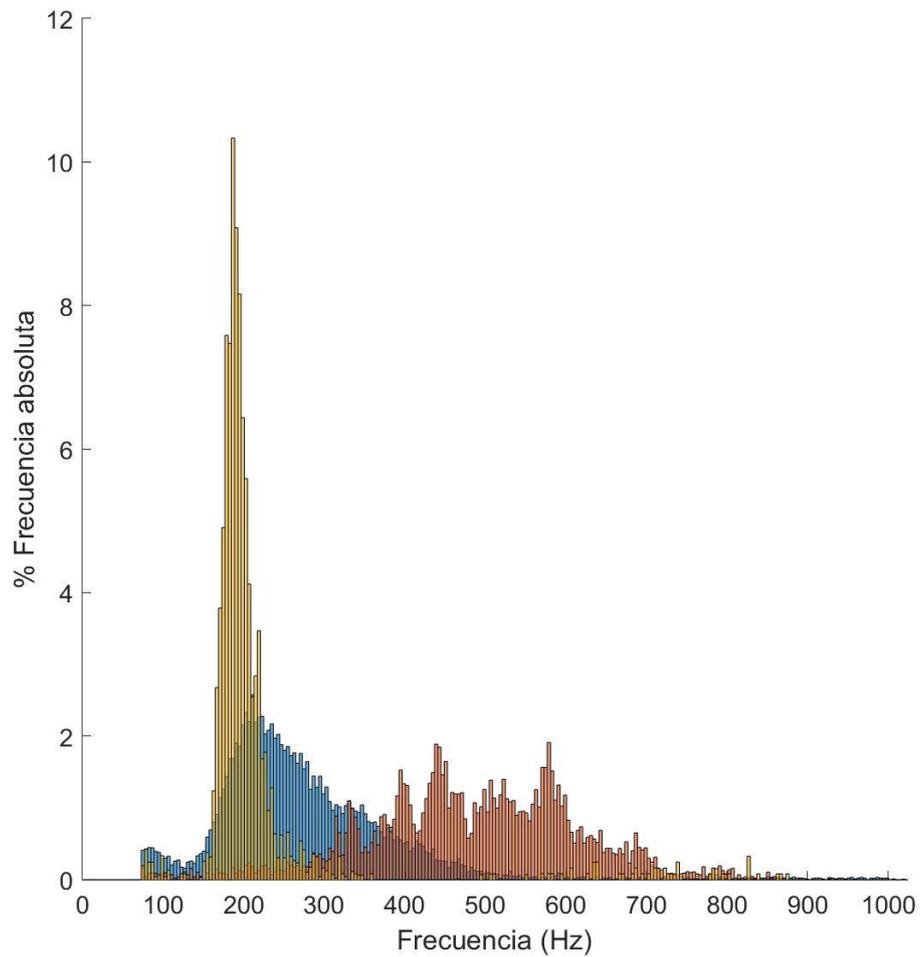


Figura 62: Histograma de frecuencias fundamentales del día 7 del sujeto 1. En color azul se muestran las frecuencias fundamentales registradas durante periodo de uso de la voz durante la jornada laboral. En color naranja se muestran las frecuencias fundamentales durante el periodo de canto. En color amarillo se muestran las frecuencias fundamentales en la conversación después de la sesión de canto.

La Figura 62 representa claramente las diferencias en los perfiles de frecuencias fundamentales en cada uno de los periodos estudiados.

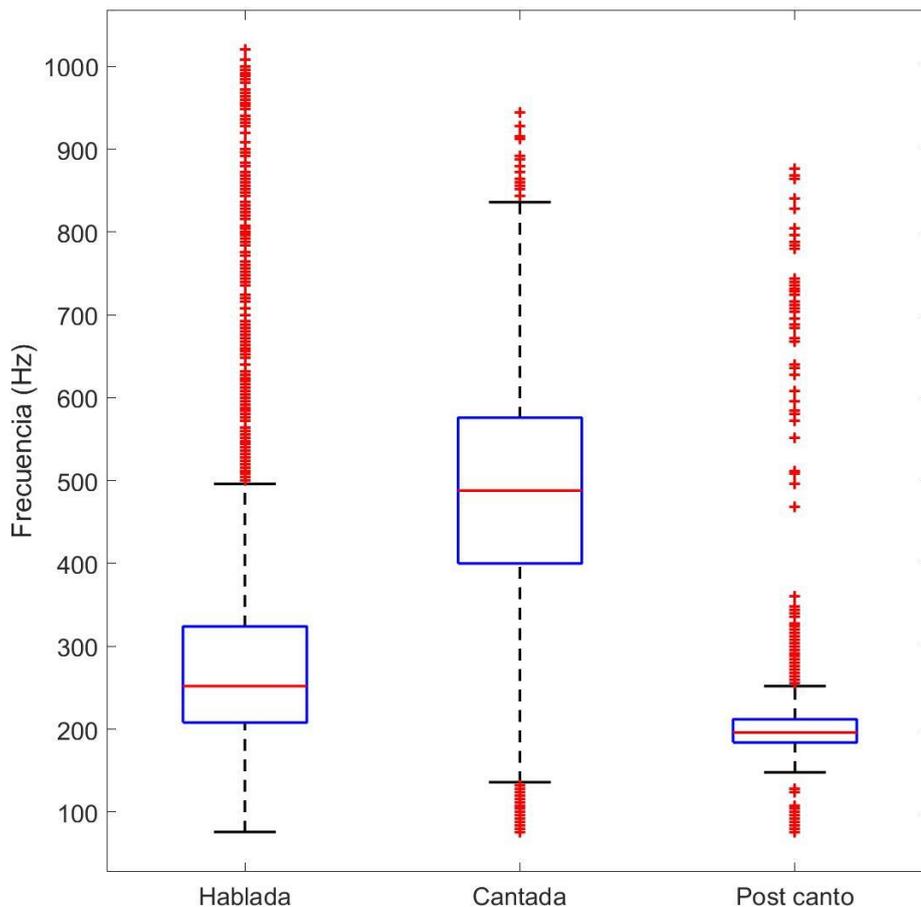


Figura 63: Boxplot de frecuencias fundamentales del día 7 del sujeto 1 según las 3 divisiones temporales explicadas anteriormente. De izquierda a derecha: voz hablada durante la jornada laboral, voz cantada en una sesión de canto y voz hablada durante conversación justo después de la sesión de canto.

La mediana de las frecuencias fundamentales son 252 Hz durante el primer periodo, 488 Hz durante el segundo y 196 Hz durante el tercero (Figura 63).

Los histogramas de intensidades relativas para estos mismos periodos se muestran en la Figura 64. En ellos se observa la similitud entre el periodo de canto y de voz hablada durante la jornada laboral en cuanto a la intensidad relativa. Sin embargo, la diferencia con respecto a las intensidades relativas en el periodo de conversación después del canto es considerable.

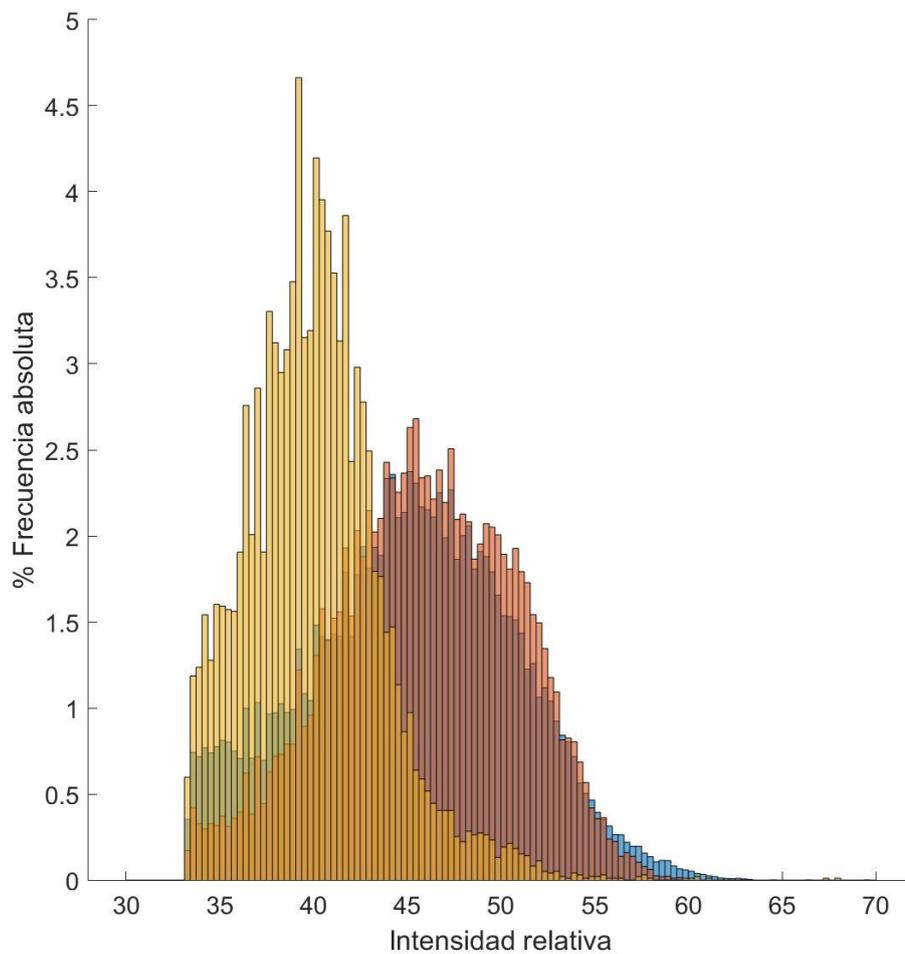


Figura 64: Histograma de intensidades relativas del día 7 del sujeto 1. En color azul se muestran las intensidades relativas registradas durante un periodo de uso de la voz durante la jornada laboral. En color naranja se muestran las intensidades relativas es durante el periodo de canto. En color amarillo se muestran las intensidades relativas en la conversación después de la sesión de canto.

El boxplot de las intensidades relativas en los tres periodos confirma lo mencionado el párrafo anterior. La mediana de las intensidades relativas del primer periodo es de 45,34, 45.93 para el segundo y 39.91 para el tercero (Figura 65).

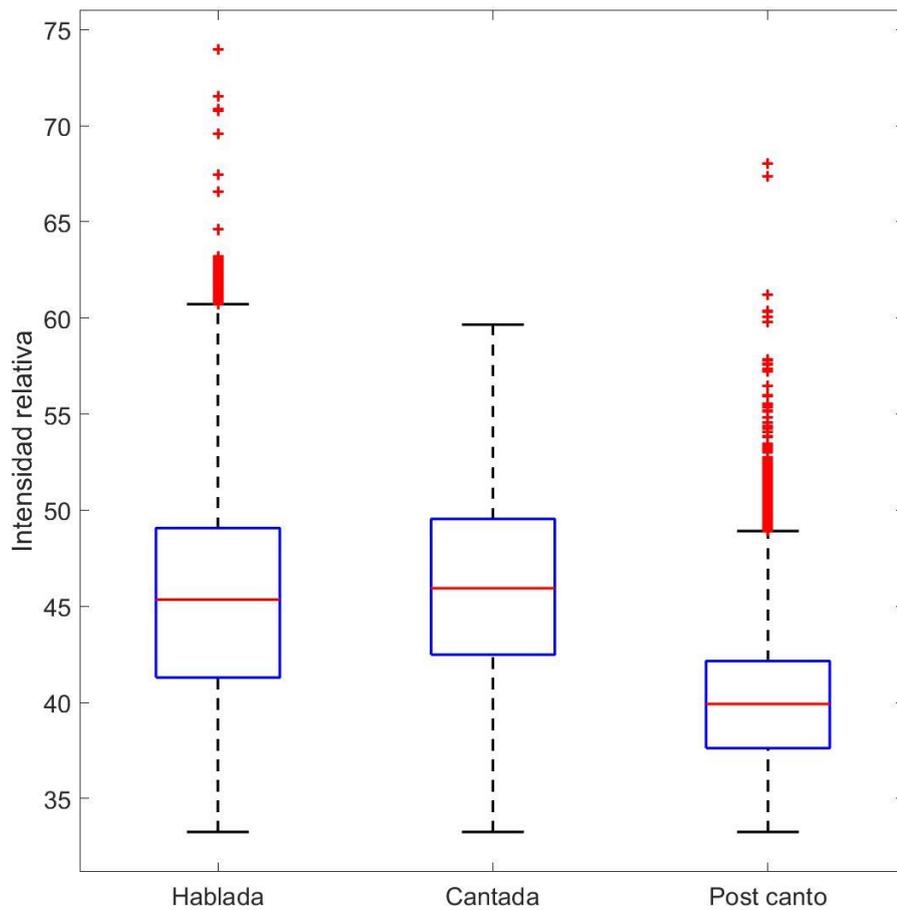


Figura 65: Boxplot de intensidades relativas del día 7 del sujeto 1 según las 3 divisiones temporales explicadas anteriormente. De izquierda a derecha: voz hablada durante la jornada laboral, voz cantada en una sesión de canto y voz hablada durante conversación

3.2. Sujeto 2

Los registros del sujeto 2 se llevaron a cabo durante 11 días. La duración de estas grabaciones se muestra en la Figura 66.

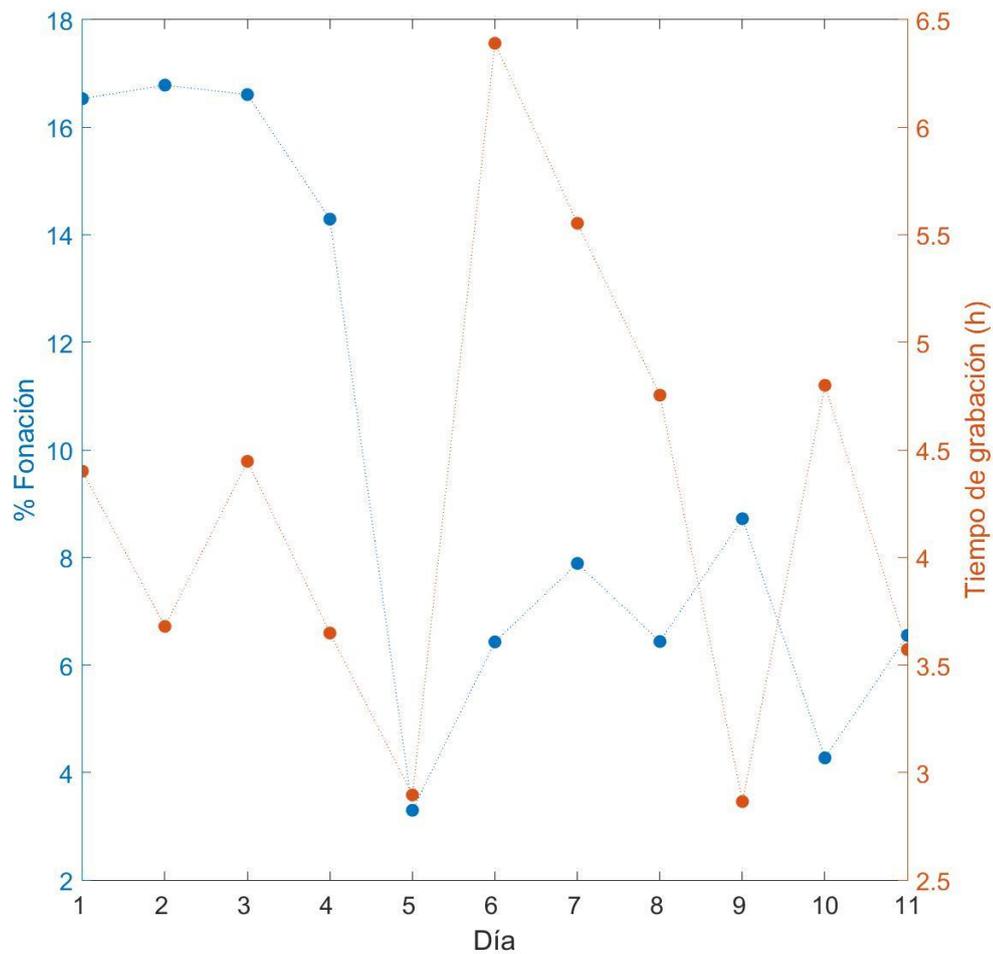


Figura 66: Registros de larga duración del sujeto 2. Rojo se muestra el tiempo de grabación y en azul el porcentaje de ese tiempo que el dispositivo ha contabilizado como fonación.

El tiempo de registro oscila entre 2.8 y 6.4 horas y el porcentaje de tiempo de fonación entre 3.2 y 16.8 %. El sujeto 2 presenta los primeros 4 días porcentajes de tiempo de fonación mayores que el resto de días. Esto se debe a que esos 4 días de registro tuvo varias reuniones, y el resto de días de registro fueron de trabajo en el laboratorio.

Los datos registrados durante los 11 días se condensan en la Figura 67 y Figura 68 donde se muestran los histogramas de frecuencias fundamentales e intensidades relativas, respectivamente.

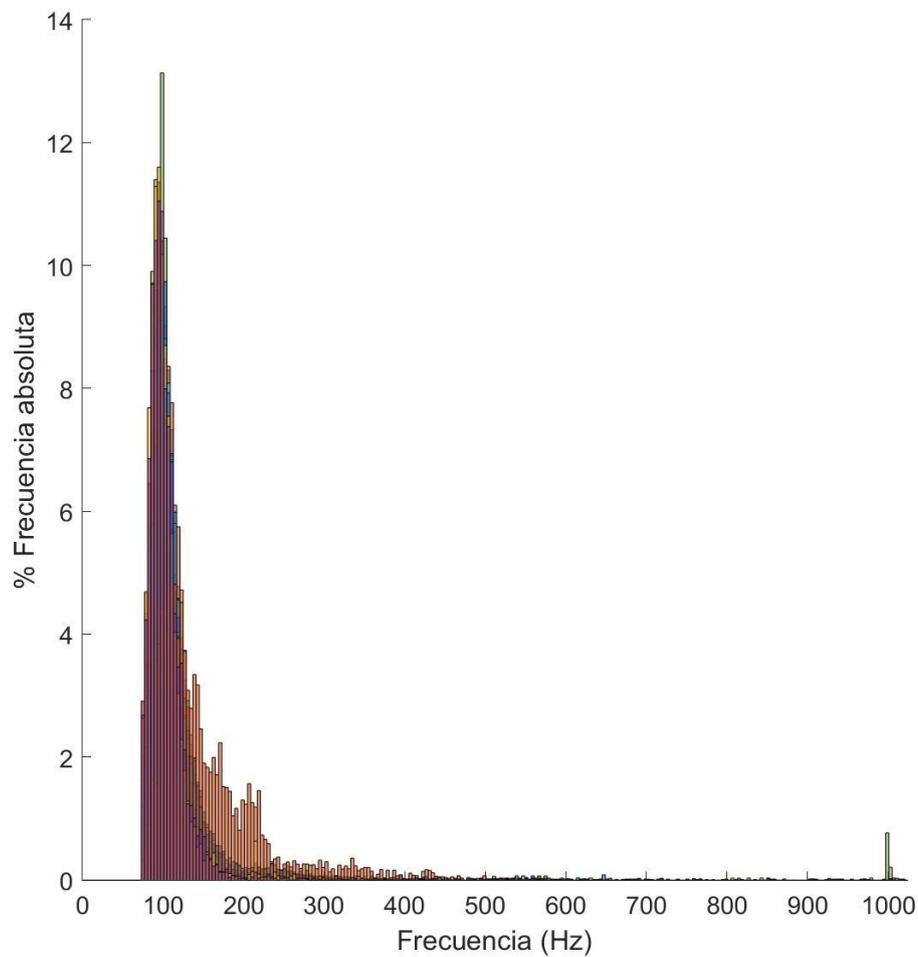


Figura 67: Histogramas de frecuencias de los 11 días de registro del sujeto 2.

Como en el caso del sujeto 2, en el histograma de frecuencias fundamentales todos los días se superponen. Sin embargo, se observa cómo uno de los histogramas contiene mayor número de frecuencias por encima de los 150 Hz. Este día corresponde con el día que el sujeto 2 realizó un patrón hiperfuncional consistente en canto a lo largo del registro.

En el histograma de intensidades relativas el día de patrón hiperfuncional no destaca y se observa mayor variación entre los diferentes días.

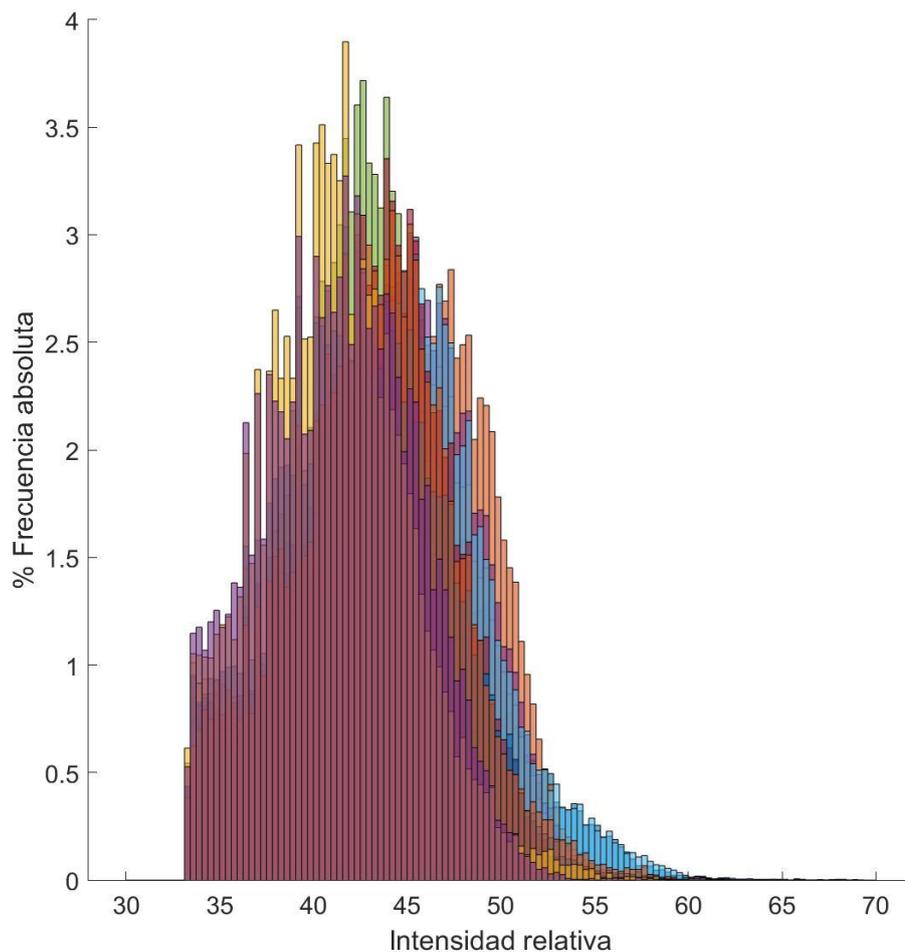


Figura 68: Histogramas de intensidad relativa de los 11 días de registro del sujeto 2.

El boxplot comparativo de frecuencias fundamentales e intensidades relativas de todos los días se encuentra en el ANEXO XIII.

3.2.1. Análisis de 3 días

El sujeto 2 durante los 11 días de registro se encontró en 3 situaciones. La primera situación hace referencia a los primeros 4 días de registro en los cuales estuvo en diferentes ambientes y reuniones. La segunda situación fue la generada por el sujeto 2 con un patrón hiperfuncional. Y la tercera situación es la de trabajo en el laboratorio desempeñado el resto de días. Tomando el día 2 como ejemplo de la primera situación, el día 9 que es el de patrón hiperfuncional

y el día 11 como ejemplo de la tercera situación, se muestran de nuevo los histogramas (Figura 69).

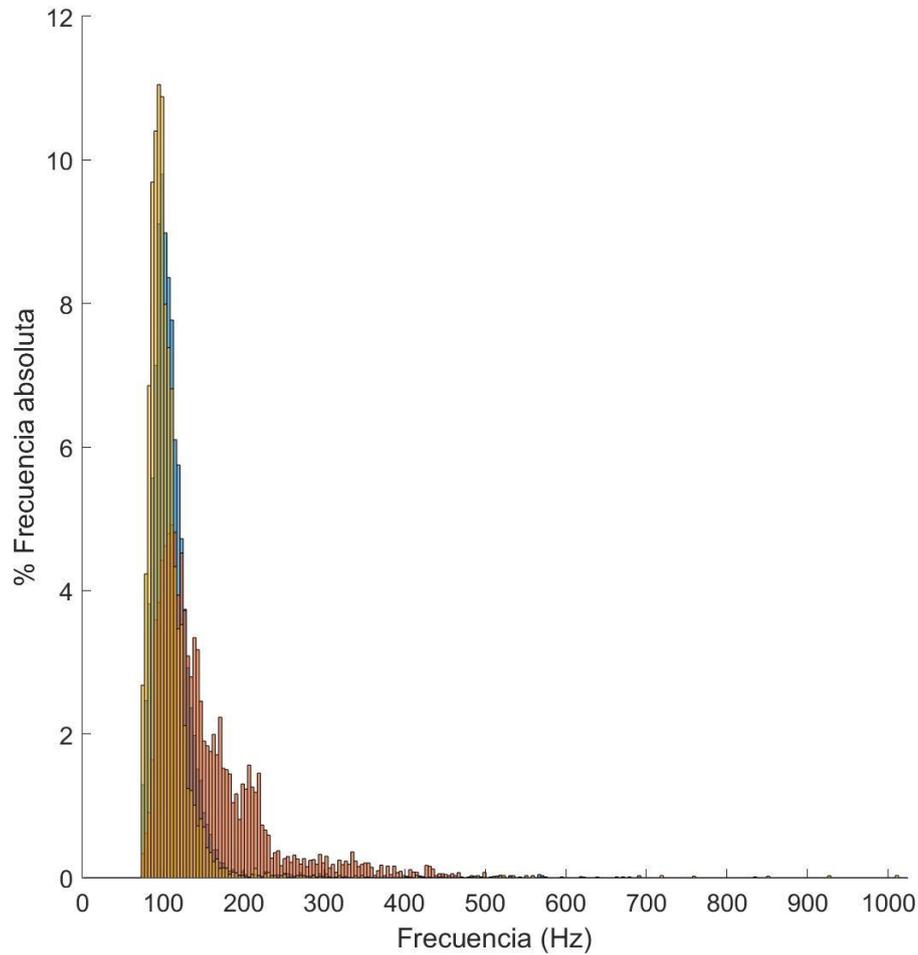


Figura 69: Histograma de frecuencias fundamentales de los días 2, 9 y 11 del sujeto 2. En color azul se muestran las frecuencias fundamentales registradas durante el día 2. En color naranja se muestran las frecuencias fundamentales durante el día 9. En color amarillo se muestran las frecuencias fundamentales durante el día 11.

El histograma del día 11 es prácticamente el mismo al del día 2 pero un poco más desplazado hacia frecuencias mayores. El histograma del día 9 contiene una proporción mayor de frecuencias fundamentales que superan los 150 Hz. Esta diferencia se observa mejor en la Figura 70.

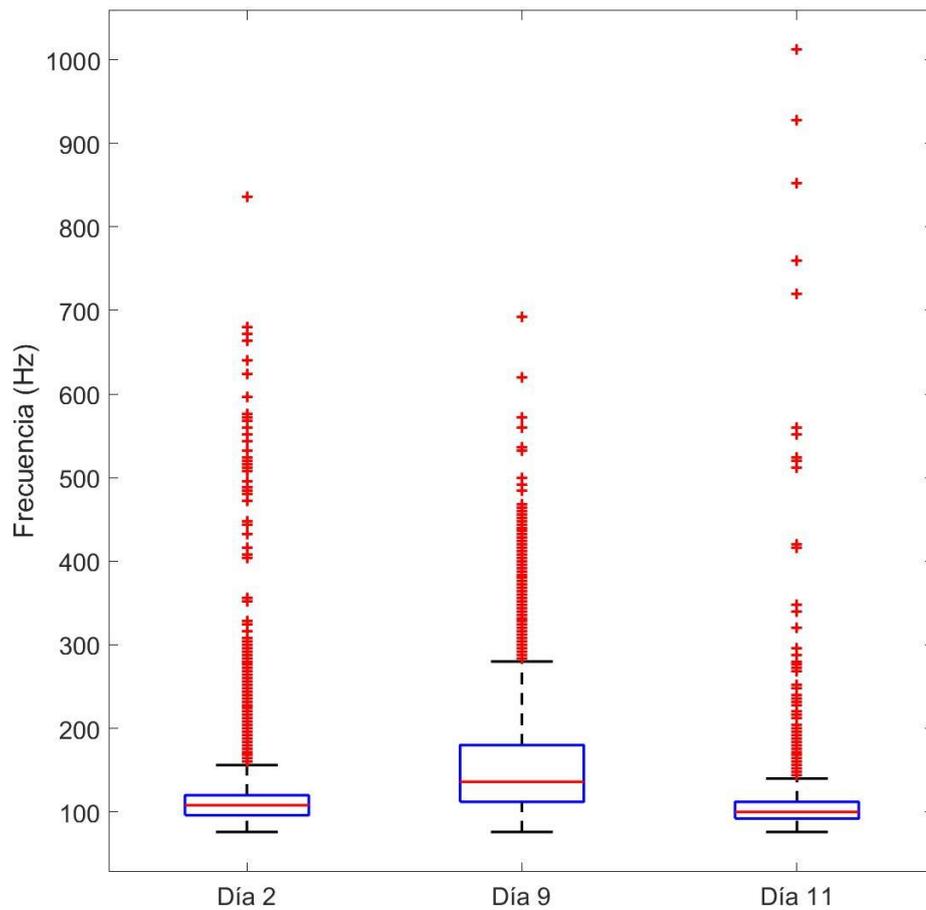


Figura 70: Boxplot de frecuencia fundamental de tres días de registro del Sujeto 2.

El registro de intensidades relativas durante el día 2 muestra un patrón de intensidades relativas elevadas en comparación con el día 11. El día 9, en el que se realizó un patrón hiperfuncional muestra intensidades relativas que se sitúan entre los dos días comentados antes (Figura 72).

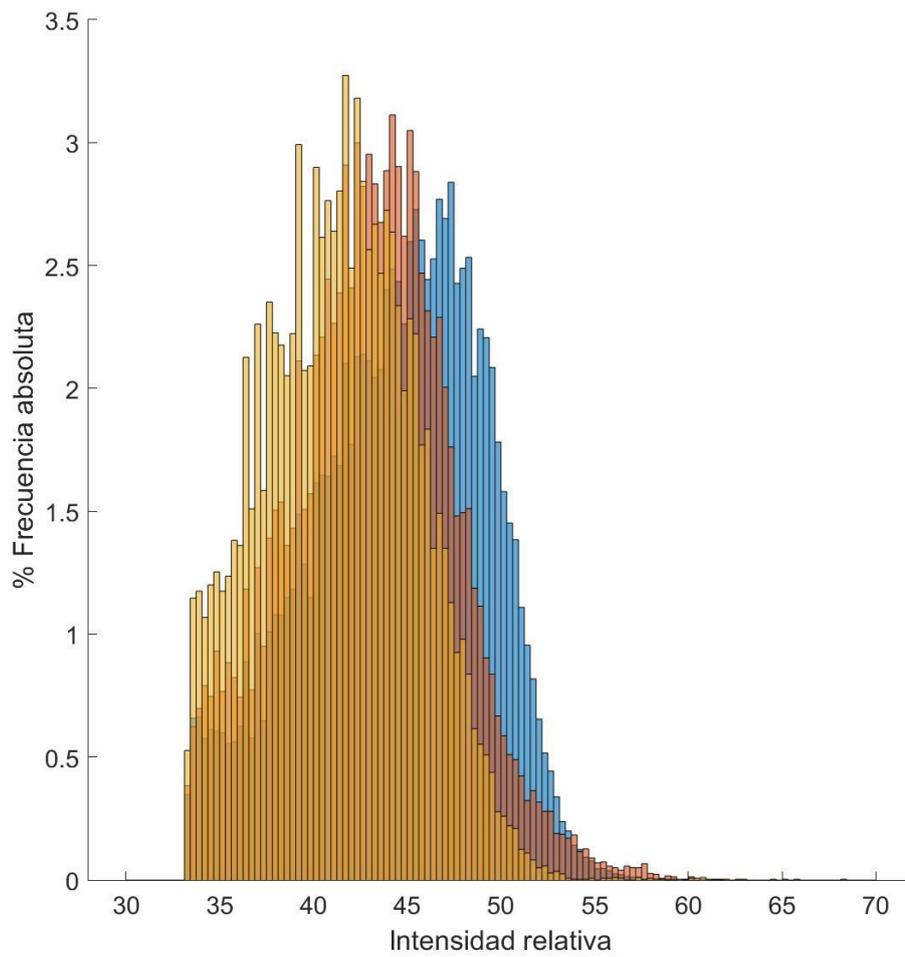


Figura 71: Histograma de frecuencias fundamentales de los días 2, 9 y 11 del sujeto 2. En color azul se muestran las frecuencias fundamentales registradas durante el día 2. En color naranja se muestran las frecuencias fundamentales durante el día 9. En color amarillo se muestran las frecuencias fundamentales durante el día 11.

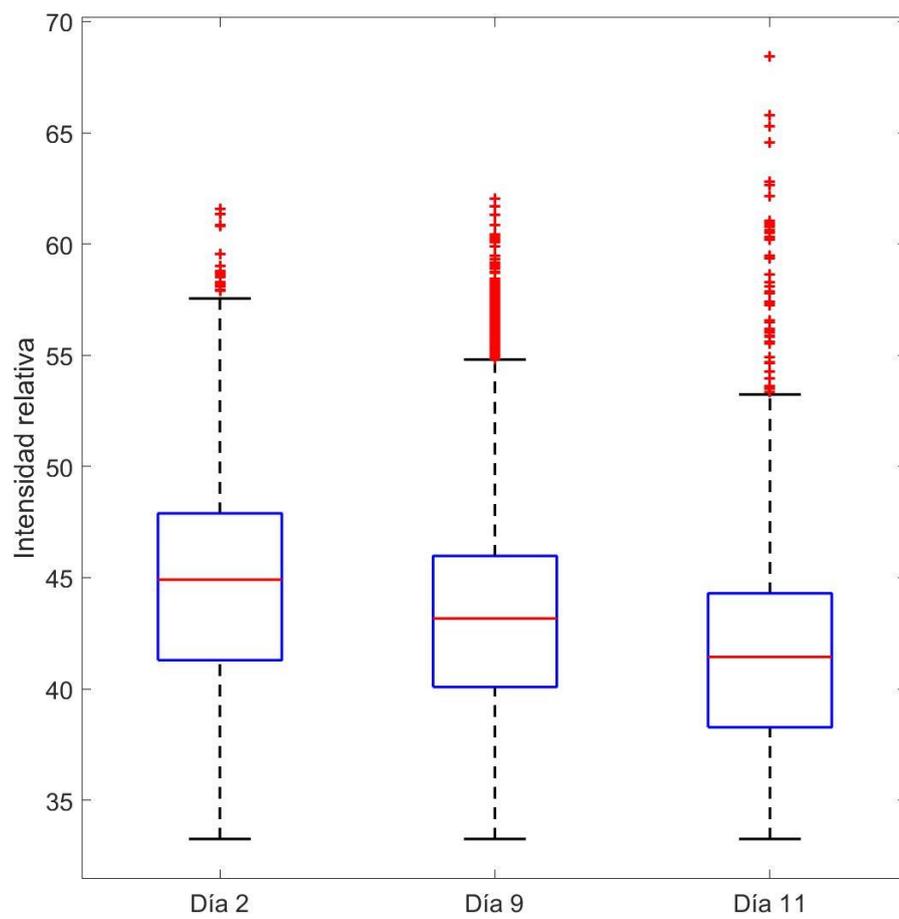


Figura 72:Boxplot de intensidades relativas de tres días de registro del Sujeto 2.

VIII. DISCUSIÓN

1. INTRODUCCIÓN

Como se ha comentado en capítulos previos, la incidencia y prevalencia de trastornos de la voz es elevada. Han sido varios los intentos de desarrollo de un dispositivo para la monitorización de la voz como se muestra en la Tabla 2. De los que llegaron a comercializarse en algún momento (Figura 1 y Figura 2), se identificó un factor que consideramos clave para que no tuvieran éxito además del alto coste económico: estos dispositivos eran portables pero no vestibles. Por ello consideramos que pese a ser utilizados como herramienta de investigación no llegan a cumplir los requisitos que se buscan para el entorno clínico o el usuario individual.

Estos dispositivos están pensados para el registro durante largos periodos de tiempo, por lo que la comodidad y practicidad pueden marcar la diferencia para su uso si los resultados son equiparables.

Todos los dispositivos mencionados, menos uno, conectaba el sensor con algún tipo de caja o grabador mediante un cable. Este cable, por sí solo, ya es molesto y puede interferir en los hábitos del usuario del dispositivo. El usuario debe llevar consigo una caja o grabador que debe colocarse en un bolsillo o colgado. El único de los dispositivos comercializados que no requiere de cable ni caja o grabador es el Vocalog2, por lo que se puede considerar como vestible por su tamaño y peso. No obstante, este presenta dos inconvenientes muy importantes. Primero, no registra frecuencia fundamental. Y segundo, registra las vibraciones a través de un micrófono de contacto y no mide la frecuencia fundamental. Aunque el dispositivo no esté pensado para ello, el hecho de poseer un micrófono que pueda registrar la voz del usuario cuestiona de manera importante la privacidad. Este aspecto de la privacidad también es un problema en el dispositivo Voice-Care que usa micrófono de contacto, y en el dispositivo presentado por Mehta et al. que posee un micrófono para medir el ruido ambiente⁵¹ y en el dispositivo Voxlog que también dispone de micrófono para la medición de ruido ambiente y autocalibración de la intensidad.

En dispositivos pensados para registros de larga duración, 24 horas de registro son escasas. En estos casos, el usuario ha de cargar la batería del dispositivo o vaciar la memoria requiriendo intervención sobre el dispositivo para obtener registro de más de

24 horas. De hecho, no se tiene constancia de ningún estudio que utilice este tipo de dispositivos en registros de más de dos semanas de duración.

2. HARDWARE

El hardware utilizado ha permitido desarrollar un dispositivo potencialmente vestible, esto quiere decir que su tamaño y peso permite llevarlo encima. Un peso estimado de menos de 40 g y unas dimensiones reducidas (Figura 19), el dispositivo será cómodo y liviano. Para estas estimaciones se ha tenido en cuenta una batería de 400 mAh, suficiente para aproximadamente 164 horas de registro (7 días ininterrumpidos), pero podría variar según los requisitos. Una batería de 1000 mAh pesa 7 g más y tiene un volumen 40 % superior y duraría 17 días. Si por lo contrario con 24 – 48 horas de registro es suficiente, contando con que el usuario cargaría el dispositivo a diario, una batería de 110 mAh sería suficiente reduciendo el peso en 6 g y un volumen 35 % menor con respecto a la batería de 400 mAh.

El único dispositivo en la bibliografía similar en tamaño y peso es el propuesto en 2017 por Metha et al.⁵¹ Sin embargo, este está formado por dos componentes: el primero se encuentra pegado mediante adhesivo al cuello y el segundo es un receptor que debe encontrarse en un rango de distancia establecido por la tecnología que utiliza (Bluetooth 5.0). El hecho de requerir un dispositivo receptor reduce su comodidad y practicidad; es un dispositivo más que debe llevar el usuario con él, puede haber problemas de envío de datos, y ambos dispositivos deben cargar sus baterías. El dispositivo diseñado trabaja de manera independiente y no necesita estar conectado a ningún otro para operar.

Para los registros realizados en este trabajo, el prototipo utilizado poseía unas características más parecidas al resto de dispositivos (Figura 1), el sensor conectado por cables con una caja (Figura 17). Los comentarios de los sujetos ratificaron la incomodidad del cable y la caja.

El tamaño y peso del dispositivo está determinado por el bajo consumo del dispositivo. La información sobre el consumo de otros dispositivos es escasa, solo se puede calcular en tres trabajos. El primero, presentado por H. Cheyne et al. (2003), describen que su

dispositivo (basado en el Compaq iPAQ 3765) cuya batería es de 950 mAh es capaz de hacer registros de 2-3 horas, lo que significa un consumo de entre 475 y 316 mA.⁵ El segundo detalla el dispositivo VHM que dispone de una batería de 3900 mAh que dura 18 horas, por lo que el consumo es de 216 mA.⁴⁵ El tercero y último, en el trabajo más reciente de un nuevo dispositivo de dosimetría (2017), describen una capacidad de 110 mAh y una duración de 8 horas, lo que significa un consumo de 13.65 mA.⁵¹ De estos tres dispositivos solo los dos primeros realizan el cálculo de F_0 e I, el tercero únicamente transmite los datos mediante bluetooth a un receptor que los almacena. Si en el tercero tenemos en cuenta el consumo del receptor (1000 mAh y 24 horas de uso son 41.6 mA de consumo) el consumo total del dispositivo es de 55.31 mA. Y este dispositivo no realiza el cálculo de F_0 en el propio receptor, sino que solo almacena información para ser procesada en un ordenador.⁴⁶ Frente a estos datos, el dispositivo presentado en este documento tiene una importante ventaja pues su consumo es de 2.44 mA. Comparado con los 3 mencionados es 194, 89 y 23 veces más eficiente respectivamente. En igualdad de parámetros con los dos primeros (calculan en el mismo dispositivo F_0 e I) y con mayor trabajo computacional que el tercero que no calcula internamente F_0 .

3. VALIDACIÓN

Con el objetivo de estudiar el procesado de los datos en el microcontrolador se analizaron cinco señales generadas con características conocidas. Utilizando el mismo algoritmo se compararon los resultados obtenidos utilizando Matlab y el microcontrolador. Como se muestra en los resultados, el cálculo de la magnitud de la FFT es muy parecido, sin embargo, en el filtrado y la correlación existe un desfase. Este desfase, determinado por las limitaciones de procesado del microcontrolador, se tiene en cuenta para no incurrir en un error constante en el cálculo de la frecuencia fundamental.

Una vez comprobado que el mismo algoritmo obtiene como resultados los mismos valores usando Matlab y usando el microcontrolador, estos se comprueban con los esperados. La Tabla 6 muestra la coincidencia entre ambos (Matlab y dispositivo) y con lo real, salvo en la primera señal generada (90 Hz con 3 armónicos de amplitudes relativas 1,1 y 1). La resolución frecuencial del dispositivo propuesto es de 4 Hz, y esto

conlleva errores como este, en el que lo calculado y lo real no coincide. Como se ha explicado en el apartado 1.2.2.2 del capítulo VI, esta resolución está determinada por la frecuencia de muestreo elegida y el número de muestras del intervalo analizado. Esta resolución frecuencial puede parecer baja si se están analizando cortos periodos de fonación. Teniendo en mente el objetivo de monitorización de la voz durante largos periodos, la resolución es aceptable puesto que obtendremos un valor de frecuencia por cada 31.25 ms de fonación. Esto quiere decir que en un minuto de fonación tenemos 2000 valores de frecuencia fundamental. Es mucha información cuando vamos a estar registrando durante días. Al igual que en la Figura 53 varios trabajos, en varios trabajos representan los datos de frecuencia o intensidad con la mediana de segmentos de varios minutos,^{7,45,66} por lo que por cada 10000 valores se toma uno para representarlos.

Las señales generadas para la validación del algoritmo son mucho más simples que las registradas por el acelerómetro sobre el sujeto, por lo que se valida posicionándolo sobre la piel y registrando voz. Como se muestra en la Figura 39, la correlación entre las frecuencias fundamentales calculadas por el dispositivo y las calculadas con el SoundScope es muy alta ($R = 0.9998$). Además, esta correlación se cumple en todo el rango de frecuencias analizado (86-550 Hz). Este detalle muestra cómo la estrategia implementada en el algoritmo combinando la información de la correlación y la magnitud de la FFT es muy eficaz. Analizando los datos del artículo presentado por Bottalico et al. (2018) se obtiene que para el fonema /a/ el error medio de los dispositivos VoxLog, Voice-Care y APM es 7.66, 1.83 y 4 Hz respectivamente⁴⁷, y para el dispositivo presentado en este documento el error absoluto medio es de 1.94 Hz. Estos dispositivos basan su algoritmo para la detección de la frecuencia fundamental en un único método, el procesado basado en FFT usado por el VoxLog y la autocorrelación usada en los dispositivos Voice-Care y APM⁶⁷.

En el cálculo del valor de intensidad la principal diferencia con el resto de dispositivos que miden esta variable es que en este trabajo se plantea la ausencia de calibración puesto que se prescinde de calcular el valor absoluto de intensidad. Los dispositivos que cuentan con un micrófono son capaces de autocalibrarse, pero los que se basan en acelerometría necesitan calibración para poder obtener los resultados en las mismas unidades que los sonómetros. En nuestro planteamiento, consideramos que los valores

en sí mismos no son relevantes, sino que el interés de la medición de intensidad en largos periodos reside en el análisis de la variación de esta a lo largo del tiempo. Con ello se evita la necesidad de usar micrófono en el dispositivo para mantener la privacidad y se prescinde del proceso de calibración simplificando su uso.

De todos modos, a pesar de no realizarse la calibración, se comprobó la relación de los valores obtenidos por el dispositivo en relación a las mediciones realizadas con un sonómetro. Los resultados concuerdan con lo publicado por A. Astolfi et al. (2013)⁶⁷ y más detallado por A. Carullo et al. (2015).⁶⁸ Calculando el logaritmo en base 10 del valor de intensidad relativa calculado por el dispositivo sí observamos una correlación lineal (Figura 42) con $R = 0.96$. Esto significa que la calibración afectaría a los valores y las unidades utilizadas, pero no a su comportamiento. Ahora bien, la Figura 41 es si cabe más relevante pues muestra que esta alta correlación disminuye poco ($R = 0.86$) tras analizar las 77 mediciones de intensidad. Durante estas mediciones el dispositivo se quitó de su posición varias veces, por lo que se muestra que los valores de intensidad relativa mantienen la correlación entre diferentes sesiones. Esto es importante para poder comparar las mediciones obtenidas a lo largo de distintos días de registro del usuario.

El último apartado de la validación corresponde al tiempo de fonación. El hecho de que los tiempos medidos usando micrófono y usando el dispositivo correspondan cuando solo se están registrando sonidos sonoros (vocales) demuestra que el dispositivo funciona. Las diferencias en las grabaciones de los 3 sujetos entran dentro de lo esperado, el dispositivo solo cuantifica el tiempo de fonación, mientras que el micrófono contabiliza el total de la señal acústica (sorda y sonora). Esto significa una mayor precisión a la hora de determinar el número de ciclos vocales y poder relacionarlo con la carga vocal. La correcta colocación del dispositivo se ha mostrado también como punto importante, la grabación del sujeto 1 erra en la estimación del tiempo debido al incorrecto contacto del acelerómetro con la piel. Se ha visto que esto hace que el dispositivo deje de medir durante un periodo del registro.

4. REGISTROS DE VOZ

Los resultados del registro durante varios días muestran la utilidad de este dispositivo. La Figura 44 y Figura 66 muestran información sobre el porcentaje de fonación respecto al tiempo registrado. Estos datos permiten objetivar y cuantificar la sobrecarga vocal o patrones fonatorios que supongan riesgo de lesión orgánica. Poderlo asociar a determinados días o patrones fonatorios puede permitir la reducción de la carga vocal y mejorar la salud de aparato fonador dado que los trastornos de la voz más comunes son crónicos o recurrentes como resultado del abuso o mal uso de la voz.²⁸

De estas dos figuras mencionadas también se destaca la variabilidad en el porcentaje de fonación dentro de cada sujeto. Esta variabilidad muestra la importancia de un análisis durante largos periodos, el uso de la voz es muy variable y a partir de muestras de corta duración o en condiciones de laboratorio no se pueden inferir resultados relevantes. Con este razonamiento, si nos fijamos en los registros del sujeto 2 (Figura 66) las conclusiones a las que llegaríamos analizando los primeros 4 días de registro serían totalmente distintas a las de los siguientes 7 días. La información que aporta el registro durante largos periodos de tiempo puede ser muy importante tanto para el propio usuario como para el clínico (en el caso en el que el usuario sea un paciente).

En los histogramas que muestran la frecuencia fundamental de los registros realizados con el sujeto 1 y 2 (Figura 45 y Figura 67) se aprecia la diferencia de rango frecuencial de cada sujeto. Esta información es relevante si queremos conocer cómo es el uso de la voz. En estos histogramas se observan datos acumulados en frecuencias elevadas (en torno a 500 Hz en el sujeto 1 y en 1024 Hz en el sujeto 2). Como se ha mencionado con anterioridad, la señal registrada procede propiamente de la actividad de las cuerdas vocales, de la ondulación y contacto de las mismas. En el intervalo de 250 ms utilizado para el cálculo de la frecuencia fundamental puede haber cambios frecuenciales que propicien errores como estos en el cálculo de la frecuencia. Por este motivo también se registran más periodos de frecuencias menores de 100 Hz en el sujeto 1, hecho improbable al tratarse de una mujer cuyo rango vocal normal se encuentra entre los 180 y los 250 Hz.¹⁰ Este patrón de frecuencias inferiores a lo esperado también se presenta en otros dispositivos.⁶⁷

En cuanto a los histogramas de intensidad relativa cabe mencionar el corte de la cola izquierda en la Figura 46, y que no destaca tanto en la Figura 68. Este corte se debe al uso de umbral de intensidad para considerar un intervalo como fonación. Este límite no pudo reducirse más puesto que se observó que empezaban a reconocerse como intervalos con fonación aquellos que no la tenían. Esta acumulación en bajas intensidades ocurre por dos motivos. El primero de ellos lo asociamos a los comienzos y finales de fonación. Aunque la intensidad sea elevada, esta no pasa de cero al valor de la intensidad de manera súbita, si no que asciende con mayor o menor velocidad. Al analizar subintervalos de 31.25 ms para el cálculo de la intensidad, el intervalo puede registrar solo el ascenso, por lo que se registrará una intensidad baja. Este efecto sucede en cada comienzo y final de fonación durante todo el registro. El segundo motivo lo asociamos a que aceleraciones no producidas por fonación se registren como tal. Para que ocurra algo así debe darse que el intervalo se determine como intervalo con fonación y que en un subintervalo que no contenga señal de fonación una aceleración haga que el cálculo de RMS del subintervalo supere el umbral establecido para diferenciar los subintervalos que contienen fonación de los que no. Por lo analizado visualmente en varios conjuntos de datos, se considera el primero de estos motivos como el principal. Además, este efecto en el histograma se presenta de igual manera en dispositivos como el APM y el Voice-Care.⁶⁷

Toda esta información puede utilizarse para comparar la fonación durante varios días como se muestra en el apartado 3.2.1 del capítulo VII, pero también pueden analizarse de una manera más concreta, analizando un único día. La información mostrada en la Figura 61 es muy útil porque permite identificar cómo se distribuye el tiempo de fonación durante el registro. Una fonación acumulada de 20 minutos en un intervalo de registro de 30 minutos no supone la misma sobrecarga vocal que esa misma fonación acumulada distribuida a lo largo de un registro de 300 minutos y, por lo tanto, no supone el mismo riesgo de trauma. Incluso si se identifican diferentes actividades, pueden analizarse de manera separada partiendo de un histograma como el de la Figura 57 para obtener uno como el de la Figura 62.

5. LIMITACIONES Y LÍNEAS FUTURAS

La elección del microcontrolador ha sido clave para hacer posible que el dispositivo cumpla el objetivo marcado, pero el microcontrolador también ha supuesto los mayores retos. Las limitaciones técnicas de memoria y de no poder enviar los datos mientras se registran son las mayores limitaciones del microcontrolador. Además, al ser un procesador moderno se dispone de poca información para que un programador con experiencia nula en empleo de este tipo de microcontroladores pueda aprender y resolver problemas, como ha sido el caso.

El requisito de hacer el dispositivo lo más ligero y eficiente energéticamente posible ha hecho que se descarte la amplificación y filtrado de la señal recibida del acelerómetro por hardware antes de procesarla como otros trabajos anteriores sí han hecho, complicando este procesamiento.

La versión vestible del dispositivo (Figura 19) tiene en cuenta todos los componentes fundamentales del equipo para ser funcional y se está trabajando para su fabricación. Para la tesis, todos los datos han sido tomados con el prototipo de la Figura 17.

El algoritmo es optimizable; una mejor gestión de las variables puede hacer que sea más eficiente enviando la información en vez de grabándola en la memoria micro SD disminuyendo el tiempo de escritura que es el proceso que más consumo energético conlleva. Además, sería interesante utilizar otras características como el RTC del microcontrolador para estampar la fecha de comienzo de escritura.

IX. CONCLUSIONES

- Es posible desarrollar un dispositivo vestible que permita la monitorización durante largos periodos de tiempo de los principales parámetros de la voz.
- La combinación del análisis de la magnitud de la FFT y la autocorrelación del filtrado de la magnitud de la FFT es un método válido para el cálculo de la frecuencia fundamental de la fonación registrada a través de acelerometría.
- El microcontrolador MSP430FR5994 permite el cálculo de la frecuencia fundamental, intensidad relativa y tiempo de fonación de la fonación en tiempo real.
- La combinación del microcontrolador MSP430FR5994 y el algoritmo presentado reduce considerablemente el consumo energético necesario para la medición de los parámetros fundamentales de la voz.
- El reducido precio de los componentes hace viable la comercialización del dispositivo para que un mayor rango de profesionales como logopedas, foniatras o profesores de canto entre otros, puedan acceder a él.
- Los registros realizados con el dispositivo presentado permiten diferenciar diferentes patrones de frecuencia fundamental e intensidad relativa en distintos días de registro y en diferentes horas de un día de registro.
- El dispositivo permitirá la detección de patrones fonatorios ineficientes o con riesgo de desarrollo de lesiones orgánicas a la vez que permitirá el seguimiento de la modificación de conductas fonatorias después de haber seguido tratamientos logopédicos, farmacológicos o quirúrgicos.

X. BIBLIOGRAFÍA

1. Horii Y, Fuller BF. Selected acoustic characteristics of voices before intubation and after extubation. *J Speech Hear Res.* 1990;33(3):505-510.
2. Stevens KN, Kalikow DN, Willemain TR. A miniature accelerometer for detecting glottal waveforms and nasalization. *J Speech Hear Res.* 1975;18(3):594-599.
3. Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope.* 2005;115(11):1988-1995. doi:10.1097/01.mlg.0000179174.32345.41
4. Perelló J. *La Théorie Muco-Ondulatoire de La Phonatió n.* Vol 79.; 1962.
5. Popolo PS, Svec JG, Titze IR. Adaptation of a Pocket PC for use as a wearable voice dosimeter. *J Speech Lang Hear Res.* 2005;48(4):780-791. doi:10.1044/1092-4388(2005/054)
6. Cheyne H a, Hanson HM, Genereux RP, Stevens KN, Hillman RE. Development and testing of a portable vocal accumulator. *J Speech Lang Hear Res.* 2003;46(6):1457-1467. doi:10.1044/1092-4388(2003/113)
7. Van Stan JH, Mehta DD, Zeitels SM, Burns JA, Barbu AM, Hillman RE. Average Ambulatory Measures of Sound Pressure Level, Fundamental Frequency, and Vocal Dose Do Not Differ Between Adult Females With Phonotraumatic Lesions and Matched Control Subjects. *Ann Otol Rhinol Laryngol.* 2015;124(11):864-874. doi:10.1177/0003489415589363
8. Van Stan JH, Gustafsson J, Schalling E, Hillman RE. Direct Comparison of Three Commercially Available Devices for Voice Ambulatory Monitoring and Biofeedback. *Perspect Voice Voice Disord.* 2014;24(2):80. doi:10.1044/vvd24.2.80
9. Carullo A, Vallan A, Astolfi A. A low-cost platform for voice monitoring. In: *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC).* IEEE; 2013:67-72. doi:10.1109/I2MTC.2013.6555383
10. Colton RH, Casper JK, Leonard R. *Understanding Voice Problems : A Physiological Perspective for Diagnosis and Treatment.* 4ª. Baltimore, Md. : Lippincott Williams & Wilkins; 2011.
11. Artkoski M, Tommila J, Laukkanen A-M. Changes in voice during a day in normal voices without vocal loading. *Logoped Phoniatr Vocol.* 2002;27(3):118-123. doi:10.1080/140154302760834840
12. Laukkanen A-M, Ilomäki I, Leppänen K, Vilkmán E. Acoustic measures and self-reports of vocal fatigue by female teachers. *J Voice.* 2008;22(3):283-289. doi:10.1016/j.jvoice.2006.10.001
13. Lehto L, Laaksonen L, Vilkmán E, Alku P. Changes in objective acoustic measurements and subjective voice complaints in call center customer-service advisors during one working day. *J Voice.* 2008;22(2):164-177. doi:10.1016/j.jvoice.2006.08.010
14. Rantala L, Vilkmán E. Relationship between subjective voice complaints and acoustic parameters in female teachers' voices. *J Voice.* 1999;13(4):484-495. doi:10.1016/S0892-1997(99)80004-6
15. Rantala L, Vilkmán E, Bloigu R. Voice changes during work: subjective complaints and

- objective measurements for female primary and secondary schoolteachers. *J Voice*. 2002;16(3):344-355. doi:10.1016/S0892-1997(02)00106-6
16. Jónsdóttir V, Laukkanen A-M, Vilkmán E. Changes in Teachers' Speech during a Working Day with and without Electric Sound Amplification. *Folia Phoniatr Logop*. 2002;54(6):282-287. doi:10.1159/000066149
 17. Vilkmán E, Lauri ER, Alku P, Sala E, Sihvo M. Effects of prolonged oral reading on F0, SPL, subglottal pressure and amplitude characteristics of glottal flow waveforms. *J Voice*. 1999;13(2):303-312. doi:10.1016/S0892-1997(99)80036-8
 18. Vilkmán E. Occupational safety and health aspects of voice and speech professions. *Folia Phoniatr Logop*. 2004;56(4):220-253. doi:10.1159/000078344
 19. Leena Rantala, Päivi Lindholm E. F0 change due to voice loading under laboratory and field conditions. A pilot study. *Logop Phoniatr Vocology*. 1998;23(4):164-168. doi:10.1080/140154398434068
 20. Marcus Wirebrand. Real-time monitoring of voice characteristics using accelerometer and microphone measurements. *Electr Eng*. 2011.
 21. Popolo PS, Rogge-Miller K, Svec JG, Titze IR. Technical considerations in the design of a wearable voice dosimeter. *J Acoust Soc Am*. 2002;112(5):2304-2304. doi:10.1121/1.4779276
 22. Matic A, Osmani V, Mayora O. Speech activity detection using accelerometer. *Conf Proc . Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf*. 2012;2012:2112-2115. doi:10.1109/EMBC.2012.6346377
 23. Carullo A, Vallan A, Astolfi A. Design Issues for a Portable Vocal Analyzer. *IEEE Trans Instrum Meas*. 2013;62(5):1084-1093. doi:10.1109/TIM.2012.2236724
 24. Svec JG, Titze IR, Popolo PS. Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *J Acoust Soc Am*. 2005;117(3 Pt 1):1386-1394.
 25. Mehta DD, Zanartu M, Van Stan JH, Feng SW, Cheyne HA, Hillman RE. Smartphone-based detection of voice disorders by long-term monitoring of neck acceleration features. In: *2013 IEEE International Conference on Body Sensor Networks*. IEEE; 2013:1-6. doi:10.1109/BSN.2013.6575517
 26. Searl J, Dietsch A. Testing of the VocaLog Vocal Monitor. *J Voice*. 2014;28(4):523.e27-523.e37. doi:10.1016/j.jvoice.2014.01.009
 27. Masuda T, Ikeda Y, Manako H, Komiyama S. Analysis of vocal abuse: fluctuations in phonation time and intensity in 4 groups of speakers. *Acta Otolaryngol*. 1993;113(4):547-552.
 28. Aronson AE. *Clinical Voice Disorders : An Interdisciplinary Approach*. New York :Thieme-Stratton Corp, 1985.; 1985.
 29. Lyberg Åhlander V, Pelegrín García D, Whitling S, Rydell R, Löfqvist A. Teachers' voice use in teaching environments: a field study using ambulatory phonation monitor. *J Voice*. 2014;28(6):841.e5-15. doi:10.1016/j.jvoice.2014.03.006
 30. Morrow SL, Connor NP. Comparison of voice-use profiles between elementary classroom and music teachers. *J Voice*. 2011;25(3):367-372. doi:10.1016/j.jvoice.2009.11.006
 31. Mehta DD, Van Stan JH, Hillman RE. Relationships between vocal function measures

- derived from an acoustic microphone and a subglottal neck-surface accelerometer. *IEEE/ACM Trans audio, speech, Lang Process.* 2016;24(4):659-668. doi:10.1109/TASLP.2016.2516647
32. Zañartu M, Ho JC, Mehta DD, Hillman RE, Wodicka GR. Subglottal Impedance-Based Inverse Filtering of Voiced Sounds Using Neck Surface Acceleration. *IEEE Trans Audio Speech Lang Processing.* 2013;21(9):1929-1939. doi:10.1109/TASL.2013.2263138
 33. Fryd AS, Van Stan JH, Hillman RE, Mehta DD. Estimating Subglottal Pressure From Neck-Surface Acceleration During Normal Voice Production. *J Speech Lang Hear Res.* 2016;59(6):1335-1345. doi:10.1044/2016_JSLHR-S-15-0430
 34. Holbrook A, Rolnick MI, Bailey CW. Treatment of Vocal Abuse Disorders Using a Vocal Intensity Controller. *J Speech Hear Disord.* 1974;39(3):298. doi:10.1044/jshd.3903.298
 35. Zicker JE, Tompkins WJ, Rubow RT, Abbs JH. A portable microprocessor-based biofeedback training device. *IEEE Trans Biomed Eng.* 1980;27(9):509-515. doi:10.1109/TBME.1980.326665
 36. Ryu S, Komiyama S, Kannae S, Watanabe H. A Newly Devised Speech Accumulator. *ORL.* 1983;45(2):108-114.
 37. Ohlsson AC, Brink O, Löfqvist A. A voice accumulation--validation and application. *J Speech Hear Res.* 1989;32(2):451-457.
 38. McGillivray R, Proctor-Williams K, McLister B. Simple biofeedback device to reduce excessive vocal intensity. *Med Biol Eng Comput.* 1994;32(3):348-350. doi:10.1007/BF02512536
 39. Rantala L, Haataja K, Vilkman E, Körkkö P. Practical arrangements and methods in the field examination and speaking style analysis of professional voice users. *Scand J Logop Phoniatr.* 1994;19(1-2):43-54. doi:10.3109/14015439409101073
 40. Buekers R, Bierens E, Kingma H, Marres EHMA. Vocal Load as Measured by the Voice Accumulator. *Folia Phoniatr Logop.* 1995;47(5):252-261. doi:10.1159/000266359
 41. Airo E, Olkinuora P, Sala E. A Method to Measure Speaking Time and Speech Sound Pressure Level. *Folia Phoniatr Logop.* 2000;52(6):275-288. doi:10.1159/000021545
 42. Szabo A, Hammarberg B, Håkansson A, Södersten M. A voice accumulator device: evaluation based on studio and field recordings. *Logoped Phoniatr Vocol.* 2001;26(3):102-117. doi:10.1080/14015430152728016
 43. Van Stan JH, Mehta DD, Sternad D, Petit R, Hillman RE. Ambulatory Voice Biofeedback: Relative Frequency and Summary Feedback Effects on Performance and Retention of Reduced Vocal Intensity in the Daily Lives of Participants With Normal Voices. *J Speech Lang Hear Res.* 2017;60(4):853-864. doi:10.1044/2016_JSLHR-S-16-0164
 44. Lindstrom F, Wayne KP, Södersten M, McAllister A, Ternström S. Observations of the relationship between noise exposure and preschool teacher voice usage in day-care center environments. *J Voice.* 2011;25(2):166-172. doi:10.1016/j.jvoice.2009.09.009
 45. Mehta DD, Zañartu M, Feng SW, Cheyne HA, Hillman RE. Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Trans Biomed Eng.* 2012;59(11):3090-3096. doi:10.1109/TBME.2012.2207896
 46. Chwalek PC, Mehta DD, Welsh B, et al. Lightweight, on-body, wireless system for ambulatory voice and ambient noise monitoring. In: *2018 IEEE 15th International*

- Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE; 2018:205-209. doi:10.1109/BSN.2018.8329694
47. Bottalico P, Ipsaro Passione I, Astolfi A, Carullo A, Hunter EJ. Accuracy of the quantities measured by four vocal dosimeters and its uncertainty. *J Acoust Soc Am*. 2018;143(3):1591-1602. doi:10.1121/1.5027816
 48. Instruction Manual for VocaLog 2™ Vocal Activity Monitor. http://www.vocalog.com/VL2_Help/VocaLog2_Help.html. Published 2014. Accessed April 9, 2019.
 49. Hillman RE, Mehta DD. Ambulatory Monitoring of Daily Voice Use. *Perspect Voice Disord*. 2011;21(2):56-61. doi:10.1044/vvd21.2.56
 50. Hunter EJ. Vocal Dose Measures : General Rationale , Traditional Methods and Recent Advances. In: *Communicative Sciences and Disorders Michigan State University*. ; 2016.
 51. Mehta DD, Chwalek PC, Quatieri TF, Brattain LJ. Wireless neck-surface accelerometer and microphone on flex circuit with application to noise-robust monitoring of Lombard speech. In: *Interspeech*. ; 2017.
 52. Hunter EJ, Titze IR. Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings. *J Speech Lang Hear Res*. 2010;53(4):862-875. doi:10.1044/1092-4388(2009/09-0040)
 53. Morrow SL, Connor NP. Voice amplification as a means of reducing vocal load for elementary music teachers. *J Voice*. 2011;25(4):441-446. doi:10.1016/j.jvoice.2010.04.003
 54. Gaskill CS, O'Brien SG, Tinter SR. The effect of voice amplification on occupational vocal dose in elementary school teachers. *J Voice*. 2012;26(5):667.e19-27. doi:10.1016/j.jvoice.2011.10.010
 55. Astolfi A, Puglisi GE, Cutiva LCC, Pavese L, Carullo A, Burdorf A. Associations between Objectively-measured Acoustic Parameters and Occupational Voice Use among Primary School Teachers. *Energy Procedia*. 2015;78:3422-3427. doi:10.1016/j.egypro.2015.12.330
 56. Titze IR, Hunter EJ, Svec JG. Voicing and silence periods in daily and weekly vocalizations of teachers. *J Acoust Soc Am*. 2007;121(1):469-478.
 57. Franca MC, Wagner JF. Effects of vocal demands on voice performance of student singers. *J Voice*. 2015;29(3):324-332. doi:10.1016/j.jvoice.2014.07.004
 58. Echternach M, Nusseck M, Dippold S, Spahn C, Richter B. Fundamental frequency, sound pressure level and vocal dose of a vocal loading test in comparison to a real teaching situation. *Eur Arch Otorhinolaryngol*. 2014;271(12):3263-3268. doi:10.1007/s00405-014-3200-6
 59. Gama ACC, Santos JN, Pedra E de FP, Rabelo ATV, Magalhães M de C, Casas EB de Las. Vocal dose in teachers: correlation with dysphonia. *CoDAS*. 2016;28(2):190-192. doi:10.1590/2317-1782/20162015156
 60. Van Stan JH, Mehta DD, Hillman RE. The Effect of Voice Ambulatory Biofeedback on the Daily Performance and Retention of a Modified Vocal Motor Behavior in Participants With Normal Voices. *J Speech Lang Hear Res*. 2015;58(3):713-721. doi:10.1044/2015_JSLHR-S-14-0159

61. Astolfi A, Carullo A, Pavese L, Puglisi GE. Duration of voicing and silence periods of continuous speech in different acoustic environments. *J Acoust Soc Am*. 2015;137(2):565-579. doi:10.1121/1.4906259
62. Misono S, Banks K, Gaillard P, Goding GS, Yueh B. The clinical utility of vocal dosimetry for assessing voice rest. *Laryngoscope*. 2015;125(1):171-176. doi:10.1002/lary.24887
63. Carullo A, Casassa F, Castellana A, Astolfi A, Pavese L, Puglisi GE. Performance Comparison of Different Contact Microphones Used for Voice Monitoring. 2015;(July):12-16. doi:10.13140/RG.2.1.1364.3280
64. Hunter EJ. Teacher response to ambulatory monitoring of voice. *Logoped Phoniatr Vocol*. 2012;37(3):133-135. doi:10.3109/14015439.2012.664657
65. Klatt DH. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J Acoust Soc Am*. 1976;59(5):1208-1221.
66. Mehta DD, Van Stan JH, Zañartu M, et al. Using Ambulatory Voice Monitoring to Investigate Common Voice Disorders: Research Update. *Front Bioeng Biotechnol*. 2015;3(October):155. doi:10.3389/fbioe.2015.00155
67. Astolfi A, Carullo A, Vallan A, Pavese L. Influence of classroom acoustics on the vocal behavior of teachers. *Proc Meet Acoust*. 2013;19(1):40123. doi:10.1121/1.4800427
68. Carullo A, Vallan A, Astolfi A, Pavese L, Puglisi GE. Validation of calibration procedures and uncertainty estimation of contact-microphone based vocal analyzers. *Meas J Int Meas Confed*. 2015;74:130-142. doi:10.1016/j.measurement.2015.07.011

XI. GLOSARIO

_q15

El formato de representación de números en sistemas digitales con punto fijo y 15 bits fraccionales.

Acelerómetro

Aparato que mide la aceleración de un cuerpo. En este caso mide la aceleración de la piel sobre la que se apoya, generadas por las vibraciones de las cuerdas vocales.

Armónicos

Cada una de las ondas simples en que se descompone una onda acústica compleja y periódica, según el Teorema de Fourier. La primera de estas ondas, o primer armónico, es la frecuencia fundamental de la onda acústica.

Autocorrelación

Se define como la correlación cruzada de la señal consigo misma. En este documento no se distingue de la correlación y por ello se usan indistintamente.

Bit

En informática y otras disciplinas, unidad mínima de información, que puede tener solo dos valores (cero o uno).

Byte

Conjunto de 8 bits.

Convertor analógico digital (ADC)

Dispositivo electrónico capaz de convertir una señal analógica, ya sea de tensión o corriente, en una señal digital mediante un cuantificador y codificándose código binario.

De-bugging

Depuración del programa. Proceso de localización y eliminación de errores de programación.

DMA

El acceso directo a memoria (DMA, del inglés direct memory access) permite a cierto tipo de componentes de un microcontrolador acceder a la memoria del sistema para leer o escribir independientemente de la unidad central de procesamiento (CPU) principal. Se utiliza para liberar al CPU de esas tareas y así mejorar el rendimiento.

Exacto

Que se ajusta lo más posible al valor real de una magnitud.

Fiable

Creíble, fidedigno, sin error.

Flujo transglótico

Flujo de aire que atraviesa las cuerdas vocales durante la fonación en l/s

Fonetograma

Representación gráfica de la capacidad fonatoria de la laringe. Rango dinámico de la voz.

FRAM

Tipo de memoria similar a la RAM con la peculiaridad de que no es volátil.

Frecuencia fundamental (F_0)

Número de ciclos vocales por segundo, expresado en hercios (Hz). Se considera que esta corresponde con el primer armónico

Glottis

Porción más estrecha de la luz laríngea, espacio que está limitado por las cuerdas vocales, la porción vocal de los aritenoides y el área interaritenoides.

Intensidad

Magnitud física que expresa la mayor o menor amplitud de las ondas sonoras.

Jitter

Parámetro de perturbación de la frecuencia fundamental.

kB

Un Kilobyte (abreviado como KB o Kbyte) es una unidad de medida equivalente a 1024 bytes de memoria.

El teorema de Nyquist

La reconstrucción exacta de una señal periódica continua en base a sus muestras, es matemáticamente posible si la señal está limitada en banda y la tasa de muestreo es superior al doble de su ancho de banda.

LED

Diodo semiconductor que emite luz cuando se le aplica tensión.

Lenguaje C++

Lenguaje de programación orientado a objetos que toma la base del lenguaje C.

Macro

Estos básicamente son un alias que podemos incluir en nuestro código el cual, al momento de compilar, será reemplazado por lo que hayamos definido. La definición de la macro `_Q15()` se encuentra en `DSPLib_1_30_00_02\include\DSPLib_support.h`.

Microcontrolador

Circuito integrado programable, capaz de ejecutar las órdenes grabadas en su memoria. Se diferencia de un microprocesador en que dispone de muchos subsistemas como memoria interna RAM y ROM, puertos de comunicación, ADCs, etc. en un mismo circuito integrado formando un sistema digital.

Micrófono de contacto

Sensor que capta las vibraciones de aquellos objetos con los que entran en contacto.

MicroSD

Formato para tarjetas de memoria flash para el almacenamiento de archivos digitales en dispositivos electrónicos.

Muestra

Es un valor numérico en función del tiempo. Este valor es parte de una señal continua o de una señal discreta y son extraídos de éstas para un procesamiento matemático.

Número complejo

Número que se expresa como la suma de un número real y otro imaginario.

Preciso

Que permite medir magnitudes con un error mínimo.

Presión subglótica

Presión registrada debajo de los pliegues vocales generada por el paso de aire a través de la glotis durante la fonación de una vocal sostenida medido en mmH₂O.

RAM

Siglas de Random Access Memory ('memoria de acceso aleatorio'), memoria principal de la computadora, donde residen programas y datos, sobre la que se pueden efectuar operaciones de lectura y escritura.

Relación señal ruido

Se define como la proporción existente entre la potencia de la señal que se transmite y la potencia del ruido que la corrompe.

Resistencia laríngea

Relación entre presión subglótica y flujo transglótico en mmH₂O/l/s.

Root Mean Square

Media cuadrática, valor cuadrático medio, raíz de la media cuadrática o RMS (del inglés root mean square) es una medida estadística de la magnitud de una cantidad variable. Puede calcularse para una serie de valores discretos o para una función matemática de variable continua. El nombre deriva del hecho de que es la raíz cuadrada de la media aritmética de los cuadrados de los valores.

RTC

Un Reloj en tiempo real (en inglés, real-time clock, RTC) es un reloj de un ordenador, incluido en un circuito integrado, que mantiene la hora actual.

Shimmer

Parámetro de perturbación de la intensidad.

SPI

El Bus SPI (del inglés Serial Peripheral Interface) es un estándar de comunicaciones, usado principalmente para la transferencia de información entre circuitos integrados en equipos electrónicos. El bus de interfaz de periféricos serie o bus SPI es un estándar para controlar casi cualquier dispositivo electrónico digital que acepte un flujo de bits serie regulado por un reloj (comunicación sincrónica).

SRAM

Son las siglas de la voz inglesa Static Random Access Memory, que significa memoria estática de acceso aleatorio (o RAM estática), para denominar a un tipo de tecnología de memoria RAM basada en semiconductores, capaz de mantener los datos, mientras siga alimentada, sin necesidad de circuito de refresco.

Supraglótico

Se dice de aquello que se encuentra encima de la glotis.

Tiempo de fonación

Tiempo durante en el que los pliegues vocales ondulan.

Tiempo real

Se dice del proceso cuyo tiempo de ejecución del procesado de los datos es mejor al tiempo de adquisición de los datos.

Transformada rápida de Fourier

Es un algoritmo que permite calcular eficientemente la Transforma Discreta de Fourier.

Unidireccional

Se dice de un acelerómetro que solo es sensible a la aceleración a en uno de sus ejes.

XII. ANEXOS

ANEXO I. DOCUMENTO TÉCNICO VOXLOG

SONVOX

TECHNICAL DOCUMENT: VOXLOG

2015 - 12 - 01

VoxLog

The VoxLog is a voice measurement device, capable of capturing a number of aspects of the voice. The VoxLog uses a neck collar that contains an accelerometer and a microphone, and uses the signals obtained from these during speech to extract and record parameters of the voice.



Measurements

The VoxLog measures four aspects in essence: F0 (pitch), SPL (sound pressure level) for both voice and noise and PT% (Phonation time in %). F0 is measured using an FFT-based pitch extraction method. Phonation time is measured for when the device detects speech. SPL Voice is then measured for the intervals where speech is present, and SPL Noise is then measured when speech is absent. Unlike many other dosimeters, the VoxLog does not need calibration for SPL. It works the same between different persons. If the collar is correctly placed (see above) and the vocal cords are in contact with the sensor, The VoxLog SPL dB minus 10 dB is approximately equal to SPL measured at 30 cm in front of the mouth. The VoxLog extracts the values of F0, SPL and PT directly from the signal and stores them, instead of storing raw data. This means that it is not possible to recreate the actual sound recording from the VoxLog recording, which secures the privacy of VoxLog users.



Collar

The VoxLog collar outputs a stereo signal, one channel containing the microphone signal and the other channel containing the accelerometer signal. The collar needs to be powered by 4-10V and the VoxLog supplies 6V to the collar. The collar is adjustable and allows for a snug fit for both male and female users with any neck size.

Recording and Memory

A new recording is made each time the VoxLog is turned off and on again (while not connected to a computer). VoxLog's internal memory can store up to 100 recordings, with each recording having a maximum length of ~14 hours. This accounts for a total maximum of ~1400 hours of voice data. The VoxLog can be connected to a computer, and using the software VoxLog Discovery, each individual recording can be imported and stored on the computer. Using the software, recordings can be associated with specific persons, allowing the use of a single VoxLog for many persons. After importing recordings, the recordings can be removed from the device to allow for more space if necessary.

Power

The VoxLog device is powered by a chargeable battery, and can operate a full day on an overnight charging. The battery is charged using a power plug, specified as follows — Input Voltage: 100-240V, ~0.3A — Input Frequency: 50-60 Hz — Output Voltage: 5V, 1A — Max Output Power: 6W

Settings

The VoxLog can be set to record datapoints at different time intervals. For example, a 5 second or even 60 second time window is often used for longer recordings, while a window of 100 milliseconds is suitable for short test recordings or when more granularity in the data is needed. Note that high resolution recordings (100 ms and 1s) have a maximum length of 2 hours.

The VoxLog also has settings for A-weighting, calibration offset, and feedback. The feedback functionality is in the form of vibration of the device and works for both F0 and SPL, where a low or high threshold can be set for which the device should give feedback (vibrate). Some specific settings for the vibration can also be set, such as the length of vibrations and delay between them. All settings can be changed using the software VoxLog Discovery.

Exporting / Importing data

The VoxLog Discovery software has functionality for exporting and importing databases of persons, recordings and analyses, to allow for easily creating backups or sharing the data between work stations.

Data Analysis

When the recordings has been imported into the VoxLog Discovery software, they can be analysed either individually or across many recordings and or persons. For easy analysis, recordings and specific intervals in recordings can be tagged, and analyses can be made upon certain tags or combinations of intervals, recordings and persons. Parameters such as cycle dose and statistics are calculated by the software and can be viewed using the analysis tool in VoxLog Discovery. The F0, SPL and Phonation data of recordings, as well as analysis statistics, can also easily be exported to different formats (csv, excel, pdf) to be used in further calculations outside the software.



ANEXO II. VOCAL HOLTER – PROTOTYPE MED

1.1.1



Vocal Holter Med – Prototype 1.1.1

Items

- Data Acquisition and Processing (DAP) unit (size about 15x10x5 cm)
- Contact microphone (model hx-505-1-1) – see attached picture
- Audio microphone
- Instruction manual

DAP unit - main features

- ✓ Battery powered (battery included – about eight hours of battery life)
- ✓ Two input channels for contact and audio microphones (sampling frequency 22050 Sa/s, resolution 12 bit)
- ✓ Embedded temperature and relative humidity sensor
- ✓ USB port for data transfer to a PC and for battery recharge from PC or power adapter (not provided)
- ✓ User interaction through switch and led*

Measured parameters

- ✓ Sound Pressure Level (*SPL*) – preliminary calibration required
- ✓ Fundamental frequency (*F0*)
- ✓ Voicing time percentage (*D_{t%}*)
- ✓ Air temperature (*θ*)
- ✓ Air relative humidity (*h*)

Measured parameters are stored in the internal memory of the DAP unit.

Raw samples can also be stored for post-processing purposes (maximum recording time: about 5 hours)

Price: 900,00 euro (VAT excluded)

Calibration of the reference channel not included**
Audio microphone support and spacer non provided

Payment: 50% at the order; 50% before shipment

Availability: about two months from your order

* Model equipped with touch-screen display available: price 1500 euro, available within four months

** Calibration option (traceability to international standard ensured) available – price 200,00 euro

ANEXO III. VOCAL HOLTER – PROTOTYPE MED

2.0.0



Vocal Holter Med – Prototype 2.0.0

Items

- Data Acquisition and Processing (DAP) unit (size about 15x10x5 cm)
- Contact microphone (model hx-505-1-1) – see attached picture
- Audio microphone and spacer embedded in the DAP unit
- Power adapter
- Instruction manual

DAP unit - main features

- ✓ Battery powered (rechargeable battery included – about six hours of battery life), connector for battery recharge from power adapter
- ✓ Two acquisition channels for contact (external) and audio (internal) microphones (sampling frequency 22050 Sa/s, resolution 12 bit)
 - Audio microphone is used during preliminary calibration and during long-term monitoring (estimation of the background noise level)
- ✓ Embedded temperature and relative humidity sensors
- ✓ Control/monitoring by means of smartphone or PC through remote interface

Measured parameters*

- ✓ Sound Pressure Level (*SPL*) – preliminary calibration required
- ✓ Fundamental frequency (*F0*)
- ✓ Voicing time percentage (*D_t%*)
- ✓ Local jitter (%), local shimmer (%) and CPPS (Cepstral Peak Prominence Smoothed) related to the uttering of a continuous vowel
- ✓ Air temperature (*θ*)
- ✓ Air relative humidity (*h*)

Measured parameters, which are stored in the internal memory of the DAP unit, can be transferred to a PC through a wireless connection.

Raw samples can also be stored for post-processing purposes (maximum recording time: about 6 hours)

Price: 1800,00 euro (VAT excluded)

Calibration of the reference channel not included**

Payment: 50% at the order; 50% before shipment

Availability: about two months from your order

PR.O.VOICE S.r.l. – P.IVA 11231510014
 Corso Castellidardo 30/A – 10129 Torino (Italy)
 Tel. 011.0904202 – Fax 011.0905126
 e-mail: info@provoicesrl.com – web-site www.provoicesrl.com



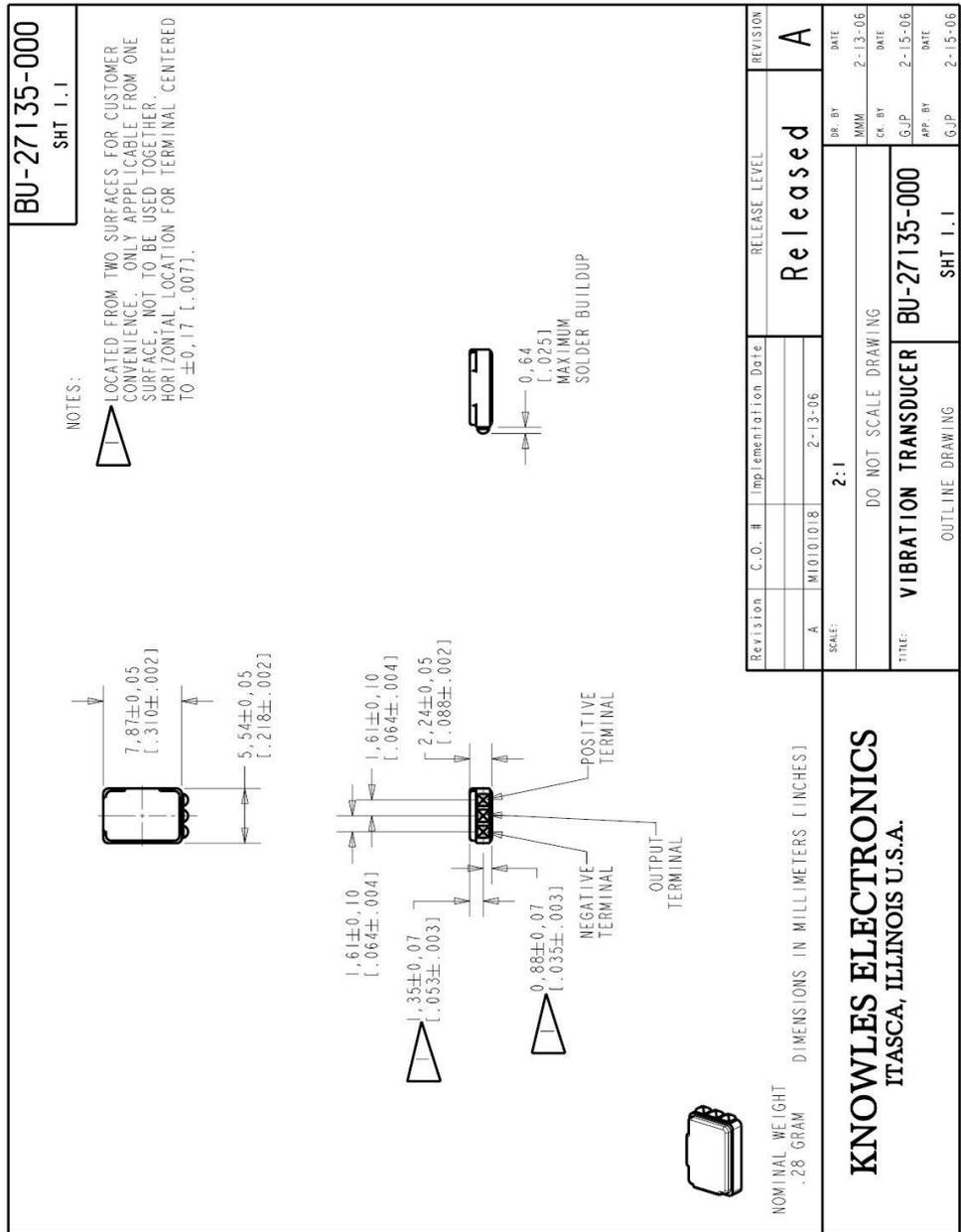
* Other parameters can be estimated according to customers' requirements; firmware upgrade will be proposed when available

** Calibration option (traceability to international standard ensured) available – price 200,00 euro

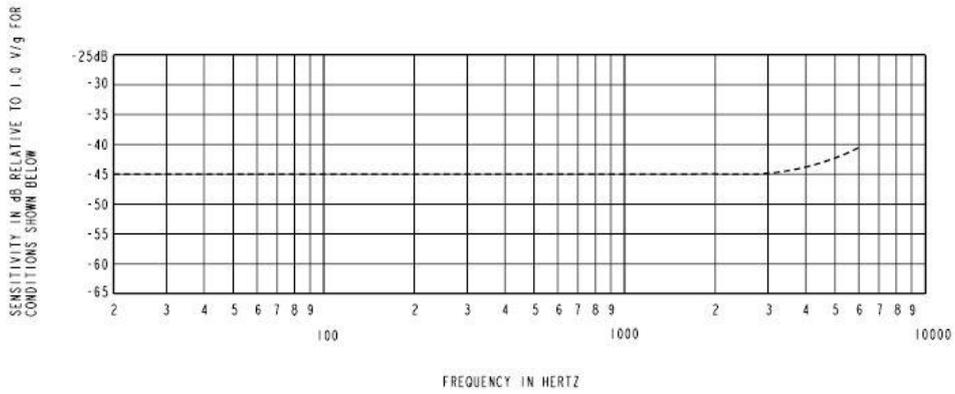


Contact microphone

ANEXO IV. DATASHEET ACELERÓMETRO



BU-27135-000
SHEET 2.1

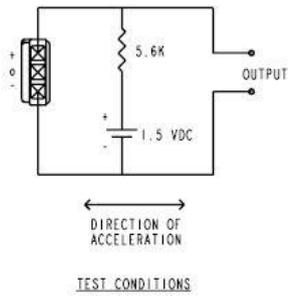


FREQUENCY	SENSITIVITY			DEVICE CONFORMITY	
	MIN.	NOM.	MAX.	RANGE OF DEVIATION FROM 3 kHz	
300	---	-45.0	---	-1.5	+1.5
1000	-49.5	-45.0	-40.5	0.0	0.0
3000	---	-44.5	---	-1.0	+2.0

NOTES:

1. ALL DATA MEASURED WITH CIRCUIT CONFIGURATION SHOWN BELOW.
2. CASE CONNECTED TO NEGATIVE TERMINAL.

DC SUPPLY	BATTERY CURRENT DRAIN	"A" WEIGHTED NOISE (RE 1.0 VOLT)	OUTPUT IMPEDANCE OHMS (AT 1 kHz)		
			MIN.	NOM.	MAX.
1.5V	0.3 mA MAX.	-103 dB MAX.	4900	5200	5500



Revision	C.O. #	Implementation Date	RELEASE LEVEL	REVISION
A	W10101B	2-13-06	Released	A

WHEN TEST LIMITS ARE USED TO ESTABLISH INCOMING INSPECTION ACCEPTANCE REJECTION CRITERIA, CORRELATION OF TEST EQUIPMENT WITH KNOWLES IS ALSO REQUIRED FOR ELIMINATION OF EQUIPMENT AND TEST METHOD VARIATION

DR. BY	DATE
MMM	2-13-06
CK. BY	DATE
GJP	2-15-06
APP. BY	DATE
GJP	2-15-06

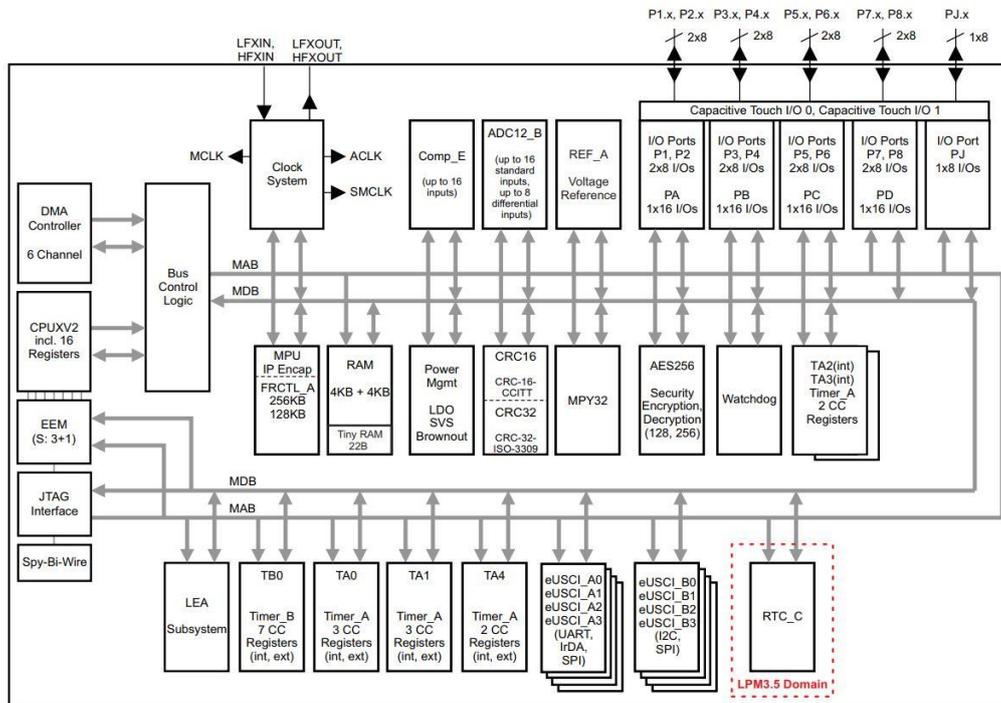
KNOWLES ELECTRONICS
ITASCA, ILLINOIS U.S.A.

TITLE: VIBRATION TRANSDUCER	BU-27135-000
PERFORMANCE SPECIFICATION	SHT 2.1

ANEXO V. DIAGRAMA DE BLOQUE DEL MICROCONTROLADOR MSP430FR5994

1.4 Functional Block Diagram

Figure 1-1 shows the functional block diagram of the devices.



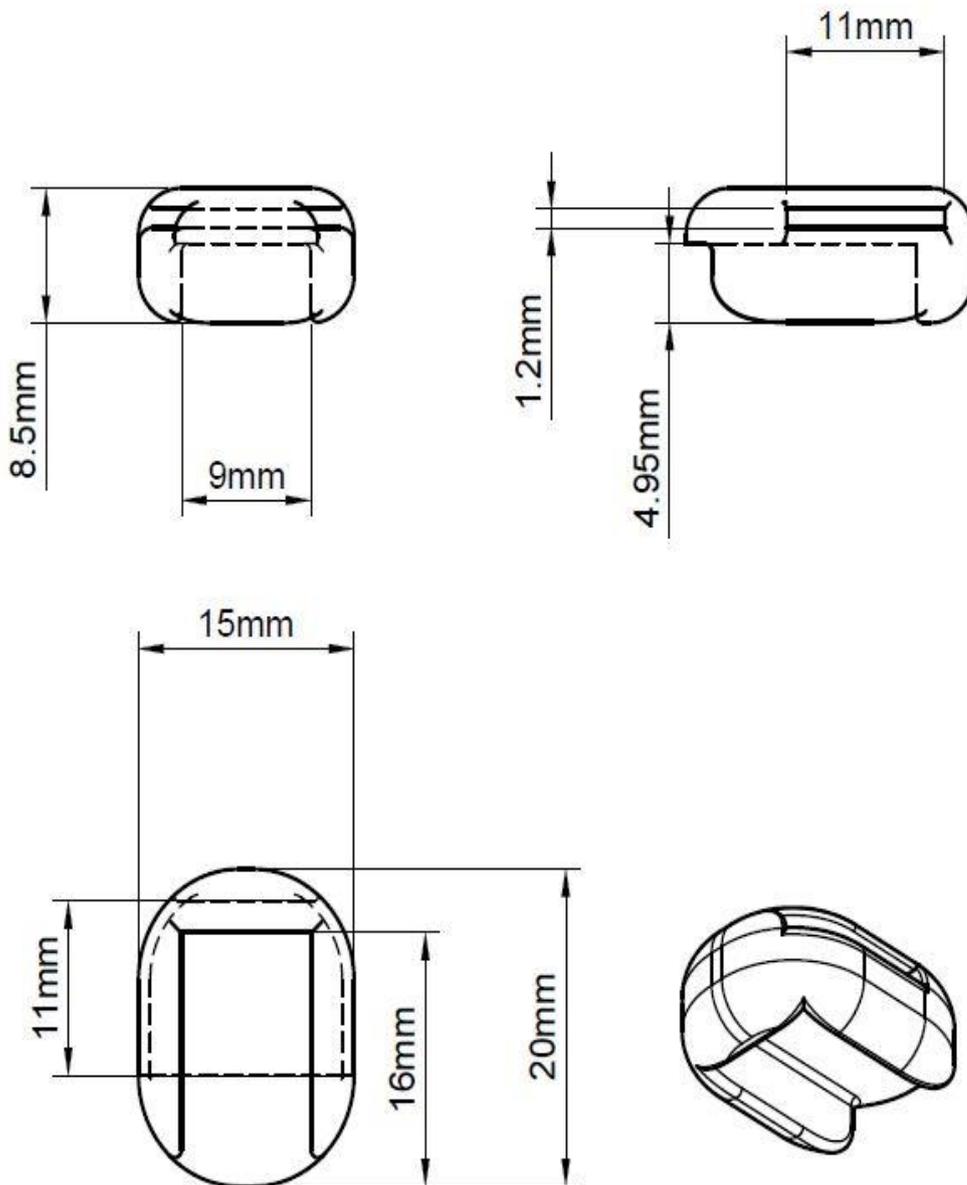
Copyright © 2016, Texas Instruments Incorporated

- A. The device has 8KB of RAM, and 4KB of the RAM is shared with the LEA subsystem. The CPU has priority over the LEA subsystem.
- B. The LEA subsystem is available on the MSP430FR599x MCUs only.

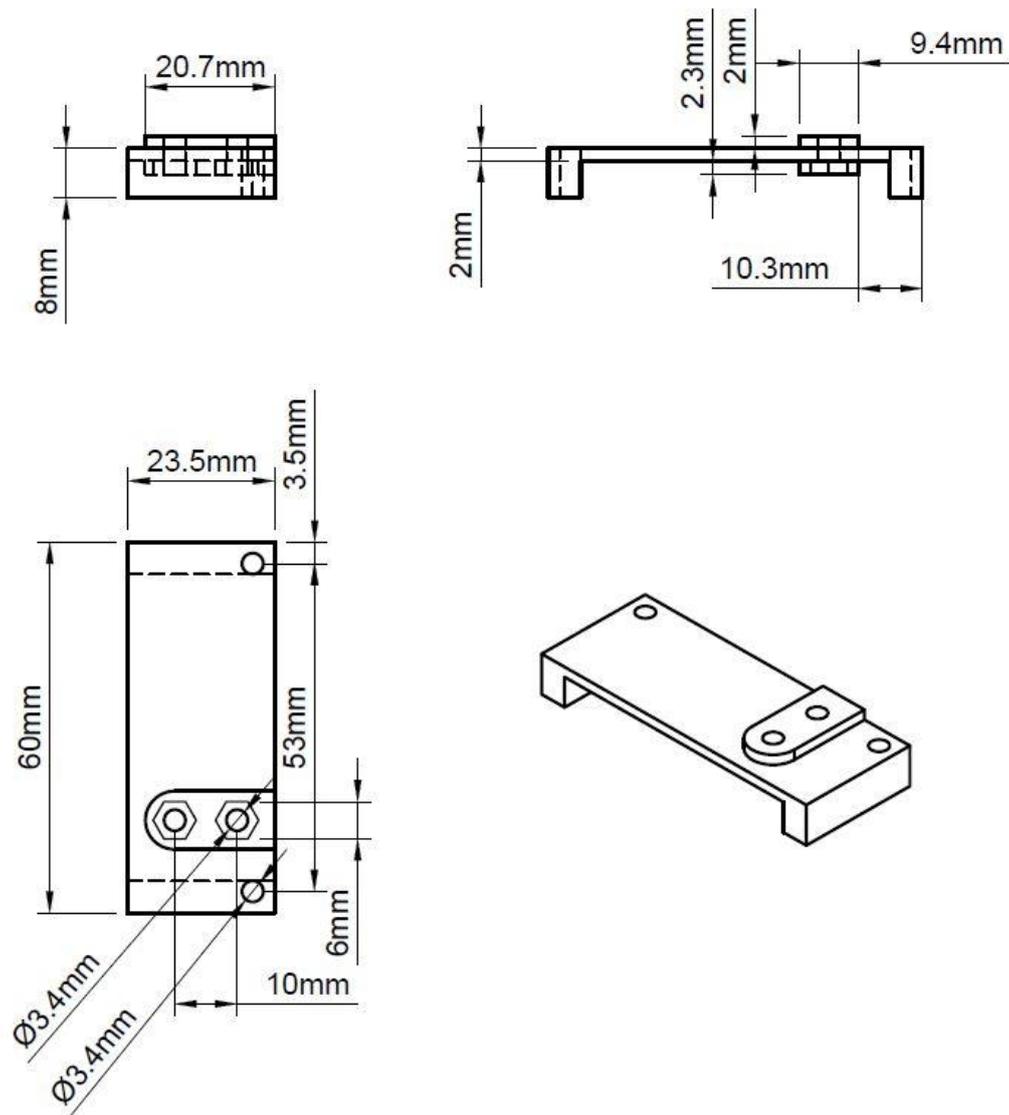
Figure 1-1. Functional Block Diagram

Datasheet completo en <http://www.ti.com/lit/ds/symlink/msp430fr5994.pdf>

ANEXO VI. DIBUJO TÉCNICO DE PROTOTIPO COLGANTE

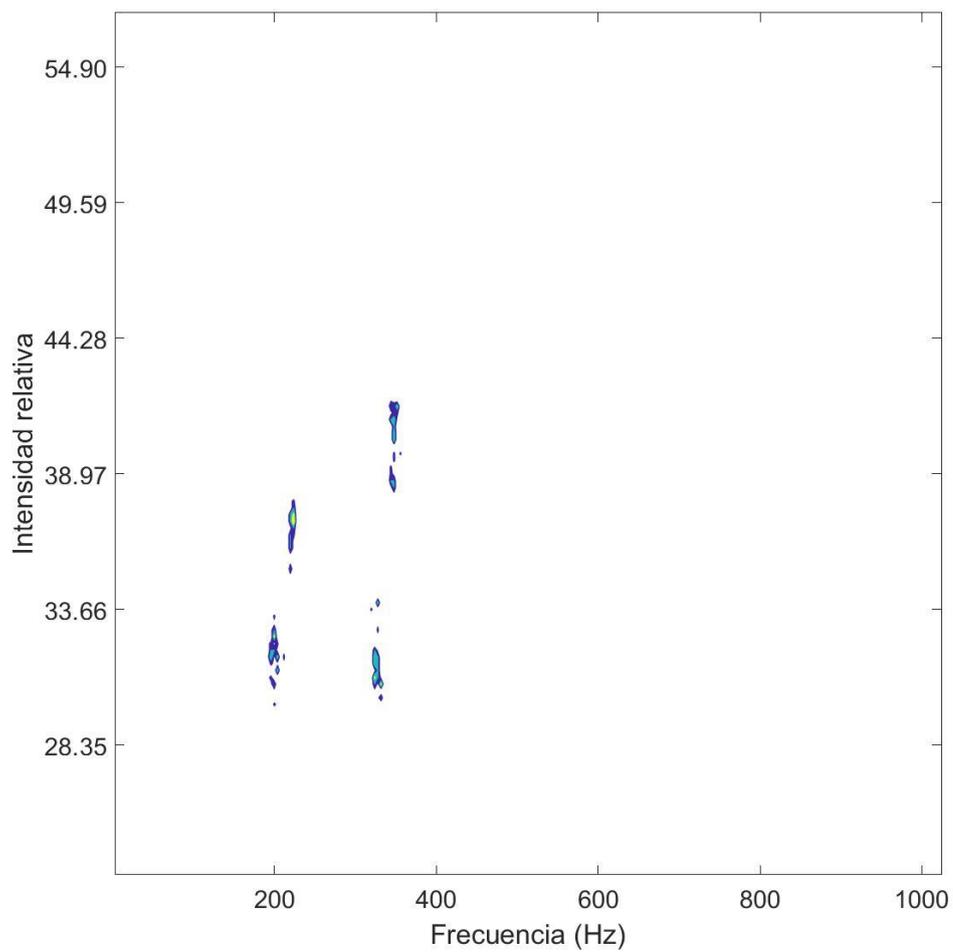


ANEXO VII. DIBUJO TÉCNICO DE SOPORTE PARA ANCLAR CIRCUITO A TARJETA DE DESARROLLO

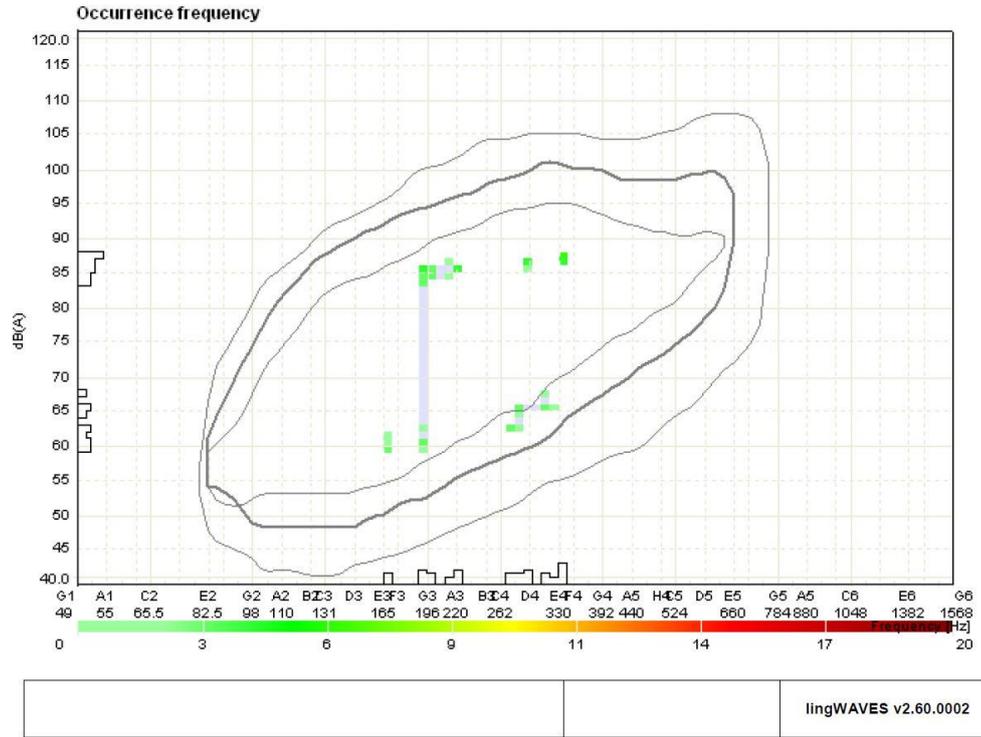


ANEXO VIII. FONETOGRAMA REGISTRADO CON EL DISPOSITIVO Y CON LINGVAWES

Fonograma registrado con el dispositivo, dos frecuencias distintas a dos intensidades distintas.



Fonetograma registrado usando sonómetro y el software LingVAWES

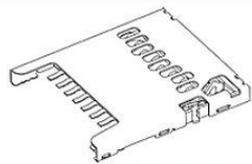


ANEXO IX. DATASHEET LECTOR DE TARJETA MICROSD



This document was generated on 04/29/2019

PLEASE CHECK WWW.MOLEX.COM FOR LATEST PART INFORMATION

Part Number:	5040771891	 <i>Series image - Reference only</i>																						
Status:	Active																							
Overview:	microSD Memory Card Connector																							
Description:	1.28mm Height Push-Pull microSD Card Socket, Surface Mount, 8 Circuits																							
Documents:																								
3D Model	Packaging Specification_SPK-504077-001-001 (PDF)																							
Drawing (PDF)	RoHS Certificate of Compliance (PDF)																							
Product Specification_PS-504077-002-001 (PDF)																								
General																								
Product Family	Memory Card Sockets																							
Series	504077																							
Overview	microSD Memory Card Connector																							
Product Name	microSD Card																							
Style	Push-Pull																							
UPC	884982892651																							
Physical																								
Card Detection Switch	Open																							
Card Entry Location	Front																							
Circuits (Loaded)	8																							
Durability (mating cycles max)	10000																							
Ejector Button Side	N/A																							
Material - Contact	Copper Alloy																							
Material - Plating Mating	Gold over Nickel																							
Material - Shell	Stainless Steel																							
Net Weight	188.245/mg																							
PCB Locator	No																							
PCB Retention	None																							
Packaging Type	Embossed Tape on Reel																							
Pitch - Mating Interface	1.10mm																							
Temperature Range - Operating	-40° to +85°C																							
Termination Interface: Style	Surface Mount																							
Electrical																								
Current - Maximum per Contact	0.5A																							
Shielded	Yes																							
Voltage - Maximum	10V AC (RMS)/DC																							
Material Info																								
Reference - Drawing Numbers																								
Packaging Specification	SPK-504077-001-001																							
Product Specification	PS-504077-002-001																							
Sales Drawing	SD-504077-001, SD-504077-002																							
<table border="1"> <tr> <td>EU ELV</td> <td>Not Relevant</td> </tr> <tr> <td>EU RoHS</td> <td>China RoHS</td> </tr> <tr> <td>Compliant</td> <td></td> </tr> <tr> <td>REACH SVHC</td> <td></td> </tr> <tr> <td colspan="2">Not Contained Per -ED/88/2018 (15 January 2019)</td> </tr> <tr> <td>Halogen-Free Status</td> <td></td> </tr> <tr> <td>Low-Halogen</td> <td></td> </tr> <tr> <td colspan="2">For more information, please visit Contact US</td> </tr> <tr> <td>China ROHS</td> <td>Green Image</td> </tr> <tr> <td>ELV</td> <td>Not Relevant</td> </tr> <tr> <td>RoHS Phthalates</td> <td>Not Contained</td> </tr> </table>			EU ELV	Not Relevant	EU RoHS	China RoHS	Compliant		REACH SVHC		Not Contained Per -ED/88/2018 (15 January 2019)		Halogen-Free Status		Low-Halogen		For more information, please visit Contact US		China ROHS	Green Image	ELV	Not Relevant	RoHS Phthalates	Not Contained
EU ELV	Not Relevant																							
EU RoHS	China RoHS																							
Compliant																								
REACH SVHC																								
Not Contained Per -ED/88/2018 (15 January 2019)																								
Halogen-Free Status																								
Low-Halogen																								
For more information, please visit Contact US																								
China ROHS	Green Image																							
ELV	Not Relevant																							
RoHS Phthalates	Not Contained																							
<table border="1"> <tr> <td>Search Parts in this Series</td> </tr> <tr> <td>504077 Series</td> </tr> </table>			Search Parts in this Series	504077 Series																				
Search Parts in this Series																								
504077 Series																								

This document was generated on 04/29/2019

PLEASE CHECK WWW.MOLEX.COM FOR LATEST PART INFORMATION

[https://www.molex.com/webdocs/datasheets/pdf/en-us/5040771891 MEMORY CARD SOCKET.pdf](https://www.molex.com/webdocs/datasheets/pdf/en-us/5040771891_MEMORY_CARD_SOCKET.pdf)

ANEXO X. DATASHEET LIPO BATERÍA

400 MAH

	DATA POWER TECHNOLOGY LIMITED	File.No: E-SPE-1106-01 Ver:1.0
	Product Specifications	Page: 4/10 Date: 2015-11-06

1. Scope

This specification shall be applied to the batteries from Data Power Technology Limited's product.

2. Product Type and Product Model

2.1 Type: Polymer Li-ion Recharged Battery

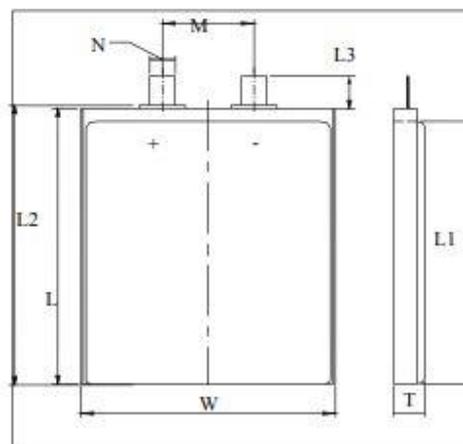
2.2 Model: DTP502535

3. Product Basic Characteristics

No	Item	Characteristics
3.1	Rated Capacity	400mAh
3.2	Minimum Capacity	400mAh
3.3	Nominal Voltage	3.70V
3.4	Charge Limited Voltage	4.20V
3.5	Discharge Cut-off Voltage	2.80V
3.6	End-of-charge Current	0.01C
3.7	Standard Charge	Charge with 0.2C(80mA) up to Limited Voltage , Charge with limited Voltage up to end-of-charge current.
3.8	Standard Discharge	Using 0.2C(80mA) constant current discharge to the Discharge Cut-off Voltage.
3.9	Maximum Continuous Charge Current	1C (400mA)
3.10	Maximum Continuous Discharge Current	1C (400mA)
3.11	Operating Temperature Range	Charge 0 – 45°C
		Discharge – 20 – 60°C
	Storage Temperature Range	-20 – 60°C
3.12	Operating And Storage Humidity Range	65 ± 20% RH
3.13	Weight	Less than 13g

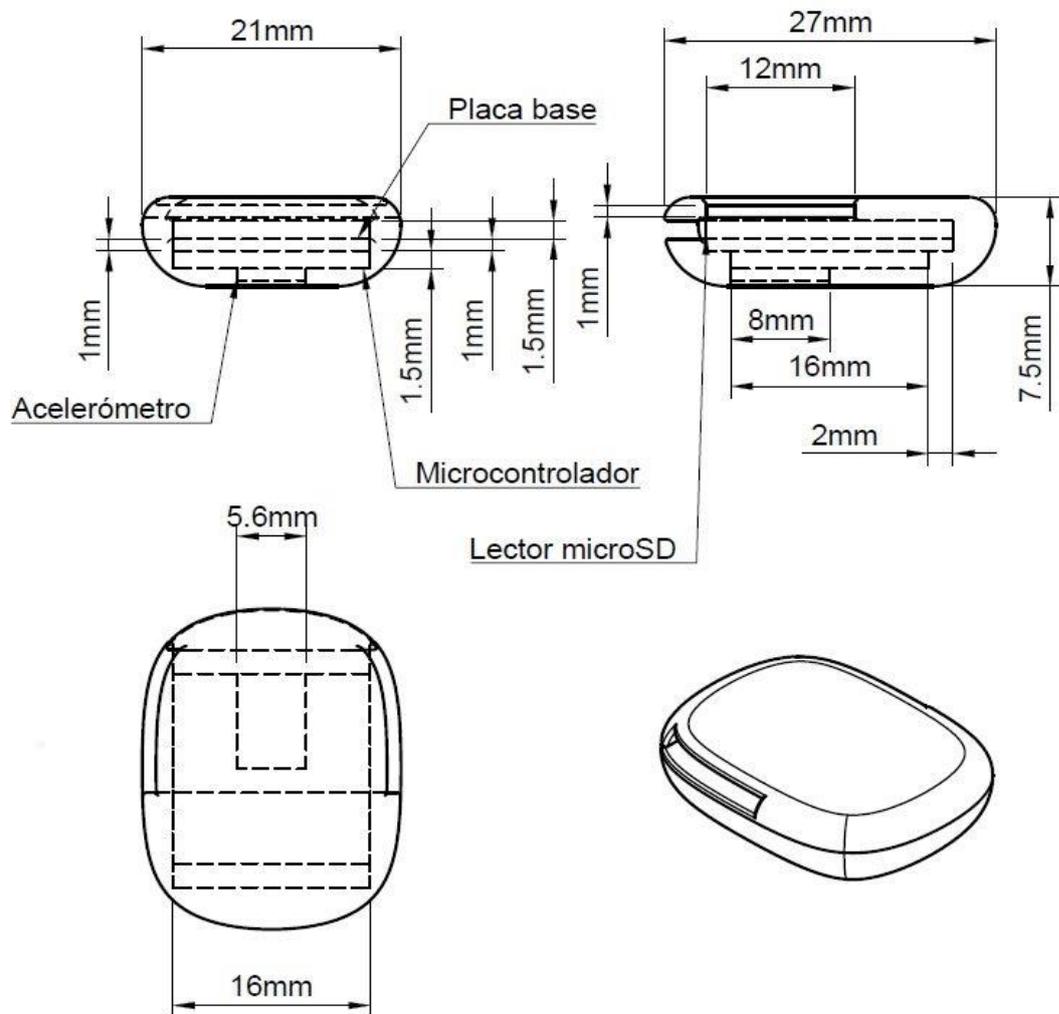
4. Cell Dimension

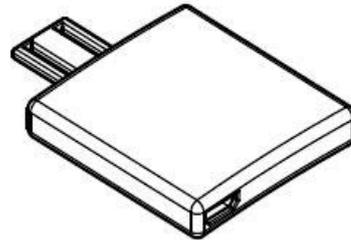
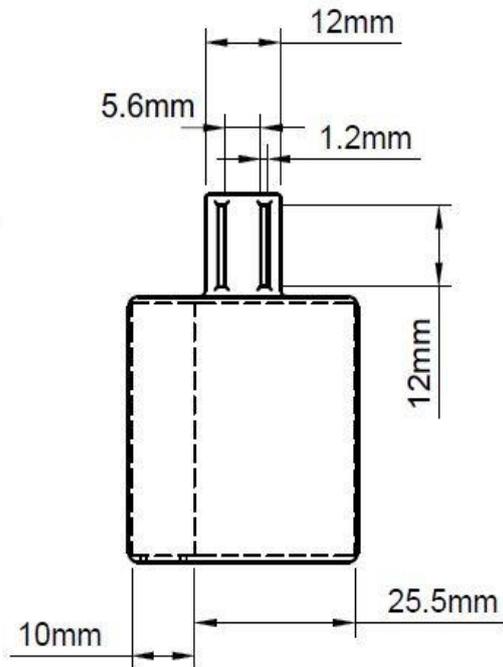
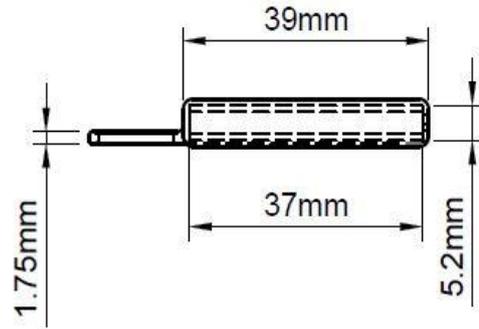
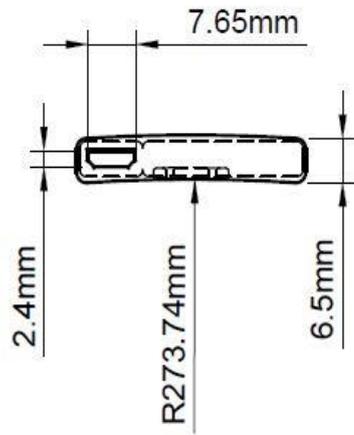
Item	Dimension (mm)
T	Max 5.0
W	Max 25.0
L	Max 35.0
L1	Max 31.0
L2	Max 35.3
L3	6.0±2.0
M	12.0±2.0
N	2.0±0.5



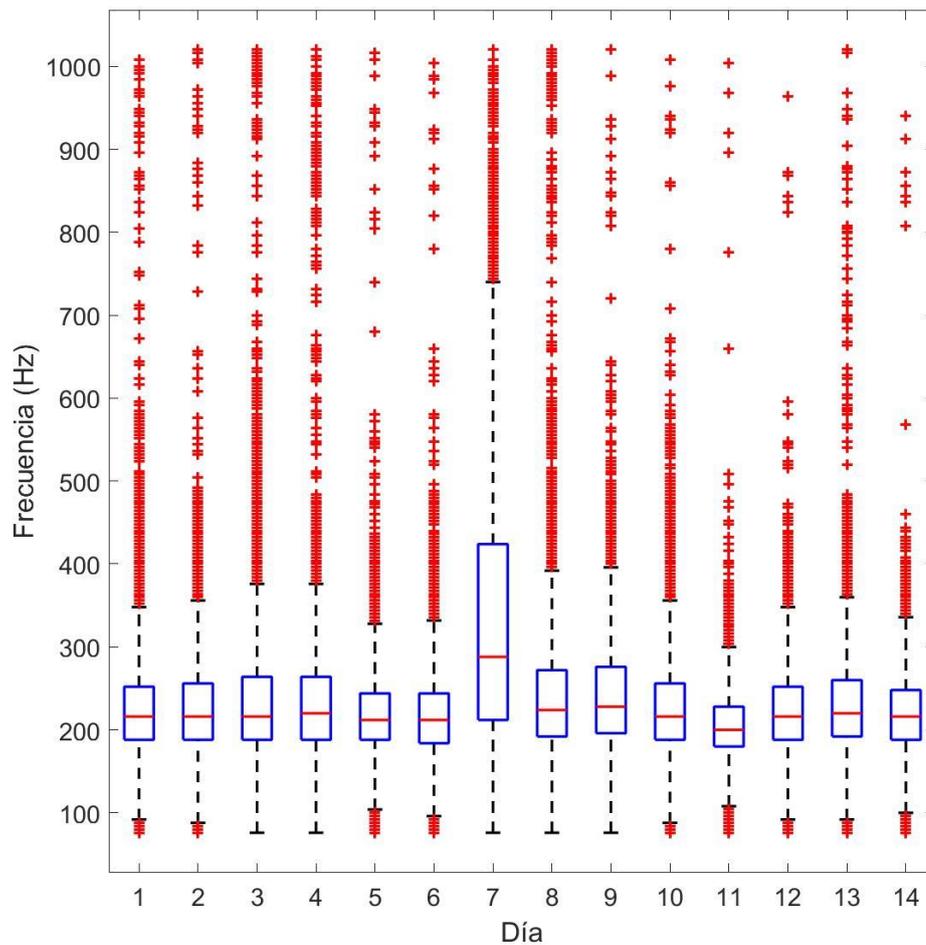
<https://cdn.sparkfun.com/datasheets/Prototyping/spe-00-502535-400mah-en-1.Over.pdf>

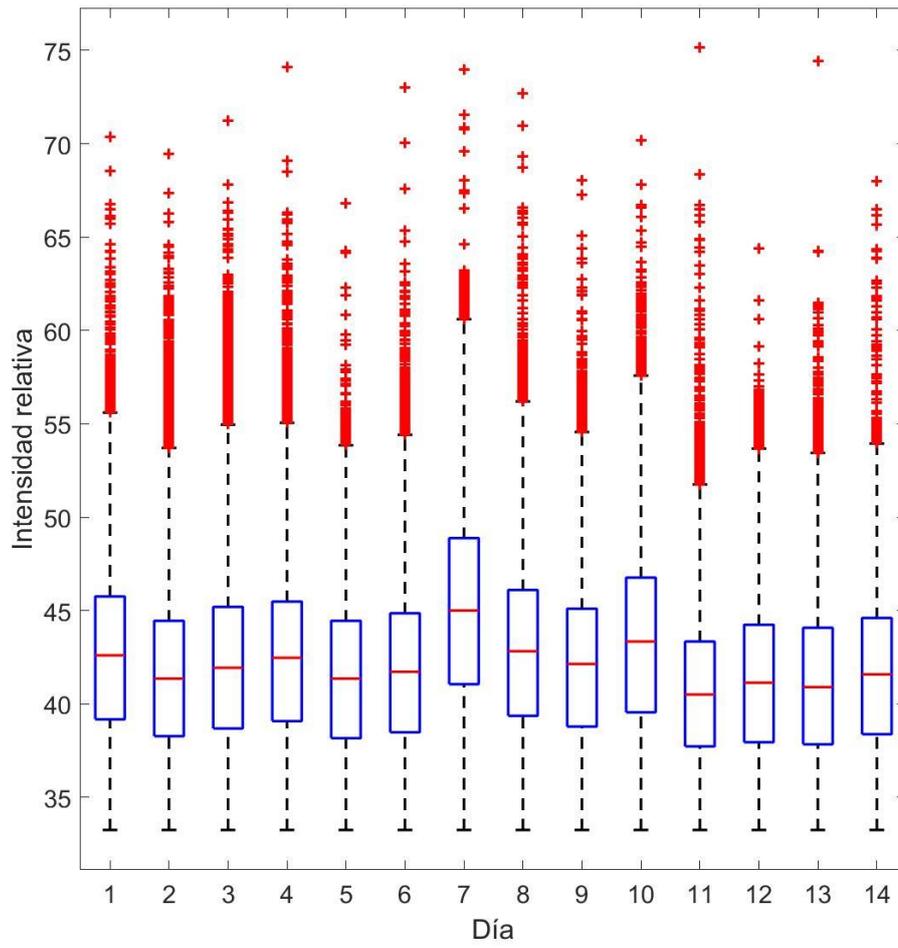
ANEXO XI. DIBUJO TÉCNICO DE DISPOSITIVO FINAL



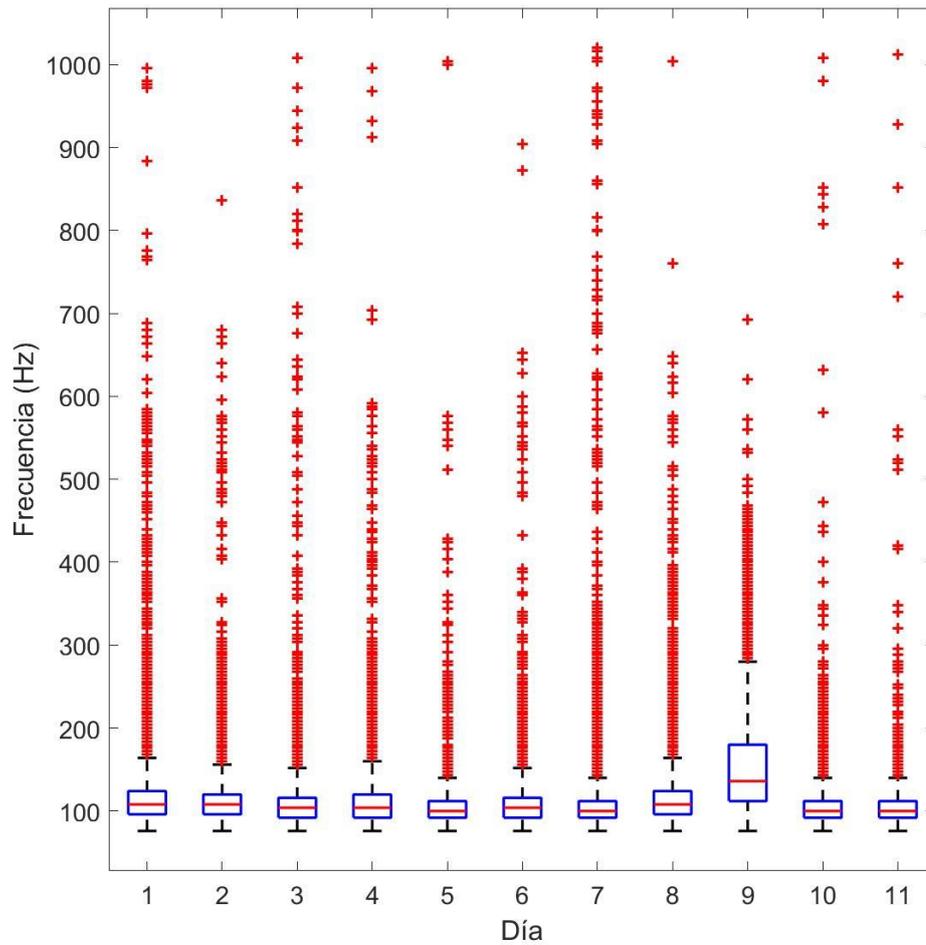


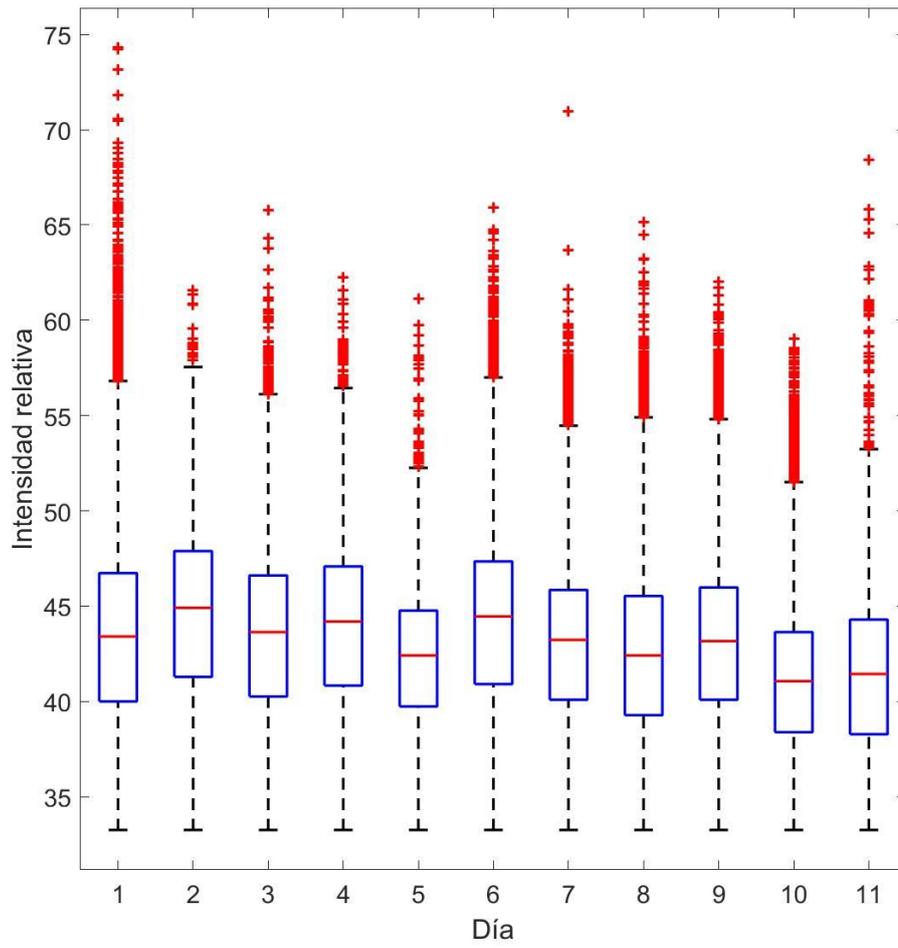
ANEXO XII. BOXPLOT DE LOS 14 DÍAS DE FRECUENCIA FUNDAMENTAL E INTENSIDAD RELATIVA DE SUJETO 1





ANEXO XIII. BOXPLOT DE LOS 11 DÍAS DE FRECUENCIA FUNDAMENTAL E INTENSIDAD RELATIVA DE SUJETO 2





ANEXO XIV. COMUNICACIONES



Marcos Llorente (Medical Engineering Laboratory, University of Navarra, Pamplona, Spain), **Secundino Fenández** (Otorhinolaryngology Department, University of Navarra, Pamplona, Spain)

BACKGROUND: Acoustic and aerodynamic analysis in the voice laboratory is a common practice in the process of assessing voice disorders. Although such analysis provides valuable information, it does not tell us how a patient uses his or her voice during different activities throughout the day. In recent years, various devices have been developed for ambulatory voice monitoring. In addition to having a phonatory signal acquisition system, most of these devices connect by cable to separate recording equipment or to a smartphone, which can process and analyse the collected signal.

AIMS: We present the prototype of a single wearable device that allows us to collect phonatory signals and process them without the need for separate recording equipment or for a smartphone. The device monitors all main phonatory parameters, such as, frequency, relative intensity and phonation time and will function independently for between 24 and 48 hours.

METHODS: The device uses the Knowles BU-21771 accelerometer, which is to be placed in contact with the skin in the suprasternal region below the larynx. The signal from this accelerometer is recorded and processed by a Texas Instruments microcontroller, and the processing results are stored on an SD memory card. The processing and analysis of the signal has been optimized for speed and low energy consumption. The prototype registers frequency and amplitude at a rate of three frames per second; from these values, phonation time is calculated. The microcontroller performs sampling and calculations in parallel, and, because the calculation time is less than the interval between samples, processing can be carried in real time.

RESULTS: Tests validated the registration procedures and verified that the device provides precise and reliable values.

CONCLUSIONS: By processing the signal with a microcontroller, it is possible to reduce the size of an ambulatory voice monitoring device to such a degree that it becomes truly wearable, whilst still monitoring all relevant acoustic voice parameters.

Póster presentado en la XII Jornada de Investigación en Ciencias Experimentales y de la Salud, Facultad de Ciencias de la Universidad de Navarra. Pamplona, 11 de abril de 2019.

Diseño de un dispositivo vestible para la monitorización de la voz

Marcos Llorente¹, Secundino Fernández^{1,2}, Adam Podhorski³.

¹ Laboratorio de Ingeniería Médica, Facultad de Medicina, Universidad de Navarra.
² Departamento de Otorrinolaringología, Clínica Universidad de Navarra.
³ Departamento de Ingeniería Biomédica y Ciencias, TECNUN, Universidad de Navarra.



Imagen 1: Comparación de dosímetros de voz¹.

INTRODUCCIÓN

El análisis acústico y aerodinámico en el laboratorio de voz es una práctica habitual en el proceso de la valoración de los trastornos de la voz. Aporta información muy valiosa pero no permite dimensionar y conocer el uso real de la voz de los pacientes. Para ello es necesaria una monitorización de los parámetros de la voz durante un largo periodo de tiempo.



Imagen 2: Dispositivo VHM².

OBJETIVOS

Prototipo de un dispositivo vestible que permita recoger la señal fonatoria y procesarla, y que sea capaz de monitorizar los principales parámetros fonatorios como la frecuencia, la intensidad relativa y el tiempo de fonación durante un tiempo prolongado.



Controlador EXP-MSP430FR5994



Acelerómetro BU-27135-000

MATERIAL Y MÉTODOS

Registro

Cálculo intensidad-relativa

FFT

Magnitud FFT

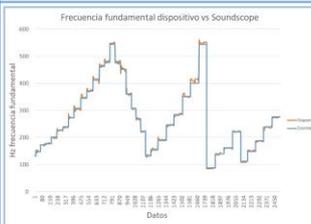
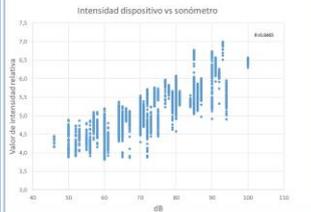
Filtrado

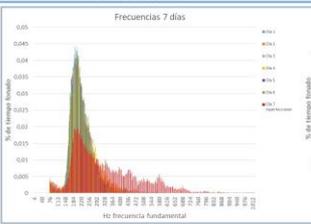
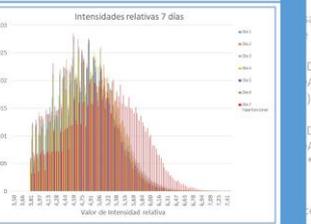
Correlación

Cálculo frecuencia fundamental

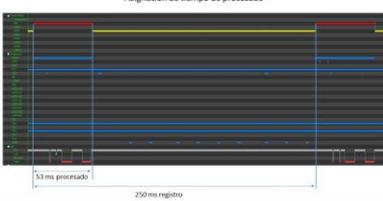
Almacenamiento de datos

RESULTADOS

Asignación de tiempo de procesado



Consumo energético

Name	Unit
System	600 ms
Energy	600x599 mJ
Power	
Mean	0,0116 mW
Min	0,0000 mW
Max	172,3723 mW
Voltage	
Mean	3,2873 V
Current	
Mean	2,4371 mA
Min	0,0000 mA
Max	32,4263 mA
Battery Life	26h34m:15.77s (est.)



Laboratorio de Ingeniería Médica
Facultad de Medicina

CONCLUSIONES

Se dispone del hardware para un dispositivo potencialmente vestible. Es capaz de medir la frecuencia fundamental e intensidad relativa. El cálculo puede realizarse a tiempo real.



Clínica Universidad de Navarra

BIBLIOGRAFÍA

1. Van Stan JH, Gustafsson J, Schalling E, Hillman RE. Direct Comparison of Three Commercially Available Devices for Voice Ambulatory Monitoring and Biofeedback. *Perspect Voice Disord*. 2014;24(2):80. doi:10.1044/vwd24.2.80.
2. Mehta DD, Lixfield RW, Cheyne II HA, et al. Duration of ambulatory monitoring needed to accurately estimate voice use. *13th Annu Conf Int Speech Commun Assoc 2012 (Interspeech 2012)*, Vols 1-3. 2012:1334-1337.

205

Póster presentado en la XI Jornada de Investigación en Ciencias Experimentales y de la Salud, Facultad de Ciencias de la Universidad de Navarra. Pamplona, 23 de marzo de 2018.

Diseño de un dispositivo vestible para la monitorización de la voz

Marcos Llorente¹, Secundino Fernández^{1,2}, Adam Podhorski³.

¹ Laboratorio de Ingeniería Médica, Facultad de Medicina, Universidad de Navarra.
² Departamento de Otorrinolaringología, Clínica Universidad de Navarra.
³ Departamento de Ingeniería Biomédica y Ciencias, TECNUN, Universidad de Navarra.

INTRODUCCIÓN

El análisis acústico y aerodinámico en el laboratorio de voz es una práctica habitual en el proceso de la valoración de los trastornos de la voz. Aporta información muy valiosa pero no permite dimensionar y conocer el uso real de la voz de los pacientes. Han sido varios los desarrollos para conseguir un sistema de monitorización.



Imagen 1: Comparación de dosímetros de voz¹.

Variables	APM	VoxLog	Vocalog	Voice-care
Intensidad (dB _{SPL})	X	X	X	X
Frecuencia a la Intensidad	X	X	X	
Frecuencia fundamental (Hz)	X	X		X
Frecuencia a la frecuencia	X	X		
Tiempo de fonación (silencios)	X	X	X	X
Ruido ambiente (dB _{SPL})			X	
Número de sílabas	X			
Distancia recorrida por las cuerdas (eventos)	X			
Calibración de Intensidad	Diala	No necesaria	Una vez	No necesaria

Tabla 1: Comparación de dosímetros de voz^{1,2}.



Imagen 2: Dispositivo VHM¹.

OBJETIVOS

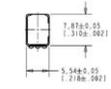
Prototipo de un dispositivo vestible que permita recoger la señal fonatoria y procesarla, y que sea capaz de monitorizar los principales parámetros fonatorios como la frecuencia, la intensidad relativa y el tiempo de fonación durante un tiempo prolongado.

MATERIAL Y MÉTODOS



Controlador Texas Instruments MSP430FR5994:

- Coprocesador para cálculo complejos (FFT y FIR).
- Registro de datos simultáneo al procesado.
- Bajo consumo.

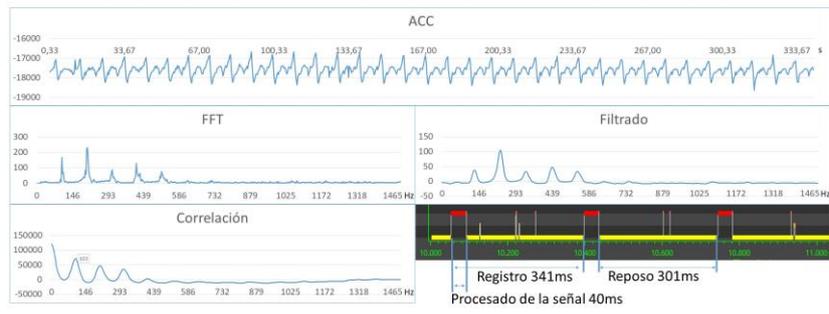


Acelerómetro BU-27135-000 Knowles Electronics:

- Acelerómetro de un solo eje.
- Respuesta lineal entre 20 Hz y 30 kHz.
- Pequeño tamaño (7,87x5,54x2,4 mm).



RESULTADOS





Laboratorio de Ingeniería Médica
Facultad de Medicina

CONCLUSIONES

Se dispone del hardware para un dispositivo potencialmente vestible.
 Es capaz de medir la frecuencia natural.
 El cálculo puede realizarse a tiempo real.



Clínica
Universidad
de Navarra

BIBLIOGRAFÍA

1. Van Stan JH, Gustafsson J, Schalling E, Hillman RE. Direct Comparison of Three Commercially Available Devices for Voice Ambulatory Monitoring and Biofeedback. *Perspect Voice Disord.* 2014;24(2):80. doi:10.1044/vvd24.2.80.
2. Carullo A, Vallan A, Astolfi A. A low-cost platform for voice monitoring. In: *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE; 2013:67-72. doi:10.1109/I2MTC.2013.6555383.
3. Mehta SD, Liffeld RW, Cheyne H HA, et al. Duration of ambulatory monitoring needed to accurately estimate voice use. *13th Annu Conf Int Speech Commun Assoc 2012 (Interspeech 2012)*. Vols 1-3. 2012:1334-1337.

206