

Sistema de análisis de la voz aplicado a fonología

Pablo D. Agüero¹, Graciela Moscardi², Juan C. Tulli¹, Esteban L. González¹, Alejandro J. Uriz¹ y Melisa G. Kuzman¹

¹Laboratorio de Comunicaciones – Facultad de Ingeniería – Universidad Nacional de Mar del Plata, {melisakuzman,ajuriz,pdaguero}@fi.mdp.edu.ar, 0223-4816600 Int 251, Juan B. Justo 4302 (7600) Mar del Plata.

²Universidad FASTA, Mar del Plata

Resumen—El análisis de la voz usando software específico constituye una herramienta importante para el profesional dedicado a la vocología. En este artículo se presenta la experiencia interdisciplinaria entre los desarrolladores de un software de análisis de la voz y los especialistas que lo utilizarán en la terapéutica. Por ello se describe brevemente el software y la experiencia de la estimación de la evaluación subjetiva RASATI usando algoritmos de aprendizaje automático. Además, también se explican las experiencias del uso de SAV por parte de los especialistas en la terapéutica.

Palabras clave— análisis, voz, software, RASATI, ORL

I. INTRODUCCIÓN

Los sistemas de análisis de la voz son herramientas importantes para el profesional dedicado a la vocología. La aplicación de la computadora en el área de la fonología tiene un rol destacado como instrumento de evaluación y seguimiento de patologías. En la clínica es útil acompañar el diagnóstico y la terapéutica con sistemas de análisis objetivo para comparar valores, confirmar tratamientos o reorientarlos.

Por otro lado, este tipo de aplicaciones también son útiles para los profesionales como instrumento de investigación. El fonólogo debe contar con diferentes pruebas para orientar al paciente y realizar una devolución durante el tratamiento, estableciendo una relación y comparación entre técnicas de análisis de la voz subjetivas como escalas comparativas y los elementos diagnósticos objetivos.

La finalidad de este trabajo es presentar la experiencia interdisciplinaria entre los desarrolladores de un sistema de análisis de la voz y los especialistas que lo utilizarán en la terapéutica. Mas específicamente, en este trabajo se describe la experiencia del desarrollo del Sistema de Análisis de la Voz (SAV), el cual fue implementado por los autores, pero que originalmente estaba orientado sólo a realizar medidas objetivas de la voz. Luego, a partir de la cooperación con la especialista se desarrolló una nueva versión del programa capaz de aproximar resultados subjetivos utilizando las mediciones objetivas realizadas.

El artículo está organizado de la siguiente manera. En la Sección II se describe el Sistema de Análisis de la Voz (SAV), software que permite el análisis objetivo de la voz.

En la Sección III se detalla el uso de las medidas objetivas para aproximar las medidas subjetivas usadas ampliamente por los especialistas. Posteriormente, en la Sección IV se presenta una descripción de la experiencia del especialista en el uso del software. Finalmente, en la Sección V se presentan las conclusiones y las perspectivas futuras.

II. SISTEMA DE ANÁLISIS DE LA VOZ

En las últimas décadas se ha difundido el uso de la computadora como herramienta objetiva para el análisis de la voz. Este tipo de sistemas de análisis procesan grabaciones de señales de voz, y permiten obtener un conjunto de medidas objetivas de la fonación. Algunos ejemplos de este tipo de herramientas son el Praat[1], el Multi-Dimensional Voice Program (MDVP)[2] y el Sistema de Análisis de la Voz (SAV)[3], este último en continuo desarrollo en el Laboratorio de Comunicaciones de la Facultad de Ingeniería de la Universidad Nacional de Mar del Plata.

SAV posee una gran variedad de parámetros, y el número de ellos está en constante expansión mediante el seguimiento de las publicaciones relacionadas con el estudio de voces patológicas.

Los parámetros disponibles son:

- Jitter y sus diferentes variantes: relativo (jitr), absoluto (jitta) y perturbación relativa promedio (jitrtrap y Jittrppq5).
- Shimmer y sus diferentes variantes: relativo (shimr), absoluto (shima) y perturbación relativa promedio (shimrap y shimppq5).
- Relación de armónicos a ruido (HNR).
- Índice de fonación débil (SPI).

Para poder evaluar la voz disfónica a través de dichos parámetros numéricos, es importante definir previamente la voz normal de aquella que no lo es, y disponer de valores normativos de comparación. Como algunos autores señalan [4], en nuestro ámbito geográfico son muy escasos los estudios llevados a cabo en este sentido. Además, algunos valores de dichos parámetros pueden ser

dependientes de los algoritmos usados por el software específico que los calcula [5, 6], lo que hace más necesario, 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 precisión y robustez adecuada que permita confiar en 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 esta 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 y realizar análisis y comparaciones válidas.

En la Fig. 1 se muestra la pantalla principal de SAV, que consta de una serie de elementos informativos tanto para el especialista como para el paciente. En la parte superior se puede observar la señal de voz, acompañada a continuación por el espectrograma de la misma.

En la posición inferior izquierda se puede observar un gráfico radial que indica el valor de los diversos parámetros de la voz que calcula SAV para la fonación visualizada. Si el parámetro de encuentra en la región verde, esto indica que su valor puede ser considerado normal. Por otra parte, si el valor del parámetro excede los límites del rango normal, este aparecerá en la región roja del gráfico radial.

En la misma pantalla también se muestra el conjunto de archivos de audio que se encuentran siendo estudiados por el especialista. Esto último resulta muy útil para poder observar la evolución del tratamiento tanto por el especialista como por el paciente, para notar el progreso.

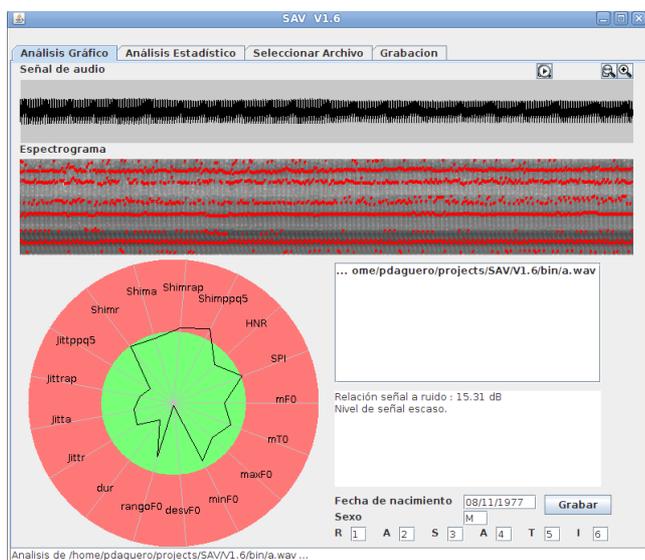


Fig. 1: Pantalla principal de SAV.

En la posición inferior derecha puede verse una evaluación por parte del programa de la calidad de la grabación. Esto es otro aspecto que debe ser tenido en cuenta por parte del especialista para cuidar la calidad de la grabación. La mala calidad de una grabación puede perturbar la precisión de los parámetros estimados por SAV, y la evaluación objetiva podría ser incorrecta.

Presionando en la solapa de análisis estadístico es posible observar (Fig. 2) en una tabla los valores de los parámetros de la voz de los diferentes audios que están siendo analizados. Esto permite realizar una comparación más precisa de la evolución mediante valores numéricos, y también con respecto a los umbrales que permiten distinguir valores normales de los que no lo son.

SAV V1.6								
Análisis Gráfico	Análisis Estadístico	Seleccionar Archivo	Grabacion					
Parámetro	Nombre	Unidad	Umbral	Audio 1	Audio 2	Audio 3	Audio 4	Audio 5
Frecuencia fundamental media	mF0	Hz	250	128.3				
Período medio	mT0	ms	5.6	7.8				
Frecuencia fundamental máxima	maxF0	Hz	250	132.0				
Frecuencia fundamental mínima	minF0	Hz	100	125.6				
Desviación de la frecuencia fundamental	desvF0	Hz	10	1.1				
Rango de frecuencia fundamental	rangoF0	st	1.5	0.9				
Duración del segmento analizado	dur	s	10	2.0				
Jitter relativo	jlttr	%	1.04	0.4				
Jitter absoluto	jltta	us	83.2	29.1				
Perturbación relativa promedio del jitter	jlttrap	%	0.68	0.2				
Pert. rel. prom. de 5 períodos del jitter	jltppq5	%	0.84	0.2				
Shimmer relativo	Shimr	%	3.81	3.6				
Shimmer absoluto	Shima	dB	0.35	0.3				
Perturbación relativa promedio del shimmer	Shimrap	%	2.0	2.2				
Pert. rel. prom. de 5 períodos del shimmer	Shimppq5	%	2.0	2.8				
Relación de armónicos a ruido	HNR	dB	15	25.0				
Índice de fonación débil	SPI	dB	10	9.5				

Fig. 2: Pantalla de análisis estadístico de SAV.

III. ESTIMANDO RASATI USANDO MEDIDAS OBJETIVAS

En el análisis de los desórdenes del habla existe una clara diferencia entre las medidas objetivas y subjetivas que forman parte de la evaluación de una fonación. Por un lado, las medidas objetivas se basan en ciertos cálculos que se hacen sobre la señal de voz tanto en el dominio del tiempo como en el de la frecuencia. Por otra parte, las medidas subjetivas se basan en escalas que abarcan conceptos más complejos, que pueden tener una correlación con un conjunto de medidas objetivas.

Un ejemplo de criterio subjetivo para la evaluación de la fonación es la escala RASATI, la cual indica la severidad de una patología usando una escala discreta que va desde cero hasta tres, considerando aspectos tales como ronquera (R), aspereza (A), soplosidad (S), astenia (A), tensión (T), e inestabilidad (I).

En los artículos de Agüero et al [7,8,9,10] se puede observar el uso de diversos parámetros objetivos para estimar los valores de la escala RASATI correspondientes a diferentes oralizaciones. Para ello se utilizó una base de datos de fonaciones en la que se incluye la valoración RASATI realizada por un especialista. Luego, mediante diferentes algoritmos de aprendizaje automático, se construyeron diferentes sistemas para estimar la valoración RASATI de las fonaciones dado un conjunto de medidas objetivas de la voz, tales como jitter, shimmer, HNR y SPI. Los algoritmos de aprendizaje automático incluyeron árboles de clasificación, redes neuronales, y regresión lineal.

Los resultados (Fig. 3) revelaron que el error en la estimación del valor RASATI está en el orden de ± 1 en el

80% de los casos. Tales errores no son severos debido a que la escala RASATI solamente posee cuatro valores posibles, y cualquier percepción ligeramente diferentes produce una fluctuación ± 1 .

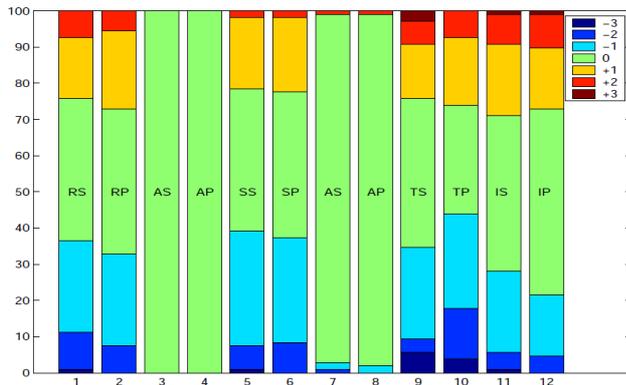


Fig. 3: Error en la estimación del valor RASATI.

IV. EL USO DE SAV EN FONOAUDIOLÓGIA

La utilización de SAV permite fundamentar objetivamente el éxito de la terapéutica fonoaudiológica, además de comparar y reafirmar los diagnósticos enviados por el médico ORL.

Es importante durante el desarrollo de la consulta comparar los valores acústicos para certificar el éxito de la terapéutica. Durante la consulta se solicita al paciente la emisión de la vocal /a/ y la vocal /e/ sostenida. Las voces evaluadas corresponden a diversas patologías:

- Disfonía orgánica con alteración de órgano emisor, como nódulos, pólipos, edemas cordales, papilomatosis, parestia, papiloma y parálisis de cuerda vocal
- Disfonía Disfuncional sin alteración del órgano emisor.

Las mediciones se realizan y comparan con escala RASATI, adaptada en Brasil por Pinho y Pontes [11] de la escala mundialmente difundida GARBAS (creada por el Comité para Test de la Función Fonatoria de la Sociedad Japonesa de Logopedia y Foniatría.

La escala GRBAS, para la evaluación vocal perceptiva a nivel glótico fue considerada por Hirano [12] escala de ronquera, donde G - Grado; R - Rough, Aspero; B - Breath, Soplado; A - Asthenic, Astenico; S - Strain, Tenso.

En las evaluaciones con escala RASATI los datos obtenidos son puramente subjetivos debiendo el evaluador poner en juego sus habilidades perceptivas acústicas, fisiológicas, y psicológicas, por eso la importancia del análisis acústico de la voz. De acuerdo a las referencias de la escala se obtiene el correlato como se detalla con los valores numéricos y la fisiología de la fonación.

La ronquera es debida a la irregularidad vibratoria de la mucosa de la cuerda vocal durante la fonación. Ante la presencia de alteración como nódulos se presenta el ronquido característico con ruidos en frecuencias bajas, modificando el trazado espectrográfico de los armónicos.

La aspereza implica rigidez de la mucosa de los pliegues vocales que causa irregularidad vibratoria. Presenta una

cualidad vocal característica con voz opaca y sin proyección. En el análisis acústico se evidencia presencia de ruidos en las frecuencias altas.

La soplosidad corresponde a la presencia de ruido de fondo audible cuyo correlato fisiológico más frecuente es la presencia de hendidura glótica.

La astenia está relacionada con la hiperfonación con poca energía. Representa quiebres o falta de potencia en la emisión vocal: se pierden armónicos agudos y la frecuencia fundamental y la amplitud se vuelven inestables.

La tensión está asociada al esfuerzo vocal por aumento de aducción glótica (hiperfunción) generalmente relacionada al aumento de la actividad de la musculatura extrínseca de la laringe, causando su elevación. Son ejemplo de ello la disfonía espasmódica de aducción, síndromes vocales tensionales, etc.

Finalmente, la inestabilidad en la emisión está asociada a la falta de sostén del soplo fonatorio, provocando una emisión irregular

Es importante resaltar que las variaciones en el tracto vocal producen modificaciones en la voz que son evaluados con SAV y la escala RASATI, comparando, analizando, y en el transcurso de la terapéutica el profesional puede valerse de los análisis con mayor precisión para observar la evolución de su paciente.

El SAV permite al fonoaudiólogo guardar los datos relativos a la escala RASATI y en una misma observación obtener un seguimiento del paciente, ofreciendo la facilidad de visualización y audición del paciente grabado.

Otro punto importante de este software es que permite comparar valores acústicos de distintos audios, lo que favorece el control de la evolución en la terapéutica, mostrar al paciente los avances y logros o no de su tratamiento, datos importantes que posteriormente el terapeuta analiza y evalúa la continuidad de la ejercitación. Además, con los valores obtenidos se puede controlar el momento de nuevos exámenes del ORL, la posibilidad del alta y la evolución de la patología.

Con el gráfico radial el paciente puede visualizar lo evaluado para que dicha observación ayude a la concientización de su patología. Es importante destacar que los sistemas de análisis de la voz como el SAV son una herramienta muy importante para los tratamientos de la voz, y deben complementar el diagnóstico mediante comparaciones con datos perceptuales y visuales.

Al comparar la situación previa y posterior al tratamiento es importante el control de la intensidad y frecuencia de la producción vocal ya que influyen en los grados de perturbación y ruido. Además los resultados de valoración perceptual deben coincidir con los diagnósticos laringológicos correspondiéndose con los hallazgos clínicos.

El análisis y la interpretación de los datos serán de utilidad cuando el terapeuta posee experiencia clínica para evaluar la información. El análisis acústico complementará el

análisis perceptual, y juntos favorecen la terapéutica en beneficio de una optimización en el trabajo fonaaudiológico.

V. CONCLUSIONES

En este artículo se presentan diversos aspectos del trabajo conjunto entre los desarrolladores de SAV y los especialistas que lo utilizarán en la terapéutica.

En el mismo se describió brevemente tanto el software como los intentos realizados para la estimación de los valores de la evaluación subjetiva RASATI usando parámetros objetivos y algoritmos de aprendizaje automático.

Los resultados observados hacen necesario continuar trabajando tanto en la adaptación del software a las necesidades que van surgiendo y corrigiendo posibles deficiencias.

El error observado en la estimación de la evaluación subjetiva RASATI implicará realizar nuevos esfuerzos en la generación de parámetros objetivos adicionales y la utilización de diferentes algoritmos de aprendizaje automático para mejorar los resultados actuales.

Por lo mencionado anteriormente, resulta necesario aclarar lo que indicamos desde que este software se hizo disponible libremente al público en el sitio web de nuestro grupo: "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".

Las experiencias y las sugerencias aportadas por los especialistas y usuarios de SAV resultaron y serán muy enriquecedoras. Esto se puede observar claramente en la evolución de SAV, que en estos momentos se encuentra en la sexta versión.

REFERENCIAS

- [1] P. Boersma and D. Weenink, "Praat: Doing Phonetics by Computer", <http://www.fon.hum.uva.nl/praat/>.
- [2] Multi-Dimensional Voice Program, www.kaypentax.com.
- [3] P. D. Agüero, J. C. Tulli, E. L. González, A. J. Uriz y F. De La Cruz Arbizu, "SAV: un sistema de análisis acústico para la evaluación de la voz", en IBERDISCAP 2011. Palma de Mallorca, España. Junio, 2011.
- [4] R. Fernández, D. Damborenea, P. Rueda, E. García, J. Leache, M.A. Campos, E. Llorente, y M. J. Naya, "Análisis acústico de la voz normal en adultos no fumadores", en Acta Otorrinolaringología, volumen 50, páginas 134-141, 1999.
- [5] V. Parsa y D. Jamieson, "A comparison of high precision fo extraction algorithms for sustained vowels", en Journal of Speech and Hearing, volumen 42, páginas 112-126, 1999.
- [6] Ch. Read, E. Buder, and R. Kent, "Speech analysis systems: An evaluation", en Journal of Speech and Hearing, volumen 35, páginas 314-332, 1992.
- [7] P. D. Agüero, J. C. Tulli, G. Moscardi, E. L. Gonzalez y A. J. Uriz, "Estimating RASATI scores using acoustical parameters", en Journal of Physics: Conference Series (JPCS). London, United Kingdom. December, 2011.
- [8] P. D. Agüero, J. C. Tulli, G. Moscardi, E. L. Gonzalez y A. J. Uriz, "Estimation of RASATI scores using objective acoustical parameters and CART", en XIV Reunión de Trabajo en Procesamiento de la Información y Control - RPIC 2011. Oro Verde, Argentina, 16-18 de Noviembre, 2011.

- [9] P. D. Agüero, J. C. Tulli, G. Moscardi, E. L. Gonzalez y A. J. Uriz, "Estimating RASATI scores using acoustical parameters", en XVIII Congreso de la Sociedad Argentina de Bioingeniería - SABI 2011. Mar del Plata, Argentina. 28-30 de Septiembre, 2011.
- [10] P. D. Agüero, J. C. Tulli, G. Moscardi, E. L. Gonzalez y A. J. Uriz, "Estimation of RASATI scores using objective acoustical parameters", AST 2011 - 40^o Jornadas Argentinas de Informática. Córdoba, Argentina. 29 y 30 de Agosto, 2011.
- [11] S. R. Pinho y P. Pontes, "Escala de evaluación perceptiva de la fuente glótica: RASAT", 2002.
- [12] M. Hirano, "Clinical examination of voice", Viena, Springer-Verlag, 1981.