

INFORME TECNICO

Sistema de Evaluación Audio-Perceptual EVAPER

Ing. Dr. Jorge A. Gurlekian

Bio Ing Dr. Humberto Torres

Resumen

EVAPER es un sistema de evaluación perceptual basado en el método psico-acústico de producción de la magnitud y “matching” (Stevens, 1975). La comparación ha mostrado ser el método más adecuado para evaluar perceptualmente el grado de aspereza de la voz de un paciente cuando se contrastan los desvíos de las respuestas obtenidos con el método clásico de estimaciones numéricas (Gurlekian, Torres y Vaccari, 2015). La diferencia en EVAPER se sustenta en el uso de las memorias externas de referencia, en contraposición al uso de la memoria propia de cada experimentador, cuando este estima los atributos de la voz. En el sistema EVAPER las dos señales a comparar son acústicas por lo que la comparación se corresponde con el método de matching intramodal definido por Stevens (1975). La producción se realiza al ajustar además de la aspereza, las referencias externas de los restantes atributos, como sople, astenia, tensión y temblor. Una de las señales corresponde a la emisión de voz del paciente. La otra señal es la señal producida artificialmente con las referencias fijas del propio sistema. El usuario puede definir su estrategia identificando las señales de referencia más próximas para finalmente realizar la comparación por pares para cada rasgo.

Uso de EVAPER

La voz del paciente es evaluada secuencialmente mediante la selección de los siguientes parámetros:

Altura tonal (*pitch*)
Sonoridad (*loudness*)
Timbre (*formant structure*)
Aspereza/ronquera (*roughness*)
Ruido o escape de aire o sople (*breathiness*)
Astenia o tensión (*asthenia-strain*)
Temblor o inestabilidad (*tremor*)
Cantidad de quiebres (*breaks*)

La Figura 1 muestra la interfase grafica del usuario.

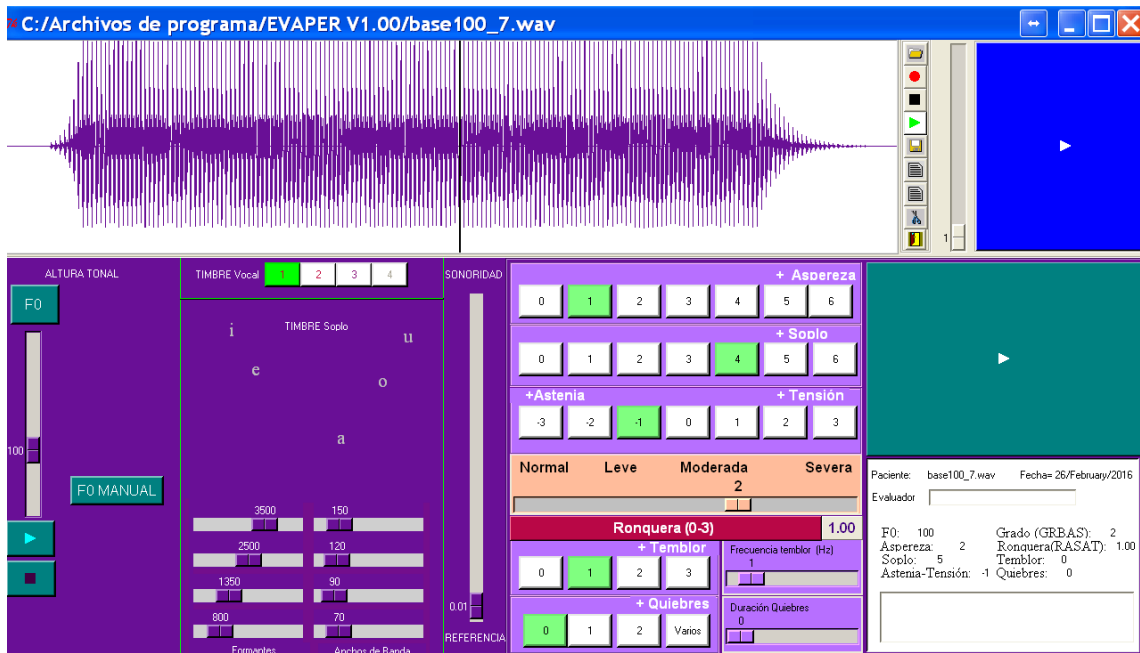


Figura 1. Pantalla principal de la interfase gráfica del usuario (IGU) del sistema EVAPER. Arriba: forma de onda correspondiente al paciente. En el centro botones de referencias externas de aspereza, soplo, astenia-tensión, grado, ronquera, temblor y quiebres.

La secuencia de producción/selección de la señal a comparar es sencilla y requiere la intervención del especialista en voz. Cada determinación de grado es arrastrada o considerada en la siguiente comparación comenzando jerárquicamente con la altura tonal, la sonoridad, el timbre y luego la aspereza, el soplo y la astenia o tensión. En cualquier instante es posible retroceder y corregir la asignación otorgada a cada atributo.

Nota: La determinación de la altura tonal puede ser asistida mediante el cálculo automático ó manual del F0. Para la determinación automática se requiere pintar un segmento extenso de la onda del paciente y oprimir el botón Captura F0 AUTO. Para la determinación manual se requiere pintar un segmento breve de la onda y oprimir el botón Captura F0 MANUAL, en una ventana emergente se podrá medir un período de la señal y visualizar el valor de F0. En ambos casos se puede escuchar el segmento de onda haciendo click derecho sobre el segmento pintado. La determinación automática del F0 permite igualar la altura tonal teniendo en cuenta el cálculo de los formantes y anchos de banda de la vocal. El espacio del triángulo vocálico permite realizar esta aproximación en forma manual para los dos primeros formantes haciendo click en el espacio vocálico.

Una vez hecha esta determinación, la vocal /a/ empleada para la igualación de la aspereza tendrá un F0 próximo al obtenido. La frecuencia y nivel de muestreo es de 50.000 Hz y 16 bits.

Una vez ajustado el nivel de F0, la sonoridad y el timbre, se busca la igualación del grado de aspereza. El botón cero corresponde a la vocal sin aspereza. Una vez determinado el nivel de aspereza el botón finalmente elegido quedará iluminado. Luego se continúa con la igualación de soplo. El nivel 0 no introduce ningún grado de soplo, por lo que se escuchará la señal con la aspereza obtenida previamente. Los otros

niveles de soplo se agregarán al nivel de aspereza originalmente elegido. Es posible que la sensación de aspereza aumente levemente debido a la presencia del soplo. Si el cambio es perceptible se recomienda disminuir un punto el grado de aspereza. El timbre del soplo puede modificarse haciendo click en el espacio del triángulo vocálico. Se determina a continuación el nivel de astenia-tensión con siete grados desde -3 a +3. Se parte del nivel 0 en el centro ó fonación modal hacia sucesivos valores de astenia hasta llegar al máximo nivel de astenia de -3. Los valores positivos corresponden a grados mayores de tensión siendo +3 el máximo. En la presente implementación, como es habitual, los niveles de aspereza y soplo se mapean para determinar el grado en una escala del tipo GRBAS.

El agregado de niveles de temblor y quiebres permite realizar una comparación más precisa.

La ronquera se calcula mediante un promedio ponderado de F0, aspereza y soplo.

Referencias

Stevens SS, Psychophysics, Wiley, New York (1975).

Rabiner LR, Shafer R W. (2006). Digital Processing of Speech Signals, Pearson Education, 6.

Rabiner LR. (1968). Digital Formant Synthesizer for Speech Synthesis Studies. J. Acoust Soc. Am. 43, 822-828.

Klatt DH, & Klatt LC. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America, 87, 820-857.

Gurlekian JA, Torres HM and Vaccari ME. Comparison of Two Perceptual Methods for the Evaluation of Vowel Perturbation Produced by Jitter. Journal of Voice. Elsevier. (2015) [doi:10.1016/j.jvoice.2015.05.009](https://doi.org/10.1016/j.jvoice.2015.05.009)

Alzamendi GA, Schlotthauer G, Rufiner HR, Torres ME, Evaluation of a new model for vowels synthesis with perturbations in acoustic parameters, Latin Am Appl Res, 43 (2013), pp. 225–230.

Shrivastav R, Camacho A, Patel S, Eddins DA. (2011). A model for the prediction of breathiness in vowels. J. Acoust Soc Am. 129(3): 1605–1615.

Technical Report

EVAPER - Audio Perceptual Evaluation System

Eng. Jorge A. Gurlekian, PhD

Bio Eng Humberto Torres, PhD

Abstract

EVAPER is a perceptual evaluation system based on the psycho-acoustic method of magnitude production and matching (Stevens, 1970). Matching has shown to be the most suitable method to perceptually assess the degree of roughness of the patient voice given the contrast with the deviations of the responses obtained with the classical method of numerical estimation produced by speech therapists (Gurlekian, Torres and Vaccari, 2015). The difference is based on the use of external reference memories over the use of the own rater memory. In EVAPER the two signals are of the same modality: one is the signal corresponding to the patient voice. The other signal is produced artificially using a set of fixed references provided by the system. Besides roughness, external references of breathiness, asthenia, tension and tremor are also available. Starting with pitch, loudness, timbre and roughness the user can define its strategy by identifying the closest reference signals to finally make the pairwise comparison for each attribute.

How to use EVAPER

Patient voice is evaluated through de successive production/selection and comparison of the following parameters:

Pitch (*Altura Tonal*)
Loudness (*Sonoridad*)
Timbre (*Estructura de formantes*)
Roughness (*Aspereza*)
Breathiness (*Soplo*)
Asthenia-Strain (*Asthenia –Tensión*)
Tremor (*Temblor*)
Breaks (*Quiebres*)

Figure 1 shows a the user graphical interface.

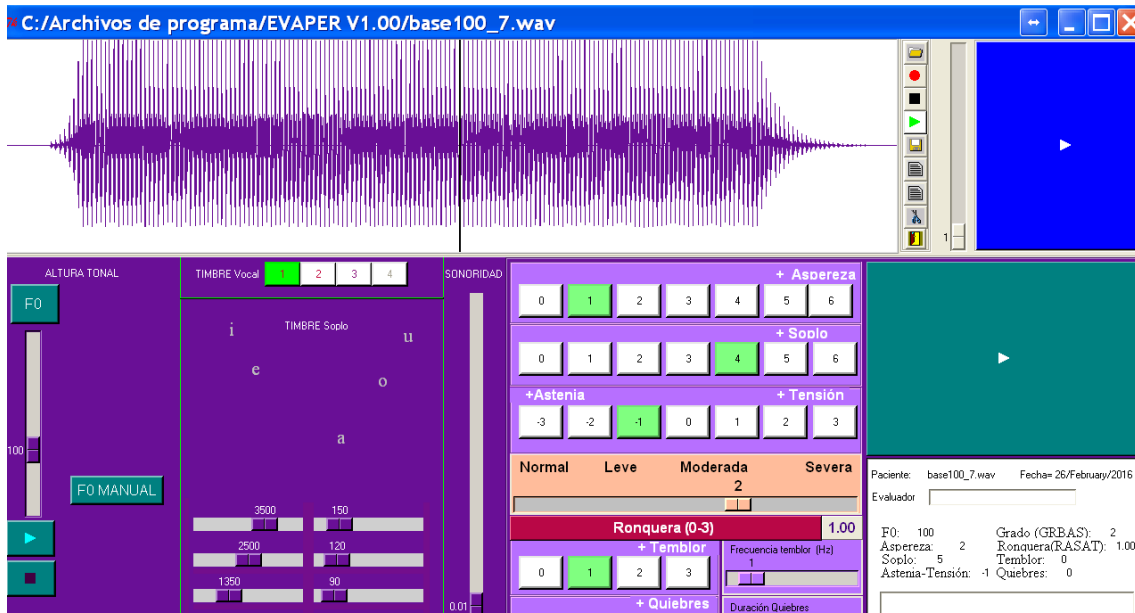


Figure 1. Main Screen of the user graphic interfase (GUI) of **EVAPER**. **Top:** Patient waveform. **Middle center:** external references for Roughness, Breathiness, Asthenia-Strain, Grade, Hoarseness, Tremor and Breaks. **Left:** capture tools for Pitch determination.

Production of the artificial signal to be compared with the patient voice is simple and requires the intervention of a voice or speech professional. Each grade determination is dragged or considered for the following feature comparison, hierarchically starting with pitch, then successively roughness, breathiness, and grade of asthenia/tension.

Note: For pitch determination, users may be assisted by the automatic or manual calculation of F0. Automatic determination requires to paint a large segment of the patient waveform and press the capture button Capture AUTO F0. For manual determination is required to paint a brief segment of the patient wave and press the Capture F0 MANUAL button then a pop-up window appear where period of the signal can be measured. In both cases users can hear the wave segment by right-clicking on the wave segment. Pitch is evaluated considering the formants and bandwidths of the patient voice. A vowel space triangle allows this approach manually for the first two formants by clicking in that space.

Once pitch determination is finished, a synthetic vowel /a/ with the corresponding F0 and a set of roughness levels are available for comparison. Vowel /a/ is synthesized by the LPC method at a sampling frequency of 50,000 Hz and 16 bits.

After adjusting the level of F0, roughness matching is sought. The zero button represents the voice without any degree of harshness. After determining the level of roughness the chosen button is highlighted. Then it continues with matching breathiness. Level 0 does not introduce any degree of breath, so that the signal with the previously obtained roughness is heard. The other breathiness levels will be added to roughness level originally chosen. You may expect roughness increase due to the

presence of breath. If the change is perceptible decrease one point of roughness. Breath timbre can be changed by clicking on the triangle vowel space.

Then, user determines the asthenia-strain level from seven degrees from -3 to +3. It starts at level 0 in the center or modal phonation to successive values of asthenia until the maximum level of -3. Positive values correspond to higher degrees of voice strain being the maximum +3. In the present implementation only roughness and breathiness are mapped to determine the extent of grade at the GRBAS like scale.

References

Stevens SS, Psychophysics, Wiley, New York (1975).

Rabiner LR, Shafer R W. (2006). Digital Processing of Speech Signals, Pearson Education, 6.

Klatt, D. H., & Klatt, L C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.

Gurlekian JA, Torres HM and Vaccari ME. Comparison of Two Perceptual Methods for the Evaluation of Vowel Perturbation Produced by Jitter. *Journal of Voice*. Elsevier. (2015) [doi:10.1016/j.jvoice.2015.05.009](https://doi.org/10.1016/j.jvoice.2015.05.009)

Alzamendi GA, Schlotthauer G, Rufiner HR, Torres ME, Evaluation of a new model for vowels synthesis with perturbations in acoustic parameters, *Latin Am Appl Res*, 43 (2013), pp. 225–230.

Shrivastav R, Camacho A, Patel S, Eddins DA. (2011). A model for the prediction of breathiness in vowels. *J. Acoust Soc Am*.129(3): 1605–1615.