

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



TRABAJO FIN DE GRADO ANÁLISIS ACÚSTICO DE TÉCNICAS DE REEDUCACIÓN DE LA VOZ

Titulación: Grado en Ingeniería en Tecnologías de la Telecomunicación

Autor: D. Francisco Javier Valerón Peñate

Tutores: Dr. D. Jesús Bernardino Alonso Hernández

Dr. D. Carlos Manuel Travieso González

Fecha: Diciembre de 2014

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



TRABAJO FIN DE GRADO

ANÁLISIS ACÚSTICO DE TÉCNICAS DE REEDUCACIÓN DE LA VOZ

HOJA DE FIRMAS

Alumno:

Fdo.: Francisco Javier Valerón Peñate

Tutor:

Fdo.: Dr. D. Jesús Bernardino
Alonso Hernández

Tutor:

Fdo.: Dr. D. Carlos Manuel
Travieso González

Fecha: Diciembre de 2014

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



TRABAJO FIN DE GRADO

ANÁLISIS ACÚSTICO DE TÉCNICAS DE REEDUCACIÓN DE LA VOZ

HOJA DE EVALUACIÓN

Calificación: _____

Presidente:

Vocal:

Secretario:

Fecha: Diciembre de 2014

Agradecimientos

Quisiera agradecer a mis tutores de este Trabajo Fin de Título, al Dr. D. Jesús Bernardino Alonso Hernández y al Dr. D. Carlos Manuel Travieso González por su ayuda, colaboración y apoyo recibido, gracias a lo cual, la realización de este trabajo ha sido mucho más grata y fructífera.

Agradezco a mi familia por todo su apoyo, sus ánimos y su estímulo, haciendo que la elaboración de este Trabajo Fin de Título haya sido más amena.

Y por último, agradecer a todos mis amigos y compañeros de la universidad por sus ánimos durante el transcurso de la realización de este Trabajo Fin de Título.

ÍNDICE

I.- MEMORIA.....	13
Capítulo 1.- Introducción.....	15
1.1.- Motivación.....	15
1.2.- Objetivo del trabajo	15
1.3.- Estructura del trabajo	16
Capítulo 2.- La Voz. El Sistema Vocal. Cualidades de la Voz.....	19
2.1.- La voz hablada.....	19
2.2.- La voz cantada.....	20
2.3.- Alteraciones del habla y del lenguaje	21
2.4.- Alteraciones de la voz.....	21
2.5.- La reeducación vocal.....	22
2.6.- El aparato vocal.....	24
2.6.1.- Los fuelles.....	25
2.6.2.- El vibrador	25
2.6.3.- Los resonadores	25
2.6.4.- Fonación	26
2.7.- Las cualidades de la voz.....	27
2.7.1.- Intensidad.....	27
2.7.2.- Altura	27
2.7.3.- Timbre	28
2.8.- Incidencia del funcionamiento laríngeo en las cualidades de la voz.....	28
2.8.1.- Elongación.....	31
2.8.2.- Grosor.....	32
2.8.3.- Masa muscular vibrante	34
2.8.4.- Presión de cierre	34

Capítulo 3.- Representación de la Señal de Voz.....	37
3.1.- La calidad de la voz. Fenómenos implicados.....	37
3.1.1.- La calidad de la voz desde diferentes dominios.....	38
3.1.2.- Dominio temporal	39
3.1.3.- Dominio espectral.....	41
3.1.4.- Dominio cepstral	46
3.1.5.- Fenómenos implicados en la calidad de la voz	50
3.2.- Cuantificación de la estabilidad de la voz.....	51
3.2.1.- Dominio temporal	51
3.2.1.1.- Perturbación de la amplitud: <i>shimmer</i>	51
3.2.2.- Dominio espectral.....	53
3.3.- Cuantificación de la riqueza espectral.....	53
3.3.1.- Dominio espectral.....	54
3.3.1.1.- Perturbación del pitch: <i>jitter</i>	54
3.3.1.2.- Grado de insonoridad	55
3.3.1.3.- Grado de periodicidad de las tramas sordas	56
3.3.1.4.- Relación subarmónico a armónico	56
3.3.1.5.- Grado de sonoridad por trama	57
3.3.2.- Dominio cepstral	57
3.3.2.1.- Medidas basadas en el cepstrum.....	57
3.3.2.1.1.- Energía cepstral.....	57
3.3.2.1.2.- Energía cepstral del pitch.....	58
3.3.2.1.3.- Energía cepstral del segundo armónico	60
3.3.2.1.4.- Cepstrum derivado espacialmente.....	61
3.3.2.1.5.- Relación aperiódico-periódico.....	62
Capítulo 4.- Metodología	63
4.1.- Estudio. Propuesta metodológica	63
4.2.- Técnica de voz con el método “Pantalla de Celofán”	64
4.2.1.- Fase previa: entrenamiento.....	66
4.2.2.- Fase final: la percepción	67
4.3.- Procedimiento de grabación y segmentación de los archivos de audio	67
4.4.- Análisis de la señal de voz	68

Capítulo 5.- Base de datos	69
5.1.- Descripción general de la base de datos	69
5.2.- Balance de muestras de voz de la base de datos	70
5.3.- Equipo utilizado en las grabaciones de los archivos de audio	71
5.4.- Captura de las señales de audio	73
Capítulo 6.- Análisis de los resultados	79
6.1.- Análisis cualitativo. Gráficas	79
6.2.- Análisis cuantitativo. Medidas	110
Capítulo 7.- Conclusiones	119
Capítulo 8.- Bibliografía	121
II.- PRESUPUESTO	123
II.1.- Costes de recursos humanos	125
II.2.- Costes de ingeniería	126
II.3.- Costes de amortización	127
II.3.1.- Costes de recursos hardware	127
II.3.2.- Costes de recursos software	127
II.4.- Otros costes	128
II.5.- Presupuesto total	128
II.6.- Herramientas hardware	129
II.7.- Paquetes software	130
III.- ANEXO: Gráficas del Análisis Cualitativo	135
GLOSARIO: Términos y definiciones	199

I.- MEMORIA



I.- MEMORIA

Capítulo 1

Introducción

1.1.- Motivación

En la sociedad actual las patologías de la voz tienen una gran incidencia, como consecuencia de los esfuerzos continuados por motivos profesionales (cantante, educador, orador, vendedor, animador, etc.) ya sean distonías, nódulos, afonías u otros, o simplemente que su emisión de voz no es la más adecuada, produciendo cansancio muscular en las cuerdas vocales y el consiguiente sobreesfuerzo vocal.

La voz es la consecuencia de una compleja mecánica en la que cada elemento posee una función e intervienen una multitud de fenómenos que interaccionan en la laringe. Es bien sabido que aprender a controlar y verificar todos los movimientos que hacen falta para producir un buen sonido es bastante complicado, porque el cantante percibe unas sensaciones internas, que no son siempre iguales y que varían dependiendo de la técnica vocal que se emplee.

La motivación de este trabajo es la reeducación de vocal y cuya comprobación se realizará a través de un análisis acústico para corroborar las mejoras en la producción vocal obtenidas tras practicar el método “Pantalla de celofán”; el cuál se detallará posteriormente en este trabajo.

1.2.- Objetivo del trabajo

El objetivo de este Trabajo Fin de Título es estudiar las mejoras en la producción vocal obtenidas tras practicar el método “Pantalla de celofán” para la reeducación vocal, a través de un análisis acústico pertinente de dos vocales fuertes o abiertas y una vocal cerrada o débil grabadas por cada locutor participante *antes* y *después* de aplicar el método.

Para cada caso de estudio, se dispondrá de un repositorio de muestras (archivos de audio), se analizará una grabación de varias vocales (/a/, /i/ y /o/) de forma sostenida tanto antes como después de aplicar el método “Pantalla de celofán” por cada locutor o participante, siendo las vocales /a/ y /o/ dos vocales fuertes o abiertas, y la vocal /i/ una vocal débil o cerrada. Para la grabación de cada una de las vocales sostenidas antes y después de aplicar el método citado anteriormente se utilizará el software de libre distribución *Audacity*.

El análisis acústico consistirá en la representación de la señal en el dominio temporal, en dominio frecuencial por medio del espectrograma de banda estrecha, la excitación glótica y una representación en el dominio cepstral. Para obtener las representaciones se utilizará el software de libre distribución *ESection*.

1.3.- Estructura del trabajo

El presente Trabajo Fin de Título está compuesto por los siguientes títulos o secciones:

- **I.** Memoria
- **II.** Presupuesto
- **III.** Anexo

En el **Título Primero (I)**, correspondiente a la **Memoria**, se incluye un capítulo introductorio sobre el proyecto, donde se describe la motivación y los objetivos del mismo. Además, se incluye el resto de capítulos; donde se recogen la descripción de las tareas realizadas, resultados obtenidos, conclusiones y bibliografía referenciada. Los capítulos adjuntos en la Memoria son los descritos a continuación:

- El **capítulo 1** de esta Memoria, corresponde con la parte introductoria de este Trabajo Fin de Título; dónde se expone cuál es la motivación del trabajo y cuáles son los objetivos y la estructura del mismo.
- En el **capítulo 2**, se explica los aspectos más significativos que conciernen tanto a la voz hablada como a la voz cantada, y se describen las alteraciones del habla y del lenguaje, así como las alteraciones que se suelen producir en la voz. Además se describe los diferentes métodos existentes para la reeducación de la voz y se realiza una descripción sobre el aparato vocal y las cualidades de la voz y como inciden estas cualidades en el funcionamiento del sistema fonatorio.

- En el **capítulo 3** se detalla cómo pueden representarse las señales de voz correspondientes en diferentes dominios (temporal, espectral, cepstral, etc.). Así mismo, se describen los fenómenos implicados en la calidad de la voz
- A continuación, en el **capítulo 4** se detalla la propuesta metodológica que se ha utilizado (“Pantalla de celofán”) para la reeducación vocal y se realizará una explicación de los pasos que se siguieron en las grabaciones realizadas, y de cómo fueron procesados estos archivos de audio.
- En el **capítulo 5** se adjunta la base de datos, dónde se ilustra diversas capturas correspondientes a las señales de audio correspondientes para cada una de las diferentes vocales a analizar. Además se describen las características que se tuvieron en cuenta a la hora de realizar las grabaciones.
- En el **capítulo 6** se analiza los resultados obtenidos tras la mejora después de aplicar el método correspondiente; tanto cualitativamente (análisis de las gráficas obtenidas en su dominio temporal, espectral y cepstral), como cuantitativamente (análisis de las medidas obtenidas de diferentes parámetros de medida de calidad de la voz).
- En el **capítulo 7** se presentan las conclusiones obtenidas del estudio tras el análisis acústico de las vocales grabadas; *antes* y *después* de aplicarse el método “Pantalla de celofán”.
- Por último, en el **capítulo 8** de esta parte 1 correspondiente a la Memoria del Trabajo Fin de Título, se detalla la bibliografía utilizada para la realización de este proyecto, indicándose los libros y fuentes de Internet consultadas.

El **Título Segundo (II)**, corresponde al **Presupuesto**, se describe el precio total del sistema de grabación y análisis de las señales de voz, que corresponde a la suma del presupuesto del equipo utilizado para la grabación de los archivos de audio y el presupuesto del coste de los paquetes software utilizados para el editado, procesado y análisis de las señales de voz. Así mismo, se recogen las características de las herramientas hardware y de los paquetes softwares utilizados.

El **Título Tercero (III)**, corresponde al **Anexo**, donde se ilustran las gráficas obtenidas para cada una de las vocales (/a/, /i/ y /o/) *antes* y *después* de aplicar el método “Pantalla de celofán”, correspondiente al análisis cualitativo.

Para finalizar, se incluye un **Glosario** con los términos más singulares que aparecen en este trabajo, así como sus respectivas definiciones.

Capítulo 2

La Voz. El Sistema Vocal. Cualidades de la Voz

2.1.- La voz hablada

El primer uso que hacemos de nuestro aparato fonador es al nacer “el llanto” que se caracteriza por altas frecuencias, el ataque del sonido es brusco, de fuerte intensidad y modulación reducida. La extensión va desde 787 a 1.318 Hz. A partir de ahí las primeras referencias auditivas que tenemos son las de nuestro entorno familiar, ésta es la llamada “etapa pre-lingüística” a partir de la que se producirá un proceso escalonado e interrumpido, que se divide por etapas a efectos metodológicos, después de las cuáles se pasará al desarrollo del lenguaje oral.

Hablar se considera un acto tan natural como puede ser cualquier otro, caminar, comer, etc. pero nadie nos prepara para hablar de forma correcta, entendiendo por ello un correcto funcionamiento del aparato vocal. Aprenderemos a hablar y utilizamos nuestra voz en todos los momentos de nuestra vida pero sin saber y sin tener una técnica apropiada que nos facilite el buen uso de ella.

A griegos, y romanos admiradores de la belleza en todas sus formas, “el arte del bien decir” constituía una parte importante de su formación y para ello contaban con grandes oradores y maestros que cultivaban el arte de la oratoria.

Siguiendo la línea que nos legó el célebre orador y filósofo Demóstenes, en la actualidad las técnicas que se usan para mejorar la emisión vocal son muchas y variadas: colocarse un tapón de corcho entre los dientes, *Lucien Guitry*, se servía de un lápiz entre los dientes, *Sarah Bernhardt* recitaba versos empleando un lapicero, etc.

El estudio del mecanismo vocal, la respiración, la emisión, la colocación de la voz, es primordial no solo para cualquier persona sino aún más para los que utilizan su voz en público, pero bien es cierto que el estudio de la voz hablada no parece indispensable. Se podría decir que el estudio de la voz hablada es algo inexistente, salvo para aquellas personas que deciden asistir a profesionales de la voz por exigencias de su profesión como actores u oficios que requieren hablar en público.

Otra cuestión bien distinta es cuando se trata de voces patológicas. El papel que desempeñan médicos, logopedas, foniatras ha sido y es de vital importancia para la rehabilitación y reeducación de la voz.

En la actualidad nos encontramos en que una gran parte de la población presenta problemas de tipo vocal ya sea disfonías, nódulos, afonías u otros, o simplemente que su emisión de voz no es la más adecuada, produciendo cansancio y el consiguiente forzamiento vocal.

El aprendizaje de la voz hablada no difiere tanto de la técnica empleada en el canto, por cuanto la respiración constituye la base fundamental en ambas, además del control muscular, relajación, de técnica de soplo de verticalidad, dicción y articulación.

La producción sonora tanto en la voz hablada como cantada se produce con los mismos órganos pero las características o las cualidades del sonido: la altura, duración, intensidad y timbre son diferentes, así como el control que se ejerce sobre los distintos órganos que intervienen en la fonación. Aunque el tono y la intensidad del habla están determinados principalmente por la vibración de las cuerdas vocales, su espectro está fuertemente determinado por las resonancias del tracto vocal.

La inteligibilidad oral se debe a las altas frecuencias. Para que el habla sea comprensible, es indispensable la presencia de armónicos cuya frecuencia se halla entre 500 y 3.500 Hz [11].

2.2.- La voz cantada

En la voz cantada, la producción sonora se produce con los mismos órganos pero las características o cualidades del sonido son diferentes pero no porque se produzcan de manera natural sino que es un entrenamiento de un esquema corporal-vocal diferente del que utilizamos para el habla.

Todo el mundo puede cantar puesto que contamos con las mismas capacidades fisiológicas, depende más de cómo manejas y controlas el mecanismo vocal que de las cualidades o aptitudes que en principio se pueda tensar. Se aprende a cantar con el oído, pero es el cerebro el que gobierna la frecuencia de las vibraciones y controla la emisión de la voz.

La extensión normal de una voz cantada es de dos octavas. La voz varonil va del Do2 al Do4 y la femenina del Do3 al Do5 porque la mujer canta en una octava más elevada que el hombre.

La clasificación de la voz tiene enorme importancia pues cuando hay errores muy evidentes se pueden producir desviaciones que conllevan trastornos vocales, nódulos, etc. Ciertamente es que cuando empieza a trabajar una voz, no siempre está clara una clasificación ni tampoco hay que darle excesiva importancia siempre y cuando se esté trabajando en una tesitura cómoda y que no resulte perjudicial al alumno. El profesor de canto es el que debe descubrir junto al alumno la tesitura real y el verdadero timbre de esa voz artificial [12].

2.3. Alteraciones del habla y el lenguaje

Las causas que pueden producir una alteración del habla y el lenguaje son muy variadas: neurológicas, genéticas, congénitas, traumáticas, iatrogénicas, tumorales, metabólicas, sensoriales, etc. Se puede afectar la comunicación de la palabra de distintos modos [13]:

- Sí se altera la voz sólo, hablaremos de *disfonía*.
- Sí se altera la articulación: *disartria*.
- Sí se altera el lenguaje: *afasia*.
- Sí se altera la fluencia del lenguaje: *disfemia*.
- Sí se altera la comunicación debido a la audición: sordomudez o *presbiacusia*.

2.4. Alteraciones de la voz

Existen distintos tipos de vista para clasificar las alteraciones de la voz, nosotros utilizaremos la que tiene en cuenta los mecanismos *etiopatogénicos* implicados en su producción:

- Disfonías orgánicas: cuando existe una clara alteración anatómica o estructural
- Disfonías funcionales: cuando no puede identificarse ninguna alteración anatómica o estructural.
- Disfonías orgánico-funcionales: la alteración inicial es funcional pero, por un mal uso o abuso vocal, acaban convirtiéndose en lesiones orgánicas que no conllevan en su evolución a patología maligna (aunque en su génesis puede participar como elemento favorecedor el tabaco) y que se denominan lesiones mínimas asociadas.

Hay que tener en cuenta que entre todos los tipos de disfonía existe una interrelación entre la causa y el efecto de los diferentes mecanismos, siendo muy difícil en ocasiones poder determinar una causa etiológica concreta, como se muestra en la Figura 1.1 [14].

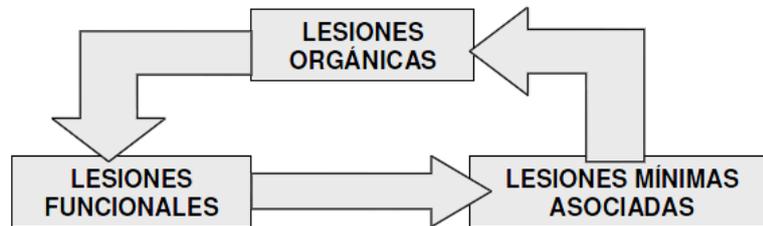


Fig. 1.1.- Interrelación de lesiones debido a las disfonías

En cuanto a las voces disfónicas; éstas pueden presentar diferentes características, siendo las más frecuentes la “voz aérea” (voz con exceso de escape de aire, sugiriendo un insuficiente cierre cordal; característico de parálisis cordales o masas de cuerda vocal que impiden el cierre de las mismas durante la fonación) y la “voz rasposa” (secundaria a alteraciones del borde libre de la cuerda vocal, producida por laringitis, pólipo o nódulo cordal, entre otros). También está la “voz tensa”, apretada, que tiene su máxima expresión en las distonías laríngeas; y la “voz asténica” o cansada, débil, frecuente de encontrar en pacientes hospitalizados por periodos largos, con pérdida de peso y masa muscular.

El concepto de mal uso vocal se refiere a un comportamiento distorsionado e involuntario del aparato fonatorio y que interfiere con la producción vocal efectiva. El concepto de abuso vocal se refiere a un uso incorrecto del tono y de la intensidad en la producción vocal. Este último posee mayor capacidad potencial para ocasionar daño sobre la mucosa laríngea [1].

2.5. La reeducación vocal

La voz es sin duda el instrumento de expresión y comunicación de trabajo más difundido en la vida moderna. Todas las personas utilizan la voz a diario, es el vehículo y condicionante del trabajo en nuestra vida, en casi todas las manifestaciones. Para algunas profesiones el uso de la voz es fundamental para el desempeño laboral, como actores, cantantes, locutores, oradores, telefonistas docentes y otros.

En la actualidad, nos encontramos en que una gran parte de la población presenta trastornos vocales como consecuencia de sobreesfuerzos del instrumento, ya sean *distonías*, *nódulos*, *afonías* u otros, o simplemente que su emisión de voz no es la más adecuada, produciendo cansancio y el consiguiente forzamiento vocal.

Entre la población docente existe un alto porcentaje de profesores que utilizan mal la voz y realizan esfuerzos musculares para obtener los sonidos con intensidad y amplificadas, para así poder ser escuchados. Hay una estrecha correlación entre éstos y los niños. Es muy frecuente encontrar a niños con dificultades en la fonación, pero vemos que donde realmente radica el problema es en los modelos vocales, o sea, en sus padres, familia o en las personas que más horas pasan con ellos, que son sus maestros.

Es también habitual, tanto en los adultos y en niños, y sobre todo en la etapa de la adolescencia, ver que no se sienten identificados con su voz, no les gusta y buscan patrones de conducta vocal cercanos a sus deseos, tendiendo a hacer el tono de voz más grave buscando resonancias más profundas, oscuras y atractivas, adaptando su voz a patrones erróneos y como consecuencia a un mal funcionamiento de su órgano vocal.

En la transmisión de comunicación en el canto es bastante frecuente utilizar el ejemplo directo del maestro. Es habitual que el maestro interprete y el discípulo ejecute, lo que resulta muy difícil aprender a controlar y verificar todos los movimientos que hacen falta para tener una buena emisión vocal, ya que las sensaciones no son iguales en todas las personas y tienen que aprender a investigar sobre ellos mismos.

En el estudio del canto, los alumnos se hacen siempre la misma pregunta: “¿Cómo sé si es correcto lo que estoy haciendo?”. Realmente no saben a ciencia cierta si lo están haciendo bien o no. Tienen la sensación que, sin la ayuda de su profesor poco o nada pueden hacer por ellos mismos, y no tienen nada tangible a lo que aferrarse. El control de la voz por el propio oído les resulta muy difícil porque carecen de una referencia clara para cerciorarse de que están emitiendo un sonido correcto, por lo tanto, el estudio y la práctica en casa se les hace mucho más engorroso. Es bastante ingrato repetir y repetir sin tener la certeza de que están trabajando adecuadamente, y sólo les queda esperar una semana, hasta la próxima clase, para que su profesor les verifique si han realizado bien los ejercicios [2].

Actualmente existen diferentes estrategias para reeducar la voz los cuales se describen a continuación:

- *Método de Voice Craft*: se basa en la fisiología y la anatomía del cuerpo. El método consiste en dividir el proceso de producir sonidos en 13 ejercicios o Figuras Obligatorias, que aíslan las estructuras independientes en el mecanismo vocal. Esto permite al cantante escoger que “ingredientes” combinan en la “receta” de cualquier sonido particular que desee hacer.

- *Método Speech Level*: creado por Seth Riggs, es una técnica de entrenamiento vocal que consiste en mezclar resonancias entre pecho y cabeza. Esta mezcla es gradual al pasar de registro. Cantar con los músculos internos de la laringe no con los externos, por lo que se necesita menos esfuerzo para producir el sonido. Se desarrolla un sonido parejo y consistente sin discrepancias al cambiar de un registro a otro.
- *Método Linklater*: fue diseñado por Kristin Linklater, su enfoque se caracteriza en que no desenraza la voz de la persona “Escuchamos la persona, no escuchamos su voz”. Consiste en una progresión de ejercicios al servicio de la libertad de expresión y cuya meta es la libertad de la voz natural.
- *Complete Vocal Technique*: creado por Cathrine Sadolin, según su método protege anatomía y el uso de la voz mediante cuatro modos vocal (neutrales, retenido/frenado, directos, extremos), los colores del sonido y la asistencia urgente; opta por una manera sana de hacer efectos vocales.
- *Método Vocal Power*: creado por Elisabeth Howard, es un método “step - by step” de alineamiento vocal que permite un mejor control de la voz y una mayor libertad expresiva.
- *Método Proel*: método propioceptivo elástico energético basado en los conceptos fisiológicos de la voz aplicados al canto: La viscoelasticidad y la lubricación de la laringe, la dinámica del velo del paladar, de la lengua y de la musculatura facial para abrir el tracto vocal.

2.6.- El aparato vocal

Una cuestión básica a considerar es que la voz humana es única y es exclusivamente aire, que después de unas modificaciones producidas por nuestro aparato vocal se convierte en un sonido que se propaga a través del espacio.

Si partimos de esta premisa, podemos llegar a concluir que el aparato vocal es como un instrumento musical que funciona gracias al paso del aire. Como un instrumento más, diremos que tiene la facultad de emitir un sonido lo suficientemente fuerte y armónico para hacerse oír solo o en compañía de otros instrumentos.

A causa de la naturaleza de la voz y de su sistema de producción, se ha comparado el aparato vocal con un instrumento de viento. Clásicamente, en él se han descrito las mismas partes principales que en el mecanismo de un órgano. Así, podemos distinguir: el fuelle o aparato productor de aire; la lengüeta, que vibra y nos da la altura del sonido, y el tubo o caja de resonancia, en el que éste se modifica y se amplía.

2.6.1.- El fuelle

Como el fuelle que suministra el aire en el órgano, el cuerpo humano utiliza los pulmones y los músculos de la respiración para dar el aire necesario para hacer vibrar las cuerdas vocales. Los mismos elementos serán utilizados, con las variaciones técnicas necesarias, por un instrumentista que toque cualquiera de los instrumentos de viento de una orquesta. Para la voz cantada se necesitará más cantidad de aire espirado que para la voz hablada. Será necesario aumentar la duración de la espiración y el control sobre la misma, mientras que la inspiración será lo más rápida posible para no interrumpir, en según qué casos, la melodía.

2.6.2.- El vibrador

El aparato vibrador, que en el órgano es la lengüeta, en el aparato fonador son las cuerdas vocales. A diferencia de lo que sucede en el órgano, las cuerdas vocales se pueden alargar o acortar a voluntad. Esto determinará que, mediante las órdenes de nuestro cerebro, y en función de la nota que queramos emitir, la tensión y la longitud de las cuerdas vocales se regulen automáticamente.

2.6.3.- Los resonadores

La caja de resonancia, que en el órgano son los tubos más o menos largos según la nota tocada, en el cuerpo humano la forman todas aquellas cavidades situadas por encima de las cuerdas vocales (*faringe*, boca y cavidad nasal). Estas estructuras presentan la ventaja, en comparación con los tubos del órgano, de poder cambiar de forma en función del sonido emitido.

La cavidad nasal es el único resonador fijo, ya que sus paredes son rígidas y no puede cambiar de volumen y forma, pero esta cavidad será utilizada como resonador de la voz únicamente en determinados sonidos nasales, tanto en el habla como en el canto. La *laringe*, que contiene las cuerdas vocales, puede desplazarse verticalmente variando, según su posición más alta o más baja, el tamaño y la longitud de la *hipofaringe*. La posición de la laringe también influirá en la posición de la lengua y, por tanto, en el volumen de la cavidad bucal. La boca, que será el principal resonador de la voz, puede variar su abertura, por medio de la lengua y los labios el sonido puede modificarse y convertirse en palabras o canto articulado.

Como en el órgano, es necesaria la coordinación de todas las partes para poder hablar o cantar de la manera más correcta posible, evitando todo tipo de tensión que pueda lastimar el aparato fonador.

2.6.4.- Fonación

La fonación es el trabajo muscular realizado para emitir sonidos inteligibles, es decir, para que exista la comunicación oral. El objetivo último de la fonación es la articulación de palabras, a través del proceso por el cual se modifica la corriente de aire procedente de los pulmones y la laringe en las cavidades *supraglóticas* (encima de la *glotis*) como consecuencia de los cambios de volumen y de forma de estas cavidades. El conjunto de las cavidades supraglóticas puede dividirse en tres partes: la faringe, la cavidad bucal y la cavidad nasal.

El sistema fonatorio, como se aprecia en la Figura 2.1 se vincula con otros sistemas. La interacción de éstos es parte activa en la función fonatoria, que se regula por el sistema nervioso central y periférico. El sistema fonatorio es parte integrante del cuerpo, vinculado con otras estructuras y sistemas. La interacción de estos sistemas participa activamente en la función fonatoria, todo regulado por el sistema nervioso central y periférico.

Es por todo esto que los procesos de atención y memoria son fundamentales en la educación, reeducación y rehabilitación de la voz. La atención focalizada permite procesar algunos estímulos sensoriales ignorando otros. Los que llegan al encéfalo en paralelo, se seleccionan o bien al inicio del procesamiento perceptivo o bien luego de ocurrido el análisis por significado. Esto ocurre en el transcurso del tratamiento de la voz.

Utilizamos el modelo de memoria de trabajo para explicar los procedimientos de la terapia vocal. Comprende el ingreso de la información del ambiente, su retención durante períodos breves y los mecanismos de ensayar tanto la información verbal como visual para introducirla en la memoria a largo plazo. Esto permite automatizar el aprendizaje realizado en el tratamiento de la voz.

Es indispensable el conocimiento de los procesos neurocognitivos; es decir, de procesos cognitivos mediante el cual las neuronas realizan sinapsis para dar como resultado el almacenamiento de la información que podrá ser luego evocada, a la comprensión acabada de la producción de la voz, la educación, reeducación y rehabilitación vocal [3].

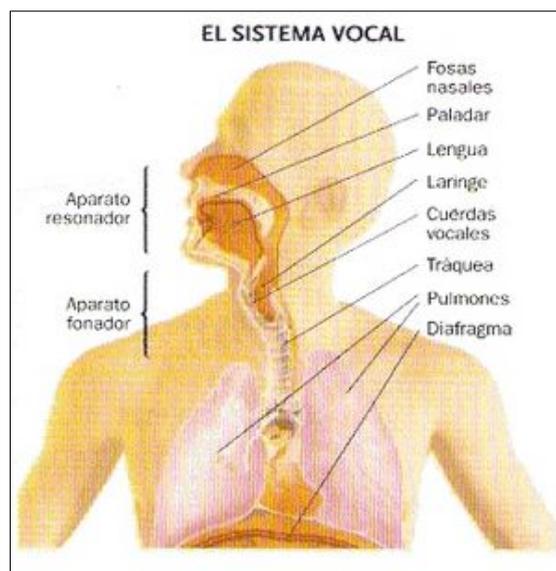


Fig. 2.1.- Sistema fonatorio

2.7. Las cualidades de la voz

En las cualidades acústicas de la voz intervienen una multitud de fenómenos que interactúan en la laringe para proporcionar a la voz humana una variabilidad notabilísima, que le permite adaptarse con precisión a las múltiples circunstancias en la que tiene que manifestarse. Sin embargo, todas las variaciones de la voz se relacionan con modificaciones de 3 cualidades acústicas del sonido vocal, como son la intensidad, la altura y el timbre, que se producen de forma aislada o, con mayor frecuencia, conjunta.

2.7.1. Intensidad

La intensidad de un sonido o volumen, por consiguiente de la voz, traduce la amplitud de la variación de presión correspondiente a dicho sonido. El sonido es el resultado de la transmisión de una variación cíclica de presión, siempre que su período se sitúe dentro del espectro audible (entre 30 y 16.000 ciclos/segundo).

2.7.2. Altura

La altura de la voz o tono, o frecuencia fundamental del sonido, depende directamente de la periodicidad del movimiento de los labios glóticos, es decir, en la práctica del número de aperturas glóticas por segundo.

La altura tonal que utiliza una persona determinada depende en parte del tamaño de su laringe; cuanto más largo sean los pliegues vocales, más grave puede ser en principio la voz.

El conjunto de frecuencias que utiliza una persona se denomina extensión de la voz. El conjunto de las frecuencias que se utiliza en la voz cantada se denomina tesitura. (En un cantante suficientemente entrenado pueden llegar a coincidir su tesitura y su extensión).

2.7.3. Timbre

Es una característica muy importante del sonido vocal. En efecto, principalmente por el timbre es como puede identificarse a una persona cuando se oye su voz. Por otra parte, desde un punto de vista estético, la calidad de una voz se basa sobre todo en las cualidades de su timbre.

Desde el punto de vista físico, el timbre de un sonido con carácter musical, como por ejemplo el de la voz cantada, depende del número y de la intensidad relativa de los armónicos que contiene dicho sonido.

Los armónicos corresponden a frecuencias múltiples del sonido fundamental que superponen al mismo según una distribución particular en la escala de frecuencias.

El timbre de la voz depende, por otra parte, de las formas de aproximación, de los pliegues vocales y, por otra, de las características anatómicas de las cavidades de resonancia (faringe, boca, y en menor grado cavidad nasal), así como de su disposición [4].

2.8. Incidencia del funcionamiento laríngeo en las cualidades de la voz

La voz es el sonido producido voluntariamente por el aparato fonatorio humano. Es el instrumento más antiguo y más natural con el que se puede hacer música.

Según las leyes de la Acústica, hay tres elementos indispensables para la producción del sonido: un cuerpo vibrante, un medio elástico que propague las vibraciones y una caja de resonancia que las amplifique, a fin de que puedan ser percibidas por el oído. El aparato fonatorio humano ilustrado en la Figura 2.2, cumple con las tres condiciones señaladas: el cuerpo que vibra son las cuerdas vocales, situadas en la laringe; el medio de propagación es el aire proveniente de los pulmones y la caja de resonancia está formada por la cavidad torácica, la faringe, las cavidades oral y nasal y una serie de elementos articulatorios, que son los labios, los dientes, el alvéolo, el paladar, el velo del paladar y la lengua.

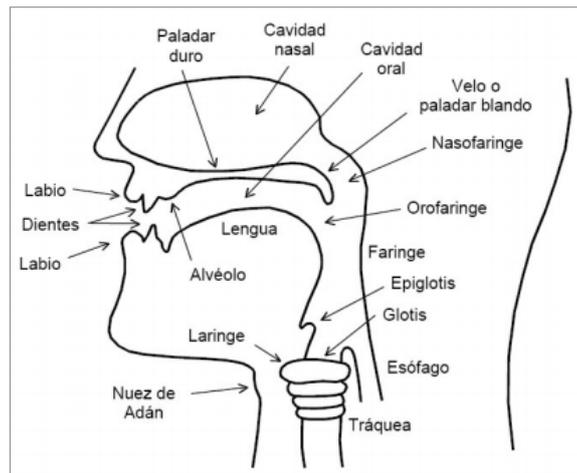


Fig. 2.2.- Corte esquemático del aparato fonatorio

La fonación se realiza durante la espiración, cuando el aliento o aire contenido en los pulmones sale de éstos, bajo la presión de los músculos abdominales, los intercostales y el diafragma, y, a través de los bronquios y la tráquea, llega a la laringe. Allí choca con las cuerdas vocales, también conocidas como pliegues vocales, que no tienen forma *cordófona*, pues se trata de unas bandas o membranas hechas de tejido muscular liso.

Las cuerdas vocales son cuatro. Las dos superiores (o bandas ventriculares) no participan en la fonación, proceso que se efectúa con las dos cuerdas inferiores. Éstas son, en realidad, dos pequeños músculos elásticos, orientados en la laringe de delante hacia atrás, tal y como se aprecia en la Figura 2.3. Por su parte anterior las cuerdas están unidas al cartílago tiroides (que puede palparse sobre el cuello, inmediatamente por debajo de la unión con la cabeza; en los varones suele apreciarse una protuberancia conocida como nuez de Adán). Por detrás, van sujetas a sendos cartílagos *aritenoides*, los cuales pueden separarse voluntariamente por acción de músculos. La abertura existente entre ambas cuerdas se denomina glotis.

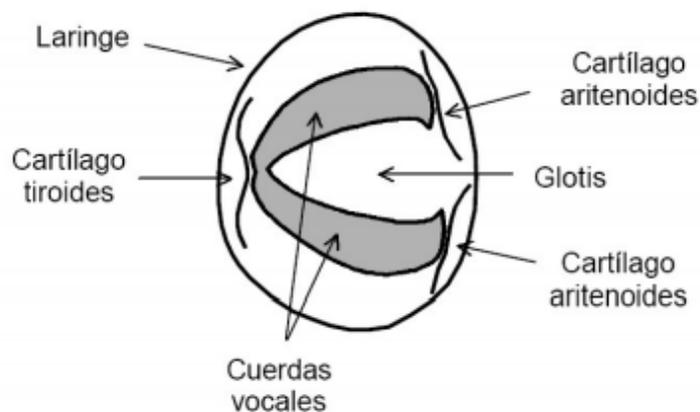


Fig. 2.3.- Corte esquemático de la laringe según un plano horizontal

Si las cuerdas vocales se encuentran separadas, la glotis adopta una forma triangular. El aire pasa libremente, sin hacer presión, lo que permite respirar y, prácticamente, no se produce sonido. Por el contrario, si la glotis comienza a cerrarse las cuerdas se juntan. El aire choca entre ellas y experimenta una turbulencia, emitiéndose un ruido de origen aerodinámico conocido como aspiración (aunque en realidad acompaña a una espiración o exhalación).

A cerrarse más, las cuerdas vibran a modo de lengüetas, lo que produce un sonido tonal. La frecuencia de este sonido depende de varios factores, entre otros del tamaño y la masa de las cuerdas vocales, de la tensión que se les aplique y de la velocidad del flujo del aire proveniente de los pulmones. A mayor tamaño, menor frecuencia de vibración, lo cual explica por qué en los varones, cuya glotis suele ser mayor que la de las mujeres, la voz es en general más grave.

Si aumenta la tensión de las cuerdas, la frecuencia vibratoria también lo hace, siendo el sonido producido más agudo. Así, para lograr emitir sonidos en el registro extremo de la voz es necesario un mayor esfuerzo vocal. A igualdad de otras condiciones, el aumento del flujo de aire también supone el aumento de la frecuencia, razón por la cual si se incrementa la intensidad de emisión se tiende a elevar espontáneamente el tono de voz.

El sonido producido en las cuerdas vocales es muy débil. Para convertirse en audible deber ser amplificado, lo que sucede en los órganos resonadores, es decir, la cavidades torácica y bucofaríngea, las fosas nasales, los senos nasales, frontales y maxilares y la cavidad craneana.

Una vez que la voz sale de los resonadores, es moldeada por los articuladores, que son una serie de elementos situados en el interior de la boca: paladar, lengua, dientes y labios. La articulación es el proceso mediante el cual alguna parte del aparato fonatorio interpone un obstáculo para la circulación del flujo de aire.

En su camino hacia el exterior, la frecuencia fundamental del sonido, derivada del frotamiento de las cuerdas vocales al paso del aire, va siendo enriquecida por los armónicos o *sobretonos* producidos por los resonadores, que son, a la postre, los que confieren a la voz una apariencia acústica, una peculiaridad y un timbre concretos.

La voz humana tiene una extensión, o respuesta en frecuencias, que oscila entre los 80 y los 1.000 Hz, aunque su eficiencia es mayor entre los 200 y los 700 Hz. En lo que se refiere a su estructura, los órganos de la fonación son iguales en el hombre y en la mujer, sólo difieren por sus dimensiones (mayor capacidad pulmonar, cuerdas vocales más largas y más sólidas en el hombre; diferente volumen de las cavidades de resonancia, etc.). Por término medio, entre adultos, las voces femeninas se sitúan una octava más alta que las masculinas.

Anteriormente se definió el tono, el cual hace mención a la gravedad de la voz. Hace referencia a las frecuencias de las vibraciones sonoras. Si se produce una baja frecuencia será una voz grave, y alta frecuencia será una voz aguda.

Cuando se produce una baja frecuencia y hay una alteración de la voz se suelen producir **nódulos**. Al proceso de cambio de la voz, se produce una inadaptación en el cambio de voz (muda vocal) tienden a darse momentos de alta frecuencia (gallitos al hablar) sobre todo se da en varones.

La voz depende de modificaciones físicas que afectan a los pliegues vocales y que conciernen a los siguientes aspectos:

2.8.1. Elongación

El tono o frecuencia fundamental es regulado a nivel glótico con los cambios de longitud, masa y elasticidad que experimentan los distintos planos de la cuerda vocal ante la acción muscular y que determinan las variaciones de frecuencia fundamental.

El acortamiento de la cuerda vocal produce un sonido más grave a disminuir la tensión y aumentar la masa de la superficie vibrátil por lo que disminuirá la frecuencia fundamental.

La elongación de la cuerda vocal produce un sonido más agudo, por tensar la cuerda vocal y disminuir la masa y la superficie de contacto. Esto se consigue predominantemente por la acción del músculo **cricotiroideo**. Si el músculo cricotiroideo está contraído y el músculo **tiroaritenideo** está relajado, la longitud de las cuerdas vocales está aumentada y su rigidez está globalmente aumentada en todas las capas. Inversamente si el músculo tiroaritenideo está contraído y el músculo cricotiroideo desactivado, la rigidez de la masa muscular aumenta y la frecuencia fundamental aumenta aunque la longitud de la cuerda vocal está disminuida.

Hay un control diferencial de la frecuencia fundamental por estos dos músculos que reciben una inervación motora diferente. El tracto vocal modifica la frecuencia de los formantes sobre todo en la región de las frecuencias altas. Una posición baja de la laringe en la vocalización de los cantantes se relaciona con emisión de frecuencias altas.

2.8.2. Grosor

Para entender adecuadamente la espectrografía de la voz se debe conocer la fisiología de la producción vocal al nivel de la glotis. Esta producción fisiológica del sonido se relaciona entonces con el gráfico del espectrograma que aparece en la pantalla. Si asumimos, para simplificar, que la glotis se abre y cierra como un obturador en un tubo uniformemente circular, cuando se abre y permite el paso del aire a través de la laringe hacia arriba se produce una presión positiva justo encima de las cuerdas.

Esta presión positiva se comunica a la glotis y ayuda a separar lateralmente las cuerdas, este movimiento lateral asistido de la cuerda vocal es equivalente a una transmisión de energía desde el aire a las cuerdas. Cuando las cuerdas se están cerrando, su tejido se mueve medialmente, no obstante todavía sigue pasando aire hacia arriba. Dado que pasa menor cantidad de aire a través de la glotis, la presión justo por encima de las cuerdas desciende y llega incluso a hacerse negativa debido a que sus moléculas se separan más y más creando una rarefacción del medio aéreo. En efecto, cuando la glotis se cierra, se crea una clara presión negativa por la falta de flujo aéreo que sale de la laringe, pero el aire continúa fluyendo hacia arriba en el tracto vocal. Durante el movimiento medial de las cuerdas, la presión *subglótica* desciende y se comunica a la glotis. Esta presión negativa (que succiona) ayuda o asiste al movimiento que cierra las cuerdas en su misma dirección y velocidad. La representación de la vibración vocal se puede apreciar en la Figura 2.4.

El concepto a resaltar es que el movimiento diferido de la masa aérea en el tracto vocal (debido a su inercia) crea presiones supraglóticas que se comunican a la glotis para asistir tanto al movimiento lateral como al medial de las cuerdas vocales.

Es importante destacar que las presiones en la glotis son asimétricas dado que son más positivas durante el movimiento lateral (hacia afuera) que en el movimiento medial (hacia adentro). Los requerimientos básicos son que la fuerza dinámica será menos positiva durante el cierre de las cuerdas que durante su apertura y que la fuerza dinámica neta a lo largo del ciclo entero sea suficiente para superar las fuerzas friccionales. La fuerza dinámica producida por las presiones debe ser suficiente para vencer toda la propiedad amortiguadora tisular, de otro modo la vibración no podría ser mantenida o quizá ni siquiera iniciada.

La presión subglótica mínima capaz de iniciar el movimiento vibratorio de las cuerdas vocales se relaciona con varios factores como son el grosor de las cuerdas vocales, fricción de las mismas y velocidad de la onda mucosa.

Así, unas cuerdas con un grosor adecuado expondrán una mayor superficie de su borde libre a interactuar con las presiones que acompañan a la columna aérea, con lo que requerirán una menor presión subglótica para comenzar a vibrar que unas cuerdas con un grosor reducido. Una menor fricción o viscosidad de las cuerdas permitirá una mayor facilidad de los tejidos en iniciar el movimiento con presiones bajas. La velocidad con que la onda mucosa viaja desde la parte inferior del borde libre de la cuerda al superior se relaciona con mayor rigidez de la mucosa, situación que se observa en la fonación con tonos altos, con lo que la emisión de sonidos con frecuencia fundamental alta se acompaña de presiones subglóticas elevadas.

El tono vocal es una función controlada por la longitud, tensión y amplitud del movimiento de las cuerdas vocales y por la actividad del músculo tiroaritenoides.

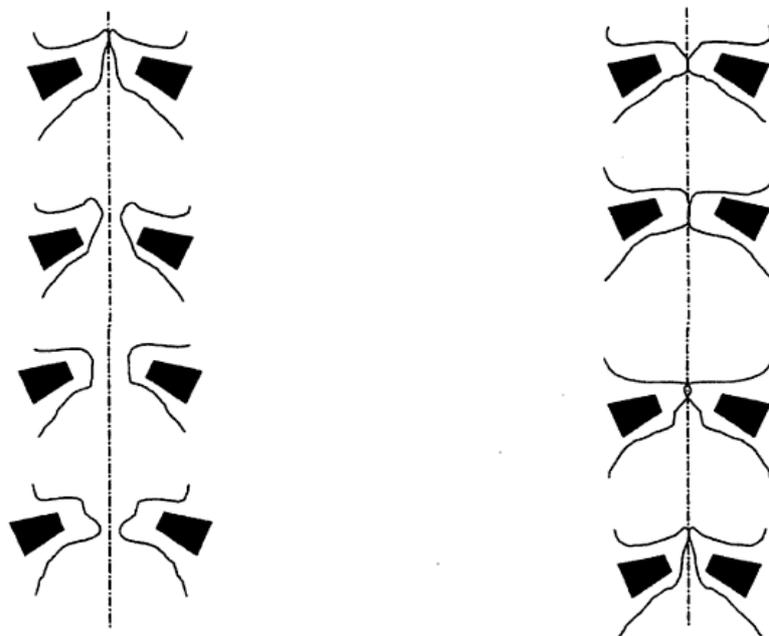


Fig. 2.4.- Representación esquemática y simplificada de la vibración vocal en un corte frontal de las cuerdas, donde se ven tanto la mucosa como el músculo

Las cuerdas vocales pueden modificar el tono de forma pasiva o de forma activa. De forma pasiva lo consigue al aumentar la tensión por medio de dos factores que producen una elongación dinámica de las mismas. El primero es la acción de los músculos cricotiroideos, el epitelio que cubre las cuerdas presenta la mayor tensión cuando las mismas se hallan elongadas por su acción, a semejanza de lo que ocurre con una cuerda de un instrumento musical que a mayor tensión ejercida sobre ella producirá sonidos más agudos. El segundo se consigue al incrementar la presión subglótica, las cuerdas alcanzan su máxima separación con lo que aumenta su tensión.

La contracción del músculo tiroaritenoides controla de forma activa el tono vocal de forma compleja según sea la actividad simultánea de los músculos cricotiroideos.

La calidad de la voz depende del grado de abducción de las cuerdas vocales, si es débil, la voz tendrá una calidad espirada, si es exagerado la tendrá áspera y si es normal será una voz brillante, al ser armónicamente rica.

La intensidad de la voz se incrementa con los incrementos de la presión subglótica, de la frecuencia fundamental y de la velocidad del flujo aéreo a través de la glotis. Al doblar la presión subglótica se obtiene un aumento de 10 dB en el nivel de presión sonora de la fonación. El grado de abducción vocal también afecta a la intensidad, demasiado débil o demasiado fuerte producen una fonación inadecuada, en circunstancias normales el grado óptimo de abducción que maximiza la intensidad resultante es cuando se observa que las apófisis vocales de los *aritenoides* casi se tocan o lo hacen con suavidad [15] [16].

2.8.3. Masa muscular vibrante

La laringe y las cuerdas vocales pueden vibrar más o menos dependiendo de la constitución de cada persona, pero también de su edad. Por otra parte, la denominada "huella vocal" está determinada por la forma del tracto laríngeo superior, que varía de un individuo a otro, por lo que la resonancia de la voz también varía. El primer grito de un bebé, según los expertos, tiene una tonalidad media entre 440 y 500 Hz (muy aguda). A los siete años de edad, la frecuencia de la voz se ha reducido en 200 Hz, y ronda los 295 Hz para las niñas y los 268 Hz para los niños. Al alcanzar la pubertad, la testosterona masculina hace que la masa muscular vocal aumente, las paredes de la laringe se engruesen, crezcan los cartílagos laríngeos, se alarguen las cuerdas vocales y la voz de los varones se vuelva más grave. El tono de voz en los chicos disminuye aproximadamente una octava. La voz de las adolescentes femeninas también cambia debido a la acción de los estrógenos sobre las cuerdas vocales. Además, se producen frecuentes "gallos" por adaptación a la nueva modulación.

En la senectud, disminuye la testosterona en los hombres y la voz se vuelve a agudizar. En las mujeres, por el contrario, se vuelve ligeramente más grave por los estrógenos. Además, las cuerdas vocales pierden elasticidad y fibras de colágeno [17].

2.8.4. Presión de cierre

La acción del diafragma empuja el aire desde los pulmones a través de las cuerdas vocales, produciendo un tren periódico de pulsos de aire. Este tren de impulsos está conformado por las resonancias del tracto vocal. Las resonancias básicas llamadas formantes de voz, pueden ser cambiadas por la acción de los articuladores, para producir sonidos distinguibles de voz como las vocales.

Situado en la base de la laringe en el tracto vocal, estos repliegues gemelos de membrana mucosa, actúan como vibrador o "lengüeta" durante la fonación. Abierto durante la respiración, estos pliegues se cierran por el volteo de los cartílagos *aritenoides*, para hablar o cantar.

La presión positiva del aire de los pulmones los obliga a abrirse momentáneamente, pero el aire a alta velocidad produce una depresión por el efecto *Bernoulli* que los vuelve a unirse. Los propios pliegues tienen una frecuencia de resonancia que determina el tono de la voz.

En un hombre adulto, las cuerdas vocales suelen ser de 17-23 mm de largo, y de 12,5-17 mm en una mujer adulta. Por acción de los músculos en la laringe, se pueden estirar 3 o 4 mm. El promedio de la frecuencia de voz masculina es alrededor de 125 Hz, mientras que este promedio en el sexo femenino es sobre 210 Hz. Las voces de los niños promedian más de 300 Hz. En la ilustración correspondiente a la Figura 2.5, se muestra los tonos aproximados de voces relacionados con un teclado de piano temperado (ajustado a los doce sonidos, o lo que es igual a la doceava parte de la octava; siendo una octava es el intervalo que separa dos sonidos cuyas frecuencias fundamentales tienen una relación de dos a uno) basado en el sistema A4 = 440 Hz.

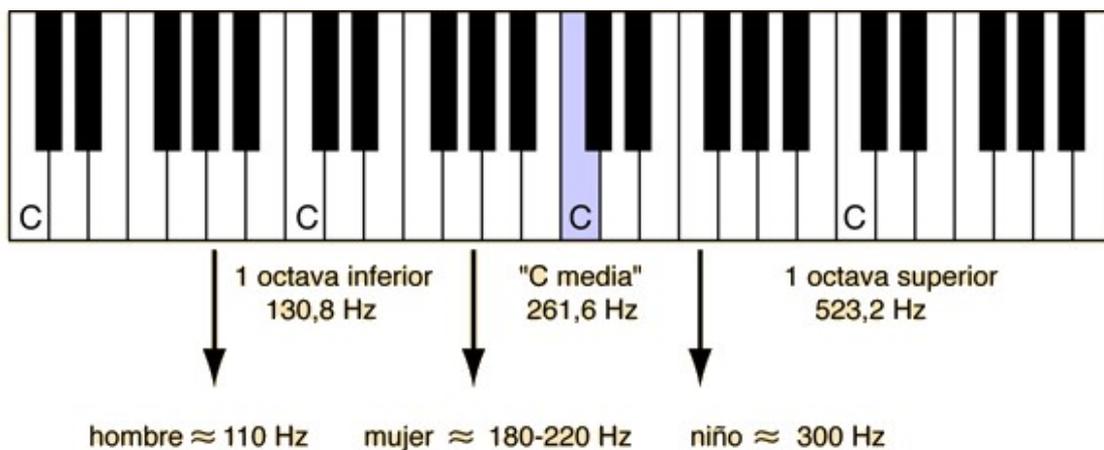


Fig. 2.5.- Tonos aproximados de voces relacionados con un teclado de piano

La parte delantera de las cuerdas vocales están unidas al cartílago tiroide, la "manzana de Adán". El extremo posterior está unido a los cartílagos *aritenoides*, que se mueven para separar los pliegues en la respiración.

El proceso de convertir la presión de aire de los pulmones en vibraciones sonoras se llama fonación. Cuando el aire pasa a través de las cuerdas vocales elásticas y las hacen vibrar, el tipo de fonación se denomina sonoración. Las cuerdas vocales dan al cantante una amplia gama de control sobre el tono del sonido producido. Aunque el término "pliegue vocal" es más descriptivo que "cuerda vocal", hay cierta similitud con las cuerdas vibrantes en donde el tono producido depende de la longitud, masa y tensión de las cuerdas vocales.

La excitación de las cuerdas vocales es sin embargo muy diferente de la excitación de una cuerda, ya que está causada por el paso del aire a través de la abertura entre los pliegues. Los músculos de la laringe cambian la elasticidad y la tensión de las cuerdas vocales para determinar el tono del sonido.

El ciclo vibratorio de las cuerdas vocales es impulsado por fenómenos aerodinámicos. La presión del aire conductor de los pulmones controla la apertura de los pliegues, y el efecto Bernoulli controla la fase de cierre. Cuando la parte superior de los pliegues se está abriendo, el fondo está en el proceso de cierre, y tan pronto como la parte superior se cierra, la acumulación de presión empieza a abrir la parte inferior.

La vibración es entonces como una onda que se desplaza desde el fondo hasta la parte superior de las cuerdas vocales. Cada vibración permite escapar un soplo breve de aire, produciendo un sonido audible a la frecuencia de la abertura; este proceso se denomina sonoración.

La intensidad de la voz se puede aumentar con el flujo de aire de los pulmones y aumentando la resistencia de las cuerdas vocales. Las cuerdas vocales son arrastradas para separarse más, y estar separadas más tiempo durante el ciclo. Esto aumenta la amplitud de la onda de presión de sonido producido [18].

Capítulo 3

Representación de la Señal de Voz

3.1.- La calidad de la voz. Fenómenos implicados

En este capítulo se realiza un recorrido por los diferentes dominios de representación en los que se puede estudiar la señal de voz con el fin de describir las diferentes manifestaciones que nos permiten valorar la calidad de la voz. Cada una de las diferentes manifestaciones mencionadas caracteriza un fenómeno físico que interviene en la generación de la voz. Sin embargo el mismo fenómeno físico puede manifestarse en diferentes dominios, en forma de diferentes manifestaciones. Por tanto, en este capítulo se va a identificar el conjunto de fenómenos físicos necesarios, para realizar una correcta documentación de la calidad de voz de un locutor.

Medir la calidad de la voz consiste en cuantificar diferentes características acústicas de la voz que permitan compararla con respecto a ciertos patrones de referencia. En la bibliografía actual no existe una única medida que documente todas las características acústicas de la voz.

Una medida de la calidad acústica de la voz basada en una valoración auditiva es intrínsecamente difícil de establecer como referencia comparativa entre distintas voces y distintos evaluadores. En la bibliografía actual se han realizado varios intentos de objetivar la medida de la calidad de voz por medio de medidas clínicas multidimensionales basadas en métodos perceptuales. Existen ejemplos bien conocidos como son: la escala **GRABS** de Japón y su formulario extendido, desarrollado y aplicado en Europa, o el utilizado en Suecia donde se ha trabajado en otro juego de descriptores clínicos de la voz referentes a la percepción, o como en el que se introduce un juego de características fonéticas que intentan agregar información de la excitación del tracto vocal en la medida de la calidad de la voz. Estos procedimientos de medida de la calidad de la voz pretenden alcanzar una medida objetiva a partir de una valoración subjetiva.

Existen diferentes trabajos en los se proponen medidas objetivas de la calidad de la voz obtenidas a partir de una grabación de voz. Una característica común de la mayoría de estos trabajos es la de utilizar para el cálculo de las medidas de la voz la grabación de un sonido sonoro sostenido, principalmente una vocal.

Esto es debido a que durante la fonación de este tipo de sonidos el sistema de producción de voz pone en funcionamiento gran parte de sus mecanismos (flujo glótico de aire constante, vibración de las cuerdas vocales de forma continuada, etc.) permitiendo captar cualquier anomalía de estos mecanismos. El valor que toma cada una de las medidas de calidad de la voz depende de la vocal analizada. En estos trabajos, los distintos autores presentan diferentes conjuntos de medidas con la finalidad de cuantificar objetivamente la calidad de voz. Todos ellos tienen en común la necesidad de realizar diferentes medidas de la voz con la finalidad de recoger en el proceso de medida los distintos aspectos de sus características acústicas [5].

3.1.1. La calidad de la voz desde diferentes dominios

Una grabación de voz aporta diferentes características respecto a la calidad de la voz del locutor. Esta señal de voz registrada es posible representarla en diferentes dominios, lo cual resulta interesante ya que cada dominio en el que se puede representar la señal, manifiesta de forma preferencial alguna de las características de la voz del locutor. Los principales dominios estudiados son los siguientes:

- Dominio temporal
- Dominio espectral
- Dominio cepstral

Sin duda alguna gran cantidad de los trabajos sobre procesado digital de señales de voz existentes en la bibliografía actual centran sus estudios en estos dominios. Sin embargo, nuevas líneas de trabajo presentan nuevos dominios derivados de los anteriores. Así pues, resulta de interés estudiar nuevos dominios que cada vez toman mayor importancia en el procesado de señal: los dominios que manifiestan el comportamiento no lineal de la señal de voz.

A continuación se describen las características más relevantes de cada uno de los diferentes dominios. En este punto resulta necesario introducir el concepto de trama, ya que tanto para la visualización como para el análisis de estas características, resulta necesario segmentar la señal con el fin de optimizar la obtención de información que aporta cada dominio. A cada una de estas unidades fruto de la segmentación se la denomina trama. La duración temporal de las tramas, dependiendo de cada característica, suele ser constante (por ejemplo 20, 25, 100, 300 msec), o bien en algunos casos depende de una relación directa con respecto a algún parámetro fisiológico (por ejemplo, tres veces la duración de un ciclo de apertura y cierre de la cuerdas vocales) [6].

3.1.2. Dominio temporal

En el dominio temporal se manifiestan ciertas características de la calidad de la voz las cuales permiten diferenciar entre una muestra de voz de alta calidad y una muestra de voz de baja calidad. En la Figura 3.1 se muestra una representación de los cinco golpes de voz correspondientes a las cinco vocales castellanas pronunciadas de forma sostenida (se han eliminado los silencios existentes entre cada vocal) para una voz de buena calidad y para una voz de mala calidad.

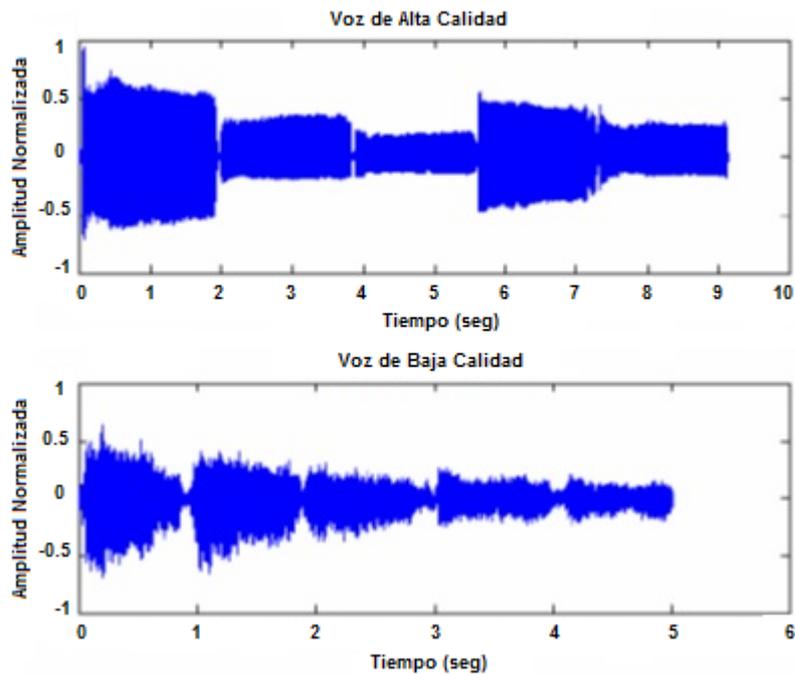


Figura 3.1.- Representación de la señal de voz en el dominio temporal correspondiente a la fonación de cinco vocales de forma sostenida de un locutor con voz de alta calidad (superior) y de un locutor con voz de baja calidad (inferior)

En la Figura 3.1 puede apreciarse que envolvente de la señal de voz de un locutor con voz de alta calidad resulta más regular que la envolvente de señal de voz del locutor con voz de baja calidad. Este hecho resulta más significativo en intervalos de tiempo muy pequeños. Los fenómenos más característicos que permiten diferenciar entre voz de alta calidad y voz de baja calidad son:

- La energía de la señal contenida en un intervalo de tiempo pequeño, varía mucho de un intervalo al siguiente en las señales de voz de baja calidad con respecto a las fluctuaciones de la de alta calidad. Esta característica de variabilidad también se manifiesta en variaciones en la forma de onda entre dos periodos consecutivos. Este fenómeno se puede observar en la Figura 3.2.

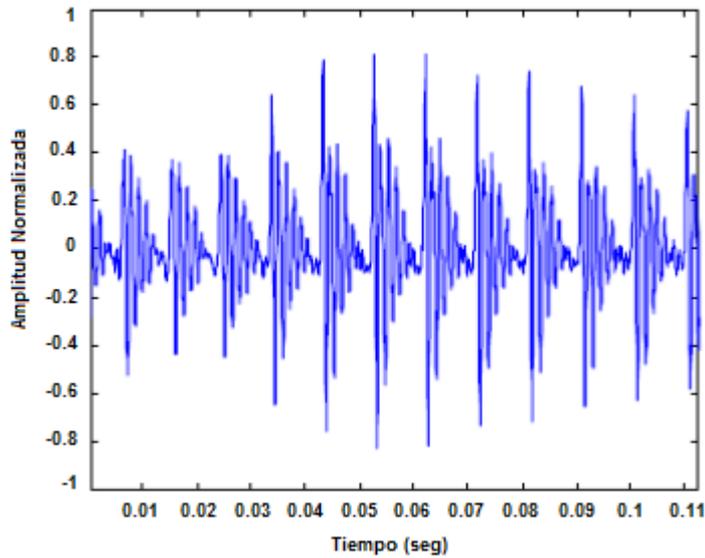


Figura 3.2.- Representación en el dominio temporal de un sonido sonoro sostenido

- Otra característica que presenta las voces de baja calidad es la existencia de intervalos carentes de periodicidad, lo que se identifica como la existencia de tramas sordas, durante fonaciones sonoras sostenidas. Un ejemplo de este fenómeno se muestra en la Figura 3.3.

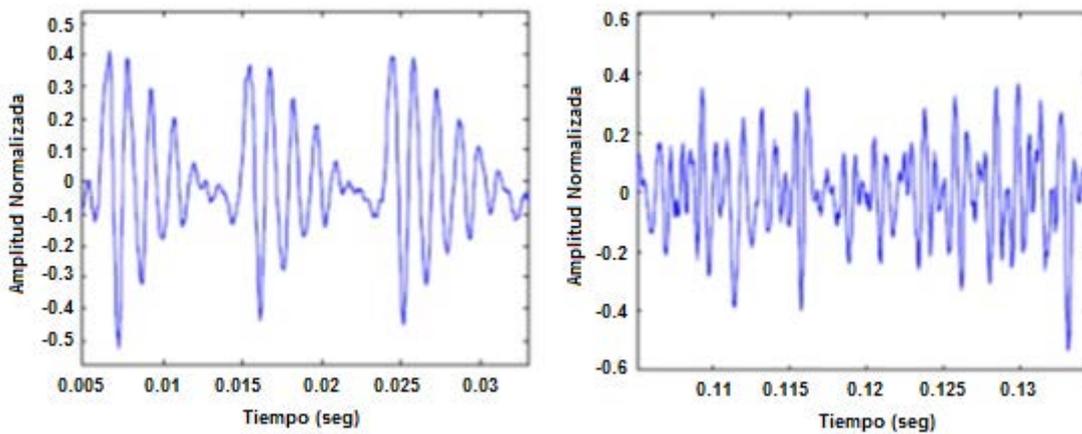


Figura 3.3.- Representación en el dominio temporal de un sonido sonoro sostenido durante un corto intervalo de tiempo para una voz de alta calidad (izquierda) y para una voz de baja calidad (derecha)

3.1.3. Dominio espectral

En este dominio es posible estudiar la contribución de potencia de cada una de las frecuencias (componentes espectrales) que forman la señal. Este estudio se realiza a partir de una nueva señal denominada espectro. El espectro se calcula a partir de una trama de voz.

El espectro de una trama de voz se define según la siguiente ecuación:

$$S_x(\omega) = \sum_{m=-\infty}^{+\infty} R_x(m) e^{-j\omega m} \quad [\text{Ec. 3-1}]$$

siendo $R_x(m)$ la función de autocorrelación, que se define como:

$$R_x(m) = E\{x(n)x(n+m)\} \quad [\text{Ec. 3-2}]$$

debido a que esta definición es estadística y no se conoce la función densidad de probabilidad de la señal de voz para cada una de las tramas, se asume que la trama de voz es estacionaria y ergódica, y se utiliza el siguiente estimador temporal del espectro:

$$\hat{S}_x(\omega) = \frac{1}{N} \cdot |X(\omega)|^2 \quad [\text{Ec. 3-3}]$$

donde $X(\omega)$ es la transformada de Fourier de cada una de las tramas. Existen otros estimadores temporales no paramétricos del espectro como el estimador de Welch y Blackman-Tukey, además de los diferentes métodos basados en los estimadores paramétricos del espectro) [7].

De los parámetros que caracterizan una señal de voz desde el punto de vista espectral están:

- *Formantes*: son picos en el espectro de voz consecuencia del efecto resonador debido a la acción articuladora de la cavidad vocal. Están caracterizados por su frecuencia central y su ancho de banda. Para un adulto las resonancias están centradas en torno a 500, 1500, 2500 y 3500 Hz. Existen pequeñas diferencias en las frecuencias de los formantes entre locutores de distinto género. Generalmente las frecuencias de los formantes de locutores femeninos son mayores que las frecuencias de locutores masculinos.

- *Frecuencia fundamental* o *Pitch*: es la frecuencia de vibración de las cuerdas vocales. En realidad no hay periodicidad total, este es el motivo de hablar de “casi periodicidad” en la frecuencia fundamental.
- *Timbre* o *matiz*: el movimiento vibratorio generador del sonido es, en general, un fenómeno complejo en el que intervienen simultáneamente, un movimiento vibratorio principal (vibración de las cuerdas vocales) y por otro lado, uno o más movimientos vibratorios secundarios. Los tonos secundarios (armónicos) resultan de las resonancias que el tono principal (armónico principal o *pitch*) produce en la cavidad o cavidades formadas en el tracto vocal. En este conjunto sonoro de tono fundamental (*pitch*) y tonos secundarios, el resonador predominante determina el timbre o matiz característico de cada sonido. Los sonidos son por su timbre, así como por su tono, agudos o graves, según la altura de la nota que corresponde a su resonador predominante.

El dominio espectral permite diferenciar las muestras de voz de alta calidad y las muestras de voz de baja calidad. En la Figura 3.4 se muestra una representación del espectro en escala logarítmica de una vocal sostenida producida por un locutor con voz de alta calidad y el espectro en escala logarítmica de la misma vocal producida por un locutor con voz de baja calidad.

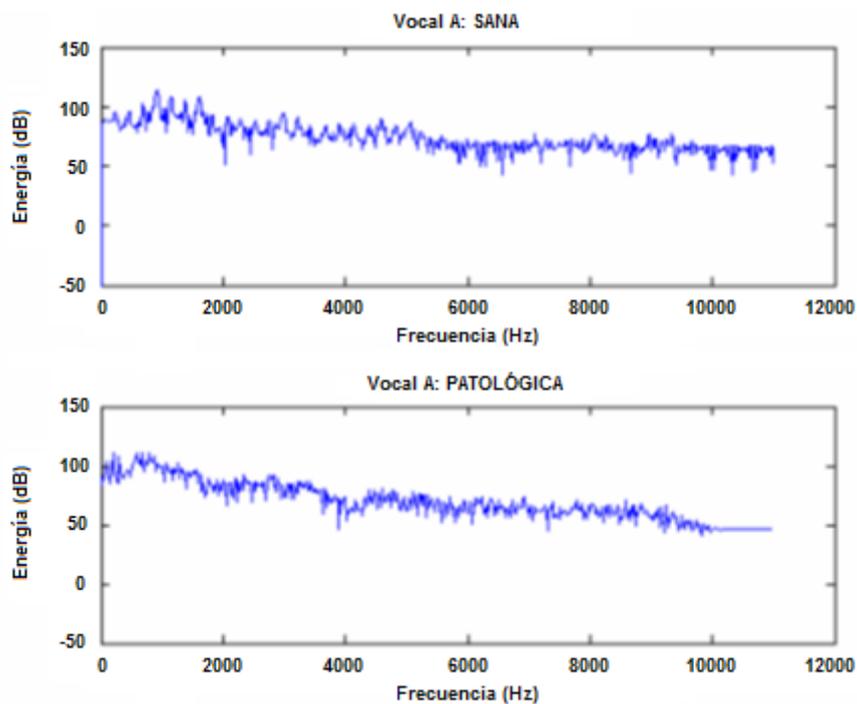


Figura 3.4.- Representación del espectro en escala logarítmica de una vocal sostenida producida por un locutor con voz de alta calidad (superior) y el espectro en escala logarítmica de la misma vocal producida por un locutor con voz de baja calidad (inferior)

Frente a las voces de alta calidad, en las muestras de voz de un sonido sonoro sostenido producidas por un locutor con voz de baja calidad se presentan las siguientes características diferenciales:

- Menor regularidad de la envolvente del espectro, principalmente en las bajas frecuencias.
- Mayor porcentaje de energía en las bajas frecuencias respecto a la energía total
- Aumento de energía en las altas frecuencias, que es debida a la presencia de ruido glótico.
- Gran variación del espectro de una trama con respecto a las tramas contiguas resultando significativa la variación del espectro entre tramas contiguas con duración de tres periodos de *pitch*.

En la siguiente figura se muestra la envolvente del espectro para un sonido sonoro sostenido. Cada uno de los picos de la envolvente del espectro se corresponde con los diferentes formantes. Los cuatro primeros formantes han sido utilizados ampliamente en trabajos de reconocimiento de voz. En la Figura 3.5 se muestra la posición de los cuatro primeros formantes para un sonido sonoro sostenido correspondiente a una vocal.

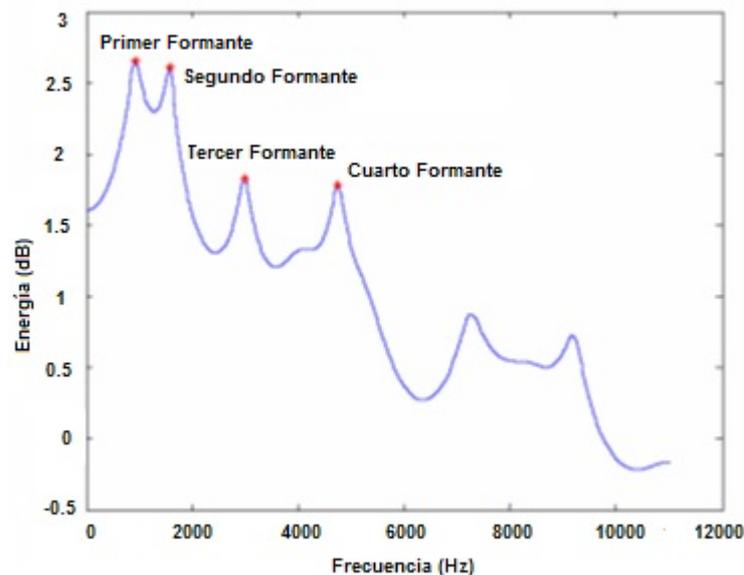


Figura 3.5.- Envolvente del espectro de una trama de voz donde aparecen señalados los cuatro primeros formantes

Una característica de las voces de alta calidad es la energía del espectro concentrada alrededor de determinados formantes, principalmente el primer y el tercer formante, mientras que las voces de baja calidad se caracterizan por presentar gran componente de ruido alrededor de los mismos.

El problema de una representación espectral basada en el espectro es la pérdida de información en lo referente a su evolución temporal, es decir, el espectro no representa la variación espectral correspondiente a las diferentes tramas. Sin embargo, existe otra forma de representar la información espectral denominada espectrograma, la cual solventa esta carencia, realizando una representación del espectro a lo largo del tiempo. Para ello, representa la potencia de las diferentes componentes espectrales por medio de la intensidad de color utilizada en la representación, utilizando los ejes para representar las frecuencias y el tiempo.

En la Figura 3.6 se muestra el espectrograma de una muestra de voz de alta calidad y una muestra de voz de baja calidad durante la fonación de las cinco vocales. Existen dos tipos de espectrogramas, espectrogramas de banda ancha y espectrogramas de banda estrecha, diferenciándose por la duración del intervalo temporal de la señal que se estudia. En las representaciones de la Figura 3.6 se utilizan ventanas de duración temporal significativas, y por lo tanto se obtienen espectrogramas de banda estrecha.

En el espectrograma de voz se aprecian características que identifican particularidades útiles para valorar la calidad de voz. Durante la fonación de un sonido sonoro sostenido se aprecian las siguientes características:

- Las voces con alta calidad de la voz se caracterizan por presentar líneas paralelas, correspondientes a las componentes armónicas, principalmente en las bajas frecuencias. En las altas frecuencias también es común la existencia de las componentes armónicas, siendo cada vez menos paralelas. Esta característica nos da idea de la riqueza espectral de la voz. Por el contrario, las voces de baja calidad presentan muy poca componente armónica, estando ésta concentrada en las componentes de muy baja frecuencia.
- Las voces de baja calidad presentan zonas de energía irregulares, lo cual indica la presencia de cantidades anormales de ruido. Este fenómeno se manifiesta principalmente en altas frecuencias y alrededor de los formantes, lo cual es significativo de la presencia de ruido glótico.

Otra característica típica de las voces de baja calidad es la de presentar, durante la fonación de un sonido sonoro sostenido, variaciones en el ritmo de vibraciones de las cuerdas vocales, lo cual es equivalente a decir que existe variaciones de la frecuencia de *pitch*. Para el cálculo de la variación del valor de *pitch* es necesario el uso de un detector de *pitch* exacto, debido a que diferentes medidas de calidad de voz se fundamentan en los valores de *pitch*. Existe mucha bibliografía sobre detectores de *pitch*, destacando entre ellas [7].

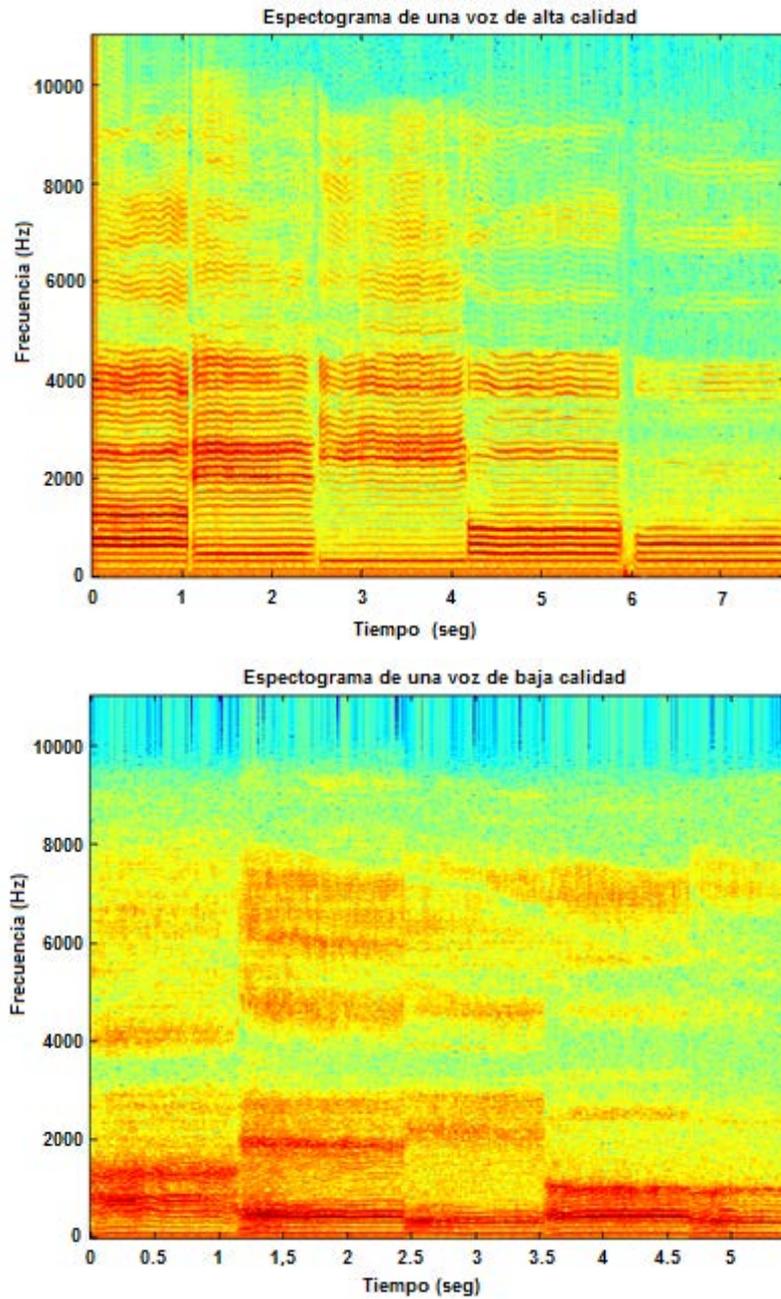


Figura 3.6.- Espectrograma (banda estrecha) estimado para una frecuencia de muestreo de 22050 Hz de una muestra de voz de alta calidad (superior) y una muestra de voz de baja calidad (inferior) durante la fonación de la cinco vocales

En la Figura 3.7 se muestra los valores que toma la frecuencia de *pitch* para las distintas tramas de una muestra de voz sana y de una muestra de voz patológica.

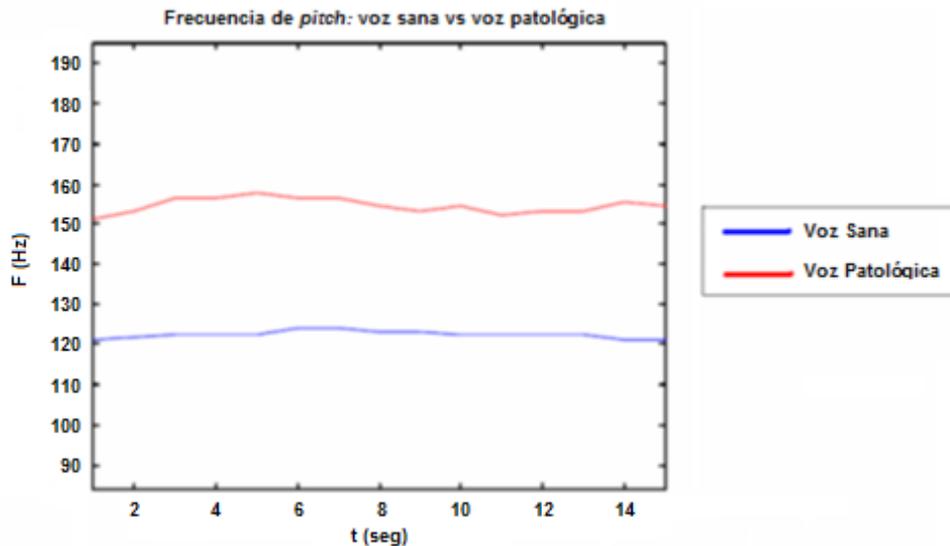


Figura 3.7.- Valores de la frecuencia de pitch para las distintas tramas de una muestra de voz de alta calidad (color azul) y de una muestra de voz de baja calidad (color rojo). Ambas muestras corresponden a la vocal "a" sostenida de un locutor masculino

Existen variantes al dominio espectral, como por ejemplo el dominio melespectro, en el que la escala de frecuencias no es lineal permitiendo simular la percepción acústica, o el dominio de los LP (*Linear Predictor*), en el cual se estudia el efecto de la envolvente del espectro, o bien una representación utilizando MBE (*Multi-Band Excitation*) en el que se pretende identificar qué partes de espectro corresponde a parte armónica y qué partes corresponde a las componentes de ruido [6].

3.1.4. Dominio cepstral

Los *cepstrum* complejos y los *cepstrum* de potencia han sido ampliamente utilizados en el procesado digital de señales. Los *cepstrum* complejos de una secuencia discreta $x[n]$ de longitud N, cuya transformada de *Fourier* es $X(\omega)$, se definen como:

$$C_x(n) \equiv F^{-1}\{\log(X(\omega))\} \quad [\text{Ec. 3-4}]$$

donde $F^{-1}\{\}$ denota la transformada inversa de Fourier.

Los *cepstrum* de potencia, también llamados *cepstrum*, se definen como:

$$p_x(n) \equiv F^{-1}\{\log(|X(\omega)|^2)\} \quad [\text{Ec. 3-5}]$$

La relación que existe entre los *cepstrum* complejos y los *cepstrum* de potencia es la siguiente:

$$p_x(n) \equiv c_x(n) + c_x(-n) \quad [\text{Ec. 3-6}]$$

El dominio de los *cepstrum* de potencia es el comúnmente conocido como dominio *Cepstral*. Por medio de una representación *cepstral* es posible obtener información sobre la envolvente del espectro, la riqueza espectral e identificar las componentes armónicas y las componentes de ruido.

En la Figura 3.8 se muestra una representación de los *cepstrum* para una muestra de voz de alta calidad y para una muestra de voz de baja calidad.

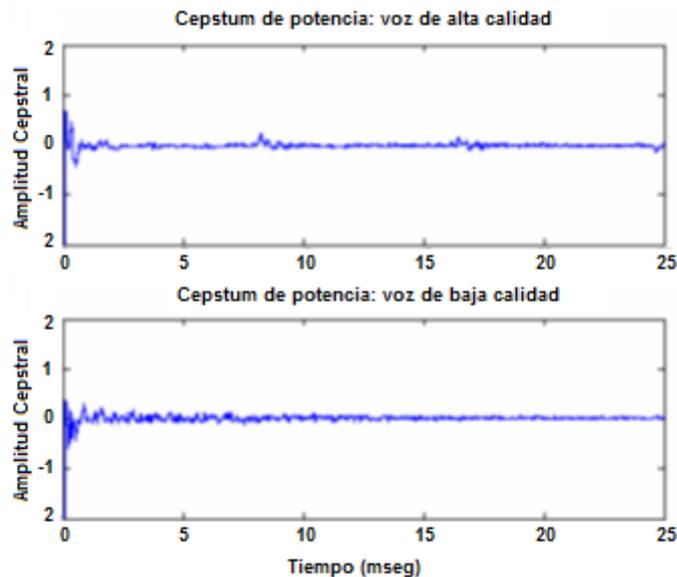


Figura 3.8.- Representación en el dominio de los *cepstrum* de potencia de una muestra de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior)

Es posible apreciar de forma más descriptiva la información que aportan los *cepstrum* de potencia realizando un rectificado, es decir, igualando los valores negativos a cero. En la Figura 3.9 se muestra una representación de los *cepstrum* de potencia rectificados aplicando un zoom en la zona de mayor interés.

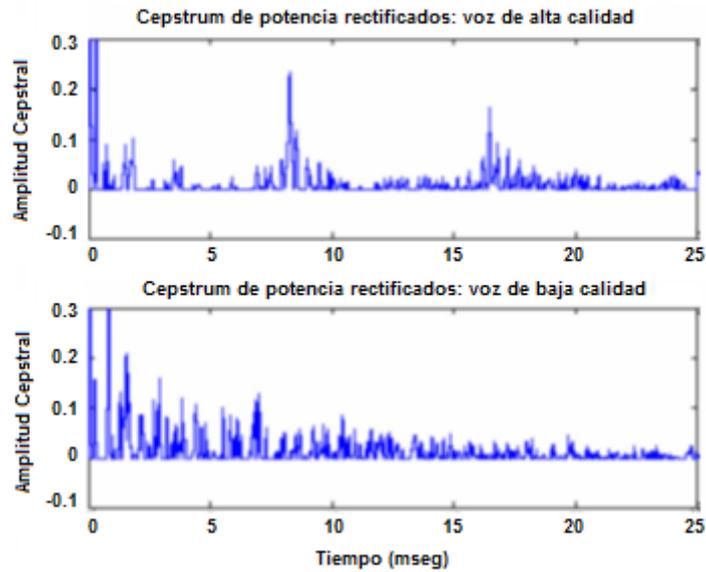


Figura 3.9.- Representación en el dominio de los cepstrum de potencia rectificados de una muestra de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior)

Por medio de una representación cepstral es posible identificar características que permiten valorar la calidad de la voz. En el caso de una fonación de un sonido sonoro sostenido, realizando una representación cepstral de diferentes tramas con duración equivalente a tres periodos de *pitch*, pueden identificarse las siguientes características:

- La riqueza espectral de una muestra de voz puede cuantificarse por medio de la amplitud y anchura de la componente cepstral correspondiente al *pitch*. La existencia de un pico de una amplitud considerable indica la presencia notable de energía en dicha componente armónica, siendo ésta una característica de las voces de alta calidad. A estos picos claramente identificados se les denomina *armónicos*. También, una anchura reducida del pico cepstral, correspondiente al *pitch*, para una trama de tres periodos de *pitch*, indica la alta estabilidad de la frecuencia de *pitch* para tres periodos consecutivos, también siendo ésta una característica de las voces de alta calidad.
- La presencia de ruido en diferentes regiones del *cepstrum* puede cuantificarse por medio de la relación entre: la componente armónica (las componentes cepstrales correspondientes al *pitch* y sus armónicos), y la componente de ruido (el resto de las componentes cepstrales).

En el dominio *cepstral* se manifiesta con claridad la presencia de componentes armónicas. En la Figura 3.10 se aprecia la localización del primer armónico (el *pitch*) y del segundo armónico.

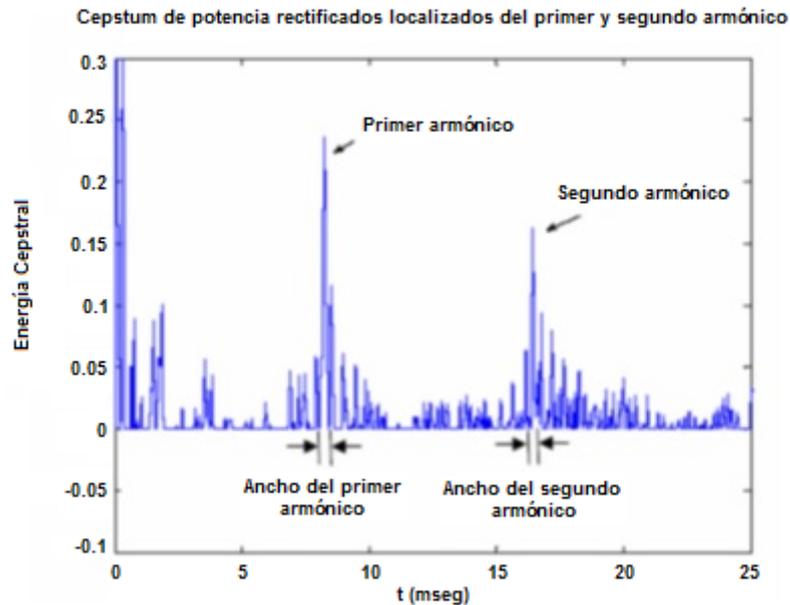


Figura 3.10.- Localización del primer y segundo armónico en el dominio de los cepstrum de potencia rectificadas

Características como la amplitud y la estrechez del pico *cepstral* correspondiente al segundo armónico son, al igual que ocurre con el pico del *pitch*, aspectos que diferencian a las muestras de voz de alta y baja calidad.

La presencia de un pico *cepstral* correspondiente al primer armónico que resulte más estrecho que el pico *cepstral* correspondiente al segundo armónico resulta una característica propia de las voces de alta calidad. Esta característica matemática en el dominio de los *cepstrum* se identifica con el fenómeno físico consistente en la capacidad de tener una frecuencia de vibración de las cuerdas vocales (frecuencia del *pitch*) muy estable, sumándose la capacidad del locutor de generar armónicos de frecuencia del *pitch* con variaciones a corto intervalo de tiempo. Un ejemplo de este fenómeno se aprecia en la voz proyectada, donde el locutor es capaz de mantener el *pitch* constante y para cada uno de sus armónicos presentar ligeras variaciones respecto a los múltiplos exactos de la frecuencia de *pitch*, manifestándose más significativamente este fenómeno en los armónicos de altas frecuencias. Esta característica se denomina *vibrato*.

Existen otros dominios similares como los *delta-cepstrum*, los *mel-cepstrum*, y los *mel-delta-cepstrum*, etc., los cuales permiten destacar ciertas particularidades que se manifiestan en el dominio cepstral [6].

3.1.5. Fenómenos físicos implicados en la calidad de la voz

En los apartados anteriores se ha realizado un recorrido por los diferentes dominios de representación en los que se puede estudiar la señal de voz, describiendo las diferentes manifestaciones, propias de cada dominio, que permiten valorar la calidad de la voz. Cada una de las diferentes manifestaciones mencionadas, caracteriza un fenómeno físico que interviene en la generación de la voz y sin embargo el mismo fenómeno físico puede manifestarse en diferentes dominios, en forma de diferentes manifestaciones. En este apartado del capítulo se va a identificar el conjunto de fenómenos físicos necesarios para realizar una correcta documentación de la calidad de voz de un locutor.

Existen cuatro fenómenos físicos cuantificables que permiten caracterizar la calidad de la voz. Los cuatro fenómenos físicos identificados son los siguientes:

- *Estabilidad de la voz*: es la capacidad de un locutor de producir, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales. Este fenómeno físico se cuantifica a partir de medidas de estabilidad de la señal de voz.
- *Riqueza espectral*: se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora del tracto vocal con gran cantidad de componentes espectrales. Este fenómeno físico se cuantifica por medio del cálculo de la estabilidad de la frecuencia de *pitch* y por la presencia de armónicos con energía elevada en las diferentes bandas de frecuencias.
- *Presencia de ruido*: se define como la presencia de ruido glótico en la voz, durante la fonación de un sonido sonoro sostenido, debida a carencias en la fase de cierre de las cuerdas vocales. Este fenómeno físico se cuantifica por medio de la presencia de ruido no estacionario en la voz.
- *Comportamiento no lineal*: este fenómeno físico es ocasionado por un funcionamiento anómalo de las cuerdas vocales, debido a irregularidades en las masas que intervienen en la fase de cierre de las cuerdas. También se debe a la existencia de asimetrías en el movimiento de las cuerdas y factores relacionados con la mucosa. Estos fenómenos son cuantificables por medio de la identificación de comportamientos no lineales en la señal de voz.

Una voz de calidad baja presenta al menos uno de los valores correspondientes a la cuantificación de los cuatro fenómenos físicos fuera de los rangos de normalidad. Este procedimiento de cuantificación de la calidad de la voz permite identificar calidades de voz anómalas de diferente origen, es decir, la voz de un locutor con baja calidad debida a un estado incipiente de una patología orgánica presentará probablemente la alteración de la *estabilidad de la voz* mientras que otro locutor con un estado incipiente de una patología funcional presentará valores anormales de “presencia de ruido”. A medida que la calidad de la voz de un locutor empeora aumentará el número de fenómenos físicos y la desviación de los valores de la cuantificación de dichos fenómenos respecto a los valores de normalidad.

Estos cuatro diferentes fenómenos físicos son cuantificables en diferentes dominios de representación de la señal de voz, existiendo diferentes medidas objetivas de calidad que cuantifican con mayor o menor precisión un mismo fenómeno físico [8].

3.2. Cuantificación de la estabilidad de la voz

La estabilidad de la voz es la capacidad de un locutor de generar, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales.

Este fenómeno físico se cuantifica a partir de medidas de estabilidad de la señal de voz. Estas medidas cuantifican la variación de la intensidad de la señal de voz a lo largo del tiempo, durante la fonación de un sonido sonoro sostenido.

A continuación se presentan clasificadas en los diferentes dominios de representación las diferentes medidas de calidad de la voz englobadas en este grupo.

3.2.1. Dominio temporal

3.2.1.1. Perturbación de la amplitud: shimmer

La medida de la perturbación de la amplitud, también llamada *shimmer*, caracteriza la variación del valor del pico positivo máximo de cada trama de voz en el dominio temporal. Se determina mediante la variación de la amplitud máxima entre tramas adyacentes.

La perturbación de la amplitud, el *shimmer*, ha sido definida como:

$$AP = \left(\frac{1}{N-1}\right) \left(\frac{1}{|SA_{max}|}\right) \sum_{i=1}^{N-1} |SA(i+1) - SA(i)| \quad [\text{Ec. 3-7}]$$

donde *AP* es el valor de la perturbación de la amplitud a lo largo del tiempo, *SA* representa el valor de pico positivo máximo de cada trama de voz, siendo *i* el índice de cada trama y *N* es el número de tramas en las que se divide el audio que se analiza. Otros autores calculan el valor del *shimmer* midiendo la perturbación periodo a periodo.

Existen trabajos como, en los que se utiliza un estimador diferente para el *shimmer*. Este estimador se calcula de la siguiente manera:

$$AP = \left(\frac{100\%}{N-1}\right) \sum_{i=2}^{N-1} \left| \frac{SA(i) - SA(i-1)}{SA(i)} \right| \quad [\text{Ec. 3-8}]$$

De forma similar se ha definido una medida genérica denominada Perturbación Promedio Relativa (*Relative Average Perturbation - RAP*) que se define según:

$$RAP = \left(\frac{1}{N-2}\right) \sum_{i=1}^{N-2} \left| \frac{[P(i-1) + P(i) + P(i+1)]/3 - P(i)}{\frac{1}{N} \sum_{i=1}^N P(i)} \right| \quad [\text{Ec. 3-9}]$$

cuando esta medida se aplica al valor máximo *P(i)* de la *i*-ésima trama, el valor de la medida cuantifica el mismo fenómeno físico que el *shimmer*. Esta medida ha sido utilizada en trabajos como. En estos trabajos, al valor de la RAP aplicado al valor máximo de la trama se la ha denominado Cociente de Perturbación de la Amplitud (*Amplitude Perturbation Quotient - APQ*).

Con el mismo fin, y como mejora respecto a tomar el valor máximo de cada trama, en el dominio temporal se ha propuesto el uso del Valor Cuadrático Medio (VMC) para cuantificar la variación de la energía de la trama. El valor cuadrático medio se define para una trama de la siguiente forma:

$$VCM = \frac{1}{L} \sum_{i=1}^L (x(i) - \mu_x)^2 \quad [\text{Ec. 3-10}]$$

donde VCM es el valor cuadrático medio de una trama, $x(n)$ es la trama de voz la cual tiene una longitud de L muestras, y $x \mu$ es el valor medio de la trama.

Además, se ha definido la medida denominada intensidad que cuantifica el mismo fenómeno que el *shimmer*. Esta medida se estima como el valor de la energía expresado en dB, para ventanas de corta duración, recomendándose en el trabajo citado ventanas de 10 mseg.

Otra medida, denominada Frecuencia del Temblor de Amplitud (Fatr) y similar al *shimmer*, cuantifica la frecuencia con mayor intensidad dentro de las componentes de baja frecuencia de la señal moduladora de la modulación de amplitud presente en la señal de voz. Esta modulación es posible obtenerla por medio de un filtrado paso-banda alrededor del primer formante. La medida Fatr, se obtiene por medio de técnicas lineales. Existen estudios donde se realiza un procesado no lineal de esta modulación de amplitud con el fin de obtener unas medidas de la calidad de la voz.

El *shimmer* se ha estudiado también en diferentes dominios, como es el caso en el que se estima a partir de la señal obtenida tras aplicar a la señal de voz un filtrado inverso.

3.2.2. Dominio espectral

Existe una medida de la calidad de la voz basada en la cuantificación del peso de la Energía Espectral del Primer Armónico (*Normalized First Harmonic Energy* - NFHE), estimándose como el cociente entre la energía concentrada en el primer armónico frente a la energía concentrada en el segundo y sucesivos armónicos [9].

3.3. Cuantificación de la riqueza espectral

La riqueza espectral: se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora con gran riqueza de componentes armónica. Este fenómeno físico se cuantifica por medio del cálculo de la estabilidad de la frecuencia de *pitch* y por la presencia de armónicos con energía elevada en las diferentes bandas de frecuencias.

A continuación, al igual que en el capítulo anterior, se presentan las diferentes medidas de calidad de la voz pertenecientes al grupo de *riqueza espectral*.

3.3.1. Dominio espectral

3.3.1.1. Perturbación del pitch: jitter

Esta medida cuantifica la variación a lo largo del tiempo del valor del *pitch*. Se define la perturbación del valor de *pitch*, “*jitter*”, como:

$$PP = \left(\frac{1}{N-1}\right) \left(\frac{1}{|SP_{max}|}\right) \sum_{i=1}^{N-1} |SP(i+1) - SP(i)| \quad [\text{Ec. 3-11}]$$

donde *PP* es el valor de la perturbación del *pitch* a lo largo del tiempo, *SP* es el valor del *pitch* para cada una de la trama *i* y *N* es el número de tramas en las que se divide el audio que se analiza. De igual forma existen autores que estiman el valor de esta medida calculando la variación del valor del *pitch* periodo a periodo.

Existen diferentes definiciones de esta medida tal como la definida según la siguiente expresión:

$$PP = \left(\frac{100\%}{N-1}\right) \sum_{i=2}^N \left| \frac{SP(i) - SP(i-1)}{SP(i)} \right| \quad [\text{Ec. 3-12}]$$

o una versión modificada en la que se sustituye la sustracción del valor del *pitch* de la ventana anterior por el valor promediado de *pitch* de *K* ventanas, tal y como se describe en la siguiente expresión:

$$PP = \left(\frac{100\%}{N-K}\right) \sum_{i=\frac{K}{2}}^{N-\frac{K}{2}} \left| \frac{SP(i) - \frac{1}{k} \sum_{d=-\frac{k}{2}}^{\frac{k}{2}} SP(i+d)}{\frac{1}{k} \sum_{d=-\frac{k}{2}}^{\frac{k}{2}} SP(i+d)} \right| \quad [\text{Ec. 3-13}]$$

De forma similar se ha definido una medida genérica denominada Perturbación Promedio Relativa (*Relative Average Perturbation* - RAP) la cual se define según:

$$RAP = \left(\frac{1}{N-2}\right) \sum_{i=1}^{N-2} \left| \frac{[P(i-1) + P(i) + P(i+1)]/3 - P(i)}{\frac{1}{N} \sum_{i=1}^N P(i)} \right| \quad [\text{Ec. 3-14}]$$

Cuando esta medida se aplica a la frecuencia del *pitch* $P(i)$ de la i -ésima trama, el valor de la medida cuantifica el mismo fenómeno físico que el *jitter*. Al valor de la RAP aplicado a la frecuencia del *pitch* se la ha denominado cociente de la Perturbación del Periodo de Pitch (*Pitch Perturbation Quotient* -PPQ).

Este parámetro es ampliamente utilizado para la cuantificación de la calidad de la voz. Se ha estudiado en diferentes dominios, como en el caso en el que se estima el valor del *jitter* a partir de la señal obtenida tras aplicar un filtro inverso a la señal de voz.

Existen trabajos en los que a partir de la serie temporal que se obtiene al estimar la frecuencia de vibración de las cuerdas vocales para cada una de las diferentes tramas, se divide esta señal en dos: la señal correspondiente a las altas frecuencias denominada partir de la cual se calcula la medida denominada Variación del *Pitch* de Alta Frecuencia (HFPV) y la señal correspondiente a las bajas frecuencias a partir de la cual se estima la medida denominada temblor (*tremor*).

Otra medida, similar al *jitter*, es la que cuantifica la frecuencia con mayor intensidad dentro de las componentes de baja frecuencia de la señal moduladora de la modulación en frecuencia presente en la señal de voz. Esta medida se denomina Temblor de f_0 (Fftr). Esta modulación de la señal de voz es posible obtenerla por medio de un filtrado paso-banda alrededor del primer formante. El parámetro Fftr, se obtiene por medio de técnicas lineales.

3.3.1.2. Grado de Insonoridad

El Grado de Insonoridad (*Degree of Unvoiceness* - DUV) cuantifica la presencia de periodos de tiempo durante la fonación de un sonido sostenido en los que las cuerdas no vibran o vibran de forma muy irregular.

Se calcula como la relación entre el número de tramas sordas y el número total de tramas durante la fonación de una vocal sostenida. El DUV se calcula mediante la siguiente ecuación:

$$DUV = \frac{N_{tramas\ sordas}}{N_{tramas}} \quad [Ec. 3-15]$$

donde $N_{tramas\ sordas}$ es el número de tramas sordas y N_{tramas} es el número total de tramas.

3.3.1.3. Grado de Periodicidad de la Tramas Sordas

El Grado de Periodicidad de la Tramas Sordas (*Unvoiced Periodicity Ratio* - UPR) complementa la denominada grado de insonoridad (DUV). Cuantifica la presencia de periodos de tiempo durante la fonación de un sonido sostenido en los que las cuerdas no vibran o vibran de forma muy irregular. La UPR cuantifica la frecuencia de las ocurrencias de periodos de tiempo en los que ocurre el comportamiento irregular de las cuerdas vocales.

La medida UPR cuantifica si la producción de tramas sordas a lo largo de la fonación de un sonido sonoro se produce de forma consecutiva o de forma alternada. Esta medida, presentada en, cuantifica el número de transiciones entre “trama sorda – trama sonora” o bien entre “trama sonora – trama sorda” respecto al número de tramas totales.

$$UPR = \frac{\text{número de transiciones}}{N_{tramas}} \quad [\text{Ec. 3-16}]$$

3.3.1.4. Relación Subarmónico a Armónico

Existen trabajos en los que se indica la presencia de *subarmónicos* del *pitch*. La medida de calidad de la voz demonizada Relación Subarmónico a Armónico (*Subharmonic-to-Harmonic Ratio* – SHR) se define como:

$$SHR = \frac{SS}{SH} \quad [\text{Ec. 3-17}]$$

donde ambos términos se definen según las siguientes expresiones:

$$SH = \sum_{n=1}^N X(nf_0) \quad [\text{Ec. 3-18}]$$

$$SS = \sum_{n=1}^N X((n - \frac{1}{2})f_0) \quad [\text{Ec. 3-19}]$$

siendo X el espectro de la señal de voz, f_0 la frecuencia fundamental, nf_0 representa los diferentes armónicos teóricos y los $(n - 1)f_0$ representa a los posibles subarmónicos o los valles entre los armónicos teóricos. $N \cdot f_0$ identifica al mayor de los armónicos representables.

3.3.1.5. Grado de Sonoridad por Trama

La medida de la calidad de la voz denominada Grado de Sonoridad por Tramas (GST) está basada en el funcionamiento del codificador paramétrico de voz *Multiband Excitation Vocoder* (MBE). Esta medida cuantifica la riqueza espectral de cada trama. Para ello, identifica las componentes del espectro que pertenece al *pitch* y todos sus armónicos. Esto nos permite dividir el espectro de cada trama en diferentes regiones, cada una de ellas correspondientes a un armónico (Figura 3.11). Posteriormente se etiqueta cada una de las diferentes bandas en “banda armónica” o “banda no armónica” en función de la energía máxima de la banda que supera un determinado umbral. Este etiquetado permite calcular la medida GST, que se define como el porcentaje de bandas del espectro de una trama que son etiquetadas como armónicas con respecto al número total de bandas en la trama [9].

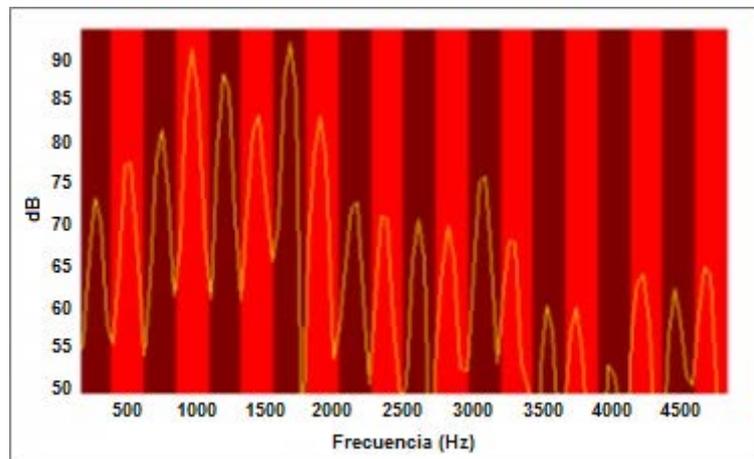


Figura 3.11.- Representación de las diferentes bandas en las que se divide el espectro de una trama de voz en el proceso de estimación del grado de sonoridad por trama

3.3.2. Dominio cepstral

3.3.2.1. Medidas basadas en el cepstrum

3.3.2.1.1. Energía Cepstral

Por medio de la medida Energía Cepstral (*Cepstral Energy* – CE) se estudia la información que proporcionan los *cepstrum* complejos acerca de la excitación de una voz sonora. Estudios realizados demuestran que la región de altas *quefreny*6 contiene información sobre el grado de periodicidad de la señal. Este grado de periodicidad se cuantifica por medio de la amplitud máxima del pico cepstral principal en el rango de las altas *quefreny*.

En la Figura 3.12 se muestra los *cepstrum* complejos de una trama de una muestra de voz de alta calidad (imagen superior) y una trama de voz de baja calidad (imagen inferior). Ambas muestras corresponden a locutores masculinos. En dicha figura se puede apreciar como el valor máximo de los *cepstrum* complejos se obtiene para la muestra de voz de alta calidad.

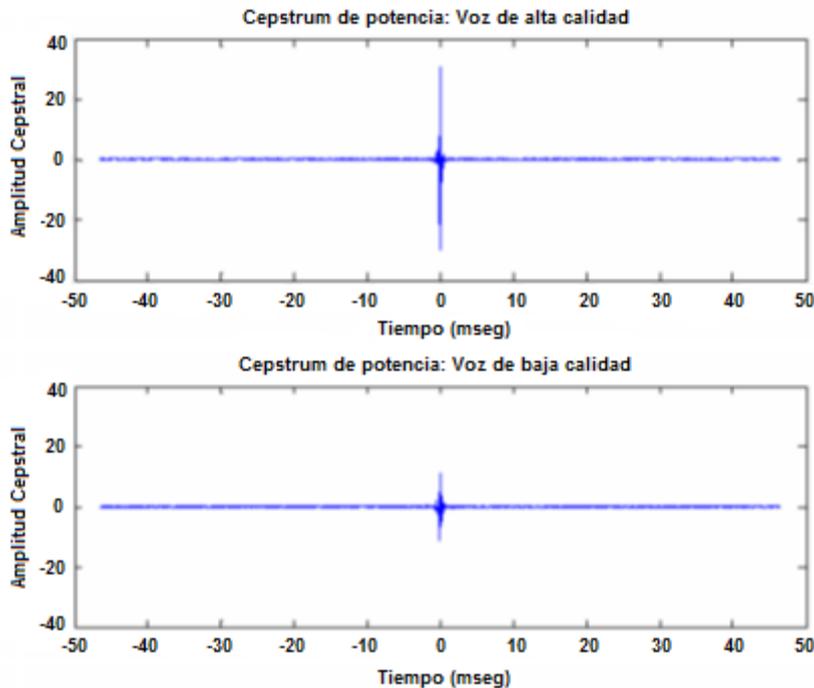


Figura 3.12.- Cepstrum complejos de una muestra de voz sana y de una muestra de voz patológica

La medida de Energía Cepstral se calcula mediante la siguiente expresión:

$$CE = \frac{1}{N_{tramas}} \sum_{n=1}^{N_{tramas}} PC_n \quad [Ec. 3-20]$$

donde PC_n es la amplitud máxima del pico cepstral principal de los *cepstrum* complejos de la trama n , y N_{tramas} el número total de tramas en las que se divide la señal de voz sonora que se pretende analizar.

3.3.2.1.2. Energía Cepstral del Pitch

La medida Energía Cepstral del Pitch (*Pitch Energy Cepstral Measure* -PECM) cuantifica la capacidad del locutor para hacer vibrar las cuerdas vocales a la misma frecuencia a lo largo de la fonación. Esta medida calcula dicha capacidad en cortos periodos de tiempo. Además, cuantifica la relación porcentual de ruido existente en la voz. Su cálculo se realiza en el dominio de los *cepstrum* de potencia.

En la Figura 3.13 se muestra una representación de los *cepstrum* de potencia para una muestra de voz de alta calidad y para una muestra de voz de baja calidad.

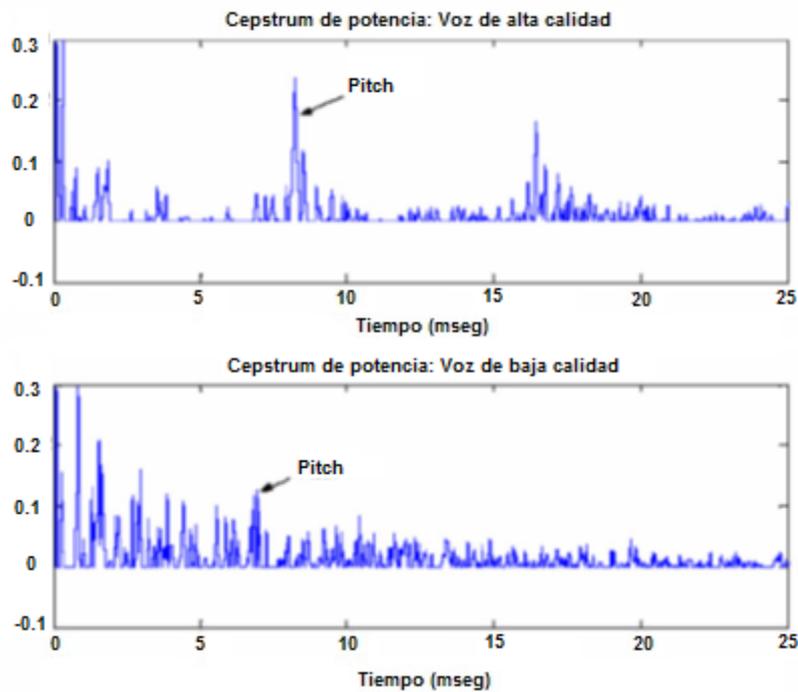


Figura 3.13.- Representación en el dominio de los *cepstrum* de potencia de una muestra de voz de voz de alta calidad (superior) y para una muestra de voz de baja calidad (inferior)

La medida de energía cepstral de *pitch* PECM evalúa la relación entre la energía *cepstral* concentrada en el pulso del *pitch* y la energía cepstral total de cada ventana en el contexto de los *cepstrum* de potencia rectificadas. Para localizar el pulso del *pitch* es necesario previamente haberlo localizado mediante un detector de *pitch*. El PECM se calcula para cada trama. La expresión del PECM es la siguiente:

$$PECM = \frac{E_{\text{pulso del pitch}}}{E_{\text{total de la trama}}} \quad [\text{Ec. 3-21}]$$

donde la energía cepstral del pulso del *pitch* se define como:

$$E_{\text{pulso del pitch}} = \sum_{n=n_{\text{inicio}}}^{n_{\text{fin}}} C(n) \quad [\text{Ec. 3-22}]$$

siendo: $c(n)$ los coeficientes cepstrales que forman el pulso del *pitch*, n_{inicio} es el número de la muestra después del último cruce por cero que se produce a la izquierda de la posición cepstral del *pitch*, y n_{fin} es el número de la muestra antes del primer cruce por cero que se produce a la derecha de la posición cepstral del *pitch*. La energía cepstral total definida como:

$$E_{total\ de\ la\ trama} = \sum_{n=0}^L C(n) \quad [Ec. 3-23]$$

donde L es el número total de coeficientes *cepstrum* de cada trama.

La medida útil para cuantificar la calidad de la voz es la perturbación a lo largo del tiempo de PCEM durante la fonación de un sonido sostenido. Esta medida no se puede estimar durante los fragmentos de tiempo en los que se produce una ausencia de vibración de las cuerdas vocales.

3.3.2.1.3 Energía Cepstral del Segundo Armónico

De forma similar al PCEM, la Energía Cepstral del Segundo Armónico (*Second Harmonic Energy Cepstral Measure - SHECM*), cuantifica la capacidad del locutor para producir de forma constante a lo largo del tiempo que dura la fonación de un sonido sonoro sostenido, un segundo armónico con estabilidad y con energía significativamente diferente respecto al componente de ruido de la señal. En la Figura 3.14 se aprecia la localización del primer armónico (correspondiente al *pitch*) y del segundo armónico en el dominio de los *cepstrum* de potencia.

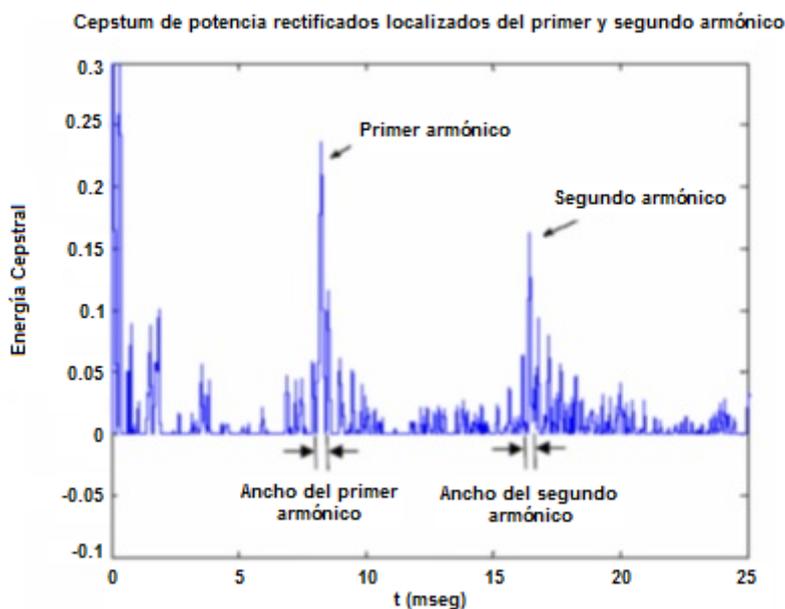


Figura 3.14.- Localización del primer y segundo armónico en el dominio de los cepstrum reales

La medida SHCEM se define de forma similar al PECM, con el hecho diferencial de que cuantifica la energía del pico cepstral centrado en el segundo armónico respecto a la energía total de la ventana de los *cepstrum*. Al igual que ocurre con la PECM, la perturbación a lo largo del tiempo de la medida SHCEM durante la fonación de un sonido sonoro sostenido, resulta útil para medir la calidad de la voz.

3.3.2.1.4. Cepstrum derivado espacialmente

Con el fin de cuantificar las manifestaciones del fenómeno de *riqueza espectral* que se obtienen de una representación cepstral de la señal de voz, se ha evaluado la derivada espacial de una ventana cepstral, estudiándose el comportamiento del primer armónico (*pitch*) y del segundo armónico en el dominio de la derivada espacial de los *cepstrum*. Para la obtención de la derivada espacial de los *cepstrum* se ha utilizado una aproximación a un diferenciador ideal mediante un filtro de fase lineal, el cual se define de la siguiente manera:

$$s(n) \equiv \frac{d}{dt} s(n) \approx \sum_{m=N_d}^{N_d} m \cdot s(n + m) \quad [\text{Ec. 3-24}]$$

donde es práctica habitual tomar $d N$ igual a 5 muestras, para una frecuencia de muestreo 22050 Hz.

El resultado obtenido en la transformación de una trama de voz al dominio de los *cepstrum* derivados espacialmente se muestra en la Figura 3.15, donde además se indica la localización de los dos primeros picos positivos, que corresponden a las pendientes de subida del primer armónico y del segundo armónico, siendo proporcionales a los respectivos anchos de los pulsos.

Al igual que sucedía con el espectro, es significativo los cambios que presentan tramas adyacentes de tres periodos de *pitch* de duración en una representación cepstral, principalmente en el caso de las voces de baja calidad.

Las medidas de calidad que cuantifican la riqueza espectral de la voz propuestas en una representación en el dominio de los *cepstrum* derivados espacialmente son el valor de pico de primer armónico, el valor de pico del segundo armónico y la diferencia de amplitudes entre el primer y el segundo armónico. Como medidas de calidad que cuantifican la riqueza espectral de la voz, resultan de interés tanto los valores que toman las magnitudes descritas, como la variación en el tiempo de los valores de dichas magnitudes.

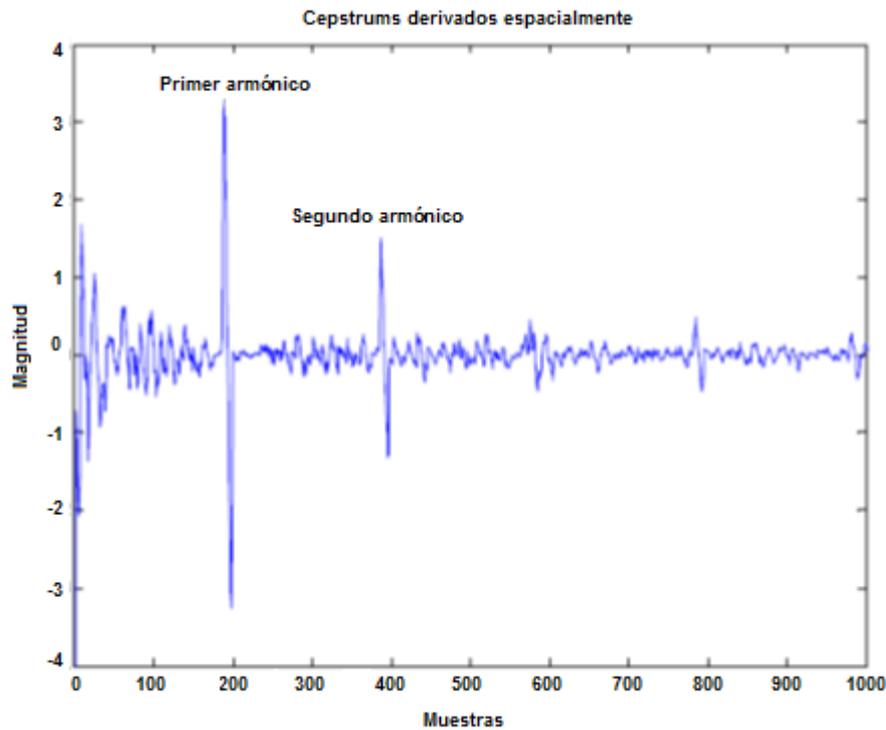


Figura 3.15.- Localización del primer y segundo armónico en el dominio de los cepstrum derivados

3.3.2.1.5. Relación Aperiódico-Periódico

La medida denominada Relación Aperiódico-Periódico (*Aperiodic Periodic Ratio* - APR) proporciona una medida de la relación de la parte armónica con respecto al ruido. En este caso se relaciona la energía cepstral del *pitch* y del segundo armónico con respecto a la energía procedente del tracto vocal, los coeficientes *cepstrales* inferiores al pico de *pitch* [9].

$$APR = \frac{\sum_{n=\text{inicio}_{pitch}}^{\text{fin}_{pitch}} c(n) + \sum_{n=\text{inicio}_{2^{\circ} \text{ armónico}}}^{\text{fin}_{2^{\circ} \text{ armónico}}} c(n)}{\sum_{n=1}^{\text{inicio}_{pitch}} c(n)} \quad [\text{Ec. 3-25}]$$

Capítulo 4

Metodología

4.1. Estudio. Propuesta metodológica.

En este presente trabajo se pretende realizar el análisis acústico (cuantitativo y cualitativo) de unas señales de voz para comprobar las mejoras en la producción vocal obtenidas tras aplicar la técnica o método “Pantalla de Celofán”; el cuál se describirá en el apartado siguiente. Se realizará el análisis de las señales de voz correspondientes a las vocales grabadas (/a/, /i/ y /o/) *antes* y *después* de aplicar el método.

Para el estudio se ha se ha dispuesto de un número de 20 locutores; los cuáles han sido partícipes en la realización de las grabaciones de audio. Para cada locutor o participante, se ha realizado la grabación de las vocales (/a/, /i/ y /o/) *antes* y *después* de aplicar el tratamiento o procedimiento “Pantalla de celofán”.

Se ha realizado la grabación de un repositorio de 6 muestras por cada locutor, consistente cada muestra en una grabación de una vocal sostenida (/a/, /i/ y /o/), tanto *antes* como *después* de aplicar el método correspondiente.

El análisis acústico es la exploración que cuantifica los distintos componentes de la emisión vocal. Estos sonidos (señal microfónica) se capturan mediante un micrófono, se digitalizan mediante un convertidor analógico/digital, y se analizan con un programa de análisis acústico. Con esta exploración se pueden obtener muchos parámetros, pero los más interesantes son que nos permiten valorar la evolución del paciente tras cualquier tratamiento. Entre los parámetros que pueden medirse tras realizar un análisis acústico se encuentran:

- *Frecuencia fundamental (F_0)*; que es el número de veces que las cuerdas vocales vibran en un segundo. Se mide en ciclos/segundo o hertzios (Hz). Valores normales: 125 Hz para el hombre, 250 Hz para la mujer, 350 Hz para el niño. La correlación física de la frecuencia es el tono. Su variación se relaciona con los cambios estructurales de la cuerda vocal: la F_0 aumenta al aumentar su rigidez, y disminuye al aumentar la masa de la cuerda vocal.
- *Jitter*: es una medida que refleja las variaciones involuntarias que de la frecuencia se producen ciclo a ciclo. A mayor valor, mayor sensación de ronquera en la voz.
- *Shimmer*: es una medida que refleja las variaciones involuntarias que de la intensidad se producen ciclo a ciclo. A mayor valor, mayor aspereza de la voz.
- *Intensidad*: refleja la amplitud o fuerza con la que le produce el tono. Se expresa en decibelios (dB). Depende de la vibración de las cuerdas vocales y de la presión subglótica del flujo de aire pulmonar. Su disminución puede deberse a: enfermedad pulmonar, defecto de cierre glótico o rigidez excesiva de los tejidos de la cuerda vocal.

El análisis acústico de los resultados de las diferentes medidas objetivas de calidad de la voz que cuantifican cada uno de los diferentes fenómenos físicos se verá en capítulos posteriores. Para cada medida de calidad de la voz se presenta una descripción y una valoración de su calidad para cuantificar el fenómeno físico que mide.

4.2. Técnica de voz con el método “Pantalla de Celofán”.

La propuesta metodológica para la reeducación de la voz del “Pantalla de celofán”, tanto en el habla como en el canto, tiene como objetivo proporcionar una herramienta sencilla y práctica a todas aquellas personas que necesiten mejorar su emisión vocal.

La propuesta metodológica para la reeducación de la voz “Pantalla de celofán” responde a una necesidad que se plantea habitualmente en la docencia de cómo hacer que los estudiantes de canto consigan discernir cuándo un sonido está bien emitido, entendiéndolo por estabilidad de la voz, riqueza espectral, riqueza de armónicos y facilidad de la emisión, sin tensiones.

Este trabajo propone una técnica de entrenamiento vocal y la información de cómo usarla. La técnica de voz llamada “Pantalla de celofán” sirve para que la práctica vocal no sólo sea por imitación y por vía auditiva, sino por sensaciones internas que puedan fijar a través de la utilización con la “Pantalla de celofán”. El objetivo es proporcionar una herramienta sencilla y práctica a todas aquellas personas que necesiten mejorar su emisión vocal.

El método “Pantalla de celofán” consiste en hacer vibrar un papel de celofán transparente de bajo gramaje, aproximadamente de 20 gr/m², colocado sobre los labios y sosteniéndolo con una o ambas manos (Figura 4.1).



Fig. 4.1.- Pantalla de celofán

El papel de celofán debe tener aproximadamente unas medidas de 20 cm. de ancho por 15 cm. de alto, tal que pueda ser sostenido con las manos sin inconveniente, procurando que los hombros se encuentren distendidos y los orificios nasales no estén taponados. El papel de celofán no debe estar rígido, de tal manera que cada persona elegirá el grado de tensión que necesita para hacerlo vibrar. Se da el caso de quienes necesitarán taparse prácticamente la boca para vibrar, y otros simplemente rozando delicadamente los labios. Algo importante es la colocación de los labios, dándose por experiencia el resultado más favorable cuando se coloca la boca en forma de <U>. (Figura 4.2).

Para comprobar la incidencia en el resultado sonoro con la utilización de la “Pantalla de celofán” se ha realizado una evaluación preliminar para ver las condiciones de partida y posteriormente se ha efectuado el registro de la voz.

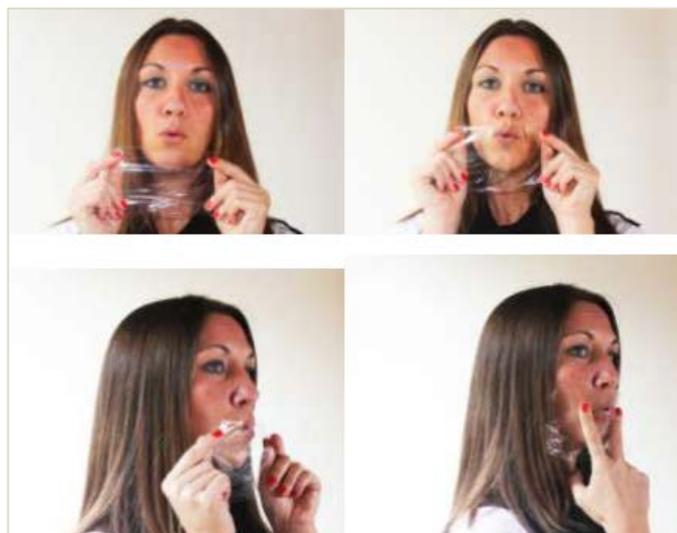


Fig. 4.2.- Uso de la pantalla de celofán

A cada uno de los locutores o participante se les ha explicado qué es la “Pantalla de celofán” y cómo usarla, también cuál va a ser la dinámica para tomar las muestras necesarias y los instrumentos de recogida de datos, así como el tiempo que se ha dedicado a la práctica.

Con anterioridad a la grabación de los registros de voz (*vocales /a/, /i/ y /o/*), se ha ensayado con sonidos sostenidos, arpegios y práctica de sonidos *onomatopéyicos*. Se ha educado a la participante en el uso, alternando con la “Pantalla de celofán” y sin ella. Y finalmente, se han realizado los registros de voz antes y después de aplicar éste método. Para el análisis acústico de los archivos de voz se ha utilizado el programa de libre distribución *SFS/ESection*.

Aunque el tiempo necesario que se empleará en cada sesión dependerá de las necesidades de cada individuo, en el caso que nos ocupa no ha superior a 30 minutos.

4.2.1. Fase previa: entrenamiento

Para hacerlo sonar se emite un sonido, haciendo vibrar el papel de celofán. Resulta más conveniente dar la nota <La4> o <Si4>, depende también del tipo de voz que sea, si es hombre, mujer o niño, o simplemente la nota en la que se encuentren más cómodo.

La punta de la lengua debe estar contra las encías o contra los dientes, y los bordes deben contactar con las encías o con las arcadas dentales superiores. La cara superior de la lengua puede ahuecarse más o menos hasta formar un canal mediante el enrollamiento lateral de sus bordes.

Una vez que se ha conseguido hacer vibrar el papel de celofán se harán ejercicios de vocalización (canto sin texto). Su finalidad es colocar el aparato fonador en las máximas condiciones de flexibilidad y funcionamiento para obtener una correcta emisión de la voz.

Se debe empezar con una vocal que más espontáneamente produzca las cualidades vocales de la persona, aunque se recomienda para este trabajo comenzar a vocalizar con la <U>, porque es más sencillo hacer vibrar el papel de celofán, facilita la emisión y sirve como referencia y punto de partida para las demás vocales.

Cuando hay dificultades en hacer vibrar el papel de celofán se pueden hacer varios ejercicios:

- Con una mano se sostiene el papel sobre los labios y con la otra mano se alterna el dedo índice y mediano dando golpecitos al papel a un ritmo rápido, y cantando la vocal <u>. También pueden colocar la palma de la mano de manera cóncava para aquellas personas que no les sale con los dedos alternados.
- Otra manera de hacerlo vibrar es pronunciar el vocablo <pr> y mantener el sonido onomatopéyico del motor de un coche <rrrrr>.
- Imitar el sonido de la sirena de una ambulancia también da muy buenos resultados.
- Para aquellas personas que tengan una voz débil resulta muy apropiado imitar a un perro con un ladrillo fino y chillón <gua gua gua> o también a una gallina (co co co co).

4.2.2. Fase final: La percepción

Primero se practicará haciendo sonidos onomatopéyicos, aquellos de los que hemos hablado anteriormente y que resulten más adecuados para el alumno. Una vez se haya conseguido hacer vibrar el papel con soltura, se comenzará a practicar con sonidos sostenidos sobre una nota. Se debe iniciar los ejercicios comenzando de notas agudas a graves, manteniendo la misma posición y la energía espiratoria, para pasar a ejercicios de extensión gradual hacia los extremos de la voz, y así con todas las vocales.

Una vez que el locutor ha conseguido hacerlo vibrar con desenvoltura y haya podido hacer los ejercicios correspondientes, se le pedirá que retire el papel manteniendo la misma posición y la misma sensación que tenía cuando lo hacía con el papel, de esta manera logrará una emisión fácil y correcta sin necesidad de buscar dónde debe colocar el sonido, se trata de hacer exactamente lo mismo que cuando se canta con el papel [10].

4.3. Procedimiento de grabación y segmentación de los archivos de audio

En el presente Trabajo Fin de Título se ha realizado la grabación de las vocales (/a/, /i/ y /o/) *antes* y *después* de aplicar el método “Pantalla de Celofán”, para cada uno de los 20 locutores o participantes, tratándose por tanto de 6 grabaciones o muestras por cada locutor como se ha mencionado anteriormente. Cada muestra ha consistido en una grabación de una vocal sostenida (/a/, /i/ y /o/), siendo las vocales /a/ y /o/ dos vocales fuertes o abiertas, y la vocal /i/ una vocal débil o cerrada.

La grabación de los archivos de audio y edición de los mismos se ha realizado con el software de libre comercialización *Audacity*. Las señales de audio de cada una de las vocales grabadas han tenido una duración en torno a los 10 segundos. Estas señales de audio de cada una de las vocales (/a/, /i/ y /o/) han sido posteriormente segmentadas en tramas de 2,5 segundos; esto se ha hecho para realizar el análisis acústico (cualitativo y cuantitativo) sobre un tramo de la señal donde el ruido es prácticamente nulo y donde la señal de voz es más limpia, siendo la voz en ese tramo más nítida en su reproducción.

4.4. Análisis de la señal de voz

Después de haber realizado el proceso de grabación y segmentado de cada uno de los archivos de audio para cada una de las vocales grabadas (/a/, /i/ y /o/) *antes* y *después* de aplicar el método, se ha procedido a realizar el correspondiente análisis acústico, teniendo por un lado el análisis cualitativo y por otro lado el análisis cuantitativo.

El análisis *cualitativo* ha consistido en la representación de la señal en el dominio temporal, en dominio frecuencial por medio del espectrograma de banda estrecha, la excitación glótica y una representación en el dominio cepstral. Para obtener las representaciones se ha utilizado el software de libre distribución *Esection* con el que se ha procedido a importar y procesar los archivos de audio.

EL análisis *cuantitativo* ha consistido en la medida de cuatro parámetros de medidas de calidad de la voz; cuyos valores óptimos se encuentran dentro de unos márgenes establecidos. Para realizar este tipo de análisis se ha utilizado un programa creado y editado con el software de libre distribución *Matlab* (herramienta de software matemático).

Capítulo 5

Base de Datos

5.1.- Descripción general de la base de datos

En la base de datos se recoge los archivos de audio correspondientes a las grabaciones de cada una de las vocales (/a/, /i/ y /o/) realizadas *antes* y *después* de aplicar el método “Pantalla de Celofán” a cada uno de los 20 locutores participantes que se han escogido para la realización de este estudio o análisis acústico.

Es de notar que cada una de las grabaciones se han guardado en formato WAV; el cuál es un formato de audio digital sin compresión de datos de mayor calidad y peso que el formato MP3; consiguiéndose así que la calidad en la señal de voz sea mayor y apreciándose de forma más notable la diferencia en el sonido a la hora de reproducir tanto una vocal antes de aplicar el método “Pantalla de Celofán”, cómo después de aplicar dicho método.

5.2. Balance de muestras de voz de la base de datos

En total se recogen en la base de datos 120 grabaciones o archivos de audio en formato WAV; donde las señales de audio de cada una de las vocales grabadas han tenido una duración en torno a los 10 segundos. Para cada uno de las 20 personas participantes en este estudio se han grabado 6 archivos de audio, correspondientes a la grabación de las vocales (/a/, /i/ y /o/) *antes* y *después* de aplicar el método “Pantalla de Celofán”.

Se ha rellenado un cuestionario por cada uno de los 20 locutores que han sido partícipes en las grabaciones de las vocales (/a/, /i/ y /o/) *antes* y *después* de aplicar el tratamiento o procedimiento “Pantalla de celofán”. Los datos que se han tomado son referidos al sexo, a la edad y la fecha de grabación. En la Tabla 5.1 se recoge la información correspondiente a cada locutor:

Locutor	Sexo	Edad	Fecha de Grabación
1	Mujer	50 en adelante	20/03/2014
2	Mujer	50 en adelante	21/03/2014
3	Hombre	entre 35 y 50	21/03/2014
4	Mujer	50 en adelante	27/03/2014
5	Hombre	50 en adelante	28/03/2014
6	Mujer	50 en adelante	29/03/2014
7	Mujer	50 en adelante	31/03/2014
8	Mujer	entre 25 y 35	03/04/2014
9	Hombre	entre 25 y 35	03/04/2014
10	Mujer	50 en adelante	04/04/2014
11	Hombre	entre 25 y 35	11/04/2014
12	Mujer	50 en adelante	11/04/2014
13	Mujer	50 en adelante	22/04/2014
14	Mujer	50 en adelante	25/04/2014
15	Mujer	50 en adelante	25/04/2014
16	Hombre	entre 35 y 50	25/04/2014
17	Mujer	entre 35 y 50	26/04/2014
18	Mujer	entre 35 y 50	04/05/2014
19	Hombre	50 en adelante	04/05/2014
20	Mujer	entre 35 y 50	04/05/2014

Tabla 5.1.- Datos de los locutores participantes

En la Tabla 5.2 se muestra el número de locutores participantes separados por su rango de edad y sexo, denotándose que la mayor parte de los participantes corresponden al género femenino.

Rango de edad	Sexo	
	Hombres	Mujeres
Entre 25 y 35	2	1
Entre 35 y 50	2	10
De 50 en adelante	2	3
Total participantes por sexo:	6	14
Total participantes:	20	

Tabla 5.2.- Número de participantes desagregados por sexo

Como se describió anteriormente en apartados anteriores, las señales de audio de cada una de las vocales (/a/, /i/ y /o/) han sido posteriormente segmentadas en tramas de 2,5 segundos; esto se ha hecho para realizar el análisis acústico (cualitativo y cuantitativo) más minucioso sobre una señal de voz limpia y sin ruido, siendo el balance de muestras de voz de la base de datos la expuesta a continuación en la Tabla 5.3.

Grupo de Muestras de Voces	Muestras en Hombres	Muestras en Mujeres	Total Muestras
Muestras de vocales (/a/, /i/ y /o/) ANTES de aplicar el método “Pantalla de Celofán”	18	42	
Muestras de vocales (/a/, /i/ y /o/) DESPUÉS de aplicar el método “Pantalla de Celofán”	18	42	
Muestras de vocales (/a/, /i/ y /o/)	36	84	120

Tabla 5.3.- Balance de muestras de voz de la base de datos

5.3. Equipo utilizado en las grabaciones de los archivos de audio

Para realizar el análisis acústico de cada señal de voz, se ha realizado previamente la grabación de cada archivo de audio para cada vocal (“o”, “a” e “i”). Las grabaciones han presentado las siguientes características:

- La grabación consiste en la fonación de forma sostenida de la vocal “o”, “a” e “i”, durante varios segundos (en torno a 5 segundos). Ha de ser una fonación no susurrada.
- La grabación de audio se realiza con un PC de propósito general equipado con una tarjeta de sonido convencional. La grabación se realiza con una frecuencia de muestreo de 44.100 muestras por segundo y una resolución de 32 bits por muestra. (forma predeterminada que ofrece el software Audacity).

- Se utiliza un micrófono, con al menos un ancho de banda sin distorsión de 11 KHz, pudiéndose elegir un micrófono convencional de sobremesa. Se recomienda que el micrófono esté a una distancia de 15 centímetros del locutor.

Para la grabación de los archivos de audio se ha tenido en cuenta los siguientes parámetros en el software *Audacity* (Ver Tabla 5.4).

Canal:	Monofónico – (1)	
Formato de audio:	WAV (audio digital sin compresión de datos)	
Muestreo:	Frecuencia de muestro predefinida:	44.100 Hz
	Formato de muestra predeterminado:	32-bit float

Tabla 5.4.- Parámetros utilizados en las grabaciones

5.4. Captura de las señales de audio

A continuación se muestran a modo de ejemplo, algunas capturas de las señales de audio correspondientes a las grabaciones de las vocales “o”, “a” e “i” realizadas en las diferentes sesiones de grabación, *antes* y *después* de aplicar el método correspondiente de reeducación de la voz obtenidas con el software *Audacity* para el caso de diferentes locutores participantes. Estas capturas corresponden a las señales de audio en el *dominio temporal*.

Las señales de audio de cada una de las vocales grabadas han tenido una duración en torno a 10 segundos. Estas señales de audio de cada una de las vocales (“o”, “a” e “i”) como se ha descrito anteriormente, han sido posteriormente segmentadas en tramas de 2,5 segundos cada una de ellas; para así quedarse con un tramo de señal de voz sin ruidos ni silencios, y así llevar a cabo posteriormente el análisis acústico pertinente.

El análisis acústico cualitativo (representación en diferentes dominios) y cuantitativo (medidas de parámetros de calidad de la voz), de las señales de voz para las vocales grabadas (“o”, “a” e “i”) se realizará en el capítulo siguiente.

❖ Grabación de las vocales “o”, “a” e “i” para un locutor tipo con el Audacity:

Descripción: Voz de una señora de 50 años en adelante que presenta en su emisión una señal fluctuante y bronca al principio, y que empieza a estabilizarse posteriormente.

- Vocales “o”, “a” e “i” grabadas *antes* de aplicar el método “Pantalla de celofán”.

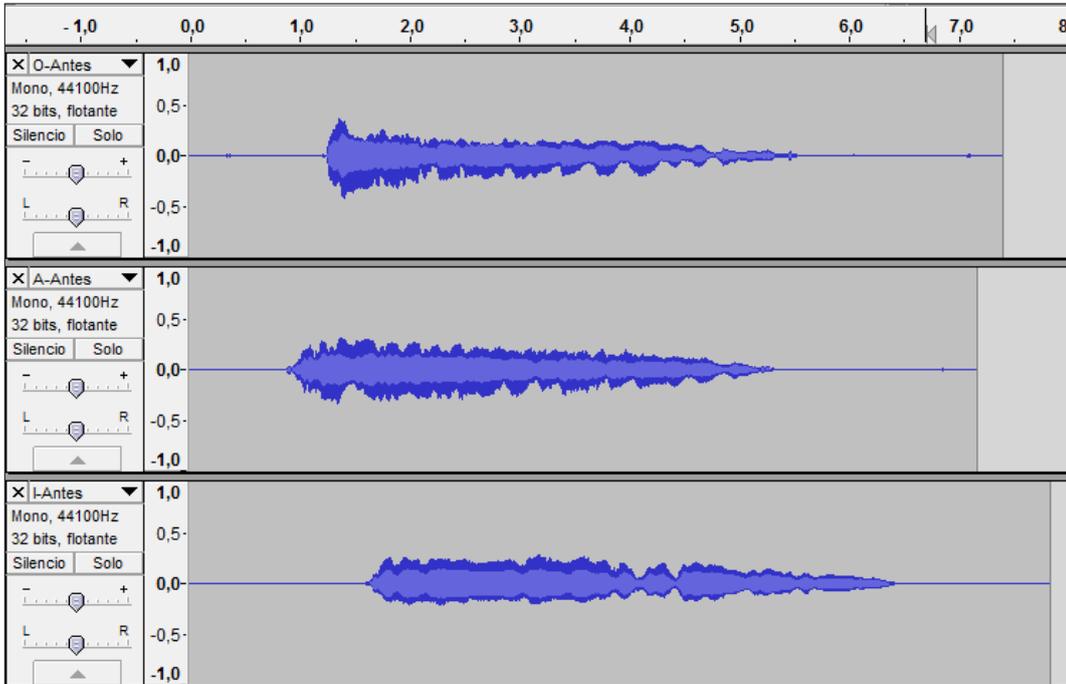


Fig. 5.1.- Vocales “o”, “a” e ”i” *antes* de aplicar el método

- Vocales “o”, “a” e “i” grabadas *después* de aplicar el método “Pantalla de celofán”.

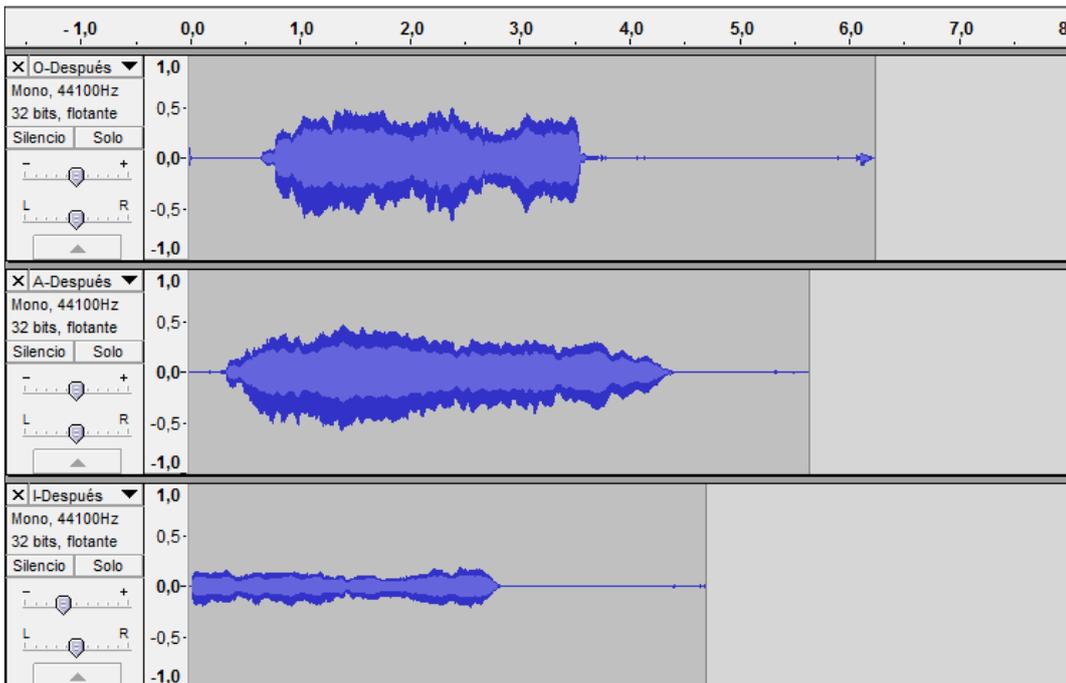


Fig. 5.2.- Vocales “o”, “a” e ”i” *después* de aplicar el método

- Vocales “o”, “a” e “i” segmentadas *antes* de aplicar el método “Pantalla de celofán”.

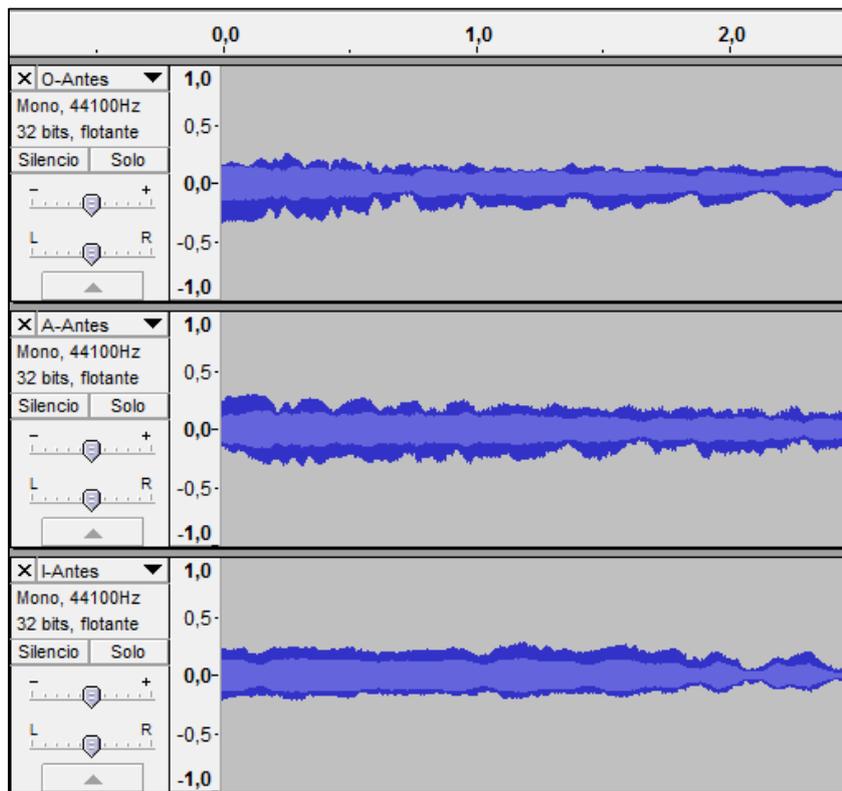


Fig. 5.3.- Vocales “o”, “a” e ”i” segmentadas *antes* de aplicar el método

- Vocales “o”, “a” e “i” segmentadas *después* de aplicar el método “Pantalla de celofán”.

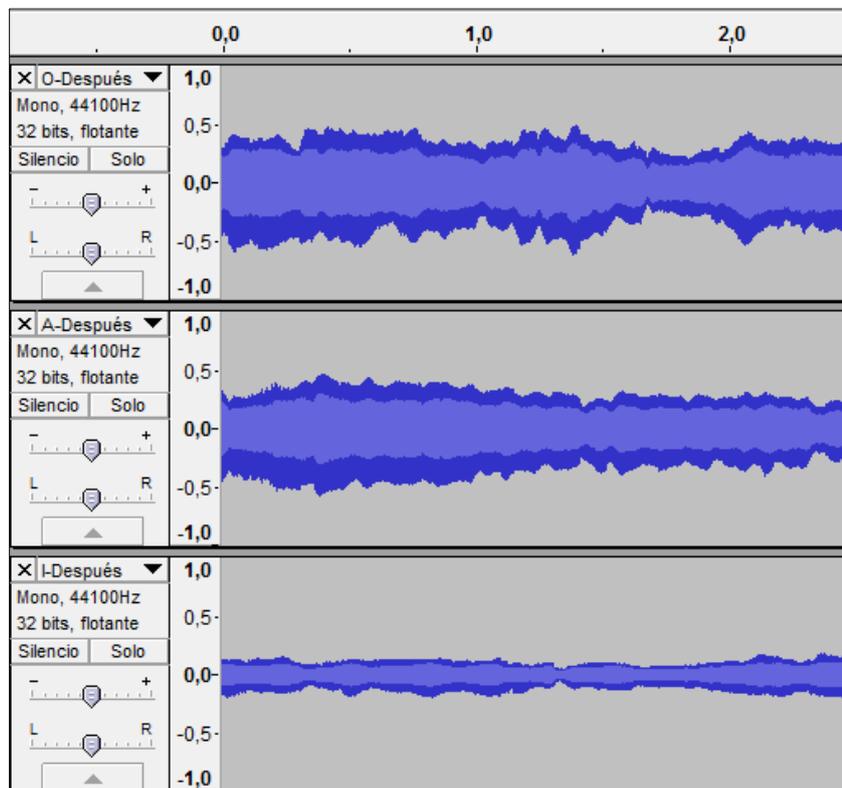


Fig. 5.4.- Vocales “o”, “a” e ”i” segmentadas *después* de aplicar el método

❖ Grabación de las vocales “o”, “a” e “i” para un locutor tipo con el Audacity:

Descripción: Voz de un varón, de entre 35 y 50 años de edad, con aparentes signos de ronquera o afonía y de fatiga vocal.

- Vocales “o”, “a” e “i” grabadas *antes* de aplicar el método “Pantalla de celofán”.

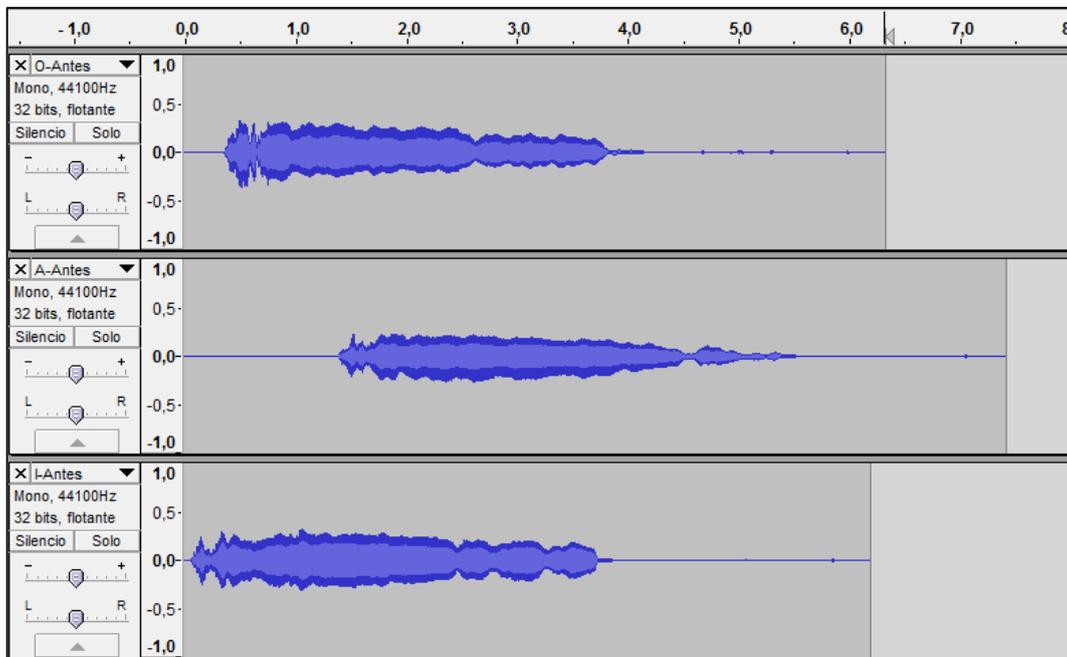


Fig. 5.5.- Vocales “o”, “a” e ”i” *antes* de aplicar el método

- Vocales “o”, “a” e “i” grabadas *después* de aplicar el método “Pantalla de celofán”.

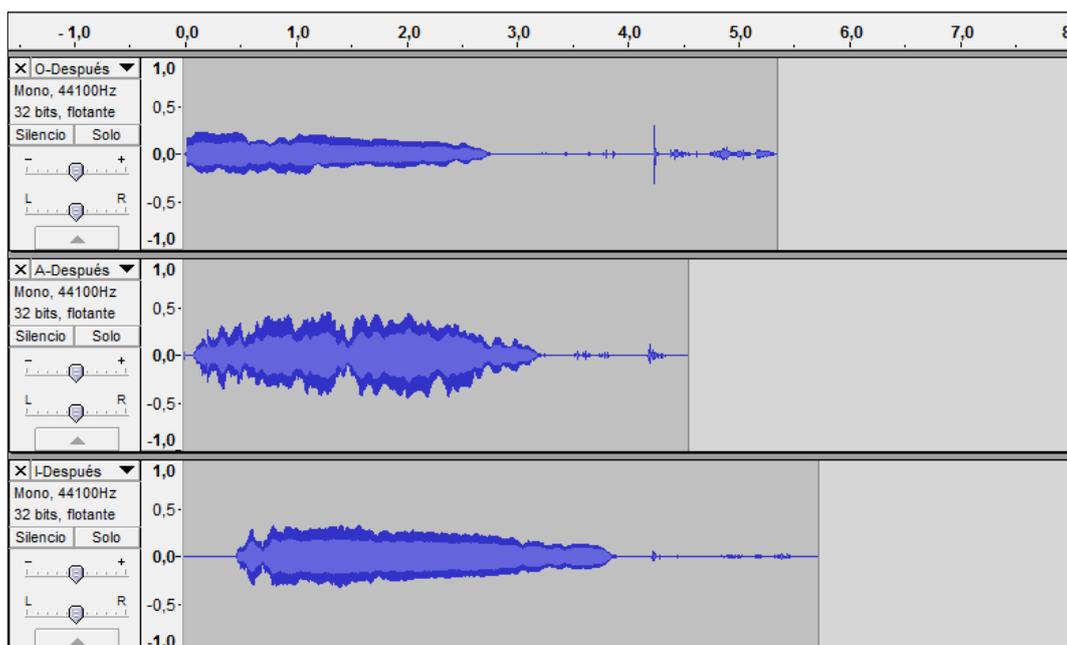


Fig. 5.6.- Vocales “o”, “a” e ”i” *después* de aplicar el método

- Vocales “o”, “a” e “i” segmentadas *antes* de aplicar el método “Pantalla de celofán”.

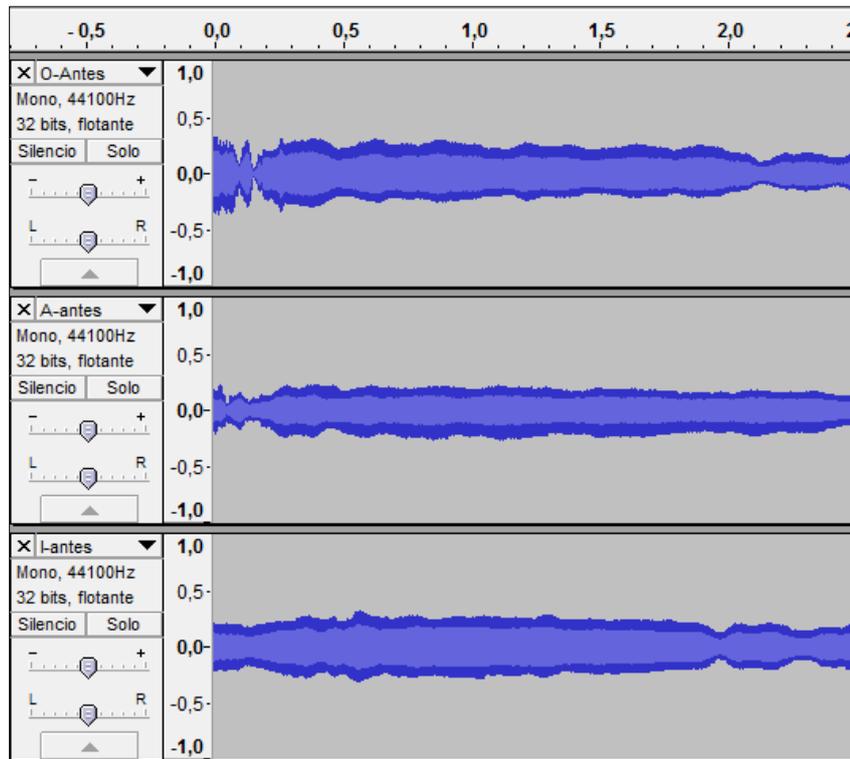


Fig. 5.7.- Vocales “o”, “a” e ”i” segmentadas *antes* de aplicar el método

- Vocales “o”, “a” e “i” segmentadas *después* de aplicar el método “Pantalla de celofán”.

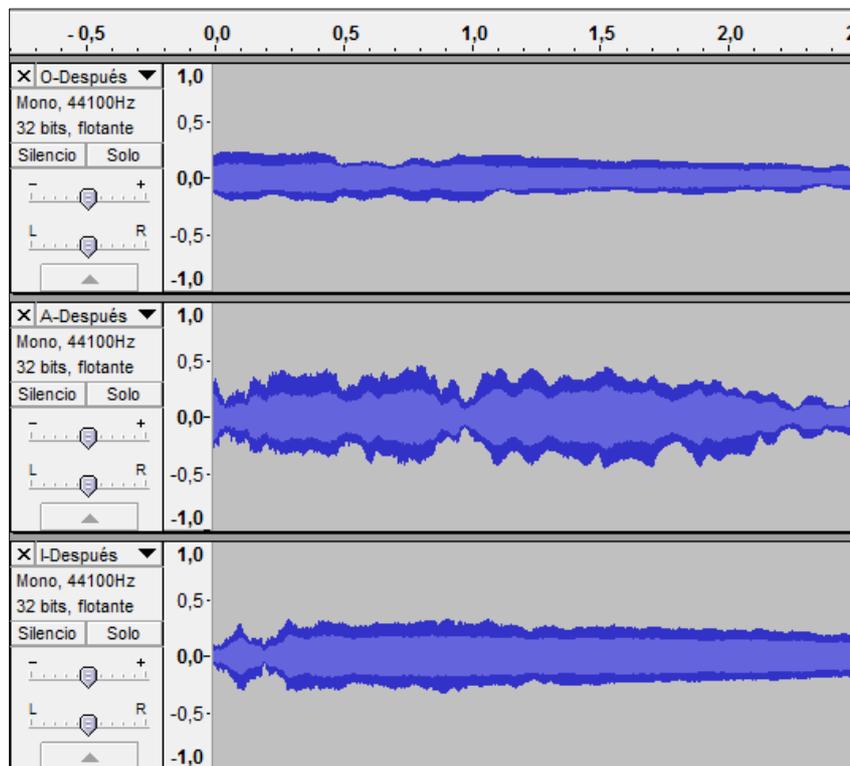


Fig. 5.8.- Vocales “o”, “a” e ”i” segmentadas *después* de aplicar el método

❖ Grabación de las vocales “o”, “a” e “i” para un locutor tipo con el Audacity:

Descripción: Voz de un varón, de entre 25 y 35 años de edad, que presenta signos de debilidad e irregularidad en la voz.

- Vocales “o”, “a” e “i” grabadas *antes* de aplicar el método “Pantalla de celofán”.

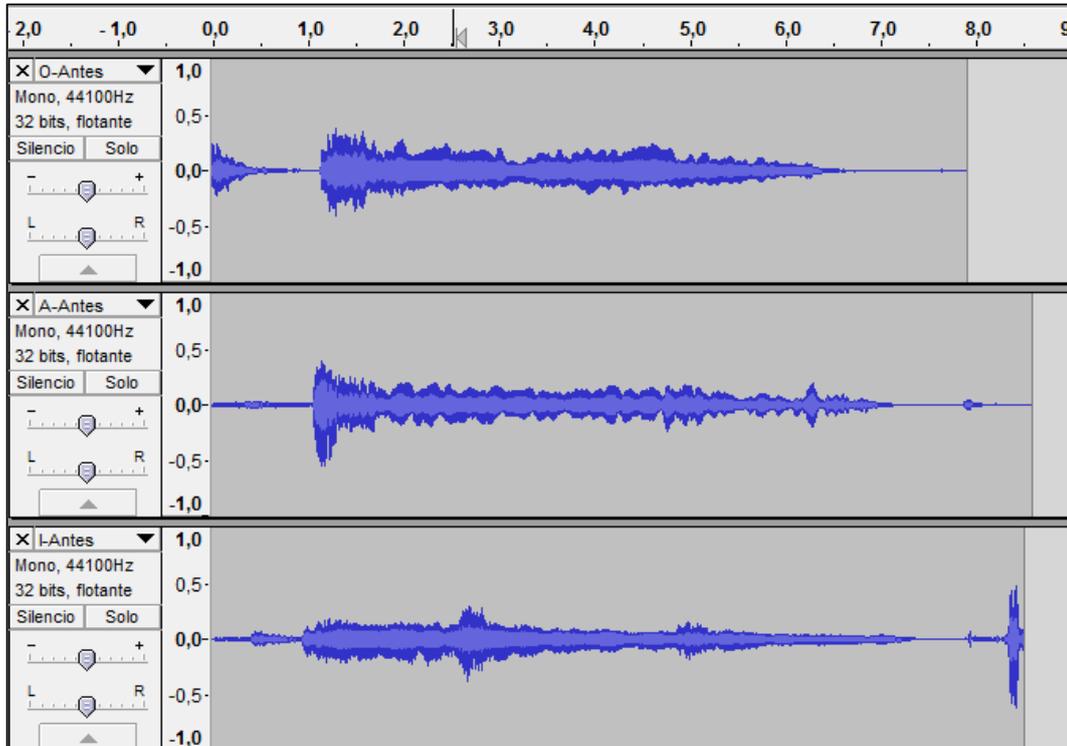


Fig. 5.9.- Vocales “o”, “a” e “i” *antes* de aplicar el método

- Vocales “o”, “a” e “i” grabadas *después* de aplicar el método “Pantalla de celofán”.

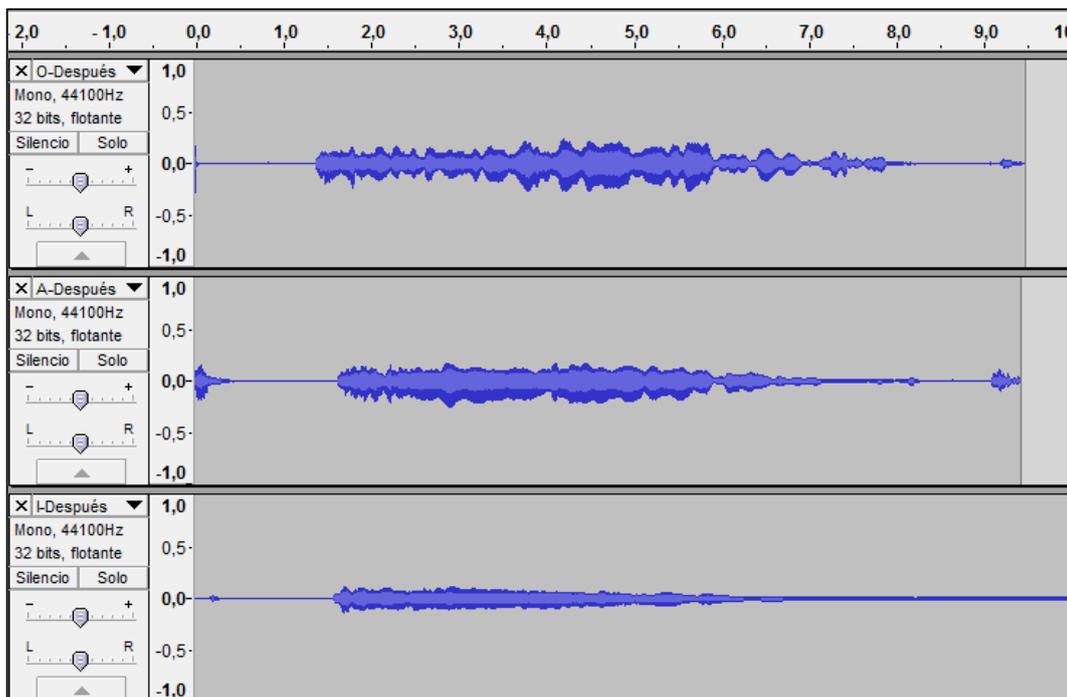


Fig. 5.10.- Vocales “o”, “a” e “i” *después* de aplicar el método

- Vocales “o”, “a” e “i” segmentadas *antes* de aplicar el método “Pantalla de celofán”.

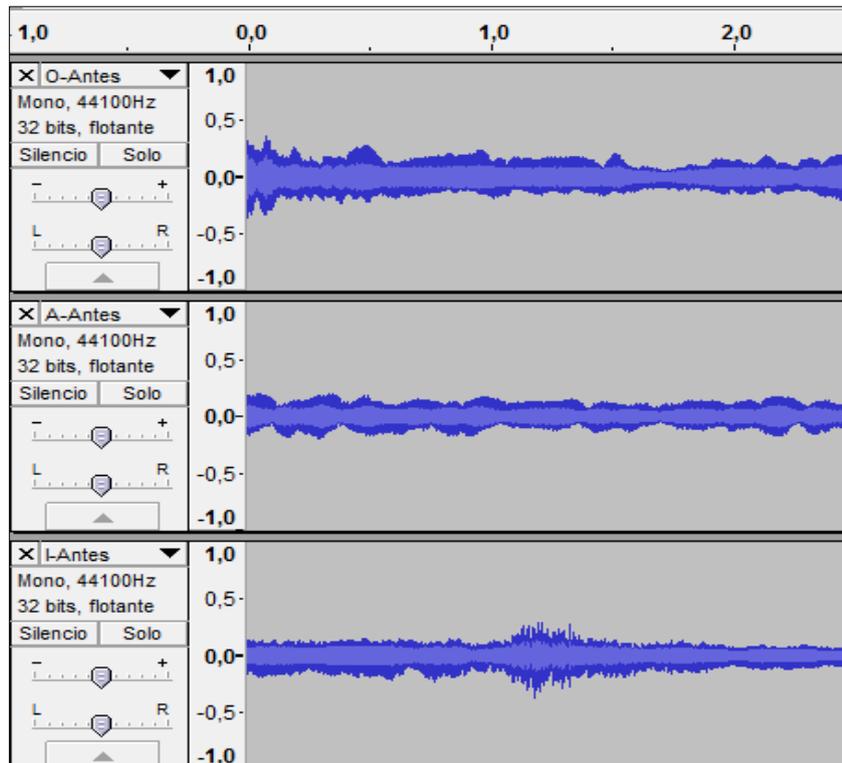


Fig. 5.11.- Vocales “o”, “a” e “i” segmentadas *antes* de aplicar el método

- Vocales “o”, “a” e “i” segmentadas *después* de aplicar el método “Pantalla de celofán”.

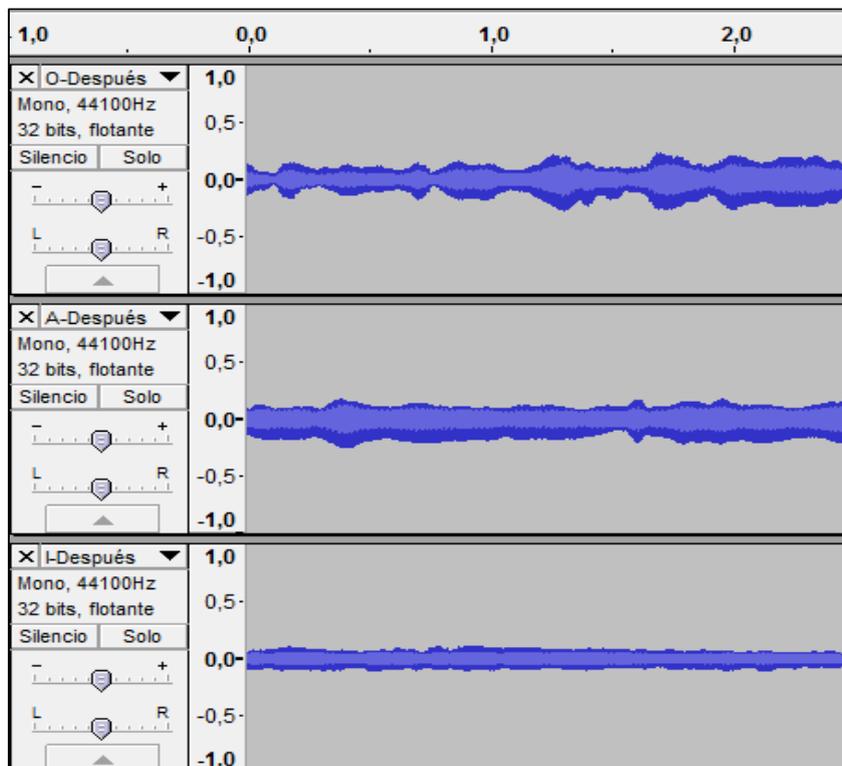


Fig. 5.12.- Vocales “o”, “a” e “i” segmentadas *después* de aplicar el método

Capítulo 6

Análisis de los Resultados

6.1.- Análisis cualitativo. Gráficas

En este análisis acústico cualitativo se puede comprobar que existe mejoría en la mayoría de las sesiones donde las vocales (“o”, “a” e “i”) fueron grabadas *después* de aplicar el método correspondiente, con respecto a la sesiones donde las vocales fueron grabadas antes de aplicar el método “Pantalla de celofán”.

Para constatar la mejoría producida, a continuación se realizará la comparación de las vocales (“o”, “a” e “i”) grabadas antes aplicar el tratamiento con las vocales grabadas después de aplicar el procedimiento (“o”, “a” e “i”), para cada vocal por separado. A continuación se expone el estudio de 3 casos diferentes, correspondiendo cada caso con un tipo de locutor distinto, apreciándose en estos casos la mejoría en la producción vocal de forma significativa. Para cada caso se ha analizado la grabación de la vocal correspondiente /o/, /a/ e /i/ de forma sostenida tanto antes como después del tratamiento.

El análisis ha consistido en la representación de la señal en el dominio temporal, del espectrograma de banda estrecha, la excitación glótica y una representación en el dominio cepstral. Para obtener las representaciones ha sido utilizado el software de libre distribución *ESection*.

Las gráficas correspondientes a cada una de las vocales /o/, /a/ e /i/ que han sido grabadas *antes* y *después* de aplicar el método “Pantalla de celofán” para cada uno de los 20 locutores, se recogen en el **Anexo** de este Trabajo.

❖ **Caso 1: Análisis de voz para las vocales “o”, “a” e “i” correspondiente al locutor nº 1.**

Descripción: Voz de una señora de 50 años en adelante que presenta en su emisión una señal fluctuante y bronca al principio, y que empieza a estabilizarse posteriormente.

✓ **Análisis de la vocal “O” correspondiente al locutor nº 1.**

➤ **Vocal “o” antes:**

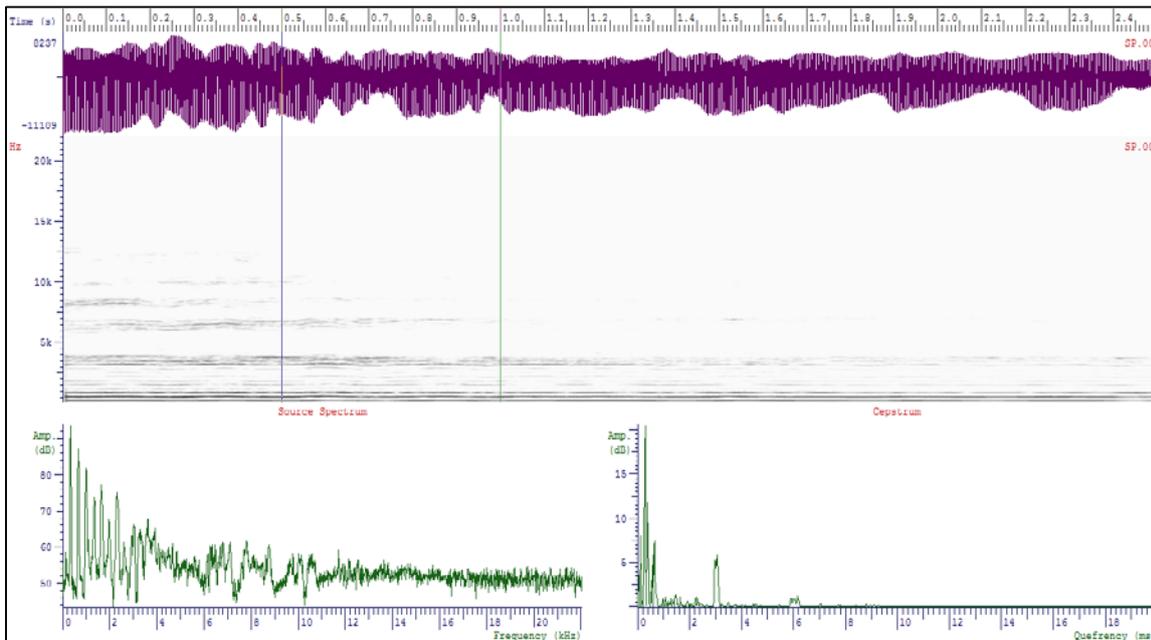


Fig. 6.1.- Vocal “o” antes de aplicar el método

➤ **Vocal “o” después:**

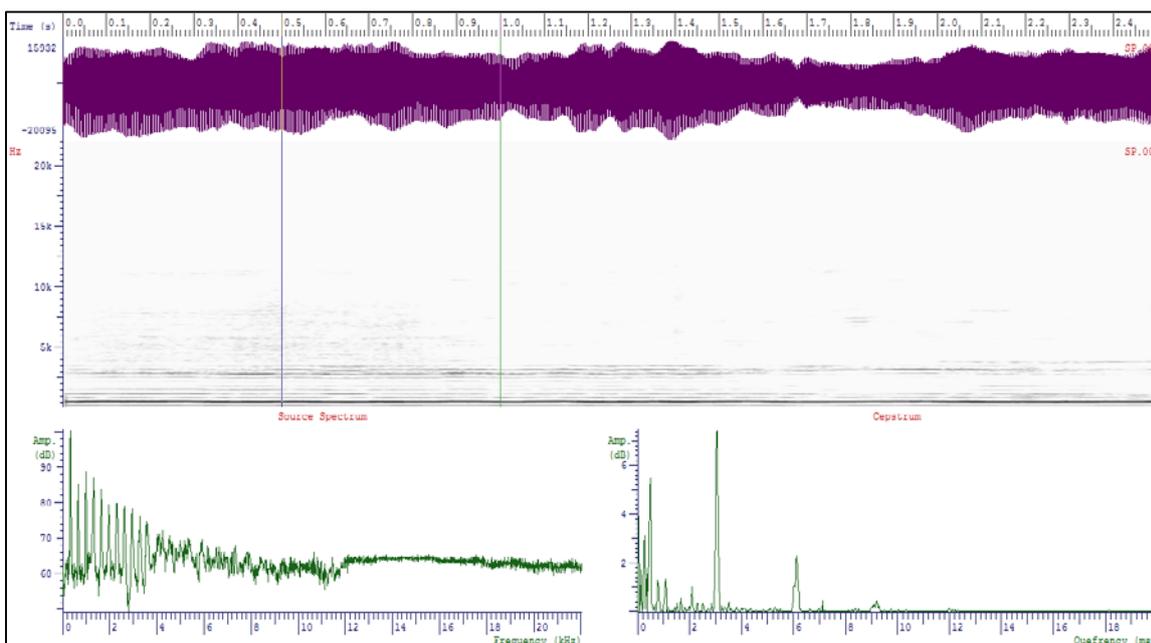


Fig. 6.2.- Vocal “o” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “o”, se aprecia en la parte superior de cada imagen (*antes* y *después* de aplicar el procedimiento) la señal correspondiente representada en el dominio temporal. Se trata de una voz oscilante y áspera al principio, y que empieza a estabilizarse posteriormente *después* de aplicar el método.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

A continuación, debajo de la señal en el dominio del tiempo se observa la señal en el dominio frecuencial por medio del espectrograma de banda estrecha. Para ello, representa la potencia de las diferentes componentes espectrales por medio de la intensidad de color utilizada en la representación, utilizando los ejes para representar las frecuencias y el tiempo

Se puede comprobar que en la **vocal “o” antes** de aplicar el método apenas se puede apreciar alguna banda de frecuencia en concreto donde los formantes estén más definidos. Además se comprueba que existe ruido en frecuencias comprendidas a frecuencias altas, **superiores a 3KHz** (Ver Figura 6.3).

Sin embargo, en la **vocal “o” después** de aplicar el método se aprecia un número significativos de armónicos en torno a los **3 KHz** siendo está la banda de frecuencia donde se concentra la mayor parte de la energía sonora del sonido y la presencia de ruido es prácticamente escasa (Ver Figura 6.3).

Para ello, se ha partido de que un formante es el pico de intensidad en el espectro de un sonido; se trata de concentración de energía (amplitud de onda) que se da en una determinada frecuencia. Técnicamente los formantes son bandas de frecuencia donde se concentra la mayor parte de la energía sonora de un sonido.

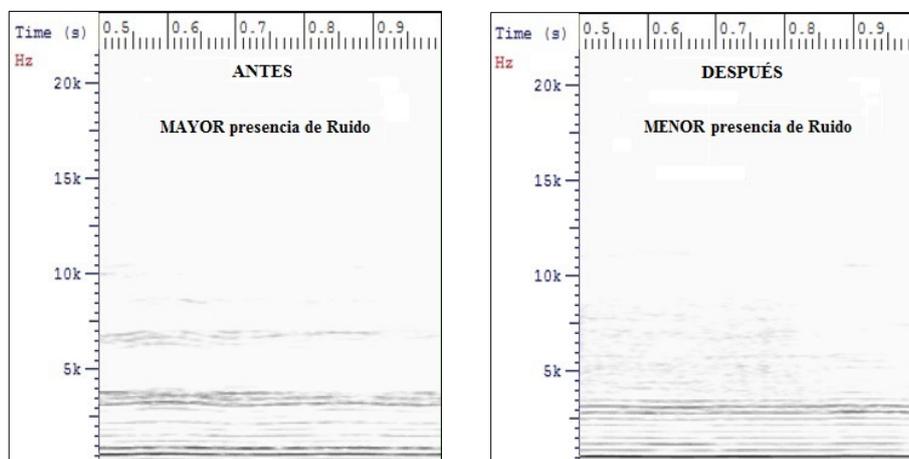


Figura 6.3.- Zoom aumentado de la señal en el dominio frecuencial (espectrograma) para la vocal “o” antes (izquierda) y después (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la parte inferior izquierda de cada imagen se observa una *representación del espectro de la señal de salida* de la vocal correspondiente.

En el capítulo 3 se vio que frente a las voces de alta calidad, en las muestras de voz de un sonido sonoro sostenido producidas por un locutor con voz de baja calidad se presentan las siguientes características diferenciales:

- Menor regularidad de la envolvente del espectro, principalmente en las bajas frecuencias.
- Mayor porcentaje de energía en las bajas frecuencias respecto a la energía total.
- Aumento de energía en las altas frecuencias, que es debida a la presencia de ruido glótico.
- Gran variación del espectro de una trama con respecto a las tramas contiguas.

Se puede constatar que en la *vocal “o” antes* de aplicar el método, el número de armónico está mucho menos definido, hay menor regularidad de la envolvente del espectro, y por tanto se aprecia una gran variación del espectro de una trama con respecto a las tramas contiguas. Así, se pueda observar que la posición de los formantes para el sonido sonoro sostenido de la *vocal “o” antes* de aplicar el método *empieza a denotar una significativa variación del espectro a partir de los 3 KHz* (Ver Figura 6.4).

Por otro lado, en la *vocal “o” después* de aplicar el método se aprecia que la variación del espectro de una trama con respecto a las tramas contiguas no se produce aproximadamente *hasta los 4 KHz*, siendo la señal hasta dicha frecuencia menos variable y con los armónicos más definidos, *habiendo mayor regularidad de la envolvente del espectro* (Ver Figura 6.4).

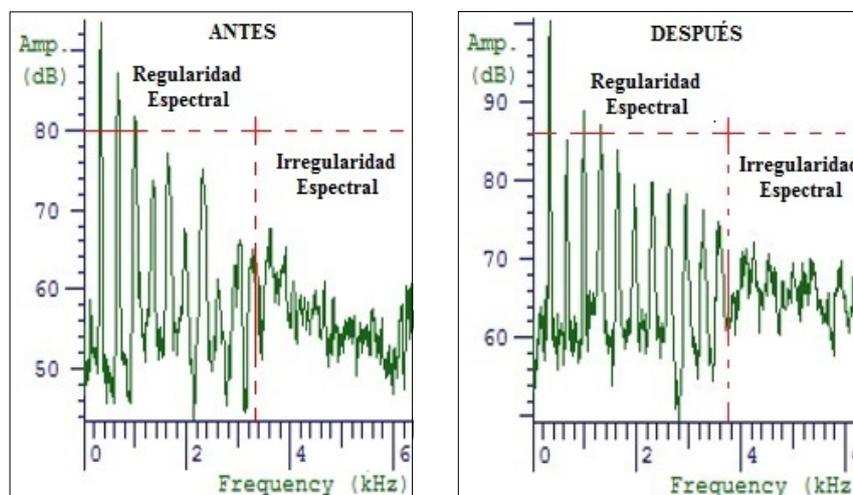


Figura 6.4.- Zoom aumentado de la señal en el dominio frecuencial (excitación glótica) para la vocal “o” antes (izquierda) y después (derecha) de aplicar el método

➤ **Dominio Cepstral: Cepstrum de potencia**

En la parte inferior derecha de cada imagen se observa una representación en *el dominio cepstral* de la señal de salida de la vocal correspondiente.

Como se describió en el capítulo 3, por medio de una representación *cepstral* es posible obtener información sobre la envolvente del espectro, la riqueza espectral e identificar las componentes armónicas y las componentes de ruido. Por medio de una representación cepstral es posible identificar características que permiten valorar la calidad de la voz.

En el caso de una fonación de un sonido sonoro sostenido, realizando una representación cepstral de diferentes tramas con duración equivalente a tres periodos de *pitch*, pueden identificarse las siguientes características:

La riqueza espectral de una muestra de voz puede cuantificarse por medio de la amplitud y anchura de la componente cepstral correspondiente al *pitch*. La existencia de un pico de una amplitud considerable indica la presencia notable de energía en dicha componente armónica, siendo ésta una característica de las voces de alta calidad. A estos picos claramente identificados se les denomina *armónicos*.

También, una anchura reducida del pico cepstral, correspondiente al *pitch*, para una trama de tres periodos de *pitch*, indica la alta estabilidad de la frecuencia de *pitch* para tres periodos consecutivos, también siendo ésta una característica de las voces de alta calidad.

La presencia de ruido en diferentes regiones del *cepstrum* puede cuantificarse por medio de la relación entre: la componente armónica (las componentes cepstrales correspondientes al *pitch* y sus armónicos), y la componente de ruido (el resto de las componentes cepstrales).

En la *vocal “o”* antes de aplicar la “Pantalla de celofán” se puede ubicar el primer armónico (el *pitch*) en torno a **3 KHz**. La amplitud del pico del *pitch* es de aproximadamente **5 dB** y no hay presencia de armónicos secundarios. Además, se observa la presencia significativa de cantidades de ruido a frecuencias próximas a la del *pitch*, y sobre todo a frecuencias bajas (Ver Figura 6.5).

En la *vocal “o” después* de aplicar la “Pantalla de celofán” el primer armónico (el *pitch*) se localiza igualmente en torno a **3 KHz**, localizándose próximo a él un armónico secundario en torno a **6 KHz**. La amplitud del pico del *pitch* es en este caso *superior a 7 dB* y el del armónico secundario de aproximadamente **2 dB**. Por tanto, se encuentran dos picos de una amplitud considerable; indica la presencia notable de energía en dicha componente armónica, siendo ésta una característica de una voz de alta calidad. Además, la presencia de cantidades de ruido sobre todo a frecuencias bajas es mucho menor en este caso (Ver Figura 6.5).

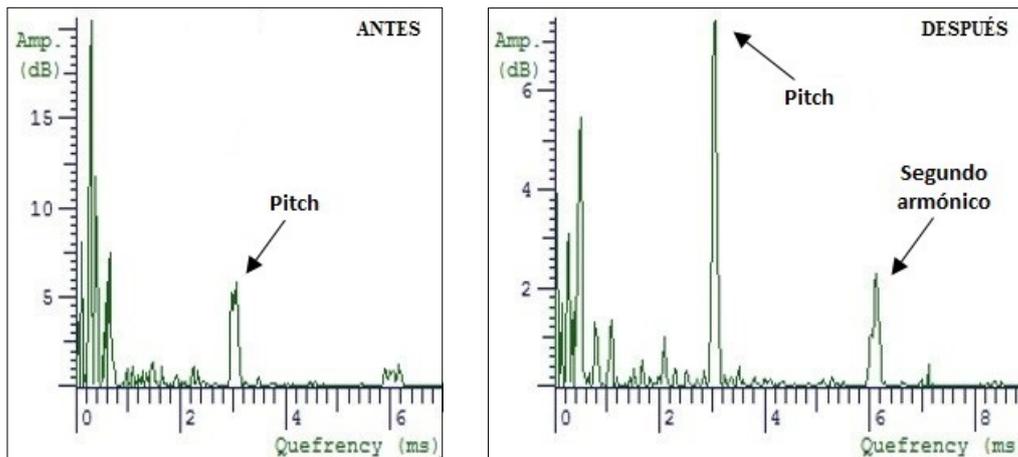


Figura 6.5.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “o” *antes* (izquierda) y *después* (derecha) de aplicar el método

✓ **Análisis de la vocal “A” correspondiente al locutor nº 1.**

➤ **Vocal “a” antes:**

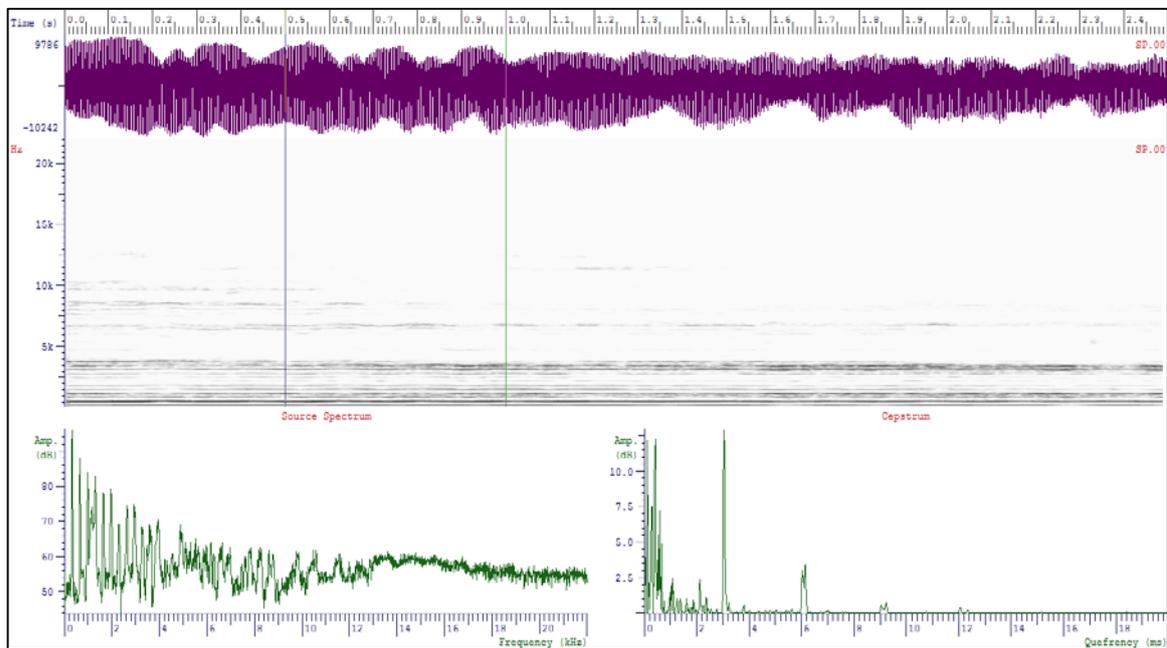


Fig. 6.6.- Vocal “a” antes de aplicar el método

➤ **Vocal “a” después:**

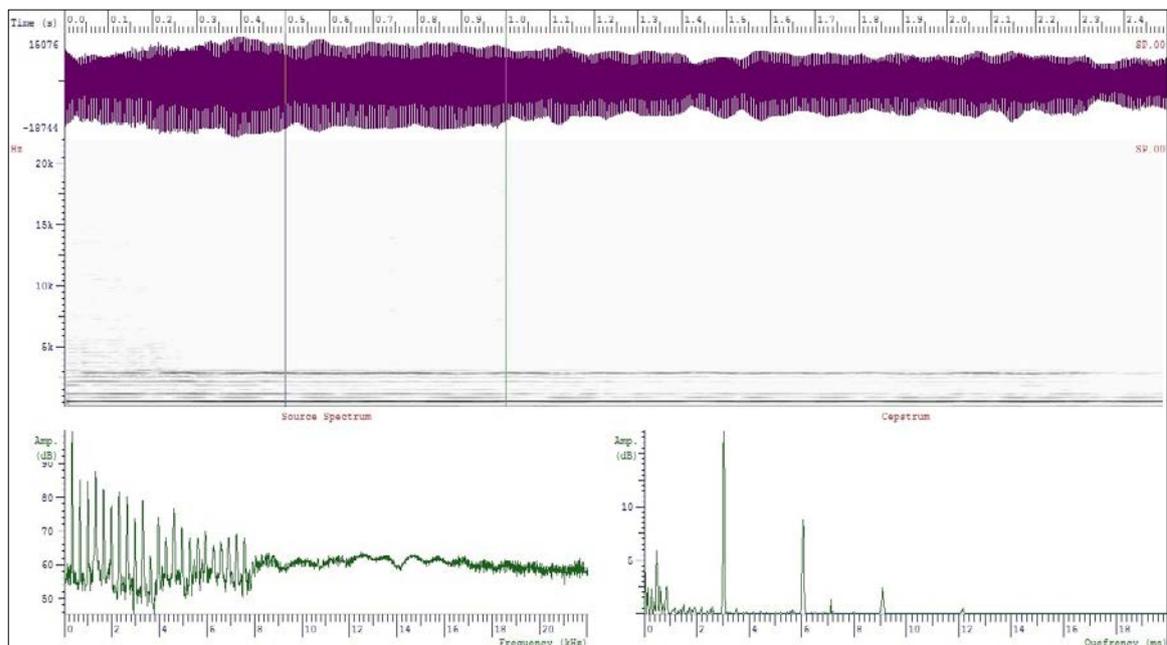


Fig. 6.7.- Vocal “a” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “a”, se obtiene una señal de voz *antes* de aplicar el método con pocas fluctuaciones; siendo esta señal más abrupta al principio; y siendo al final de su recorrido más endeble. *Después* de aplicar el método empieza a estabilizarse y tener una frecuencia parcialmente homogénea hasta el final, siendo esta señal más constante en su recorrido.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

En la *vocal “a” antes* de aplicar el método apenas se comprueba que existe bastante ruido a frecuencias bajas y hasta los **5 KHz**, siendo hasta dicha frecuencia donde se denota la presencia de cantidades anormales de ruido y por tanto, hay menores componentes armónicas (Ver Figura 6.8).

En la *vocal “a” después* de aplicar el método, se observa que la presencia de ruido es mucho menor a bajas frecuencias y se empieza a apreciar la banda donde los formantes están más definidos en torno a los **3,5 KHz** (Ver Figura 6.8).

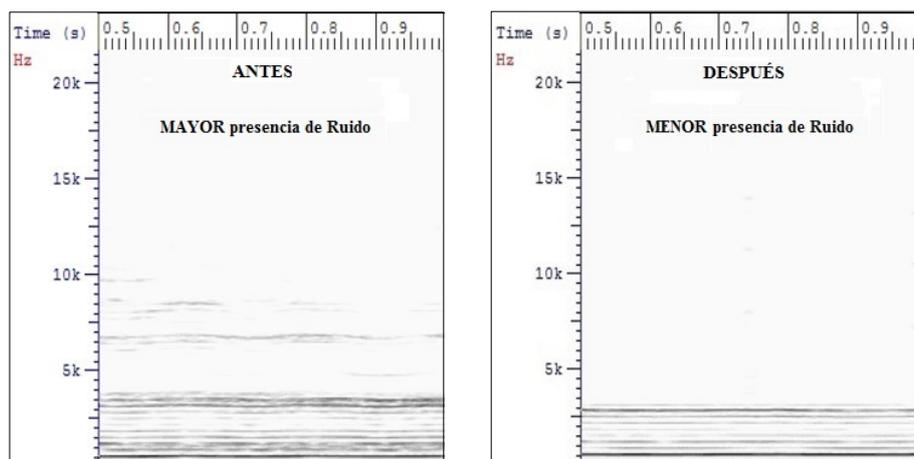


Figura 6.8.- Zoom aumentado de la señal en el dominio *frecuencial* (espectrograma) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la *vocal “a” antes* de aplicar el método, hay menor regularidad de la envolvente del espectro y los armónicos están mucho menos definidos a partir de los **4 KHz** *aproximadamente* (Ver Figura 6.9).

En la *vocal “a” después* de aplicar el método se aprecia que la variación del espectro de una trama con respecto a las tramas contiguas no se produce *aproximadamente hasta los 8 KHz*, *habiendo mayor regularidad de la envolvente del espectro* hasta dicha frecuencia y siendo la señal menos variable y con los armónicos más definidos (Ver Figura 6.9).

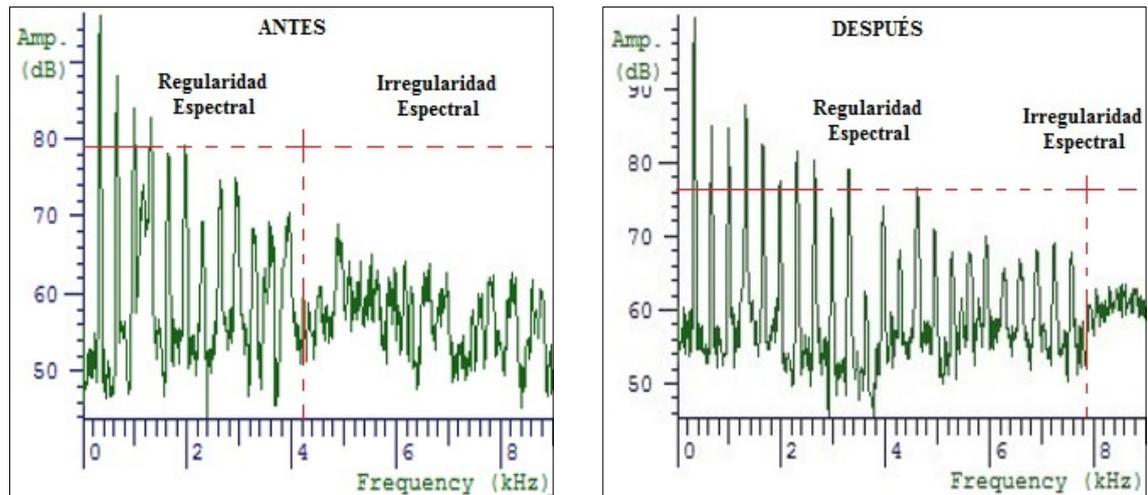


Figura 6.9.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Cepstral: Cepstrum de potencia**

En la *vocal “a” antes* de aplicar el método “Pantalla de celofán” se puede localizar el primer armónico (el *pitch*) en torno a **3 KHz**. La amplitud del pico del *pitch* es de aproximadamente **12,5 dB**, encontrándose próximo a este *pitch* un segundo armónico en torno a los **6 KHz** con una amplitud aproximada de **3 dB**. En este caso además, se observa que la presencia de ruido a frecuencias próximas e inferiores a la del *pitch* es considerable (Ver Figura 6.10).

En la *vocal “a” después* de aplicar el método, el primer armónico (el *pitch*) se localiza igualmente en torno a **3 KHz**, localizándose próximo a él igualmente un armónico secundario en torno a **6 KHz**. La amplitud del pico del *pitch* es en este caso **de 17 dB** aproximadamente (mayor presencia de energía en dicha componente armónica) y el del armónico secundario de aproximadamente **7,5 dB**. Además, nótese un tercer armónico presente, en torno a los **9 KHz** con una amplitud aproximada de **2 dB** (Ver Figura 6.10).

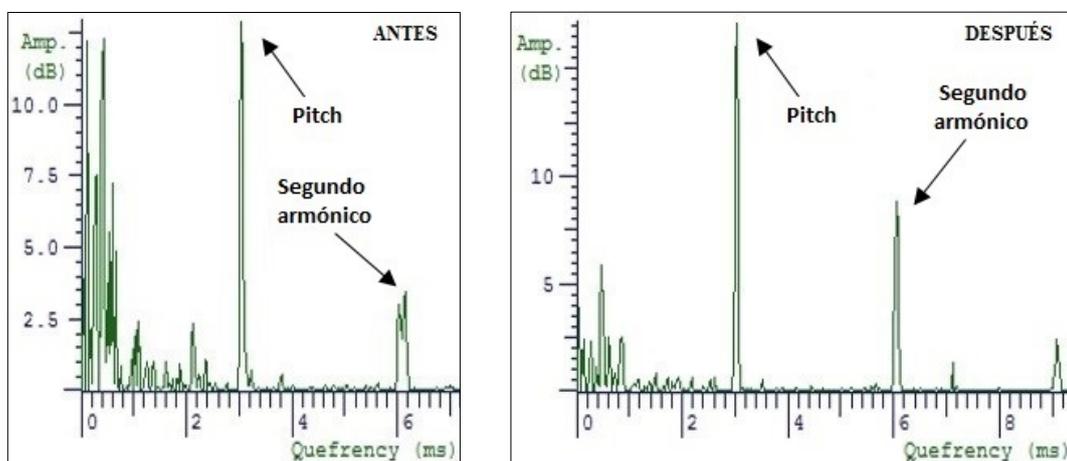


Figura 6.10.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

✓ **Análisis de la vocal “I” correspondiente al locutor nº 1.**

➤ **Vocal “i” antes:**

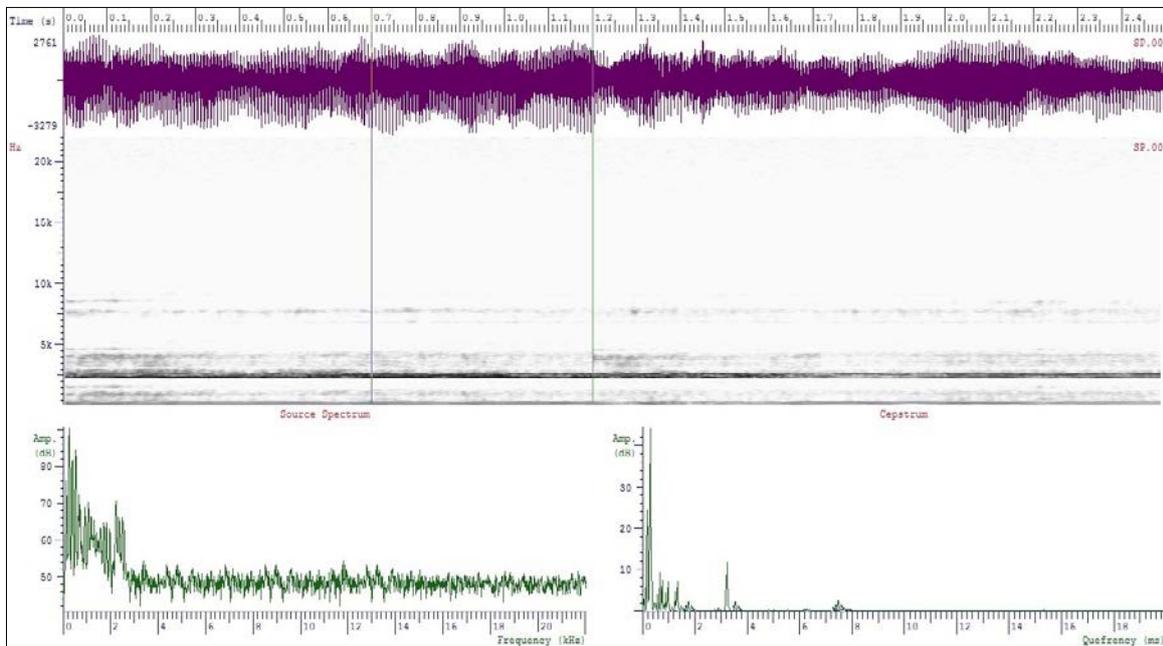


Fig. 6.11.- Vocal “i” antes de aplicar el método

➤ **Vocal “i” después:**

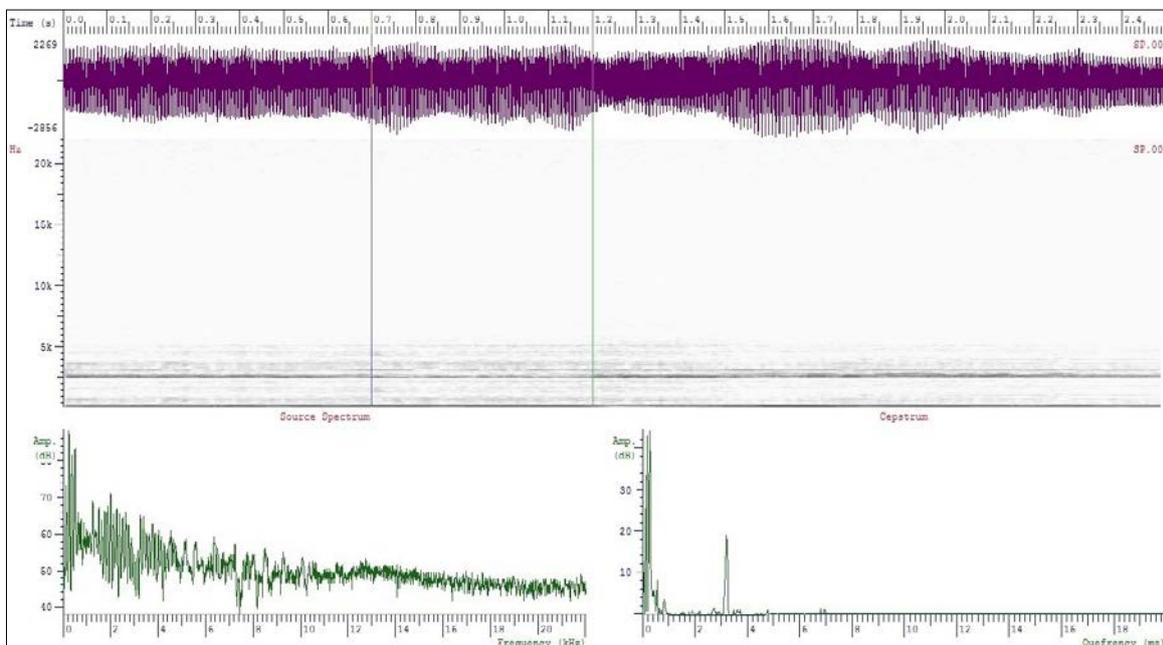


Fig. 6.12.- Vocal “i” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “i”, se obtiene una señal de voz inestable; donde la envolvente de la señal moduladora no se encuentra bien definida *antes* de aplicar el método, y que posteriormente *después* de aplicar el método se convierte en una señal de voz más homogénea, con menos ruido y donde aprecia de forma más significativa su envolvente.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

En la *vocal “i” antes* de aplicar el método apenas se comprueba que existen importantes cantidades anormales de ruido a frecuencias bajas inferiores a **5 KHz**. Además, se observa que se aprecie una banda significativa donde hay mayor energía concentrada en torno a los **3 KHz**, pero estando ésta acompañada de cantidades de ruido a frecuencias muy próximas. A frecuencias superiores a **5KHz** también se denotan ciertas cantidades de ruido (Ver Figura 6.13).

En la *vocal “i” después* de aplicar el método se observa que cantidad de ruido es mucho menor a las mismas frecuencias que antes y que el formante o banda de frecuencia más definida donde se concentra la mayor parte de energía sonora, estaría en torno a los **3 KHz** (Ver Figura 6.13).

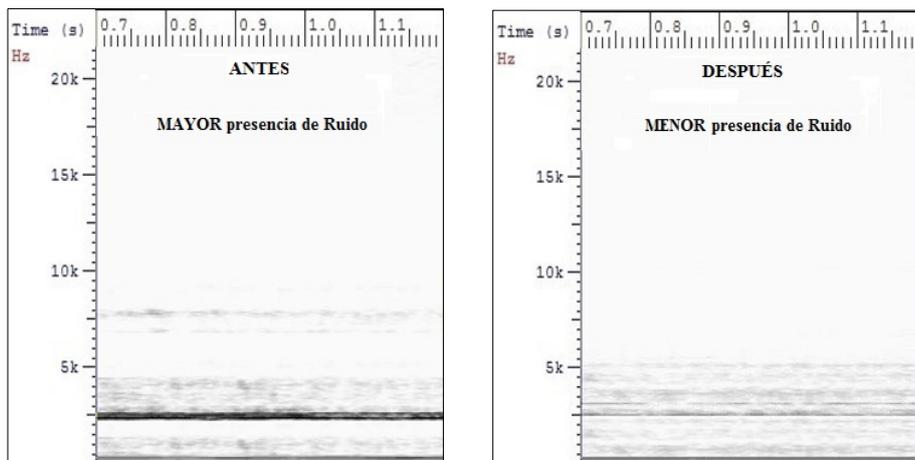


Figura 6.13.- Zoom aumentado de la señal en el dominio *frecuencial* (espectrograma) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la *vocal “i” antes* de aplicar el método, la envolvente del espectro es menos regular y que existe poca variación del espectro de una trama con respecto a las tramas contiguas hasta los **3 KHz aproximadamente**, siendo los armónicos hasta dicha frecuencia más definidos. (Ver Figura 6.14).

En la *vocal “i” después* de aplicar el método, la regularidad de la envolvente del espectro dura aproximadamente **hasta los 4,5 KHz**. A partir de dicha frecuencia la forma de onda se estrecha y por tanto a aumentar la energía, esto es debido a la presencia de ruido glótico. Por tanto, hasta los **4,5 KHz** los armónicos están más definidos y la variación de señal es menor. (Ver Figura 6.14).

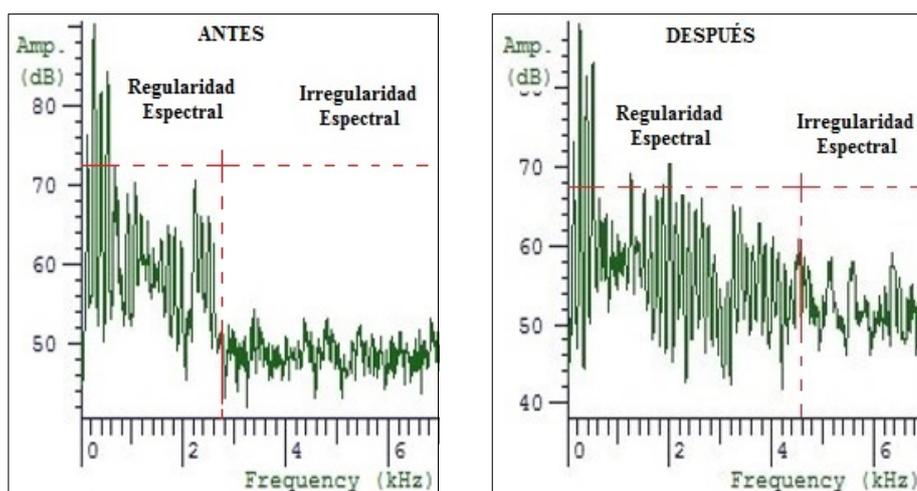


Figura 6.14.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ *Dominio Cepstral: Cepstrum de potencia*

En la *vocal “i” antes* de aplicar el método “Pantalla de celofán” se puede localizar el primer armónico (el *pitch*) en torno a **3 KHz**. La amplitud del pico del *pitch* es de aproximadamente **12 dB**, no encontrándose próximo al mismo un segundo armónico definido. Se aprecia además, la significativa presencia de ruido a frecuencias próximas e inferiores a la del *pitch*. (Ver Figura 6.15).

En la *vocal “i” después* de aplicar el método, el primer armónico (el *pitch*) se localiza igualmente en torno a **3 KHz**. La amplitud del *pitch* es en este caso de **20 dB** aproximadamente (mayor presencia de energía en dicha componente armónica), no habiendo un armónico secundario próximo al *pitch* que esté definido. (Ver Figura 6.15). Se deduce por tanto, la existencia de un *pitch* de una amplitud considerable. Esto indica la presencia notable de energía en dicha componente armónica, siendo ésta una característica de las voces de alta calidad. Además, se observa que la anchura del pico cepstral del *pitch* es reducida, lo que indica que la frecuencia de este armónico fundamental es altamente estable. (Ver Figura 6.15).

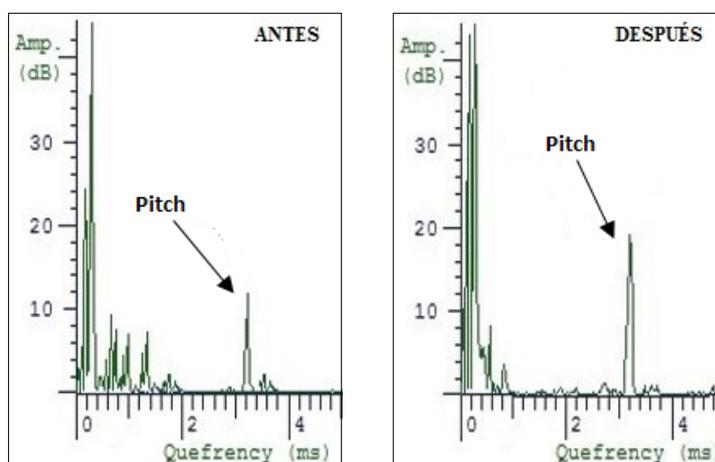


Figura 6.15.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

❖ **Caso 2: Análisis de voz para las vocales “o”, “a” e “i” correspondiente al locutor nº 3.**

Descripción: Voz de un varón, de entre 35 y 50 años de edad, con aparentes signos de ronquera o afonía y de fatiga vocal.

✓ **Análisis de la vocal “O” correspondiente al locutor nº 3.**

➤ **Vocal “o” antes::**

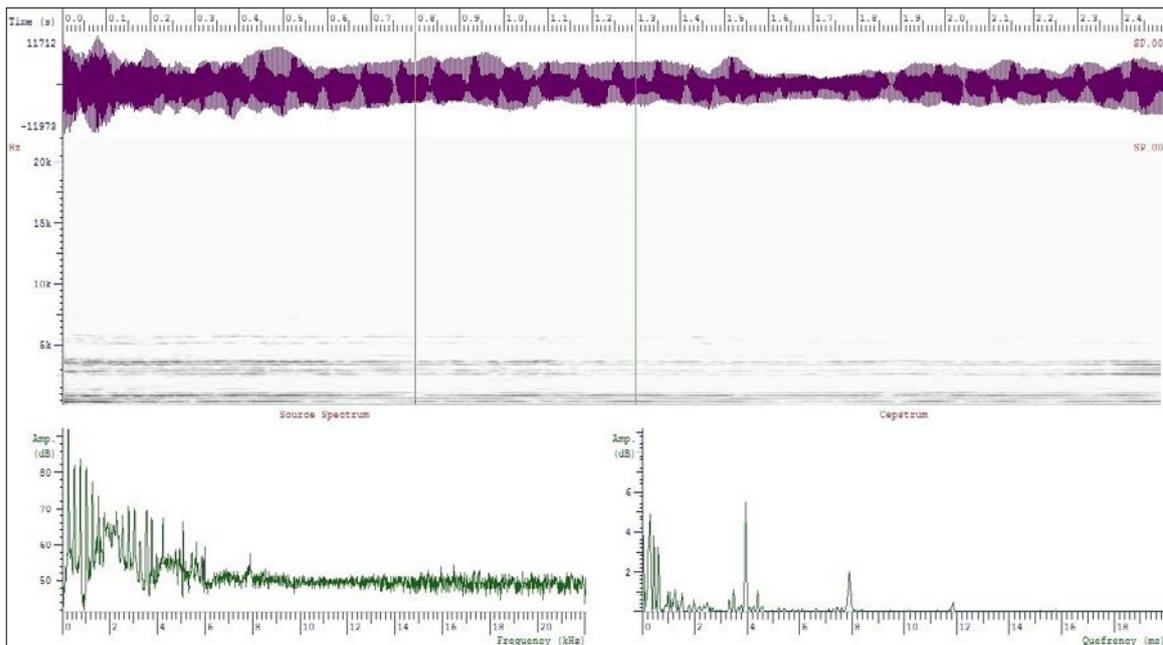


Fig. 6.16.- Vocal “o” antes de aplicar el método

➤ **Vocal “o” después:**

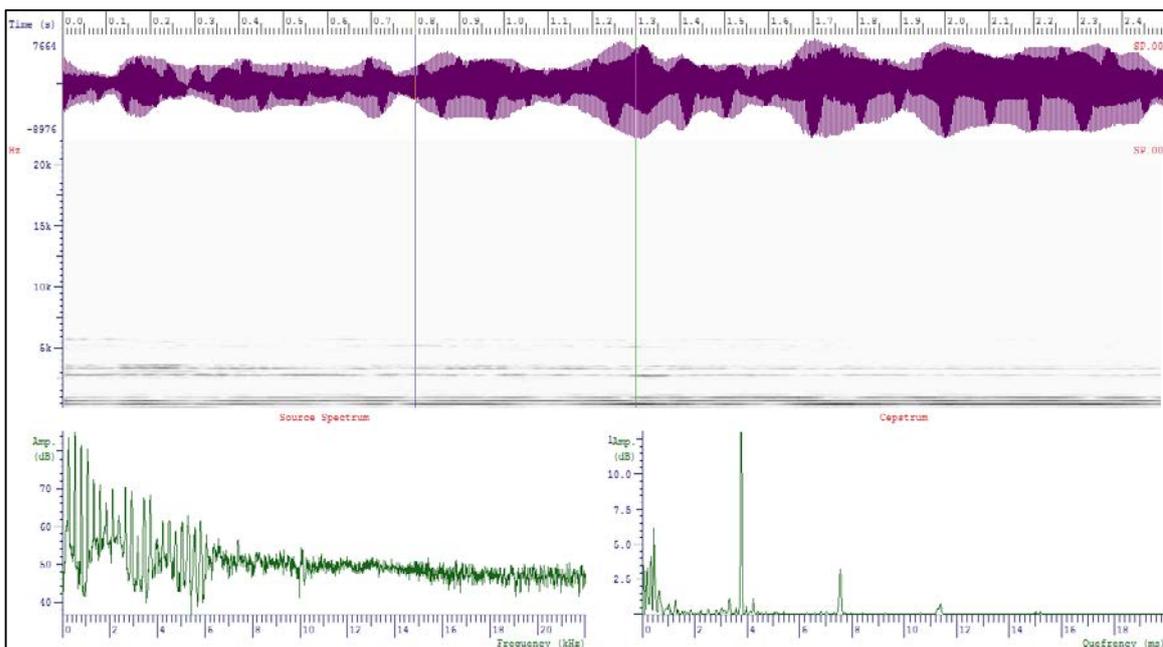


Fig. 6.17.- Vocal “o” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “o”, se obtiene una señal de voz endeble y con pocas oscilaciones *antes* de aplicar el método, y que posteriormente *después* de aplicar el método se convierte en una señal de voz menos débil y algo más abrupta en sus instantes finales.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

Se constata que en la *vocal “o” antes* de aplicar el método apenas se comprueba que existe bastante ruido a frecuencias bajas y hasta los **5 KHz** (Ver Figura 6.18).

No obstante, en la *vocal “o” después* de aplicar el método la presencia de ruido es mucho menor a bajas frecuencias y se empieza a apreciar la banda donde los formantes (bandas de frecuencia donde se concentra la mayor parte de la energía sonora de un sonido) empiezan a definirse en torno a los **2,5~3 KHz** (Ver Figura 6.18).

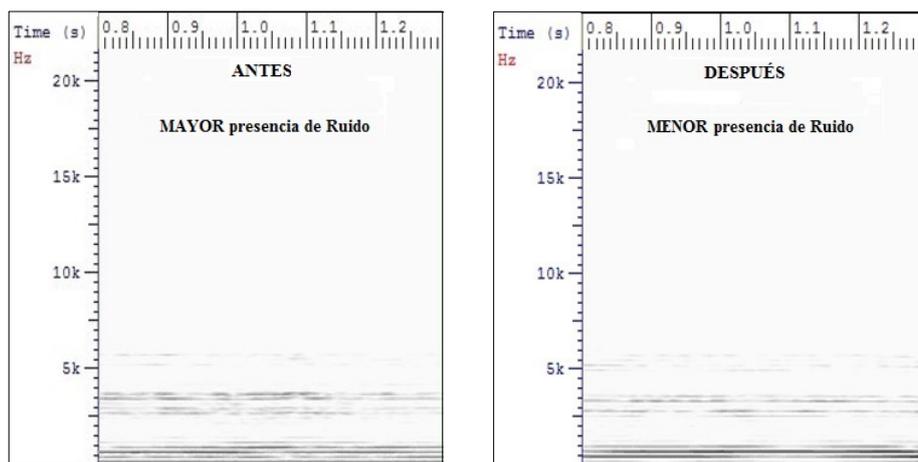


Figura 6.18.- Zoom aumentado de la señal en el dominio *frecuencial* (espectrograma) para la *vocal “o” antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la *vocal “o” antes* de aplicar el método, hay menor regularidad de la envolvente y el número de armónicos está mucho menos definido a partir de los **4 KHz** *aproximadamente* (Ver Figura 6.19).

En la *vocal “o” después* de aplicar el método se aprecia que el intervalo de periodicidad es mayor en dicha señal, donde la variación del espectro de una trama con respecto a las tramas contiguas no se produce *aproximadamente hasta los 6 KHz*, *habiendo mayor regularidad de la envolvente del espectro* hasta dicha frecuencia y siendo la señal menos cambiante y de más calidad, y con mayor número de armónicos establecidos (Ver Figura 6.19).

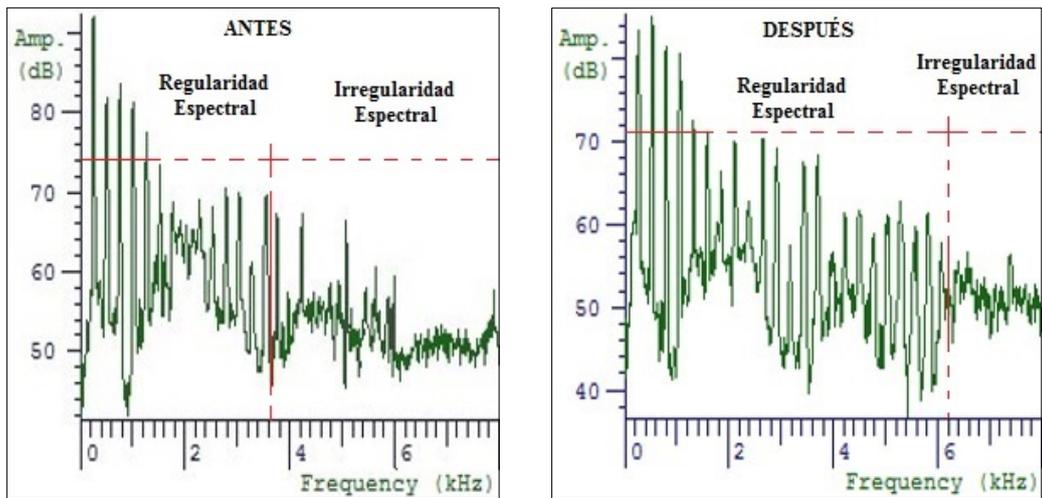


Figura 6.19.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “o” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Cepstral: Cepstrum de potencia**

En la *vocal “o” antes* de aplicar el método “Pantalla de celofán” se puede localizar el *pitch* en torno a **4 KHz**. La amplitud del pico del *pitch* es de aproximadamente **5,5 dB**, encontrándose próximo a este *pitch* un segundo armónico en torno a los **8 KHz** con una amplitud aproximada de **2 dB**. Se aprecia demás, que existe bastante cantidad de ruido a frecuencias próximas e inferiores a la del *pitch* (Ver Figura 6.20).

En la *vocal “o” después* de aplicar el método, el *pitch* se localiza igualmente en torno a **4 KHz**, localizándose próximo a él igualmente un armónico secundario en torno a **8 KHz**. La amplitud del pico del *pitch* es en este caso **de 13 dB** aproximadamente (mayor presencia de energía en dicha componente armónica) y el del armónico secundario de aproximadamente **3 dB**. La cantidad de ruido en bajas frecuencia en este caso es mucho menor, comprobándose que se trata de una voz de mayor calidad que la del caso anterior (Ver Figura 6.20).

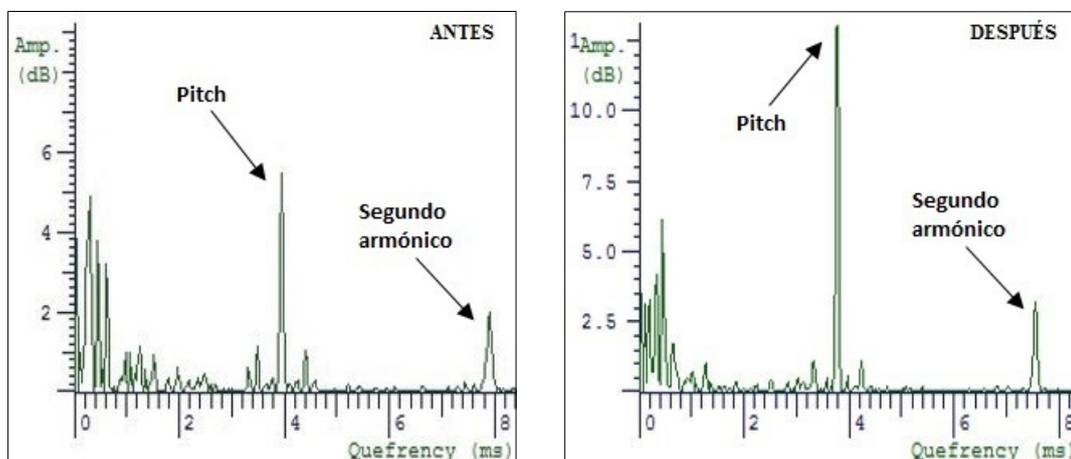


Figura 6.20.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “o” *antes* (izquierda) y *después* (derecha) de aplicar el método

✓ **Análisis de la vocal “A” correspondiente al locutor nº 3.**

➤ **Vocal “a” antes:**

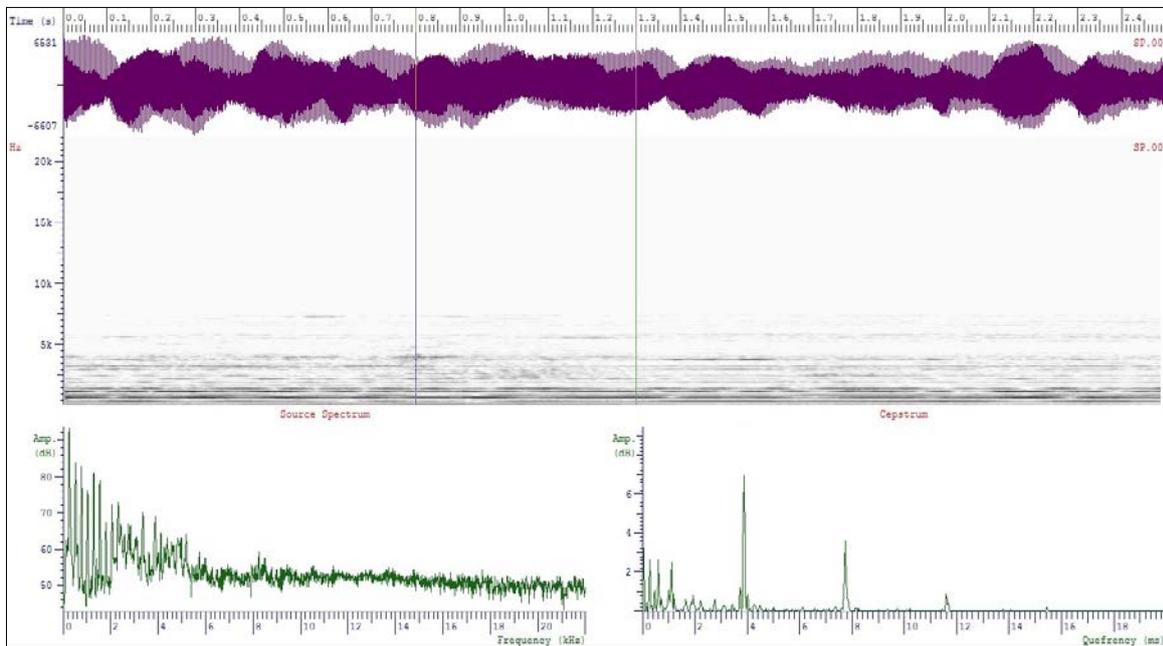


Fig. 6.21.- Vocal “a” antes de aplicar el método

➤ **Vocal “a” después:**

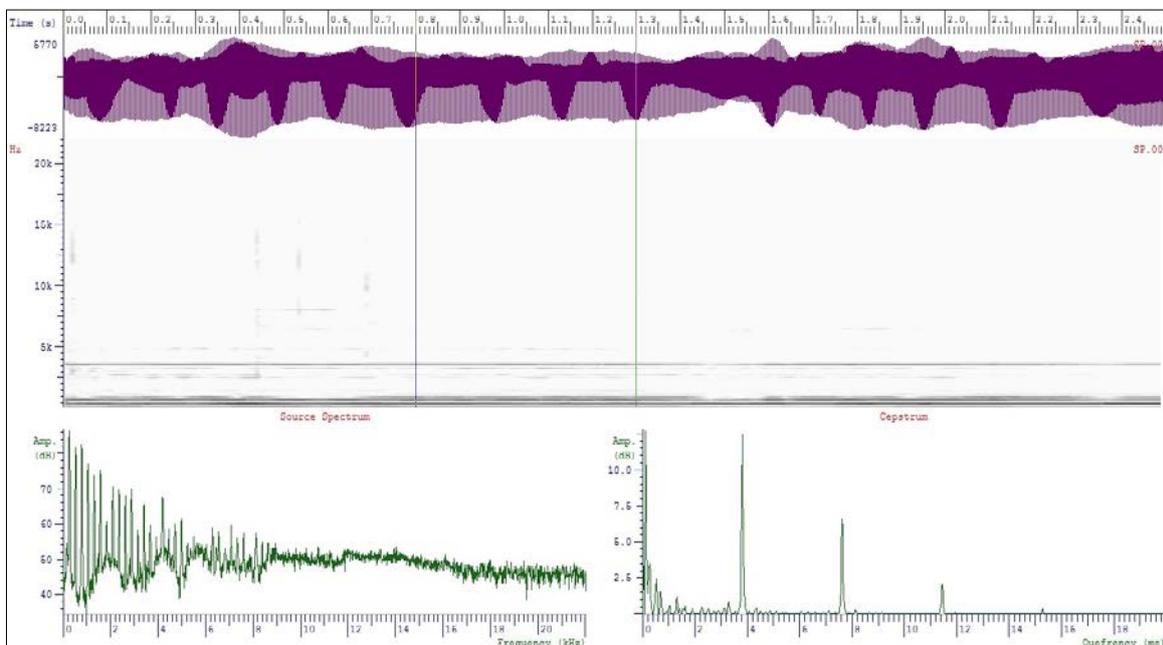


Fig. 6.22.- Vocal “a” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “a”, se obtiene una señal de voz que presenta signos de irregularidad y donde la envolvente de la señal de voz no es estable *antes* de aplicar el método. *Después* de aplicar el método la envolvente de la señal moduladora empieza a estabilizarse produciéndose algunas oscilaciones en su recorrido.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

En la *vocal “a” antes* de aplicar el método, se puede observar que existe bastante ruido a frecuencias inferiores a **4.5 KHz** y alguna que otra presencia de ruido pero de forma más escasa por encima de los **5 KHz** aproximadamente (Ver Figura 6.23).

En la *vocal “a” después* de aplicar el método, es notorio que la cantidad de ruido presente es menor que en el caso anterior, empezándose a apreciar en torno a los **3,5 KHz** una línea paralela donde las componentes armónicas se encuentran más definidas (Ver Figura 6.23).

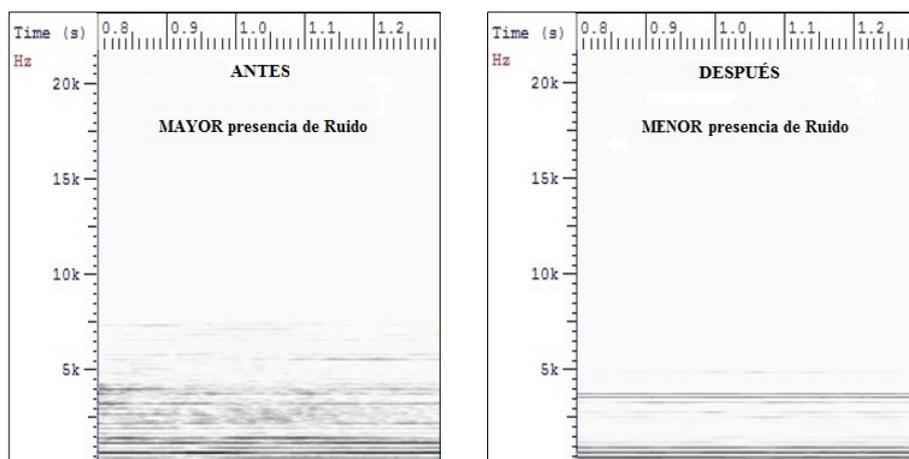


Figura 6.23.- Zoom aumentado de la señal en el dominio *frecuencial* (espectrograma) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la *vocal “a” antes* de aplicar el método, hay menor regularidad de la envolvente y el número de armónicos está mucho menos definido a partir de los **7 KHz** *aproximadamente* (Ver Figura 6.24).

En la *vocal “a” después* de aplicar el método se aprecia que la variación del espectro de una trama con respecto a las tramas contiguas no se produce *aproximadamente hasta los 8,5KHz*, *habiendo mayor regularidad de la envolvente del espectro* hasta dicha frecuencia y siendo la señal menos variable y con los armónicos más definidos (Ver Figura 6.24).

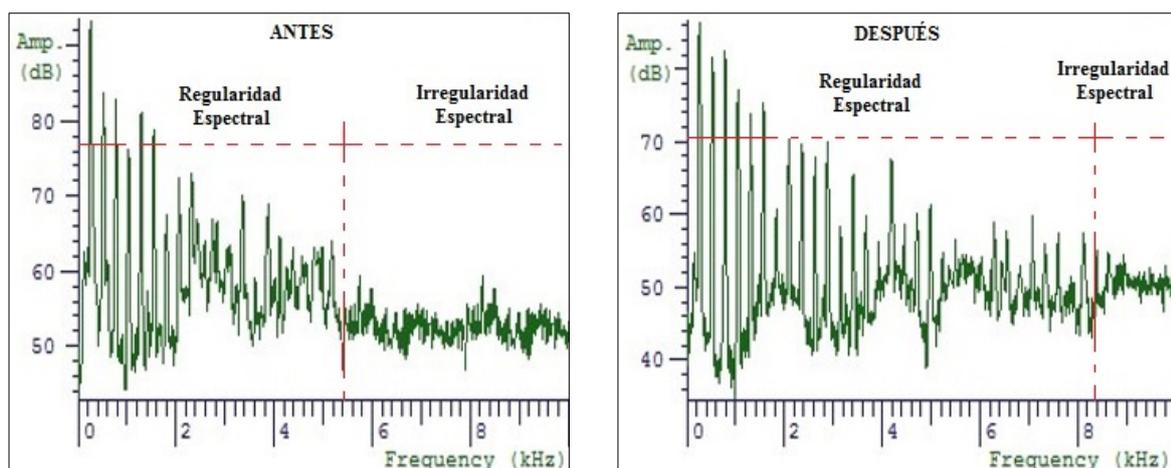


Figura 6.24.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ *Dominio Cepstral*: Cepstrum de potencia

En la *vocal “a” antes* de aplicar el método “Pantalla de celofán” se localiza el *pitch* en torno a **4 KHz**. La amplitud del pico del *pitch* es de aproximadamente **9,5 dB**, encontrándose próximo a este *pitch* un segundo armónico en torno a los **8 KHz** con una amplitud aproximada de **4 dB**, y ya en torno a los **12 KHz** signos de otro tercer armónico. En este caso además, se observa que la presencia de ruido a frecuencias próximas e inferiores a la del *pitch* es considerable (Ver Figura 6.25).

En la *vocal “a” después* de aplicar el método, el *pitch* se localiza igualmente en torno a **4 KHz**, localizándose próximo a él igualmente un armónico secundario en torno a **8 KHz**. La amplitud del pico del *pitch* es en este caso **de 12 dB** aproximadamente (mayor presencia de energía en dicha componente armónica) y el del armónico secundario de aproximadamente **7 dB**. Además, nótese en este caso que el tercer armónico está más definido; teniendo mayor amplitud. La presencia de ruido en este caso es mucho menor a frecuencias más bajas. Se deduce pues que se trata de una voz de mayor calidad que la del caso anterior (Ver Figura 6.25).

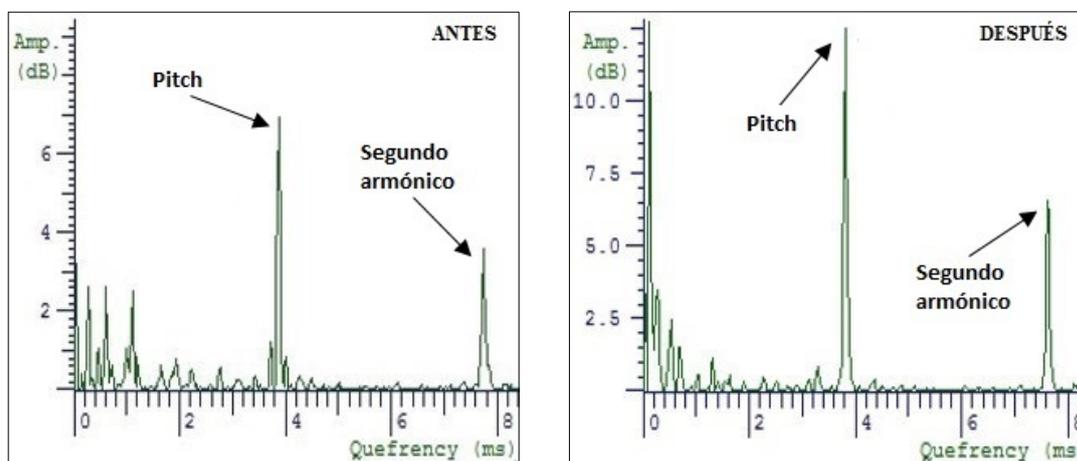


Figura 6.25.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

✓ **Análisis de la vocal “I” correspondiente al locutor nº 3.**

➤ **Vocal “i” antes::**

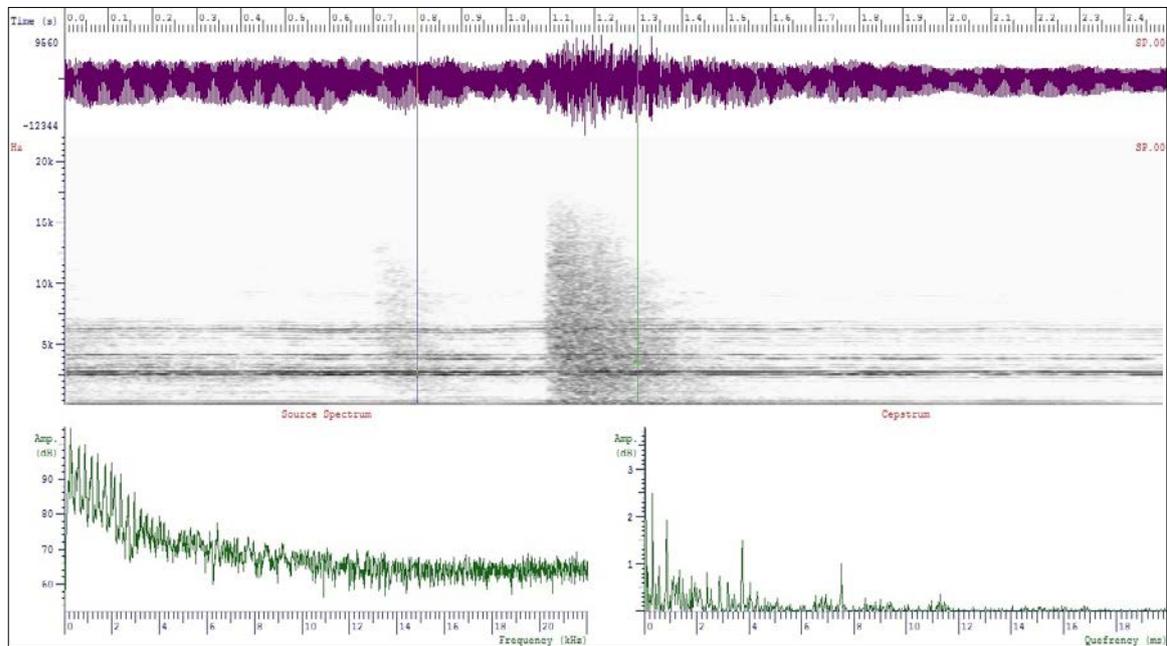


Fig. 6.26.- Vocal “i” antes de aplicar el método

➤ **Vocal “i” después:**

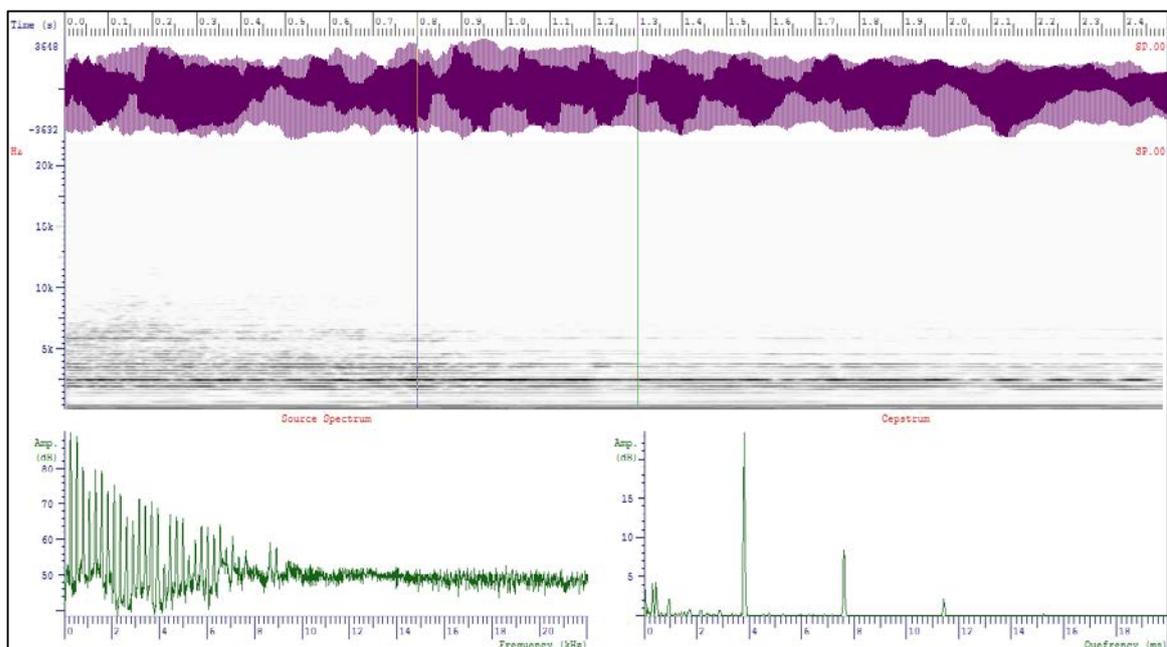


Fig. 6.27.- Vocal “i” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “i”, se obtiene una señal de voz frágil de poca amplitud en frecuencia, con ciertas irregularidades *antes* de aplicar el método, y que posteriormente *después* de aplicar el método se convierte en una señal de voz más energética y donde se aprecia de forma más significativa la envolvente de la señal moduladora.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

Se puede comprobar que en la **vocal “i”** *antes* de aplicar el método se comprueba que existe bastante ruido y las líneas paralelas no están bien definidas y por tanto, las componentes armónicas son escasas. Se denota una cantidad de ruido considerable en el intervalo de tiempo comprendido entre los 1 y 1,5 segundos, para las frecuencias inferiores a **15 KHz** (Ver Figura 6.28).

Sin embargo, en la **vocal “i”** *después* de aplicar el método la presencia de ruido es mucho menor a bajas frecuencias y se empieza a apreciar la banda donde los formantes empiezan a definirse en torno a los **2,5~3 KHz**, siendo en esta línea paralela donde existe mayores componentes armónicas y acompañada esta banda de menos ruido a frecuencias próximas (Ver Figura 6.28).

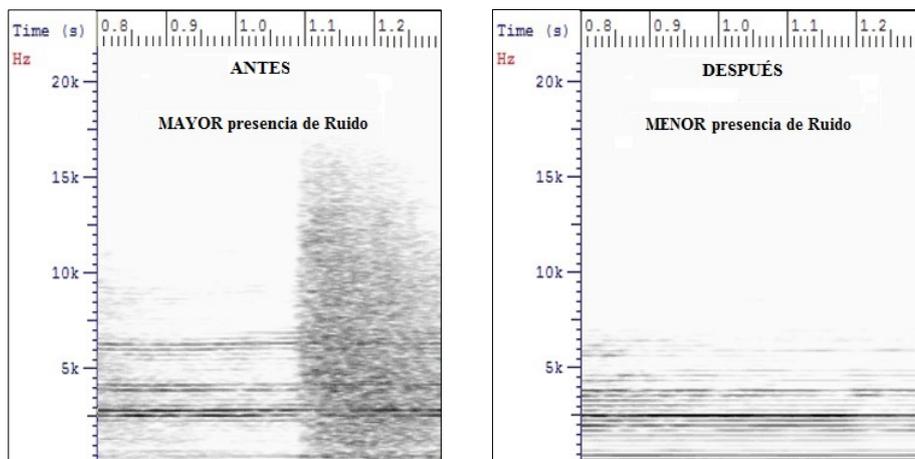


Figura 6.28.- Zoom aumentado de la señal en el dominio *frecuencial* (espectrograma) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la **vocal “i”** *antes* de aplicar el método, hay menor regularidad de la envolvente y el número de armónicos está mucho menos definido a partir de los **3 KHz** *aproximadamente* (Ver Figura 6.29).

En la **vocal “i”** *después* de aplicar el método se aprecia que la variación del espectro de una trama con respecto a las tramas contiguas no se produce *aproximadamente hasta los 6,5KHz*, *habiendo mayor regularidad de la envolvente del espectro* hasta dicha frecuencia y siendo la señal menos variable y con los armónicos más definidos (Ver Figura 6.29).

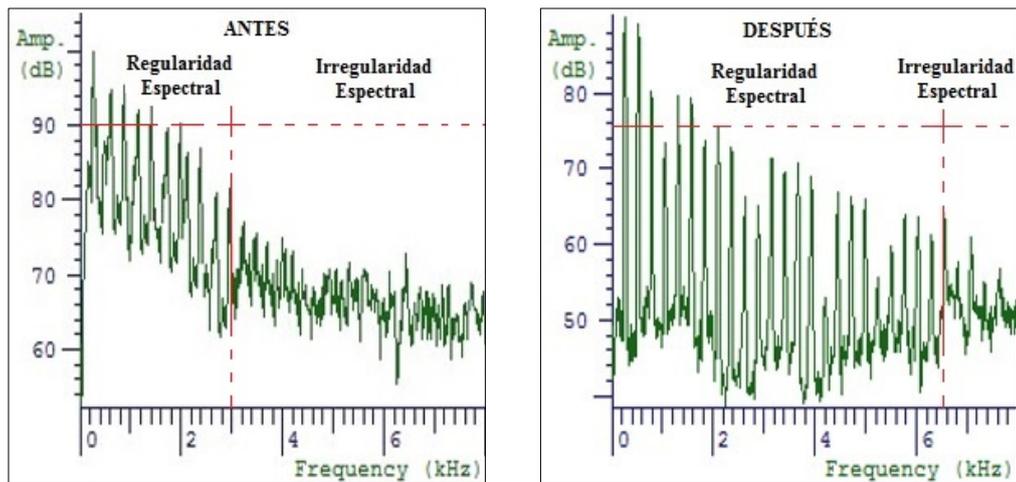


Figura 6.29.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Cepstral: Cepstrum de potencia**

En la *vocal “i” antes* de aplicar el método “Pantalla de celofán” se puede localizar el *pitch* en torno a **4 KHz**. La amplitud del pico del *pitch* es de aproximadamente **1,5 dB**, encontrándose próximo a este *pitch* un segundo armónico en torno a los **8 KHz** con una amplitud aproximada de **4 dB**, y ya en torno a los **12 KHz** signos de otro tercer armónico. En este caso además, se observa que la presencia de ruido a frecuencias próximas e inferiores a la del *pitch* es considerable (Ver Figura 6.30).

En la *vocal “i” después* de aplicar el método, el *pitch* se localiza igualmente en torno a **4 KHz**, localizándose próximo a él igualmente un armónico secundario en torno a **8 KHz**. La amplitud del pico del *pitch* es en este caso **de 12 dB** aproximadamente (mayor presencia de energía en dicha componente armónica) y el del armónico secundario de aproximadamente **7 dB**. Además, nótese en este caso que el tercer armónico está más definido; teniendo mayor amplitud. La presencia de ruido en este caso es mucho menor a frecuencias más bajas. Se deduce pues que se trata de una voz de mayor calidad que la del caso anterior (Ver Figura 6.30).

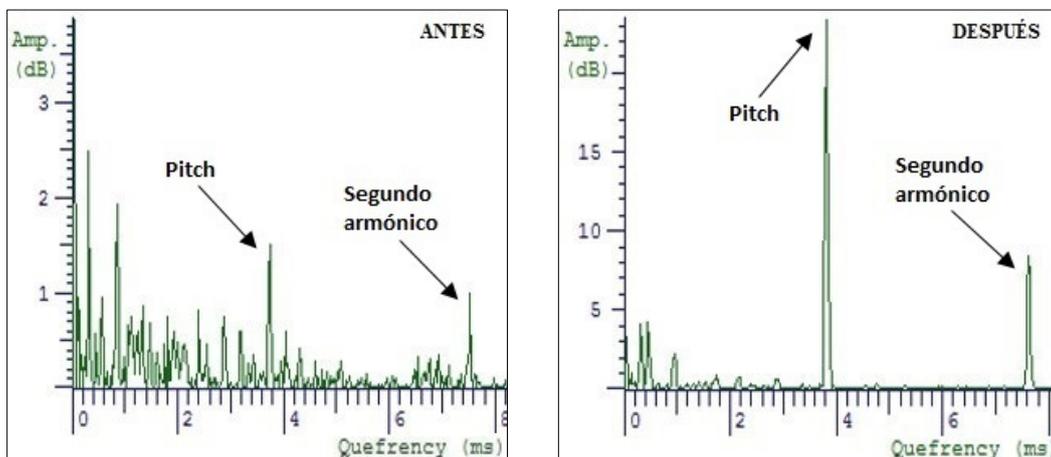


Figura 6.30.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

❖ **Caso 3: Análisis de voz para las vocales “o”, “a” e “i” correspondiente al locutor nº 11.**

Descripción: Voz de un varón, de entre 25 y 35 años de edad, que presenta signos de debilidad e irregularidad en la voz.

✓ **Análisis de la vocal “O” correspondiente al locutor nº 11.**

➤ **Vocal “o” antes::**

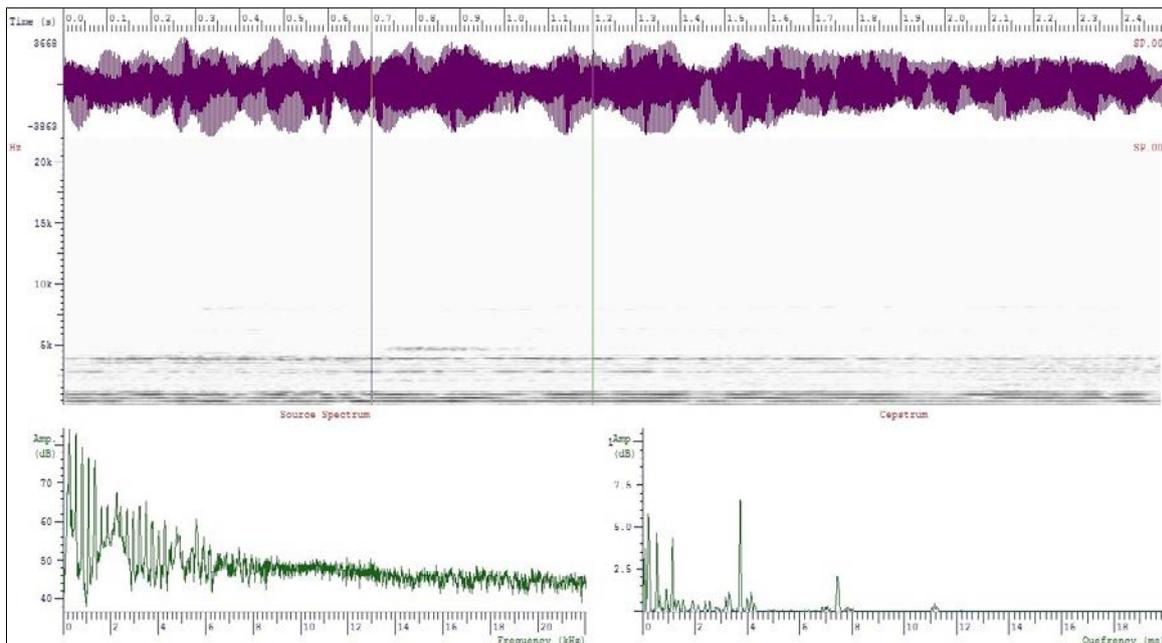


Fig. 6.31.- Vocal “o” antes de aplicar el método

➤ **Vocal “o” después:**

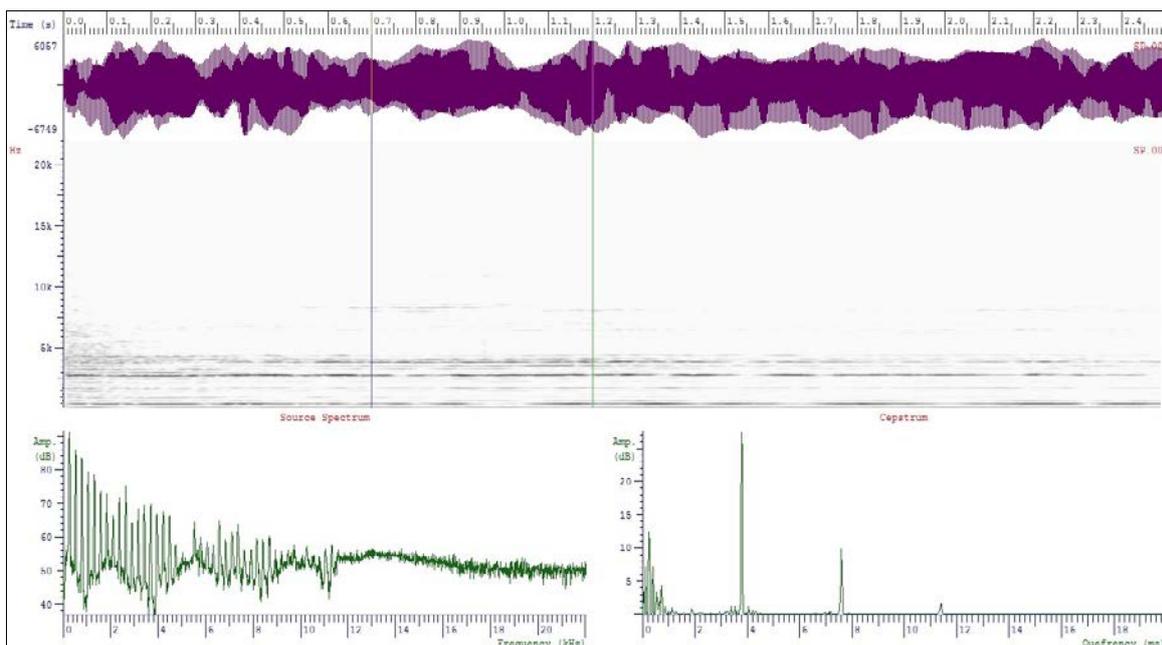


Fig. 6.32.- Vocal “o” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “o”, se obtiene una señal de voz endeble y con irregularidades donde la envolvente de la señal moduladora no está definida *antes* de aplicar el método, y que posteriormente *después* de aplicar el método se convierte en una señal de voz con mayor amplitud en frecuencia y donde la envolvente de la señal está más definida.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

En la *vocal “o” antes* de aplicar el método existe bastante ruido a frecuencias bajas; lo cual indica que hay zonas irregulares de energía hasta los **5 KHz** aproximadamente (Ver Figura 6.33).

Por otro lado, en la *vocal “o” después* de aplicar el método la presencia de ruido es mucho menor a bajas frecuencias y se empieza a apreciar una banda donde los formantes empiezan a definirse en torno a los **3,5 KHz** aproximadamente; por tanto, las componentes armónicas están más definidas y con menos presencia de ruido (Ver Figura 6.33).

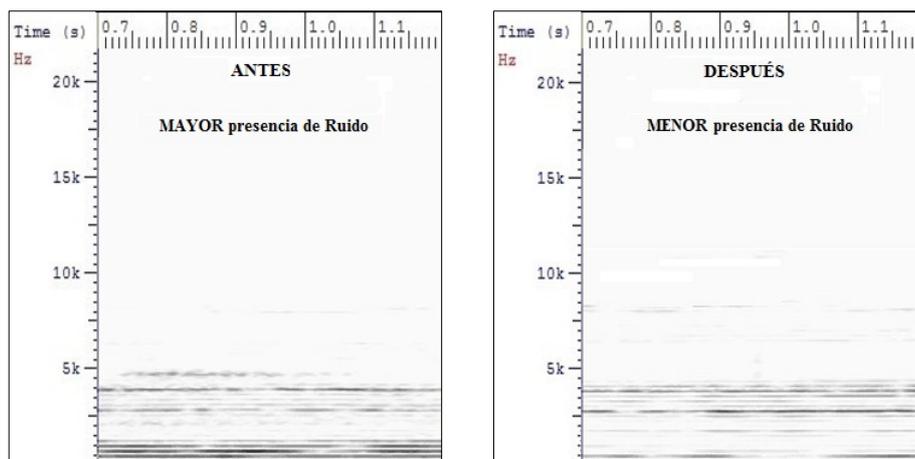


Figura 6.33.- Zoom aumentado de la señal en el dominio frecuencial (espectrograma) para la vocal “o” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la *vocal “o” antes* de aplicar el método, hay menor regularidad de la envolvente y el número de armónicos está mucho menos definido a partir de los **6,5 KHz** aproximadamente (Ver Figura 6.34).

En la *vocal “o” después* de aplicar el método se aprecia que la variación del espectro de una trama con respecto a las tramas contiguas no se produce aproximadamente **hasta los 9 KHz**, habiendo *mayor regularidad de la envolvente del espectro* hasta dicha frecuencia y siendo la señal menos variable y con los armónicos más definidos (Ver Figura 6.34).

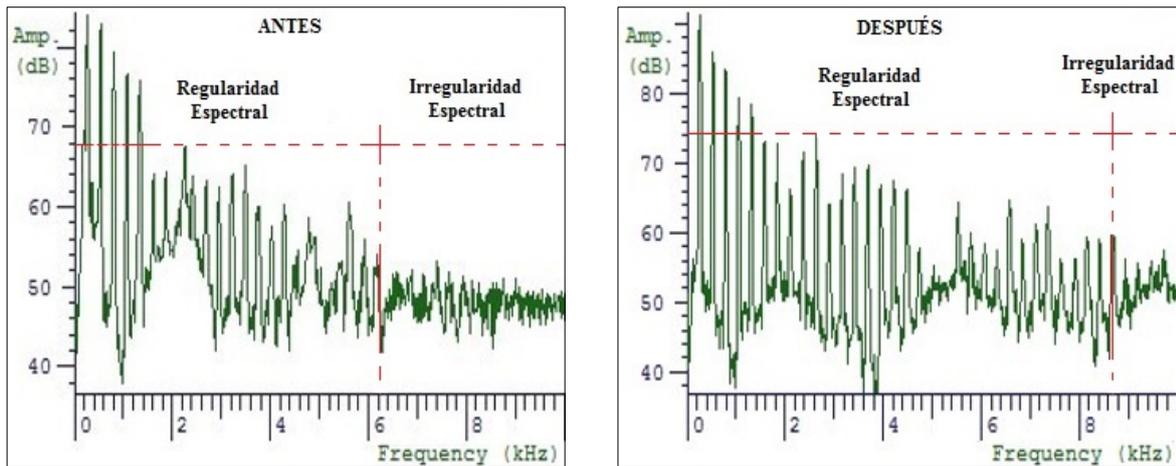


Figura 6.34.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “o” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Cepstral: Cepstrum de potencia**

En la *vocal “o” antes* de aplicar el método “Pantalla de celofán” se puede localizar el primer armónico (el *pitch*) en torno a **4 KHz**. La amplitud del pico del *pitch* es de aproximadamente **6,5 dB**, encontrándose próximo a este *pitch* un segundo armónico en torno a los **7,5 KHz** con una amplitud aproximada de **2 dB**. Es de notar también, que existe una cantidad anormal de ruido a frecuencias próximas e inferiores a la del *pitch* (Ver Figura 6.35).

En la *vocal “o” después* de aplicar el método, el *pitch* se localiza igualmente en torno a **4 KHz**, localizándose próximo a él igualmente un armónico secundario en torno a **7,5 KHz**. La amplitud del pico del *pitch* es en este caso **de 27 dB** aproximadamente (mayor presencia de energía en dicha componente armónica) y el del armónico secundario de aproximadamente **9 dB**. Además, se aprecia que la cantidad de ruido a bajas frecuencias es mucho menor. Se deduce pues que se trata de una voz de mayor calidad que la del caso anterior (Ver Figura 6.35).

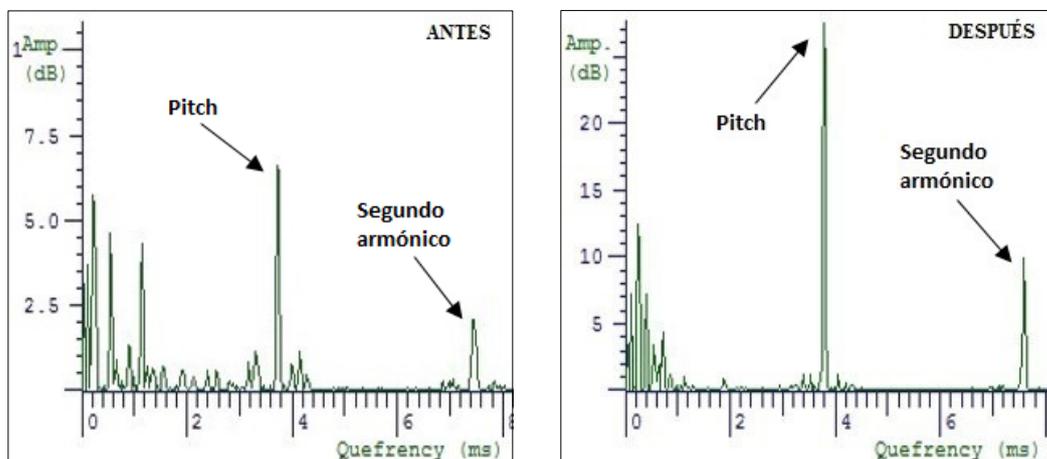


Figura 6.35.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “o” *antes* (izquierda) y *después* (derecha) de aplicar el método

✓ **Análisis de la vocal “A” correspondiente al locutor nº 11.**

➤ **Vocal “a” antes:**

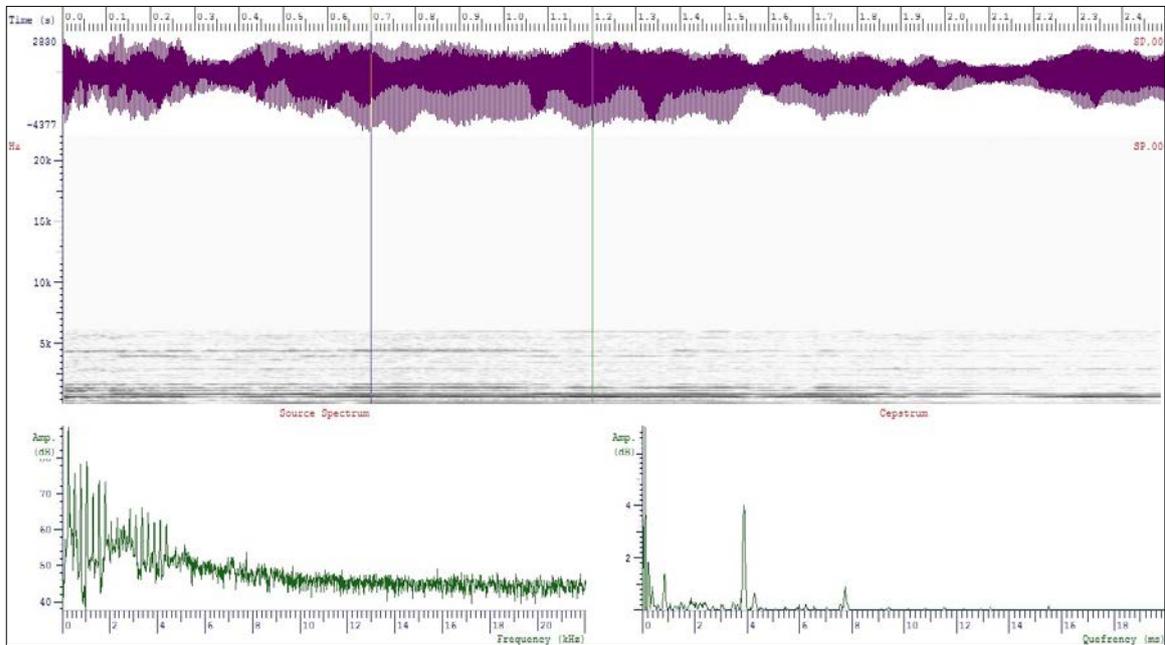


Fig. 6.36.- Vocal “a” *antes* de aplicar el método

➤ **Vocal “a” después:**

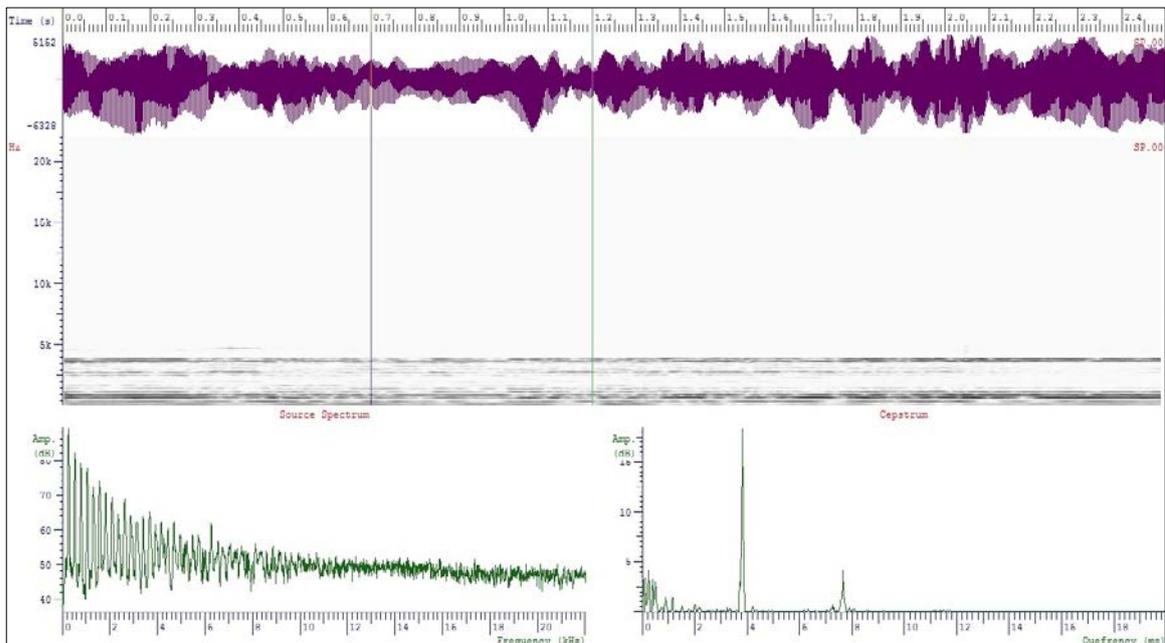


Fig. 6.37.- Vocal “a” *después* de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “a”, se obtiene una señal de voz débil y con fluctuaciones durante su período de tiempo *antes* de aplicar el método. *Después* de aplicar el método la envolvente de la señal de voz se denota de forma más significativa pero con demasiadas oscilaciones en su recorrido.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

En la *vocal “a” antes* de aplicar el método apenas se comprueba que existe bastante ruido a frecuencias bajas y hasta los **7 KHz**, y además no existe una línea paralela característica definida donde las componentes armónicas estén mayormente definidos (Ver Figura 6.38).

En la *vocal “a” después* de aplicar el método la presencia de ruido es considerablemente menor a bajas frecuencias y se aprecia que en torno a los **4 KHz** aproximadamente se tienen una línea paralela con poca cantidad de ruido en frecuencias próximas: lo cual indica, que se trata de una zona donde se encuentran las componentes armónicas de la señal de voz y donde hay menos zonas presentes de energía irregulares (Ver Figura 6.38).

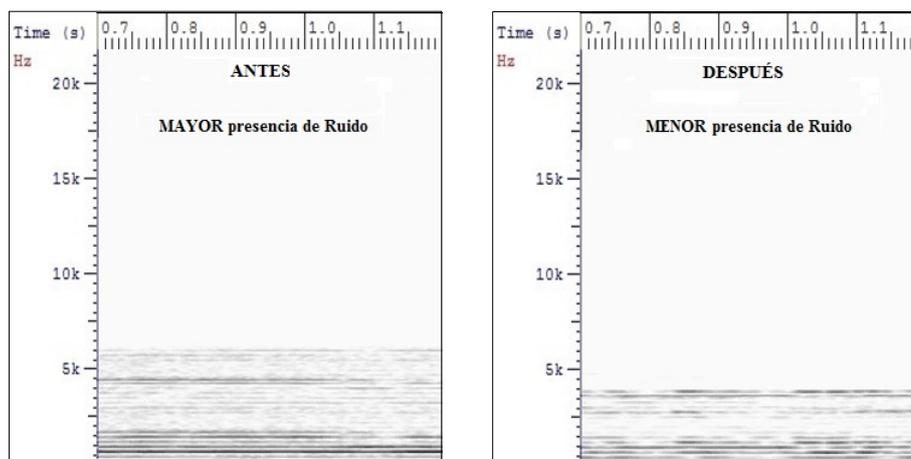


Figura 6.38.- Zoom aumentado de la señal en el dominio *frecuencial* (espectrograma) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la *vocal “a” antes* de aplicar el método, hay menor regularidad de la envolvente y el número de armónicos está mucho menos definido a partir de los **7 KHz** *aproximadamente* (Ver Figura 6.39).

En la *vocal “a” después* de aplicar el método se aprecia que la variación del espectro de una trama con respecto a las tramas contiguas no se produce *aproximadamente hasta los 8,5KHz*, *habiendo mayor regularidad de la envolvente del espectro* hasta dicha frecuencia y siendo la señal menos variable y con los armónicos más definidos (Ver Figura 6.39).

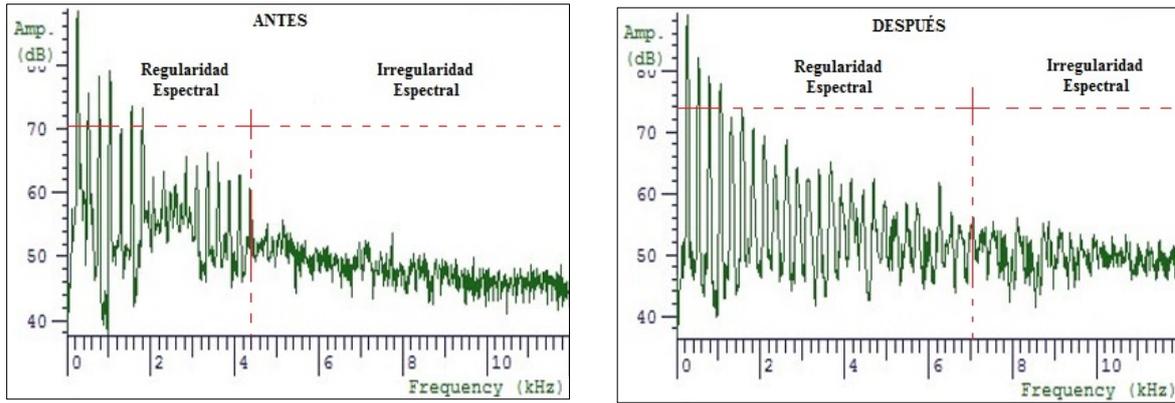


Figura 6.39.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Cepstral: Cepstrum de potencia**

En la *vocal “a” antes* de aplicar el método “Pantalla de celofán” se puede localizar el primer armónico (el *pitch*) en torno a **4 KHz**. La amplitud del pico del *pitch* es de aproximadamente **4 dB**, encontrándose próximo a este *pitch* un segundo armónico en torno a los **8 KHz** con una amplitud aproximada de **1 dB**. Se observa que la presencia de ruido a frecuencias próximas e inferiores a la del *pitch* es considerable (Ver Figura 6.40).

En la *vocal “a” después* de aplicar el método, el *pitch* se localiza igualmente en torno a **4 KHz**, localizándose próximo a él igualmente un armónico secundario en torno a **8 KHz**. La amplitud del pico del *pitch* es en este caso **de 18 dB** aproximadamente (mayor presencia de energía en dicha componente armónica) y el del armónico secundario de aproximadamente **3 dB**. En este caso se aprecia que la cantidad de ruido a frecuencias bajas es mucho menor. Se deduce pues que se trata de una voz de mayor calidad que la del caso anterior (Ver Figura 6.40).

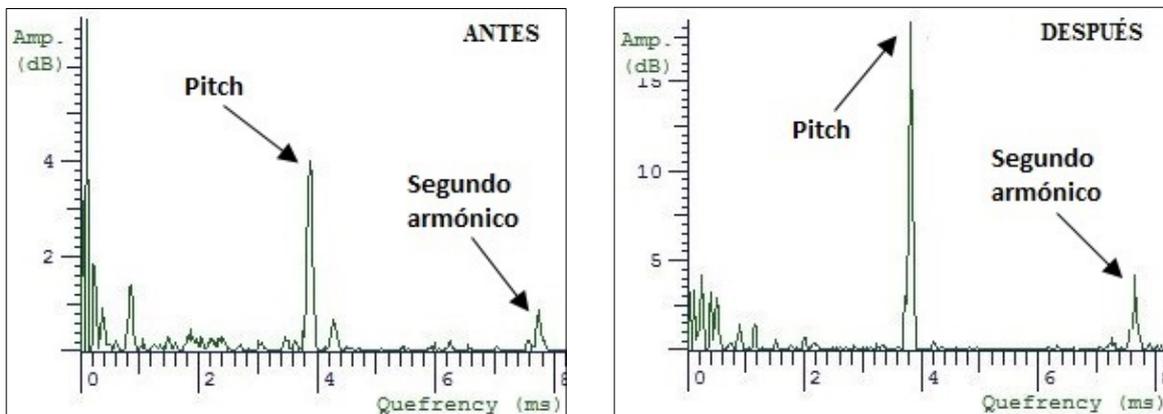


Figura 6.40.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “a” *antes* (izquierda) y *después* (derecha) de aplicar el método

✓ Análisis de la vocal “P” correspondiente al locutor nº 11.

➤ Vocal “i” antes::

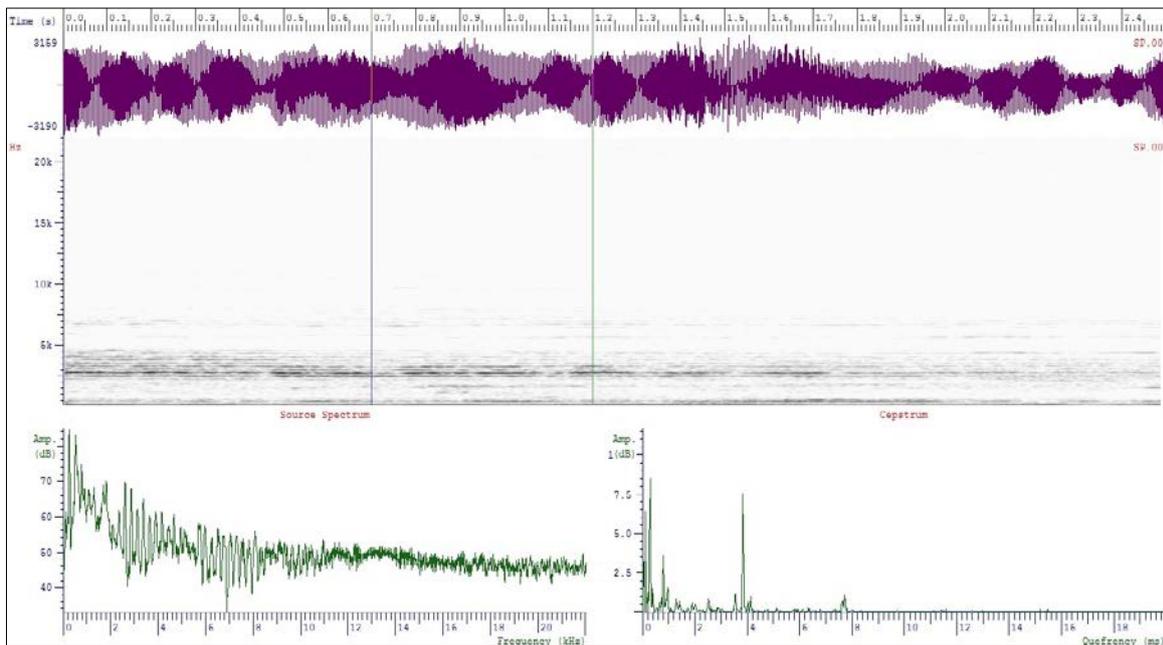


Fig. 6.41.- Vocal “i” antes de aplicar el método

➤ Vocal “i” después:

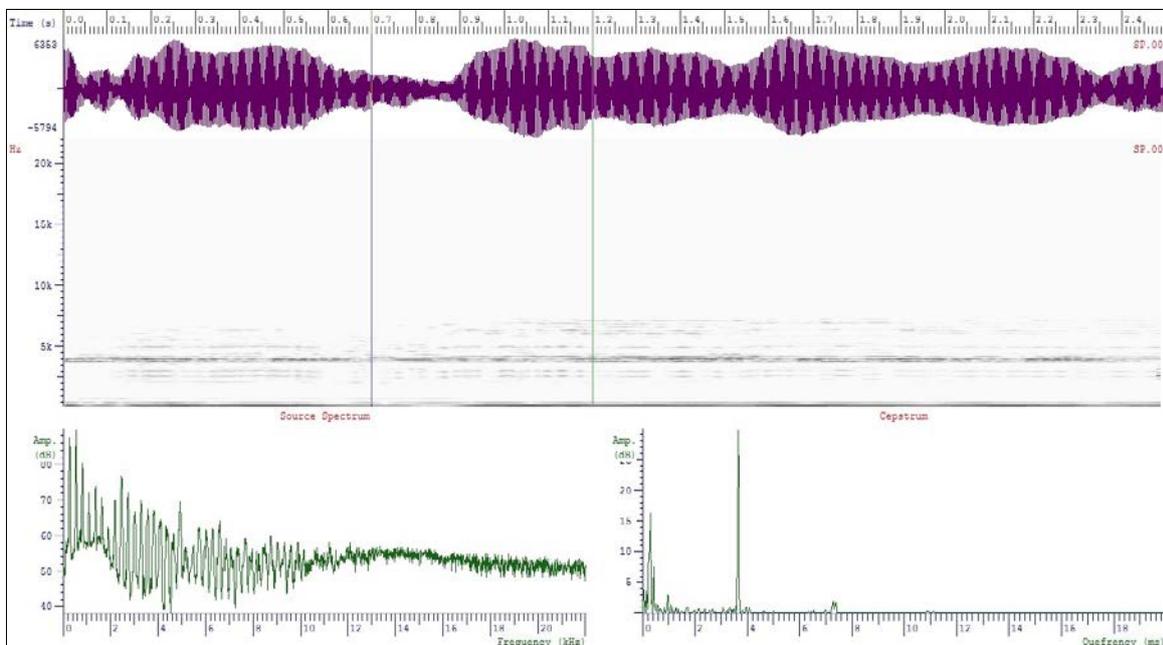


Fig. 6.42.- Vocal “i” después de aplicar el método

➤ **Dominio Temporal: Señal de voz**

En la pronunciación de la vocal “i”, se obtiene una señal de voz débil y con aparentes signos de irregularidad *antes* de aplicar el método. Se observa que *después* de aplicar el método la señal senoidal está más definida.

➤ **Dominio Frecuencial: Espectrograma de banda estrecha**

En la *vocal “i” antes* de aplicar el método apenas se verifica que se trata de una señal de voz donde existen zonas de energía irregulares, lo cual indica la presencia de cantidades anormales de ruido, y donde no existen líneas paralelas claramente definidas, en el tramo comprendido en torno a los **2~5 KHz** (Ver Figura 6.43).

En la *vocal “i” después* de aplicar el método se denota que la presencia de ruido es escaso y que en la banda comprendida entre los **3,5~4 KHz**, se define una línea paralela claramente definida, donde se encuentran las componentes armónicas de la señal de voz (Ver Figura 6.43).

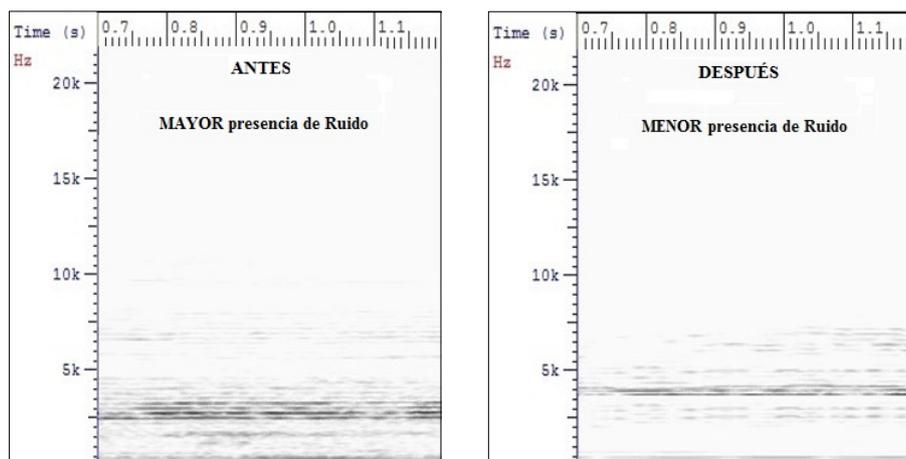


Figura 6.43.- Zoom aumentado de la señal en el dominio *frecuencial* (espectrograma) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Frecuencial: Excitación glótica**

En la *vocal “i” antes* de aplicar el método, se denota que existe un intervalo regular de periodicidad hasta los **8,5 KHz aproximadamente**, donde la envolvente del espectro presenta un tramo de regularidad y donde la variación de una trama con respecto a las tramas contiguas no es tan considerable (Ver Figura 6.44).

En la *vocal “i” después* de aplicar el método se aprecia que es hasta los **10 KHz aproximadamente**, donde la variación del espectro de una trama con respecto a las tramas contiguas no es tan significativa, siendo hasta dicha frecuencia donde hay mayor regularidad de la envolvente, tratándose pues de una señal de voz de mayor calidad que la del caso anterior (Ver Figura 6.44).

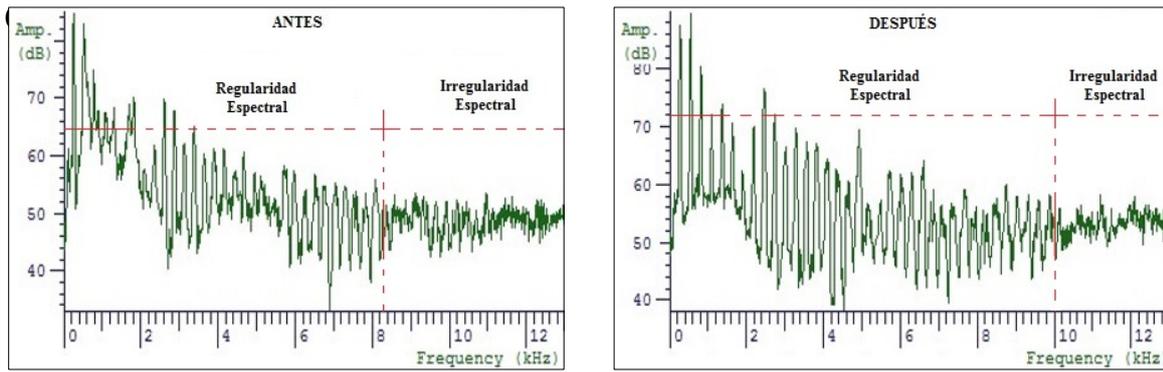


Figura 6.44.- Zoom aumentado de la señal en el dominio *frecuencial* (excitación glótica) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

➤ **Dominio Cepstral: Cepstrum de potencia**

En la *vocal “i” antes* de aplicar el método “Pantalla de celofán”, el *pitch* se ubica en torno a **4 KHz**. La amplitud del pico del *pitch* es de aproximadamente **7,5 dB**. A frecuencias próximas e inferiores a la del *pitch* se observa que la presencia de ruido es considerable (Ver Figura 6.45).

En la *vocal “i” después* de aplicar el método, se aprecia que el *pitch* se encuentra ubicado en torno a **4 KHz** de igual forma que el caso anterior, pero en este caso con una amplitud de **30 dB** aproximadamente (mayor presencia de energía en dicha componente armónica). A frecuencias bajas, se observa que *después* de aplicar el método correspondiente, la cantidad de ruido presente es bastante menor en comparación con la misma vocal *antes* de aplicar el método citado anteriormente. Se deduce por tanto, la existencia de un *pitch* de una amplitud considerable. Esto indica la presencia notable de energía en dicha componente armónica, siendo ésta una característica de las voces de alta calidad. Además, se observa que la anchura del pico cepstral del *pitch* es reducida, lo que indica que la frecuencia de este armónico fundamental es altamente estable, correspondiéndose con una voz de más alta calidad (Ver Figura 6.45).

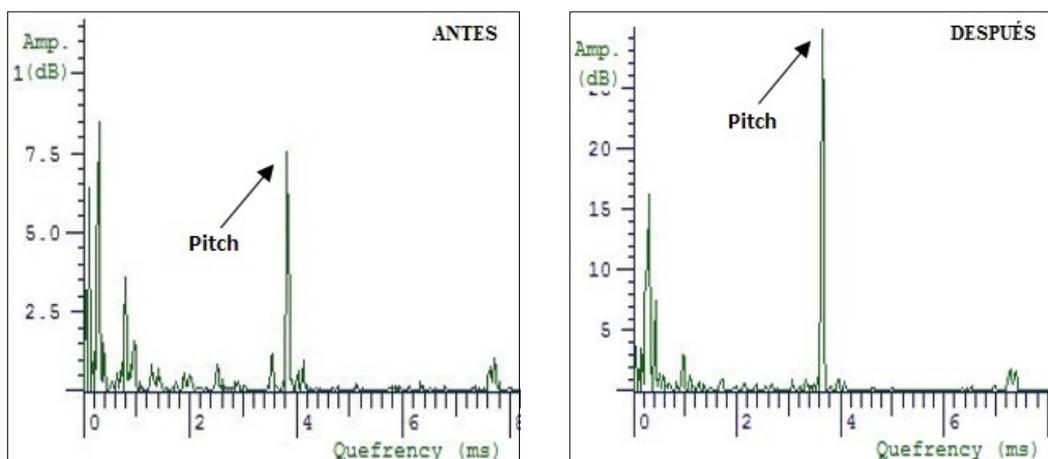


Figura 6.45.- Zoom aumentado de la señal en el dominio *cepstral* (cepstrum de potencia) para la vocal “i” *antes* (izquierda) y *después* (derecha) de aplicar el método

A modo de resumen, *después* de realizar el *análisis cualitativo* se puede concluir lo siguiente:

- Las voces *antes* de la aplicación del método “Pantalla de celofán” se caracterizan por ser aéreas (*voz con exceso de escape de aire*) y especialmente caracterizadas por la presencia de ruido en altas frecuencias.
- Las voces *después* del tratamiento se caracterizan por ser más estables y menos aéreas y con una mayor definición y rango de los armónicos.
- En las *vocales “o”, “a” e “i” antes* de aplicar la “Pantalla de celofán”, de forma general el número de formantes donde se concentra la mayor parte de la energía (amplitud de onda) sonora es menor. Además, hay menor regularidad de la envolvente del espectro, encontrándose los picos menos definidos y por tanto habiendo menor cantidad de armónicos. En este caso la variación del espectro de una trama con respecto a las tramas contiguas son mayores y la presencia de ruido es mayor.
- Por otro lado en las *vocales “o”, “a” e “i” después* de aplicar el método, se tiene que la variación del espectro de una trama con respecto a las tramas contiguas son menores y los armónicos se encuentran más definidos y en consecuencia la presencia de ruido es menos significativa.

El análisis *cualitativo* en el *dominio temporal*, en el *dominio frecuencial por medio del espectrograma de banda estrecha*, en el *dominio espectral (la excitación glótica)* y en el *dominio cepstral (cepstrum de potencia)*, se ha realizado con el software de libre distribución *ESection*; siendo éste el programa de análisis más idóneo y exhaustivo para la representación de dichos dominios.

6.2. Análisis cuantitativo. Medidas

A partir de las grabaciones obtenidas, también se ha determinado realizar un estudio de las medidas de calidad de la voz de forma cuantitativa a través de ciertos parámetros que permiten analizar la señal de voz.

En el capítulo 3 del Título I, se vio que existen cuatro fenómenos físicos cuantificables que permiten caracterizar la calidad de la voz en diferentes dominios de representación de la señal de voz, existiendo diferentes medidas objetivas de calidad que cuantifican con mayor o menor precisión un mismo fenómeno físico. Los cuatro fenómenos físicos identificados, se describen a continuación junto con sus medidas de calidad de la voz pertinentes:

- *Estabilidad de la voz*: es la capacidad de un locutor de producir, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales. Este fenómeno físico se cuantifica a partir de medidas de estabilidad de la señal de voz. Así, en el dominio temporal tenemos la medida de perturbación de amplitud o *shimmer* caracteriza la variación del valor del pico positivo máximo de cada trama de voz en el dominio temporal. Se determina mediante la variación de la amplitud máxima entre tramas adyacentes.
- *Riqueza espectral*: se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora del tracto vocal con gran cantidad de componentes espectrales. Este fenómeno físico se cuantifica por medio del cálculo de la estabilidad de la frecuencia de *pitch* y por la presencia de armónicos con energía elevada en las diferentes bandas de frecuencias. Así, una medida de calidad de la voz perteneciente a este grupo es el *jitter*; que cuantifica la variación a lo largo del tiempo del valor del *pitch*. A través de la relación subarmónico-armónico se puede medir los *problemas de cierre*.
- *Presencia de ruido*: se define como la presencia de ruido glótico en la voz, durante la fonación de un sonido sonoro sostenido, debida a carencias en la fase de cierre de las cuerdas vocales. Este fenómeno físico se cuantifica por medio de la presencia de ruido no estacionario en la voz. A través del valor promedio del primer pico cepstral de los cepstrum derivados espacialmente se obtiene la medida del *ritmo de golpeo*.
- *Comportamiento no lineal*: este fenómeno físico es ocasionado por un funcionamiento anómalo de las cuerdas vocales, debido a *irregularidades en las masas* que intervienen en la fase de cierre de las cuerdas. También se debe a la existencia de asimetrías en el movimiento de las cuerdas y factores relacionados con la mucosa.

En el protocolo de *cuantificación de la calidad de la voz*, se indica la necesidad de medir estas cuatro características o fenómenos de la voz, cada una de ellas cuantificada por una única medida. En la Tabla 6.1 se indican las medias utilizadas para cuantificar cada una de las cuatro características.

Parámetro	Medida
Estabilidad de la Voz	Cociente de perturbación de amplitud (APQ)
Problemas de Cierre	Promedio de la relación subarmónico-armónico
Ritmo de Golpeo	Valor promedio del primer pico cepstral de los cepstrum derivados espacialmente
Irregularidades en las Masas	Promedio del área inferior del índice de coherencia integrado

Tabla 6.1.- Medidas para cuantificar los distintos fenómenos físicos implicados en la medida de la calidad de la voz

Los cuatro parámetros identificados, asociados a los fenómenos físicos antes enumerados y que permiten realizar una medición objetiva de la calidad de la voz son:

- **Estabilidad de la voz:** Este fenómeno caracteriza el flujo de aire que es exhalado por los pulmones y que da lugar al movimiento vibratorio de las cuerdas vocales (movimientos de apertura y cierre). Para evaluar este fenómeno se analiza la capacidad de un locutor para producir, durante la fonación de un sonido sonoro sostenido, un flujo de aire con una intensidad constante.
- **Problemas de cierre:** Este fenómeno caracteriza el contacto de ambas cuerdas durante la fonación. En el caso de no existir un correcto cierre de las mismas se identifica la presencia de ruido en la señal de voz.
- **Ritmo de golpeo:** La estructura armónica de la señal de voz viene determinada por el ritmo y la fuerza del golpeo (el impacto entre las cuerdas vocales). La existencia de un ritmo irregular pone de manifiesto la presencia de determinadas patologías. Igualmente, si el golpeo no se produce con la suficiente intensidad, la voz producida presentará una pobre y anormal sonoridad.
- **Irregularidades en las masas:** Para realizar un correcto golpeo de las cuerdas resulta necesario que ambas cuerdas realicen de forma sincronizada las diferentes fases del movimiento vibratorio necesario que da lugar a la sucesiva apertura y cierre de las cuerdas. Este fenómeno cuantifica la descoordinación de las diferentes fases del movimiento entre cuerdas cuando estas presentan irregularidades. Además, cuantifica el deterioro de las propiedades visco-elásticas de la mucosa que recubre las cuerdas vocales.

Una voz de calidad anormal o inferior presenta al menos uno de los valores correspondientes a la cuantificación de los cuatro fenómenos físicos fuera de los rangos de normalidad.

Este protocolo de evaluación *cuantifica* la calidad de la voz permitiendo identificar calidades de voz anómalas de diferentes orígenes. A medida que la calidad de la voz de un locutor empeora, aumentará el número de fenómenos físicos y la desviación de los valores de la cuantificación de dichos fenómenos respecto a los valores de normalidad.

Se ha estimado los rangos preliminares de normalidad de las cuatro medidas de calidad de la voz, utilizando la base de datos del estudio y tomando el 95% central de los valores de la población de muestras sanas. En la Tabla 6.2 se muestran los rangos de normalidad de cada una de las medidas.

Para este análisis acústico cuantitativo se ha utilizado un programa específico desarrollado con el software *Matlab*; el cuál es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M) y servicio de especie y que está disponible para las plataformas Unix, Windows, Mac OS X y GNU/Linux.

Medida de Calidad de la Voz	Rango de Normalidad
Estabilidad de la Voz	[0.002 – 0.136]
Problemas de Cierre	[0.919 – 1.761]
Ritmo de Golpeo	[0.189 – 0.561]
Irregularidades en las Masas	[0.531 – 0.755]

Tabla 6.2.- Rango de normalidad de las diferentes medidas de la calidad de la voz

En las tablas siguientes se muestran los resultados del *análisis cuantitativo* para cada una de las 3 vocales (“o”, “a” e “i”) que han sido grabadas para cada locutor (20 locutores en total) *antes* y *después* de aplicar el método “Pantalla de celofán”.

En las mismas tablas se puede comprobar de forma general la mejoría obtenida en las vocales (“o”, “a” e “i”) grabadas *después* de aplicar el método pertinente. Se verifica que realizando el análisis cuantitativo de las vocales, **las cuatro medidas de calidad de la voz experimentan una mejoría en las vocales que fueron grabadas después de aplicar el método “Pantalla de celofán”.**

✓ **Vocal “O”:** Análisis cuantitativo antes y después de aplicar el método

Vocal "O"	ANTES				DESPUÉS				Incremento (CANTIDAD)				Incremento (PORCENTAJE)				
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4	
Locutor 1	0,13	0,41	0,51	0,5	0,08	0,42	0,63	0,49	0,05	-0,01	-0,12	0,01	38,5%	-2,4%	-23,5%	2,0%	
Locutor 2	0,09	0,4	0,67	0,52	0,04	0,32	0,55	0,5	0,05	0,08	0,12	0,02	55,6%	20,0%	17,9%	3,8%	
Locutor 3	0,12	0,74	0,46	0,46	0,14	0,61	0,47	0,48	-0,02	0,13	-0,01	-0,02	-16,7%	17,6%	-2,2%	-4,3%	
Locutor 4	0,04	0,59	0,54	0,51	0,06	0,51	0,48	0,51	-0,02	0,08	0,06	0,00	-50,0%	13,6%	11,1%	0,0%	
Locutor 5	0,07	1,07	0,49	0,52	0,1	0,79	0,49	0,49	-0,03	0,28	0,00	0,03	-42,9%	26,2%	0,0%	5,8%	
Locutor 6	0,18	0,85	0,46	0,58	0,19	1,02	0,36	0,57	-0,01	-0,17	0,10	0,01	-5,6%	-20,0%	21,7%	1,7%	
Locutor 7	0,14	0,65	0,39	0,56	0,09	0,67	0,33	0,57	0,05	-0,02	0,06	-0,01	35,7%	-3,1%	15,4%	-1,8%	
Locutor 8	0,14	0,55	0,35	0,59	0,19	0,56	0,3	0,55	-0,05	-0,01	0,05	0,04	-35,7%	-1,8%	14,3%	6,8%	
Locutor 9	0,28	0,59	0,28	0,49	0,11	0,61	0,37	0,52	0,17	-0,02	-0,09	-0,03	60,7%	-3,4%	-32,1%	-6,1%	
Locutor 10	0,17	1,13	0,61	0,47	0,16	1,03	0,37	0,54	0,01	0,10	0,24	-0,07	5,9%	8,8%	39,3%	-14,9%	
Locutor 11	0,25	0,66	0,48	0,5	0,17	0,68	0,4	0,51	0,08	-0,02	0,08	-0,01	32,0%	-3,0%	16,7%	-2,0%	
Locutor 12	0,12	0,54	0,51	0,52	0,06	0,62	0,5	0,49	0,06	-0,08	0,01	0,03	50,0%	-14,8%	2,0%	5,8%	
Locutor 13	0,35	0,38	0,42	0,53	0,12	0,42	0,46	0,53	0,23	-0,04	-0,04	0,00	65,7%	-10,5%	-9,5%	0,0%	
Locutor 14	0,17	0,89	0,43	0,5	0,12	0,92	0,38	0,5	0,05	-0,03	0,05	0,00	29,4%	-3,4%	11,6%	0,0%	
Locutor 15	0,1	0,5	0,65	0,56	0,14	0,61	0,61	0,54	-0,04	-0,11	0,04	0,02	-40,0%	-22,0%	6,2%	3,6%	
Locutor 16	0,12	0,59	0,48	0,53	0,15	0,63	0,46	0,52	-0,03	-0,04	0,02	0,01	-25,0%	-6,8%	4,2%	1,9%	
Locutor 17	0,18	0,59	0,57	0,5	0,06	0,5	0,57	0,49	0,12	0,09	0,00	0,01	66,7%	15,3%	0,0%	2,0%	
Locutor 18	0,03	0,6	0,63	0,48	0,05	0,73	0,4	0,47	-0,02	-0,13	0,23	0,01	-66,7%	-21,7%	36,5%	2,1%	
Locutor 19	0,12	1,2	0,58	0,51	0,17	0,95	0,52	0,51	-0,05	0,25	0,06	0,00	-41,7%	20,8%	10,3%	0,0%	
Locutor 20	0,08	0,72	0,54	0,47	0,08	0,7	0,53	0,48	0,00	0,02	0,01	-0,01	0,0%	2,8%	1,9%	-2,1%	
Promedio (Cantidad):									Promedio (Porcentaje):								
0,030									5,80%								

LEYENDA:

Parámetro:	Rango de Normalización
M1: Estabilidad de la Voz	[0,002 - 0,136]
M2: Ritmo de Golpeo	[0,919 - 1,761]
M3: Problemas de Cierre	[0,189 - 0,561]
M4: Irregularidades en las Masas	[0,531 - 0,755]

Valores dentro del rango normalizado
 Valores fuera del rango normalizado
 Mejora de los parámetros de calidad de la voz dentro del rango (cantidad y porcentaje)
 Parámetros de calidad de la voz dentro del rango sin mejora (cantidad y porcentaje)

Tabla 6.3.- Análisis cuantitativo de la Vocal “O” antes y después de aplicar el método

✓ **Vocal “A”:** Análisis cuantitativo *antes* y *después* de aplicar el método

Vocal "a"	ANTES				DESPUÉS				Incremento (CANTIDAD)				Incremento (PORCENTAJE)				
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4	
Locutor 1	0,09	0,47	0,39	0,5	0,04	0,31	0,49	0,51	0,05	0,16	-0,10	-0,01	55,6%	34,0%	-25,6%	-2,0%	
Locutor 2	0,08	0,58	0,55	0,52	0,08	0,44	0,42	0,5	0,00	0,14	0,13	0,02	0,0%	24,1%	23,6%	3,8%	
Locutor 3	0,2	1,23	0,4	0,49	0,06	1,03	0,43	0,53	0,14	0,20	-0,03	-0,04	70,0%	16,3%	-7,5%	-8,2%	
Locutor 4	0,14	1,22	0,24	0,5	0,1	1,12	0,44	0,52	0,04	0,10	-0,20	-0,02	28,6%	8,2%	-83,3%	-4,0%	
Locutor 5	0,17	1	0,34	0,54	0,08	1,08	0,32	0,52	0,09	-0,08	0,02	0,02	52,9%	-8,0%	5,9%	3,7%	
Locutor 6	0,1	0,84	0,22	0,67	0,09	1,14	0,25	0,57	0,01	-0,30	-0,03	0,10	10,0%	-35,7%	-13,6%	14,9%	
Locutor 7	0,09	0,56	0,61	0,54	0,08	0,7	0,31	0,51	0,01	-0,14	0,30	0,03	11,1%	-25,0%	49,2%	5,6%	
Locutor 8	0,06	0,65	0,35	0,52	0,16	0,5	0,37	0,57	-0,10	0,15	-0,02	-0,05	-	23,1%	-5,7%	-9,6%	
Locutor 9	0,08	0,85	0,35	0,53	0,04	0,99	0,32	0,57	0,04	-0,14	0,03	-0,04	50,0%	-16,5%	8,6%	-7,5%	
Locutor 10	0,17	0,99	0,43	0,5	0,11	1,08	0,21	0,54	0,06	-0,09	0,22	-0,04	35,3%	-9,1%	51,2%	-8,0%	
Locutor 11	0,15	0,73	0,48	0,45	0,24	0,58	0,42	0,48	-0,09	0,15	0,06	-0,03	-60,0%	20,5%	12,5%	-6,7%	
Locutor 12	0,11	1,23	0,38	0,57	0,08	1,16	0,32	0,54	0,03	0,07	0,06	0,03	27,3%	5,7%	15,8%	5,3%	
Locutor 13	0,11	0,51	0,54	0,5	0,1	0,43	0,54	0,5	0,01	0,08	0,00	0,00	9,1%	15,7%	0,0%	0,0%	
Locutor 14	0,17	1,06	0,35	0,52	0,08	1,09	0,31	0,55	0,09	-0,03	0,04	-0,03	52,9%	-2,8%	11,4%	-5,8%	
Locutor 15	0,11	1,95	0,63	0,51	0,12	0,69	0,59	0,46	-0,01	1,26	0,04	0,05	-9,1%	64,6%	6,3%	9,8%	
Locutor 16	0,15	1,06	0,62	0,56	0,13	0,8	0,42	0,51	0,02	0,26	0,20	0,05	13,3%	24,5%	32,3%	8,9%	
Locutor 17	0,17	0,76	0,52	0,49	0,19	0,98	0,33	0,49	-0,02	-0,22	0,19	0,00	-11,8%	-28,9%	36,5%	0,0%	
Locutor 18	0,13	0,87	0,48	0,57	0,03	1,21	0,39	0,54	0,10	-0,34	0,09	0,03	76,9%	-39,1%	18,8%	5,3%	
Locutor 19	0,15	0,67	0,48	0,51	0,12	0,67	0,47	0,51	0,03	0,00	0,01	0,00	20,0%	0,0%	2,1%	0,0%	
Locutor 20	0,14	0,86	0,45	0,51	0,15	0,83	0,42	0,53	-0,01	0,03	0,03	-0,02	-7,1%	3,5%	6,7%	-3,9%	
Promedio (Cantidad):									Promedio (Porcentaje):								
0,025									12,92%								
0,063									3,76%								
0,052									7,25%								
0,003									0,08%								

LEYENDA:

Parámetro:	Rango de Normalización
M1: Estabilidad de la Voz	[0.002 - 0.136]
M2: Ritmo de Golpeo	[0.919 - 1.761]
M3: Problemas de Cierre	[0.189 - 0.561]
M4: Irregularidades en las Masas	[0.531 - 0.755]

- Valores dentro del rango normalizado
- Valores fuera del rango normalizado
- Mejora de los parámetros de calidad de la voz dentro del rango (cantidad y porcentaje)
- Parámetros de calidad de la voz dentro del rango sin mejora (cantidad y porcentaje)

Tabla 6.4.- Análisis cuantitativo de la Vocal “A” *antes* y *después* de aplicar el método

✓ **Vocal “I”:** Análisis cuantitativo *antes* y *después* de aplicar el método

Vocal “I”	ANTES				DESPUÉS				Incremento (CANTIDAD)				Incremento (PORCENTAJE)				
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4	
Locutor 1	0,14	0,38	0,47	0,52	0,28	0,39	0,54	0,51	-0,14	-0,01	-0,07	0,01	-100,0%	-2,6%	-14,9%	1,9%	
Locutor 2	0,04	0,44	0,44	0,62	0,04	0,43	0,42	0,55	0,00	0,01	0,02	0,07	0,0%	2,3%	4,5%	11,3%	
Locutor 3	0,13	0,7	0,61	0,5	0,06	0,91	0,42	0,5	0,07	-0,21	0,19	0,00	53,8%	-30,0%	31,1%	0,0%	
Locutor 4	0,16	0,66	0,53	0,48	0,11	0,75	0,39	0,54	0,05	-0,09	0,14	-0,06	31,3%	-13,6%	26,4%	-12,5%	
Locutor 5	0,12	1,18	0,32	0,59	0,07	1,07	0,44	0,52	0,05	0,11	-0,12	0,07	41,7%	9,5%	-37,5%	11,9%	
Locutor 6	0,14	0,73	0,37	0,5	0,14	0,6	0,33	0,54	0,00	0,13	0,04	-0,04	0,0%	17,8%	-10,8%	-8,0%	
Locutor 7	0,17	0,49	0,4	0,67	0,24	0,5	0,25	0,59	-0,07	-0,01	0,15	0,08	-41,2%	-2,0%	37,5%	11,9%	
Locutor 8	0,05	0,63	0,36	0,56	0,06	0,38	0,56	0,52	-0,01	0,25	-0,20	0,04	-20,0%	39,7%	-55,6%	7,1%	
Locutor 9	0,09	0,79	0,51	0,5	0,13	0,98	0,27	0,55	-0,04	-0,19	0,24	-0,05	-44,4%	-24,1%	47,1%	-10,0%	
Locutor 10	0,23	0,52	0,56	0,47	0,05	0,41	0,67	0,48	0,18	0,11	-0,11	-0,01	78,3%	21,2%	-19,6%	-2,1%	
Locutor 11	0,08	0,55	0,59	0,48	0,1	0,55	0,43	0,53	-0,02	0,00	0,16	-0,05	-25,0%	0,0%	27,1%	-10,4%	
Locutor 12	0,12	0,7	0,33	0,5	0,14	0,65	0,35	0,5	-0,02	0,05	-0,02	0,00	-16,7%	7,1%	-6,1%	0,0%	
Locutor 13	0,13	0,98	0,77	0,49	0,15	0,92	0,75	0,53	-0,02	0,06	0,02	-0,04	-15,4%	6,1%	2,6%	-8,2%	
Locutor 14	0,14	0,7	0,32	0,59	0,12	0,73	0,29	0,54	0,02	-0,03	0,03	0,05	14,3%	-4,3%	9,4%	8,5%	
Locutor 15	0,07	0,61	0,31	0,6	0,02	0,62	0,44	0,5	0,05	-0,01	-0,13	0,10	71,4%	-1,6%	-41,9%	16,7%	
Locutor 16	0,06	0,78	0,34	0,63	0,03	0,61	0,41	0,56	0,03	0,17	-0,07	0,07	50,0%	21,8%	-20,6%	11,1%	
Locutor 17	0,09	0,89	0,87	0,48	0,09	0,93	0,37	0,5	0,00	-0,04	0,50	-0,02	0,0%	-4,5%	57,5%	-4,2%	
Locutor 18	0,05	0,33	0,62	0,45	0,02	0,47	0,44	0,49	0,03	-0,14	0,18	-0,04	60,0%	-42,4%	29,0%	-8,9%	
Locutor 19	0,09	0,83	0,3	0,48	0,07	0,69	0,46	0,5	0,02	0,14	-0,16	-0,02	22,2%	16,9%	-53,3%	-4,2%	
Locutor 20	0,17	0,63	0,33	0,52	0,15	0,47	0,5	0,5	0,02	0,16	-0,17	0,02	11,8%	25,4%	-51,5%	3,8%	
Promedio (Cantidad):									Promedio (Porcentaje):								
0,010									8,60%								

LEYENDA:

Parámetro:	Rango de Normalización
M1: Estabilidad de la Voz	[0,002 - 0,136]
M2: Ritmo de Golpeo	[0,919 - 1,761]
M3: Problemas de Cierre	[0,189 - 0,561]
M4: Irregularidades en las Masas	[0,531 - 0,755]

Valores dentro del rango normalizado

Valores fuera del rango normalizado

Mejora de los parámetros de calidad de la voz dentro del rango (cantidad y porcentaje)

Parámetros de calidad de la voz dentro del rango sin mejora (cantidad y porcentaje)

Tabla 6.5.- Análisis cuantitativo de la Vocal “I” *antes* y *después* de aplicar el método

Después de realizar el análisis cuantitativo para cada una de las 3 vocales (/o/, /a/ e /i/) que han sido grabadas para cada locutor (20 locutores en total) *antes y después* de aplicar el método “Pantalla de celofán”, se deduce que:

- El número de parámetros de medida de calidad de la voz que se encuentran dentro del rango de normalidad es mayor en las grabaciones realizar después de aplicar el método “Pantalla de celofán”. En consecuencia, el número de parámetros de medida que se encuentra fuera del rango de normalidad es menor.
- Se produce una mejora de los parámetros de calidad de la voz que se encuentran dentro del rango establecido ya que los valores obtenidos para las grabaciones realizar después de aplicar el método disminuyen considerablemente y están próximos a sus mínimos fijados (**Nota:** Los valores de los parámetros de medida de calidad de la voz han de ser los mínimos posibles sin salirse del rango establecido, al tratarse de parámetros de medida de perturbación de la voz).

A modo de estadística, para cada una de las vocales a analizar (“o”, “a” e “i”), se ha realizado un promedio de las valores obtenidos en cantidad y en porcentaje de la mejora por cada uno de los valores de los parámetros de calidad de la voz por separado (**M1, M2, M3 y M4**) que afectan a las 20 grabaciones en su conjunto (siendo 1 grabación por cada locutor).

De tal estimación, se deduce cual es el valor medio del incremento o decremento de mejora que se obtiene para cada parámetro de calidad (**M1, M2, M3 y M4**) y que cómo se ha comentado anteriormente se obtiene del promedio de las 20 grabaciones realizadas para cada vocal (“o”, “a” e “i”).

Realizándose el estudio para cada caso o vocal (/o/, /a/ e /i/) por separado, se obtienen las siguientes conclusiones:

✓ **Vocal “O”:**

- Los cuatro parámetros de medida de calidad de la voz *después* aplicar el método “Pantalla de celofán” mejoran sus valores. Análogamente se observa que el parámetro **M3** (ritmo de golpeo) es el que obtiene mayor número de valores dentro de sus márgenes de calidad establecidos para la voz de calidad.

- En cuanto al promedio realizado para cada parámetro (**M1**, **M2**, **M3** y **M4**), nótese que los valores obtenidos han sido positivos para los cuatro parámetros **M1**, **M2**, **M3** y **M4**. Nótese el caso del parámetro **M3**, cuyo valor promedio en porcentaje es de **7,08%**, siendo este el parámetro que mayor nivel de mejora experimenta (valor promedio en porcentaje que se obtiene de los 20 valores obtenidos del parámetro **M3**; correspondiéndose cada valor al tanto por ciento de la mejora o empeoramiento que experimenta cada parámetro en cada locutor).
- ✓ **Vocal “A”:**
- Al igual que los dos casos anteriores se produce una mejoría en los cuatro parámetros de medida de calidad de la voz *después* aplicar el método. El parámetro **M3** (ritmo de golpeo) es el que obtiene mayor número de valores dentro de sus márgenes de calidad establecidos para la voz de calidad, observándose también que para el parámetro **M1** (estabilidad de la voz) se obtiene una gran mejoría en sus valores obtenidos *después* de aplicar el método para cada vocal /a/ correspondiente a cada locutor.
 - Del promedio realizado para cada parámetro (**M1**, **M2**, **M3** y **M4**) se obtiene que los valores son positivos, por tanto se deduce que existe un promedio de mejora de cada parámetro (en cantidad y porcentaje) para los 20 locutores en su conjunto. Se obtiene que el parámetro **M1**, cuyo valor promedio en porcentaje es de **12,92%**, es el que mayor nivel de mejora experimenta (valor promedio en porcentaje que se obtiene de los 20 valores obtenidos del parámetro **M1**; correspondiéndose cada valor al tanto por ciento de la mejora o empeoramiento que experimenta cada parámetro en cada locutor).
- ✓ **Vocal “T”:**
- Los valores obtenidos para los cuatro parámetros de calidad *después* aplicar el método “Pantalla de celofán” mejoran con respecto a los valores obtenidos *antes* de aplicar el método. Además, para el parámetro **M3** (ritmo de golpeo), los valores obtenidos son los que más se encuentran dentro de los márgenes fijados o establecido para una voz de calidad.
 - En cuanto al promedio realizado para cada parámetro (en cantidad y porcentaje), nótese que los valores obtenidos han sido positivos para **M1**, **M2** y **M4**; lo cual indica la mejoría efectuada en la calidad de la voz después de aplicarse el método “Pantalla de celofán”. Para el caso del parámetro **M3**, nótese que el valor promedio de mejora para el parámetro **M3** es de **0,031**, correspondiéndose en porcentaje a un **- 0,90 %** de promedio de mejora (correspondiéndose este valor al promedio de los 20 valores obtenidos de dicho parámetro correspondientes a los 20 locutores que realizaron la grabación de la vocal). Este porcentaje negativo para **M3** indica que la mejora (reducción del valor de este parámetro) no ha sido muy amplia para algún locutor en concreto o que en algún o en algunos casos o locutores simplemente no se ha experimentado mejoría para este parámetro de calidad de la voz.

Capítulo 7

Conclusiones

Una vez realizadas las grabaciones de las vocales (/o/, /a/ e /i/) *antes* y *después* de aplicar el método “Pantalla de celofán” para los 20 locutores o participantes que se han sugerido para este estudio, y su posterior edición y segmentación de los archivos de audio grabados a través del software grabador y editor de audio *Audacity*, se deducen las siguientes conclusiones.

Tras el análisis acústico *cualitativo* puede deducirse que existe mejoría en la calidad de la voz después de aplicar el método “Pantalla de celofán”. Así, se tiene que las voces antes del tratamiento se caracterizan por poco aéreas, especialmente caracterizadas por la presencia de ruido en altas frecuencias. Las voces después del tratamiento se caracterizan por una mayor definición de los armónicos, mayor rango de armónicos y una mayor definición de los formantes.

En este análisis acústico *cualitativo*, se ha podido constatar de forma gráfica las mejorías producidas en la voz a través del software correspondiente de libre distribución *Esection*, mediante la representación de la señal en el dominio temporal, del espectrograma de banda estrecha (dominio frecuencial), la excitación glótica (dominio espectral) y representación en el dominio cepstral.

En cuanto al análisis acústico *cuantitativo* se ha demostrado también de forma eficiente la mejora en la calidad de la voz, a través de las medidas de cuatro parámetros característicos de medida de calidad de la voz (*estabilidad de la voz, ritmo de golpeo, problemas de cierre e irregularidades en las cuerdas vocales*); tomándose estas medidas a través de un programa creado y editado en *Matlab* (herramienta de software matemático). Los resultados obtenidos para las vocales grabadas *después* de aplicar el método han sido muy satisfactorios y habiéndose reducido de forma considerable los valores de estos cuatro parámetros de medida de perturbación de la voz.

En cuanto a la aceptabilidad de los locutores o usuarios, la práctica de la “Pantalla de celofán” según su opinión, es que han experimentado una mayor facilidad a la hora de emitir y vocalizar las diferentes vocales que fueron grabadas después de haber hecho vibrar la pantalla de celofán.

Como conclusión final, considero que el método utilizado constituye una herramienta muy práctica, sencilla y útil para hacer que una voz pueda mejorar su calidad, con el simple hecho de pronunciar de forma sostenida una vocal durante unos segundos después de haber utilizado la pantalla de celofán.

El análisis acústico pertinente ha demostrado la fiabilidad de esta técnica ya que los resultados han sido satisfactorios y la voz después de aplicar el método ha mejorado su calidad de forma satisfactoria y notable. Por tanto, considero que este trabajo puede servir como ejemplo para seguir futuras líneas de trabajo e investigación sobre las mejoras en la calidad vocal y su posterior evaluación acústica.

Capítulo 8

Bibliografía

Las bases de consulta utilizadas para la realización del presente Trabajo Fin de Título han sido:

Libros

- [1] *Tratamiento de los problemas de la voz: nuevos enfoques*. Inés Bustos Sánchez. Ed.: CEPE, 1995.
- [2] *Tratamiento de los trastornos de la voz*. Murray Morrison, Linda Rammage, Ed.: Masson, 1996.
- [3] *Anatomía para la voz*. Blandine Calais-Germain, François Germain, Ed.: La liebre de marzo, 2014.
- [4] *La voz y el habla. Principios de educación y reeducación*. Luis Diego Cruz Conejo, Ed.: EUNED, 2000.
- [5] *Clinical Examination of Voice*, Hirano, M., New York, Springer-Verlag, 1981.
- [6] *Signal modeling techniques in speech recognition*, Picone, J. W., Proceedings of the IEEE. 1993.
- [7] *Digital Spectral Analysis with Applications*, S. Lawrence Marple, Jr., Ed.: Prentice Hall Signal Processing Series, 1987.

- [8] *Metodología del Análisis Acústico para la Evaluación Clínica de la Voz*, Jesús B. Alonso Hernández, Universidad de Las Palmas de Gran Canaria, 2006.
- [9] *Automatic Detection of pathologies in the voice by Hos based parameters*, Alonso, J. B.; de León, J.; Alonso, I.; Ferrer, M. A., 2001.
- [10] *Terapéutica de los trastornos vocales (Tomo 4)*. Le Huche, F. y Allali, A. (2003), Editorial: Masson, 2003.

Fuentes de Internet

A continuación se indican las principales direcciones de Internet que han sido consultadas para la realización de este trabajo.

- [11] <http://www.ehu.es/acustica/espanol/musica/vohues>
La voz humana. Universidad del País Vasco.
- [12] <http://centrodeartigo.com/articulos-enciclopedicos>
Música vocal. La voz humana y su clasificación. Centro de Artigos.
- [13] www.psicopedagogia.com/patologias-lenguaje
Patologías del lenguaje. Psicopedagogía.
- [14] <http://cirugia.uah.es/pregrado/docs>
Alteraciones de la voz y el lenguaje. Departamento de Cirugía de la Universidad de Alcalá
- [15] <http://definicion.de/laringe/>
Definición de Laringe.
- [16] <http://www.otorrinoweb.com/>
Anatomía de la laringe.
- [17] <http://es.scribd.com/doc/129496431/Fisiologia-de-Faringe-y-Traquea>
Fisiología de la faringe y de la tráquea. Scribd.
- [18] <http://www.laringeyvoz.com/laringeyvoz.htm>
Laringe y voz.

II.- PRESUPUESTO



II.- PRESUPUESTO

En este apartado se va a presentar una estimación detallada de los costes de realización de este trabajo de fin de grado. Estos costes se presentan agrupados en diversas categorías.

Los grupos en los que se han dividido los costes son:

- Costes del personal involucrado en el proyecto (recursos humanos).
- Costes de ingeniería.
- Costes de amortización
- Otros costes.

Para el cálculo del presupuesto se deben tener en cuenta los siguientes aspectos:

- Los costes que se mostrarán estarán expresados en euros (€) a no ser que se especifique lo contrario.
- Se tomarán dos decimales para las cantidades económicas, redondeando en el caso de que sea necesario.

Además, al final del cálculo del presupuesto se describirán las herramientas hardware y los paquetes software que se han utilizado para la realización del proyecto, ya que en el coste de los equipos aparecen dichos elementos.

II.1.- Costes de recursos humanos

Para calcular el coste del personal se van a detallar las personas que han participado en el proyecto y el peso que han llevado en él. El coste por hora de cada persona del proyecto se define por un valor aproximado de lo que se está cobrando en la actualidad.

En la siguiente Tabla P.1, se muestra el ingeniero que ha realizado el proyecto, su categoría, su coste a la hora, el número de horas totales que han dedicado al proyecto y finalmente el coste total.

Nombre	Categoría	Euros/hora	Horas Totales	Coste Total
Francisco Javier Valerón Peñate	Ingeniero Técnico	22	300	6600 €
TOTAL:				6600 €

Tabla P.1.- Costes de recursos humanos

II.2.- Costes de ingeniería

Se ha invertido un total de 6 meses a tiempo completo en el desarrollo de este proyecto. Durante este tiempo se han desarrollado las tareas de formación, estudio y análisis, medidas, análisis de resultados y documentación; como muchas de estas tareas se solapan en el tiempo se han incluido los períodos de tiempo en función de la tarea predominante. En este presupuesto se ha incluido los costes de formación al ser específicos para el desarrollo de este proyecto. Tomando la tarifa por costes de ingeniería aplicada a un ingeniero junior en España, asciende a un total de 2.232,16 €/mes englobando en la misma el salario, la Seguridad Social, y el IRPF (Impuesto sobre la Renta de las personas físicas). La Tabla P.2 muestra los costes de ingeniería desglosados.

Descripción	Tiempo (meses)	Gastos (€/mes)	Gastos (€)
Formación y estudio	1	2232,16	2232,16
Realización del análisis acústico de las señales de voz	2	2232,16	4464,32
Realización del proceso de grabación y edición de los archivos de audio	2	2232,16	4464,32
Realización de la memoria	1	2232,16	2232,16
TOTAL			13392,96

Tabla P.2.- Costes de ingeniería

II.3.- Costes de amortización

En este apartado se exponen los costes derivados del uso de paquetes *software* y material *hardware*. Estos equipos *hardware* y paquetes *software* presentan un coste de amortización, en función del periodo de tiempo usado.

II.3.1.- Costes de recursos hardware

En la Tabla P.3 se detalla el coste total por amortización y mantenimiento para los recursos *hardware* utilizados. Para este tipo de equipos se considera el periodo de amortización de tres años.

Descripción	Precio (€)	Período de amortización (meses)	Tiempo de uso (meses)	Gastos (€)
Impresora Hewlett Packard Laserjet PRO CP 1025	180	18	6	60,00
Ordenador de sobremesa Medion® PC 8750	449	18	6	149,67
TOTAL				209,67

Tabla P.3.- Costes de recursos hardware

II.3.2.- Costes de recursos software

El cálculo de los gastos totales de los recursos *software* bajo licencia utilizados, tienen un coste estimado de amortización y mantenimiento según el período de uso (referenciados a un período de un año). Estos costes de licencias pertenecen a licencias universitarias, no siendo éstas comerciales y se detallan en la Tabla P.4.

Descripción	Precio (€)	Período de amortización (meses)	Tiempo de uso (meses)	Gastos (€)
Sistema operativo Microsoft Windows 8 Pro	59,99	12	6	30,00
Microsoft Office 2013 (Home and Student)	269	12	6	134,50
MATLAB Student Suite R2014a	69	12	6	34,50
TOTAL				199,00

Tabla P.4.- Costes por recursos software

A continuación, la Tabla P.5 muestra los gastos por costes de amortización.

Descripción	Gastos (€)
Costes por recursos hardware	209,67
Costes por recursos software	199,50
TOTAL	409,17

Tabla P.5.- Gastos por costes de amortización

II.4.- Otros costes

En este apartado se incluyen los costes debidos al uso de internet, material fungible y a la elaboración del documento final del proyecto. Se muestra en la Tabla P.6.

Descripción	Unidades	Coste unidad (€)	Gastos (€)
Paquetes papel DIN_A4 80 gr/m ²	3	4,5	13,50
CDs	3	3,00	9,00
Otros			100,00
TOTAL			122,50

Tabla P.6.- Otros costes

II.5.- Presupuesto total

Para finalizar, la Tabla P.7 se recoge el coste total del proyecto en función de los costes parciales comentados en las secciones anteriores.

Descripción	Gastos (€)
Costes de recursos humanos	6600
Costes de ingeniería	13392,96
Costes de amortización	409,17
Otros costes	122,50
PRESUPUESTO FINAL	20524,63
TOTAL (IGIC 7%)	21961,35

Tabla P.7.- Presupuesto total

D. Francisco Javier Valerón Peñate declara que el Trabajo Fin de Título “Análisis acústico de técnicas de reeducación de la voz” asciende a un total de 21961,35 euros.

Fdo.: Francisco Javier Valerón Peñate

DNI: 54079563-P

Las Palmas de Gran Canaria, a 09 de diciembre de 2014

II.6.- Herramientas hardware

En este apartado se especifican las herramientas hardware que se han utilizado para el desarrollo del trabajo de fin de grado y se definen las características del equipo utilizado en el proceso de grabación los archivos de audio.

- **Ordenador de sobremesa**

	Descripción: Medion® PC 8750
	Procesador Intel Core i3-3240
	Memoria RAM de 8 GB
	Disco duro de 500 GB
	Gráficos Intel HD 2500
Conexiones: USB 2.0, HDMI, entradas de audio y auriculares, conexión RJ-45 para el cable de internet y VGC para conectar el monitor.	

Tabla P.8.- Característica del ordenador de sobremesa

- **Impresora láser**

	Descripción: HP Laserjet PRO CP 1025
	Impresora láser
	Velocidad de impresión (Negro): Hasta 16 ppm
	Velocidad de impresión (Color): Hasta 4 ppm
	Potencia: 295 vatios
Fuente de alimentación: Red	

Tabla P.9.- Característica de la impresora láser

- **Micrófono y soporte**

Micrófono Audio Spectrum OSM 800 Platinum Series. Micrófono de condensador con una gran presencia de agudos, suaves medios y graves completos, este micrófono versátil proporciona una reproducción del sonido segura y natural. Es suficientemente flexible para admitir distintas fuentes de sonido lo que, junto a un precio atractivo, le hace ideal tanto para estudios caseros como profesionales. Completo, con un diafragma largo de 1" y un previo FET integrado, el OSM800 es una opción para voces o para un amplio rango de instrumentos acústicos, tanto en directo como en estudio. Incluye un interruptor bajo el cobertor, un soporte de montaje y estuche de transporte. Patrón cardioide. Frecuencia de respuesta: 20-20000KHz. Impedancia \leq 200Ohm a 1khz. Sensibilidad: -12mv/pa = 38dB (0dB = 1v/Pa). Conector: 3-Pin XLR. Necesita alimentación Phantom +48v.

Micrófono	
	Descripción: On-Stage Stands® OSM 800
	Patrón cardioide
	Respuesta de frecuencia: 20-20000 KHz
	Impedancia \leq 200 Ohm a 1kHz
	Sensibilidad: -12 mv/pa = 38 db (0 dB = 1 v/Pa)
	Conector: 3-Pin XLR
	Diafragma largo de 1" y un previo FET integrado
	Requisito de energía: + 48 v Phantom Power
	Interruptor de encendido/apagado
	Montaje de micrófono y estuche incluido
Soporte de araña para micrófono con pantalla anti-pop	
	Descripción: LD Systems® DSM 400
	Montaje de choque para micrófonos de estudio con filtro pop
	Pantalla doble nylon con marco de plástico negro resistente
	El montaje de choque diámetro: 43 – 49 mm
	Pantalla: Ø 130 mm
	

Tabla P.10.- Características del micrófono y del soporte

II.7.- Paquetes software

En este apartado se especifican los paquetes softwares que se han utilizado durante el proceso de grabación, edición y análisis de los archivos de audio correspondientes a las señales de voz de las vocales (/o/, /a/ e /i/) sometidas a estudio en este proyecto.

Se describen las características del software *Audacity*, que se ha utilizado para la grabación y segmentación de las señales de voz correspondientes a las vocales (/o/, /a/ e /i/) *antes* y *después* de aplicar el método “Pantalla de celofán” para cada uno de los 20 locutores o participantes que se han sugerido para este estudio del análisis acústico.

Además, se describen las características del software *Esection*, que se ha empleado para realizar el procesado y análisis acústico *cualitativo*; donde se realiza la representación de la señal en el dominio temporal, del espectrograma de banda estrecha (dominio frecuencial), la excitación glótica (dominio espectral) y representación en el dominio cepstral.

Para finalizar, se describen las principales características de la herramienta de software matemático *Matlab*; donde a través de un programa creado y editado con dicho software, se han obtenido los cuatro parámetros de medida de calidad de la voz (*estabilidad de la voz, ritmo de golpeo, problemas de cierre e irregularidades en las cuerdas vocales*); tomándose estas medidas a través de un programa creado y editado en *Matlab* (herramienta de software matemático).

▪ **Audacity**

Audacity es un editor de audio libre, de código abierto y multiplataforma. Es gratuito y fácil de usar, multi-pista de audio y grabador para Windows, Mac OS X, GNU / Linux y otros sistemas operativos. La interfaz está traducida a muchos idiomas. Puede usar Audacity para:

- ✓ Grabar audio en vivo.
- ✓ La reproducción en cualquier equipo de grabación de Windows Vista o posterior de la máquina.
- ✓ Convertir cintas y grabaciones a sonido digital o CD.
- ✓ Editar WAV, AIFF, FLAC, MP2, MP3 o archivos de sonido Ogg Vorbis.
- ✓ AC3, M4A / M4R (AAC), WMA y otros formatos soportados utilizando bibliotecas opcionales.
- ✓ Cortar, copiar, unir o mezclar sonidos.
- ✓ Numerosos efectos como el cambio de la velocidad o el tono de una grabación.

El software libre no es sólo libre de costo (como "cerveza gratis"). Es libre como en libertad (como "libertad de expresión"). El software libre le da la libertad de usar un programa, estudiar cómo funciona, mejorarlo y compartirlo con los demás.

Los programas como Audacity también se llaman software de código abierto, debido a que su código fuente está disponible para cualquier persona que estudie o uso. Hay miles de otros programas de código libre y abierto, incluyendo el navegador web Firefox, las suites de oficina LibreOffice o Apache OpenOffice y sistemas operativos basados en Linux como Ubuntu enteras.

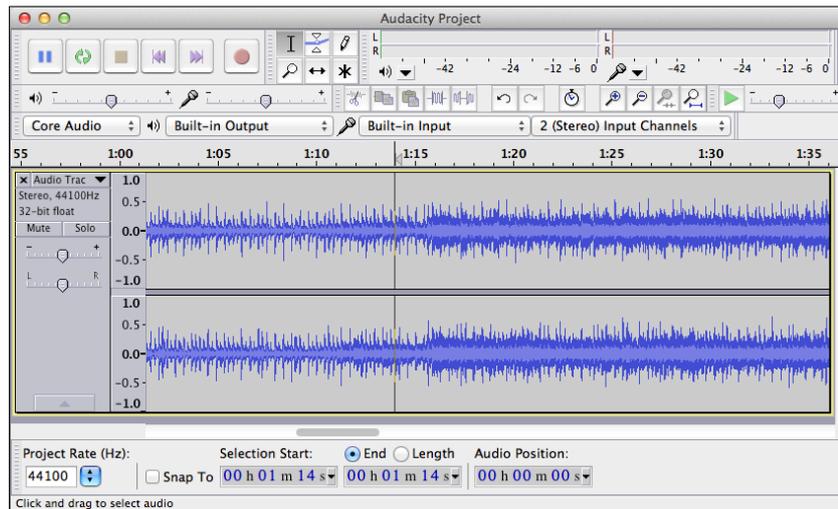


Figura P.1.- Rango de normalidad de las diferentes medidas de la calidad de la voz

▪ Esection

ESection es un programa para visualizar un número de diferentes análisis de frecuencia de una señal de voz. Para un tiempo particular en una grabación de voz, o en un rango dado de veces, el programa puede mostrar una sección transversal espectral, un espectro de predicción lineal, una autocorrelación o un análisis cepstral.

Esection puede mostrar la señal de voz de entrada como una forma de onda o un espectrograma, y estima automáticamente los valores fundamentales y de formantes de frecuencia para cada sección transversal.

ESection es un programa libre para mirar espectrales secciones transversales de las regiones de una señal de voz. Con *Esection* se puede mostrar una forma de onda o de un espectrograma de una señal de voz y mirar el espectro y otros análisis relacionados en un momento determinado o en una región particular de la señal. *ESection* es una sencilla aplicación que es completo en sí mismo, pero que también está diseñado para ser compatible con las herramientas del sistema de archivo de voz (SFS) para la investigación del habla.

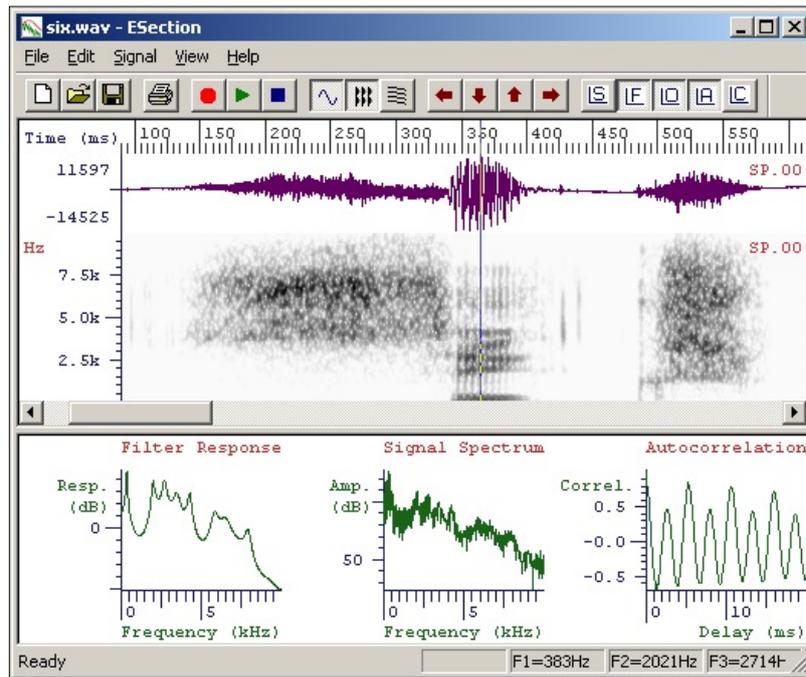


Figura P.2.- Rango de normalidad de las diferentes medidas de la calidad de la voz

▪ MATLAB

MATLAB (abreviatura de *MATrix LABoratory*, "laboratorio de matrices") es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M) y servicio de especie. Está disponible para las plataformas Unix, Windows, Mac OS X y GNU/Linux .

Entre sus prestaciones básicas se hallan: la manipulación de matrices, la representación de datos y funciones, la implementación de algoritmos, la creación de interfaces de usuario (GUI) y la comunicación con programas en otros lenguajes y con otros dispositivos hardware. El paquete *MATLAB* dispone de dos herramientas adicionales que expanden sus prestaciones, a saber, Simulink (plataforma de simulación multidominio) y GUIDE (editor de interfaces de usuario - GUI). Además, se pueden ampliar las capacidades de MATLAB con las *cajas de herramientas (toolboxes)*; y las de Simulink con los *paquetes de bloques (blocksets)*.

MATLAB es un programa de cálculo numérico diseñado para trabajar con matrices. Por tanto, será más eficiente si se diseñan los algoritmos en términos de matrices y vectores. Es un software muy usado en universidades y centros de investigación y desarrollo.

En los últimos años ha aumentado el número de prestaciones, como la de programar directamente procesadores digitales de señal o crear código VHDL. *MATLAB* se puede utilizar en una gran variedad de aplicaciones, tales como procesamiento de señales y comunicaciones, procesamiento de imagen y vídeo, sistemas de control, pruebas y medidas, finanzas computacionales y biología computacional. Más de un millón de ingenieros y científicos de la industria y la educación utilizan *MATLAB*, el lenguaje del cálculo técnico.

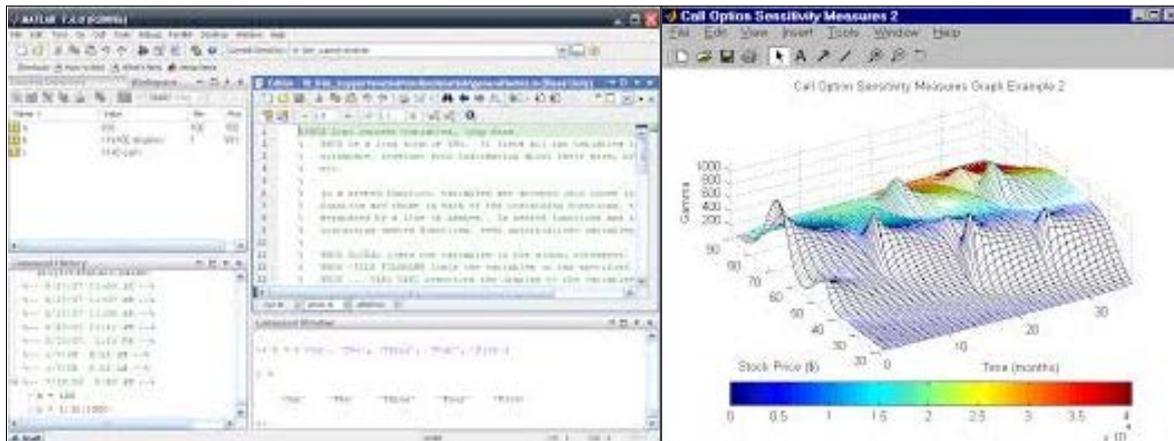


Figura P.3.- Rango de normalidad de las diferentes medidas de la calidad de la voz

III.- ANEXO:

Gráficas del Análisis Cualitativo



III.- ANEXO: Gráficas del Análisis Cualitativo

A continuación, se ilustran las gráficas correspondientes a cada una de las vocales /o/, /a/ e /i/ que han sido grabadas *antes* y *después* de aplicar el método “Pantalla de celofán” para cada uno de los 20 locutores participantes, y que han sido analizadas con el software de libre distribución *Esection*.

En el análisis acústico *cualitativo*, se analiza la señal de cada vocal grabada (/o/, /a/ e /i/) mediante la representación de la señal en el dominio temporal, del espectrograma de banda estrecha (dominio frecuencial), la excitación glótica (dominio espectral) y representación en el dominio cepstral.

❖ **Locutor 1:**

➤ *(Vocal “o” antes):*

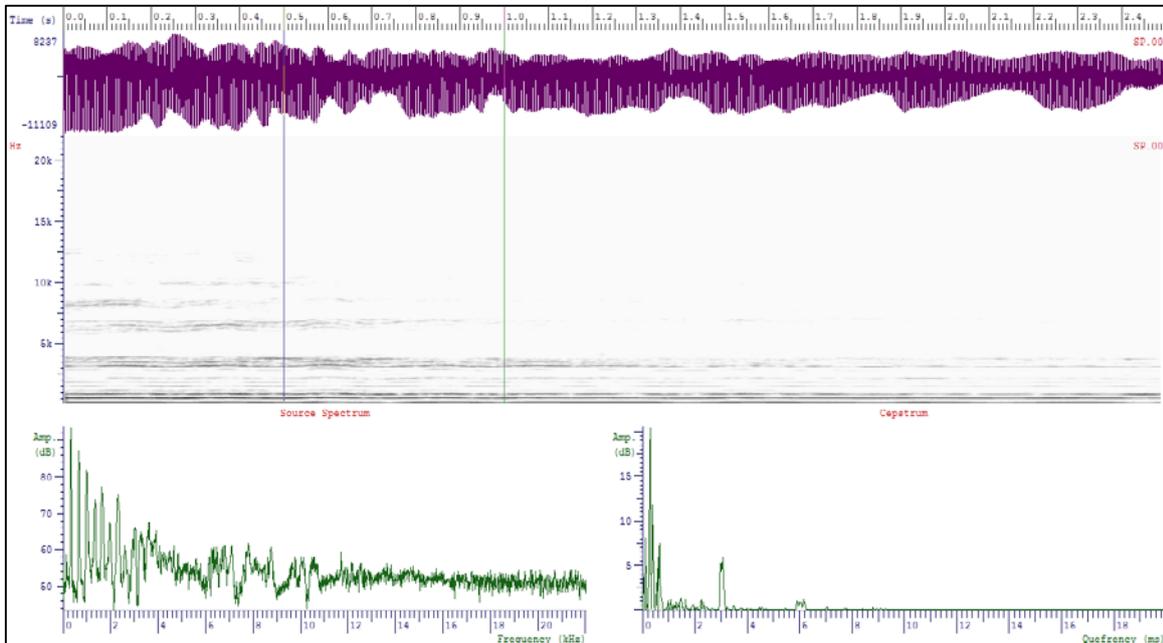


Fig. A.1.- Vocal “o” *antes* de aplicar el método para el Locutor 1

➤ *(Vocal “o” después):*

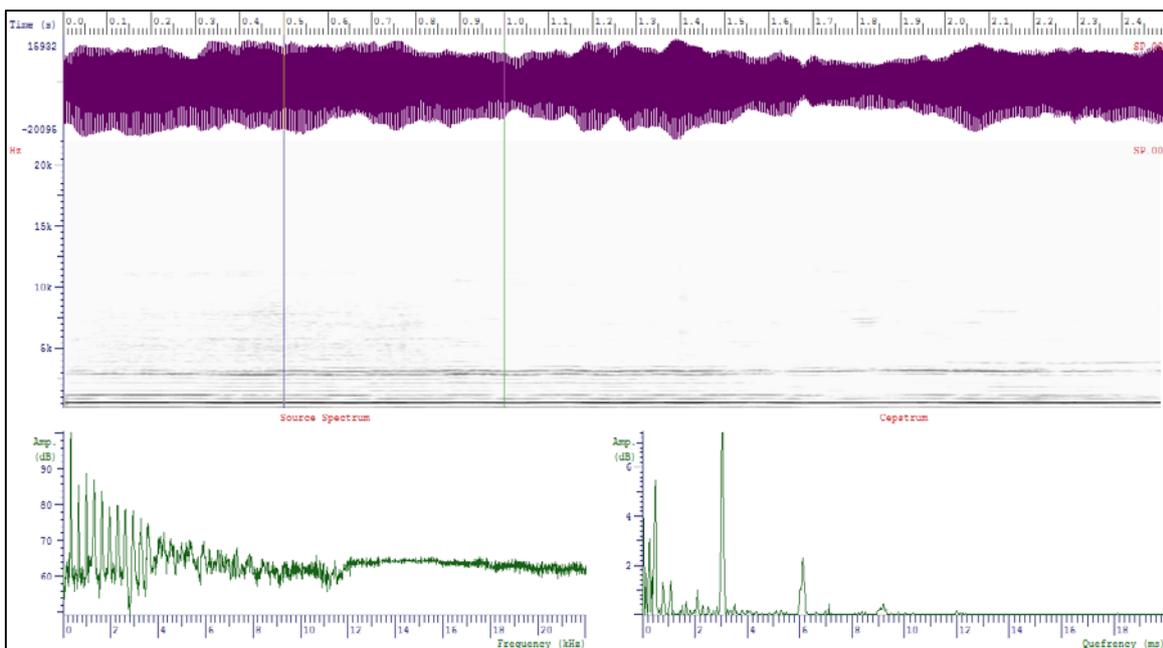


Fig. A.2.- Vocal “o” *después* de aplicar el método para el Locutor 1

➤ (Vocal “a” antes):

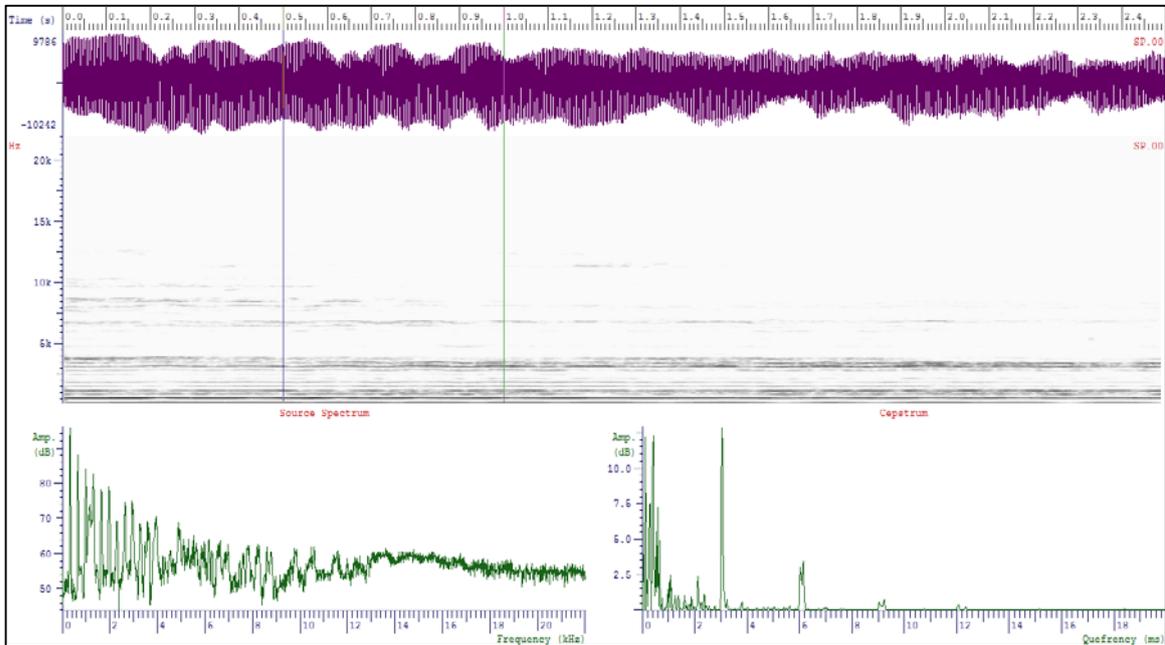


Fig. A.3.- Vocal “a” antes de aplicar el método para el Locutor 1

➤ (Vocal “a” después):

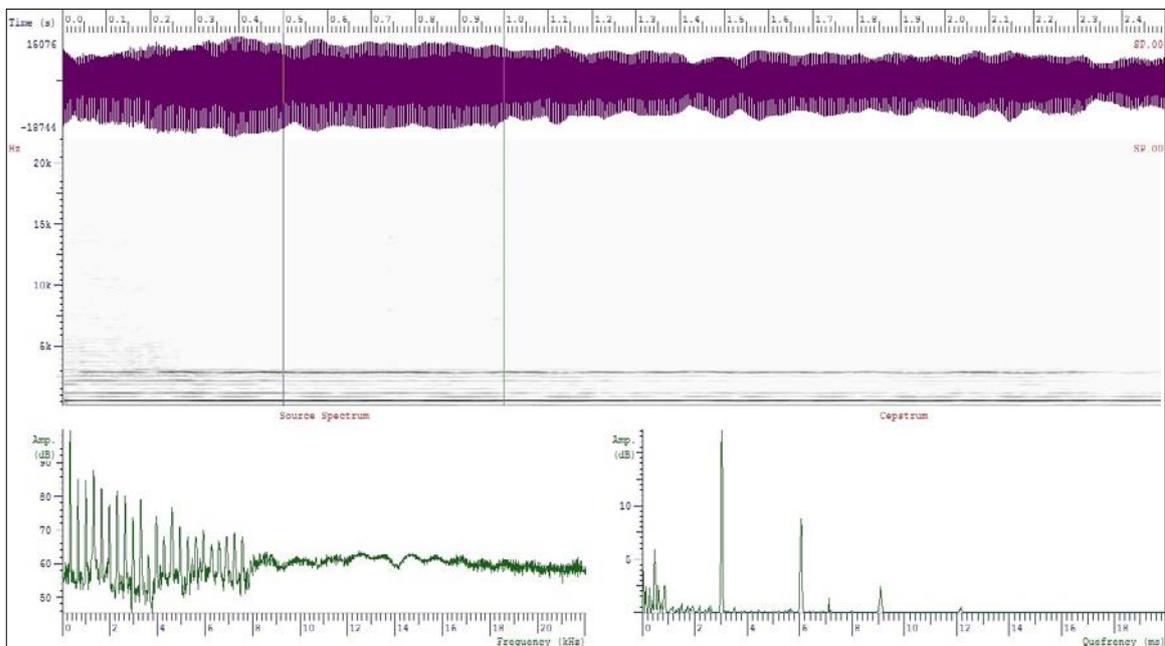


Fig. A.4.- Vocal “a” después de aplicar el método para el Locutor 1

➤ (Vocal “i” antes):

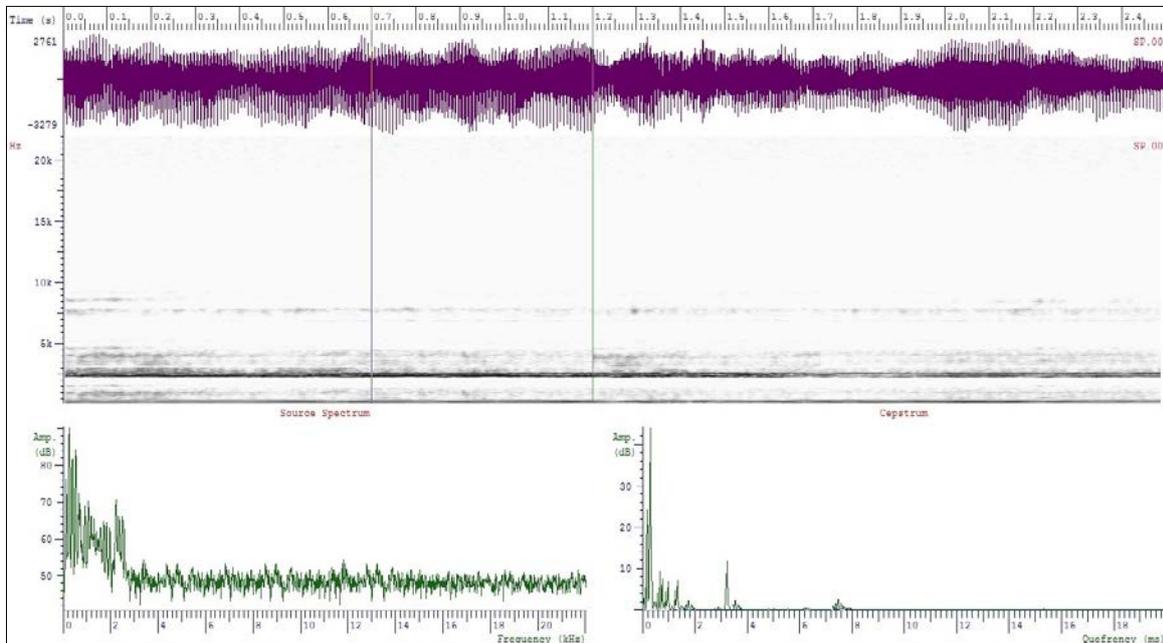


Fig. A.5.- Vocal “i” antes de aplicar el método para el Locutor 1

➤ (Vocal “i” después):

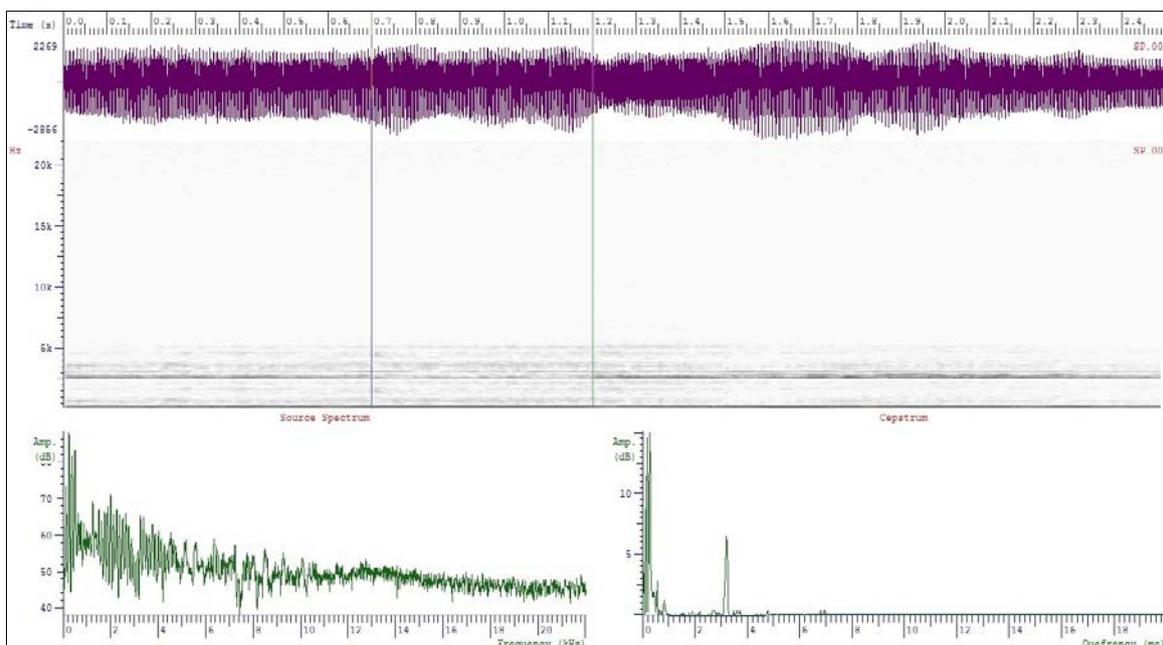


Fig. A.6.- Vocal “i” después de aplicar el método para el Locutor 1

❖ **Locutor 2:**

➤ *(Vocal “o” antes):*

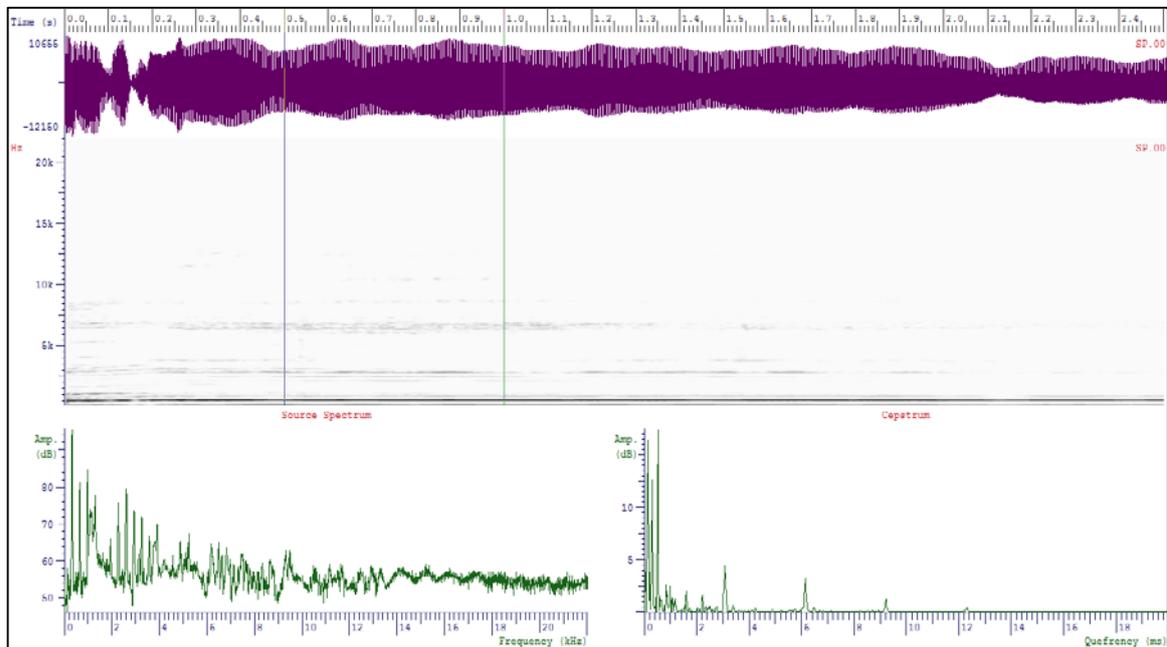


Fig. A.7.- Vocal “o” *antes* de aplicar el método para el Locutor 2

➤ *(Vocal “o” después):*

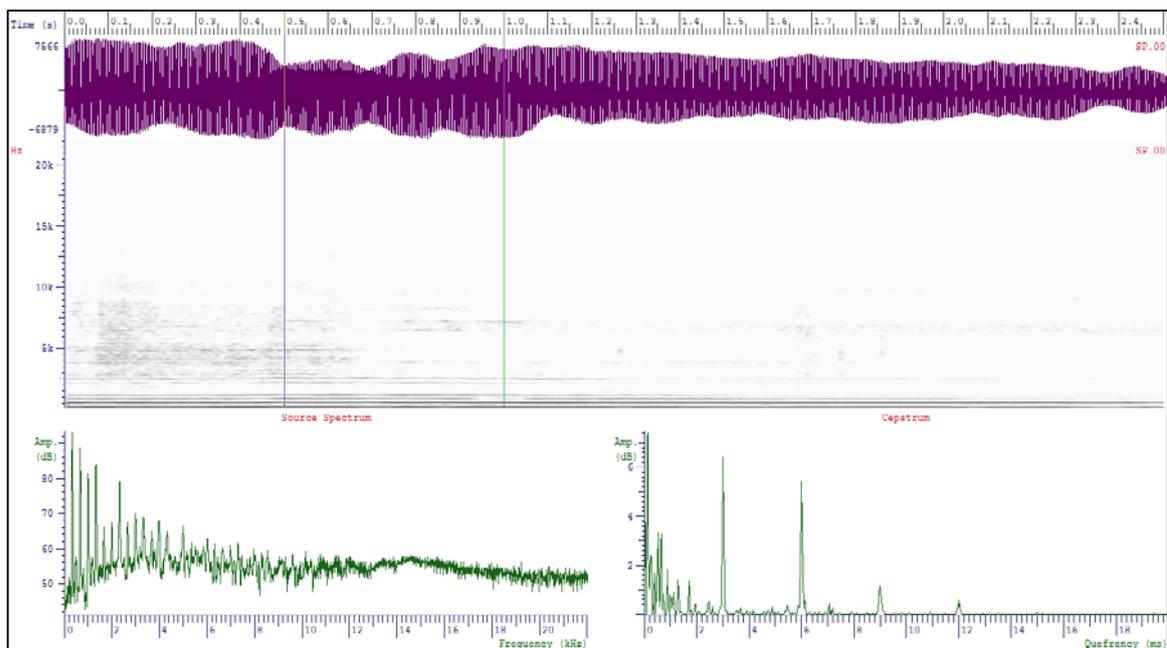


Fig. A.8.- Vocal “o” *después* de aplicar el método para el Locutor 2

➤ (Vocal “a” antes):

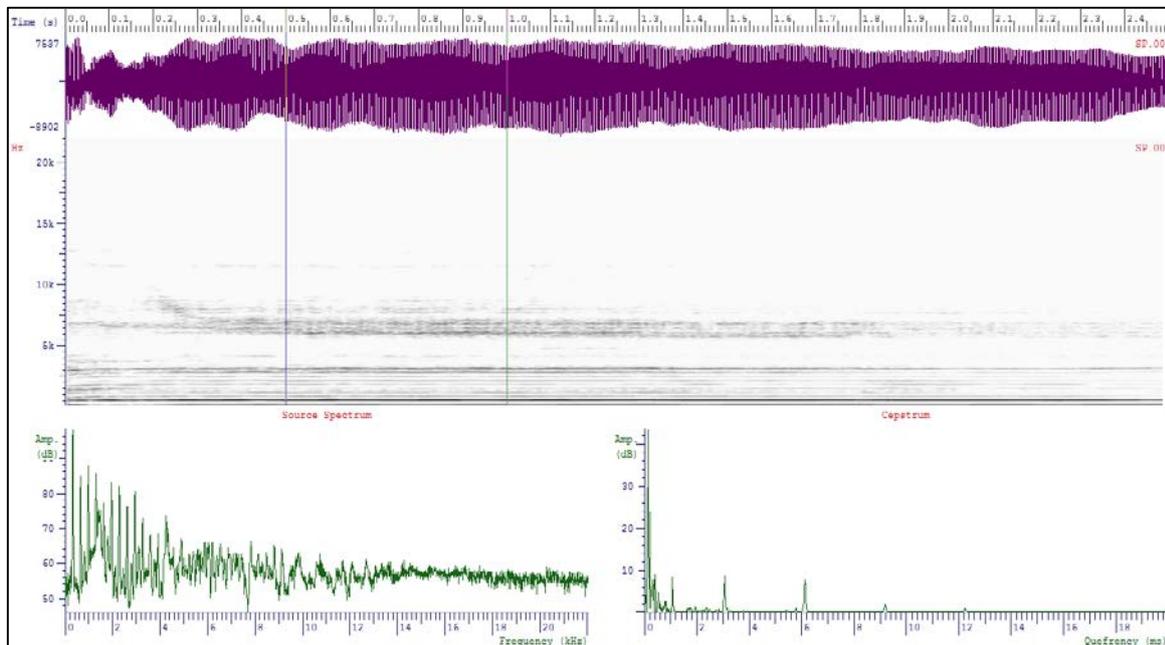


Fig. A.9.- Vocal “a” antes de aplicar el método para el Locutor 2

➤ (Vocal “a” después):

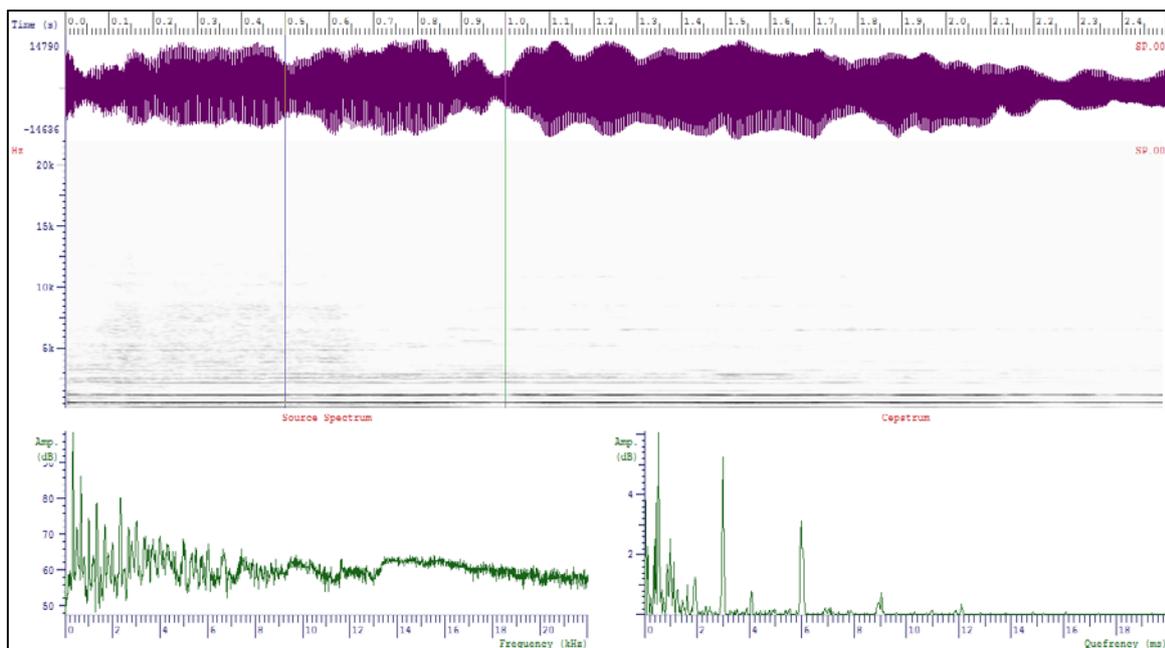


Fig. A.10.- Vocal “a” después de aplicar el método para el Locutor 2

➤ (Vocal “i” antes):

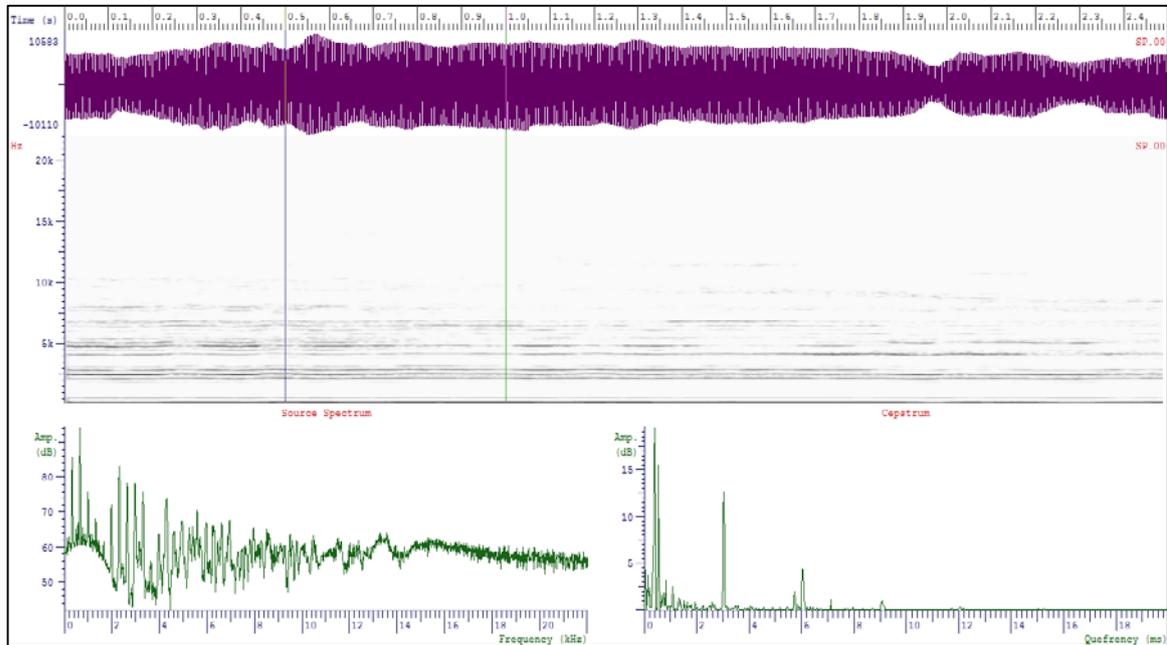


Fig. A.11.- Vocal “i” antes de aplicar el método para el Locutor 2

➤ (Vocal “i” después):

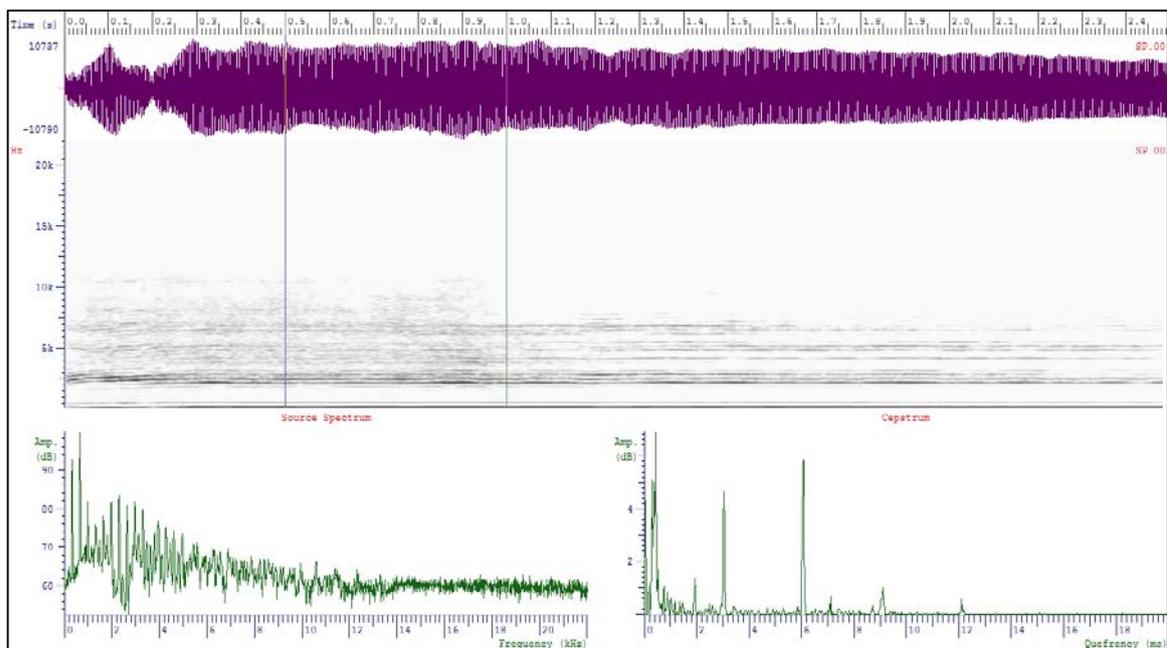


Fig. A.12.- Vocal “i” después de aplicar el método para el Locutor 2

❖ **Locutor 3:**

➤ *(Vocal “o” antes):*

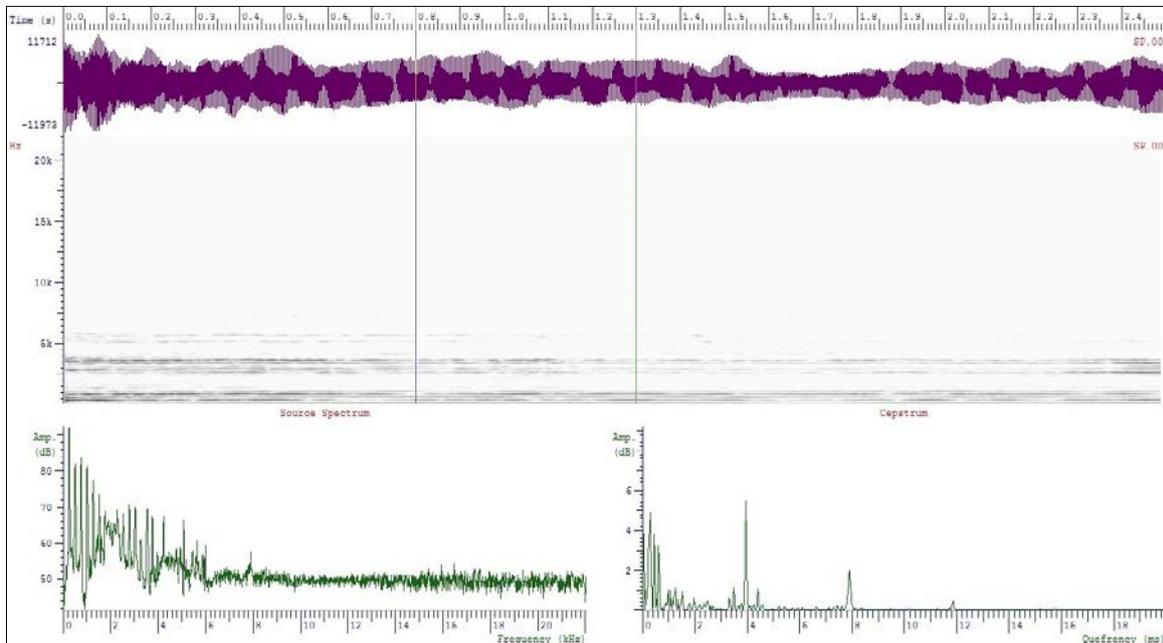


Fig. A.13.- Vocal “o” *antes* de aplicar el método para el Locutor 3

➤ *(Vocal “o” después):*

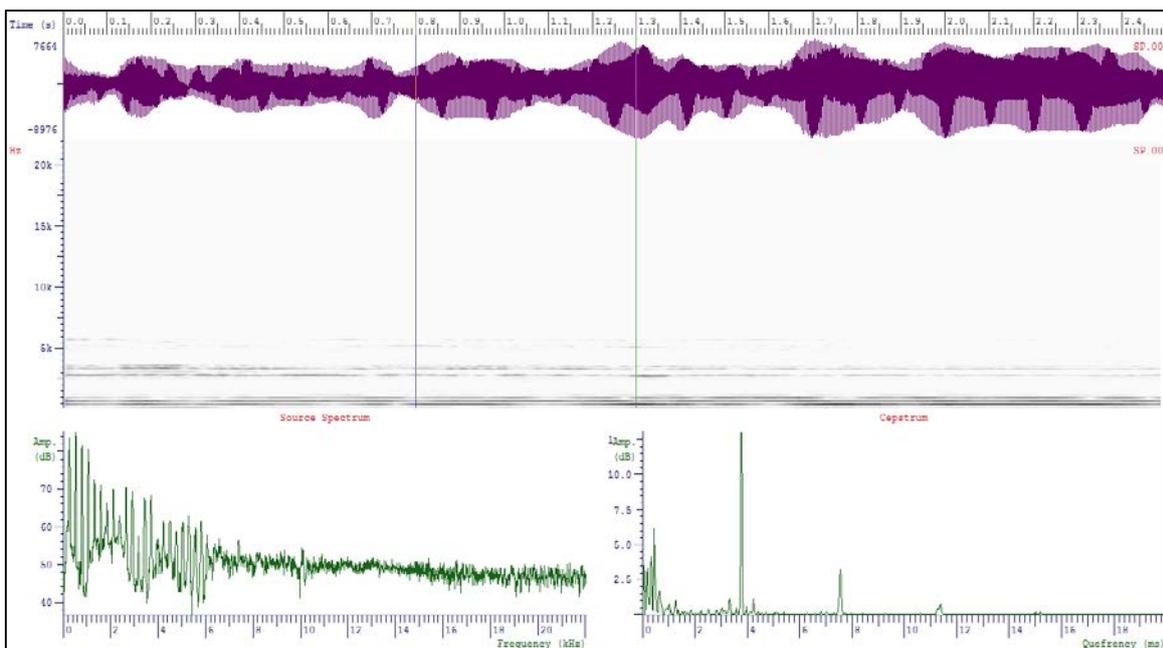


Fig. A.14.- Vocal “o” *después* de aplicar el método para el Locutor 3

➤ (Vocal “a” antes):

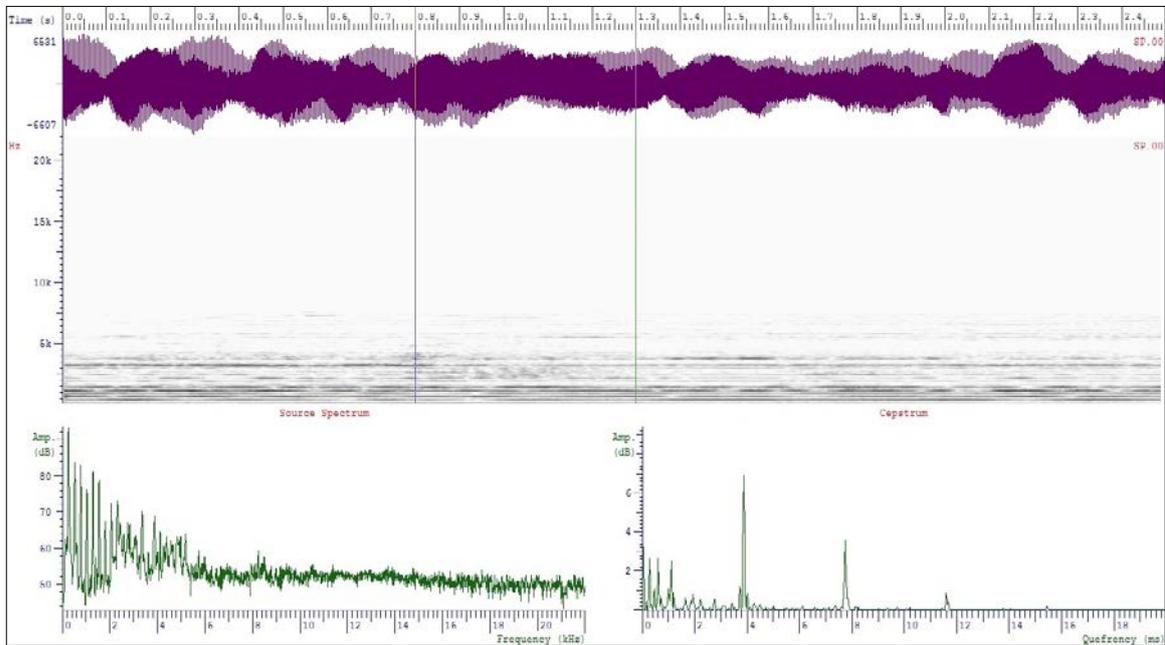


Fig. A.15.- Vocal “a” antes de aplicar el método para el Locutor 3

➤ (Vocal “a” después):

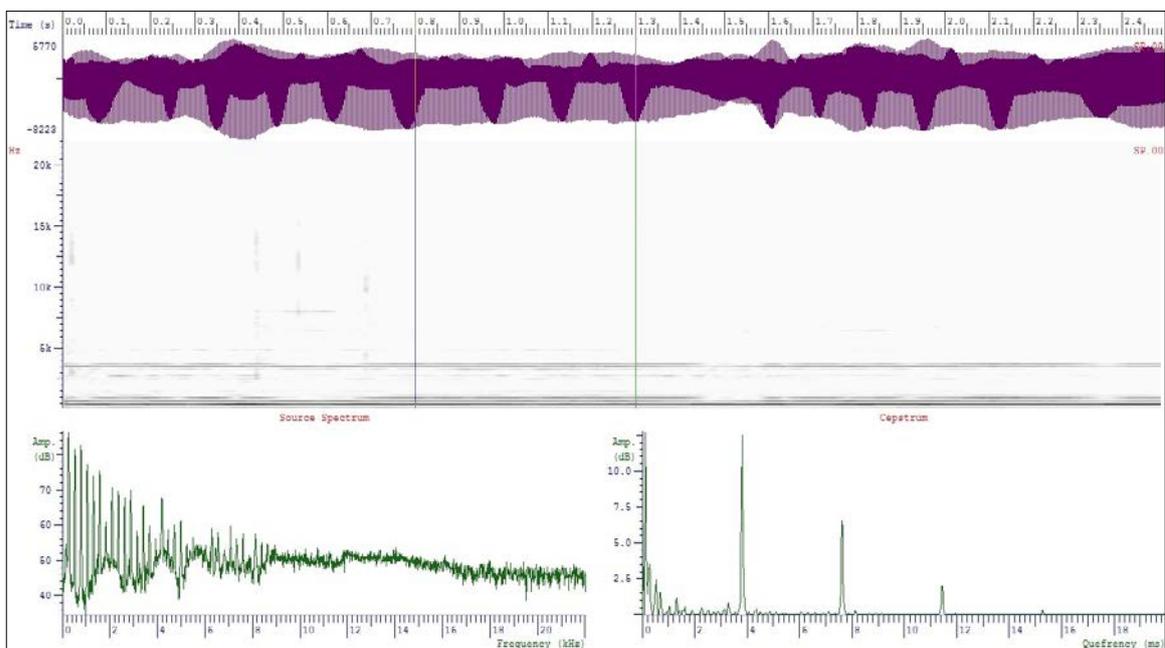


Fig. A.16.- Vocal “a” después de aplicar el método para el Locutor 3

➤ (Vocal “i” antes):

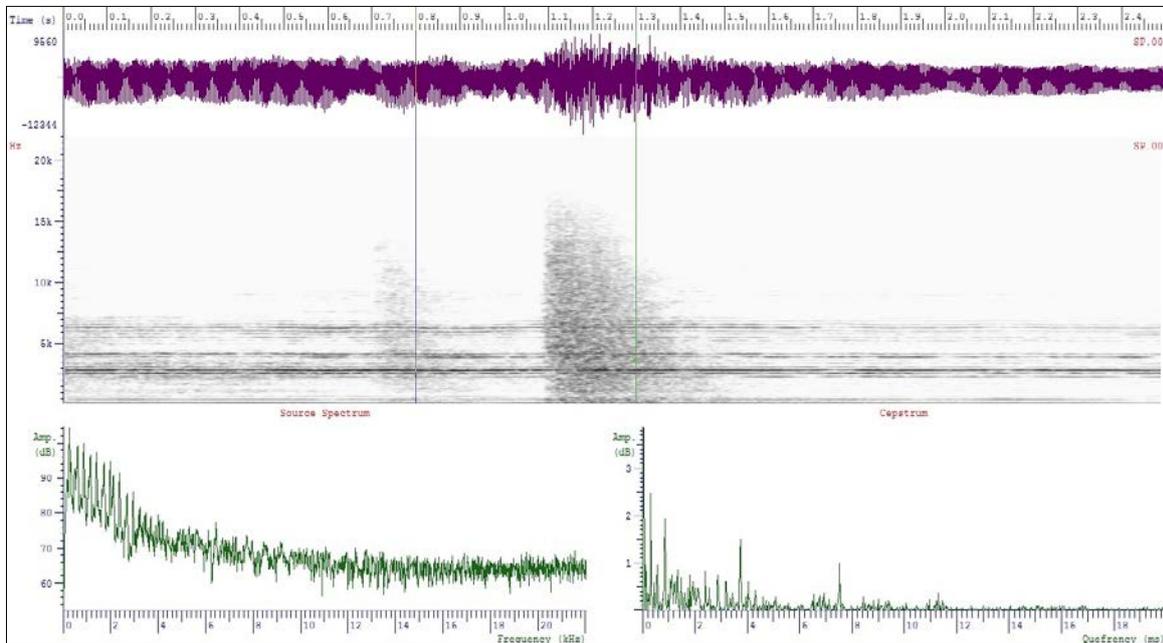


Fig. A.17.- Vocal “i” antes de aplicar el método para el Locutor 3

➤ (Vocal “i” después):

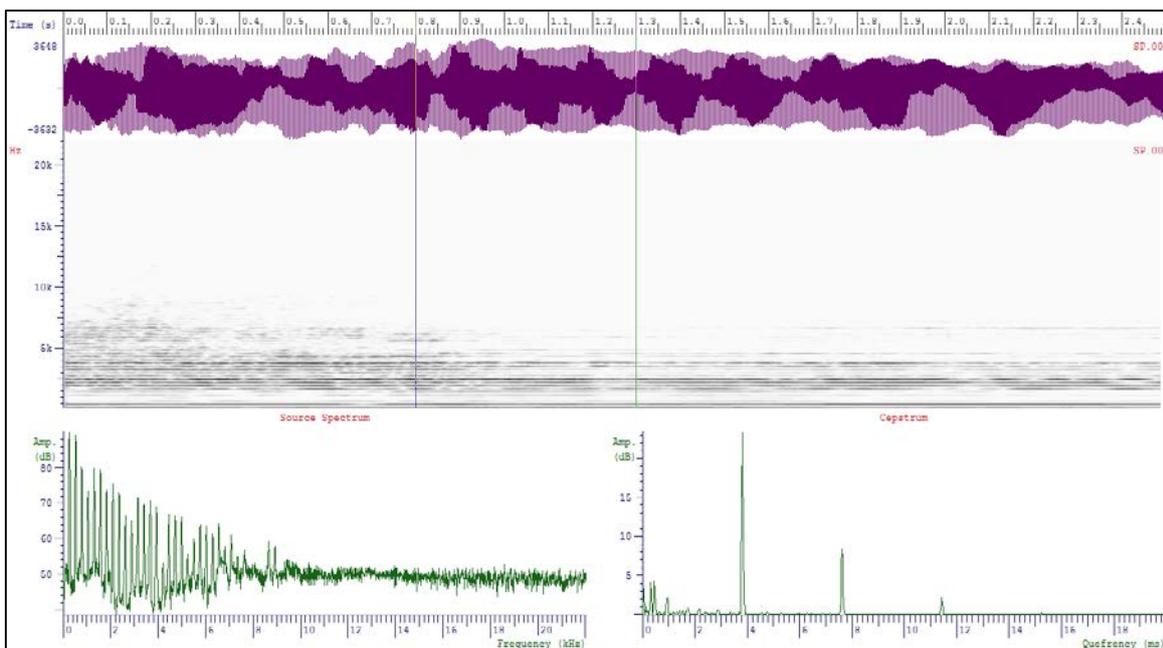


Fig. A.18.- Vocal “i” después de aplicar el método para el Locutor 3

❖ **Locutor 4:**

➤ *(Vocal “o” antes):*

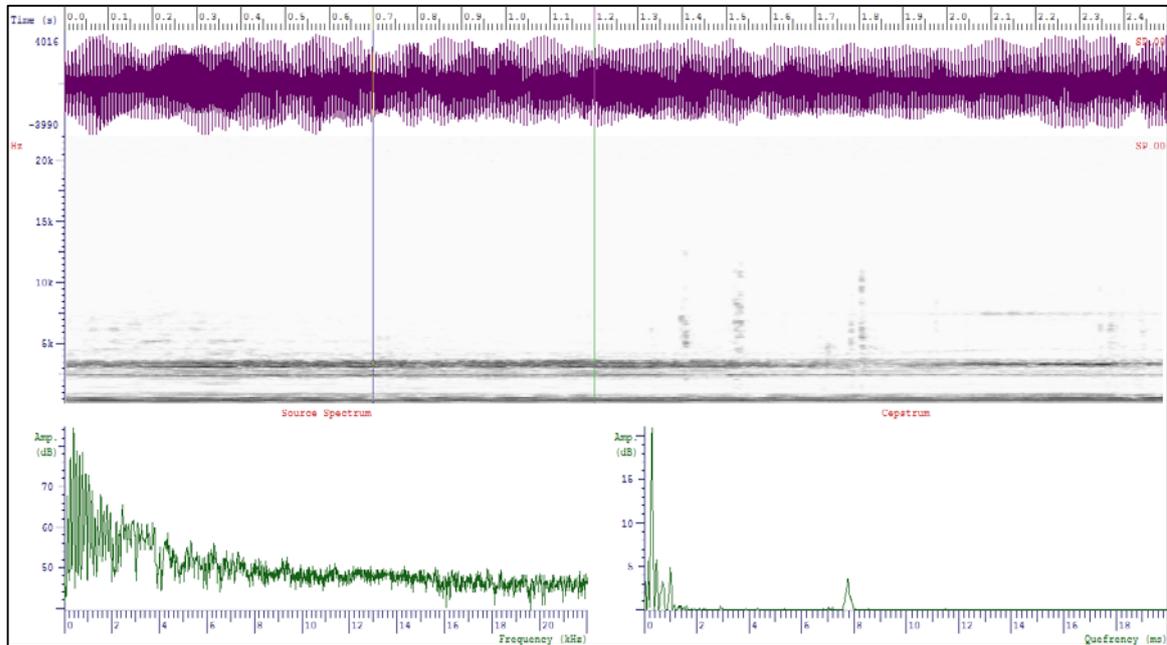


Fig. A.19.- Vocal “o” antes de aplicar el método para el Locutor 4

➤ *(Vocal “o” después):*

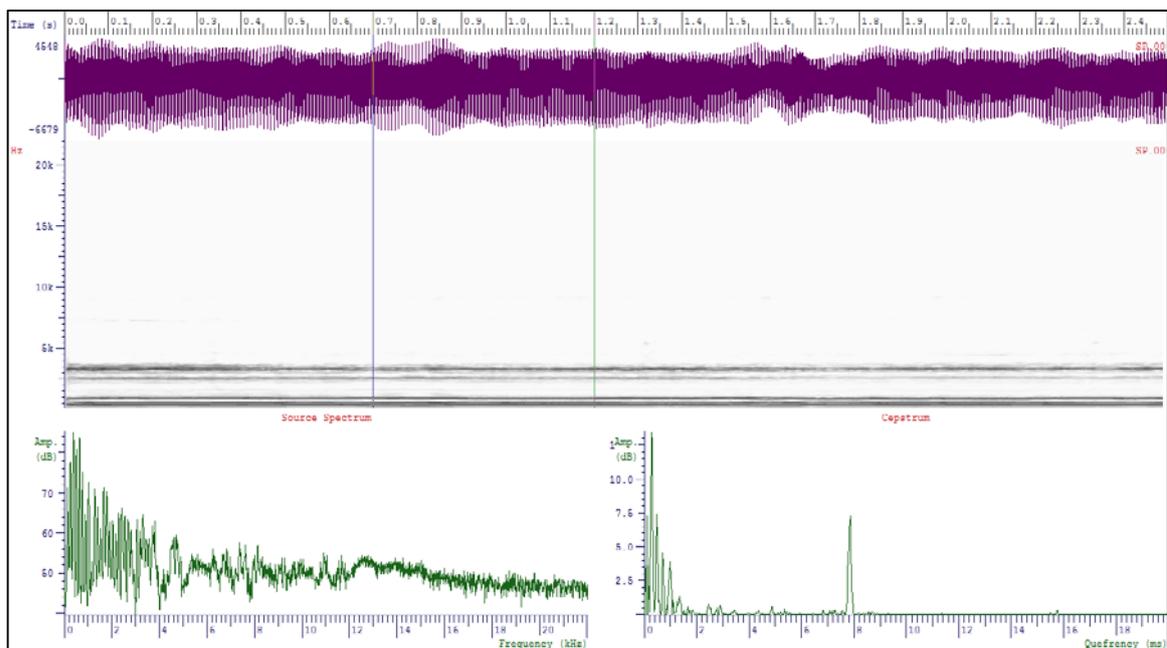


Fig. A.20.- Vocal “o” después de aplicar el método para el Locutor 4

➤ (Vocal “a” antes):

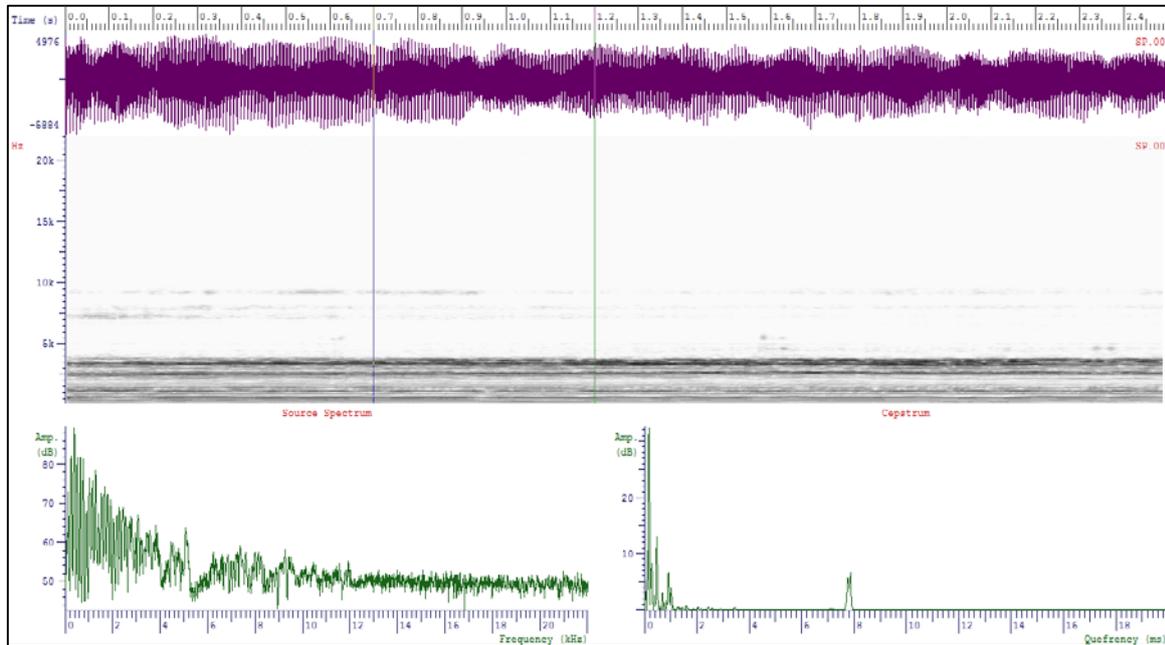


Fig. A.21.- Vocal “a” antes de aplicar el método para el Locutor 4

➤ (Vocal “a” después):

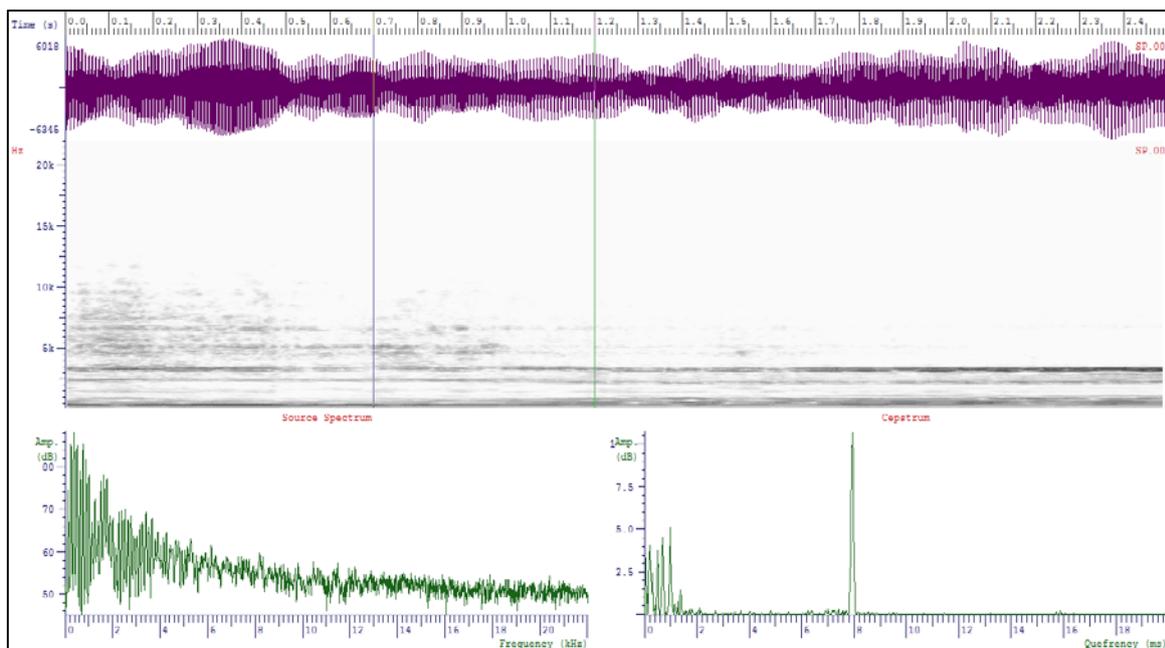


Fig. A.22.- Vocal “a” después de aplicar el método para el Locutor 4

➤ (Vocal “i” antes):

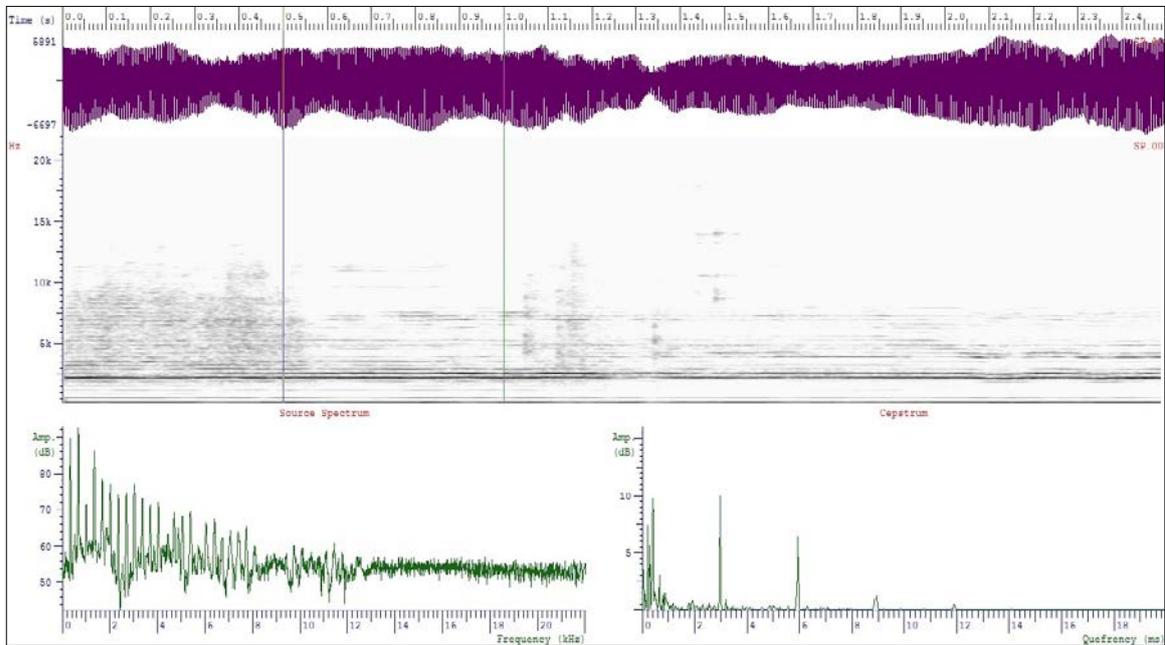


Fig. A.23.- Vocal “i” antes de aplicar el método para el Locutor 4

➤ (Vocal “i” después):

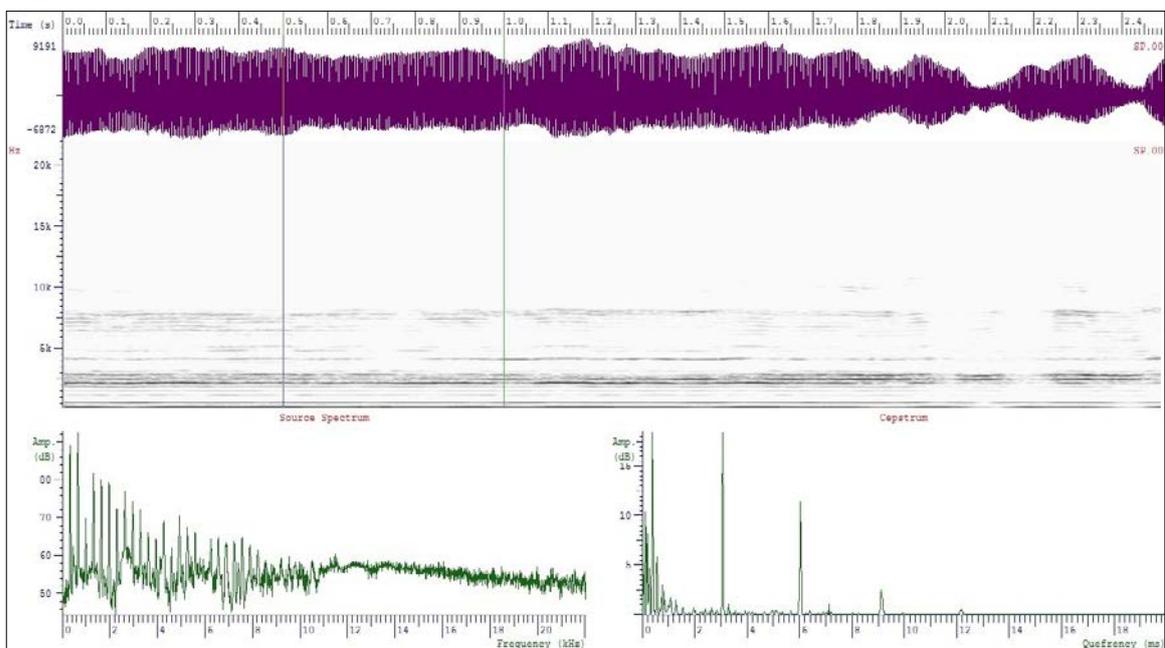


Fig. A.24.- Vocal “i” después de aplicar el método para el Locutor 4

❖ **Locutor 5:**

➤ *(Vocal “o” antes):*

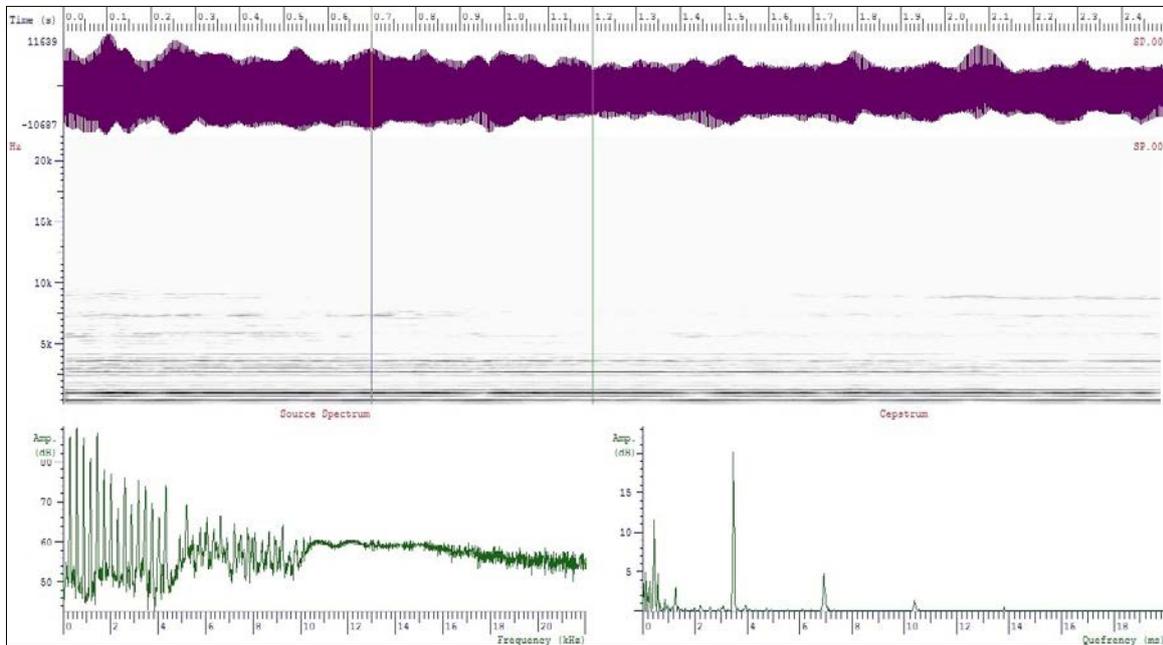


Fig. A.25.- Vocal “o” *antes* de aplicar el método para el Locutor 5

➤ *(Vocal “o” después):*

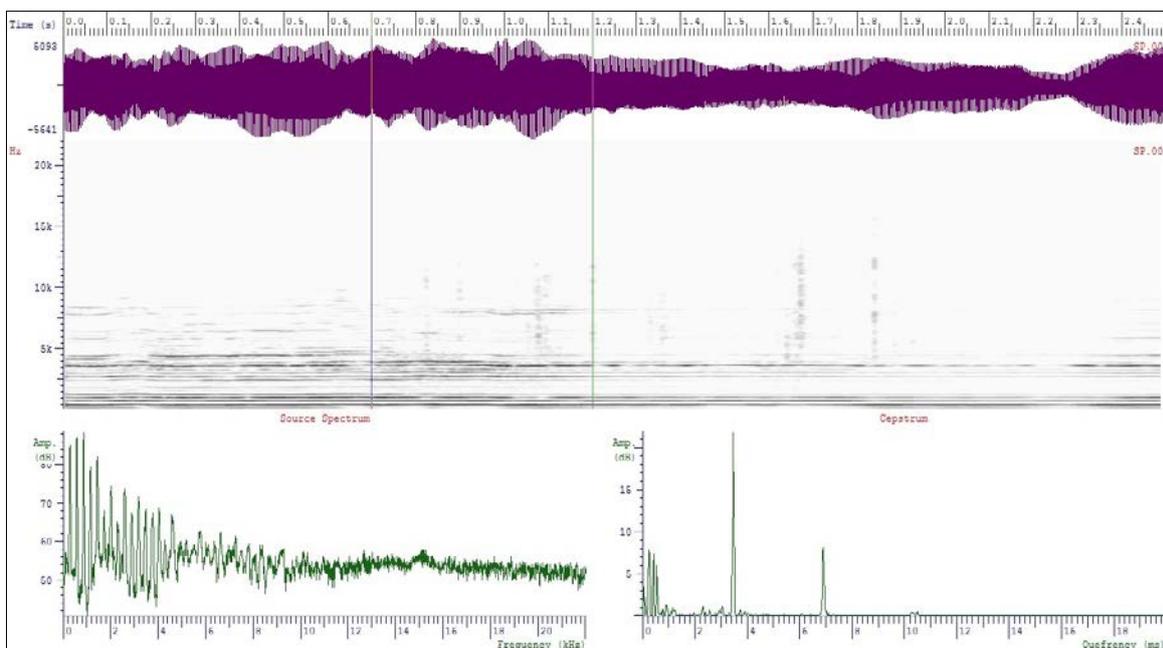


Fig. A.26.- Vocal “o” *después* de aplicar el método para el Locutor 5

➤ (Vocal “a” antes):

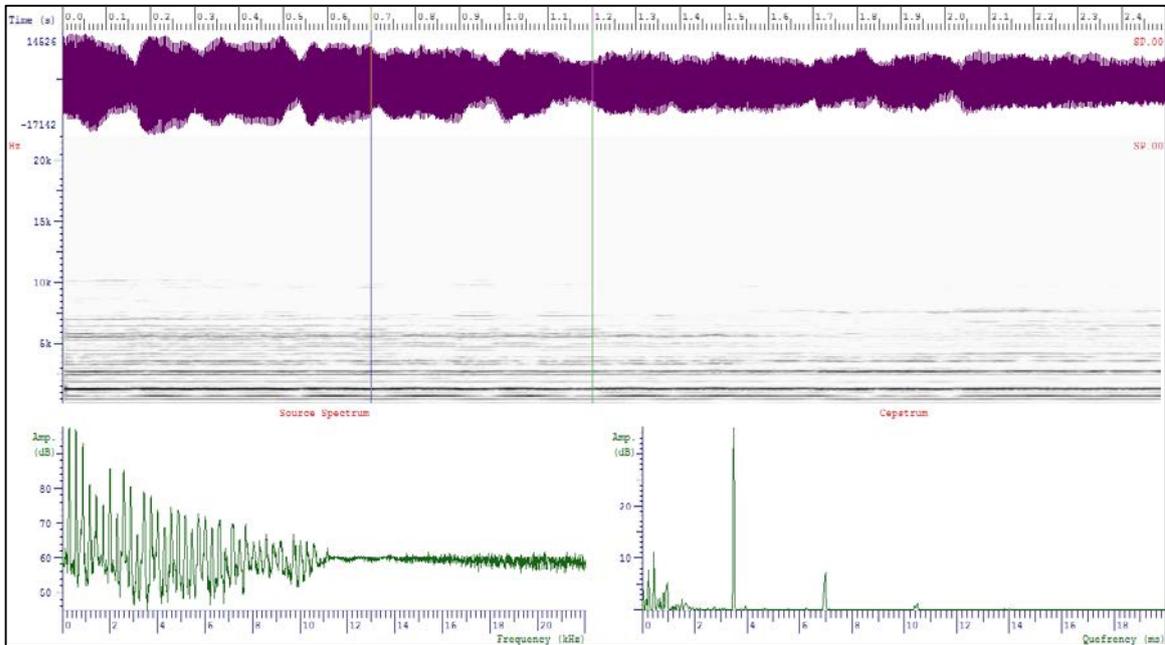


Fig. A.27.- Vocal “a” antes de aplicar el método para el Locutor 5

➤ (Vocal “a” después):

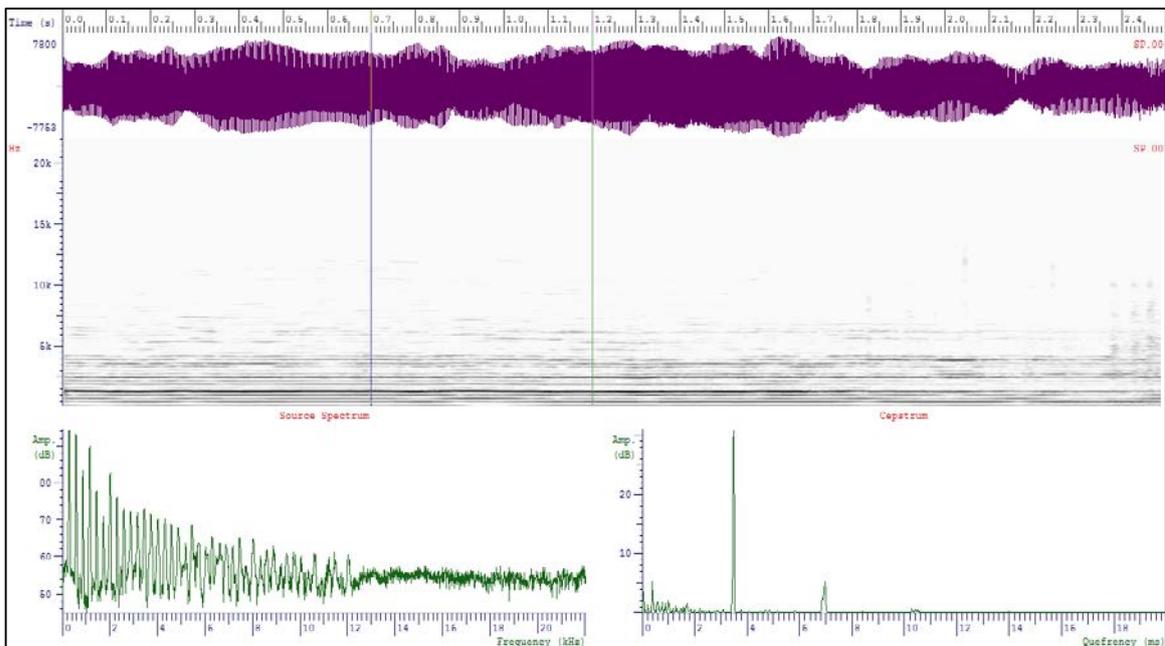


Fig. A.28.- Vocal “a” después de aplicar el método para el Locutor 5

➤ (Vocal “i” antes):

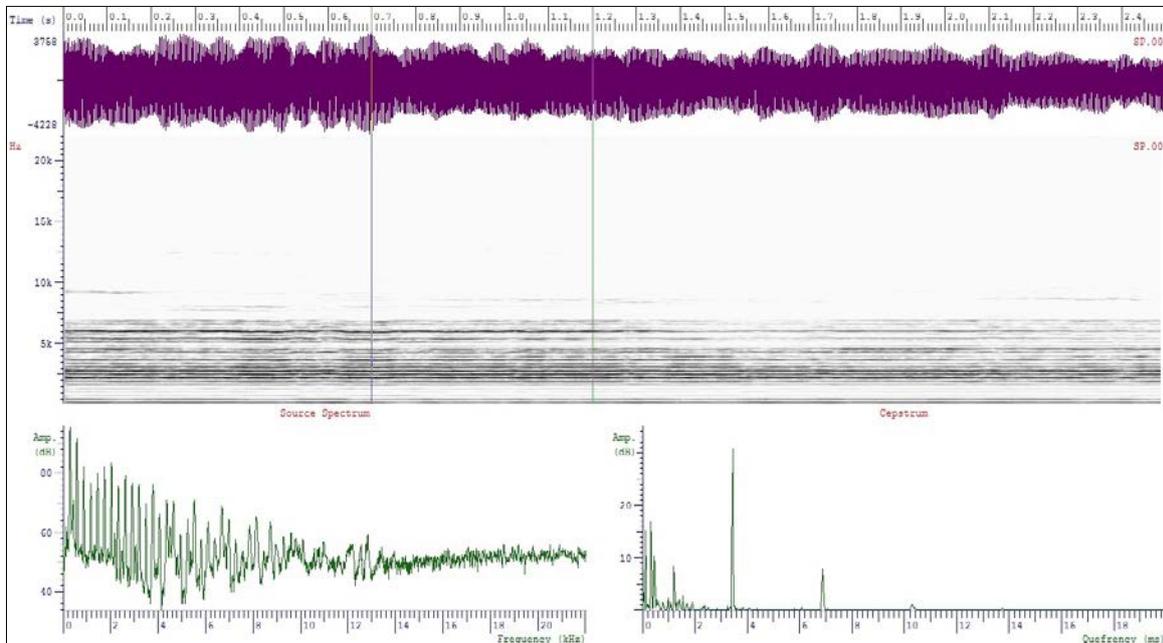


Fig. A.29.- Vocal “i” antes de aplicar el método para el Locutor 5

➤ (Vocal “i” después):

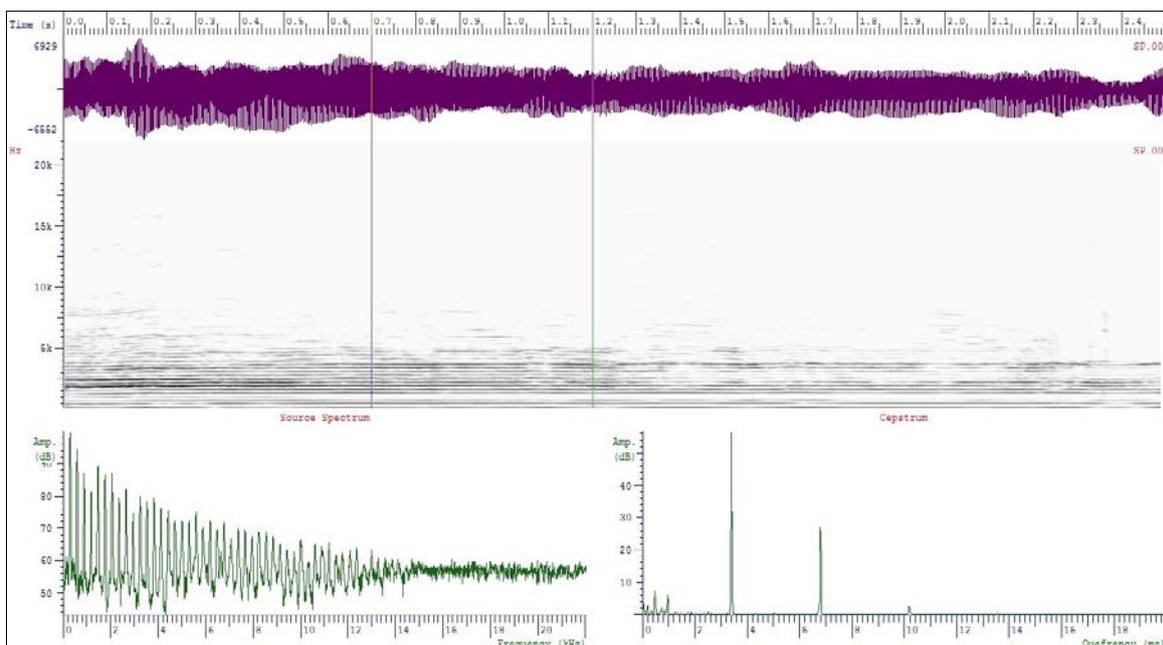


Fig. A.30.- Vocal “i” después de aplicar el método para el Locutor 5

❖ **Locutor 6:**

➤ *(Vocal “o” antes):*

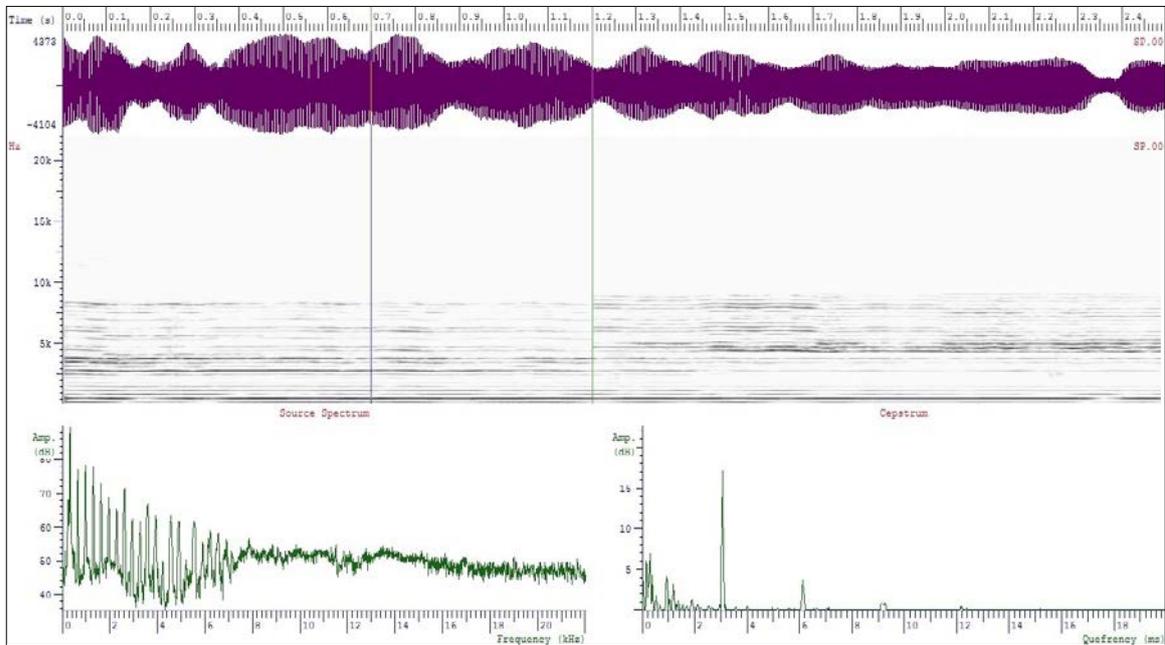


Fig. A.31.- Vocal “o” antes de aplicar el método para el Locutor 6

➤ *(Vocal “o” después):*

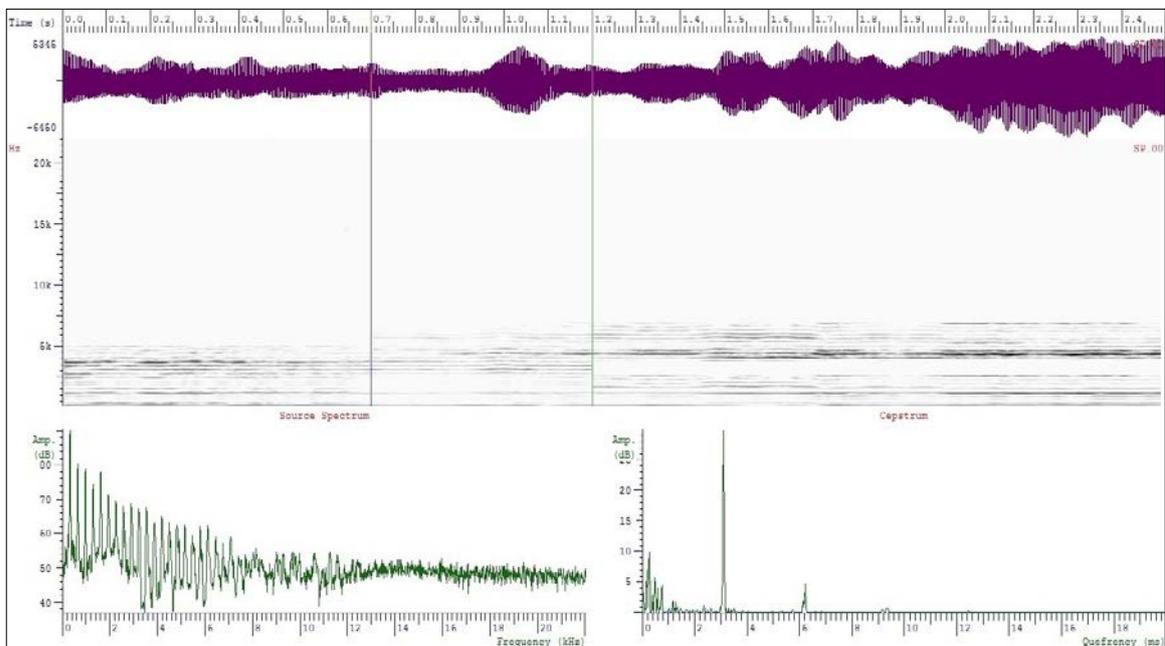


Fig. A.32.- Vocal “o” después de aplicar el método para el Locutor 6

➤ (Vocal “a” antes):

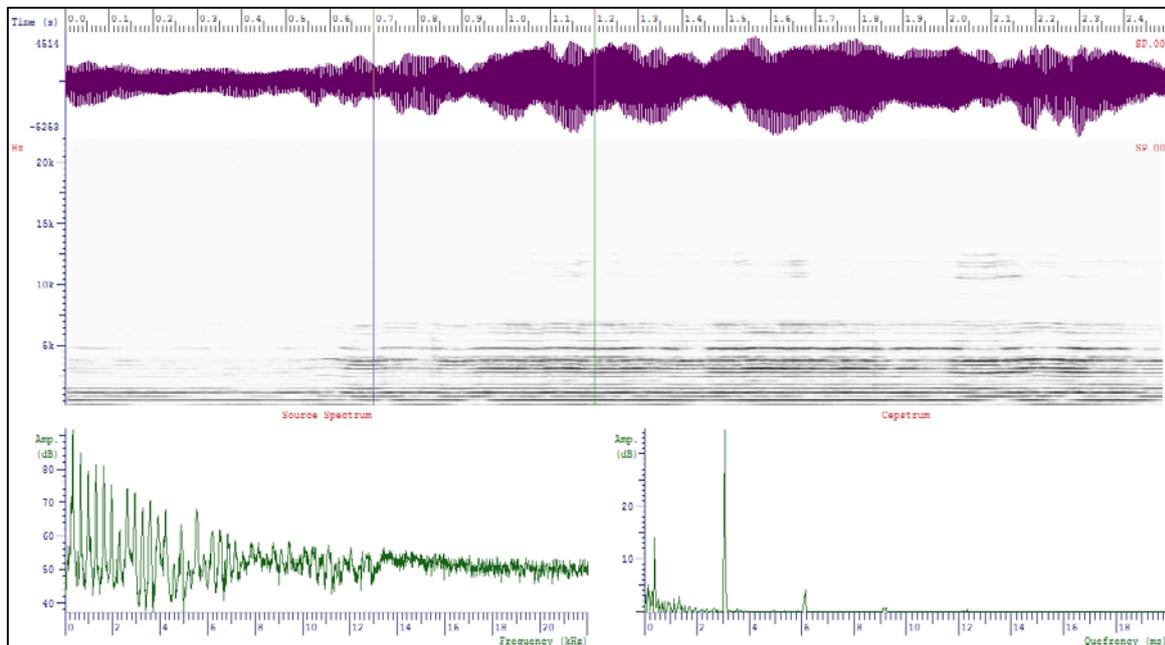


Fig. A.33.- Vocal “a” antes de aplicar el método para el Locutor 6

➤ (Vocal “a” después):

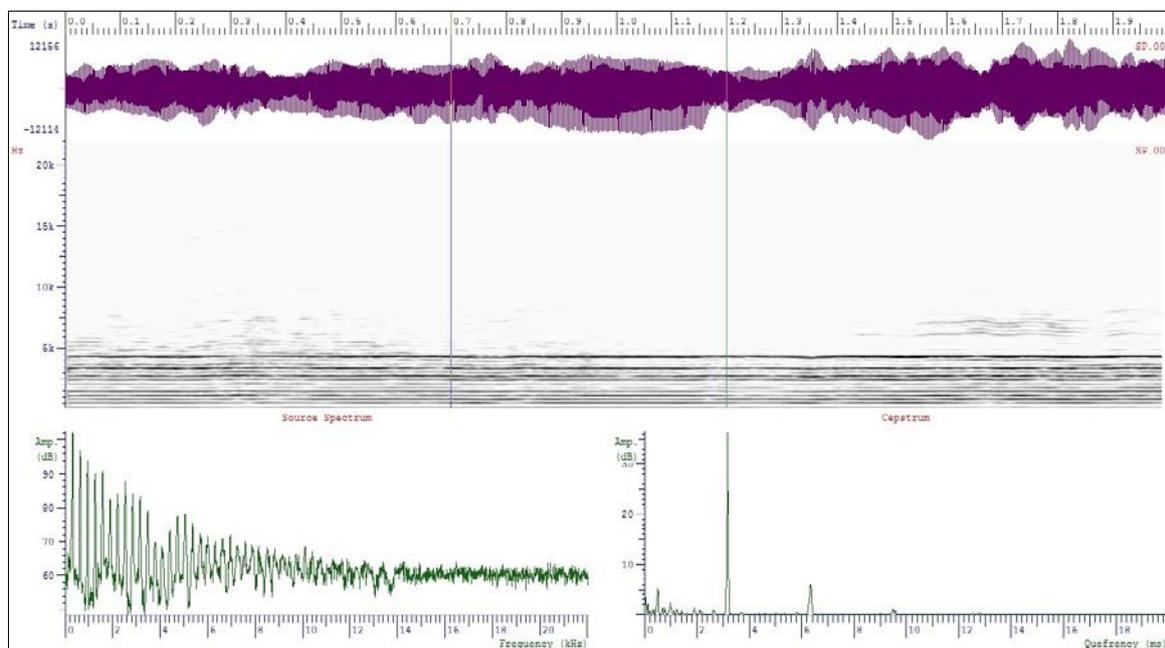


Fig. A.34.- Vocal “a” después de aplicar el método para el Locutor 6

➤ (Vocal “i” antes):

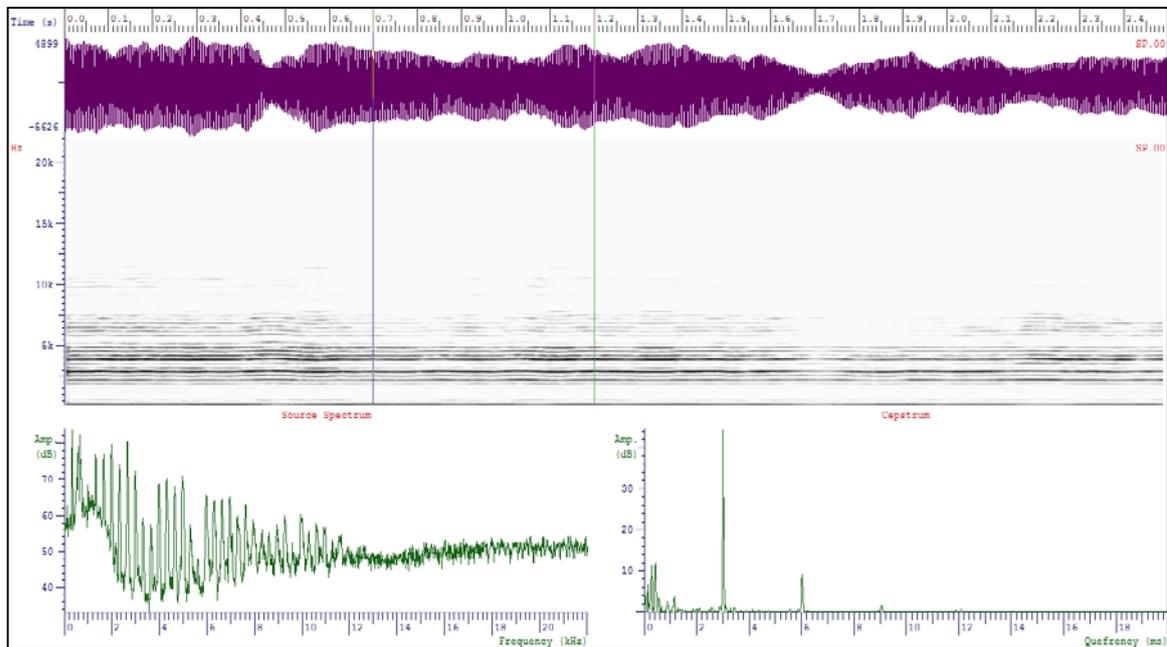


Fig. A.35.- Vocal “i” antes de aplicar el método para el Locutor 6

➤ (Vocal “i” después):

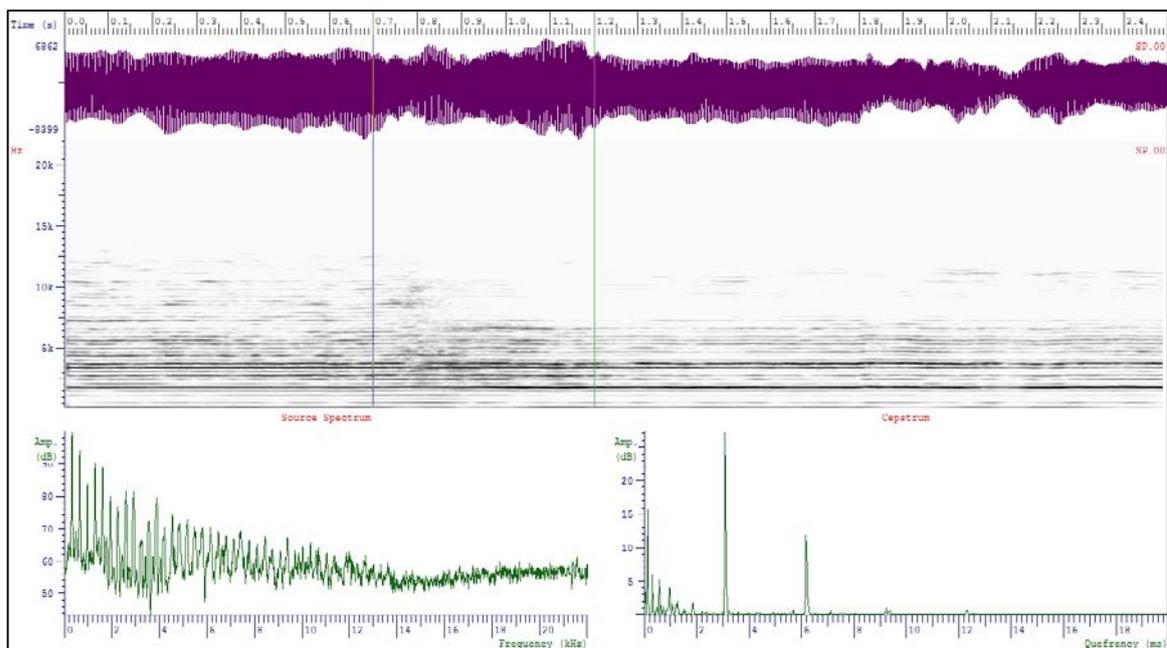


Fig. A.36.- Vocal “i” después de aplicar el método para el Locutor 6

❖ **Locutor 7:**

➤ *(Vocal “o” antes):*

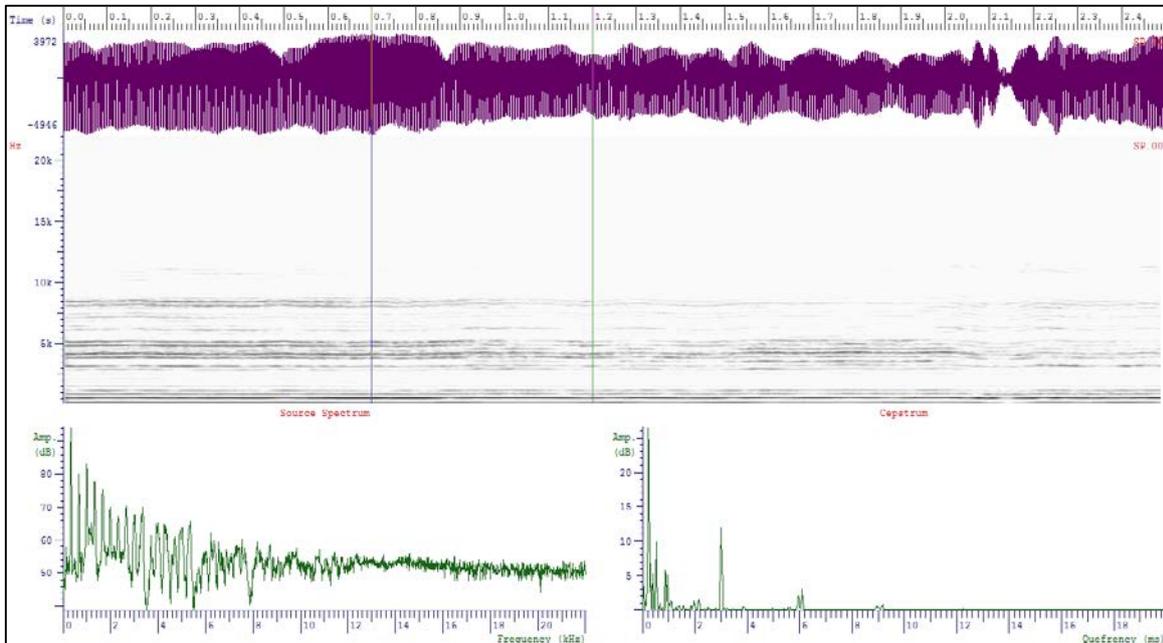


Fig. A.37.- Vocal “o” *antes* de aplicar el método para el Locutor 7

➤ *(Vocal “o” después):*

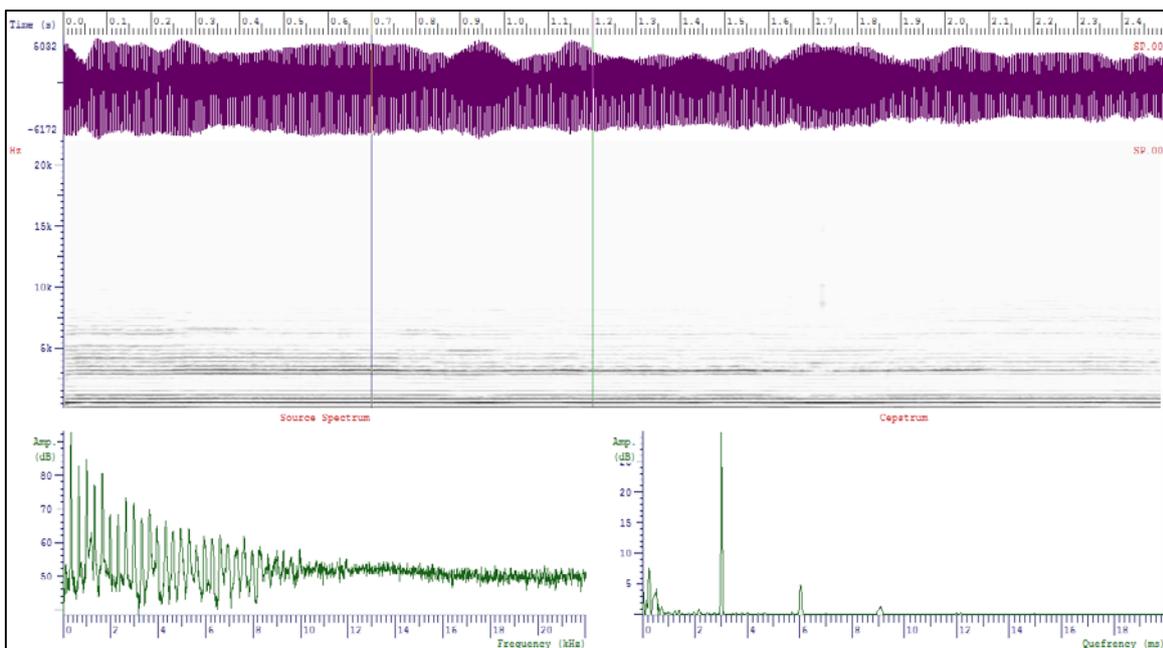


Fig. A.38.- Vocal “o” *después* de aplicar el método para el Locutor 7

➤ (Vocal “a” antes):

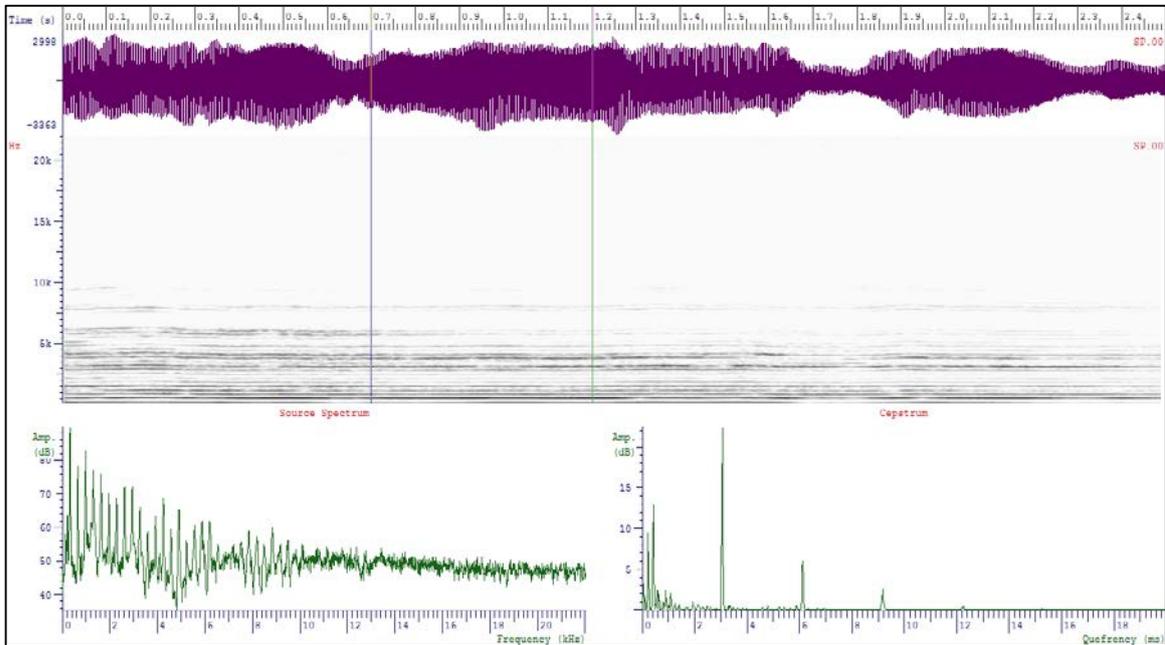


Fig. A.39.- Vocal “a” antes de aplicar el método para el Locutor 7

➤ (Vocal “a” después):

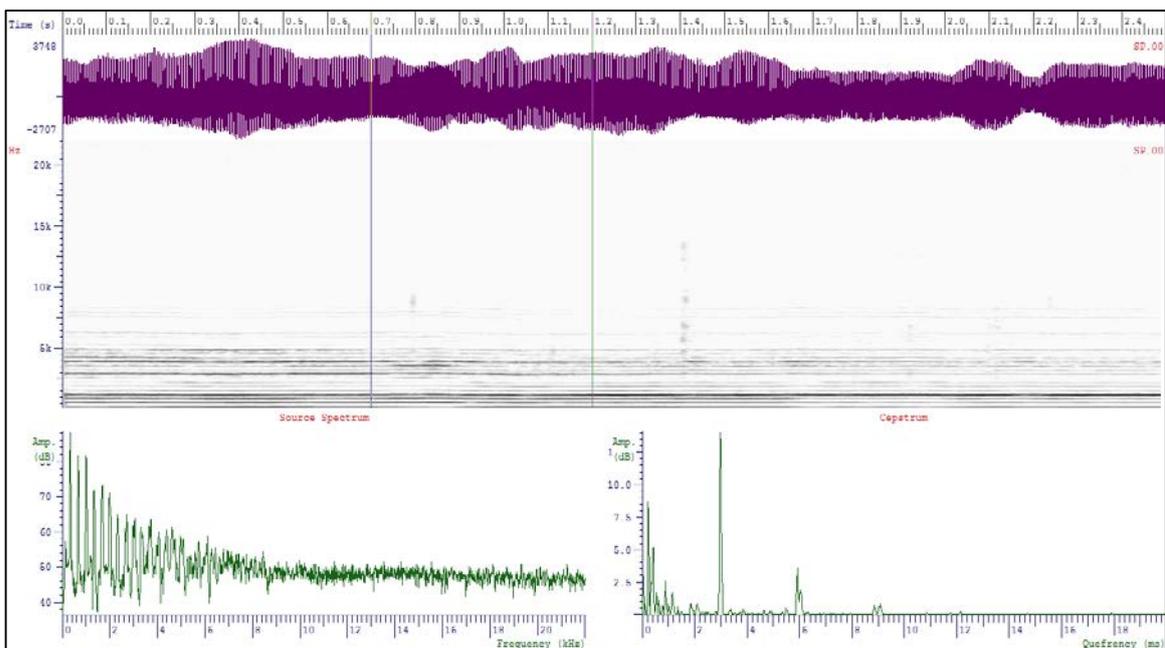


Fig. A.40.- Vocal “a” después de aplicar el método para el Locutor 7

➤ (Vocal “i” antes):

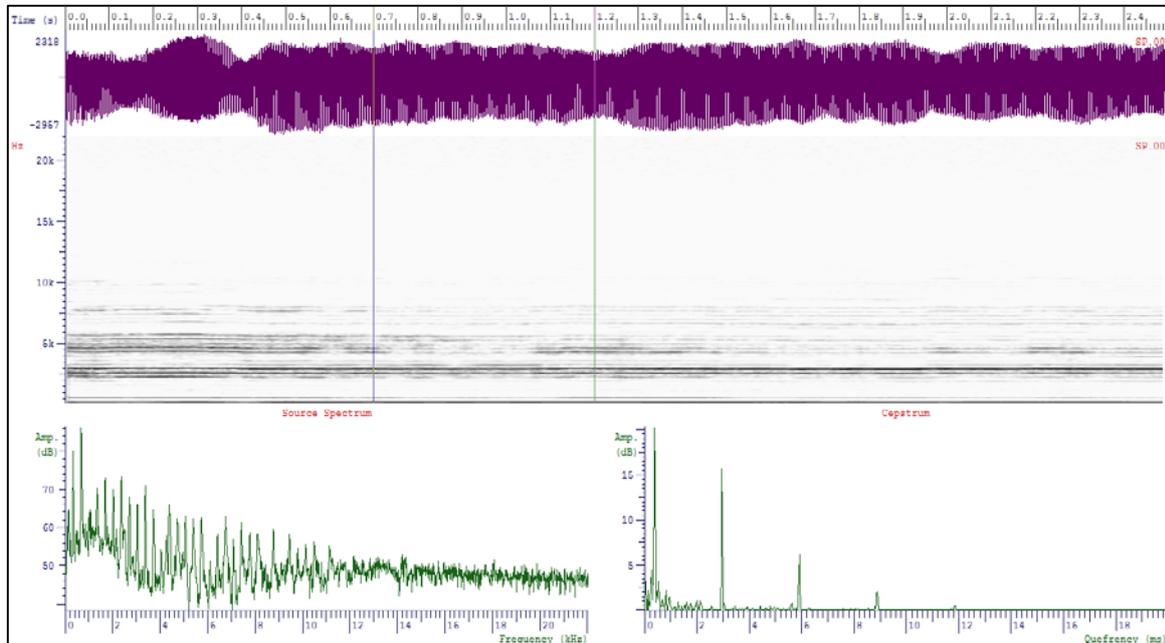


Fig. A.41.- Vocal “i” antes de aplicar el método para el Locutor 7

➤ (Vocal “i” después):

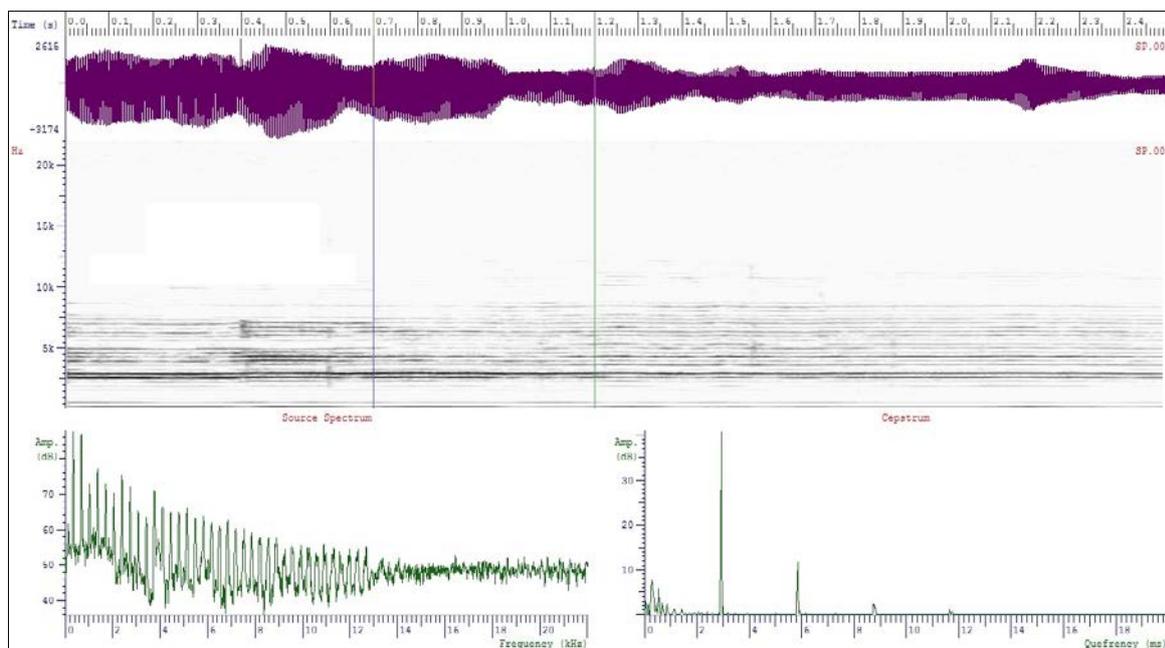


Fig. A.42.- Vocal “i” después de aplicar el método para el Locutor 7

❖ **Locutor 8:**

➤ *(Vocal “o” antes):*

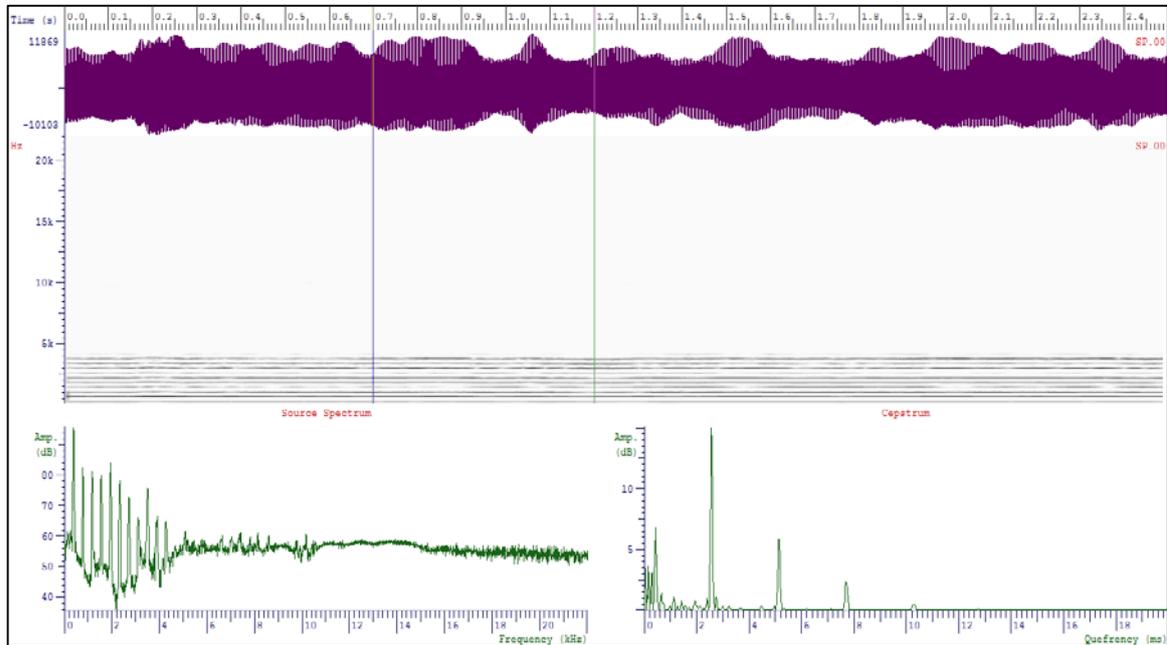


Fig. A.43.- Vocal “o” antes de aplicar el método para el Locutor 8

➤ *(Vocal “o” después):*

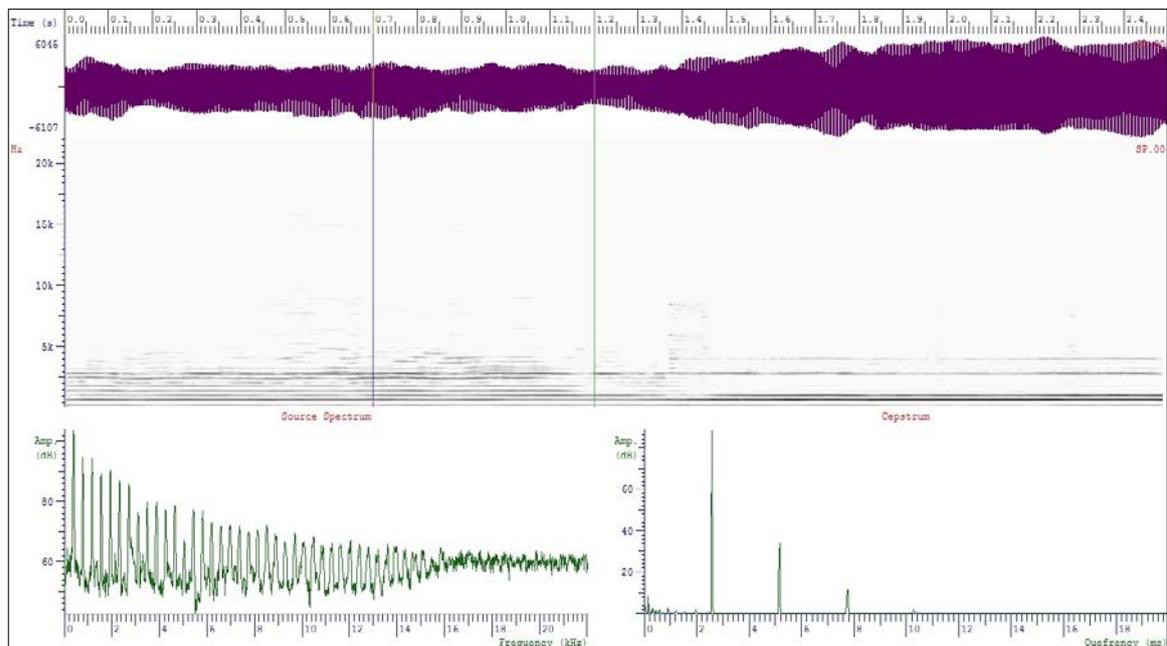


Fig. A.44.- Vocal “o” después de aplicar el método para el Locutor 8

➤ (Vocal “a” antes):

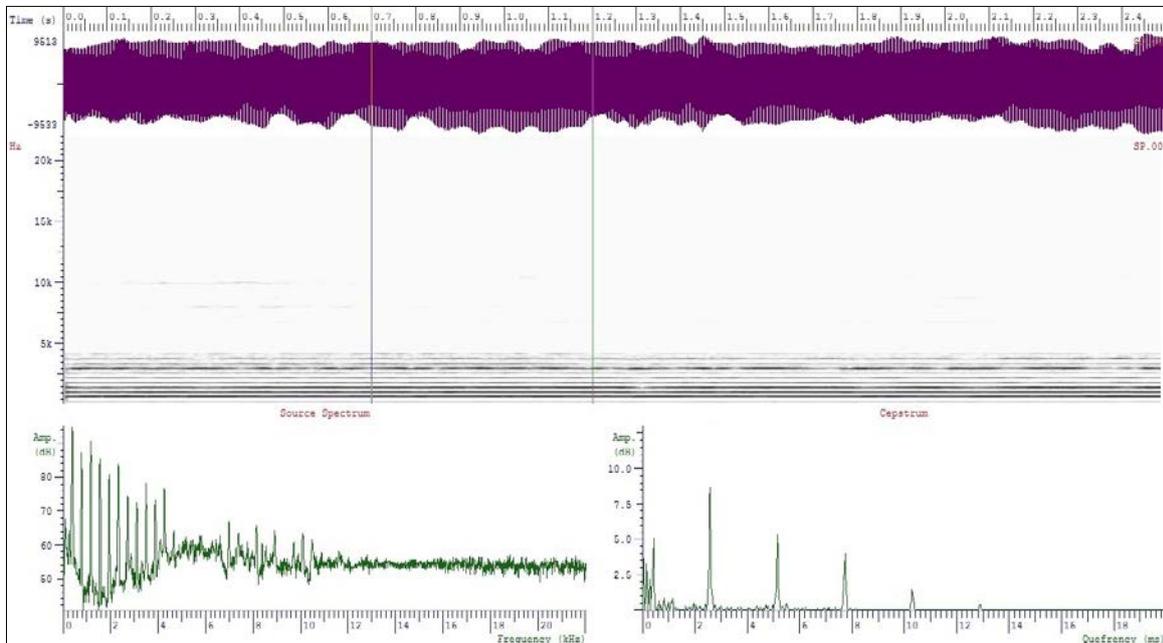


Fig. A.45.- Vocal “a” antes de aplicar el método para el Locutor 8

➤ (Vocal “a” después):

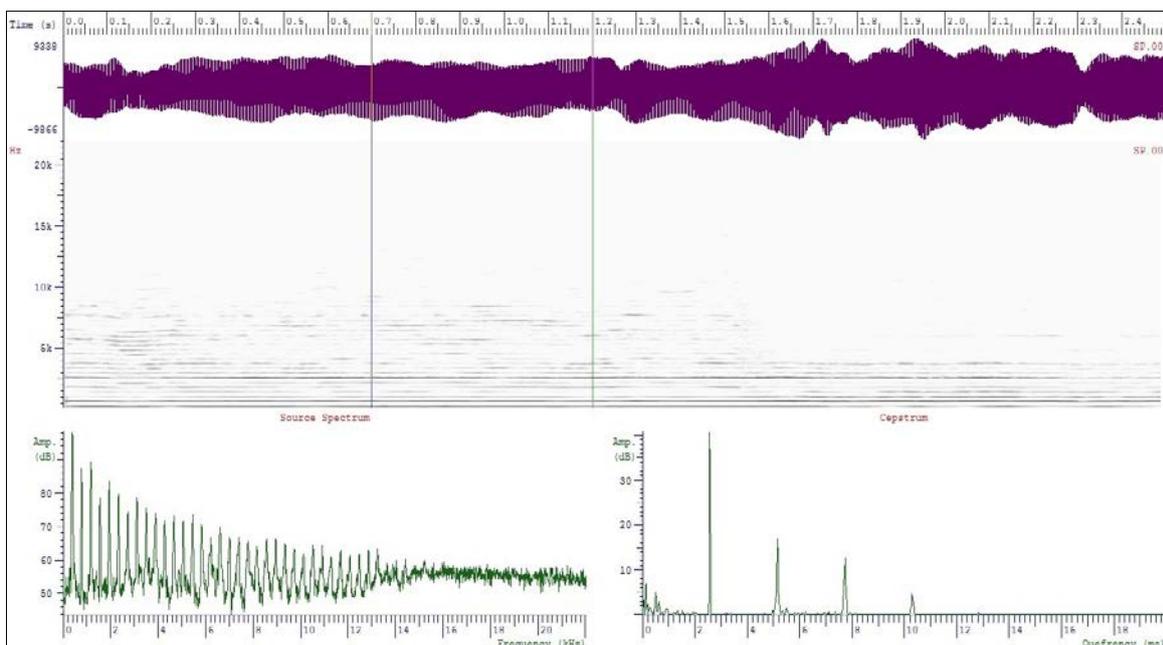


Fig. A.46.- Vocal “a” después de aplicar el método para el Locutor 8

➤ (Vocal “i” antes):

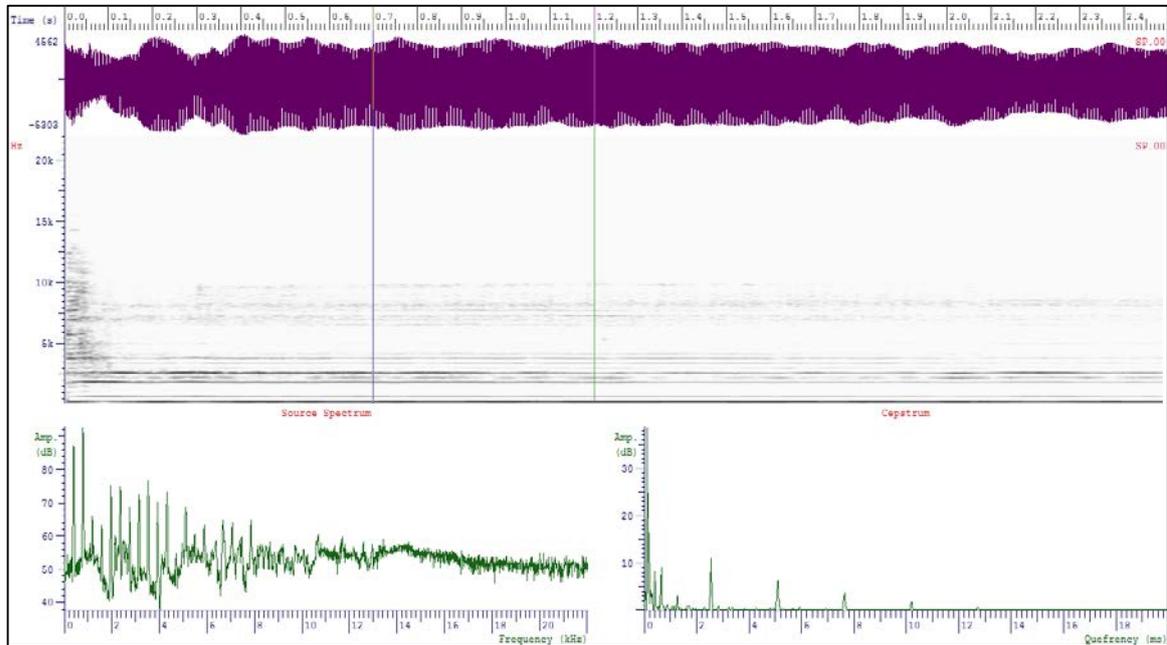


Fig. A.47.- Vocal “i” antes de aplicar el método para el Locutor 8

➤ (Vocal “i” después):

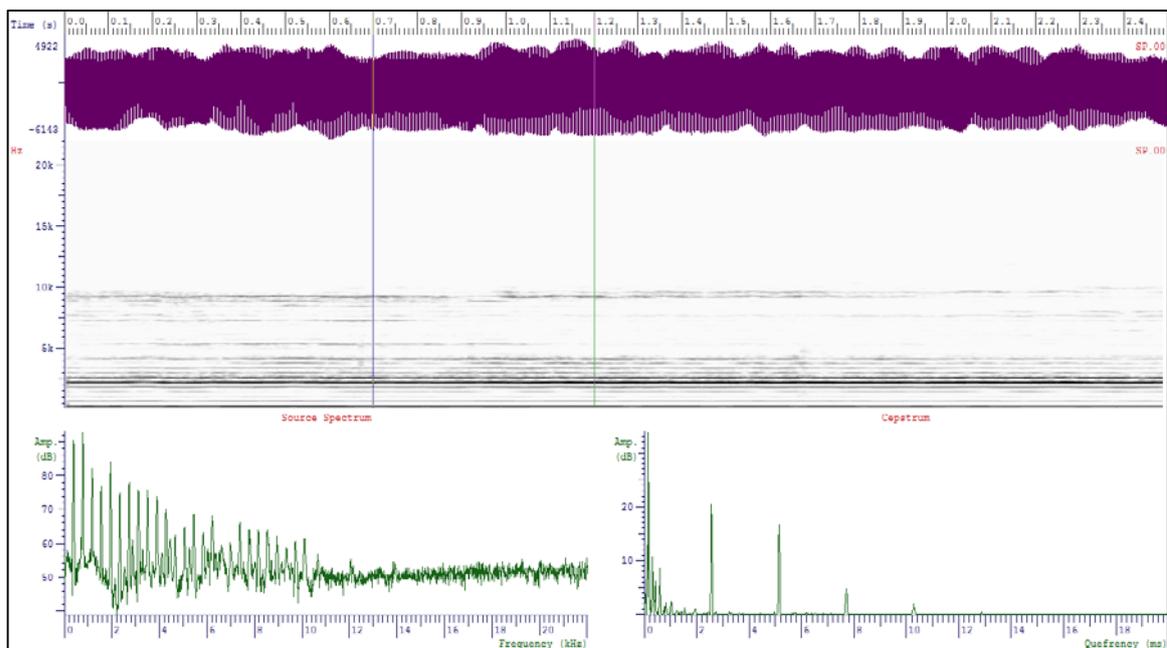


Fig. A.48.- Vocal “i” después de aplicar el método para el Locutor 8

❖ **Locutor 9:**

➤ *(Vocal “o” antes):*

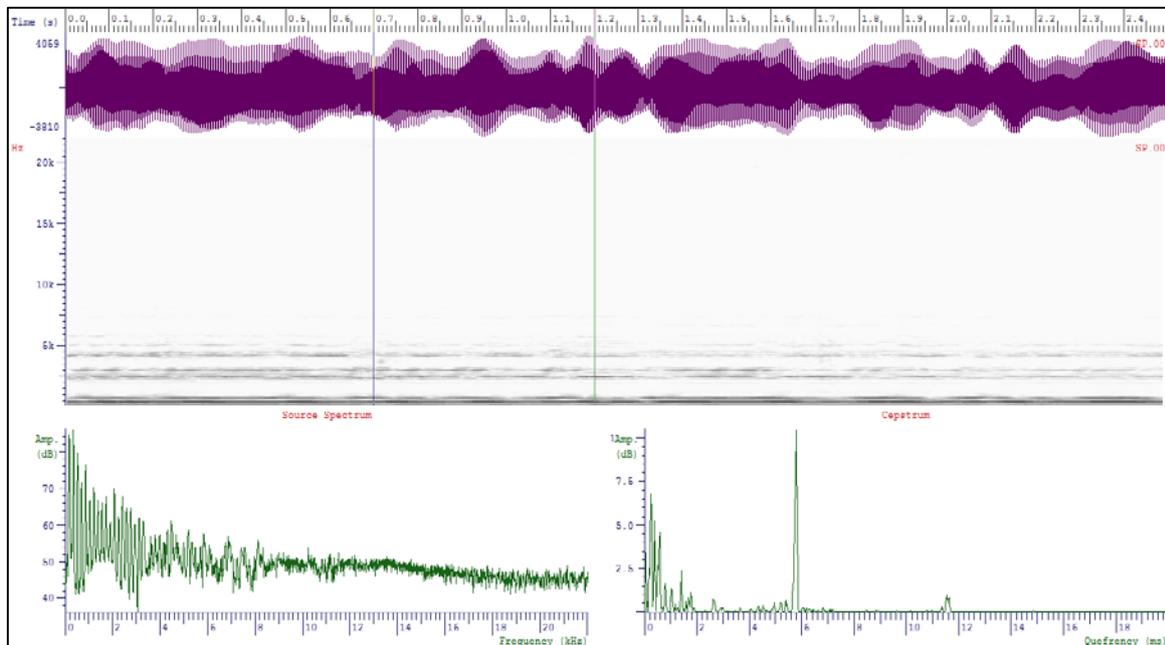


Fig. A.49.- Vocal “o” *antes* de aplicar el método para el Locutor 9

➤ *(Vocal “o” después):*

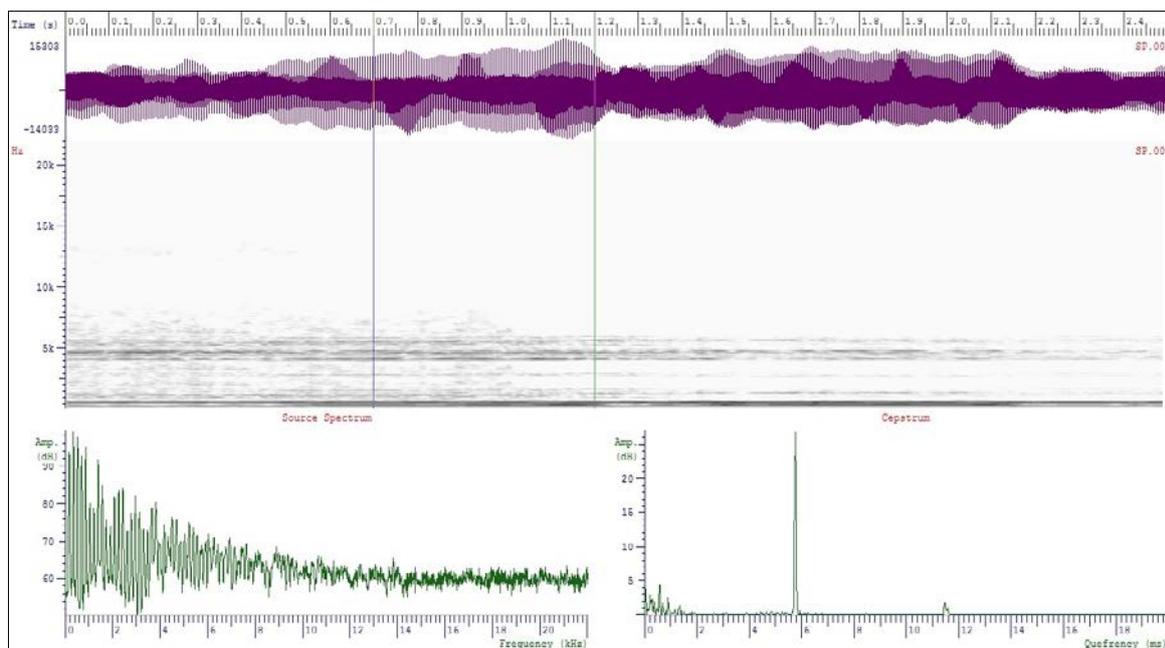


Fig. A.50.- Vocal “o” *después* de aplicar el método para el Locutor 9

➤ (Vocal “a” antes):

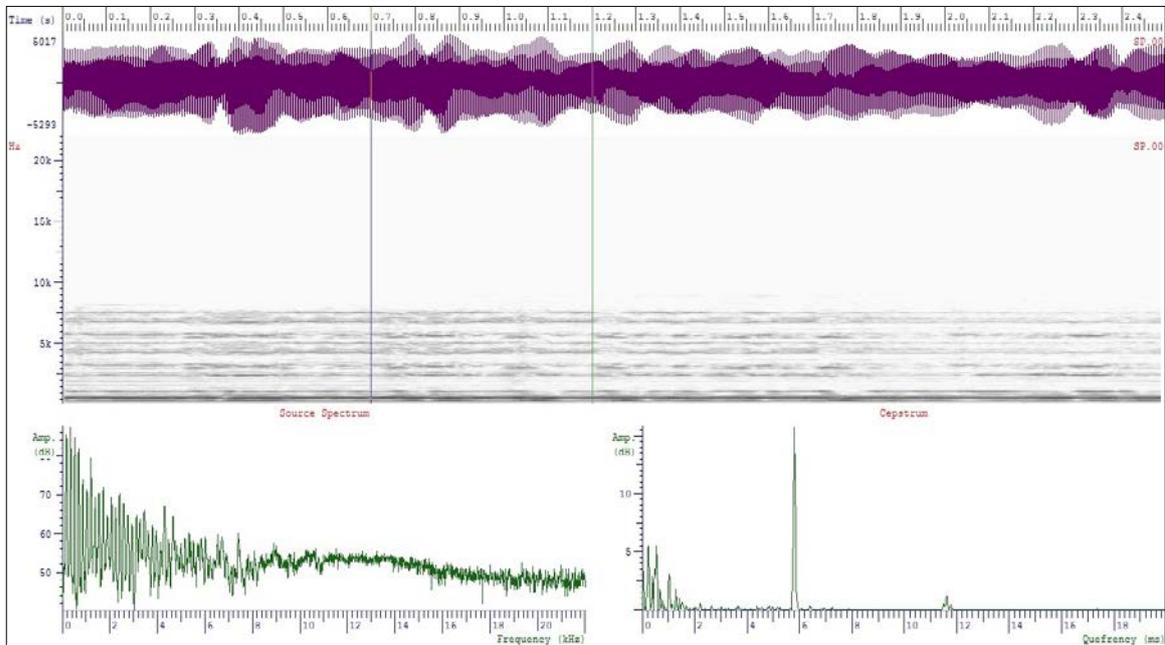


Fig. A.51.- Vocal “a” antes de aplicar el método para el Locutor 9

➤ (Vocal “a” después):

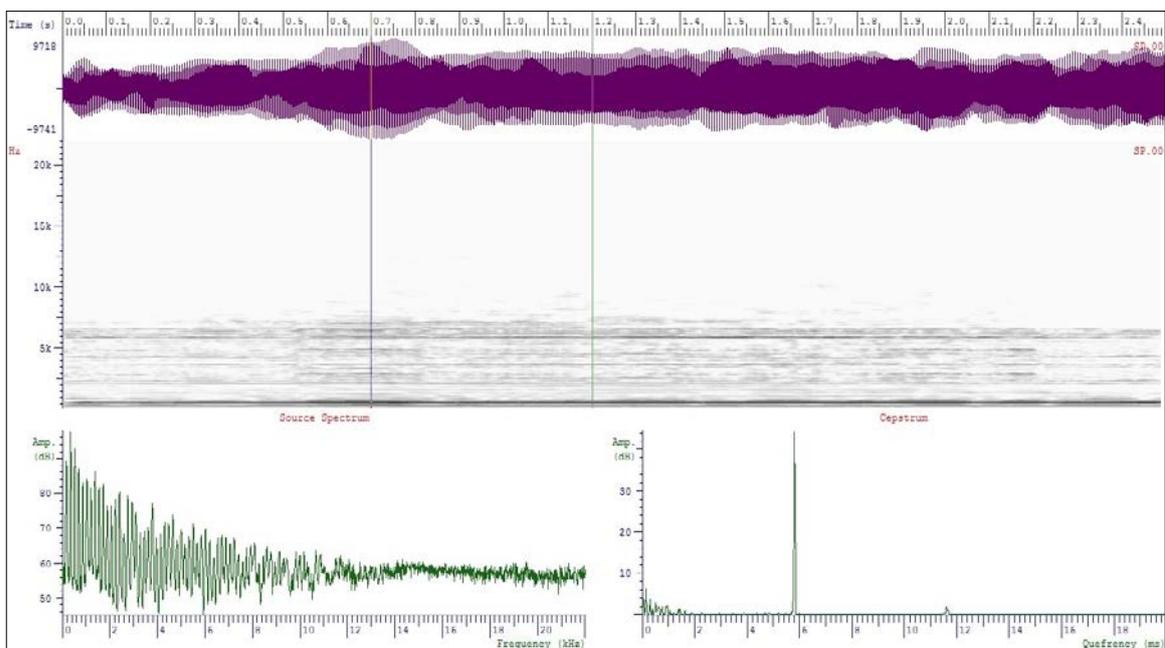


Fig. A.52.- Vocal “a” después de aplicar el método para el Locutor 9

➤ (Vocal “i” antes):

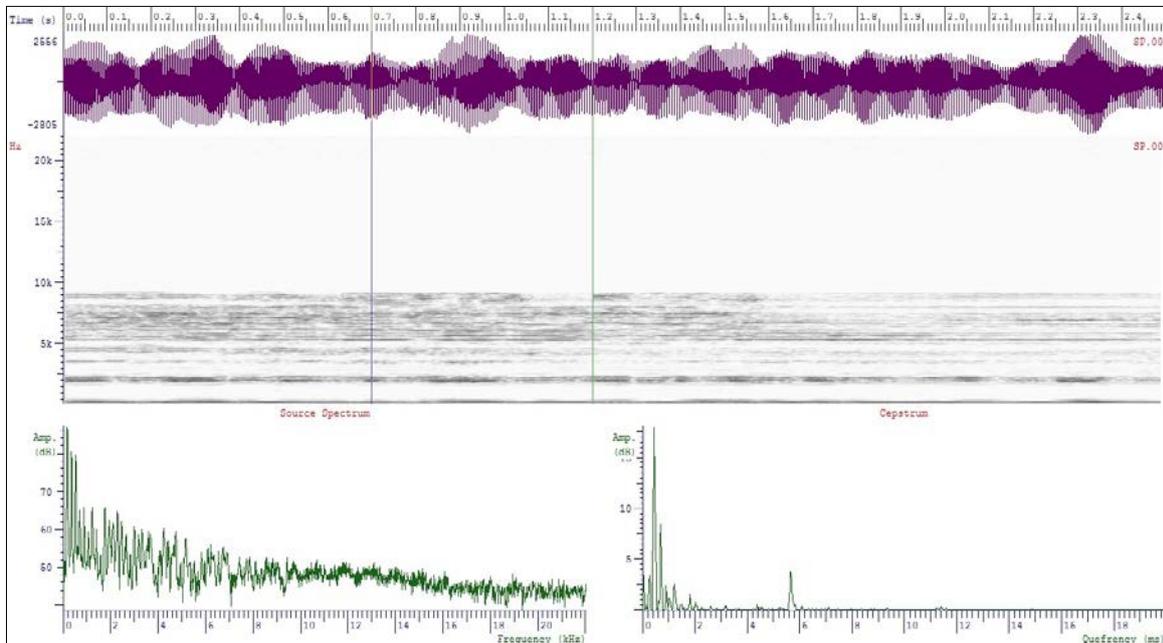


Fig. A.53.- Vocal “i” antes de aplicar el método para el Locutor 9

➤ (Vocal “i” después):

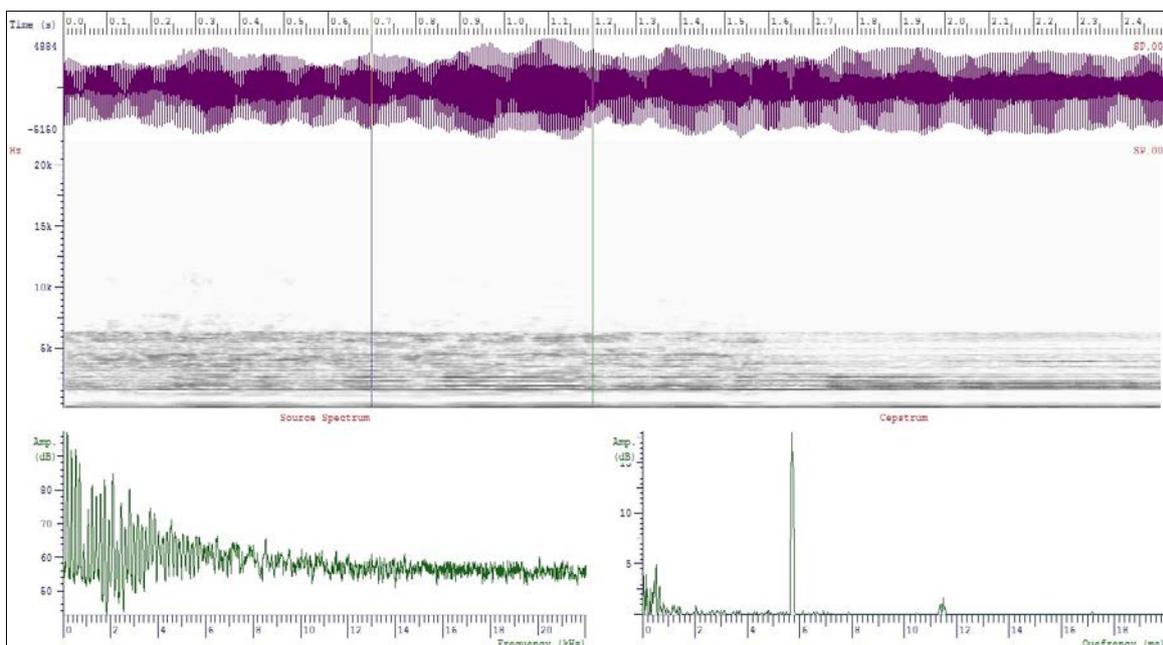


Fig. A.54.- Vocal “i” después de aplicar el método para el Locutor 9

❖ **Locutor 10:**

➤ *(Vocal “o” antes):*

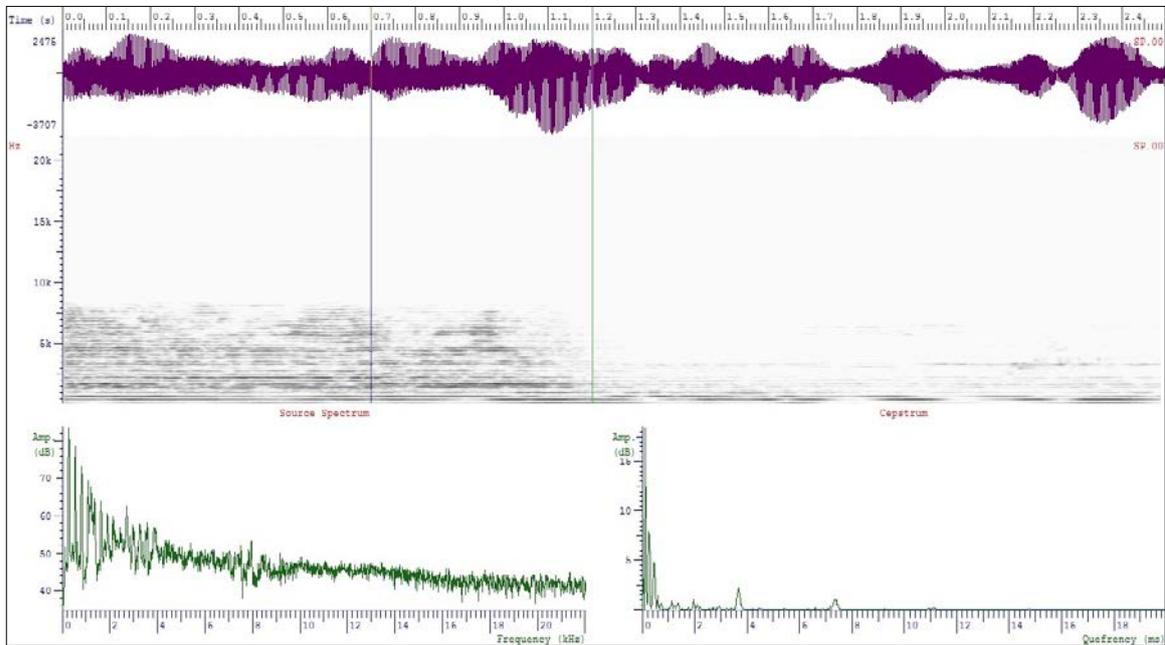


Fig. A.55.- Vocal “o” *antes* de aplicar el método para el Locutor 10

➤ *(Vocal “o” después):*

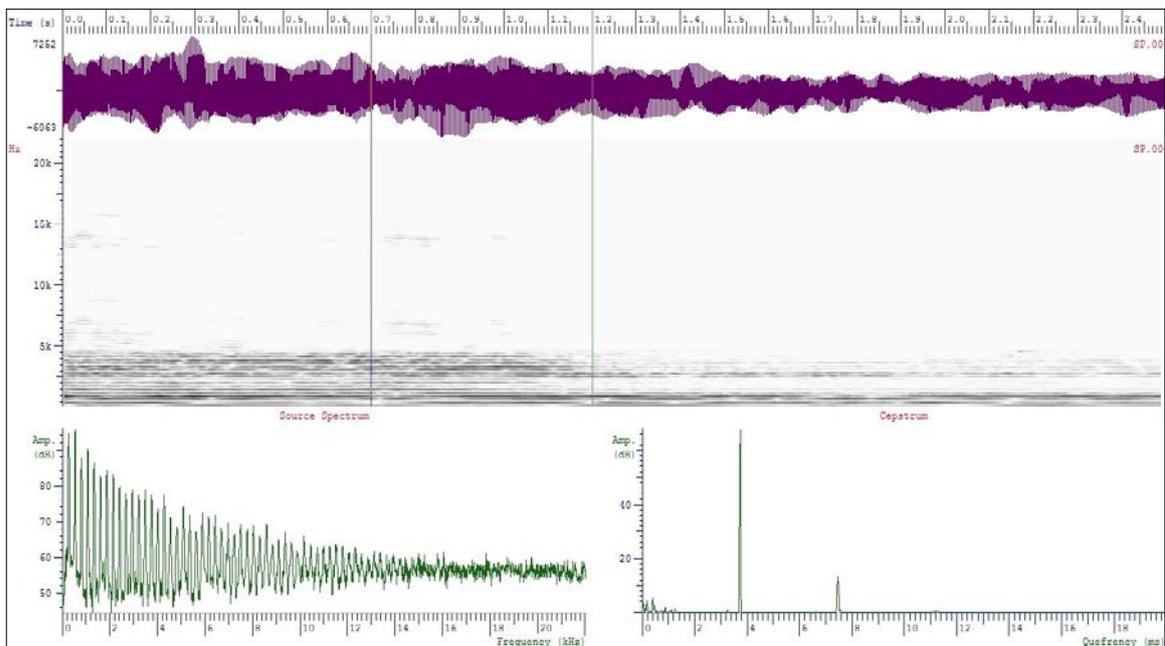


Fig. A.56.- Vocal “o” *después* de aplicar el método para el Locutor 10

➤ (Vocal “a” antes):

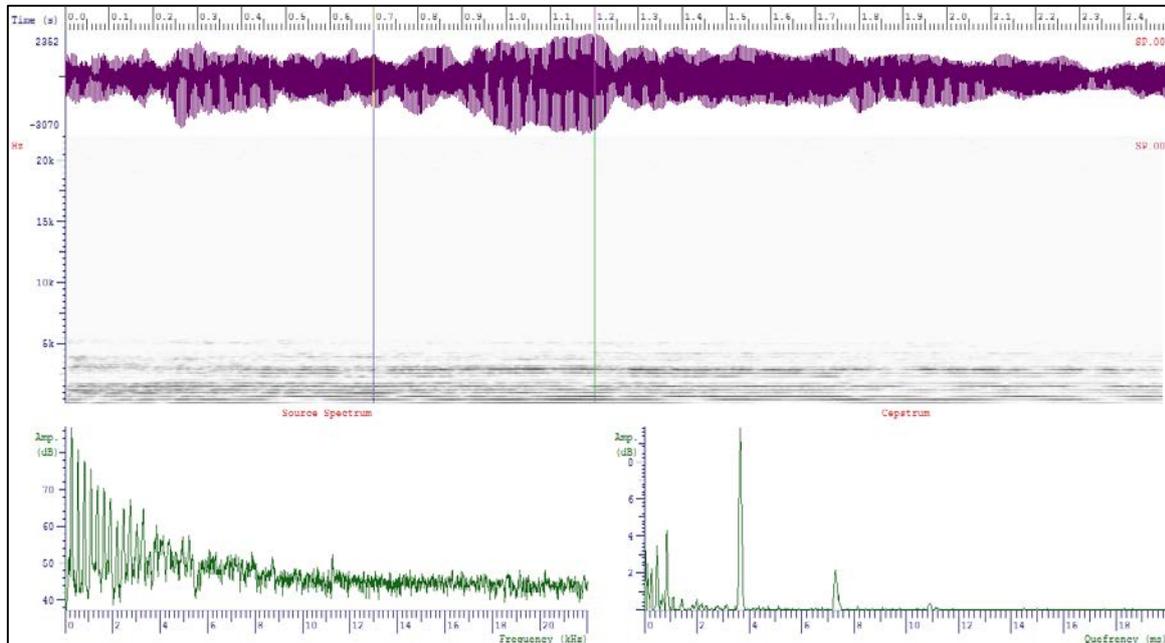


Fig. A.57.- Vocal “a” antes de aplicar el método para el Locutor 10

➤ (Vocal “a” después):

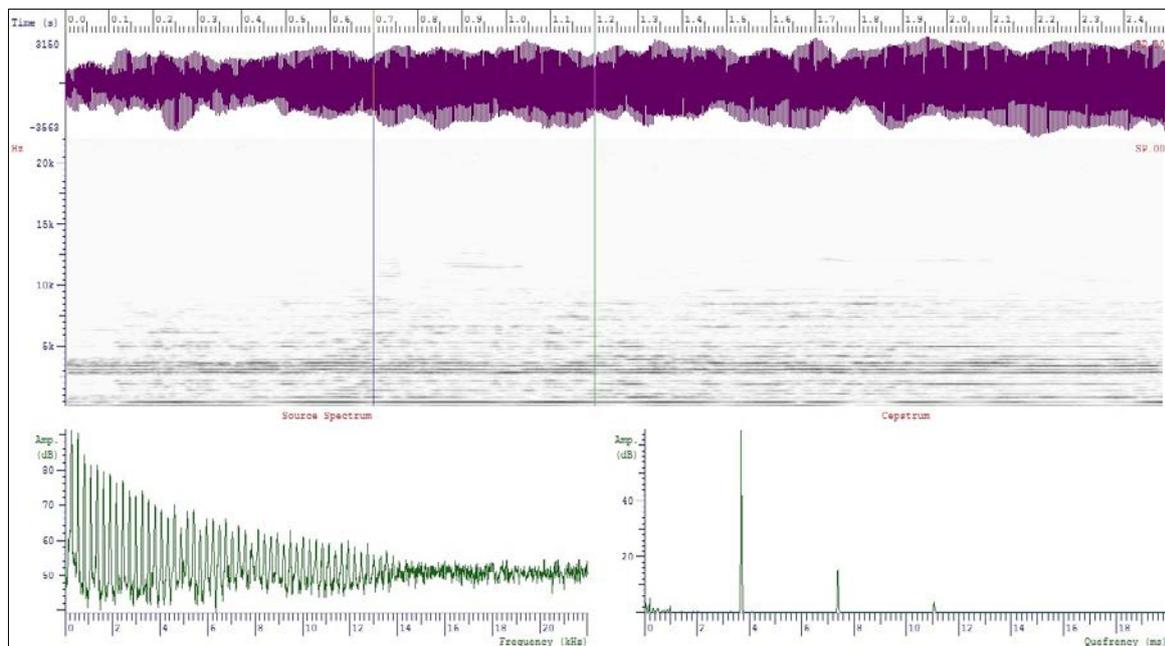


Fig. A.58.- Vocal “a” después de aplicar el método para el Locutor 10

➤ (Vocal “i” antes):

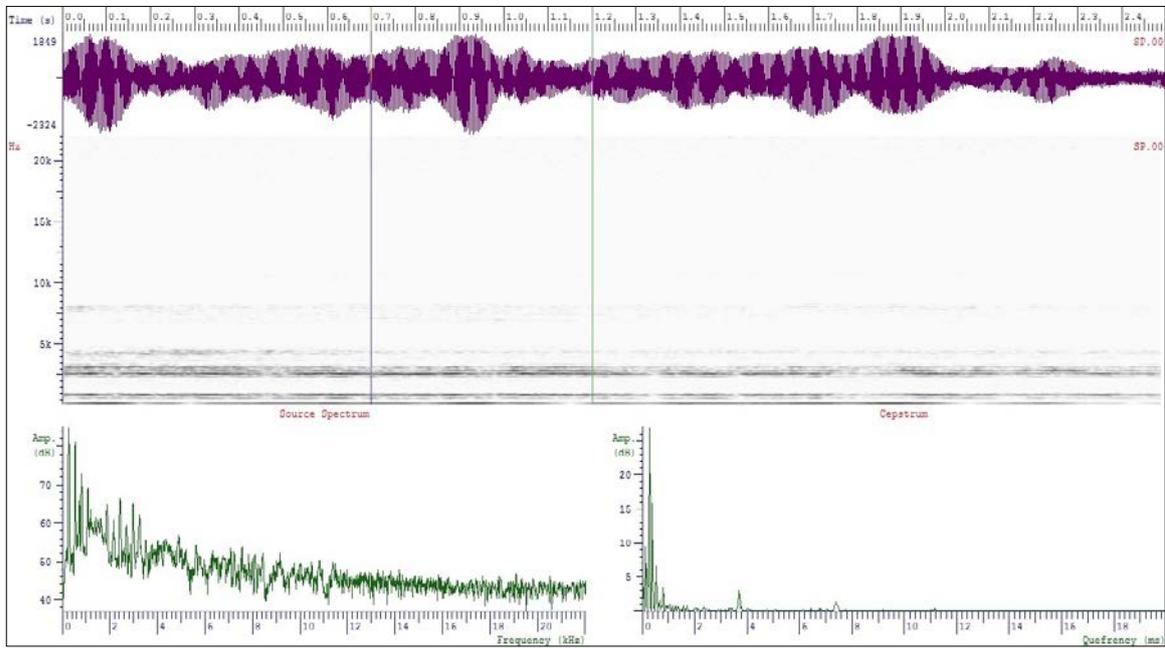


Fig. A.59.- Vocal “i” antes de aplicar el método para el Locutor 10

➤ (Vocal “i” después):

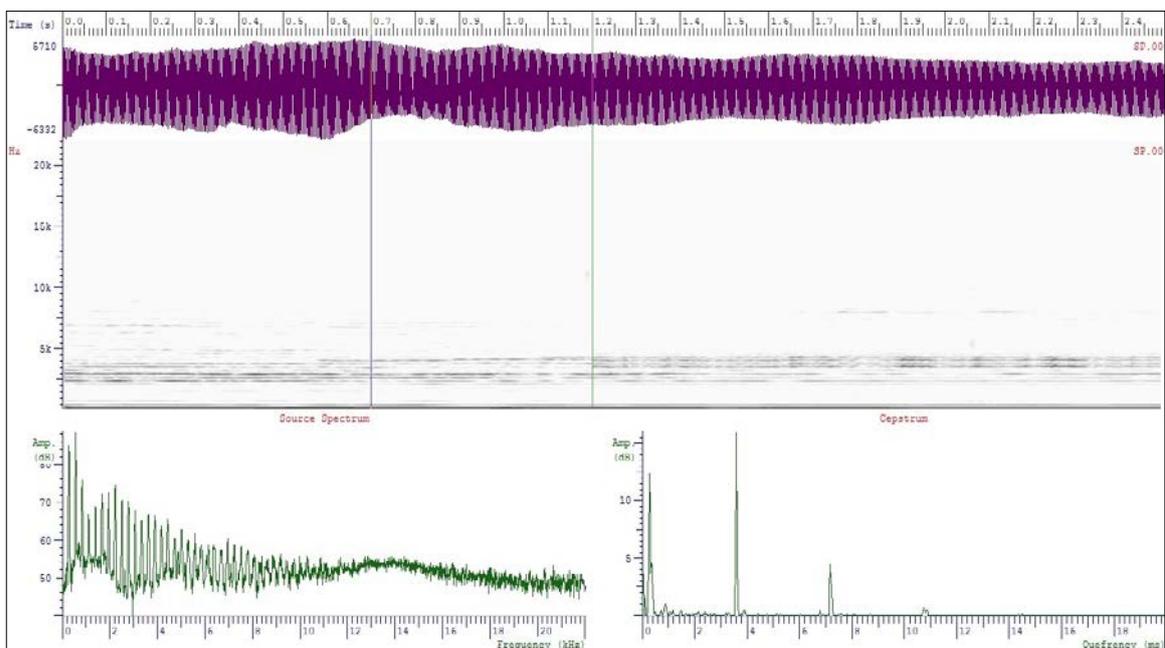


Fig. A.60.- Vocal “i” después de aplicar el método para el Locutor 10

❖ **Locutor 11:**

➤ *(Vocal “o” antes):*

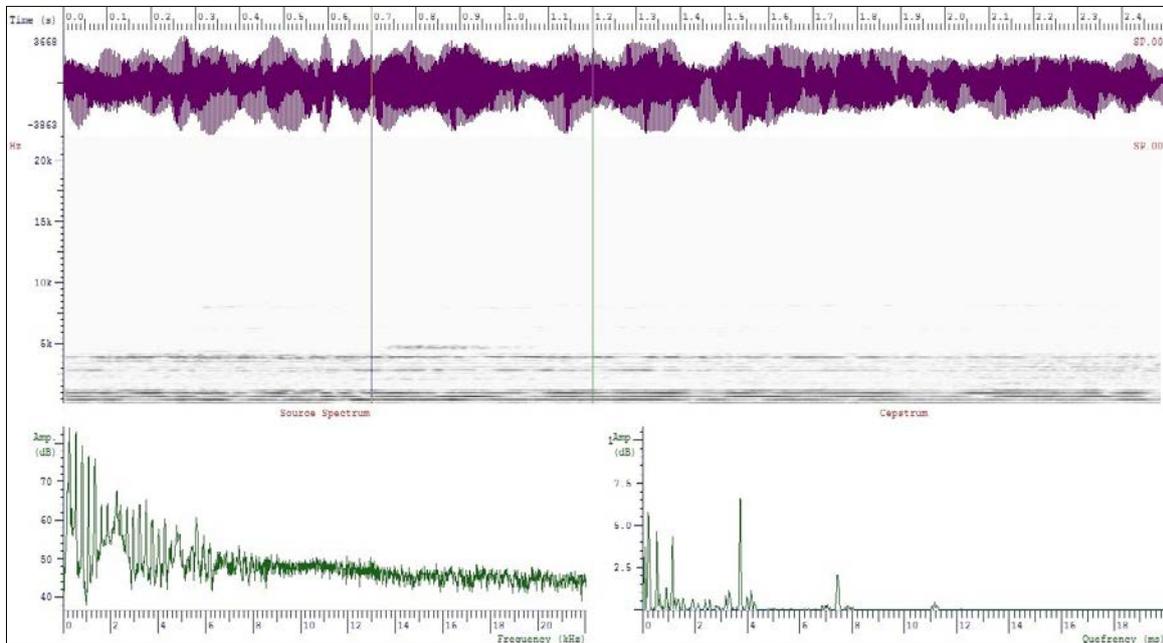


Fig. A.61.- Vocal “o” antes de aplicar el método para el Locutor 11

➤ *(Vocal “o” después):*

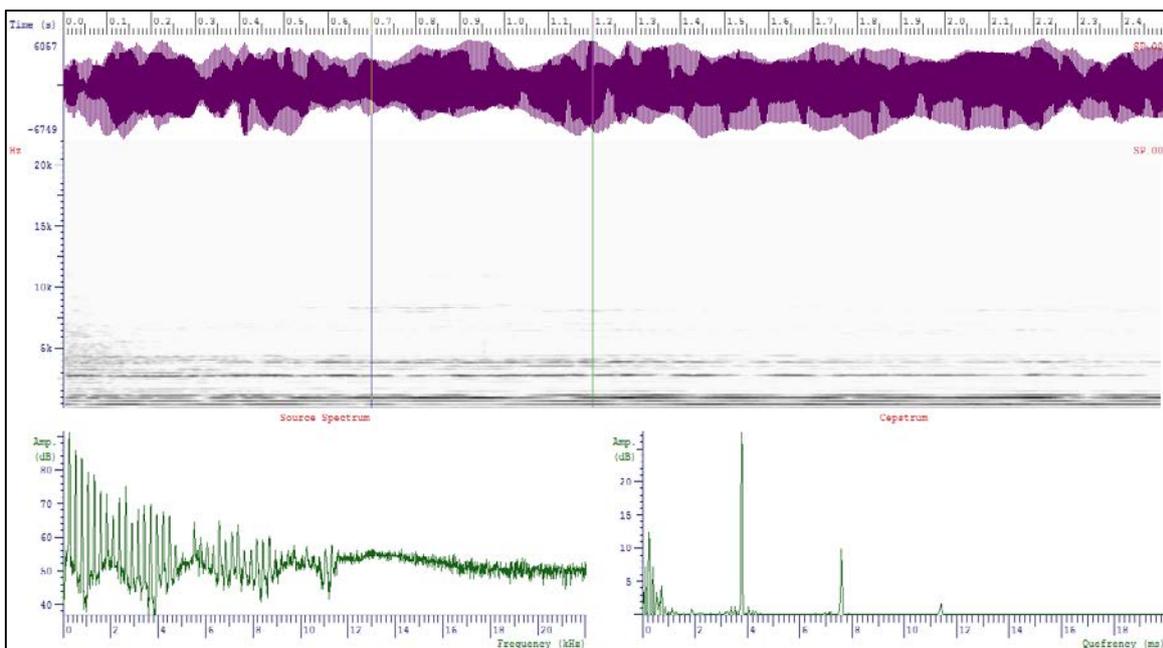


Fig. A.62.- Vocal “o” después de aplicar el método para el Locutor 11

➤ (Vocal “a” antes):

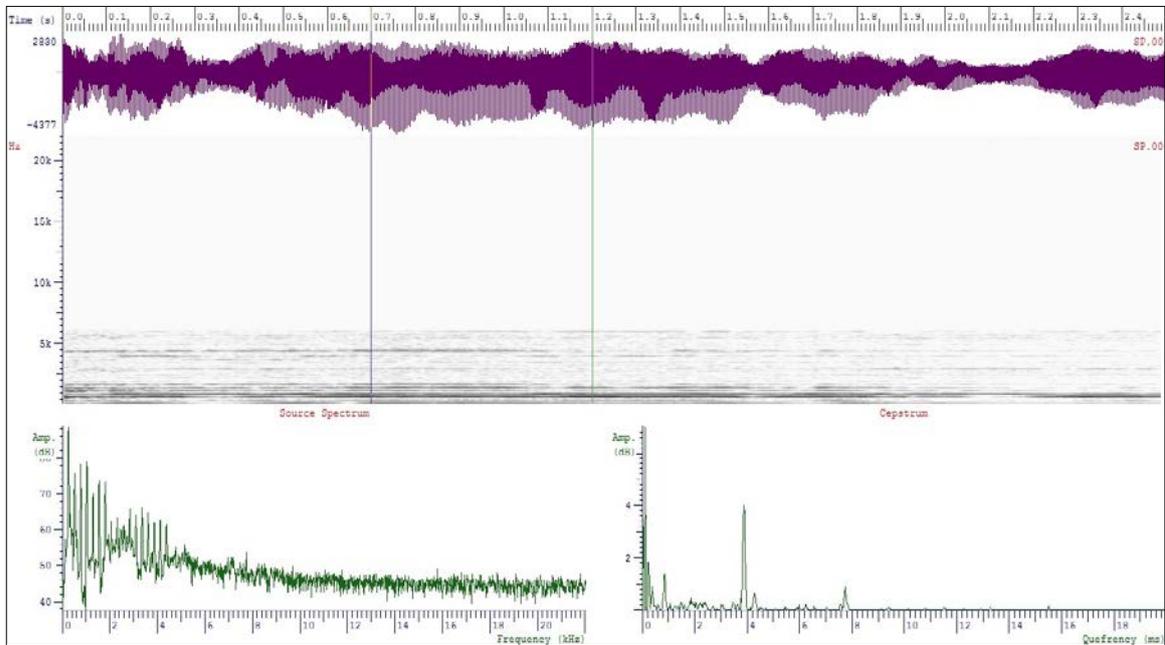


Fig. A.63.- Vocal “a” antes de aplicar el método para el Locutor 11

➤ (Vocal “a” después):

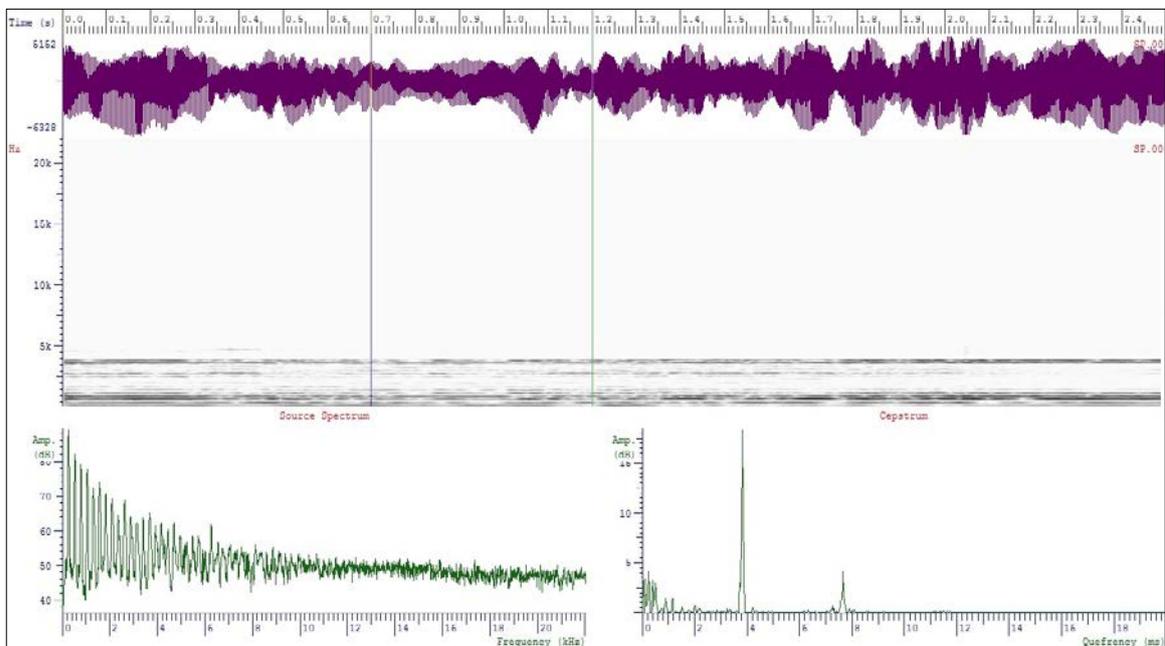


Fig. A.64.- Vocal “a” después de aplicar el método para el Locutor 11

➤ (Vocal “i” antes):

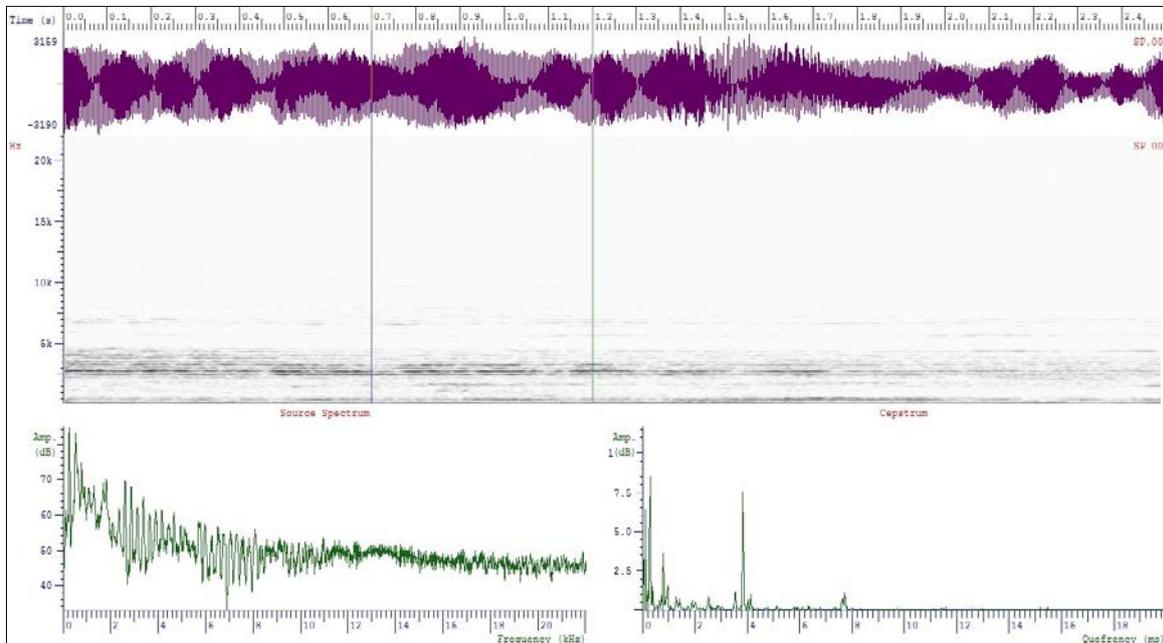


Fig. A.65.- Vocal “i” antes de aplicar el método para el Locutor 11

➤ (Vocal “i” después):

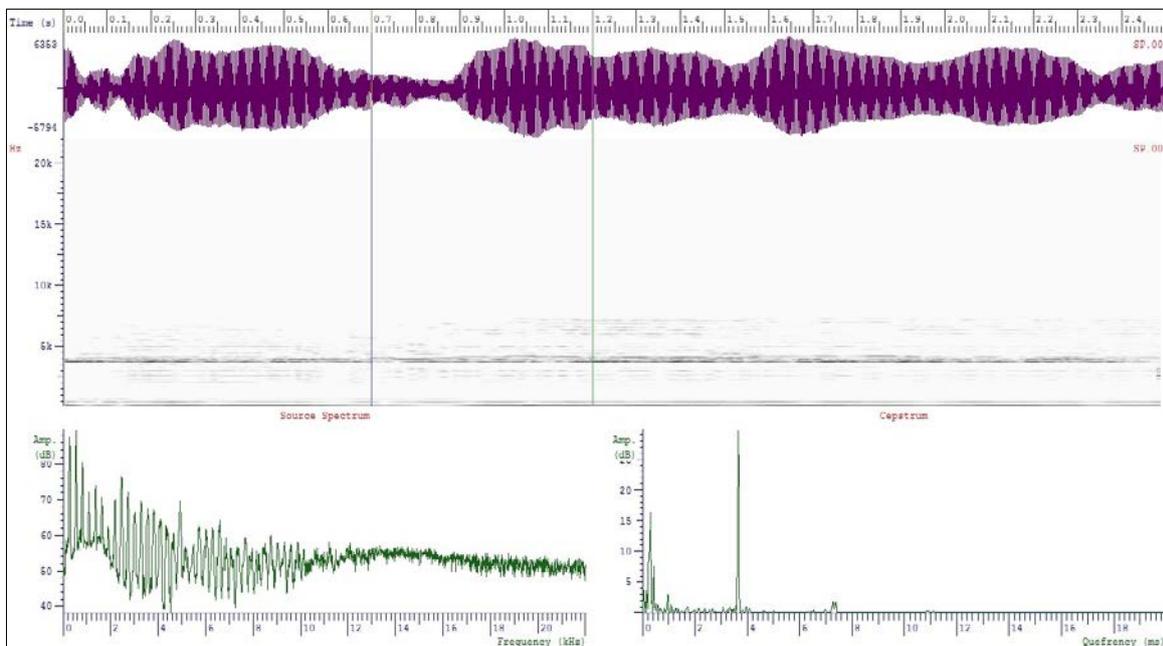


Fig. A.66.- Vocal “i” después de aplicar el método para el Locutor 11

❖ **Locutor 12:**

➤ *(Vocal “o” antes):*

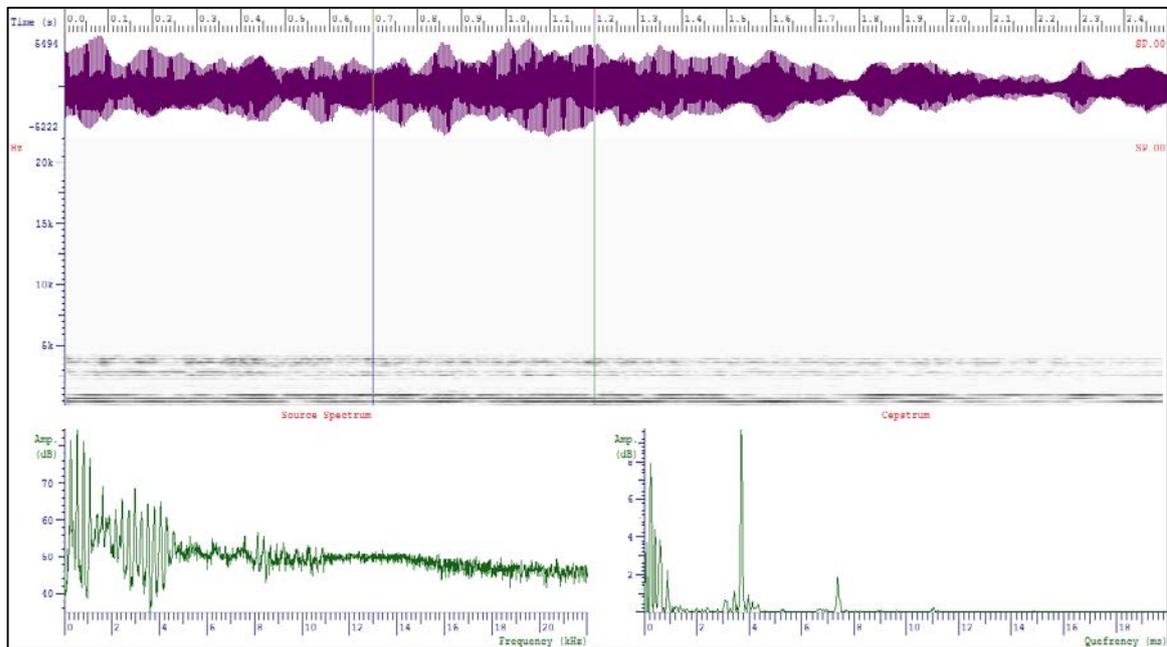


Fig. A.67.- Vocal “o” *antes* de aplicar el método para el Locutor 12

➤ *(Vocal “o” después):*

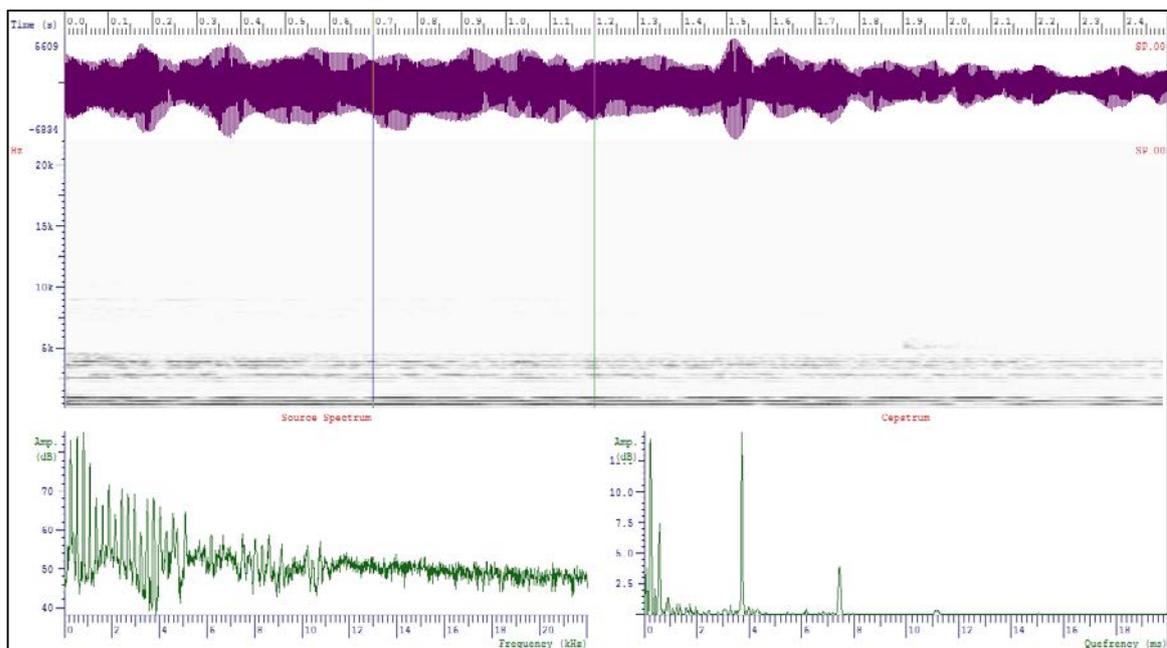


Fig. A.68.- Vocal “o” *después* de aplicar el método para el Locutor 12

➤ (Vocal “a” antes):

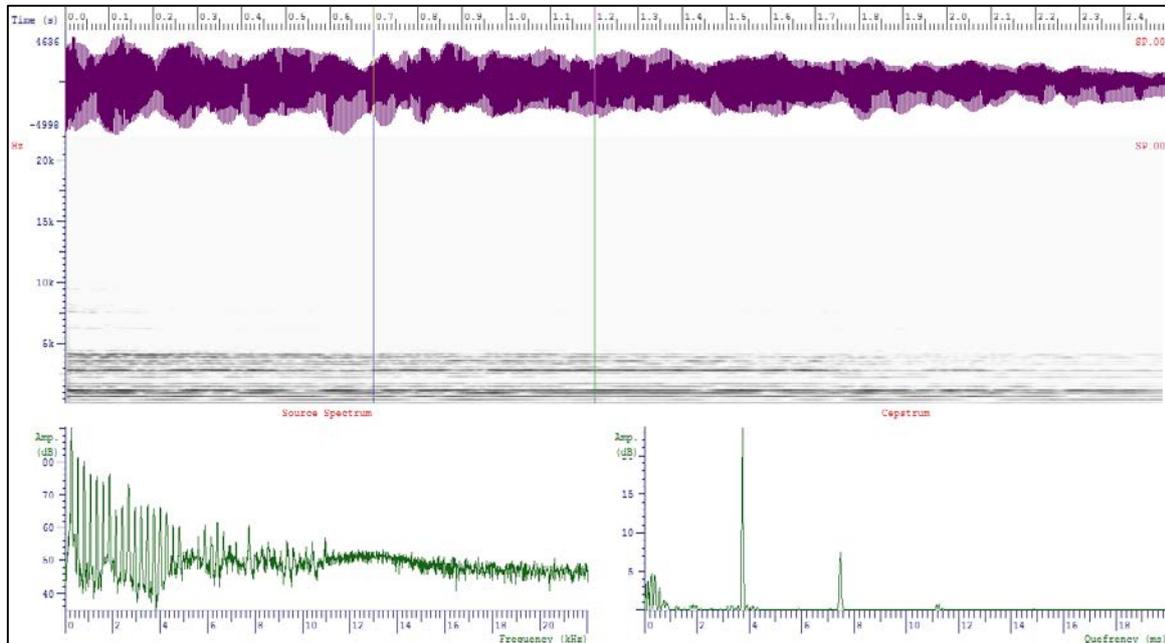


Fig. A.69.- Vocal “a” antes de aplicar el método para el Locutor 12

➤ (Vocal “a” después):

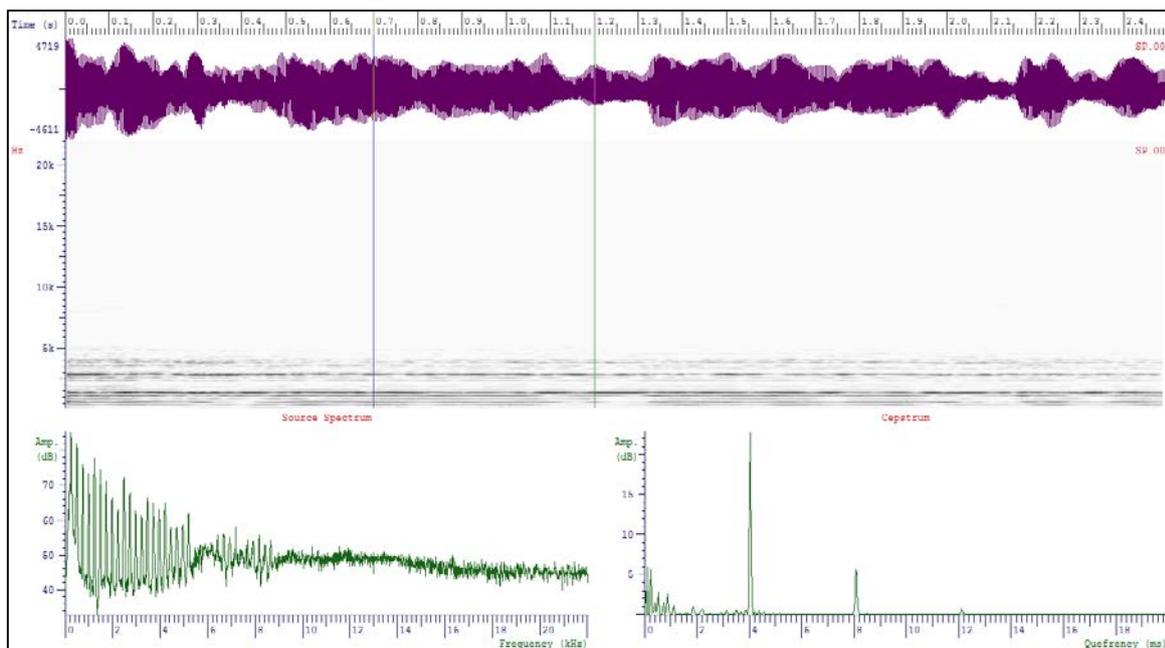


Fig. A.70.- Vocal “a” después de aplicar el método para el Locutor 12

➤ (Vocal “i” antes):

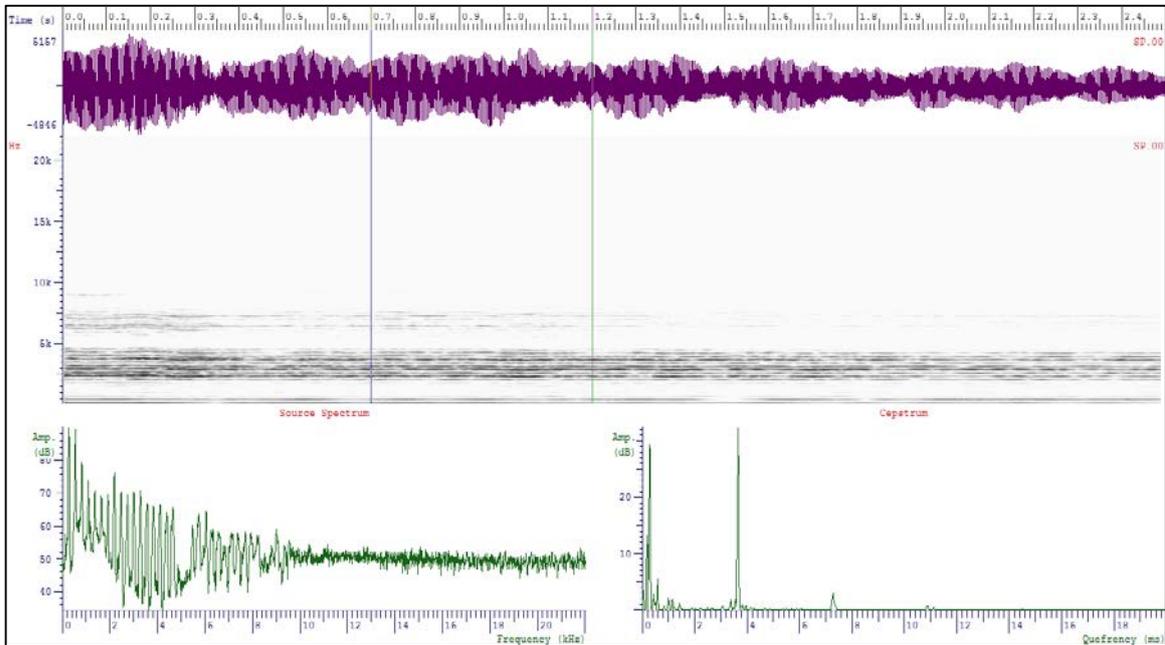


Fig. A.71.- Vocal “i” antes de aplicar el método para el Locutor 12

➤ (Vocal “i” después):

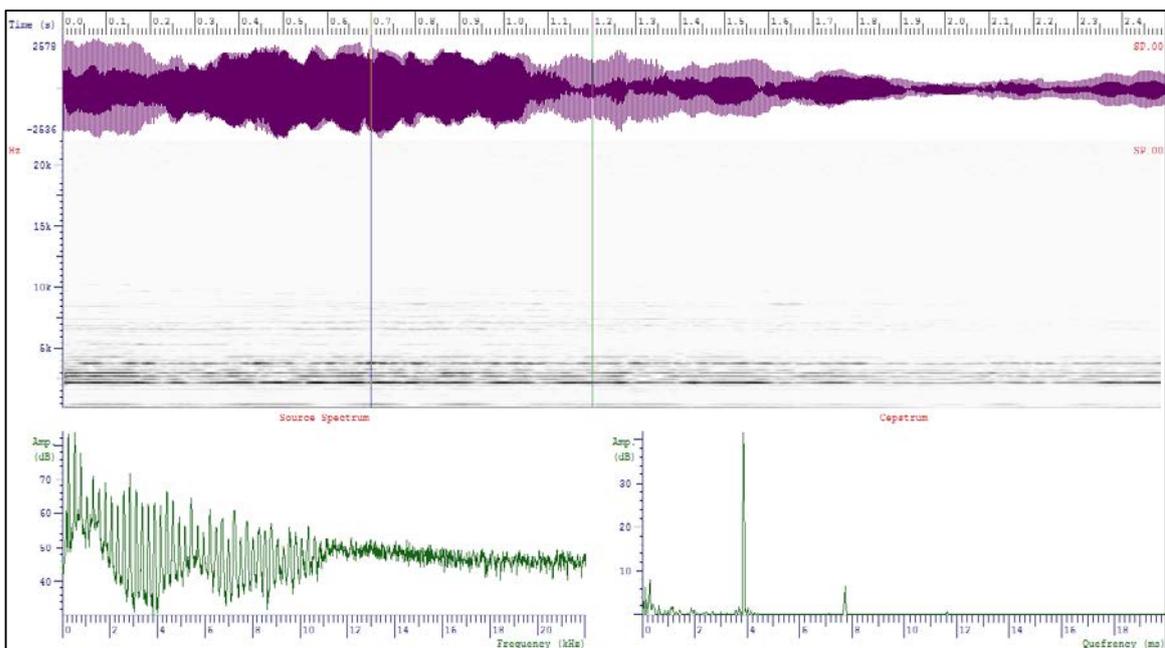


Fig. A.72.- Vocal “i” después de aplicar el método para el Locutor 12

❖ **Locutor 13:**

➤ *(Vocal “o” antes):*

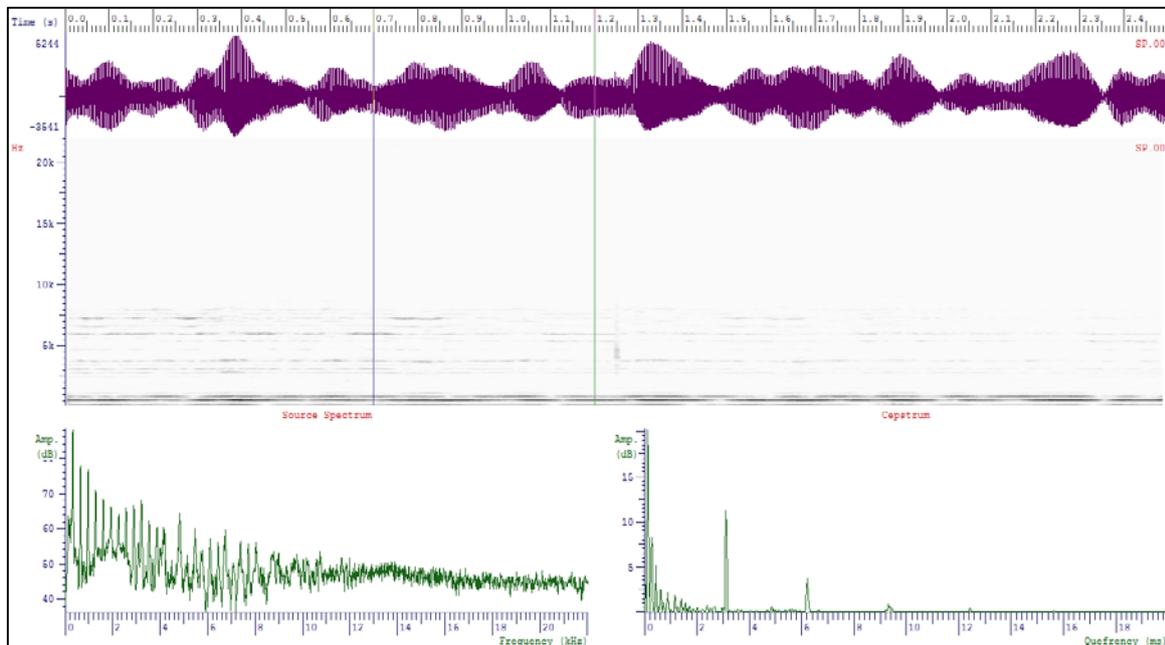


Fig. A.73.- Vocal “o” *antes* de aplicar el método para el Locutor 13

➤ *(Vocal “o” después):*

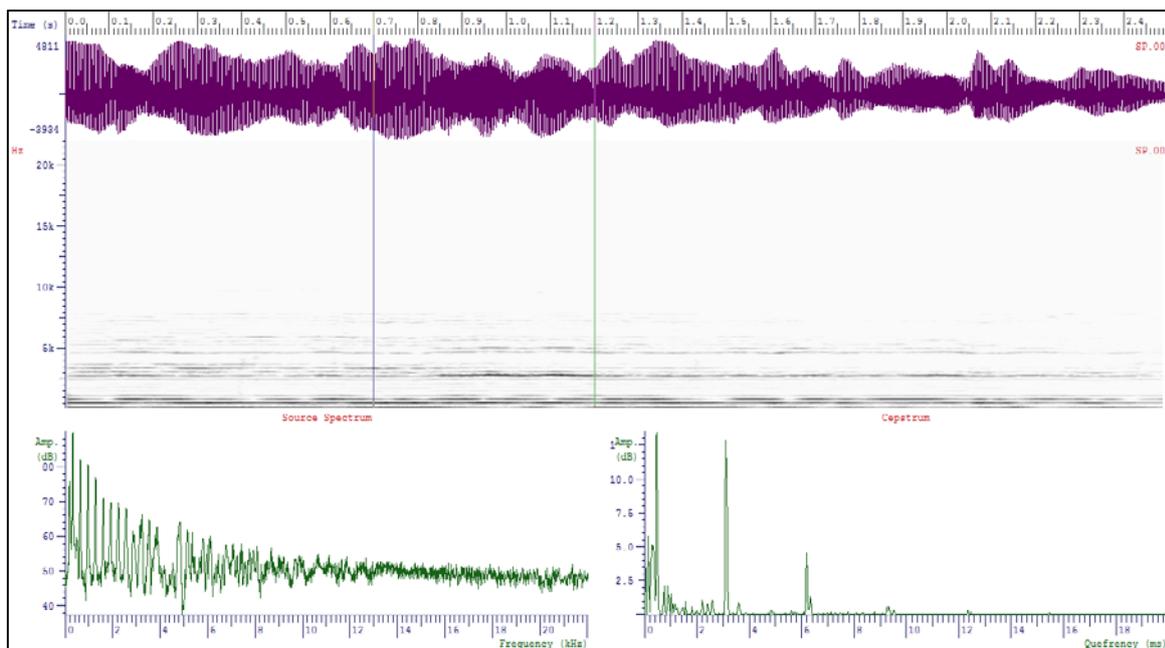


Fig. A.74.- Vocal “o” *después* de aplicar el método para el Locutor 13

➤ (Vocal “a” antes):

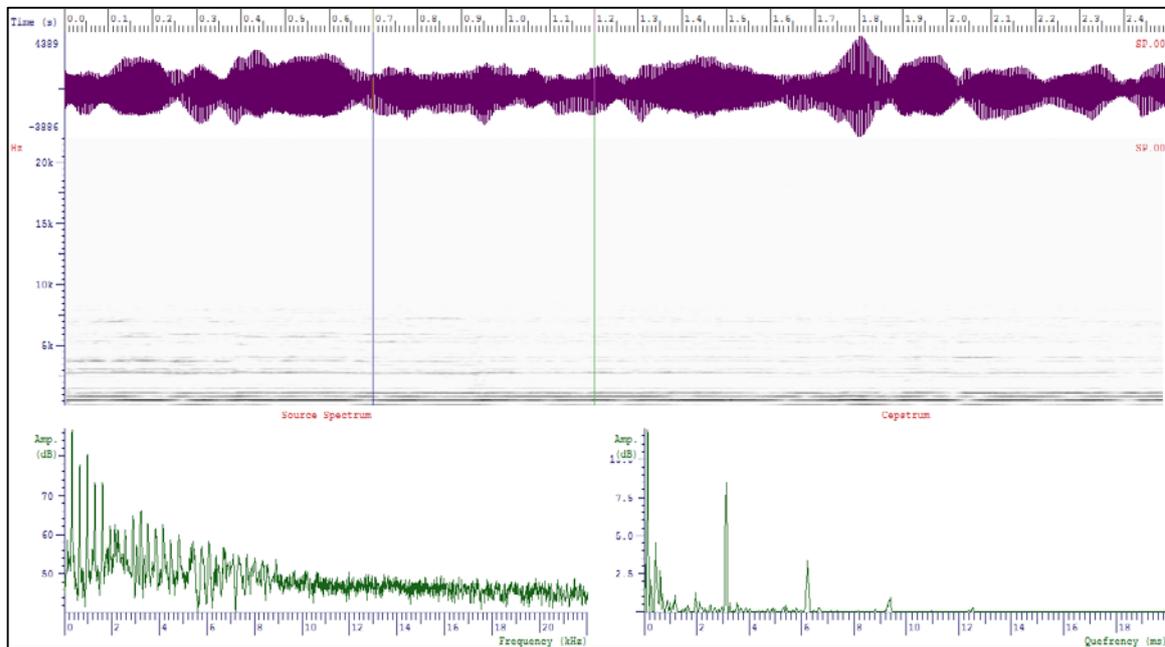


Fig. A.75.- Vocal “a” antes de aplicar el método para el Locutor 13

➤ (Vocal “a” después):

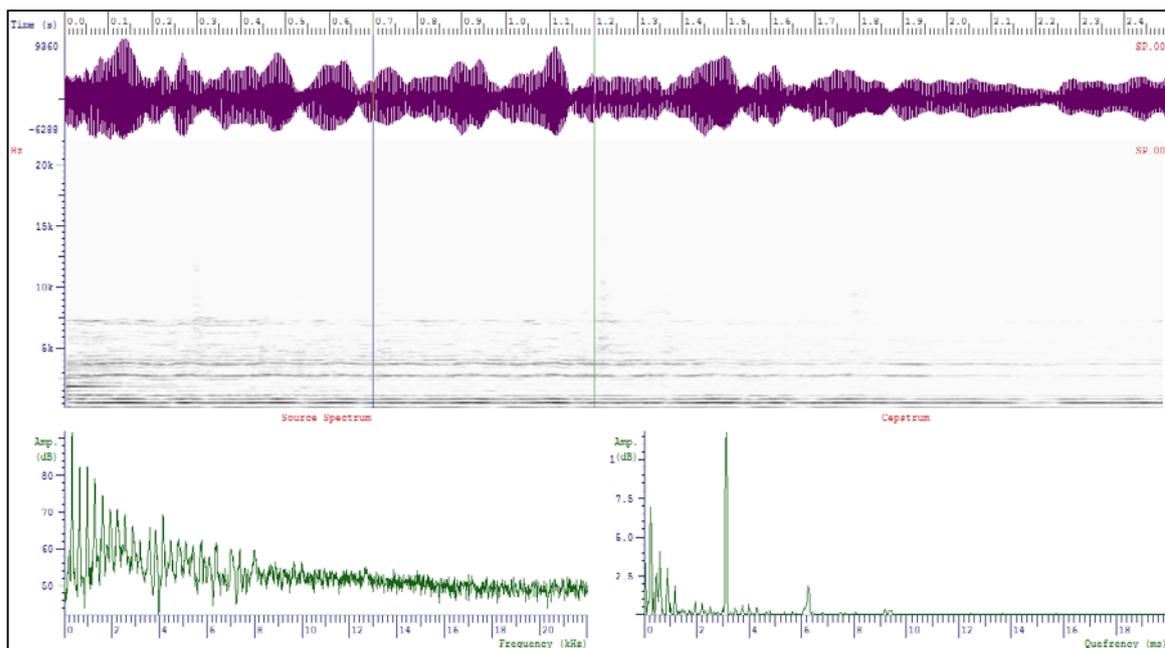


Fig. A.76.- Vocal “a” después de aplicar el método para el Locutor 13

➤ (Vocal “i” antes):

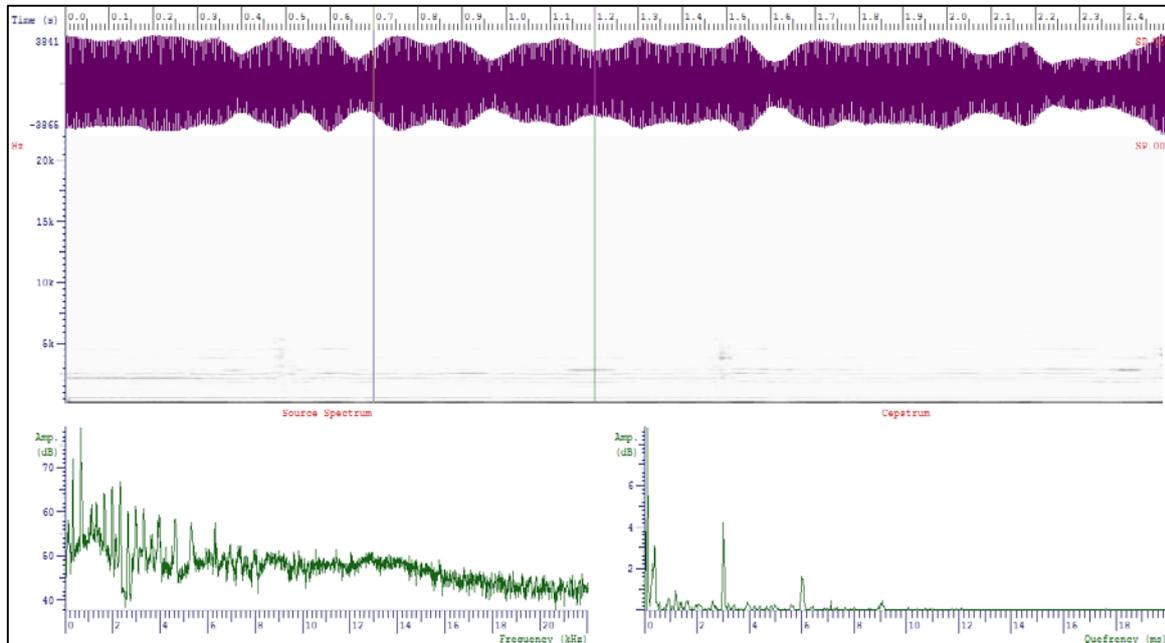


Fig. A.77.- Vocal “i” antes de aplicar el método para el Locutor 13

➤ (Vocal “i” después):

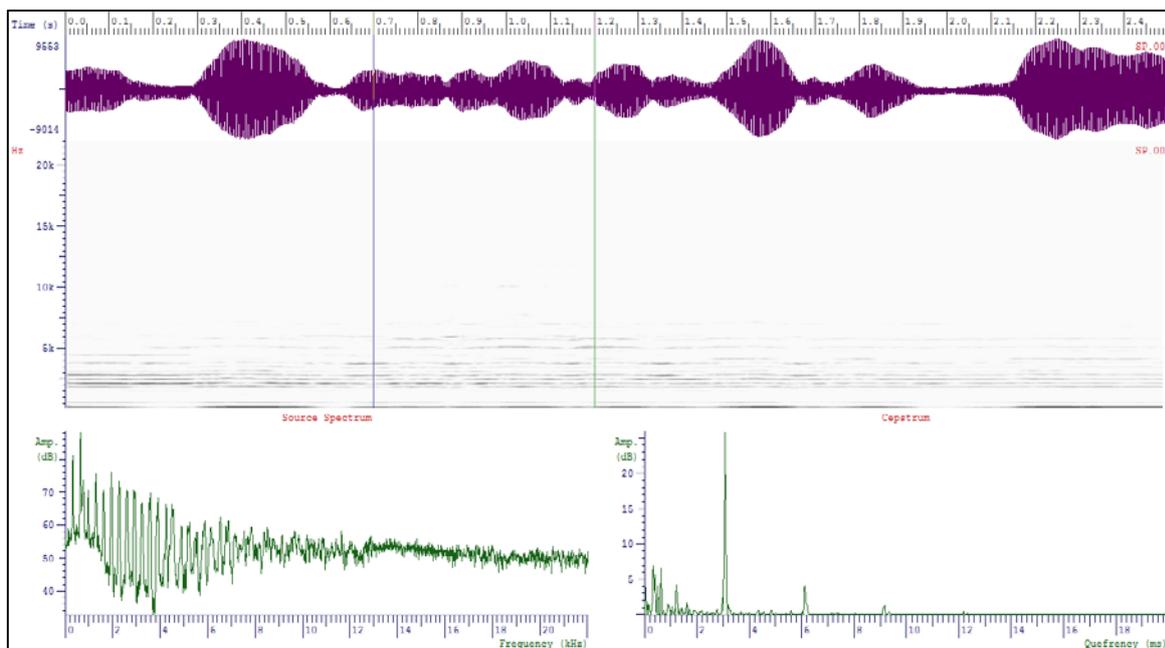


Fig. A.78.- Vocal “i” después de aplicar el método para el Locutor 13

❖ **Locutor 14:**

➤ *(Vocal “o” antes):*

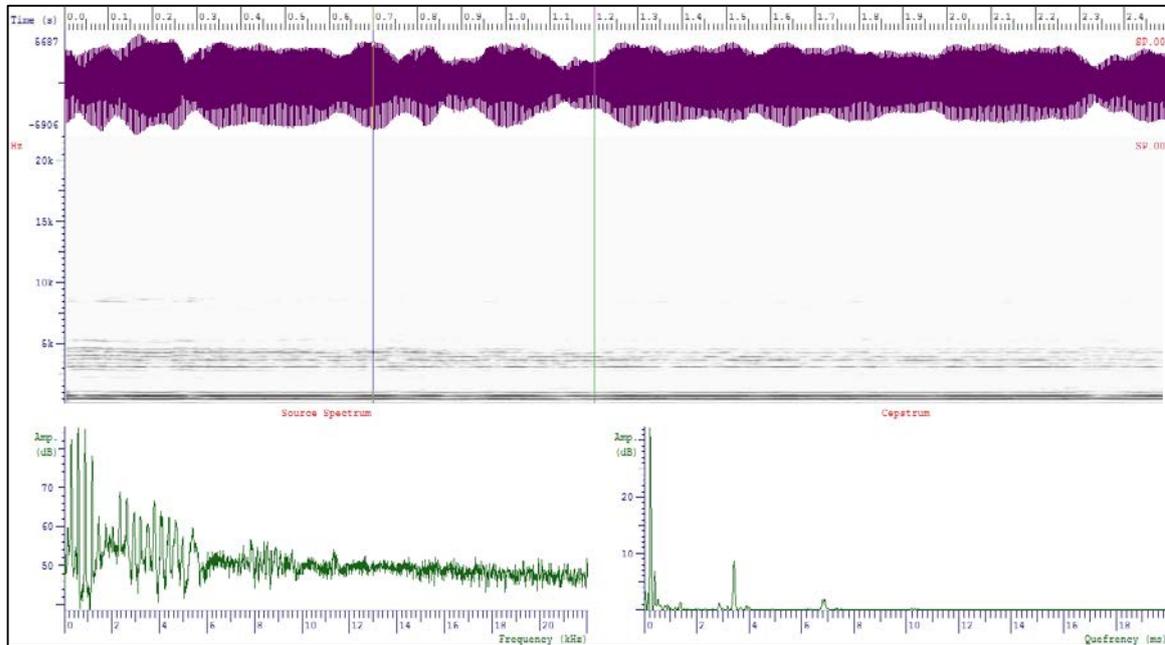


Fig. A.79.- Vocal “o” *antes* de aplicar el método para el Locutor 14

➤ *(Vocal “o” después):*

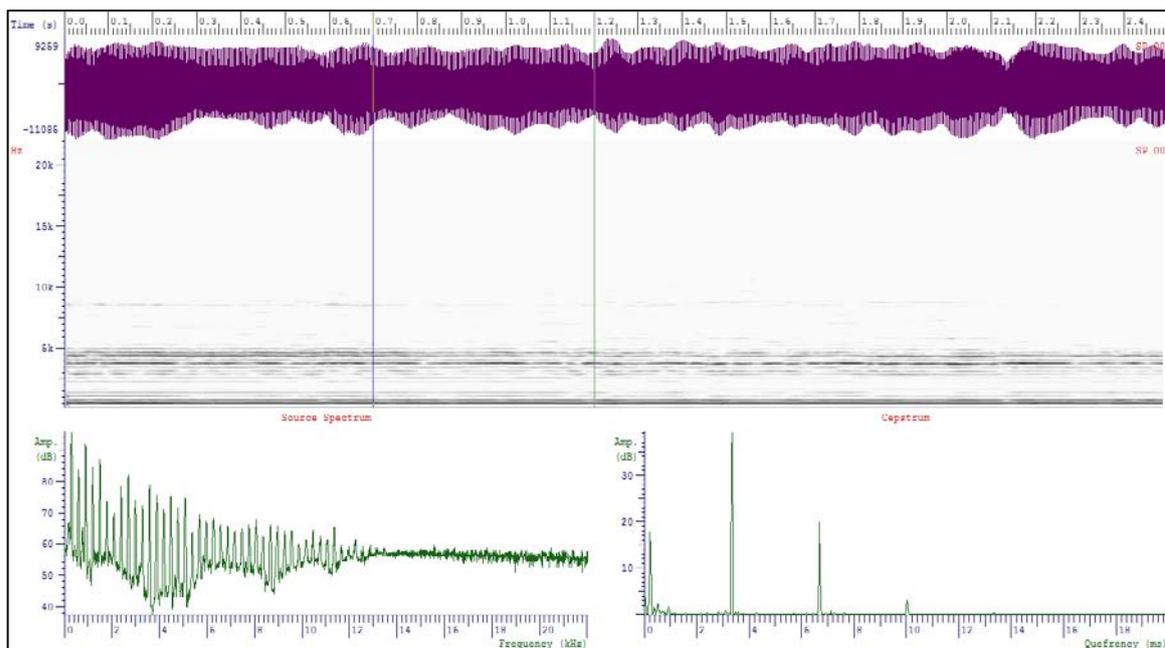


Fig. A.80.- Vocal “o” *después* de aplicar el método para el Locutor 14

➤ (Vocal “a” antes):

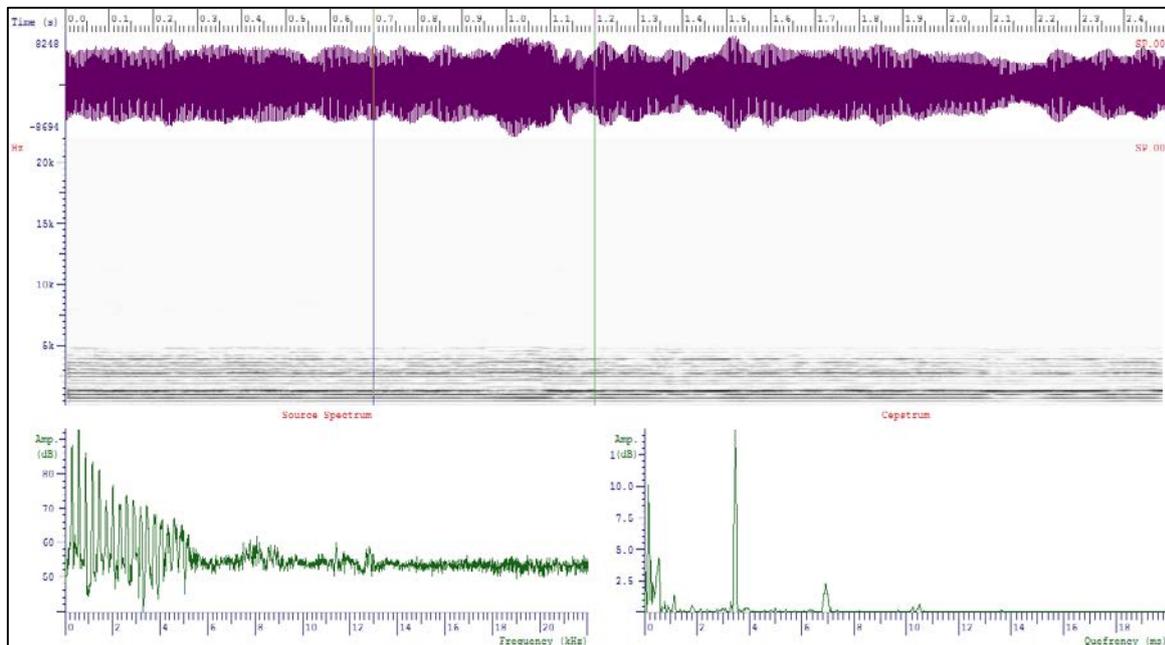


Fig. A.81.- Vocal “a” antes de aplicar el método para el Locutor 14

➤ (Vocal “a” después):

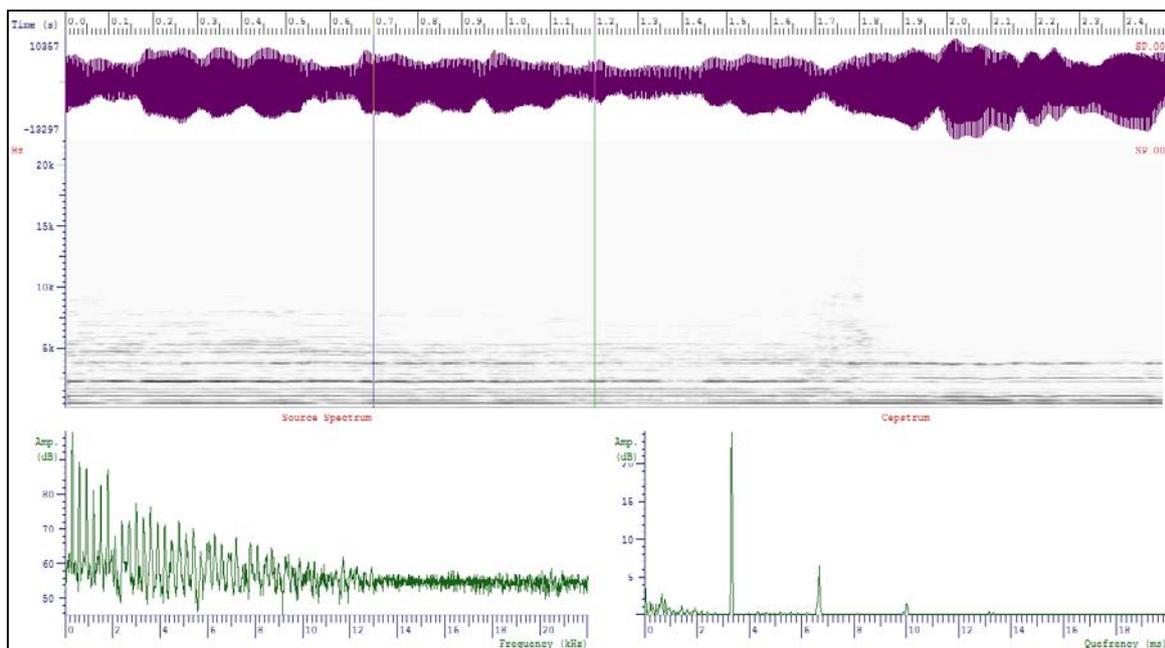


Fig. A.82.- Vocal “a” después de aplicar el método para el Locutor 14

➤ (Vocal “i” antes):

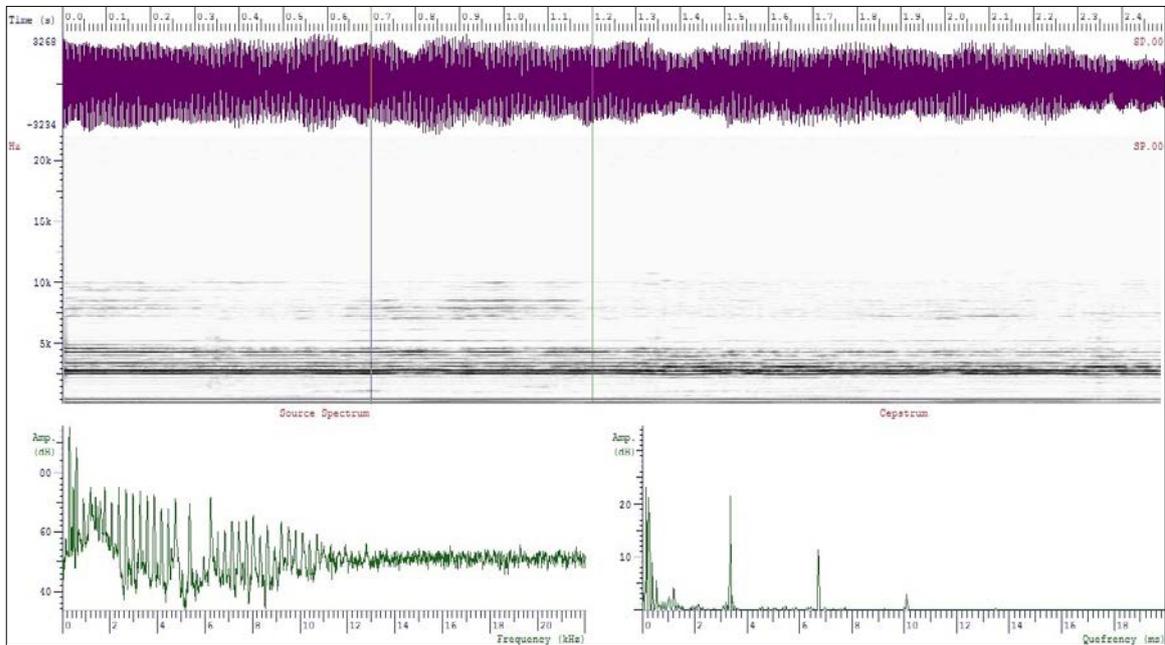


Fig. A.83.- Vocal “i” antes de aplicar el método para el Locutor 14

➤ (Vocal “i” después):

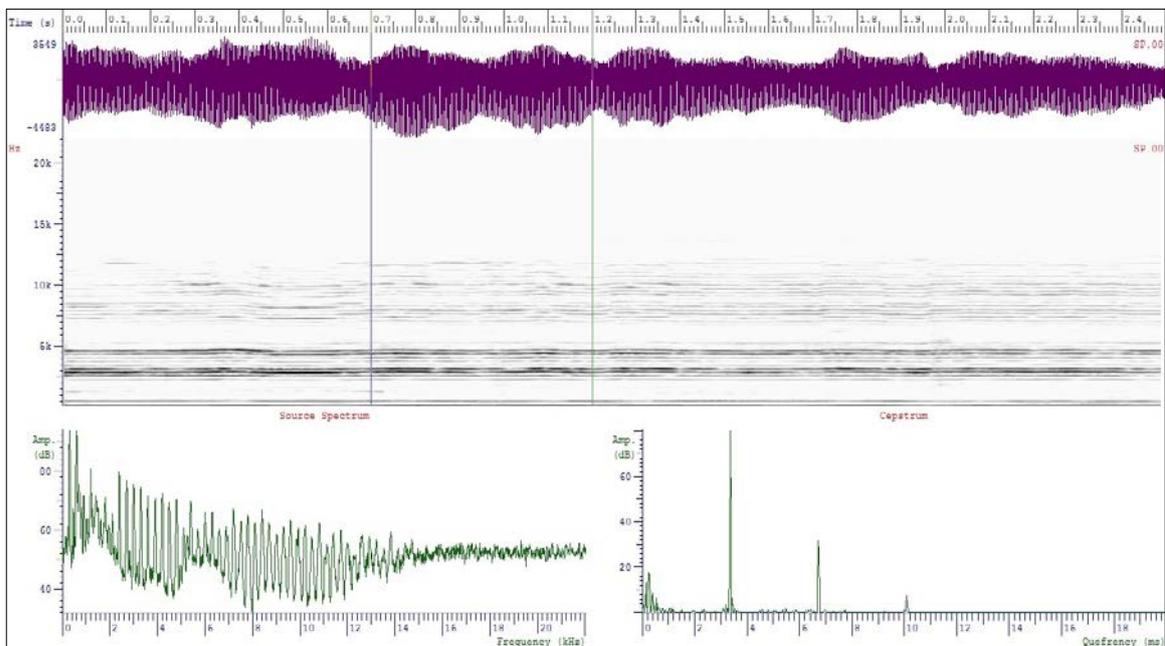


Fig. A.84.- Vocal “i” después de aplicar el método para el Locutor 14

❖ **Locutor 15:**

➤ *(Vocal “o” antes):*

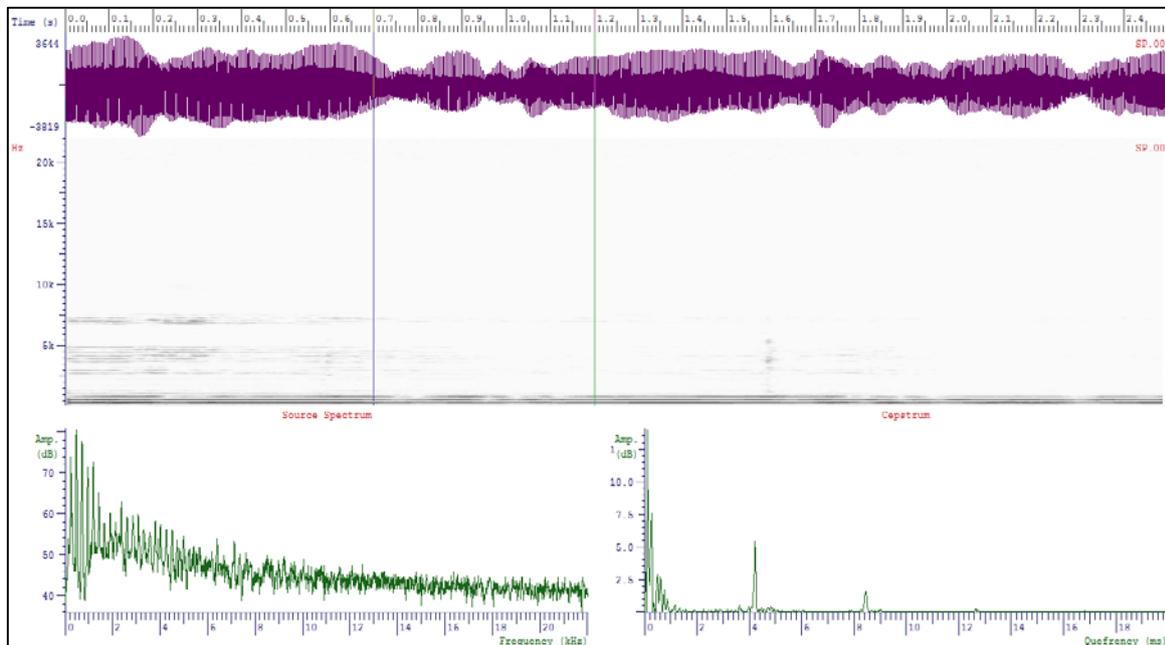


Fig. A.85.- Vocal “o” antes de aplicar el método para el Locutor 15

➤ *(Vocal “o” después):*

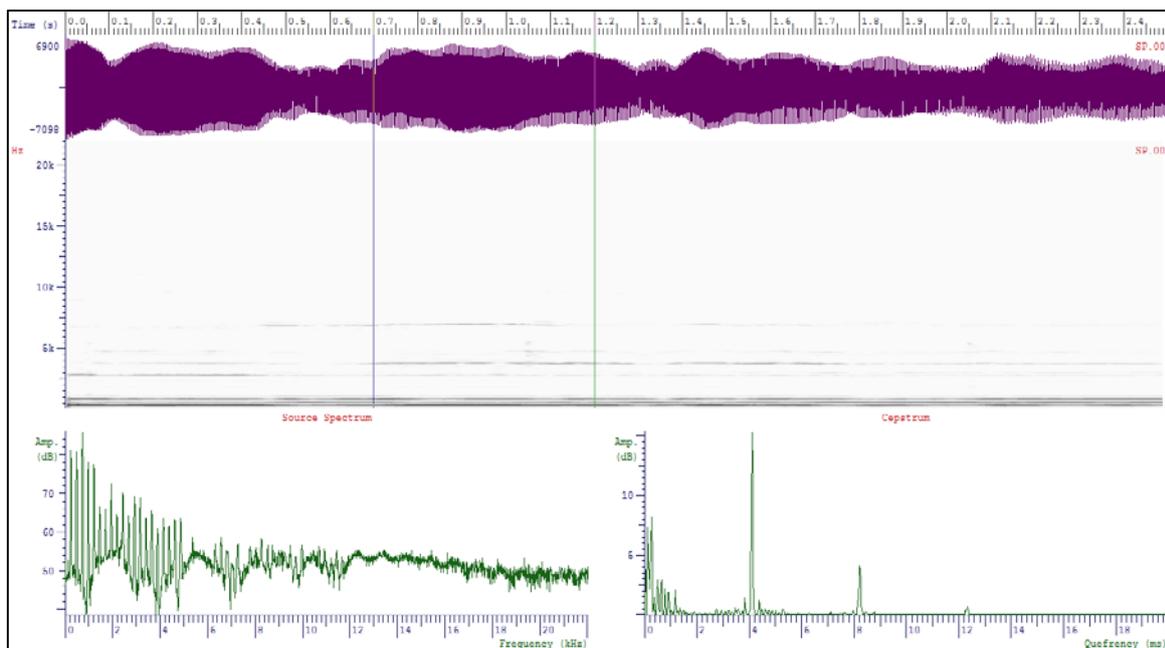


Fig. A.86.- Vocal “o” después de aplicar el método para el Locutor 15

➤ (Vocal “a” antes):

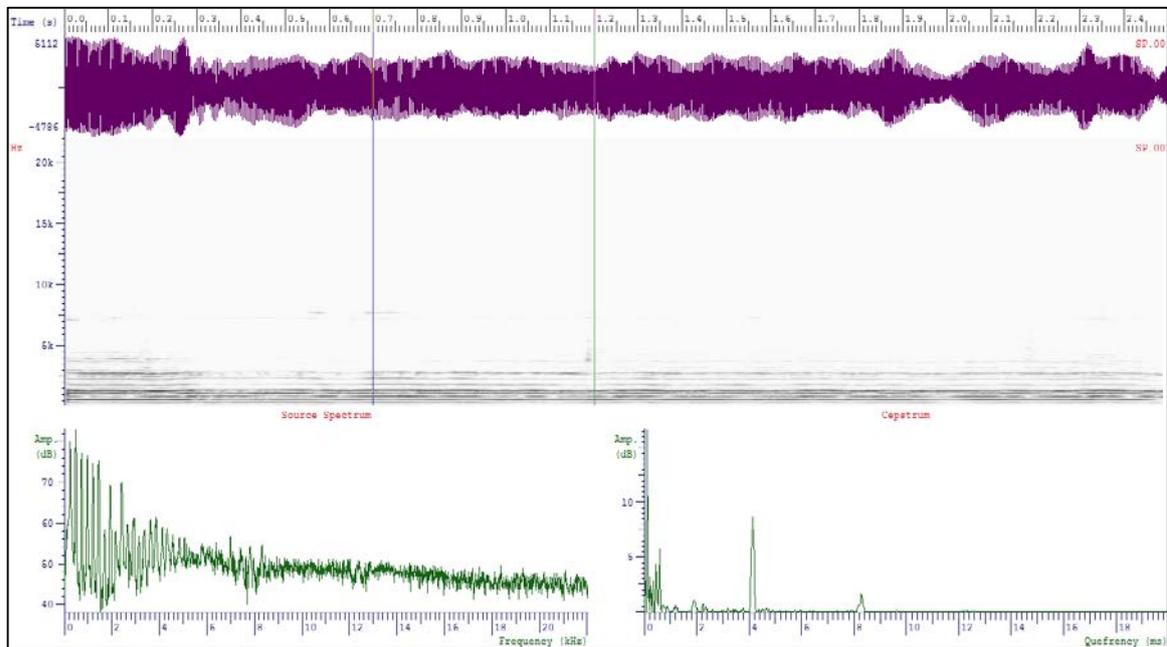


Fig. A.87.- Vocal “a” antes de aplicar el método para el Locutor 15

➤ (Vocal “a” después):

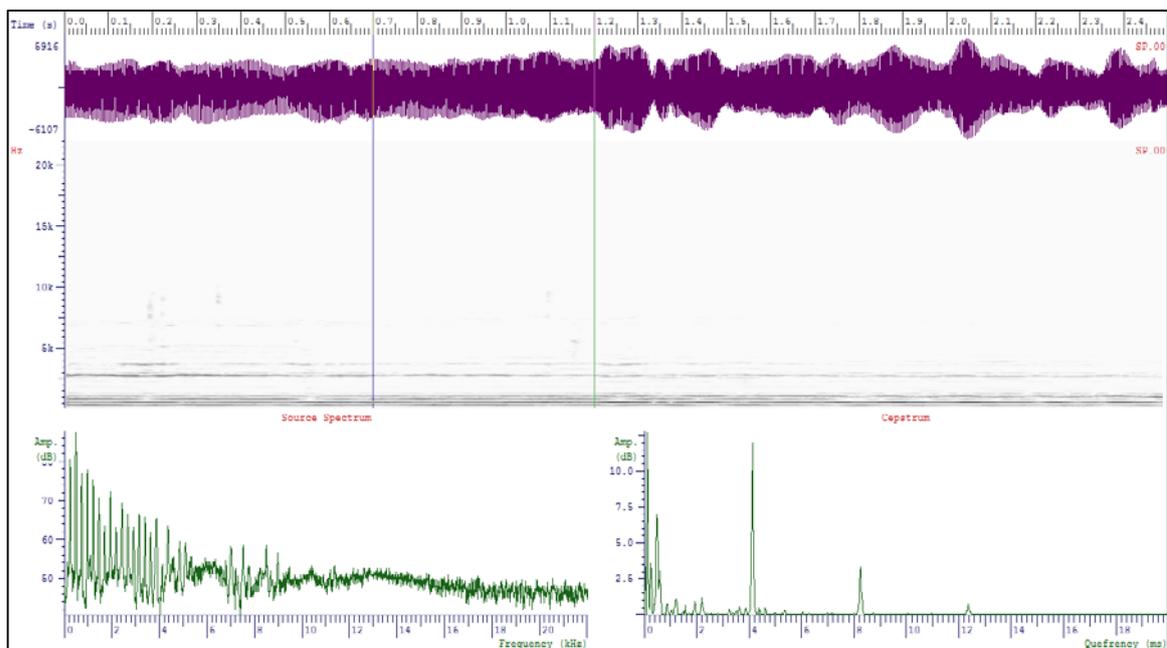


Fig. A.88.- Vocal “a” después de aplicar el método para el Locutor 15

➤ (Vocal “i” antes):

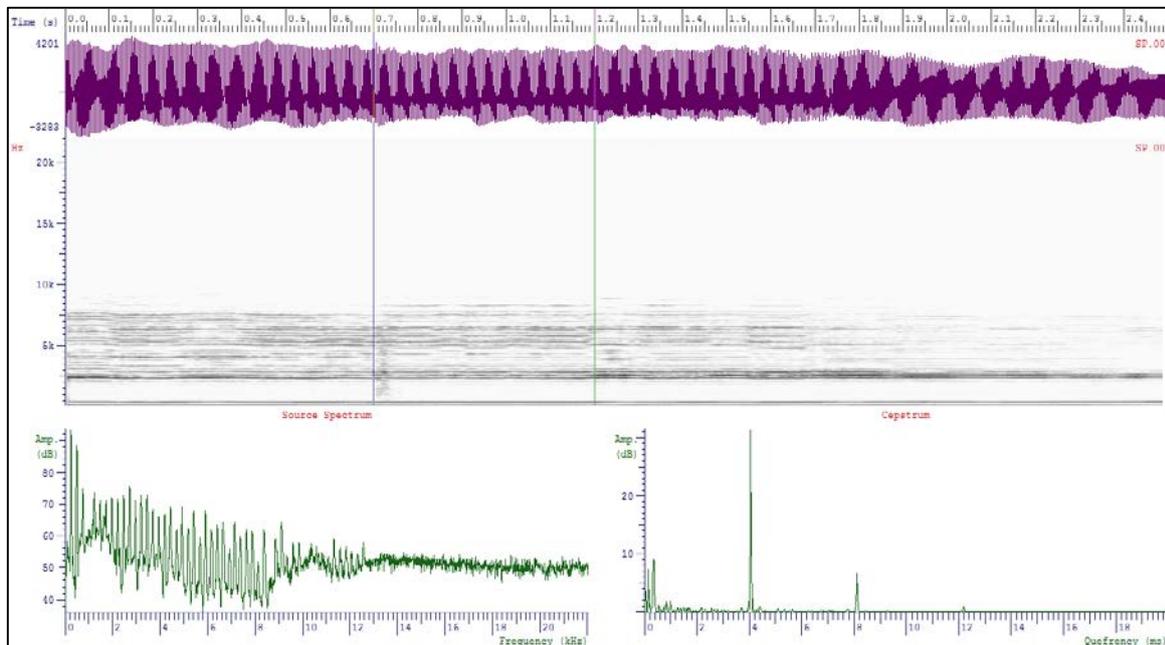


Fig. A.89.- Vocal “i” antes de aplicar el método para el Locutor 15

➤ (Vocal “i” después):

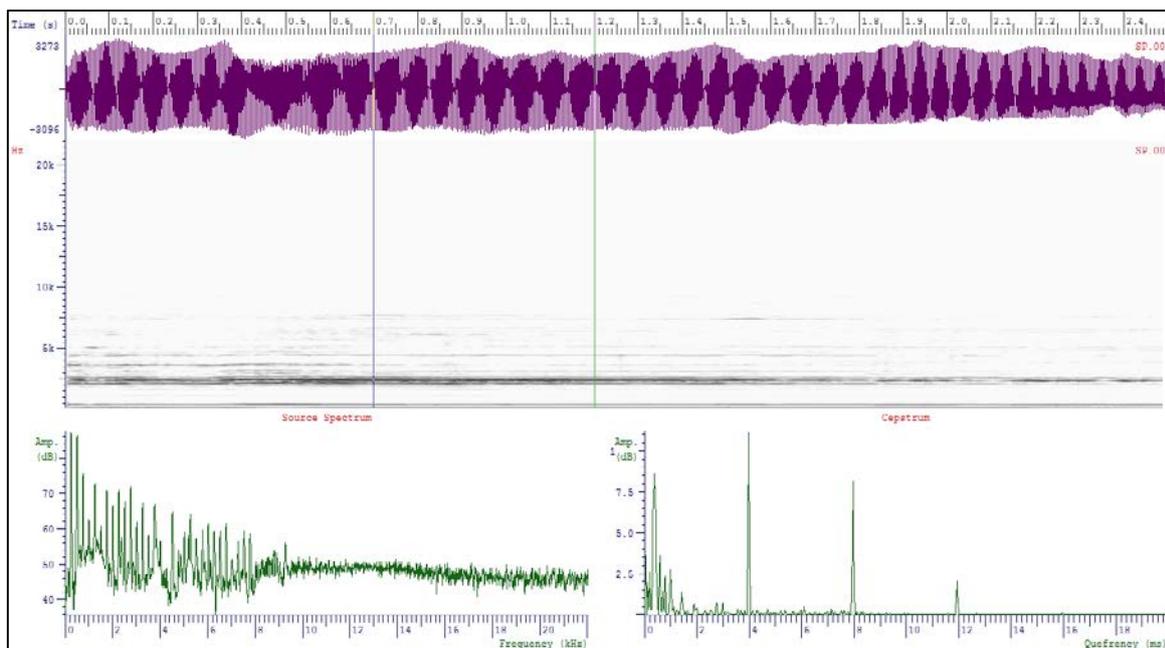


Fig. A.90.- Vocal “i” después de aplicar el método para el Locutor 15

❖ **Locutor 16:**

➤ *(Vocal “o” antes):*

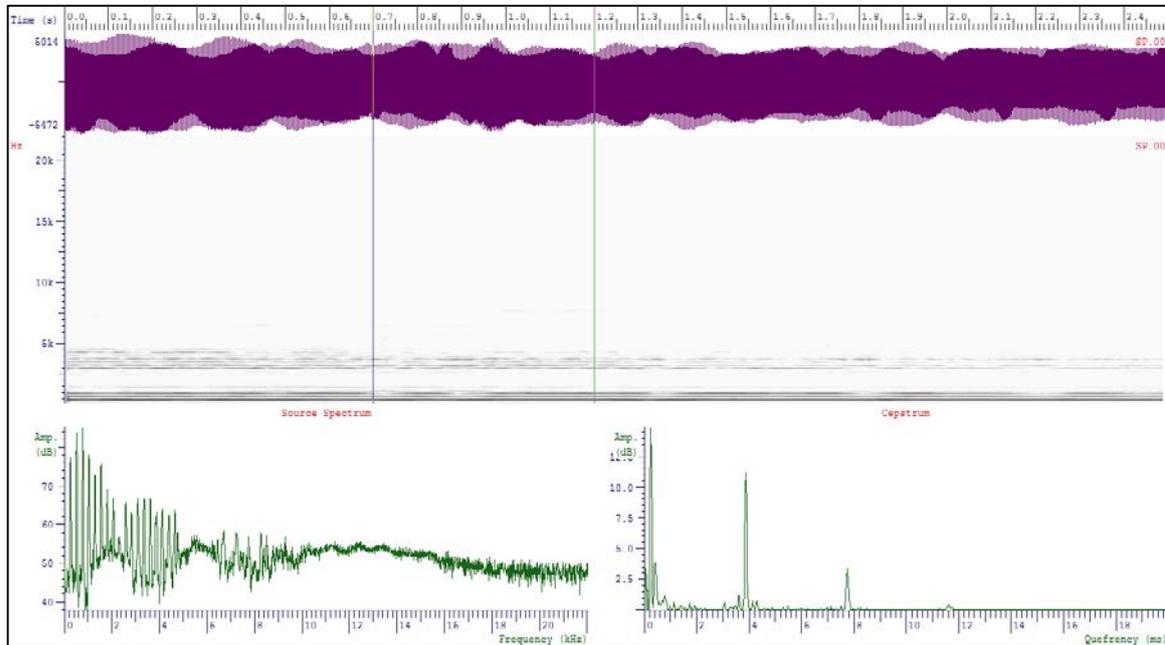


Fig. A.91.- Vocal “o” *antes* de aplicar el método para el Locutor 16

➤ *(Vocal “o” después):*

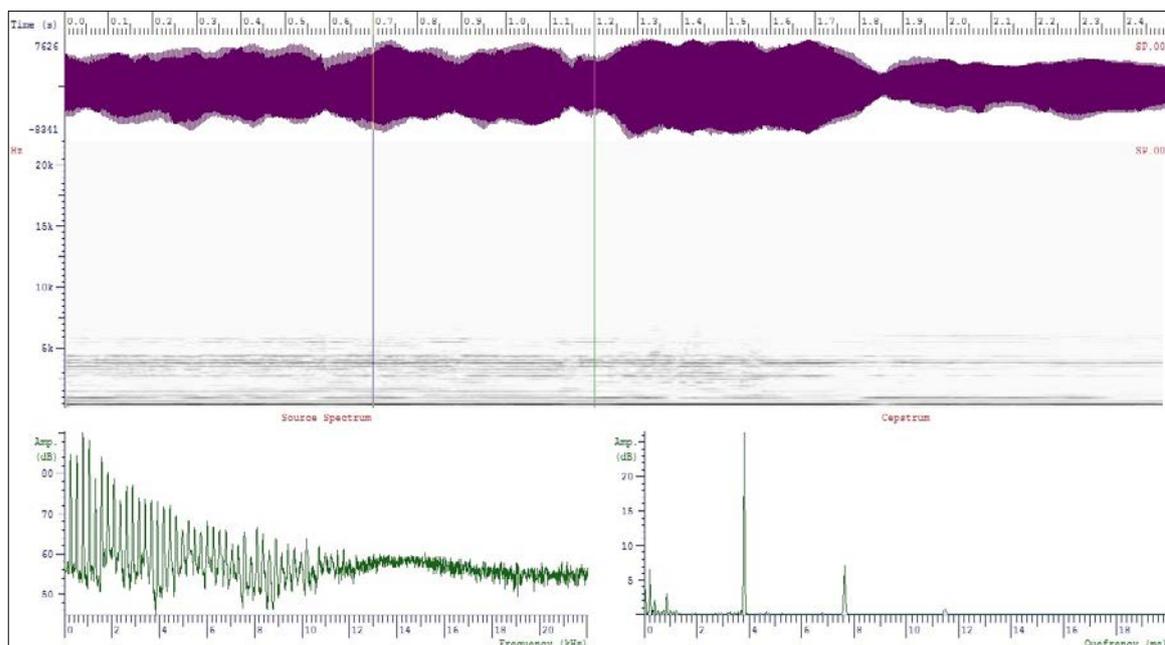


Fig. A.92.- Vocal “o” *después* de aplicar el método para el Locutor 16

➤ (Vocal “a” antes):

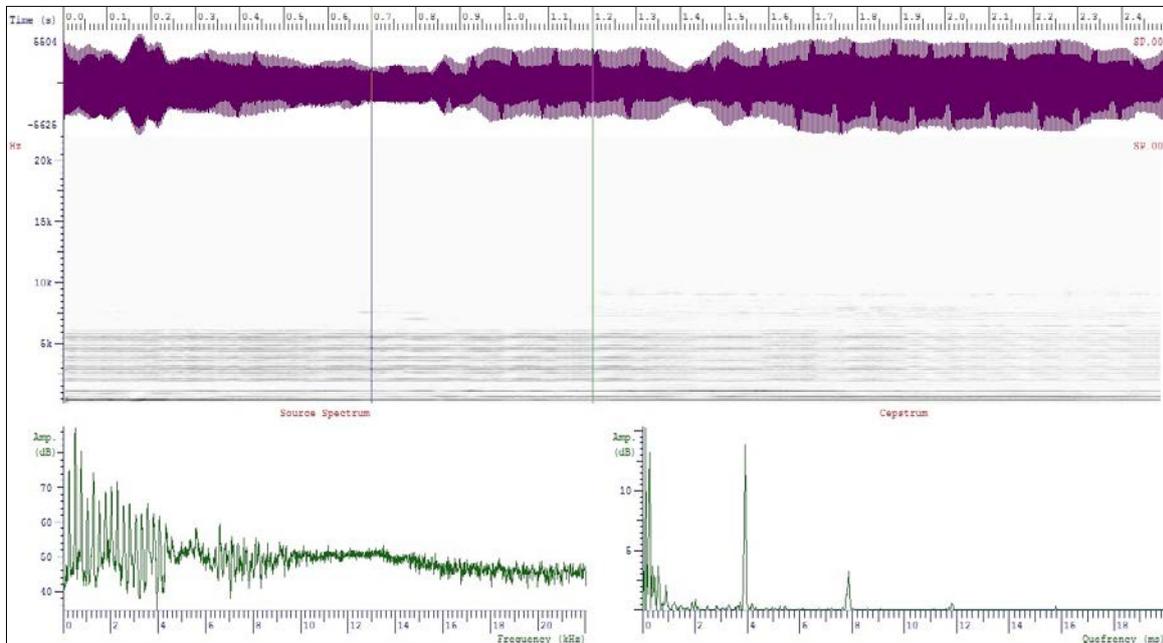


Fig. A.93.- Vocal “a” antes de aplicar el método para el Locutor 16

➤ (Vocal “a” después):

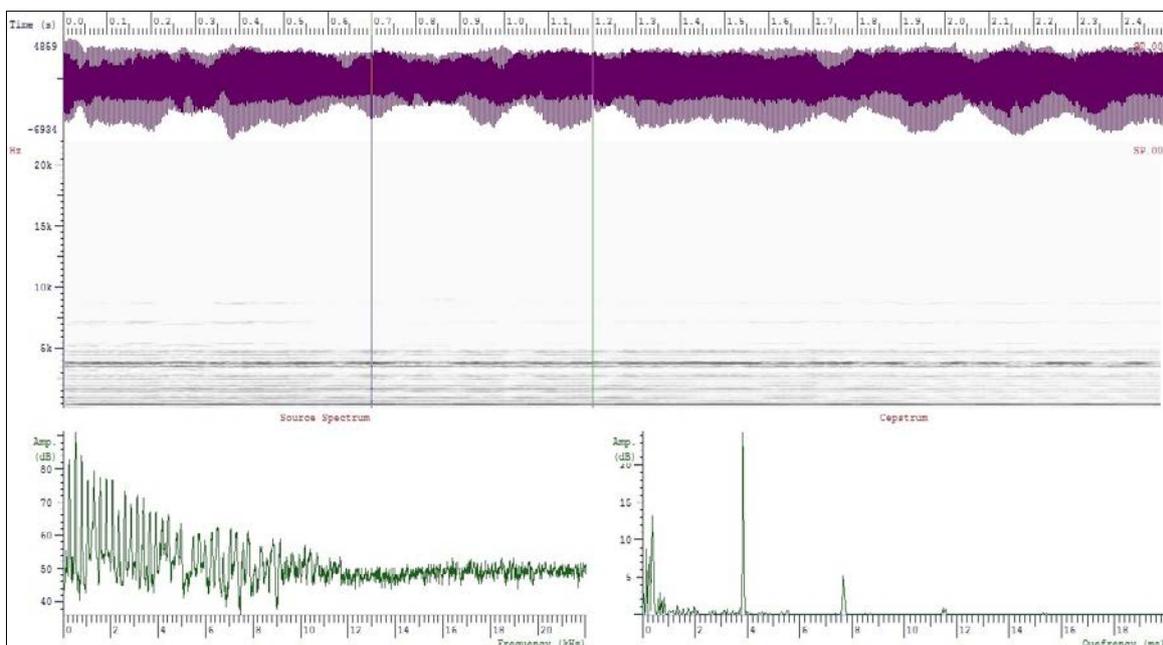


Fig. A.94.- Vocal “a” después de aplicar el método para el Locutor 16

➤ (Vocal “i” antes):

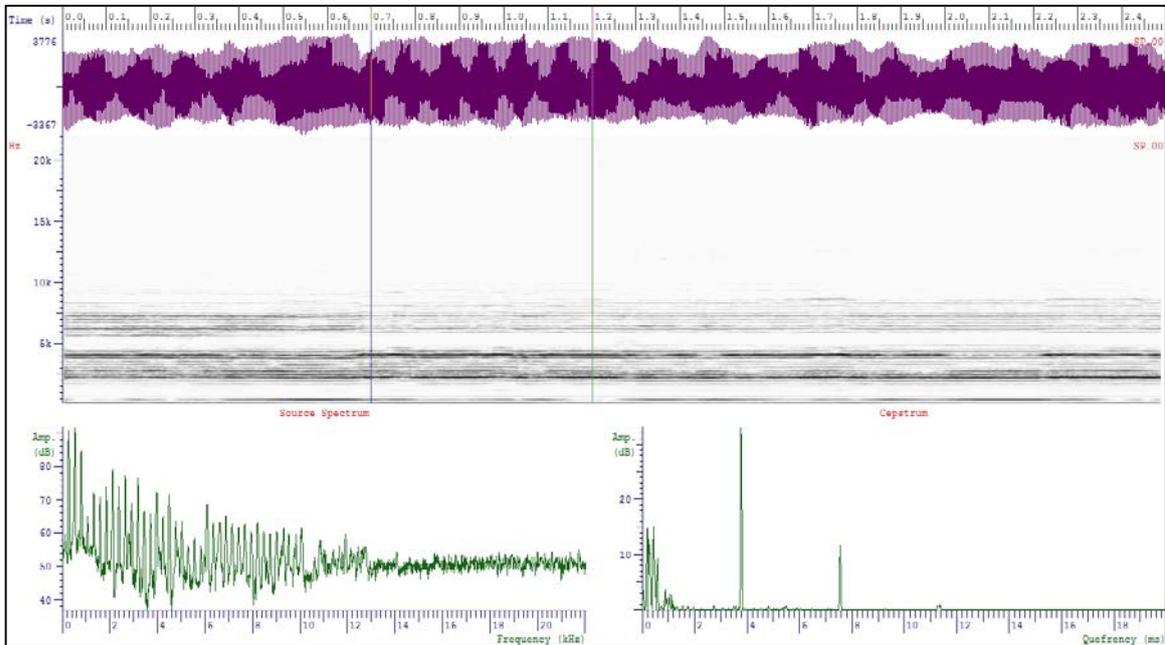


Fig. A.95.- Vocal “i” antes de aplicar el método para el Locutor 16

➤ (Vocal “i” después):

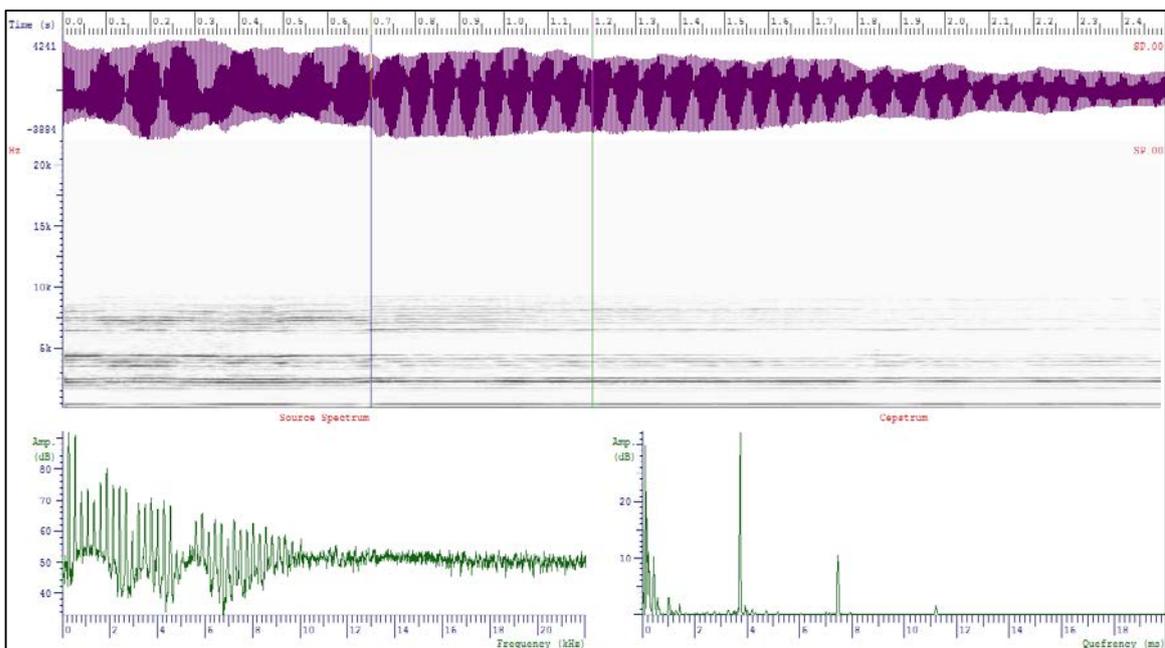


Fig. A.96.- Vocal “i” después de aplicar el método para el Locutor 16

❖ **Locutor 17:**

➤ *(Vocal “o” antes):*

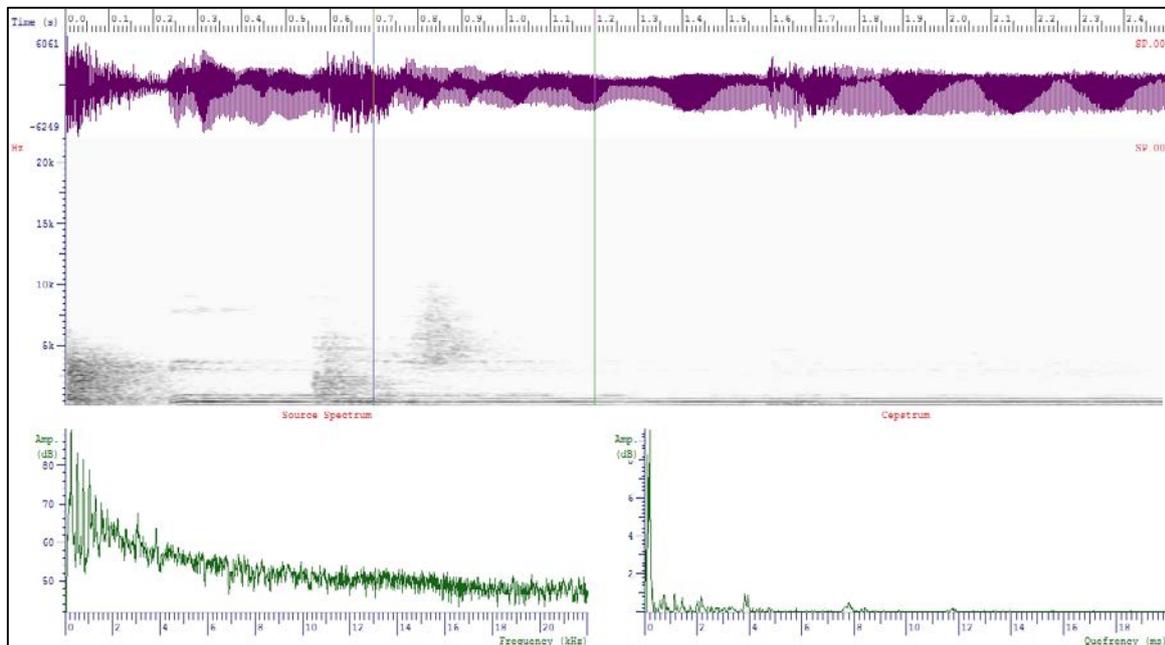


Fig. A.97.- Vocal “o” *antes* de aplicar el método para el Locutor 17

➤ *(Vocal “o” después):*

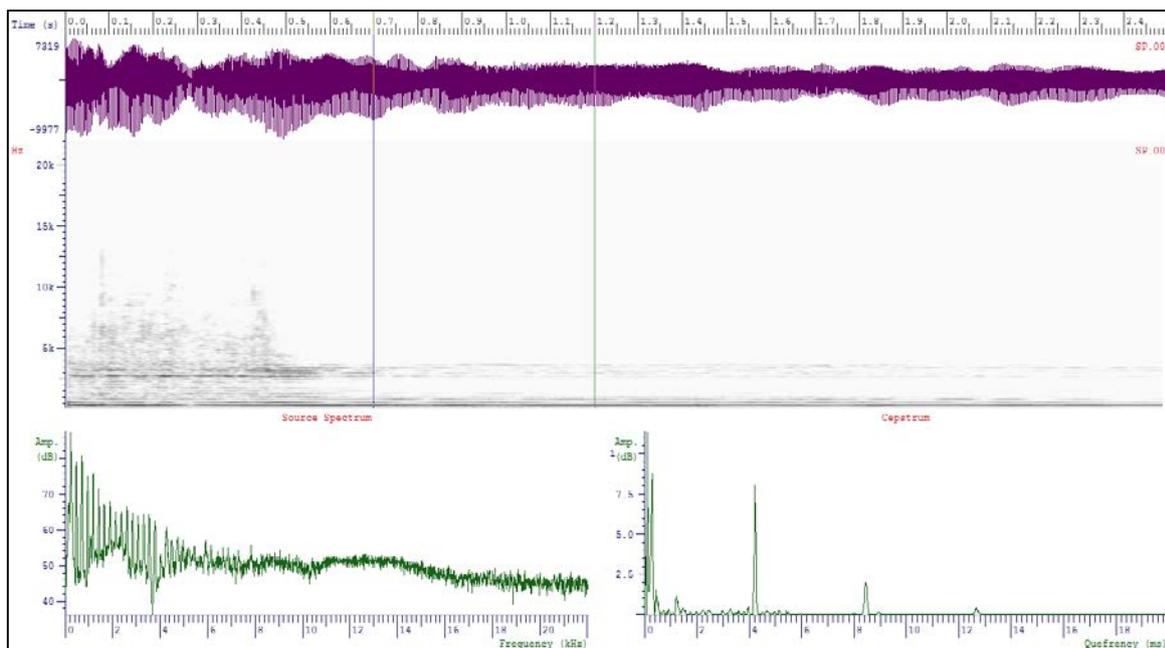


Fig. A.98.- Vocal “o” *después* de aplicar el método para el Locutor 17

➤ (Vocal “a” antes):

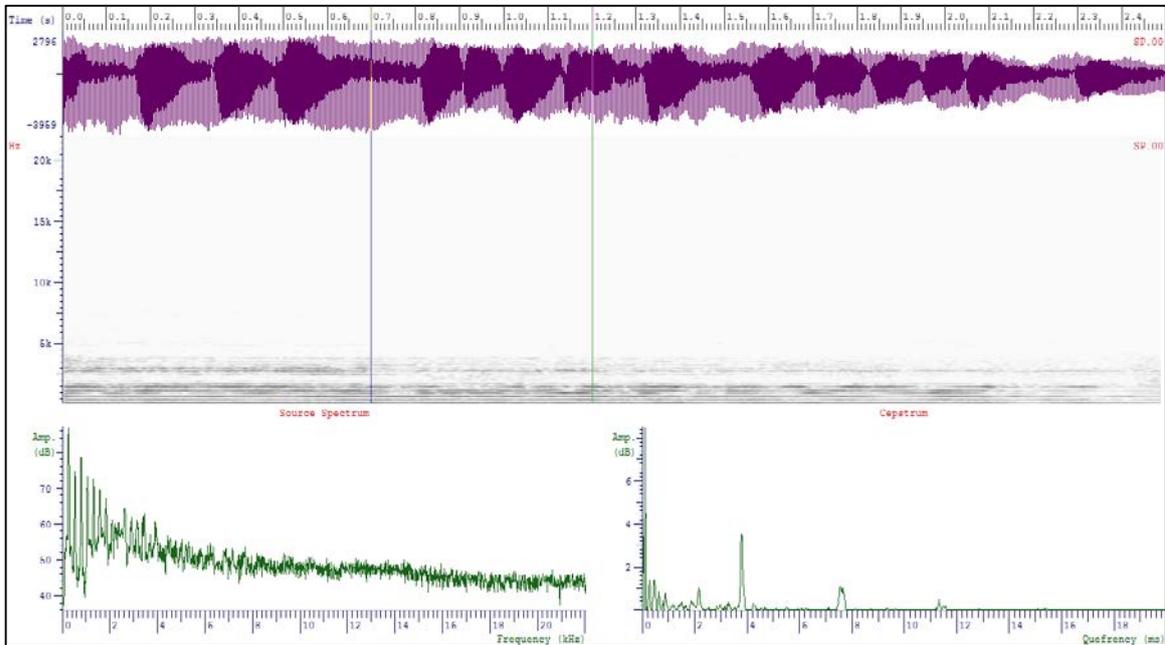


Fig. A.99.- Vocal “a” antes de aplicar el método para el Locutor 17

➤ (Vocal “a” después):

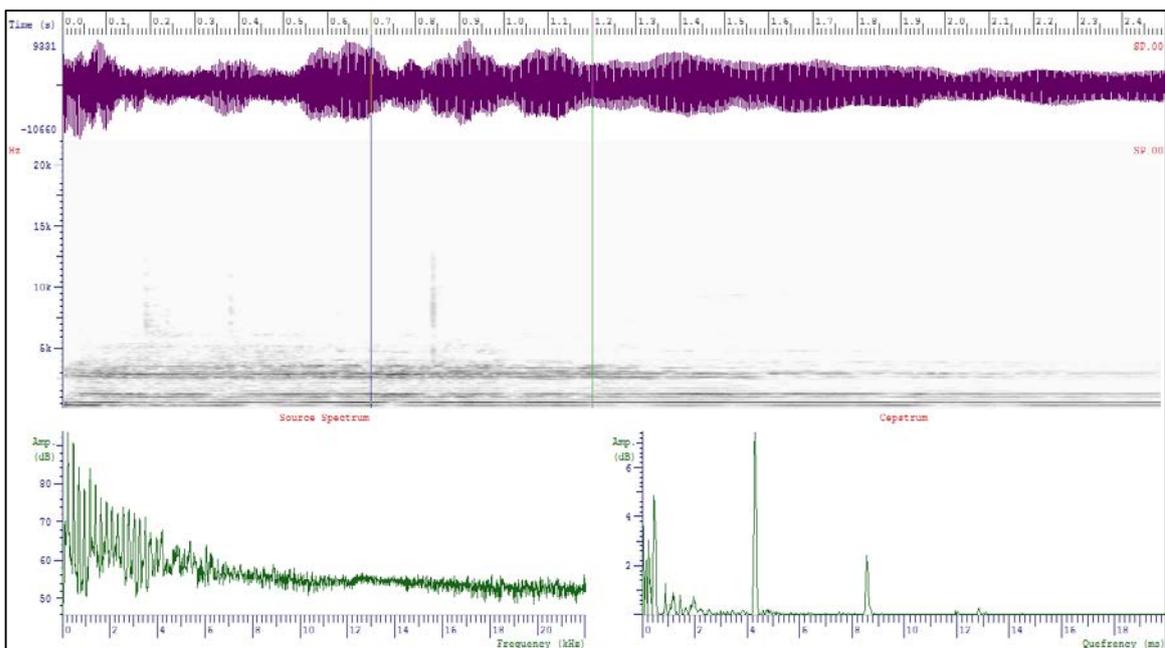


Fig. A.100.- Vocal “a” después de aplicar el método para el Locutor 17

➤ (Vocal “i” antes):

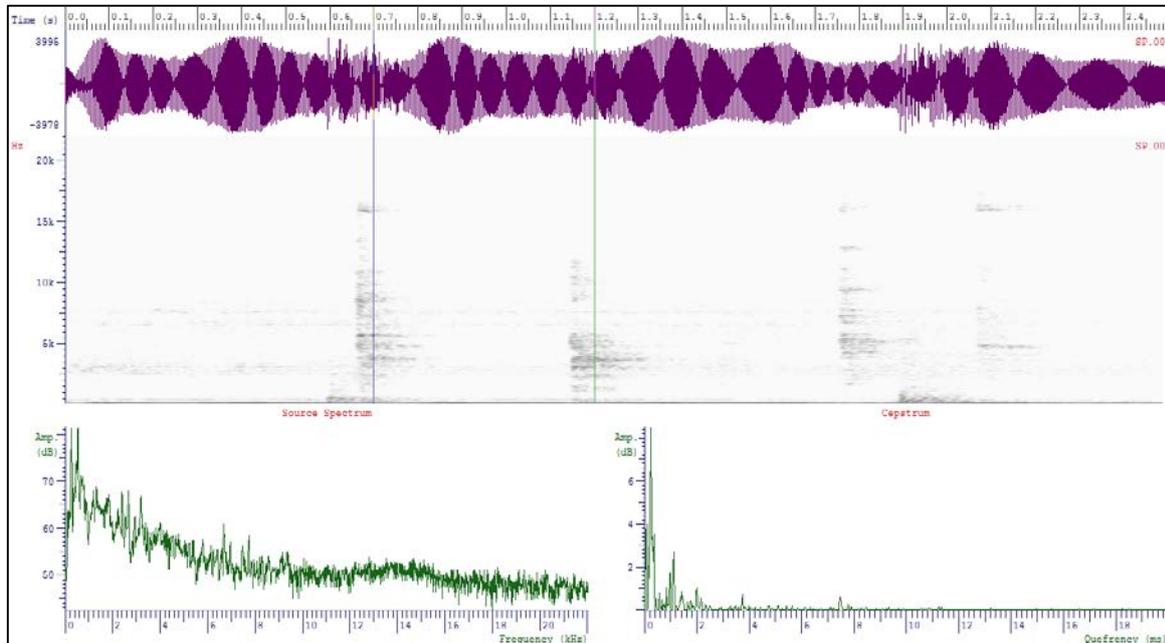


Fig. A.101.- Vocal “i” antes de aplicar el método para el Locutor 17

➤ (Vocal “i” después):

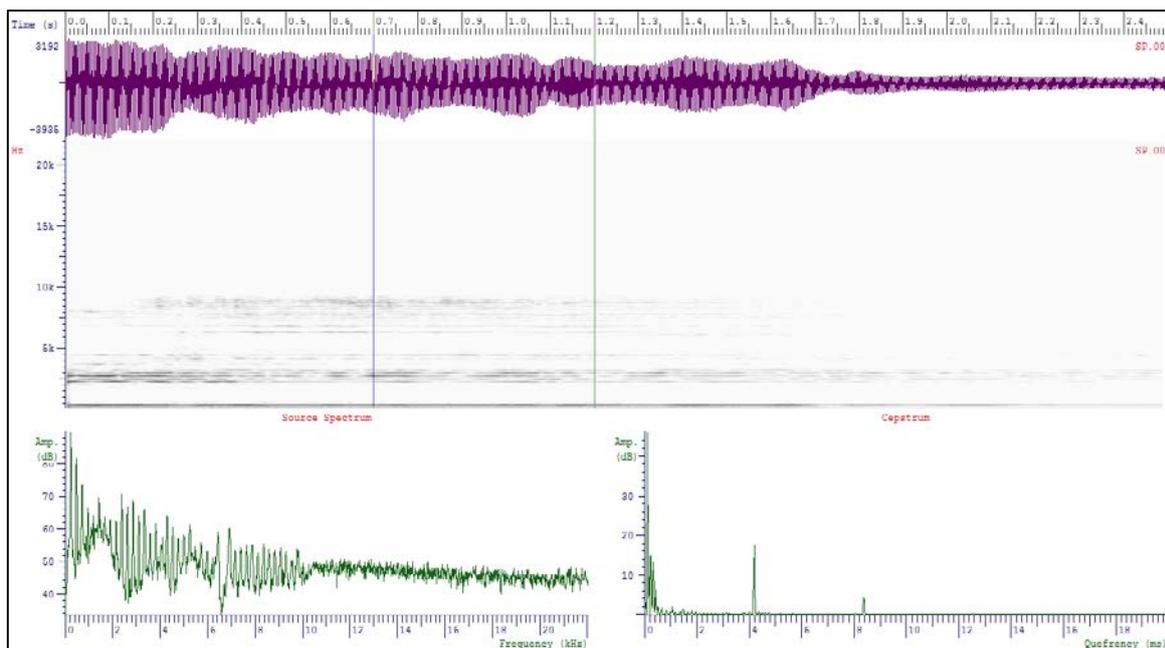


Fig. A.102.- Vocal “i” después de aplicar el método para el Locutor 17

❖ **Locutor 18:**

➤ *(Vocal “o” antes):*

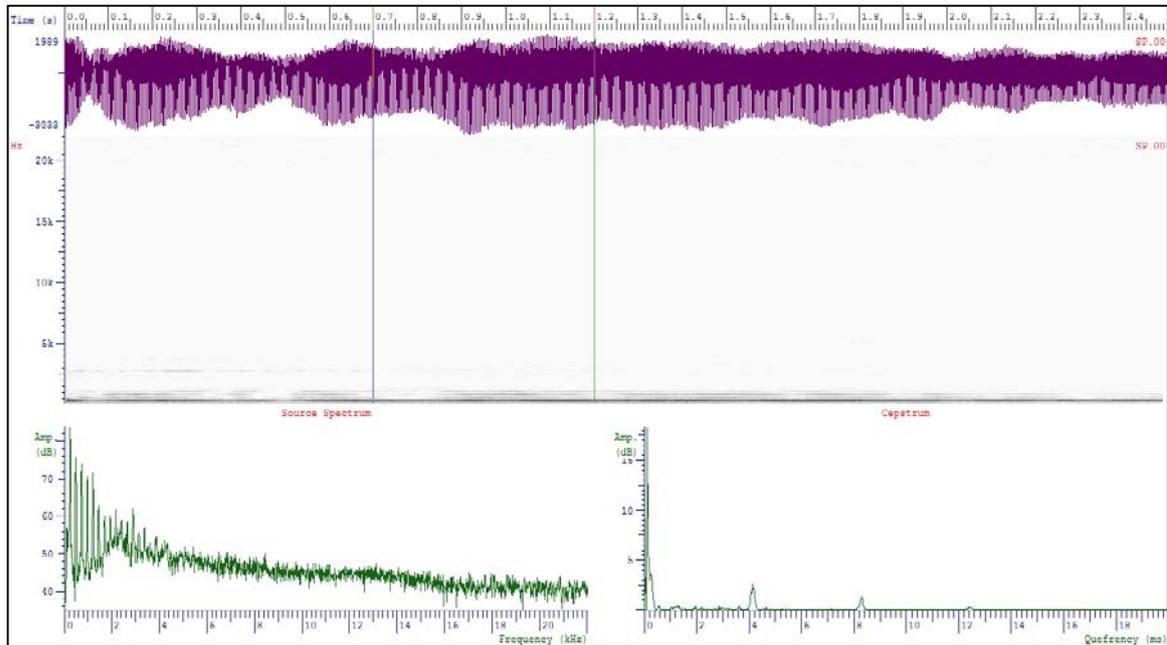


Fig. A.103.- Vocal “o” antes de aplicar el método para el Locutor 18

➤ *(Vocal “o” después):*

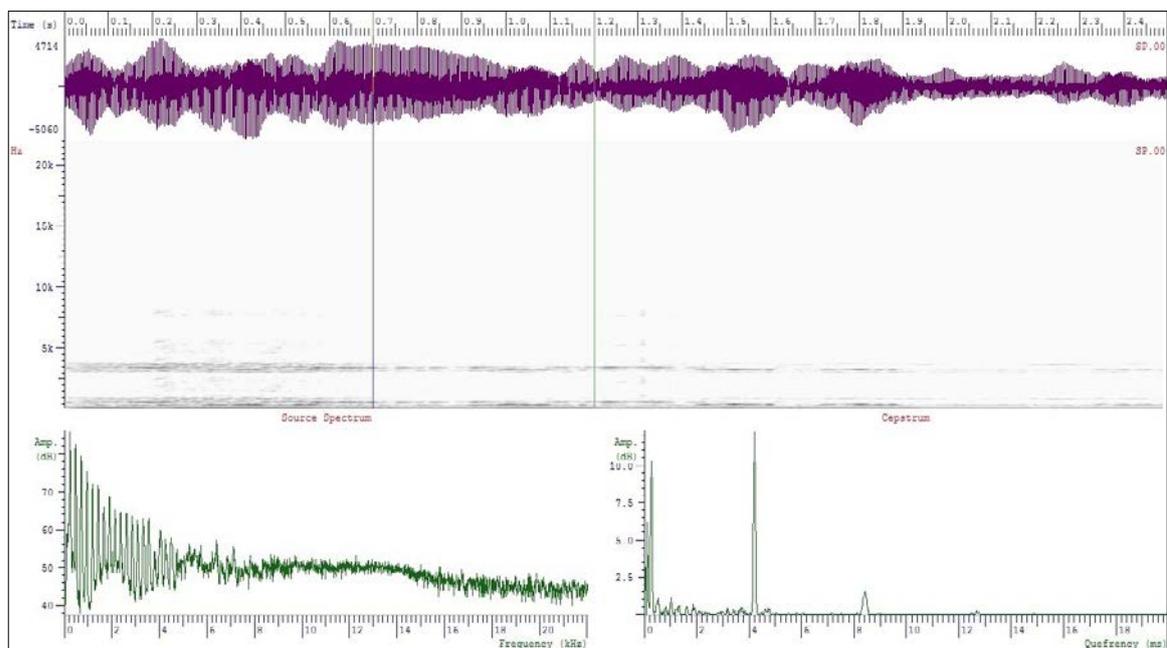


Fig. A.104.- Vocal “o” después de aplicar el método para el Locutor 18

➤ (Vocal “a” antes):

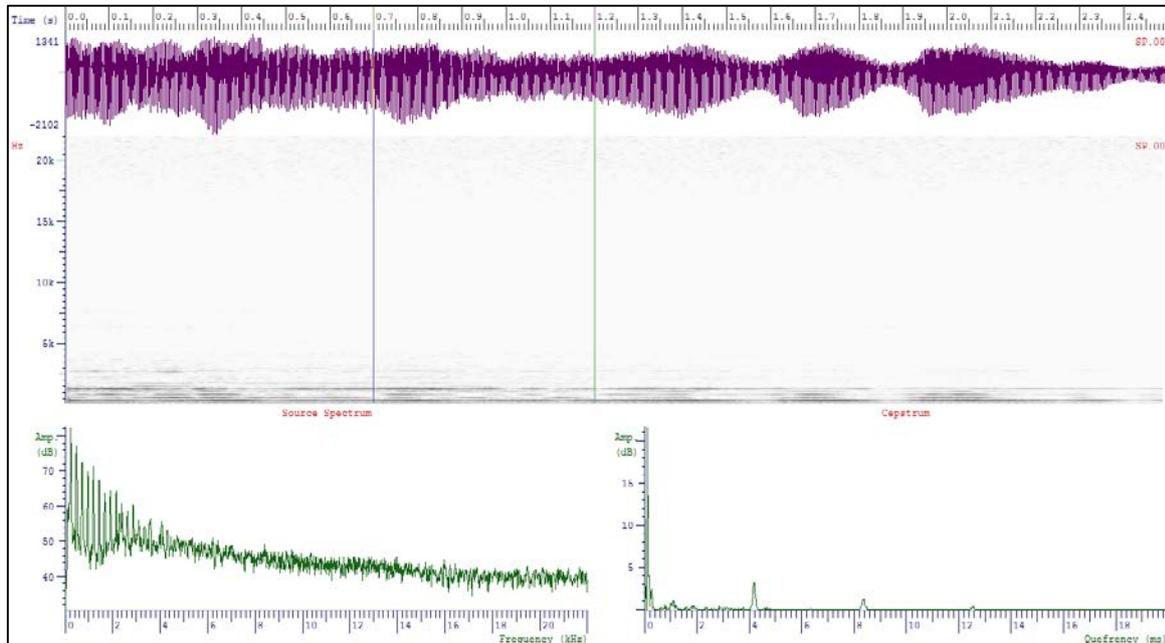


Fig. A.105.- Vocal “a” antes de aplicar el método para el Locutor 18

➤ (Vocal “a” después):

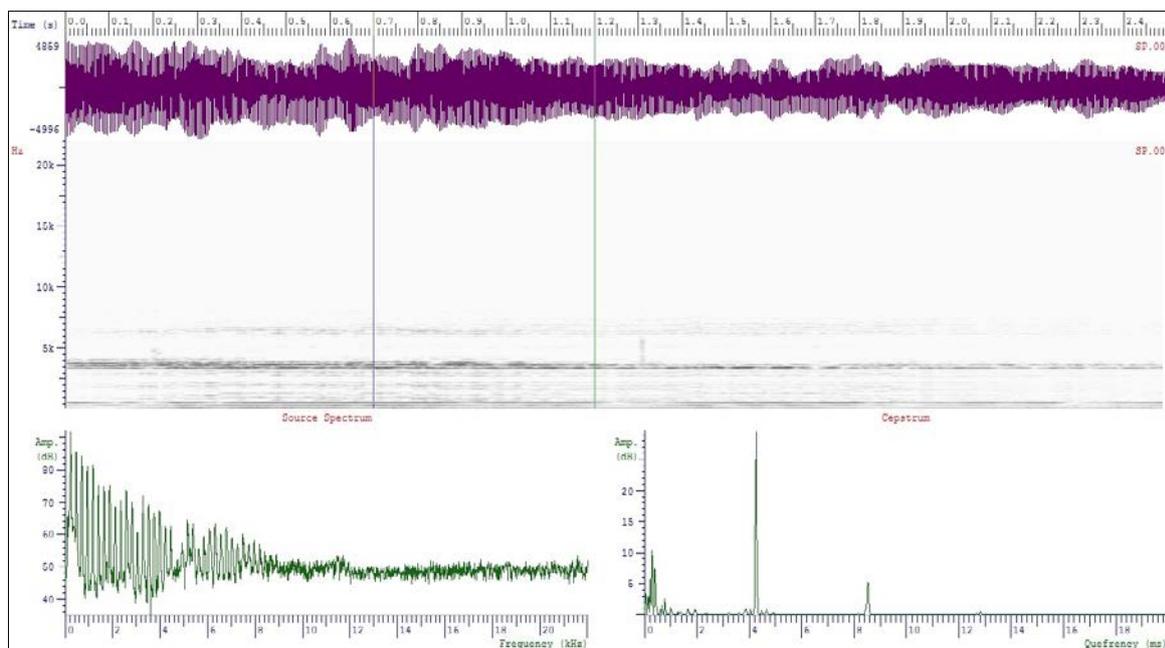


Fig. A.106.- Vocal “a” después de aplicar el método para el Locutor 18

➤ (Vocal “i” antes):

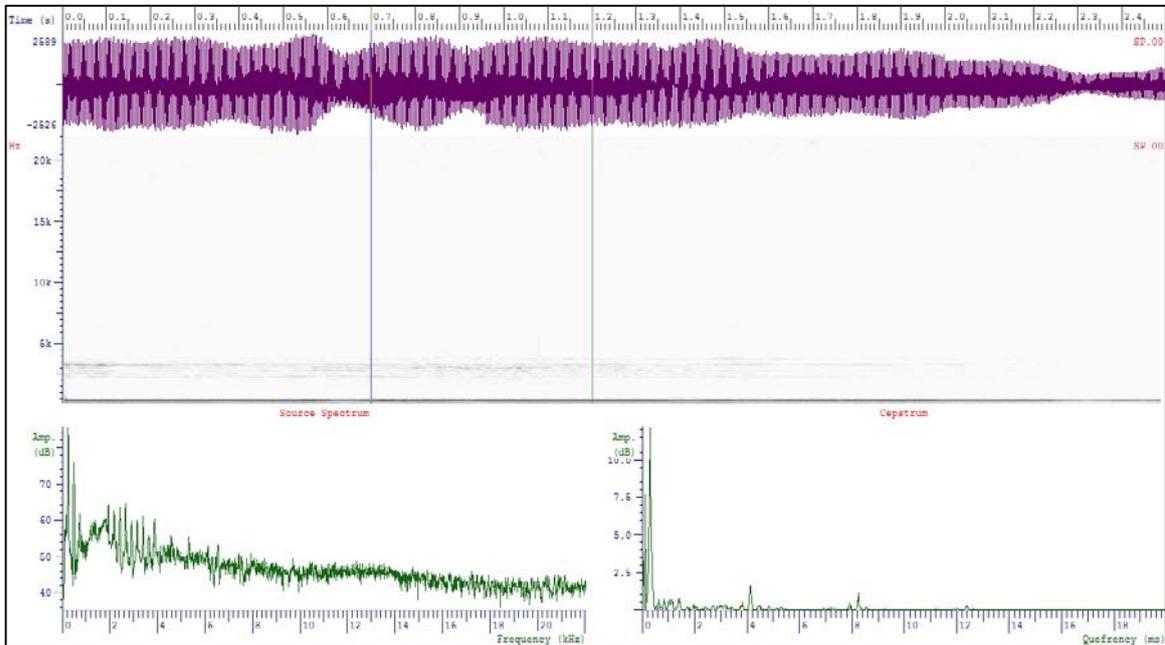


Fig. A.107.- Vocal “i” antes de aplicar el método para el Locutor 18

➤ (Vocal “i” después):

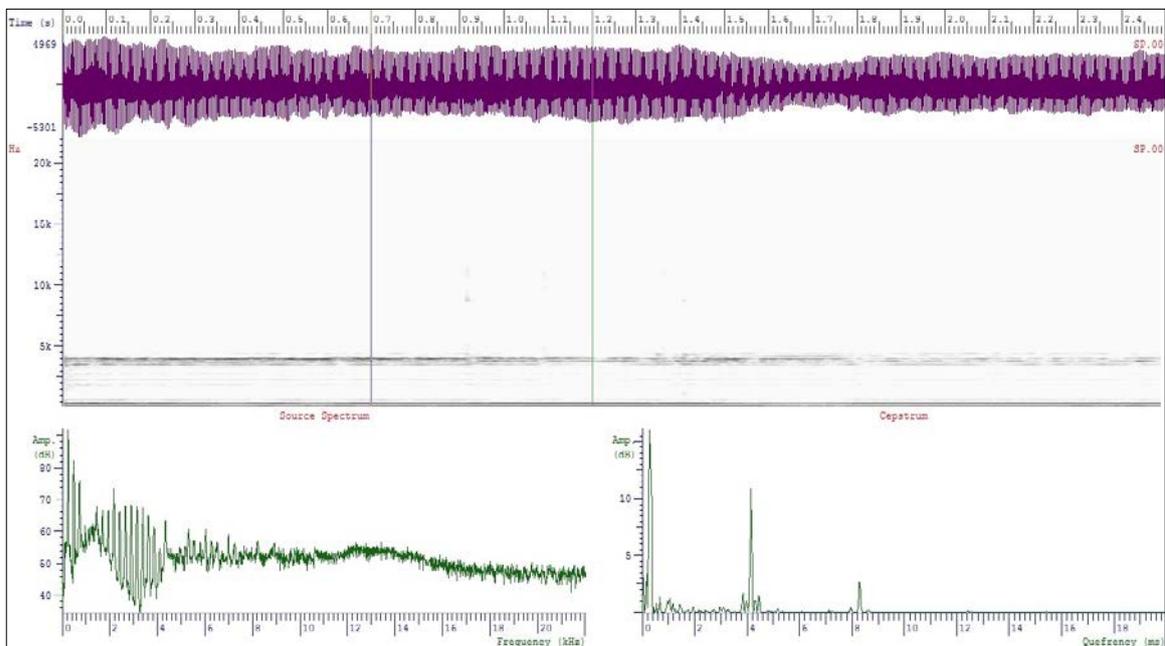


Fig. A.108.- Vocal “i” después de aplicar el método para el Locutor 18

❖ **Locutor 19:**

➤ *(Vocal “o” antes):*

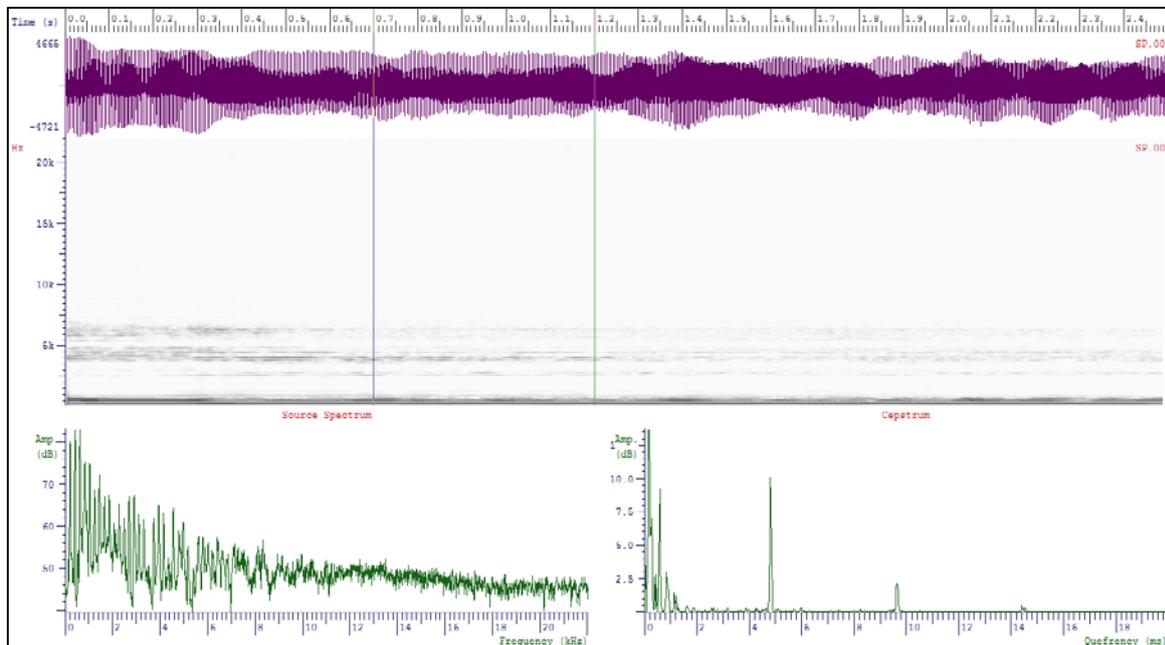


Fig. A.109.- Vocal “o” *antes* de aplicar el método para el Locutor 19

➤ *(Vocal “o” después):*

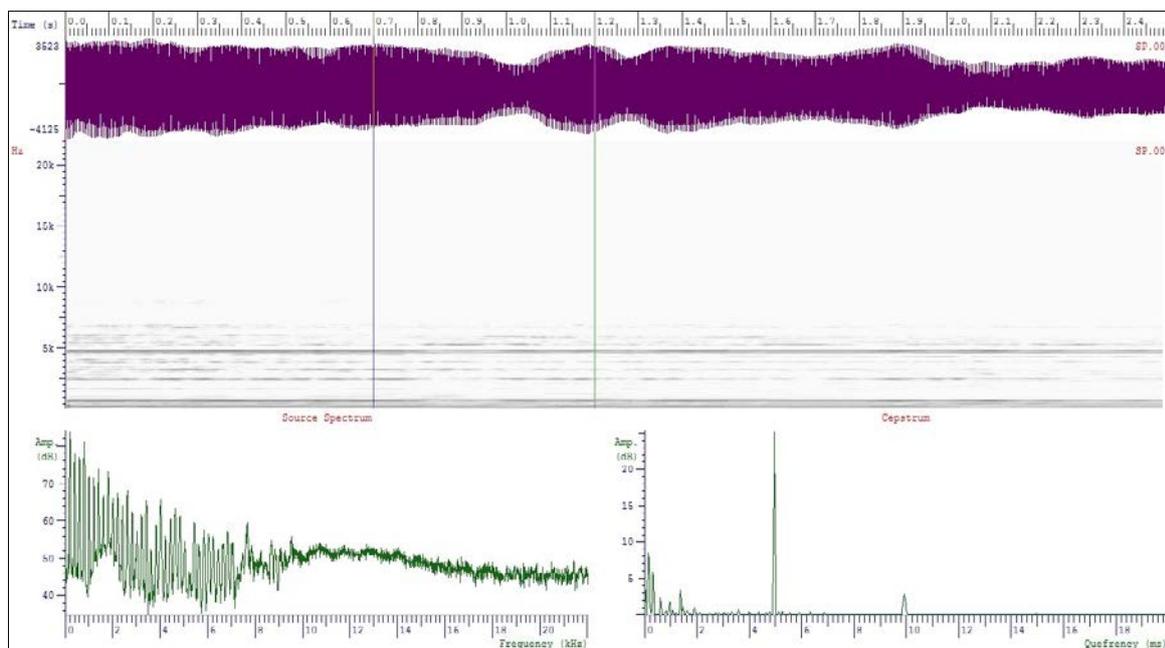


Fig. A.110.- Vocal “o” *después* de aplicar el método para el Locutor 19

➤ (Vocal “a” antes):

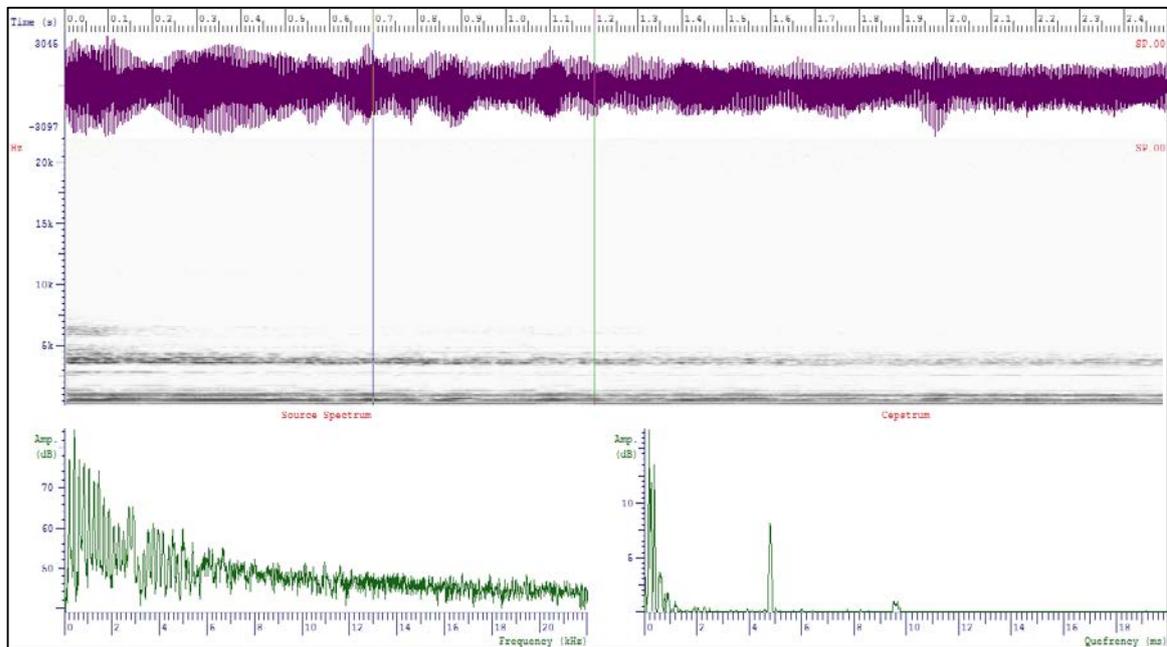


Fig. A.111.- Vocal “a” antes de aplicar el método para el Locutor 19

➤ (Vocal “a” después):

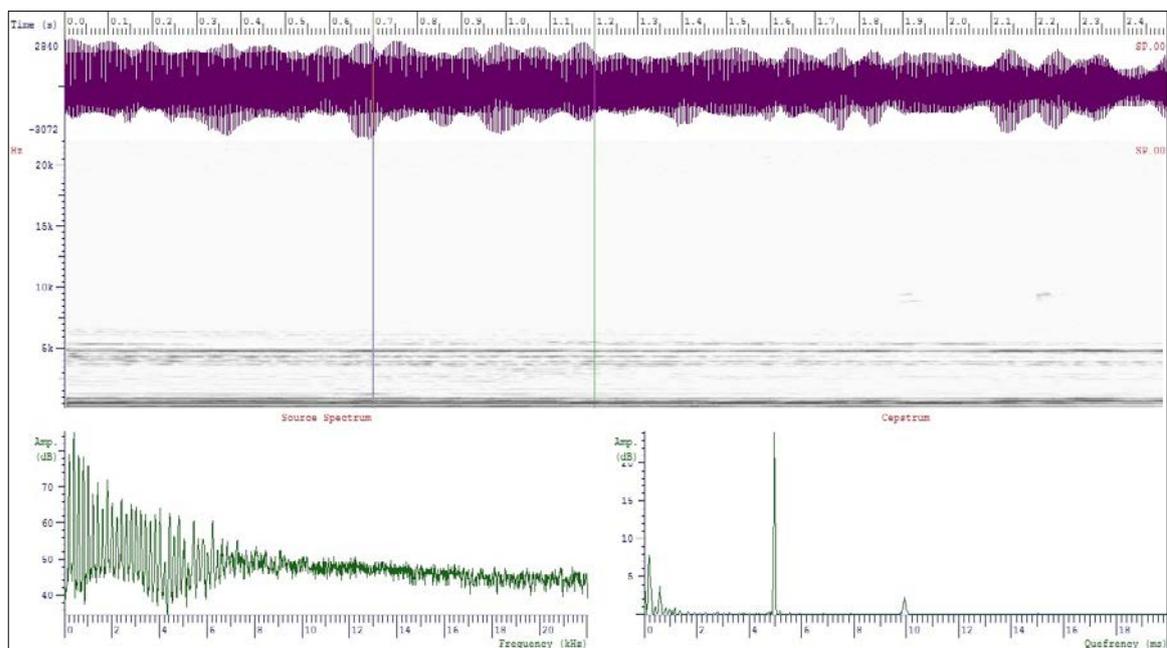


Fig. A.112.- Vocal “a” después de aplicar el método para el Locutor 19

➤ (Vocal “i” antes):

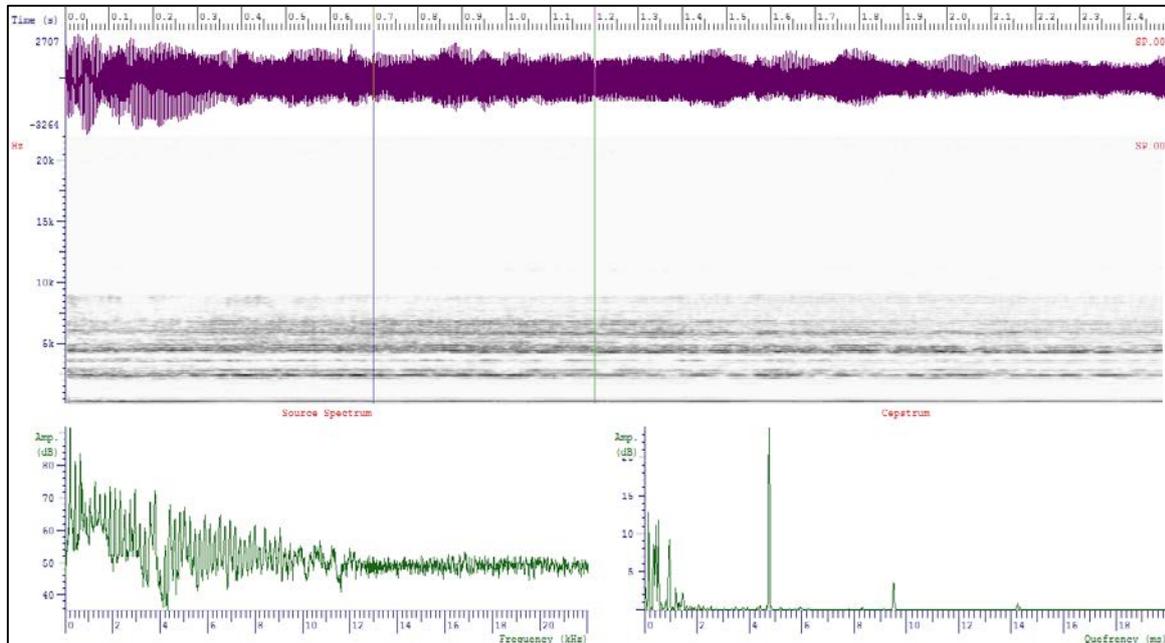


Fig. A.113.- Vocal “i” antes de aplicar el método para el Locutor 19

➤ (Vocal “i” después):

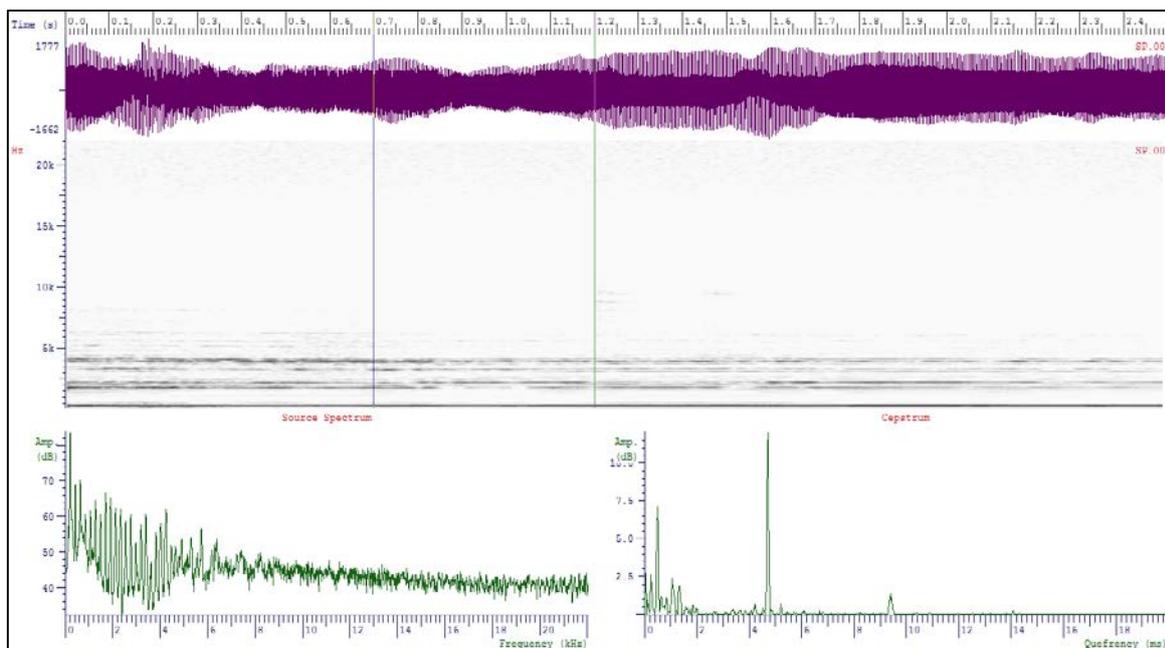


Fig. A.114.- Vocal “i” después de aplicar el método para el Locutor 19

❖ **Locutor 20:**

➤ *(Vocal “o” antes):*

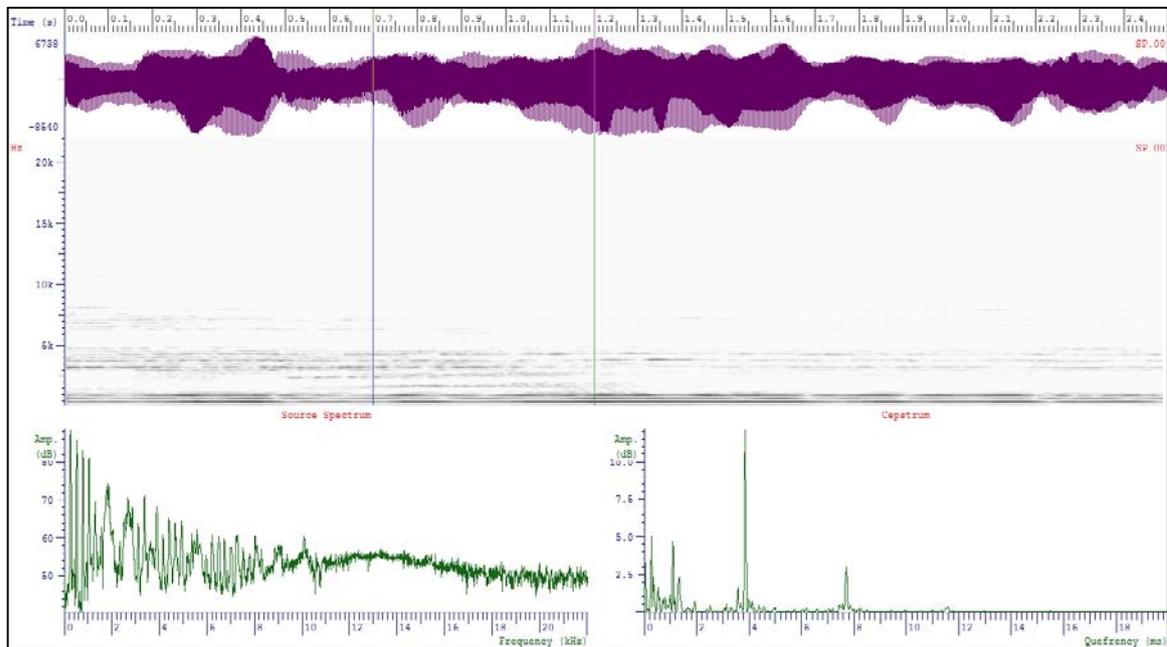


Fig. A.115.- Vocal “o” *antes* de aplicar el método para el Locutor 20

➤ *(Vocal “o” después):*

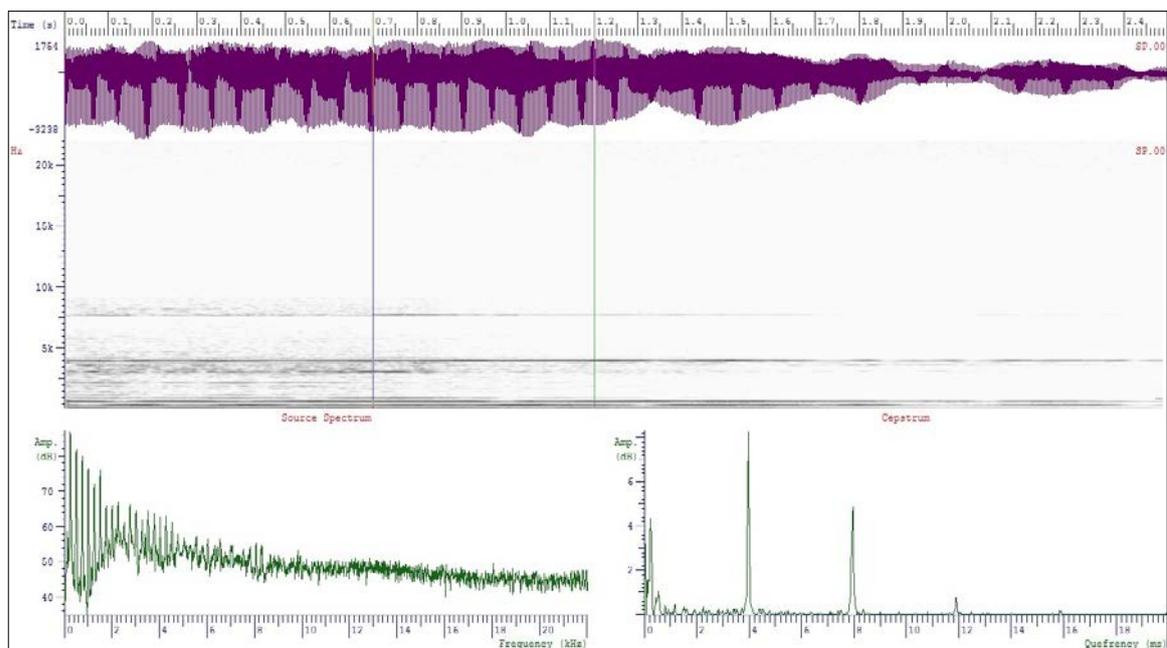


Fig. A.116.- Vocal “o” *después* de aplicar el método para el Locutor 20

➤ (Vocal “a” antes):

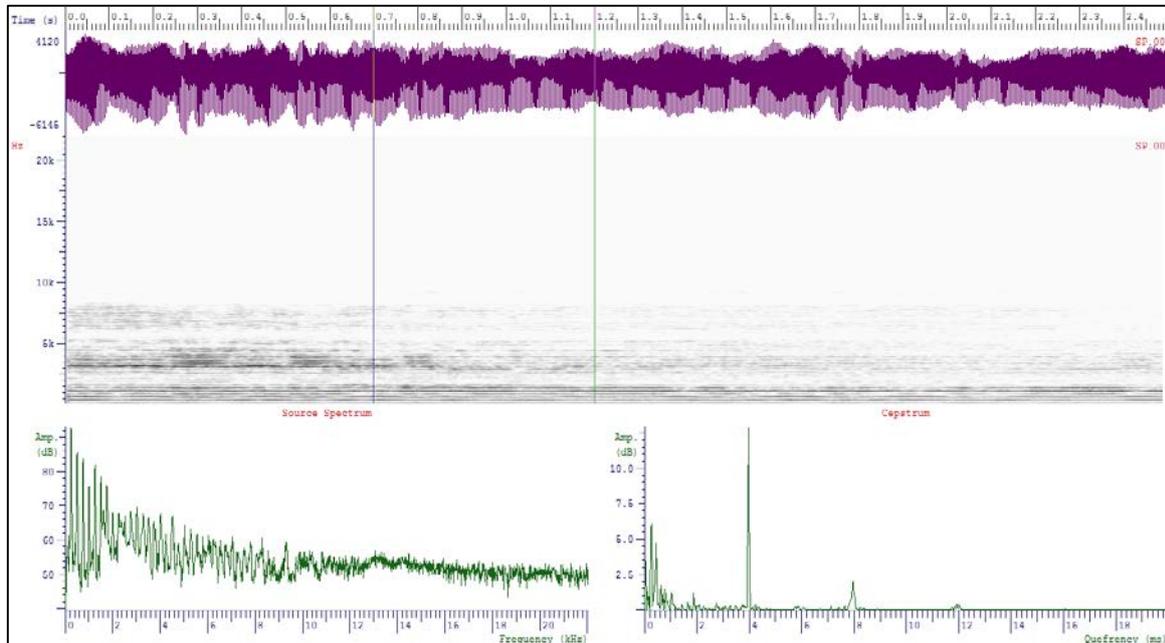


Fig. A.117.- Vocal “a” antes de aplicar el método para el Locutor 20

➤ (Vocal “a” después):

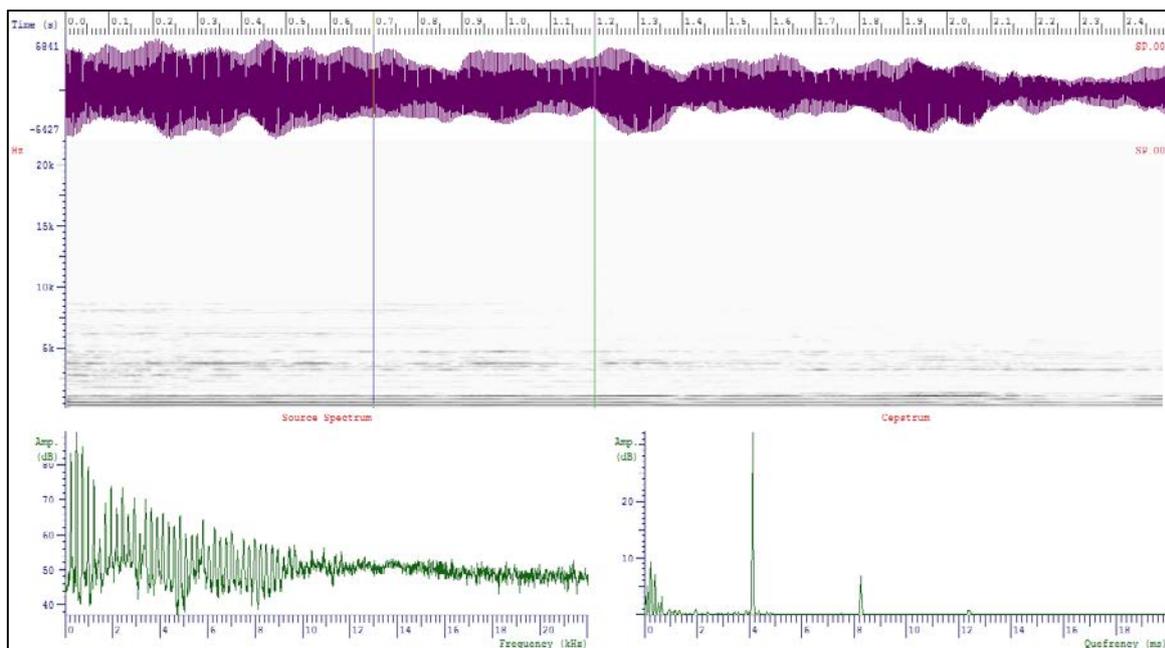


Fig. A.118.- Vocal “a” después de aplicar el método para el Locutor 20

➤ (Vocal “i” antes):

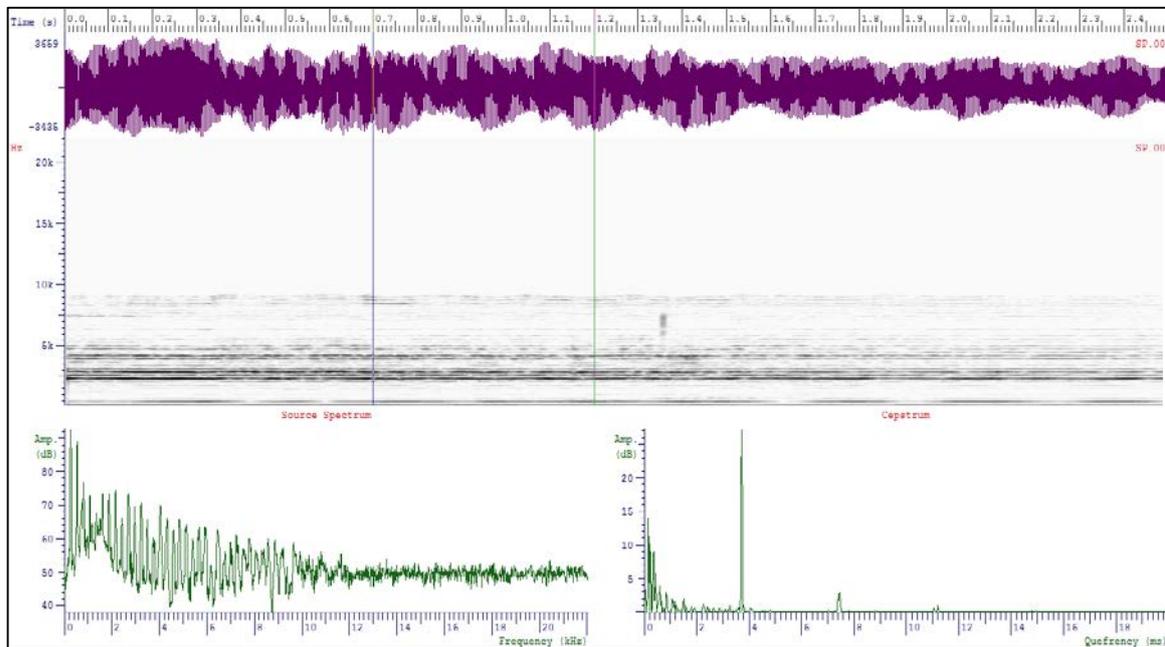


Fig. A.119.- Vocal “i” antes de aplicar el método para el Locutor 20

➤ (Vocal “i” después):

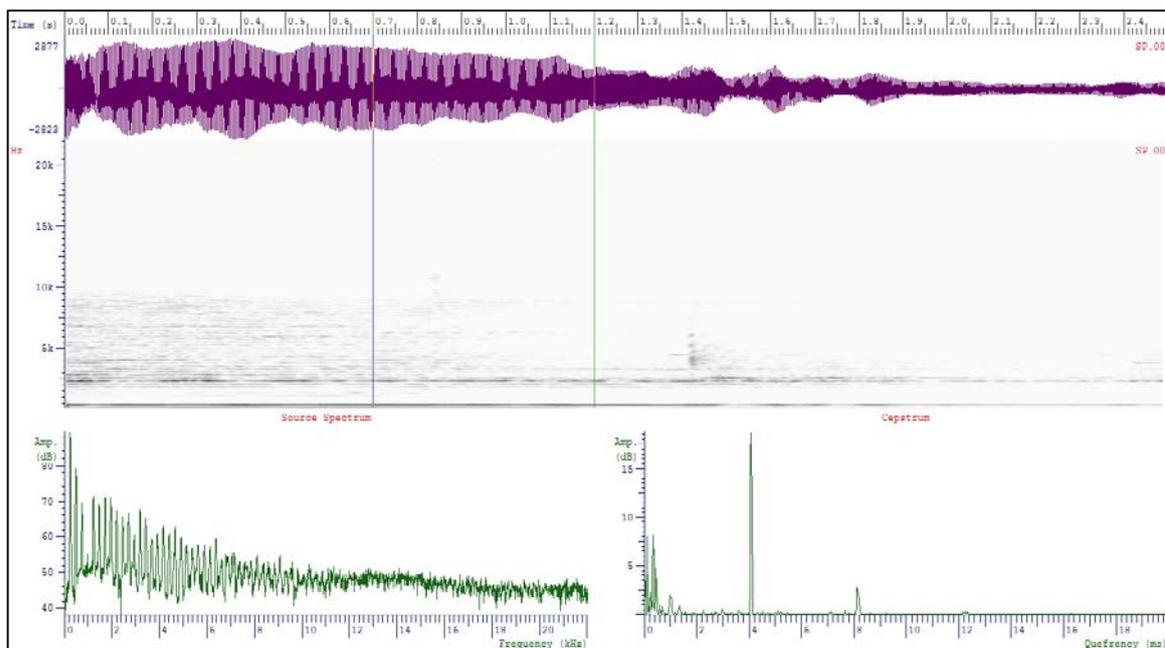


Fig. A.120.- Vocal “i” después de aplicar el método para el Locutor 20

GLOSARIO:

Términos y Definiciones



GLOSARIO

Términos y Definiciones

En el siguiente glosario se incluyen aquellos términos menos conocidos o de difícil interpretación. Cada uno de estos términos viene acompañado de su respectiva definición.

Afonía: Es el término médico usado para referirse a la incapacidad de hablar. Se le considera más grave que la disfonía.

Afasia: Es el trastorno del lenguaje que se produce como consecuencia de una lesión o daño cerebral.

Aritenoides: Se trata de dos cartílagos localizados en la parte posterior, sobre los cuernos del cricoides, sujetan por su apófisis vocal ambos extremos posteriores de las cuerdas y sus apófisis musculares se sostienen sobre el cricoides. Los aritenoides cuentan con gran movilidad, gracias a ella las cuerdas vocales pueden realizar cualquier movimiento.

Armónico: En acústica y telecomunicaciones, un armónico de una onda es un componente sinusoidal de una señal. Su frecuencia es un múltiplo de la fundamental. La amplitud de los armónicos más altos es mucho menor que la amplitud de la onda fundamental y tiende a cero; por este motivo los armónicos por encima del quinto o sexto generalmente son inaudibles.

Astenia: Es un síntoma presente en varios trastornos, caracterizado por una sensación generalizada de cansancio, fatiga, debilidad física y psíquica; con principal incidencia entre las personas de 20 a 50 años, y mayor preponderancia en las mujeres que en los hombres.

Cepstrum: El cepstrum de una señal es el resultado de calcular la transformada de Fourier (FT, del inglés Fourier Transform) del espectro de la señal estudiada en escala logarítmica (dB). El nombre *cepstrum* deriva de invertir las cuatro primeras letras de *spectrum*. El cepstrum es complejo y, por tanto, tiene su parte real y su parte imaginaria.

Cordófono/-a: Instrumento que produce el sonido al golpear, vibrar o percutir las cuerdas

Cricotiroideo: Es un músculo de la laringe; par, triangular, de vértice superior. Se inserta, por abajo, en la cara anterior del cartílago cricoides; por arriba, en el borde inferior y caras anterior y posterior del cartílago tiroides. Lo inerva el nervio laríngeo superior. Su función es bascular el tiroides hacia delante y tensor de las cuerdas vocales.

Disartria: Corresponde a una alteración en la articulación de las palabras. Se atribuye a una lesión del sistema nervioso central y periférico. Se distingue de una afasia motora en que no es un trastorno del lenguaje, sino del habla; es decir, el paciente manifiesta dificultades asociadas con la articulación de fonemas.

Disfemia: Es un trastorno de la comunicación (no un trastorno del lenguaje) que se caracteriza por interrupciones involuntarias del habla que se acompañan de tensión muscular en cara y cuello, miedo y estrés.

Disfonía: Es el nombre que recibe todo trastorno de la voz cuando se altera la calidad de ésta en cualquier grado exceptuando el total, en cuyo caso se denomina afonía.

Etiopatogénico: Que causa o contribuye a generar una determinada patología.

Faringe: Es una estructura en forma de tubo que ayuda a respirar y está situada en el cuello y revestido de membrana mucosa; conecta la nariz y la boca con la laringe y el esófago respectivamente, y por ella pasan tanto el aire como los alimentos, por lo que forma parte del aparato digestivo así como del respiratorio. En el ser humano mide unos trece centímetros, extendida desde la base externa del cráneo hasta la 6ª o 7ª vértebra cervical, ubicada delante de la columna vertebral.

Glottis: Es la 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. En el ser humano, la laringe tiene diversas funciones interviniendo en la fonación, la deglución y como mecanismo de defensa.

GRABS: La escala GRABS (Hirano, 1981), diseñada para la valoración del grado de patología de la voz, y que en el estudio fue cumplimentada por expertos en análisis acústico.

Hipofaringe: La laringofaringe o hipofaringe es la porción más baja de la faringe y la región anatómica que comunica la garganta con el esófago. De modo que en la laringofaringe desembocan dos tubos anatómicos, la laringe por delante y el esófago por detrás.

Laringe: Es un órgano constituido por varios cartílagos en la mayoría semilunares. Además, comunica a la faringe con la tráquea y se halla delante de aquella.

Nódulo: En medicina es una pequeña agrupación de células. Puede ser tanto una lesión, como una estructura funcional fisiológica.

Onomatopéyico: De la onomatopeya o relativo a ella. La onomatopeya es la imitación lingüística o representación de un sonido natural o de otro fenómeno acústico no discursivo. Según la Real Academia Española, es la imitación o recreación del sonido de algo en el vocablo que se forma para significarlo o vocablo que imita o recrea el sonido de la cosa o la acción nombrada.

Pitch: Es la frecuencia fundamental de espectro de frecuencias.

Presbiacusia: se define como la pérdida progresiva de la capacidad para oír altas frecuencias (empezando por la frecuencias del habla; que oscilan entre los 500 y los 4000 Hertzios), debido al deterioro producido en el sistema auditivo generado por la edad, principalmente a nivel del oído interno y de nervio auditivo.

Quefreny: Es una variable independiente de un gráfico de cepstrum. La quefreny es una medida del tiempo, pero no en el sentido del dominio temporal.

Sobretono: Un sobretono es un componente sinusoidal de la forma de una onda, de mayor frecuencia que su frecuencia fundamental. Generalmente el primer sobretono es el segundo armónico, el segundo sobretono el tercer armónico, etcétera.

Subarmónico: Una frecuencia submúltiplo de una dada. Suele aparecer como producto de distorsión sustractivo cuando un sistema no lineal se excita con dos tonos cuyas frecuencias están en proporción $n:m$, con n y m enteros (por ejemplo dos tonos de 500 Hz y 600 Hz darán como resultado un subarmónico de 100 Hz, que es el quinto subarmónico de 500 Hz y el sexto subarmónico de 600 Hz).

Subglótica: Situada por debajo de la glotis.

Supraglótica: Situada por encima de la glotis.

Tiroaritenoides: El músculo tiroaritenoides, también conocido como el músculo vocal, es un amplio, músculo delgado que se encuentra en paralelo con y lateral a la de las cuerdas vocales y que soporta la pared del ventrículo y su apéndice. Funciona bien en el control tonal de las cuerdas vocales.

Vibrato: es un término musical que describe la variación periódica de la altura o frecuencia de un sonido. Se trata de un efecto musical que se utiliza para añadir expresión a la música vocal e instrumental. Suele catalogarse en función de dos factores: la cantidad de variación en la altura (extensión de vibrato) y la velocidad con la cual varía la altura (velocidad de vibrato).