CARLOS ARIEL FERRER RIESGO.
MARÍA ESPERANZA HERNÁNDEZ-DÍAZ HUICI.
Centro de Estudios de Electrónica y Tecnologías de la Información.
Universidad Central de Las Villas, Santa Clara, Cuba.

Obtención de un índice objetivo de razón lenta.

Introducción

El desarrollo de algoritmos para el cálculo de parámetros de la voz que permitan diferenciar las voces normales de las patológicas ha sido y sigue siendo un campo al que se le presta un gran interés en el mundo. En una serie de estudios llevados a cabo en las Clínicas Mayo en los finales de la década de los 60's y principios de los 70, Darley, Aronson y Brown [1, 2, 3, 4, 5, 6, 7, 8] examinaron las características del habla de grupos de pacientes disártricos desde un punto de vista auditivo y desarrollaron una metodología de diagnóstico a partir de las valoraciones subjetivas de determinados rasgos perceptuales. Estos rasgos fueron evaluados en escalas de 1 a 7 en nivel ascendente de severidad a partir de la realización por el paciente de tres ejercicios: fonación de una vocal "a" sostenida, repetición de los monosílabos con consonantes explosivas "Pa", "Ta" y "Ka" y la lectura de un párrafo estándar. Los resultados de estos estudios y de trabajos relacionados estrechamente con ellos, se consideran aun hoy las bases del diagnóstico diferencial clínico de las disartrias [12 , cap 3].

El uso de técnicas de procesamiento digital de señales en la caracterización automática de voces patológicas permite obtener medidas objetivas de algunos de los rasgos descritos en dicha metodología. En el presente trabajo se propone un método totalmente automático para la obtención de una medida para la valoración objetiva del rasgo denominado Razón Lenta a partir de grabaciones del ejercicio de repetición de sílabas.

Desarrollo:

La razón lenta del habla consiste en una baja razón de unidades del lenguaje en el tiempo, ya sean sílabas por segundo, oraciones por minuto u otras. Partiendo de este hecho, resulta lo más lógico efectuar la cuantificación de este parámetro en el ejercicio del párrafo, donde se evalúa la articulación del lenguaje fluido. No obstante, la razón lenta aparece como característica principal en el *cluster* número 2 del estudio de las Clínicas Mayo, denominado Exceso Prosódico, cuya causa probable es la lentitud de los movimientos repetitivos por defecto neuromuscular. Este hecho, junto a la dificultad de la detección automática de las sílabas u oraciones en el habla fluida de un disártrico hace considerar el ejercicio de la repetición de las sílabas 'Pa'-'Ta'-'Ka' como otro material a tener en cuenta en la determinación de la razón lenta.

En un trabajo reciente [10] (1999) se halla un parámetro objetivo que resulta indicador de la razón del habla. El parámetro (RDH) se calcula como la cantidad de sílabas en un segmento del habla fluida entre la duración del mismo en segundos. El RDH se halla de forma semiautomática, pues requiere que el especialista seleccione el segmento y además especifique el número de sílabas que contiene. Al trabajar con la lectura del párrafo para caracterizar el habla fluida, el RDH pudiera verse influido por las habilidades que tenga el evaluado en la lectura o tener ligeras diferencias para distintos idiomas. Teniendo en cuenta esto, junto con el objetivo de reducir la intervención requerida del especialista, en este trabajo se propone un parámetro indicativo de la razón lenta que se denomina RL (***R**azón **L**enta*). El RL, a diferencia del RDH, se obtiene en el ejercicio de la repetición de sílabas y se calcula según la expresión:

$$RL = \frac{T_T - T_P}{N - P - 1} \qquad \textbf{(1)}$$

Donde:
T_T : Tiempo transcurrido de la primera a la última sílaba.
T_P : Tiempo consumido en las pausas.
N : Número de sílabas pronunciadas por el paciente.

P : Número de pausas efectuadas.

Como puede apreciarse en la ecuación (1), las unidades del RL (segundos/sílaba) son inversas a las del RDH, por lo que a diferencia de este alcanza un mayor valor para una mayor lentitud del habla, en correspondencia con el índice subjetivo empleado en el estudio de las Clínicas Mayo.

En el cálculo del RL se ha prestado especial interés a la eliminación del efecto que pueden provocar en el valor del parámetro las pausas que realiza el paciente, no atribuibles a la separación con que pronuncia las sílabas. En la mayoría de los casos se observa una pausa más o menos larga al pasar de una consonante a la otra, lo que hace esperar dos pausas por este motivo: las transiciones P-T y T-K. Como además de estas pausas puede haber otras (debidas a que el paciente carraspee, respire o por otro motivo) en la obtención del RL se buscan hasta dos pausas adicionales, de forma que P puede tomar valores de 2 a 4.

La obtención de todas las variables empleadas en (1) se efectúa de forma automática. T_T y N se hallan empleando un detector de sílabas sencillo (se describe someramente más adelante), siendo N el número de sílabas detectadas y T_T el tiempo entre la primera y la última. El número de pausas P y su duración T_P también se hallan una vez obtenidas las posiciones de cada una de las sílabas detectadas. El valor de P se halla de la siguiente forma: se parte de P=2 y se asume que estas dos pausas son las transiciones entre consonantes y corresponden a las dos mayores separaciones entre sílabas presentes en la grabación. Con P=2 se halla un valor temporal de RL llamado RL', correspondiente a la duración promedio de las sílabas en segundos, sin la influencia de las pausas debidas a transición. La existencia o no de pausas adicionales se determina en referencia a este valor, si existen pausas mayores de *K* veces RL'. El valor de este umbral K se fijó a 2.5, pues si se tiene en cuenta que RL' es un promedio, la existencia de valores por encima de 2.5 RL' indica una variabilidad demasiado alta. Las mayores pausas que superan esta duración se eliminan (dos como máximo) y se calcula el valor definitivo de RL.

El detector de sílabas funciona a base de obtener los máximos locales en la envolvente de la curva de intensidad de la grabación y descartar los que no constituyan sílabas. En la eliminación de los máximos que constituyen falsos candidatos a sílaba se tomaron en consideración varias condiciones que deben cumplir para no ser eliminados:

- Los máximos tienen que tener una amplitud mayor que determinado umbral. Esto elimina todos los máximos detectados en las zonas de silencio, provocados por ruido. El umbral se colocó en 5 veces el mínimo valor de la envolvente en toda la grabación (E_{min}). Se toma como referencia este mínimo al considerarlo indicador de la amplitud del ruido ambiente.
- Los máximos deben estar separados un tiempo prudencial. Para poder ser consideradas sílabas diferentes tiene que haber una diferencia de tiempo apreciable. Si dos máximos se encuentran separados a menor tiempo del estipulado se toma en consideración el de mayor amplitud. Esta condición elimina los máximos cercanos en la misma sílaba. Se determinó empíricamente que la mayor frecuencia de repetición de sílabas es de 10 por segundo, por lo que la separación mínima aceptada entre máximos se fijó a 100 milisegundos.
- Entre dos máximos de diferentes sílabas debe haber un mínimo considerable. Para que se perciba la transición entre los monosílabos se necesita que exista una disminución de nivel apreciable. Al igual que en la condición anterior, si dos máximos no presentan esta reducción del nivel entre ellos, se continúa tomando en consideración como centro de la sílaba el de mayor amplitud. Esto elimina los máximos dentro de la misma sílaba si esta está anormalmente alargada. Se usó como condición que entre los dos máximos debe haber un mínimo con amplitud menor que ¾ de la del menor máximo menos E_{min}.
- La sílaba debe tener una duración determinada. La separación entre el mínimo más pronunciado anterior y el posterior a la sílaba debe ser suficiente para que esta sea percibida como tal. Así se eliminan máximos espurios que tengan amplitud considerable. Esta separación mínima se fijó a 20 milisegundos.

Los máximos que no son eliminados al chequear el cumplimiento de estas condiciones son considerados sílabas aisladas y sus posiciones en tiempo son las empleada en la obtención del índice RL.

Resultados:

El parámetro descrito debe ser validado de forma tal que se demuestre la correspondencia entre la magnitud objetiva, obtenida mediante su uso, y la evaluación perceptual de la característica subjetiva que se pretende medir con él (Razón Lenta). En el trabajo de Fukazawa *et al* [9] se realiza una tarea similar, con un índice para medir jadeo. Fukazawa encontró una correlación entre dos especialistas de 0.714 para jadeo y de 0.589 para aspereza de la voz. Su índice tuvo una correlación de 0.734 con el promedio de las valoraciones de los especialistas para el jadeo y de 0.211 con las de aspereza. De ahí concluyen que el índice puede ser empleado en la misma medida en que se usan las valoraciones subjetivas de jadeo.

En este trabajo se sigue en lo fundamental ese método de validación, aunque se brindan también otras medidas que dan una idea de la correspondencia entre las valoraciones subjetivas (Vsubj) y el índice RL. Además del coeficiente de correlación se muestra un gráfico del índice RL vs. Vsubj (Figura 1), que también se presenta en [9], junto a la recta de mejor ajuste y los valores medios (círculos en la Figura 1) de RL por valoración subjetiva. También se muestra un intervalo de confianza (***ci*** en la Tabla 1) para el valor del coeficiente de correlación obtenido y los coeficientes b_1 y b_2 (***b*** en la Tabla 1) de la recta de mejor ajuste $RL = b_2 * Vsubj + b_1$ junto a sus respectivos intervalos de confianza (***bint*** en la Tabla 1). En la Tabla 1 se muestran además (en ***stats***) tres valores correspondientes a la regresión lineal [11] de RL-Vsubj, el primero de ellos el porciento de varianza que el modelo de regresión lineal justifica, el segundo es el valor de F en la distribución de probabilidad del mismo nombre y el último indica la probabilidad de haber obtenido una muestra como la analizada para el caso de que las dos variables RL y Vsubj no estuvieran correlacionadas. Como indicador adicional se brinda el histograma del coeficiente de correlación entre RL y Vsubj. Los resultados del índice RL se obtuvieron sobre una muestra de 47 pacientes, y se muestran en la Figura 1 , la Tabla 1 y la Figura 2.

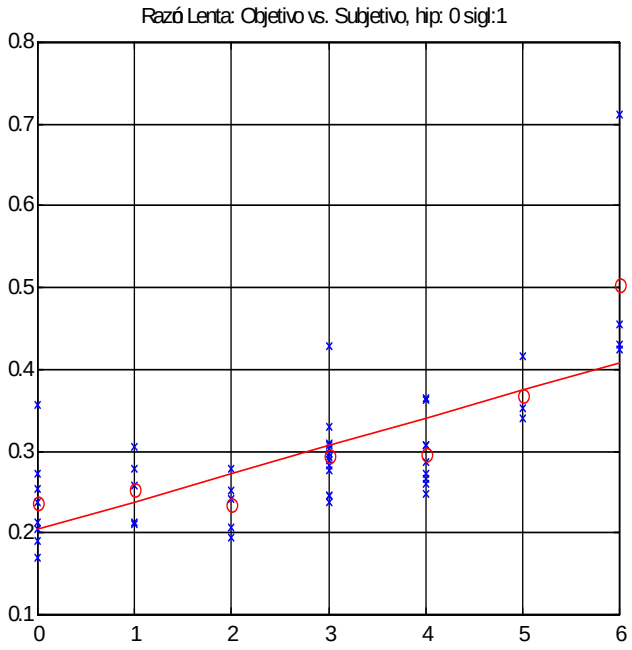

Figura 1 : Índice RL vs. percepción subjetiva de razón lenta. Recta de mejor ajuste y valores medios de RL en cada valor de percepción subjetiva.

Corr.	T-test	Regresión Lineal		
r	ci	b	bint	stats

0.6724	**0.6753**	**0.6848**	**0.2044**	**0.1675**	**0.2412**	**0.4521**	**37.134**	**0.0000**
			0.0340	**0.0228**	**0.0452**			

Tabla 1 : Resultados del índice RL.

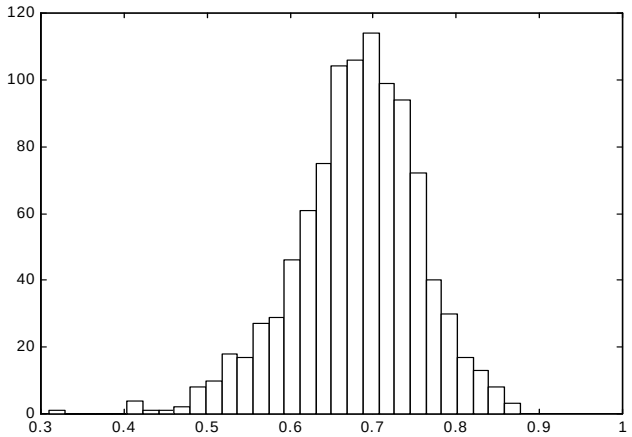

Figura 2 : Histograma del coeficiente de correlación entre el índice RL y la percepción subjetiva de Razón Lenta.

Conclusiones:

El coeficiente de correlación obtenido entre el índice propuesto y las valoraciones subjetivas (0.672) es similar al reportado por Fukazawa (0.734), y en el orden de la correlación entre especialistas para valoraciones subjetivas de características de la voz (0.589 y 0.714). En el intervalo de confianza de b_2 el valor superior apenas duplica el inferior, por lo que se puede apreciar que la relación RL-Vsubj es notable. Por otra parte, el tercer valor de stats indica que la probabilidad de que RL y Vsubj no estén correlacionadas es menor que 10-4. El histograma también muestra que la correlación entre ambas variables es notable, con casi la totalidad de los valores por encima de 0.5.

De acuerdo a estos resultados podemos afirmar que:
El ejercicio de la repetición de sílabas puede ser empleado para la determinación objetiva de la razón del habla.
El índice propuesto puede ser empleado para la valoración automática del rasgo Razón Lenta.

Referencias:

1. Aronson, A.E. “Dysarthria: Differential Diagnosis”. Mentor Seminars. Rochester. Michigan. (4 cassettes). 1993.
2. Berry, W.R.; Aronson, A.E.; Darley, F.L. & Goldstein, N.P. “Effects of penicillamine therapy and low-copper diet on dysarthria in Wilson’s disease (hepatolenticular degenerative)”. Mayo Clinic Proceedings. 49, pp 405-408. 1974.
3. Berry, W.R.; Darley, F.L.; Aronson, A.E. & Goldstein, N.P. “Dysarthria in Wilson’s disease”. Journal of Speech & Hearing Research. 17, pp 169-183. 1974.
4. 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.

5. Darley, F.L.; Aronson, A.E. & Brown, J.R. “Differential diagnostic patterns of dysarthria”. Journal of Speech & Hearing Research. 12, pp 246-269, 1969.
6. Darley, F.L.; Aronson, A.E. & Brown, J.R. “Motor Speech Disorders”. Philadelphia. Saunders. 1975.
7. Darley, F.L.; Aronson, A.E. & Brown, J.R. “Motor Speech Signs in Neurologic Disease”. Medical Clinics of North America, 52, 835-844. 1968.
8. Darley, F.L.; Brown, J.R. & Goldstein, N. “Dysarthria in multiple sclerosis”. Journal of Speech & Hearing Research. 15, pp 229-245, 1972.
9. Fukazawa, T.; El-Assuooty, A. & Honjo, I. “A new index for evaluation of the turbulent noise in pathological voice”. Journal of the Acoustical Society of America. Vol. 83, No 3. pp 1189-1193. March. 1988.
10. Ruiz Riol, R. “Programa para el cálculo de parámetros que cuantifican características de la voz”. Tesis de Grado de Ingeniero en Telecomunicaciones y Electrónica. UCLV. VC. Cuba. Julio. 1999.
11. Statistic Toobox User’s Guide. MathWorks. January. 1998.
12. Yorkston, K.M.; Beukelman, D.R. & Bell, K. “Clinical Management of Dysarthric Speakers”. PRO-ED. Austin. Texas. 1987.