### ANÁLISIS DEL SISTEMA DE FONACIÓN HUMANO

### CRISTIAN LUNA<sup>1</sup>, IGNACIO BEVACQUA<sup>1</sup> y NICOLÁS SALVAY<sup>1</sup>

<sup>1</sup>Estudiante de Ingeniería Electrónica. Universidad Tecnológica Nacional, Facultad Regional Córdoba (UTN. FRC). Maestro López esq. Cruz Roja Argentina, CP X5016ZAA, Córdoba, Argentina 51681@electronica.frc.utn.edu.ar, 51832@electronica.frc.utn.edu.ar, 52003@electronica.frc.utn.edu.ar

Resumen – Hoy en día para todo desarrollo acústico es fundamental un conocimiento amplio de la voz humana. Desde su emisión hasta cuales son las características que hacen que una persona pueda comprender el mensaje del habla sin dificultad alguna. Es por esta necesidad que el trabajo se centra en la investigación de la producción de la voz humana, las características que definen el mensaje oral y su comprensión, inteligibilidad y directividad.

### 1. INTRODUCCIÓN

El siguiente trabajo está referido al aparato de fonación en general. Tiene como objetivo analizar y comprender aspectos importantes de la anatomía como también de la voz humana como onda sonora transportadora de mensaje, el medio de trasmisión y características necesarias para transmitir un mensaje fidedigno.

Se realizó una investigación sobre la anatomía de la laringe y las cavidades resonantes del cuerpo humano y como su funcionamiento influye en el sonido proyectado. Se abordan patologías que afectan el sistema en cuestión, estudios que se pueden realizar y tratamientos. Se realizó un análisis de la voz como onda mecánica (análisis en tiempo y en frecuencia), su directividad y factores que influyen en la inteligibilidad del mensaje transmitido, como por ejemplo resonancia del medio de transmisión (ajeno al cuerpo humano).

### 2. ANATOMÍA DE LA LARINGE

La laringe es el órgano del aparato respiratorio que se encarga de la fonación, dentro de la misma encontramos las cuerdas vocales falsas (las superiores) y las cuerdas vocales verdaderas (las inferiores). Las cuerdas vocales verdaderas están formadas por el ligamento vocal y los músculos tensores intrínsecos.

Estas últimas se encargan de vibrar con el paso de aire a por la glotis. Esta vibración de las cuerdas es producida por la presión de aire emitida por los pulmones gracias a los músculos de la caja torácica. También se pueden encontrar en la laringe los músculos y nervios encargados de mover las cuerdas vocales para obtener distintos sonidos.

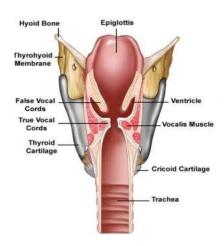


Figura 1: Corte frontal de la laringe [7].

El sonido emitido luego resonará dentro de la cavidad de la cabeza, comportándose ésta como filtro para el mismo, solo que al variar la postura se pueden variar sus características. Para que el aire no resuene tanto en la cavidad nasal el paladar blando bloquea parcialmente el paso del mismo.

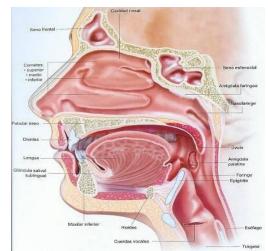


Figura 2: Anatomía de parte del sistema respiratorio [7].

Mediante el acomodamiento de la cavidad resonante se podría decir que la señal emitida por las cuerdas vocales es modulada para transportar información (el lenguaje).

Existen 2 grupos musculares ligados a la laringe, los extrínsecos, que se encargan del movimiento de la laringe en bloque (al tragar, al cantar, etc.) y los intrínsecos que se encargan del movimiento de las cuerdas vocales propiamente dichas. Estos dos grupos musculares son inervados por nervios, sensitivos y de control, que los conectan al sistema nervioso central.

#### 2.1 Emisión de la voz

Existen músculos cuya función es separar las cuerdas vocales (CV): dilatadores y otros que acercan las cuerdas vocales y cierran la glotis en la fonación: constrictores. Los que tienden las cuerdas vocales, según la altura de los sonidos, son los tensores. Los músculos fonatorios propiamente dichos son los constrictores y tensores de las cuerdas vocales.

La glotis se abre durante la respiración y se cierra durante la fonación.

En la emisión de los tonos agudos, se produce un aumento en la tensión de las CV y, en consecuencia, un adelgazamiento. En los graves, en cambio, las cuerdas disminuyen su tensión y se ensanchan. Las vocales exigen un movimiento de aducción lento y mantenido; las consonantes, en cambio, son emitidas sin mecanismo laríngeo en su mayoría y solo algunas necesitan un acercamiento de las CV brusco y de breve duración.

Existen distintos estadios dentro de la emisión de la voz, hablada, cantada o proyectada, el ataque, el sostén y el pasaje.

El ataque es el mecanismo glótico que precede la emisión del sonido laríngeo y se puede clasificar de la siguiente manera:

- Ataque soplado: el acercamiento de los labios vocales es lento y por lo tanto el sonido empieza cuanto la glotis no está todavía cerrada y deja escapar aire juntamente con él.
- *Ataque blando*: la pérdida de aire antes de la emisión sonora es mínima.
- Ataque duro: (golpe de glotis) la aducción es muy veloz, enérgica, y el contacto puede durar más tiempo, con tensión cordal excesiva. Puede llegar a producir trastornos vocales.
- Sostén: Se relaciona con la forma en la que se administra el aire por acción del diafragma y encontramos 2 formas totalmente distintas de sostén:
- *Empuje:* se realiza empujando hacia adentro la musculatura abdominal podemos utilizarlo para terminar con una emisión.
- Apoyo: requiere de la apertura de costillas y sostén con diafragma y la sensación muscular es exactamente contraria a la del empuje: hay una contracción hacia fuera. La utilizamos en vos hablada y cantada.

Pasaje es el punto crítico que se observa en cualquier sujeto cuando comienza a ascender en una escala sonora y llega al extremo de la altura. Su voz se quiebra, la laringe tira y se esfuerza. Solamente podrá seguir ascendiendo si acomoda sus cavidades resonanciales. El hombre después del pasaje obtendrá un timbre más oscuro seguirá cantando con voz de pecho, "cubre" y permite de esa forma que la laringe este baja para los sonidos agudos.

### 2.2 Parámetros de la voz

Algunos de los parámetros más importantes que definen la voz humana, así como también a cualquier sonido, son: El tono, la intensidad y el timbre. Llevado a un lenguaje más ingenieril, estos serian la frecuencia, la amplitud y la riqueza en armónicos de la onda.

La frecuencia se refiere a la cantidad de vibraciones por segundo de las cuerdas vocales y esta se puede regular mediante los músculos tensores. La frecuencia típica de emisión para el hombre es de entre 130 y 200Hz mientras que para la mujer la frecuencia normal es entre 200 y 250Hz.

La intensidad se puede regular con la presión subglótica de aire.

Los armónicos contenidos dependen de la caja de resonancia, es decir, de las dimensiones del tracto vocal y sus variaciones mediante el distinto acomodamiento del individuo.

### 3. PATOLOGÍAS

Como cualquier parte del cuerpo humano, la laringe puede sufrir una gran variedad de patologías que afecten el habla o la respiración. Aquellas que nos incumben son las disfonías. Las disfonías son cualquier factor que las perturbe el timbre, el tono o la intensidad de la voz.

### 3.1 Clasificación de las disfonías

Según grado de disfonía se puede reconocer a la disfonía como a la pérdida parcial de algún atributo y la afonía como la pérdida total de la voz.

Según el tipo de voz que se ve afectada puede clasificarse como: Disfonía del habla, disfonía de voz proyectada (Requiere un Juego pósturo-muscular y un acomodamiento que puede fallar) o disfonía de voz cantada (proviene generalmente de un elemento funcional) [1].

Otra clasificación posible es según la naturaleza del caso:

- Disfonías funcionales: Defectuosa utilización de los mecanismos que permiten la función vocal. La morfología laríngea aparentemente normal. Algunos desencadenantes: infecciones repetidas; hábitos tóxicos; labilidad psicológica; alteración del tonismo; alteraciones posturales; tipo respiratorio defectuoso; exceso de tensión muscular en zona laríngea; exigencias vocales desmesuradas.
- Disfonías por lesiones adquiridas: alteración de la conformación normal de los repliegues vocales;

puede ser ocasionadas por el abuso o sobreesfuerzo vocal. Es normal desarrollar una disfonía por irritación de las cuerdas vocales cuando se exponen al aire frio, generalmente esto se presenta durante un cuadro gripal debido a la probable infección producida en la laringe sino también. Esto puede suceder porque la congestión de la cavidad nasal nos impide respirar correctamente y nos obliga a respirar por la boca. La cavidad nasal contiene mucha irrigación sanguínea por lo que el aire que pasa a través de ella, aumenta de temperatura, por el contrario si se respira por la boca la temperatura del aire al llegar a la garganta es muy baja, esto puede alterar el PH de la zona convirtiéndola en un sitio propenso al desarrollo de bacterias o virus.

Algunos otros desencadenantes de disfonías por lesiones son: de base hormonal; por esfuerzo vocales intensos; esfuerzos a glotis cerrada (pólipos); procesos inflamatorios de la mucosa cordal; traumatismos post-intubación; hábito tabaquito y alcohólico; terreno alérgico predisponente.

• Disfonías disfuncionales: En éstas, existe una alteración de la función vocal, mantenida fundamentalmente por un trastorno del acto vocal. En la aparición de una disfonía disfuncional intervienen tres conceptos clave: el sobreesfuerzo vocal mantenido (Voz de apremio), factores desencadenantes y factores favorecedores. La Tabla 1 explica algunos de estos conceptos:

sobresfuerzo vocal	actores Factores cadenantes favorecedores
perdida de la verticalidad, crispaciones.      Acometida de golpe de glotis.      Aparición de sensaciones subjetivas especiales: perdida de eficacia, taliga vocal, dolor, tensión laringea, crispación a laringea.      Derdida de eficacia, como de laringea, crispación a como de laringea, crispación a como de laringea.      Cura como de la vertical paricipación a como de laringea, como de laringea.	stológicos. es  pelcos. amiento al.  o pre-mens- azo, interven- bdominal. le silencio por ión vocal de- to

Cuando sintesis de la concepción multifactorial y teoría de la voz de apremio de Le Huche (1994), tomo 2.

Tabla 1: Cuadro de los 3 conceptos intervinientes en disfonías disfuncionales.

La siguiente imagen ilustra algunas afecciones típicas de las cuerdas vocales:



Figura 3: Afecciones más comunes de las cuerdas vocales.

# 3.2 Lesiones y afecciones directas de la voz cantada

### 3.2.1 Dilatación venosa circunscriptas

Puede provocar hemorragia submucosas, el reposo vocal se prolonga y el uso de flavoides de epitelio protector, de coagulante.

# 3.2.2 Gotas de mucosidad sobre las cuerdas vocales en el momento de oscilación

Es una secreción que viene del ventrículo y se distribuye en múltiples gotitas simétricas que cuelgan de los labios vocales como cuentas de un collar de cristal.

### 3.2.3 Edemas circunscriptos de cuerda vocales

En individuos alérgicos o artríticos o muy nerviosos que hagan abusos vocales estos edemas son frecuentes, en un primer tiempo fugaz y suelen repetirse y persistir cada vez más. La corrección de suele ser más eficaz con reposo.

# 3.2.4 Debilidad constitucional de la laringe o de la secuela de lesiones cordales pasadas

Se produce en alumnos que recién comienzan canto. El tratamiento es aleatorio más bien podrían prevenirse si se hicieran un sistemático examen laringofoniatrico y se podría evitar.

# 3.2.5 Errores en la técnica vocal o en la enseñanza

Entre las causas que podríamos llamar pedagógicas señalamos:

- El temblor vocal
- La voz intubada
- La voz nasal
- El ataque débil
- El ataque demasiado fuerte
- La voz engolada
- La mala respiración

### 4. TRATAMIENTOS O ESTUDIOS

Un estudio muy común que se puede hacer a las cuerdas vocales es la laringoestroboscopía. El estudio

consiste en medir la frecuencia de vibración de las cuerdas vocales e iluminarlas con un flash a la misma frecuencia o a una cercana, para poder observarlas moviéndose lentamente. De esa forma es más fácil detectar problemas durante su vibración o afecciones.

En caso de encontrarse cualquier tipo de patología se puede proceder a un tratamiento no invasivo, pero en caso de necesitarse se puede intervenir quirúrgicamente. En el caso de los pólipos, por ejemplo, se procede a extirparlos y luego se debe ir a terapia a cargo de un profesional en fonoaudiología para realizar lo que se llama reeducación en la técnica vocal. A diferencia del caso anterior, por ejemplo, cuando se está en presencia de nódulos se toma a la intervención quirúrgica como última instancia viable del tratamiento.

### 5. CARACTERÍSTICAS Y PERCEPCIÓN DE LA VOZ HABLADA

Cuando una persona emite un mensaje, emplea un tiempo mayor en la emisión de las vocales que en la de las consonantes. Es por ello que las vocales constituyen el llamado *régimen permanente* del habla, mientras que las consonantes se asocian al *régimen transitorio* [6].

La duración en promedio de una vocal es del orden de 90 ms, reduciéndose a 20 ms en el caso de una consonante.

El hecho de que la duración de las vocales sea más elevada hace que el nivel de presión sonora asociado a las mismas sea, en promedio, del orden de 12 dB mayor que el correspondiente a las consonantes. Por otra parte, su contenido frecuencial es más rico en bajas frecuencias, mientras que las consonantes presentan una mayor contribución de altas frecuencias.

Por otro lado, el grado de inteligibilidad de la palabra está estrechamente relacionado con la correcta percepción de las altas frecuencias. En consecuencia, son las consonantes las que determinan la comprensión del mensaje oral. En cambio, la información contenida en las vocales es redundante. En la Tabla 2 se resume las características mencionadas.

	Duración Promedio	Contenido Frecuencial Dominante	Nivel Promedio	Contribución a la Inteligi- bilidad de la palabra
Vocal	≈ 90 ms	Bajas frec.	Nivel Vocales ≈ Nivel Con- sonantes + 12 dB	Baja
Consonante	≈ 20 ms	Altas frec.		alta

Tabla 2: Características más relevantes del mensaje oral [6]

### 5.1 Percepción del lenguaje

La percepción del lenguaje se ocupa de los procesos de extracción de información de la señal

acústica o, alternativamente, gráfica. Por tanto, se ocupa tanto de los procesos iniciales de análisis de la señal, como de los procesos más complejos donde el análisis es sintáctico/semántico. No obstante, percepción del lenguaje se refiere, habitualmente, a los procesos iniciales incluido el reconocimiento de la palabra y se reserva el término comprensión del lenguaje para los procesos más complejos [5].

Características de la percepción del lenguaje hablado:

- Mayor demanda de memoria en lo hablado que lo escrito.
- Ausencia de los límites de la palabra en la mayoría de sistemas hablados.
- El lenguaje hablado es un proceso contextualmente conducido: el mensaje se interpreta a partir de conocimiento y marcos de referencia participados.

### 5.1.1 Percepción de Vocales

Una vocal es percibida cuando es identificada como tal sonido dentro de una lengua. Se forma cuando el aire fluye con relativa libertad por el tracto vocal sufriendo modificaciones mínimas. En su definición acústica intervienen:

- La frecuencia de los formantes
- El contexto consonántico
- La duración

Frecuencia de los Formantes: Las vocales aisladas pueden clasificarse en base a la localización de sus formantes F1 y F2. F3 permanece relativamente estable a lo largo de todas las vocales.

El oyente utiliza la información de los valores relativos de F1 y F2, usa la frecuencia fundamental F0 como referencia para evaluar la localización relativa de los formantes y los valores de F1 y F2 en el contexto precedente.

El contexto consonántico: Afecta claramente a los valores de los formantes. La percepción de las vocales coarticuladas, se facilita en contexto consonántico mejor que aisladamente. El efecto del contexto es más relevante para el reconocimiento de las vocales que la familiaridad con las características del tracto vocal del hablante. La estructura silábica es una variable importante. Las sílabas CVC y VC son más facilitadoras que las CV.

Duración: En el habla natural, una vocal suele durar entre 100 y 300 ms. El contexto fonético y el patrón de entonación de la frase, la hace variar de forma considerable. Ello, dificulta valorar su importancia. La duración es particularmente importante al distinguir vocales con frecuencias de los formantes similares. En este proceso el oyente utiliza como fuente de información la duración relativa más que la absoluta. La duración suele ser redundante pero necesaria para el procesamiento del habla.

### 5.1.2 Percepción de Consonantes

Las consonantes se emiten con obstrucción total o parcial del tracto oral. Los rasgos que adopta el tracto oral determinan las características acústicas que hacen posible la identificación perceptiva de los fonemas. La distribución de energía en el espectro y su amplitud, la forma de las transiciones o conexiones de la energía entre consonantes y vocales, y la duración son los rasgos acústicos dominantes en la representación espectrográfica de las consonantes.

Perceptivamente, las principales características acústicas relacionadas con la identificación de las oclusivas son el VOT o intervalo de silencio hasta el comienzo de la vocal y la transición del segundo formante hacia la vocal siguiente.

Por su parte, las *fricativas* se definen, principalmente, por la presencia de ruido aperiódico en el espectro (Heinz y Stevens, 1961), su baja intensidad y la concentración de la energía entre los 3000 y los 5000 Hz.

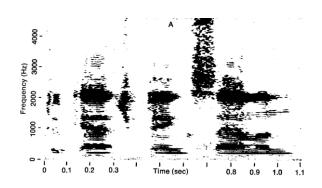
Las líquidas y nasales se producen por vibración de las cuerdas vocales con el tracto vocal cerrado en algún punto.

En las nasales, algo de aire pasa por la cavidad nasal, ello se ve acústicamente en la presencia de energía en las frecuencias bajas del espectro. F1, que domina el espectro de la señal, se sitúa en torno a los 250 Hz, F2 es muy débil y F3 se sitúa cerca de los 2200 Hz.

Las líquidas se producen por constricción parcial del tracto vocal, sus espectros son muy semejantes a los de las vocales, pero de menor intensidad, con la energía concentrada en bandas estables y duraderas.

### 5.2 Análisis Espectográfico

Mediante un análisis espectrográfico es posible extraer las características acústicas contenidas en una señal y presentarlas en un espectrograma respecto a las dimensiones de frecuencia, intensidad y tiempo. La Figura 4 muestra un ejemplo de espectograma para la frase hablada "to catch pinks salmon" [5].



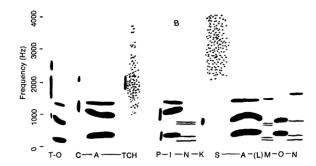
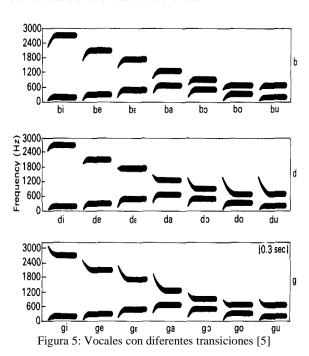


Figura 4: Sound Spectogram – Simplified version of the same phrase that serves as input to pattern playback (After Cooper et, al., 1952).

La Figura 5 muestra el espectograma de vocales sometidas a diferentes transiciones.



En la Figura 6 podemos apreciar con mayor precisión como para un mismo fonema /d/ se observan diferentes rangos.

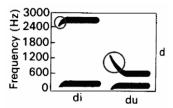


Figura 6: Rangos distintos y un mismo fonema /d/

### 6. DIRECTIVIDAD DE LA VOZ HUMANA

La directividad de la fuente de emisión sonora de la voz está dada en función de la frecuencia y por lo general aumenta cuando aumenta la misma, aunque la principal causa de la forma de directividad de la voz humana viene dado por el sistema de fonación y la forma de cabeza, siendo la dirección frontal la de mayor directividad, es por ellos que en la dirección. Para poder expresar en manera estándar se utiliza el factor de directividad Q; el mismo está definido por la relación de nivel de presión sonora de la fuente en consideración con una fuente la cual no es directiva [6]. Como ya se ha dicho anteriormente en la dirección frontal se encuentra la mayor directividad siendo esta de un valor de Q=2. En la Figura 7 puede verse dicha directividad.

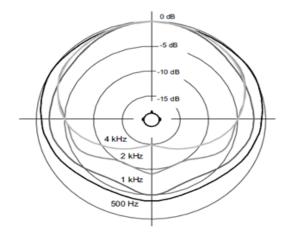


Figura 7: Directividad de la voz humana en las bandas de octavas comprendidas entre 500 Hz y 4 kHz [6]

### 7. INTELIGIBILIDAD DE LA PALABRA

La clasificación más marcada en cuanto a los diferentes sonidos que se pueden formar con el sistema fonador es sin duda la diferencia entre vocales y consonantes. Las vocales conforman lo que se conoce como sonido sonoro, ya que están generadas por vibraciones en la cuerdas vocales, en cambio la mayoría de las consonantes están conformadas por sonidos sordos, llamados así a los sonido que no se dan mediante la vibración de las cuerdas vocales sino a través de una constricción del flujo de aire que atraviesa la cavidad bucal.

Otra diferenciación entre las misma es el tiempo promedio que conlleva producir una vocal o una consonante; mientras que la primera tiene una duración aproximadamente de 90 ms en el caso de las consonantes solo se utiliza 20 ms. Esto conlleva a que el nivel de presión asociado a las vocales sea 12 dB por encima de las consonantes. Así mismo también se pueden diferenciar estos dos tipos de sonidos gracias a su contenido espectral: mientras que las vocales conllevan un alto contenido de frecuencias bajas, las consonantes presentan una mayor contribución a altas frecuencias.

A pesar de que las vocales tienen un mayor protagonismo en el mensaje oral (por su duración y nivel), el grado de inteligibilidad está dado por la percepción a altas frecuencias., por lo que son las consonantes las que determinan la compresión del mensaje que se quiere transmitir.

En la Figura 8 se puede ver la contribución de cada banda de frecuencias de octava al nivel de la voz y a la inteligibilidad de la voz. Aquí se contrasta sus diferencias, mientras que la mayor contribución a la voz de da en la zona de frecuencias medias (aprox.

500 Hz), el mayor aporte a la inteligibilidad se da a frecuencia altas entre 2 y 4 kHz.

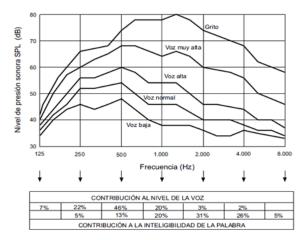


Figura 8: Contribución frecuencial al nivel de la voz y a la inteligibilidad de la palabra [6]

### 8. CONCLUSIONES

La voz humana es un sistema muy complejo, que requiere de un estudio profundo, para poder así llevar a cabo cualquier aplicación o diseño en la cual ésta se vea implicada, desde el modelado de una sala de teatro o grabación, la creación de micrófonos o altavoces, el aprendizaje del canto, la compresión de archivos, etc..

Este trabajo aporta los conocimientos necesarios sobre la anatomía humana para la conformación de la voz, describiendo a su vez las patologías y tratamientos que pueden llegarse a presentar en diversos casos y ante determinadas circunstancias.

Se puede observar como contribuyen las consonantes y las vocales con sus respectivas características intrínsecas a cada una a la inteligibilidad de la palabra y cómo a su vez la directividad es afectada por el sistema de fonación y la forma de la cabeza.

### 9. REFERENCIAS

- [1] Diamante, V, "Otorrinolaringología y afecciones conexas", Promedicina Reimpresión 2ª edición. Buenos Aires 1994
- [2] Garde, E "La voz". Lautaro. Buenos Aires 1958 [3] Jackson Menaldi, C, "Patología vocal en profesionales de la voz". Revista Fonaudiológica. 3:60-66. Buenos Aires 1994
- [4] Alexader, G. "La Eutonía". Paidós. 19 Reimpresión. Buenos Aires 1986.
- [5] Apunte Percepción del lenguaje. España
- [6] Antoni Carrión Isbert: "Diseño acústico de espacios arquitectónicos" Primera edición: Julio de 1998.
- [7] INCIAMEDIS, Instituto de Ciencias Ancestrales y Medicina Suprema, http://www.inciamedis.com.
  [8] Cuerdas Vocales,

http://borja3.glogster.com/cuerdas-vocales/

### 10. DATOS BIOGRÁFICOS

Cristian Luna, nacido en Ushuaia el 15/11/1987. Estudiante de ingeniería en electrónica, Universidad Tecnológica Nacional, Facultad Regional Córdoba. Argentina. Actualmente se encuentra cursando 6to año de la correspondiente carrera. E-mail: 51681@electrónica.frc.utn.edu.ar

Ignacio Bevacqua, nacido en la ciudad de Merlo, provincia de Bs. As., el 14/08/1987. Tres años después su familia se mudó a Bariloche donde curso la escuela primaria y secundaria. Se mudó a Córdoba para ingresar en la UTN en el 2006. Estudiante de ingeniería en electrónica, Universidad Tecnológica Nacional, Facultad Regional Córdoba. Argentina. Actualmente se encuentra cursando 6<sup>to</sup> y preparando su tesis. E-mail: 51832@electrónica.frc.utn.edu.ar

**Nicolás Salvay**, nacido en Córdoba el 02/10/1987. Terminó sus estudios secundarios en el Colegio Parroquial San Francisco de Asís. Estudiante de ingeniería en electrónica, Universidad Tecnológica Nacional, Facultad Regional Córdoba. Argentina. Actualmente se encuentra cursando 6<sup>to</sup> año de la carrera. E-mail: 52003@electrónica.frc.utn.edu.ar