

La medición de perturbaciones de frecuencia como indicador de la aspereza de la voz

C. Ferrer, M.E. Hernández-Díaz

CEETI, Facultad de Ingeniería Eléctrica, UCLV
Carretera a Camajuani, Km 5½, cferrer@uclv.etccsa.cu

This paper deals with the quantification of voice harshness in patients with voice disorders by means of digital signal processing. A previously reported index (the JF) is used to this end. It is obtained from the variability of the glottal pulse intervals found with a recently developed method. The results of the JF are compared with the subjective ratings made by two specialists of 48 recordings of the phonation of the sustained vowel 'a' by dysarthric patients. The results reveal that the JF can be a reliable index of harshness in pathological voice, in the same manner that the perceptual ratings are used.

Resumen. En el presente trabajo se aborda la cuantificación, mediante técnicas de procesamiento digital de señales, del grado de aspereza de la voz de pacientes con trastornos del habla. Para este fin se emplea un índice objetivo reportado en la literatura (conocido como JF, de *Jitter Factor*), obtenido a partir de la determinación de la posición de los pulsos glotales mediante un método desarrollado recientemente [13]. Se comparan los resultados del índice objetivo con las valoraciones subjetivas, emitidas por dos especialistas, de 48 grabaciones de la fonación por pacientes disártricos de una vocal 'a' sostenida. Los resultados obtenidos permiten afirmar que el *Jitter Factor* puede ser un indicador confiable de aspereza en voces patológicas en la misma medida en que lo son las valoraciones subjetivas.

Palabras clave: *procesamiento de voz, disartria, jitter, shimmer, aspereza.*

1. Introducción

La obtención de índices objetivos que reflejen el grado de perceptibilidad de características subjetivas del habla es una rama del procesamiento de voz al que se le dedica especial interés en el mundo. Este interés está dado por las aplicaciones que pueden tener estos índices en el diagnóstico, seguimiento y rehabilitación de personas con trastornos del habla.

Varios trabajos [3], [4], [6], [7], [8], [9], [10], [12] han mostrado que existe una correspondencia entre el tipo de trastorno y las características perceptuales del habla del enfermo. Dichas características perceptuales pueden agruparse en referidas al período fundamental, a la calidad, a la intensidad, a la articulación o a la prosodia.

Dentro de las referidas a la calidad se pueden mencionar la nasalidad, la aspereza, la ronquera y el jadeo.

La aspereza de la voz ha sido relacionada con las perturbaciones de la periodicidad de los pulsos glotales [5], [11], [16], [17] y pueden dividirse en perturbaciones de frecuencia (*jitter*) y de amplitud (*shimmer*). Estas perturbaciones de amplitud y frecuencia son provocadas por la vibración irregular de las cuerdas vocales, y rara vez parecen de forma independiente una de otra.

Como medidas de estas perturbaciones se han propuesto en la literatura [2] varios índices, entre los que se puede mencionar el *APQ* para el *shimmer*, y el *PPQ* y el *JF* (*Jitter Factor*) para el *jitter*. Todas estas medidas parten de la obtención previa de las posiciones y amplitudes de cada uno de los pulsos glotales en un sonido vocálico (preferentemente una vocal sostenida). La determinación de la posición de estos instantes de excitación en voces patológicas es poco confiable, pues las alteraciones de la periodicidad pueden ser de una magnitud tal que no se logre determinar correctamente el periodo fundamental en algunos segmentos de la vocalización. Esta falta de confiabilidad de los detectores de período fundamental en voces patológicas provoca que, aunque se acepte generalmente la relación entre la aspereza de la voz y el *jitter/shimmer*, no se hayan llegado a resultados definitivos en aplicaciones médicas.

En este trabajo se explora la correspondencia entre valoraciones subjetivas de aspereza en la voz de pacientes disártricos y el *jitter* obtenido a través de la utilización de un método de detección de período fundamental desarrollado recientemente.

2. Metodología

Se dispuso de 48 grabaciones [1] de pacientes disártricos efectuando la fonación de una vocal 'a' sostenida. Dos especialistas escucharon estas grabaciones y emitieron valoraciones subjetivas del grado de aspereza de cada una en una escala de 0 a 6, con 0 no perceptible y 6 aspereza extrema.

Como índice objetivo de aspereza se utilizó el *JF*, calculado según la expresión (1) donde con $F(i)$ se denota el inverso de la duración del pulso i -ésimo y N es la cantidad total de pulsos glotales en el segmento de voz bajo estudio.

$$JF = \frac{\frac{1}{N-1} \sum_{i=2}^N |F(i) - F(i-1)|}{\frac{1}{N} \sum_{i=1}^N F(i)} * 100 \quad (1)$$

La determinación de la posición de los pulsos glotales se efectuó mediante el método de detección del período fundamental pulso a pulso descrito en [13], que muestra mejores resultados en voces patológicas que los tradicionales. Este detector requiere de un estimado del período fundamental en un margen de $\pm 25\%$ del real. Dicho estimado fue obtenido mediante un detector básico de AMDF [15] para segmentos de

30 milisegundos. De todos los detectores de período fundamental por bloques el AMDF fue el que mostró un mejor seguimiento de la trayectoria del período fundamental en voces patológicas.

La confrontación entre los resultados del índice objetivo y las valoraciones subjetivas se efectuó siguiendo en lo fundamental el trabajo de Fukazawa *et al* [14] donde se desarrolla una tarea similar a la aquí descrita. En el mencionado trabajo se propone un índice objetivo para cuantificar el jadeo, que a su vez no estuviera significativamente correlacionado con la aspereza. Para ello usaron 24 grabaciones de vocales sostenidas de sujetos sanos y 34 de pacientes con afecciones del habla. Dos especialistas emitieron valoraciones subjetivas en una escala de 0 a 3, y a partir de estas se determinaron los coeficientes de correlación entre especialistas (0.714 en jadeo y 0.589 en aspereza), y entre el valor del índice propuesto y el promedio de las valoraciones de jadeo y aspereza de los especialistas (0.734 y 0.211 respectivamente). De aquí concluyen que el índice puede ser empleado como medida del jadeo de la voz en la misma forma en que se usa la percepción psicoacústica, ya que los resultados están en el mismo orden (0.734 y 0.714).

En el presente trabajo se decidió efectuar además un análisis de regresión lineal entre el índice objetivo y las valoraciones subjetivas, así como del histograma del coeficiente de correlación entre ambos. Estas pruebas brindan información adicional sobre la interrelación entre ambas variables.

Todos los algoritmos fueron programados en el paquete de software MatLab 5.2, incluyendo los de detección de período fundamental y el análisis estadístico.

La regresión lineal se efectuó entre el *JF* y el promedio de las valoraciones subjetivas (*VS*) de los dos especialistas. Se empleó la función *regress* del MatLab, que halla los valores b_1 y b_2 de la recta dada por:

$$JF = b_2 * VS + b_1 \quad (2)$$

además de brindar algunos datos estadísticos de interés, que se tratan con mayor detalle en las secciones 3 y 4 de este trabajo.

3. Resultados

En la Tabla 1 se muestran los coeficientes de correlación (ρ) obtenidos entre los especialistas (denominados 1 y 2 en la Tabla I) y entre el *JF* y el promedio de las valoraciones de ambos.

Tabla I
Coeficientes de Correlación Obtenidos

	1 vs. 2	<i>JF</i> vs. promedio(1,2)
ρ	0.7453	0.5996

Para tener una mejor idea de las dependencias manifestadas por el coeficiente de correlación se obtuvo un histograma del mismo. Este histograma fue hallado mediante el cálculo del coeficiente de correlación entre mil parejas de vectores (aleatoriamente conformados a partir del remuestreo de los originales) de *JF* y las valoraciones subjetivas. Los histogramas de las correlaciones entre los especialistas y de las correlaciones entre el *JF* y el promedio de los especialistas se muestran en la Figura

Los resultados de la regresión lineal se muestran en la Tabla II. La columna **b** muestra los valores de b_1 y b_2 , mientras la denominada **bint** muestra los intervalos del 95% confianza de b_1 y b_2 . En **stats** se muestran (de izquierda a derecha) el cuadrado del coeficiente de correlación y la probabilidad de haber obtenido la pareja de vectores *JF* y *VS* dado que ambos no estuvieran correlacionados.

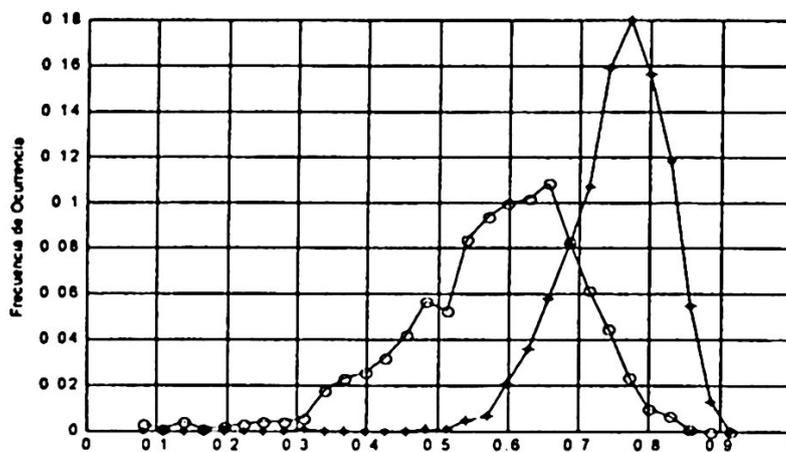


Fig. 1: Histogramas de las correlaciones entre los especialistas (-•-) y entre *JF* y el promedio de las valoraciones de los especialistas (-o-)

Tabla II

Resultados de la regresión lineal

	b	bint		stats	
1	0.6595	-0.253	1.572	0.3595	$6.67 \cdot 10^{-6}$
2	0.9946	0.6	1.3886		

4. Discusión

El valor de correlación entre los especialistas en la valoración subjetiva de la aspereza (0.745) es algo superior al reportado por Fukazawa (0.589) pero similar al obtenido por este último (0.714) en cuanto a la valoración del jadeo. La falta de un número mayor de especialistas impide emitir un criterio definitivo, pero estos resultados hacen

pensar que el intervalo de correlación de las valoraciones subjetivas oscile entre 0.6 y 0.7.

Por otra parte, la correlación entre el índice JF y el promedio de las valoraciones subjetivas (0.599) es realmente más bajo que el obtenido entre especialistas en este trabajo, pero similar al reportado por Fukazawa entre especialistas. Siguiendo el razonamiento planteado por este último, pudiera considerarse el JF como una opción de índice de aspereza en la misma medida que una valoración subjetiva, pues su correlación con estas últimas (0.599) está en el intervalo descrito (0.589-0.745) para estas.

Los histogramas mostrados en la figura 1 ratifican que la correlación entre los especialistas resultó superior. Determinar que este hecho haya sido casual o no requeriría, como se mencionó, de la existencia de un mayor número de especialistas. De hecho, es poco probable que un único índice describa a plenitud el nivel de perceptibilidad de determinada característica, debido a que en la percepción muchas veces influye más de un rasgo. En el mencionado artículo de Fukazawa se deseaba separar precisamente aspereza de jadeo, dos aspectos que tienen cierta relación en cuanto a su origen atribuido a la falta de periodicidad de los pulsos glotales. De cualquier forma, a partir de los resultados obtenidos se puede afirmar que las perturbaciones de frecuencia, cuantificadas a través del JF , constituyen un factor de la mayor relevancia en la percepción de la aspereza de la voz.

A esta afirmación contribuyen los resultados de la regresión lineal, pues puede apreciarse que el intervalo de confianza del factor b_2 , correspondiente a la pendiente de la recta que relaciona al JF con las valoraciones subjetivas, se encuentra dentro de un $\pm 40\%$ del determinado. Además, la probabilidad de haber obtenido el par de vectores JF y VS bajo la condición de que no existiera correlación es inferior a 10^{-5} , lo que denota una fuerte interrelación.

5. Conclusiones

Los resultados obtenidos ratifican la relación reportada en la literatura entre aspereza y perturbaciones de la periodicidad. El JF muestra una correlación con las valoraciones subjetivas de aspereza comparable a la existente entre especialistas.

Todo lo anterior indica que el método de detección de pulsos glotales empleado permite obtener resultados confiables en cuanto a la correlación entre las perturbaciones de la periodicidad (*jitter* y *shimmer*) y la percepción subjetiva de la aspereza en voces patológicas.

Referencias

- [1] Aronson, A.E. "Dysarthria: Differential Diagnosis". Mentor Seminars. Rochester, Michigan. (4 cassettes). 1993.
- [2] Baken, R.J. "Clinical Measurement of Speech and Voice". Singular Publishing Group, Inc. San Diego. 1996.
- [3] Berry, W.R.; Aronson, A.E.; Darley, F.L. & Goldstein, N.P. "Effects penicillamine therapy and low-copper diet on dysarthria in Wilson's disease (hepatolenticular degenerative)". Mayo Clinic Proceedings. 49, pp 405-408. 1974.
- [4] Berry, W.R.; Darley, F.L.; Aronson, A.E. & Goldstein, N.P. "Dysarthria Wilson's disease". Journal of Speech & Hearing Research. 17, pp 169-183. 1974.
- [5] Coleman, R.F. "Effect of waveform changes upon roughness perception". Folia Phoniatica. No 23. pp 314-322. 1971.
- [6] Darley, F.L.; Aronson, A.E. & Brown, J.R. "Clusters of deviant speech dimensions in the dysarthria". Journal of Speech & Hearing Research. 12, pp 462-496, 1969.
- [7] Darley, F.L.; Aronson, A.E. & Brown, J.R. "Differential diagnostic patterns dysarthria". Journal of Speech & Hearing Research. 12, pp 246-269, 1969.
- [8] Darley, F.L.; Aronson, A.E. & Brown, J.R. "Motor Speech Disorders" Philadelphia. Saunders. 1975.
- [9] Darley, F.L.; Aronson, A.E. & Brown, J.R. "Motor Speech Signs in Neurologic Disease". Medical Clinics of North America, 52, 835-844. 1968.
- [10] Darley, F.L.; Brown, J.R. & Goldstein, N. "Dysarthria in multiple sclerosis" Journal of Speech & Hearing Research. 15, pp 229-245, 1972.
- [11] Deal, R.E. & Emanuel, F.W. "Some waveform changes upon roughness perception". Journal of Speech & Hearing Research. 21, pp 250-264, 1978.
- [12] Enderby, P. "Frenchay Dysarthria Assessment". California. College Hill. 1983.
- [13] Ferrer, C., "Cuantificación de parámetros subjetivos de la voz para el diagnóstico de Disartrias". Tesis de Master, UCLV, 2000.
- [14] Fukazawa, T.; El-Assuooty, A. & Honjo, I. "A new index for evaluation of the turbulent noise in pathological voice". Journal of the Acoustical Society America. Vol. 83, No 3. pp 1189-1193. March. 1988.
- [15] Hernández-Díaz M.E., "Algoritmos para la extracción del período fundamental de la voz. Desarrollo y evaluación". Tesis doctoral. UCLV. Cuba 1996.
- [16] Wendall, R.W. "Approach to the objective evaluation of hoarseness". Folia Phoniatica. No 18. pp 98-108. 1966.
- [17] Wendall, R.W. "Laryngeal Analog Synthesis of Harsh voice Quality". Folia Phoniatica. No 15. pp 251-255. 1963.

Carlos Ariel Ferrer Riesgo; Centro de Estudios de Electrónica y Tecnologías de Información (CEETI); UCLV; Carretera a Camajuaní km 5½, Santa Clara; Villa Clara; Cuba; email: cferrer@uclv.eteccsa.cu; Fax: (53) (422)-281608; Tel: (53) (422)-281157

María Esperanza Hernández-Díaz Huici; Centro de Estudios de Electrónica y Tecnologías de la Información (CEETI); UCLV; Carretera a Camajuaní km 5½, Santa

Clara; Villa Clara; Cuba; email: mariae@uclv.etecsa.cu; Fax: (53) (422)-281608; Tel: (53) (422)-281157

Biografía corta de cada autor

Carlos Ferrer Riesgo, nació el 7 de Octubre de 1973. Graduado de Ingeniería en Telecomunicaciones y Electrónica en la UCLV en 1996. Graduado de Master en Telecomunicaciones en la UCLV en 2000. Se desempeña como investigador en el CEETI, UCLV desde 1996. Su área de interés es el procesamiento de voz para aplicaciones médicas.

María Esperanza Hernández-Díaz Huici, nació el 18 de Octubre de 1958. Graduada de Ingeniería en Telecomunicaciones en el ISPJAE en 1981. Obtiene el Título de Doctor en Ciencias en la UCLV en 1997. Labora como profesor asistente en el Centro de Estudios de Electrónica y Tecnologías de la Información (CEETI) de la UCLV desde 1992. Sus intereses principales están en el área del tratamiento digital de señales biomédicas, en particular la voz y la compresión de datos.

