

SAV versión 1.1

Pablo Daniel Agüero

Resumen

SAV (Sistema de Análisis de la Voz) es un software de análisis de audio que calcula una gran cantidad de parámetros de la voz, los cuales permitirán hacer un seguimiento de los pacientes durante su tratamiento con medidas objetivas.

Las principales motivaciones en el desarrollo de SAV son:

- Desarrollar un software gratuito que pueda ser utilizado por los especialistas de la voz.
- Proporcionar mecanismos de comunicación con los usuarios para incorporar en el software distintas funcionalidades sugeridas y corregir posibles deficiencias.
- Ofrecer estudios de casos reales y análisis del rendimiento de los distintos parámetros usando voz sintética.

SAV es un software en sus comienzos y en continuo desarrollo. Por ello estamos abiertos a críticas y sugerencias. **Nuestro software debe ser evaluado, calibrado y probado antes de poder ser utilizado en pacientes para el seguimiento de tratamientos. En consecuencia, hoy en día solo puede ser utilizado como software de prueba.**

Índice general

1. Introducción	2
1.1. Descarga	2
1.2. Instalación	3
2. Funcionamiento	4
2.1. Solapa Seleccionar Archivo	5
2.2. Solapa Análisis Gráfico	6
2.3. Solapa Análisis Estadístico	8
2.4. Solapa de Gráfico Radial	9
2.5. Análisis acústico de la voz con parámetros numéricos	9
2.6. Solapa de Grabación	11
2.7. Solapa de Cooperación	12
3. Versiones	13
Bibliografía	14

Capítulo 1

Introducción

SAV es un software de análisis de la voz programado en lenguaje JAVA[®]. Los motivos para el uso de este lenguaje de programación fueron la posibilidad de generar un software que pueda ser ejecutado en múltiples sistemas operativos, tales como Windows, Linux y Macintosh. De esta manera los usuarios no deberán utilizar complejos mecanismos de instalación para usarlo.

Otra de las razones por las cuales hemos utilizado JAVA es la posibilidad de contar con herramientas de desarrollo gratuitas, como es el caso de Netbeans, una interfaz visual de programación de aplicaciones en JAVA muy versátil.

Finalmente, la principal motivación ha sido la utilización de JAVA para ofrecer un software gratuito y de este modo llegar a la mayor cantidad de usuarios con independencia de sus posibilidades económicas.

A pesar de ser un software gratuito, SAV poseerá siempre un canal abierto para críticas y sugerencias de los usuarios, y también incorporará los avances desarrollados tanto en nuestro laboratorio como en publicaciones relacionadas con el tema.

1.1. Descarga

Para descargar SAV de nuestra página web deberán completar un formulario donde se solicita apellido, nombre, ciudad, país, ocupación y correo electrónico.



Descargar SAV para utilizarlo en su ordenador

Ingresando su nombre, apellido y correo electrónico, recibirá por e-mail un enlace para descargar SAV.

Esta información nos permitirá mantener el contacto con los usuarios de SAV ofreciéndoles información útil sobre nuevas versiones, estudios de casos reales y documentación para facilitar su uso.

Apellido y nombre:

Ciudad:

País:

Profesión:

Correo electrónico:

Una vez ingresados sus datos y luego de pulsar el botón de descarga, la siguiente página aparecerá en el navegador de su ordenador.



El enlace para descargar el software será enviado por correo electrónico luego de hacer una verificación manual previa por parte del grupo de los datos ingresados por el usuario. Si todos los datos son válidos el usuario recibirá a la brevedad un correo electrónico con la información necesaria para la descarga.

1.2. Instalación

La instalación de SAV es sencilla. Para hacerla se deberá ejecutar el archivo `setup.jar` en una carpeta de su elección (por ejemplo: el escritorio). Una vez realizada esta acción el usuario debe poder ver el siguiente contenido del archivo en la carpeta elegida.



Para comenzar a utilizar SAV debe ejecutar el archivo `SAV.jar`, el cual es un ejecutable en JAVA. El archivo `test.wav` que se ve en la pantalla es un audio de ejemplo de la vocal *a* sostenida.

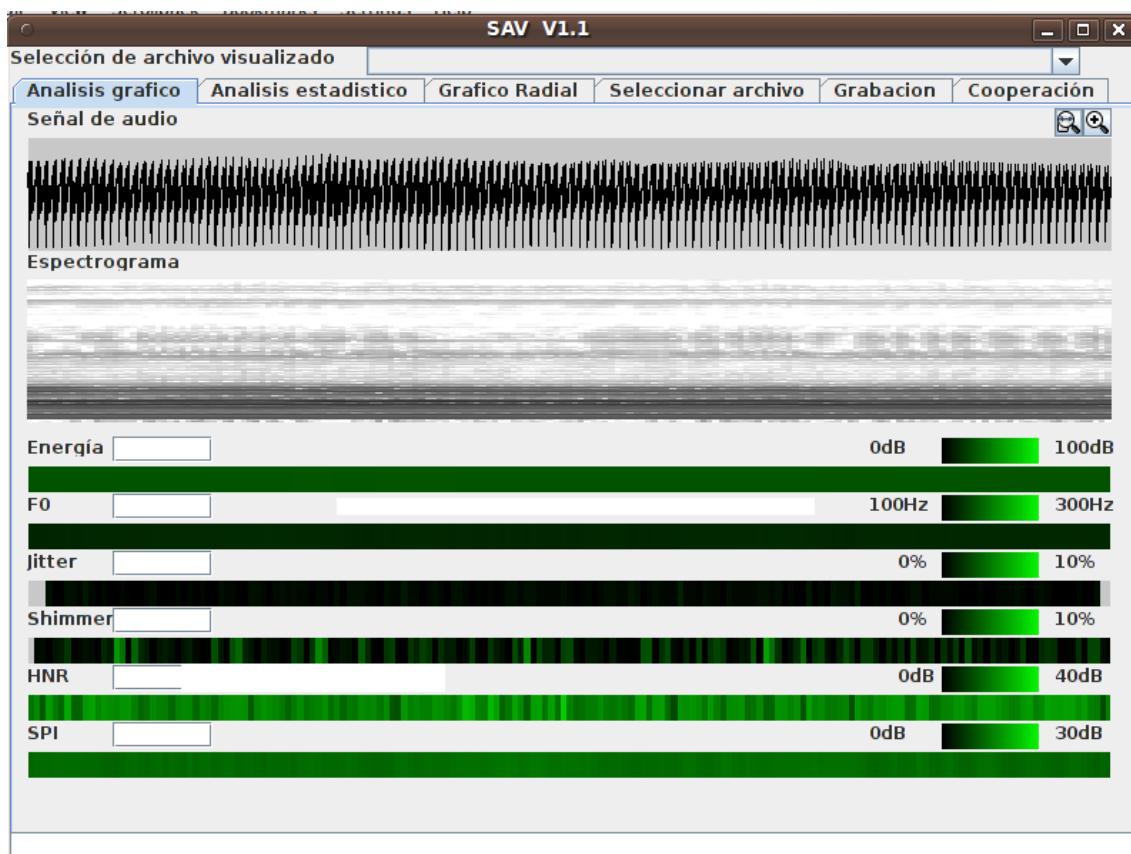
En la carpeta *lib* se encuentran un conjunto de librerías necesarias para el funcionamiento de SAV, también en lenguaje JAVA.

Capítulo 2

Funcionamiento

El funcionamiento de SAV es sencillo. En la pantalla se encuentra toda la información disponible para el usuario de manera simultánea, con el objetivo de evitar la utilización de un menú de opciones que impida observar toda la capacidad del programa con un simple vistazo.

En la pantalla inicial de SAV se puede observar la presencia de cuatro solapas: *análisis gráfico*, *análisis estadístico*, *gráfico radial* y *seleccionar archivo*.

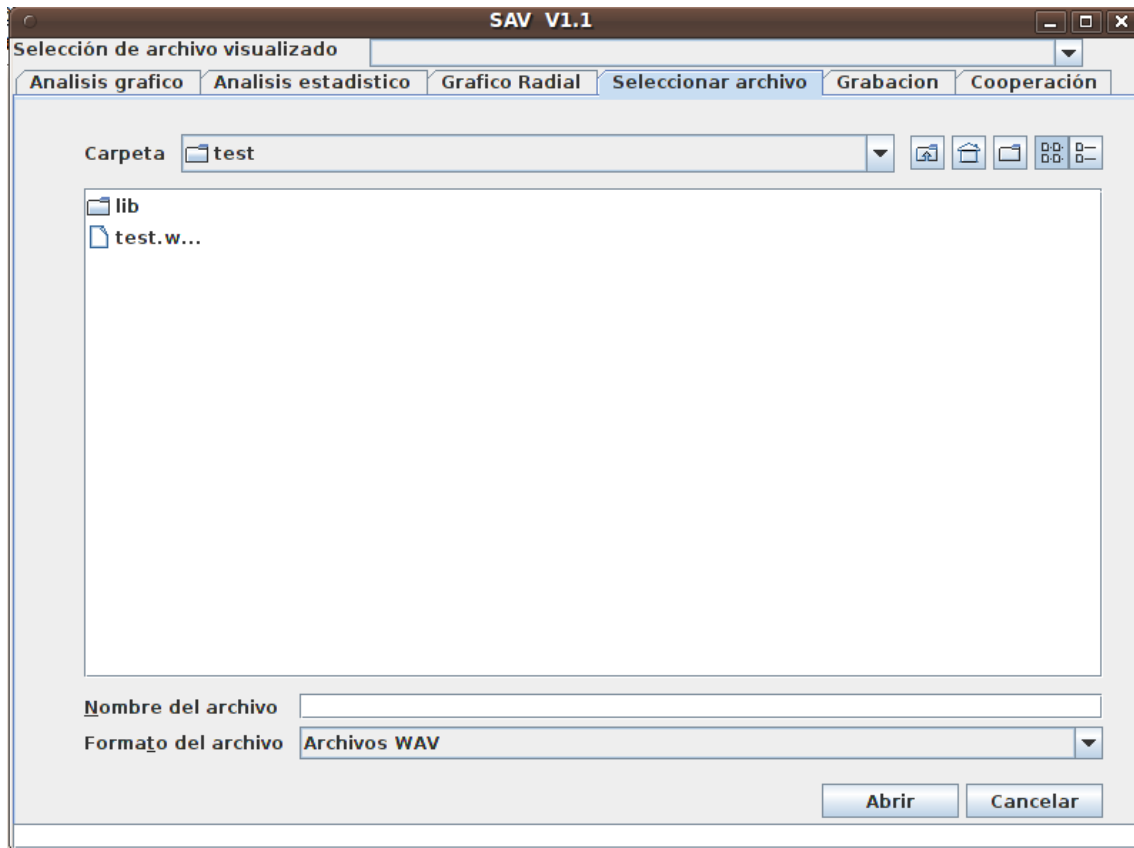


2.1. Solapa Seleccionar Archivo

Si seleccionamos la solapa *seleccionar archivo* podremos elegir el archivo que se desea analizar con SAV. En este ejemplo seleccionaremos el archivo *test.wav*, el cual está disponible cuando se descarga SAV. Una vez realizada esta operación, SAV procederá a hacer el análisis del archivo de audio y pasará automáticamente a visualizarlo en la solapa de *análisis gráfico*.

El formato del archivo de sonido debe ser WAV, grabado con una frecuencia de muestreo de 44100Hz y 16 bits. En caso de no respetar estos requisitos, el software le indicará que el formato del archivo seleccionado es incorrecto.

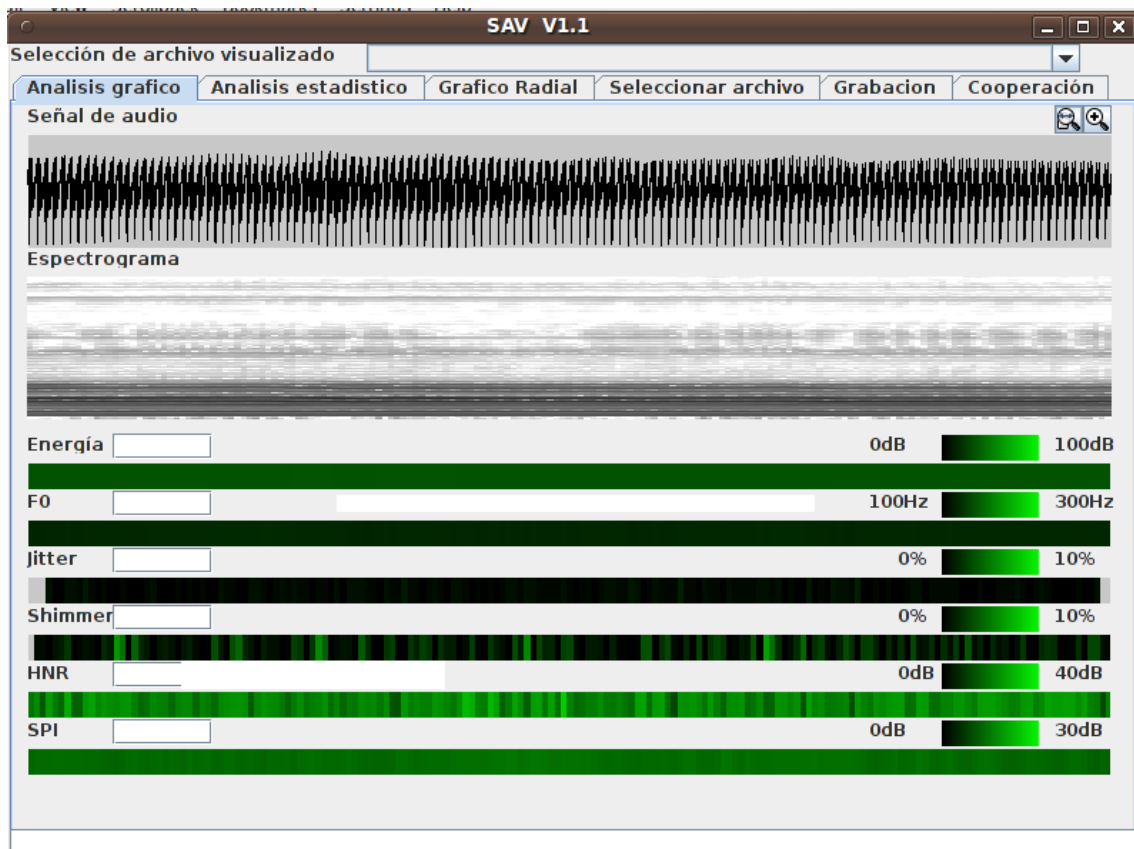
Esta nueva versión de SAV permite trabajar con múltiples archivos cargados mediante el listado que aparece en la parte superior de la ventana.



2.2. Solapa Análisis Gráfico

La pantalla de *análisis gráfico* presenta en orden descendente:

- **Señal de audio.** Es la representación visual en un gráfico XY de la señal de audio, donde el eje X corresponde al tiempo y el eje Y es la amplitud de la señal.
- **Espectrograma.** Es la representación visual de la evolución espectral de la señal de audio, donde el eje X corresponde al tiempo y el eje Y es la frecuencia. La mayor intensidad de potencia en un determinado rango de frecuencia se representa con tonos oscuros.
- **Energía.** La energía es un parámetro que permite estimar la intensidad de la fonación. Su valor está calculado en decibeles (dB), ya que esta escala permite representa un gran rango de valores con números pequeños.
- **Frecuencia fundamental.** Este parámetro estima la frecuencia instantánea de vibración de la glottis en los fonemas sonoros. Su valor está directamente vinculado con el **jitter** y el **shimmer**.
- **Jitter.** Este parámetro estima la variación del período de vibración de la glottis entre dos períodos consecutivos: $|\frac{2(T0_n - T0_{n-1})}{T0_n + T0_{n-1}}|$.
- **Shimmer.** Este parámetro estima la variación de la máxima amplitud (A) de la fonación entre dos períodos consecutivos de vibración de la glottis: $|\frac{2(A_n - A_{n-1})}{A_n + A_{n-1}}|$.
- **HNR.** La relación de armónicos a ruido (o por sus siglas en inglés HNR: Harmonics-to-Noise Ratio) estima el grado de periodicidad y armonicidad de la fonación de un fonema sonoro para un período de vibración de la glottis [2].
- **SPI.** El índice de fonación débil (o por sus siglas en inglés SPI: Soft Phonation Index) es un indicador de aducción de las cuerdas vocales y cierre glotal durante la fonación.

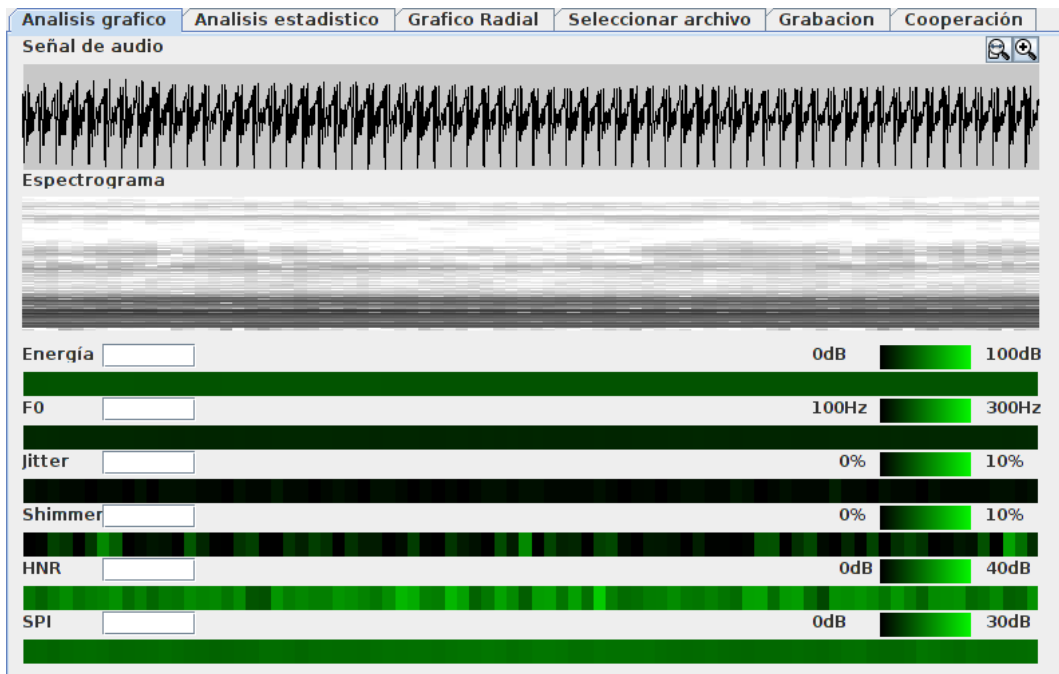


La información está representada usando una escala de intensidad de colores, donde el color verde más intenso indica un valor mayor del parámetro. Esta representación visual permite observar a intervalos regulares de 5,8ms el valor de los distintos parámetros, con lo cual se puede analizar distintos segmentos de la fonación.

Es posible ampliar una región del audio seleccionando una porción del mismo pulsando el ratón en la señal de audio para indicar el comienzo de la selección y arrastrando el ratón hasta llegar al final de la misma. Luego, liberando el botón del ratón, la porción seleccionada se resaltará con color amarillo.



La porción de audio seleccionada se puede ampliar pulsando el botón con el símbolo + en la parte superior derecha de la pantalla. Es posible visualizar nuevamente todo el audio pulsando el botón a la izquierda del botón con el símbolo +.



2.3. Solapa Análisis Estadístico

La porción de audio visualizado en la pantalla de análisis gráfico puede ser analizada estadísticamente, para lograr un conjunto de valores compacto que permita comparar distintos audios. Esa es la función de la solapa de análisis estadístico.

Parámetro	Nombre	Valor	Unidad	Norma	Desv. e...	Umbral
Frecuencia fundamental media	mF0	130.0	Hz			250
Período medio	mT0	7.7	ms			5.6
Frecuencia fundamental máxima	maxF0	132.0	Hz			250
Frecuencia fundamental mínima	minF0	128.2	Hz			100
Desviación estándar de la frecuencia fundamental	desvF0	0.8	Hz			10
Rango de frecuencia fundamental	rangoF0	0.5	semitonos			1.5
Duración del segmento analizado	dur	0.5	s			10
Jitter relativo	jitr	0.4	%			1.04
Jitter absoluto	jitta	33.0	us			83.2
Perturbación relativa promedio del jitter	jitrtrap	0.2	%			0.68
Perturbación relativa promedio de cinco períodos del jitter	jitrppq5	0.2	%			0.84
Shimmer relativo	shimr	1.8	%			3.81
Shimmer absoluto	shima	0.2	dB			0.35
Perturbación relativa promedio del shimmer	shimrap	0.7	%			2.0
Perturbación relativa promedio de cinco períodos del shimmer	shimppq5	1.0	%			2.0
Relación de armónicos a ruido	HNR	20.4	dB			15
Índice de fonación débil	SPI	12.7	dB			10

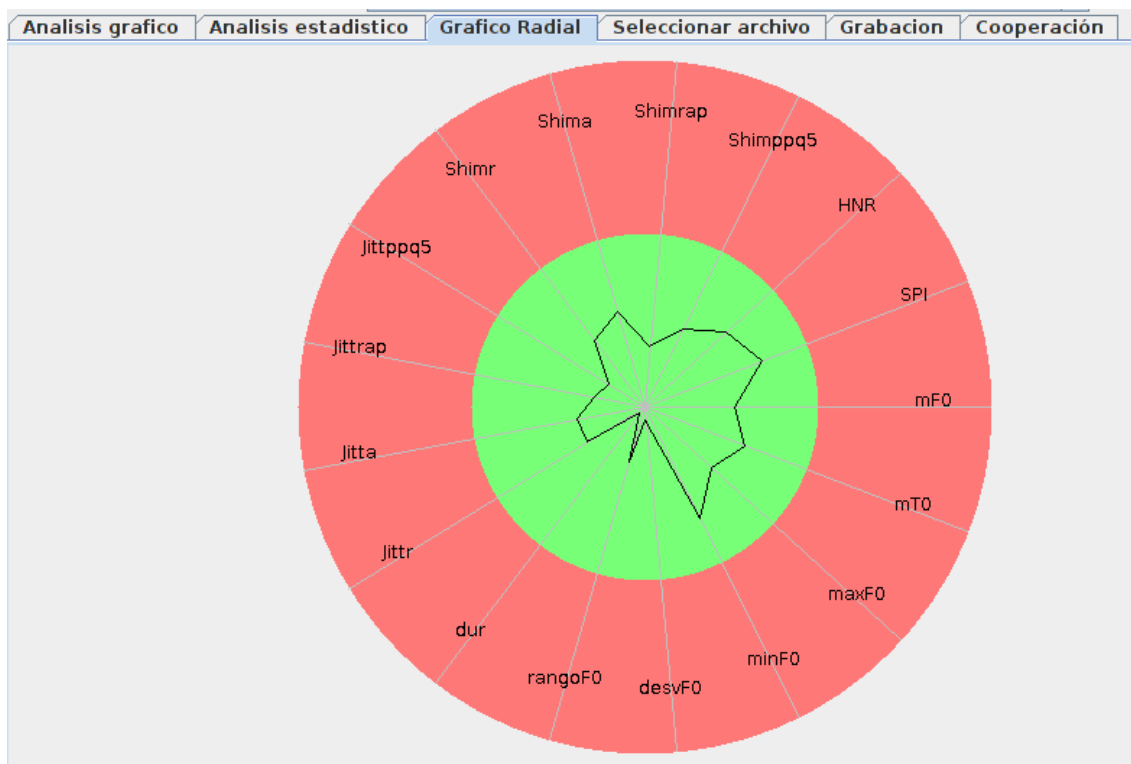
En esta pantalla se presenta un gran número de parámetros, tales como:

- **Frecuencia fundamental media.** El valor promedio de la frecuencia fundamental.
- **Período medio.** El valor promedio del período de vibración de la glotis.
- **Frecuencia fundamental máxima.** El valor máximo de la frecuencia fundamental.
- **Frecuencia fundamental mínima.** El valor mínimo de la frecuencia fundamental.
- **Desviación estándar de la frecuencia fundamental.** El valor de dispersión de la frecuencia fundamental.
- **Rango de frecuencia fundamental.** La amplitud de variación de la frecuencia fundamental en semitonos.
- **Duración del segmento analizado.** La duración de la porción de voz analizada en segundos.
- **Jitter y sus diferentes variantes.** Jitter relativo (jitr), absoluto (jitta) y perturbación relativa promedio (jitrtrap y Jitrppq5).
- **Shimmer y sus diferentes variantes.** Shimmer relativo (shimr), absoluto (shima) y perturbación relativa promedio (shimrap y shimppq5).
- **Relación de armónicos a ruido.**
- **Índice de fonación débil.**

2.4. Solapa de Gráfico Radial

Con el objeto de una visualización simple del valor de los parámetros indicando si son normales o anormales, se incorporó a SAV un gráfico radial.

En la solapa de Gráfico Radial se pueden observar los distintos parámetros del análisis estadístico, y dependiendo de su valor con respecto al umbral se ubicará en la región verde (normal) o roja (anormal).



2.5. Análisis acústico de la voz con parámetros numéricos

El análisis acústico de la voz ha alcanzado un importante desarrollo en los últimos tiempos gracias, entre otras razones, al progreso y difusión experimentados por los medios informáticos que lo hacen posible. Entre sus ventajas destaca el ser un método no invasivo de evaluación de la voz y el ofrecer la oportunidad de objetivizar la evaluación en unos parámetros numéricos¹.

Uno de los problemas principales en el diagnóstico perceptivo de la voz por el oído del clínico es que el sistema auditivo humano está preparado fundamentalmente para percibir la voz o el habla como un todo integrado, lo cual es altamente beneficioso desde el punto de vista de la comunicación lingüística. Ahora bien, esta capacidad se ve limitada cuando se trata de tomar conciencia de componentes acústicos individualizados que, sin embargo, son relevantes desde una perspectiva clínica.

En muchas ocasiones existe dificultad en determinar por un procedimiento exclusivamente perceptivo el origen de ciertas anomalías o particularidades de la voz. Por ejemplo [1], algunos rasgos del tono o pitch son más el producto de las resonancias del tracto vocal que de la frecuencia de vibración de las cuerdas vocales. La hipernasalidad percibida en una voz puede ser la consecuencia de una desincronización en los tiempos de oclusión velar antes que una oclusión incompleta. Es decir, un mismo atributo o alteración de la calidad vocal puede tener su origen en subsistemas distintos difícilmente aislables por la mera audición.

¹Texto extraído del artículo “Análisis Acústico de la voz: fiabilidad de un conjunto de parámetros multidimensionales” de J. Gonzalez y otros.

En otras ocasiones, una adecuada percepción no puede matizarse con el grado de precisión que ofrece una medida numérica. Así, en una voz percibida como soplada puede establecerse el grado de aspiración, o "breathiness", a través del parámetro pertinente (índice de turbulencia de voz). En este sentido, junto a la evaluación subjetiva por parte del clínico experimentado, el diagnóstico se enriquece y gana precisión cuando se complementa con la medida objetiva de parámetros relevantes de la voz. Las ventajas de ello se traducen en una mayor objetividad en el informe o la comunicación de los datos y en una mayor exactitud en la evaluación del progreso terapéutico, especialmente cuando éste es lento.

Para poder evaluar la voz disfónica a través de parámetros numéricos, es importante definir previamente la voz normal o no disfónica y disponer de valores normativos de comparación. Como algunos autores señalan [3], en nuestro ámbito geográfico son muy escasos los estudios llevados a cabo en este sentido.

Por otra parte, algunos valores pueden ser dependientes de los algoritmos usados por el software específico que los calcula [5, 4], lo que hace más necesario, si cabe, el disponer de normas específicas de los principales programas empleados en la clínica.

Junto a esta necesidad, una propiedad que debe disponer todo instrumento de medida es el de una fiabilidad adecuada que permita confiar en la estabilidad de los valores obtenidos.

En la extracción de parámetros de la voz no podemos esperar una coincidencia absoluta entre dos medidas sucesivas del mismo individuo y, dada la enorme variabilidad de la voz humana, tanto entre individuos como intraindividualmente, es aceptable cierta variación siempre que ésta se mantenga dentro de ciertos límites. No obstante, la robustez y validez clínica de los parámetros descansa necesariamente, como condición previa, en el grado de consistencia de sus valores. Dos medidas repetidas de la voz del mismo individuo en las mismas condiciones de registro deberían ser lo suficientemente semejantes para que podamos confiar en ellas.

2.6. Solapa de Grabación

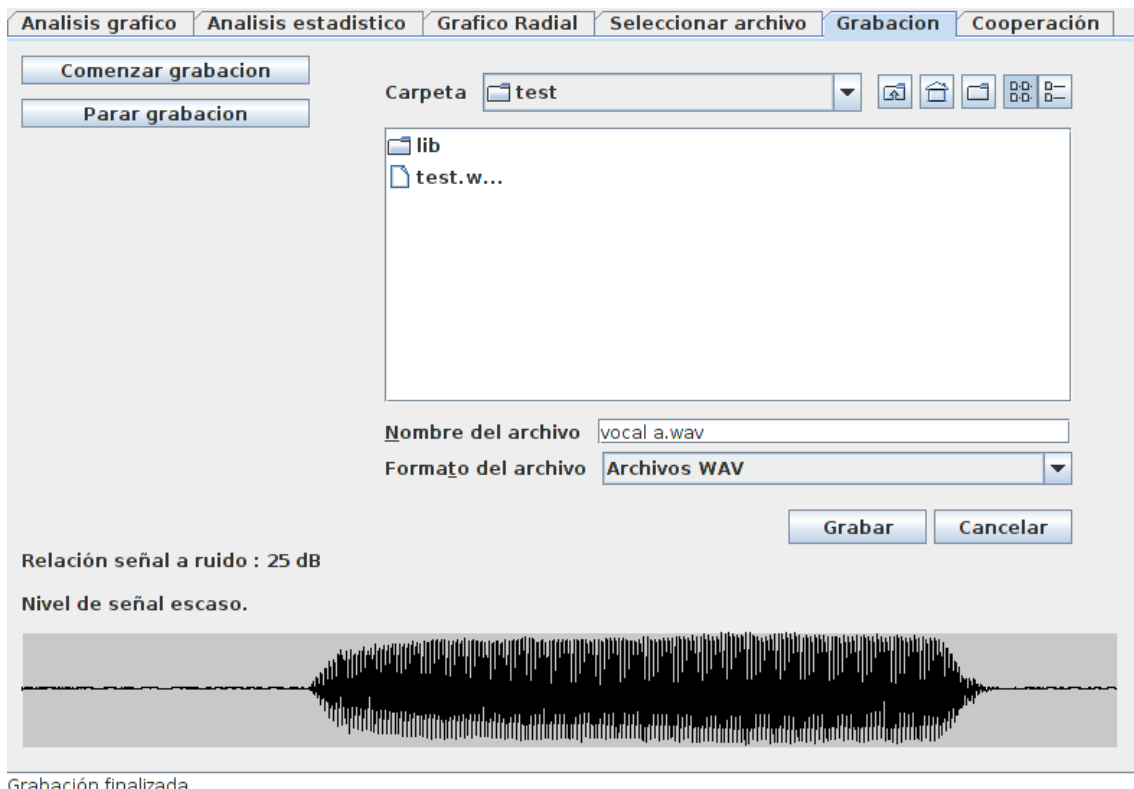
SAV ofrece la posibilidad de grabar audio mediante el uso de la solapa de *grabación*. El formato de los archivos grabados será el utilizado por SAV, con una frecuencia de muestreo de 44.100 y 16 bits por muestra.

Los botones *comenzar grabación* y *parar grabación* se utilizan para definir cuando comenzará y terminará la grabación por parte de SAV. El sistema no detectará automáticamente cuando la persona comienza a hablar, así que la duración del sonido grabado estará dada por el intervalo de tiempo transcurrido entre la presión de los dos botones.

Una vez finalizada la grabación SAV informará el nivel de señal a ruido estimado del audio, y también indicará si el nivel de volumen es adecuado. Existen cuatro mensajes diferentes acerca del nivel de sonido:

- **Nivel de señal excesivo. Grabar de nuevo.** El nivel de sonido es excesivo, y puede provocar una distorsión de la señal.
- **Nivel de señal adecuado.** El nivel de sonido es apropiado para lograr mediciones correctas de los parámetros.
- **Nivel de señal escaso.** El nivel de sonido es un poco bajo, pero todavía puede resultar adecuado para el análisis dependiendo del valor de la relación señal a ruido.
- **Nivel de señal insuficiente. Grabar de nuevo.** El nivel de sonido es insuficiente, y puede provocar mediciones incorrectas de los parámetros dependiendo del valor de la relación señal a ruido

El sonido se puede grabar en un archivo en disco para su análisis ingresando el *nombre del archivo* y luego presionando el botón *Grabar*.



2.7. Solapa de Cooperación

SAV permite a los usuarios cooperar con los desarrolladores enviando sugerencias mediante la solapa de *cooperación*. Tal como se muestra en la figura se le solicita el mensaje de la cooperación, y también ofrece la posibilidad de adjuntar un archivo de audio.

Analisis grafico **Analisis estadistico** **Grafico Radial** **Seleccionar archivo** **Grabacion** **Cooperación**

Descripción de la cooperación

Hola.

Aquí le envío un archivo de sonido de un paciente. Si necesitan mas información pueden contactarme en mi dirección de correo electrónico: minombre@miservidor.com.ar

Saludos.

Mi Nombre

Se adjuntará en la cooperación el archivo /home/priorat/projects/SAV/V1.1/version/test/vocal a.wav

Carpeta

- lib
- test.wav
- vocal a.w...

Nombre del archivo

Formato del archivo

Capítulo 3

Versiones

Mejoras de la versión 1.0

- Cambios en la visualización de los datos para evitar problemas de refresco de los gráficos.
- Mejora de la selección de archivos (filtrado, verificación del formato de 44100Hz y 16bits).
- Gráfico radial de los parámetros con regiones roja (anormal) y verde (normal).

Mejoras de la versión 1.1

- Posibilidad de analizar múltiples archivos al mismo tiempo.
- Grabación de audio, con estimación de la relación señal a ruido y la adecuación del nivel de volumen.
- Solapa de cooperación para ayudar a los desarrolladores de SAV.

Bibliografía

- [1] R. Baken and R. Orlikoff. Clinical measurement of speech and voice. In *Second Edition. San Diego, CA: Singular Publishing Group*, 2000.
- [2] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceeding of Institute of Phonetic Sciences*, volume 17, pages 97–110, 1993.
- [3] R. Fernández, D. Damborenea, P. Rueda, E. García, J. Leache, M.A. Campos, E. Llorente, and M.J. Naya. Análisis acústico de la voz normal en adultos no fumadores. In *Acta Otorrinolaringología*, volume 50, pages 134–141, 1999.
- [4] V. Parsa and D. Jamieson. A comparison of high precision fo extraction algorithms for sustained vowels. In *Journal of Speech and Hearing*, volume 42, pages 112–126, 1999.
- [5] Ch. Read, E. Buder, and R. Kent. Speech analysis systems: An evaluation. In *Journal of Speech and Hearing*, volume 35, pages 314–332, 1992.