

UN ALGORITMO PARA DETECTAR PERIODICIDAD
SIN PREFILTRADO

ING. GUALBERTO HIDALGO
ESCUELA POLITECNICA NACIONAL

ING. TANIA PEREZ
ESCUELA POLITECNICA NACIONAL

RESUMEN

En este artículo se presenta un nuevo método para detectar periodicidad en la palabra hablada. La técnica propuesta trabaja completamente en el dominio de tiempo y usa como criterio para la discriminación entre segmentos periódicos y aperiódicos la similitud que existe entre dos períodos consecutivos del lenguaje hablado. La detección de periodicidad se realiza sin ningún prefiltrado.

ABSTRACT

In this paper a new method to detect periodicity in speechwaves is presented. The proposed technique works completely in the time domain and uses as a criterium for the discrimination between periodic and aperiodic segments the similarity which exists between two consecutive periods of the speechwave. The detection of periodicity is performed without any prefiltering.

I. INTRODUCCION

Son innumerables las técnicas que se han utilizado para encontrar periodicidad en la palabra hablada. Para tener una idea de esta multiplicidad se puede consultar las referencias [1] y [2]. No obstante esta multiplicidad las mencionadas técnicas pueden agruparse de la siguiente manera:

- 1) Técnicas que analizan la onda de la voz humana en el dominio del tiempo. [3]-[9].
- 2) Técnicas que buscan periodicidad en la voz después de pasar la voz al dominio de la frecuencia. [10]-[11].
- 3) Técnicas híbridas que trabajan tanto en el dominio del tiempo como en el de la frecuencia. [12]-[13].

La nota común de todas estas técnicas es la distorsión mayor o menor a la que se somete a la onda para buscar periodicidad. De hecho la mayoría de las mencionadas técnicas prefiltran la señal para determinar el tono de la voz. Otra técnica que se usa frecuentemente para la determinación del tono de la voz es la autocorrelación, que da como resultado una señal diferente de la original. Por último las técnicas que trabajan en el dominio de la frecuencia, deben realizar primero la transformación tiempo-frecuencia, un proceso en el que la señal analizada prácticamente desaparece.

Otra característica común de los métodos arriba mencionados es la complejidad de los algoritmos. En efecto, para la mayoría de los mismos la búsqueda del tono de la voz conlleva una entera secuencia de tratamientos como: prefiltrado, autocorrelación, transformación en el dominio de frecuencia, etc.. Todas estas manipula-

ciones implican una gran cantidad de trabajo de computación. El presente algoritmo tiene la doble ventaja de ser simple desde el punto de vista computacional, y por otra parte no distorsiona la onda, pues trabaja con la onda muestreada original. (El único filtrado que se presupone es el que se requiere para el muestreo de la señal análoga). La simplicidad del algoritmo implica estar más cercanos a la condición de procesamiento en tiempo real, que es una condición que se impone para la mayoría de las aplicaciones del procesamiento de la palabra hablada. El hecho de que la onda permanece inalterada por la búsqueda del tono, hace posible usar este algoritmo como la base para otro algoritmo que segmenta la locución automáticamente y con un aceptable grado de confiabilidad en las siguientes categorías:

segmentos quasi-periódicos

segmentos aperiódicos

segmentos oclusivos

Esta segmentación puede ser de una gran ayuda para el ulterior procesamiento de la voz.

En el presente artículo se procede de la siguiente forma:

En la sección II se explican algunos conceptos básicos y criterios para la búsqueda de periodicidad. En la sección III se presenta el algoritmo utilizado con su correspondiente diagrama de flujo. Finalmente en la sección IV se dan algunos resultados que permiten evaluar el algoritmo.

II. CONCEPTOS BASICOS

Las ondas de la voz pueden ser caracterizadas de acuerdo a diferentes criterios. Si se considera la presencia o no de vibraciones de las cuerdas bucales la alocución puede ser:

voceada si requiere para su generación la vibración de las cuerdas bucales.
no voceada si es generada sin la intervención de las cuerdas bucales.

Bajo otro respecto la locución puede ser: periódica (o mejor dicho "quasi-periódica") y aperiódica. Para establecer la quasi-periodicidad de la voz, se deben dar los siguientes conceptos:

La periodicidad de una función de tiempo $f(t)$ se define con la ecuación:

$$f(t) = f(t+nT) \quad (1)$$

donde "n" es un entero positivo y "T" es el tiempo de duración de un período, esto es, el tiempo en que la función se repi-

te. Esta clase de periodicidad no existe en la locución. En efecto en la palabra hablada no existen dos periodos iguales, lo más que puede esperarse de dos "periodos" consecutivos es que estos sean similares. Por esta razón al tratar del lenguaje hablado es mejor hablar de "cuasi-periodicidad" que puede definirse como sigue:

$$f(t) \approx f(t+nT) \quad (2)$$

Esta definición cubre una entera gama de posibilidades, y está más en consonancia con la lógica imprecisa (fuzzy logic) que caracteriza al lenguaje hablado.

De acuerdo con la definición anterior la locución puede caracterizarse como "cuasi-periódica" si la ecuación (2) tiene vigencia, y como aperiódica si no la tiene. Esto no obstante, por razones de brevedad, en las páginas que siguen se hablará sólo de segmentos periódicos y aperiódicos.

Se debe admitir que el anterior criterio establece una regla evanescente, no claramente determinada para la discriminación entre segmentos periódicos y aperiódicos. Esto no obstante, los resultados en la práctica son buenos, como se podrá comprobar de los gráficos en que éstos se presentan.

Antes de terminar esta sección debe quedar establecido claramente que aquí los conceptos de periódico y voceado no son intercambiables. En efecto en las ondas que se analizan, hay segmentos que son claramente voceados sin ser periódicos.

III. EL ALGORITMO

Si, como se estableció previamente, la periodicidad de una onda puede detectarse por el hecho de que la misma obedece a la ecuación (2), entonces el núcleo de la cuestión está en encontrar algún discriminante para determinar con precisión donde la onda de locución satisface dicha condición.

Aquí cabe destacar la gran diferencia que existe entre el procesamiento paralelo de la visión humana y el procesamiento serial del computador. Así mientras al ojo humano le basta un golpe de vista para determinar los intervalos periódicos de la onda de locución, para el computador que recibe las muestras de dicha onda una por una la tarea está muy lejos de ser trivial.

La técnica que usa el presente artículo para la determinación de la periodicidad parte de la definición dada en la ecuación (2). Esto es, si para una onda periódica es verdad que $f(t) \approx f(t+nT)$, entonces cualquier magnitud correspondiente a dos periodos consecutivos será similar. Así por ejemplo los intervalos correspondientes entre cruces con cero de dos periodos consecutivos deben tener un valor semejante. Refiriéndose a la fig. 1,

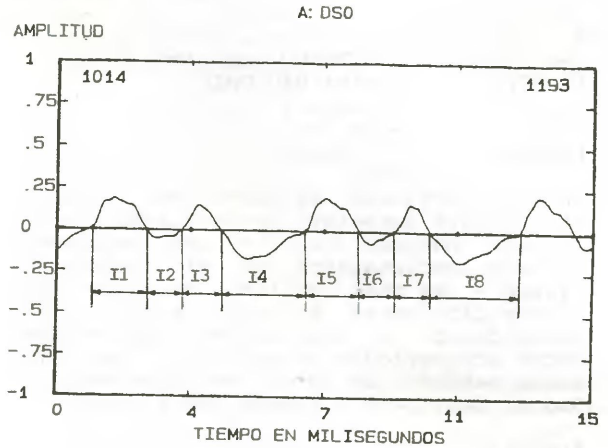


Fig. 1.

los intervalos: I1, I5
I2, I6
I3, I7
I4, I8

deben ser similares entre si en el orden indicado.

Entonces para detectar periodicidad se puede proceder de la siguiente manera:

a) obtener los intervalos entre cruces con cero para un segmento de onda mayor que dos de los periodos más largos de la voz.

b) comenzar la búsqueda de periodicidad de la siguiente manera:

Primer intento:

obtener la diferencia:

$|I1 - I3| + |I2 - I4|$; Ver fig. 2

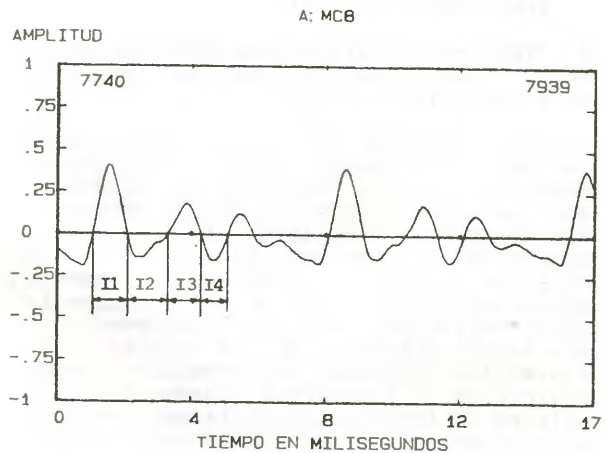


Fig. 2.

Segundo intento:

Obtener la diferencia:

$|I1 - I5| + |I2 - I6| + |I3 - I7| + |I4 - I8|$.

Ver fig. 3

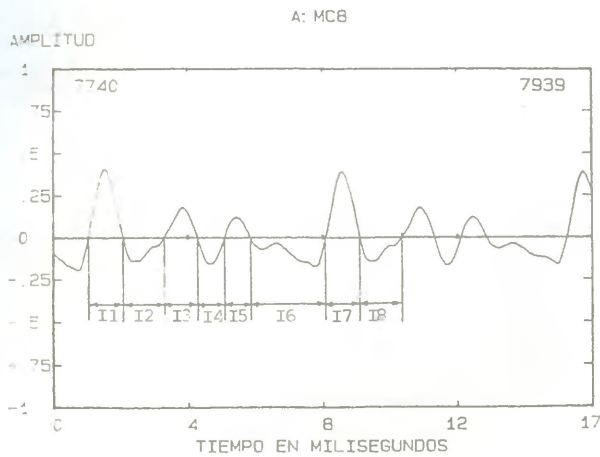


Fig. 3.

Tercer intento:

obtener la diferencia:

$|I1 - I7| + |I2 - I8| + \dots + |I6 - I12|$. Ver fig. 4

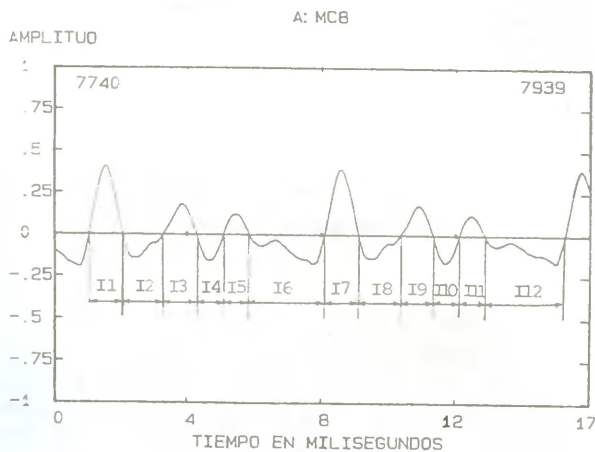


Fig. 4.

Y así en adelante.

Resulta claro que esta diferencia tendrá un mínimo definido cuando los pares de intervalos incluidos dentro de los símbolos de valor absoluto correspondan a dos periodos consecutivos. Podría objetarse con razón que este método no podrá detectar la verdadera duración de un periodo, cuando, aunque varíe la amplitud de cada uno de los picos que se dan dentro de un periodo, los intervalos entre cruces con cero son todos iguales. Este caso, que no es muy frecuente en la voz humana, se trata posteriormente. Debido al hecho de que la similitud entre periodos se debilita al comienzo y al final de un intervalo periódico; la detección de la periodicidad se producirá solamente cuando ésta ha alcanzado un nivel confiable de regularidad. Esto significa que para que sea posible captar periodicidad es necesario tener al menos dos periodos estables. Ordinariamente un

fonema periódico consta de varios periodos consecutivos. Por tanto cuando el algoritmo ha detectado periodicidad por primera vez, el algoritmo continúa el análisis de la siguiente manera:

La primera detección de periodicidad abarca el primero y el segundo periodos. La siguiente búsqueda comenzará a partir del segundo periodo para determinar si éste es seguido por un tercer periodo. En forma semejante, si existe un tercer periodo, este será el punto de partida para la búsqueda de un cuarto periodo, y así en adelante.

RIESGOS DE LA DETECCION DE PERIODICIDAD POR CRUCES CON CERO.

Aun cuando a primera vista podría pensarse que el método expuesto anteriormente trabajará adecuadamente, en la práctica hay ciertos factores que deben tomarse en cuenta si se quiere evitar errores catastróficos. La que sigue es una enumeración de los más importantes:

- a) Deben fijarse los intervalos máximo y mínimo para la búsqueda de periodicidad.
- b) Debe encontrarse una solución para prevenir que pequeñas variaciones en los cruces con cero causen grandes errores.
- c) El arracimamiento de cruces con cero que tiene lugar con ciertos fonemas, como por ejemplo con las vocales "i" y "e", debe ser tratado adecuadamente.

d) Debe encontrarse algún expediente para evitar que los cambios dramáticos de amplitud que ocurren en determinados fonemas, como ejemplo una vocal entre dos oclusivas, impida la detección de periodicidad.

a) INTERVALO MINIMO Y MAXIMO

En la locución normal existen variaciones en el tono de la voz que conllevan información, como es el caso de la elevación del tono de la voz al final de una pregunta, o dan al lenguaje hablado la variedad sin la cual éste sería insoportablemente monótono. Por otra parte se debe tomar en cuenta la gran variedad del tono medio de la voz que existe entre los diversos individuos. Para hacer frente a esta gran variedad de tonos es conveniente fijarse dos límites: un límite inferior o intervalo mínimo, a partir del cual comienza la búsqueda de periodicidad, y un límite superior, o intervalo máximo, para abandonar la búsqueda de periodicidad si éste es superado. Puesto que la búsqueda de periodicidad se realiza siempre entre dos periodos consecutivos, el límite inferior debe ser un poco mayor que dos periodos consecutivos de un locutor que posea el más alto tono medio de voz. De igual manera el límite superior debe ser un poco mayor que dos periodos consecutivos de un locutor que posea el más bajo tono medio de la voz. Teniendo en cuenta la gran diferencia que existe entre el tono medio de la voz para voces masculinas y femeninas, se consideró mejor usar dos clases de límites, unos para mujeres y otros para hombres. Puesto que el algoritmo trabaja con el

número de muestras por periodo o tono, los límites usados son los que siguen:

Límite inferior: NUMIN: 50 muestras para mujeres
104 muestras para hombres

Límite superior: NUMAX: 171 muestras para mujeres
330 muestras para hombres

b) PEQUEÑAS VARIACIONES EN LOS CRUCES CON CERO

Debido a que el algoritmo utiliza los intervalos entre cruces con cero para la detección de la periodicidad, una pequeña variación entre dos periodos como la que se indica en la fig. 5.

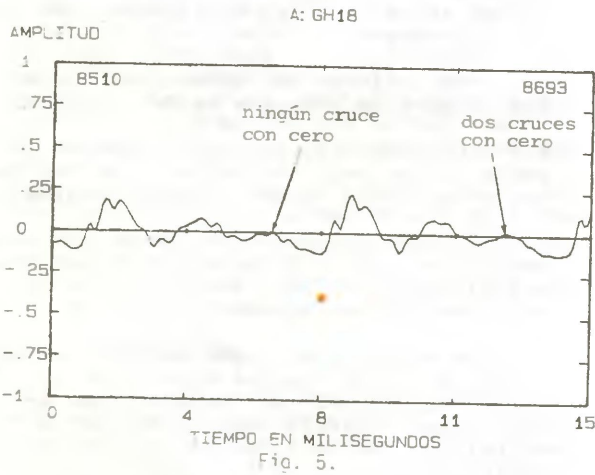


Fig. 5.

puede ser catastrófica para la búsqueda de la periodicidad. Estas variaciones son más de temerse precisamente en la vecindad de cero, donde a las naturales variaciones del lenguaje hablado deben añadirse los efectos del ruido aditivo que hace sentir más su presencia en la vecindad de cero. Para obviar este problema se consideró oportuno usar un eje horizontal con un corrimiento positivo, esto es, trazado sobre el nivel de cero. Con este propósito, se obtiene primero el máximo dentro del intervalo a ser analizado. Luego se corre el eje horizontal en una cantidad equivalente al 75 por ciento de este máximo. La razón para correr positivamente el eje es que el pico que generalmente marca el comienzo de cada periodo está más claramente definido en la región positiva de la onda. Este corrimiento positivo del eje horizontal puede ayudar también en la detección de la periodicidad en el caso, no muy frecuente en el lenguaje hablado, de una onda cuyos cruces con cero dentro de un periodo son equidistantes, ver figs. 6a y 6b.

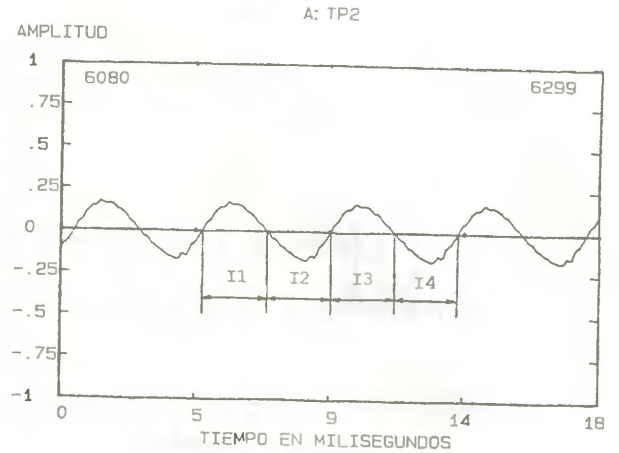


Fig. 6a.

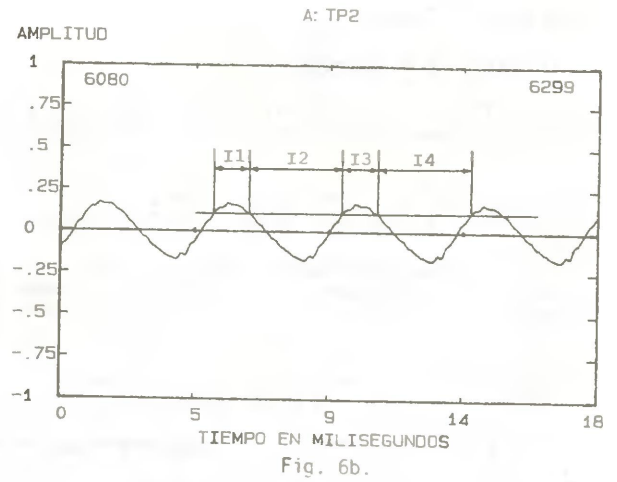


Fig. 6b.

Los cruces con cero de la figura 6a con el eje sin correr no permite una clara determinación del periodo, cosa que, en cambio, es muy fácil en la fig. 6b que tiene el eje corrido.

c) ARRACIMAMIENTO DE CRUCES CON CERO

Debido a la ausencia de un prefiltrado con un filtro pasabajos la onda conserva sus componentes de alta frecuencia, así por ejemplo no es infrecuente tener dos periodos consecutivos de la vocal "e" que tienen la forma indicada en las figs. 7a) y 7b).

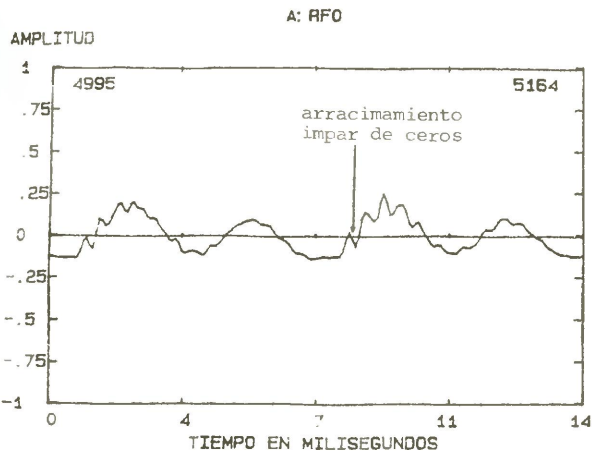


Fig. 7a.

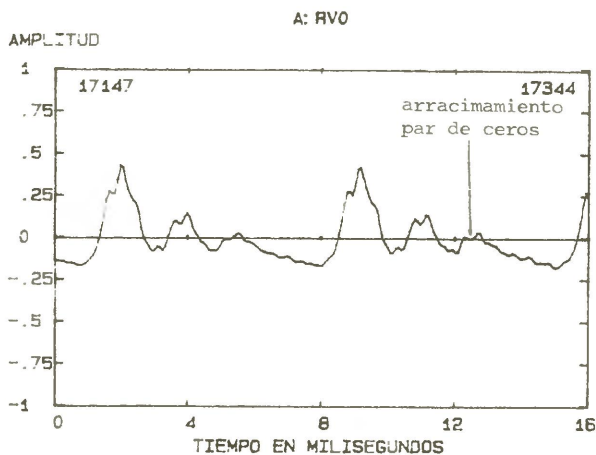


Fig. 7b.

Estas figuras presentan dos casos en los que el algoritmo para la búsqueda de la periodicidad fallará debido al arracimamiento de cruces con cero. El tratamiento dado a este problema fue diferente de acuerdo al hecho de que el arracimamiento de cruces con cero sea impar o par. Para un número impar de arracimamiento de cruces con cero la solución adoptada es la siguiente: ver fig. 8

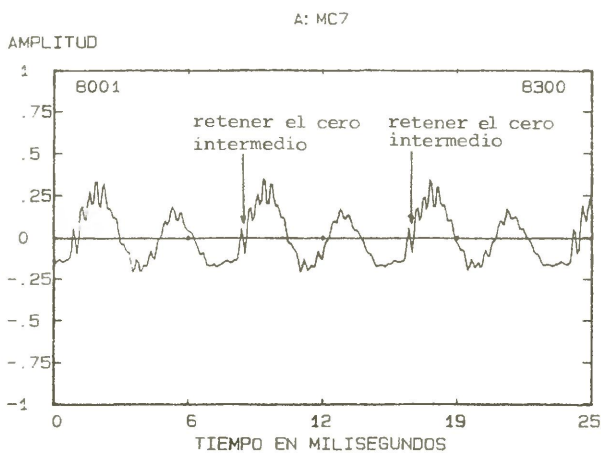


Fig. 8.

Se ignoran todos los cruces con cero arracimados excepto el central. Esto es, si hay tres cruces con cero arracimados, se toma en cuenta sólo el segundo, los demás son ignorados, si hay cinco cruces con cero arracimados sólo se toma en cuenta el tercero, los demás son ignorados, etc. Para cruces con cero arracimados de orden par la solución adoptada es la que sigue, ver fig. 9

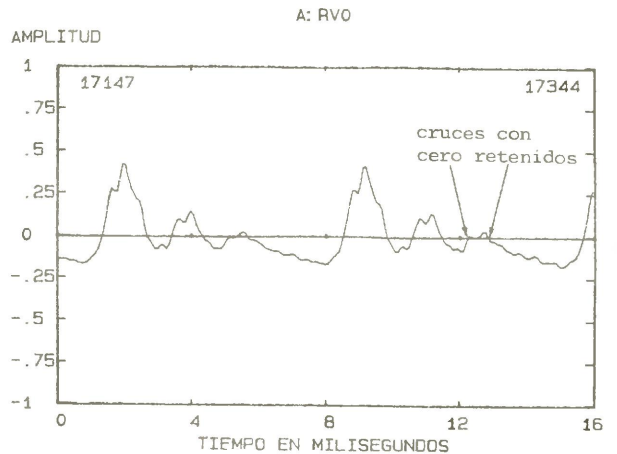


Fig. 9.

Cuando se trata de arracimamientos de orden par sólo se toman en cuenta el primero y el último cruces con cero, los demás se ignoran. Así por ejemplo si hay cuatro cruces con cero arracimados, sólo el primero y el cuarto se toman en cuenta, los demás se ignoran, etc. Cuando hay solamente dos cruces con cero arracimados se retienen los dos.

Queda por determinar cuando dos cruces con cero consecutivos deben considerarse arracimados. Como un valor tentativo, que en la práctica ha dado buenos resultados, los valores seleccionados se justifican como sigue:

para una señal muestreada a 12 KHz la separación entre muestras es $1/12000 = 83.33333$ microsegundos, en otras palabras este es el período de muestreo. Teniendo en cuenta esto, dos cruces con cero consecutivos se consideran arracimados si la separación entre los mismos es:

igual o menor que 2 períodos de muestreo para mujeres.
igual o menor que 3 períodos de muestreo para hombres.

De acuerdo a esto, el valor de la constante DIMIN que se utiliza con el propósito de detectar cruces con cero arracimados fue fijado en:

DIMIN = 2 para mujeres
3 para hombres

d) CAMBIOS RAPIDOS DE AMPLITUD

Cuando una vocal se ubica entre dos plosivas, o, en general, cuando hay una transición entre una vocal muy energética y un fonema débil o viceversa, hay un rápido cambio de energía que puede dificultar la detección de periodicidad. En este caso el desplazamiento del eje horizontal

sólo puede empeorar las cosas. Ver fig. 10.

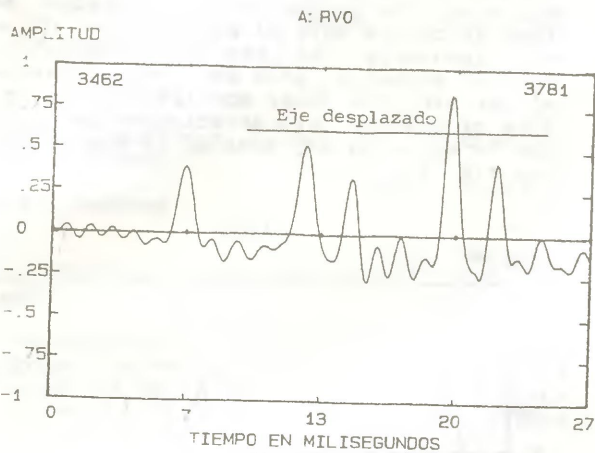


Fig. 10.

Para no perder periodicidad aun cuando haya cambios significativos de energía, se decidió usar un eje que siga aproximadamente las variaciones de energía de la onda sonora. Con este propósito la sección de la onda a ser analizada se trata de la siguiente manera:

En el intervalo NUMAX (330 muestras para hombres, 171 para mujeres) se buscan dos máximos, el primero correspondiente a los dos primeros tercios del segmento bajo análisis, y el segundo correspondiente al último tercio. El primer máximo se multiplica por el factor 0.75, y luego se traza un eje inclinado que pasa por este punto y tiene la misma pendiente que tendría la línea que pasaría por los dos máximos. Por lo tanto el eje con respecto al cual se determinan los "cruces con cero", debe aparecer como el de la fig. 11.

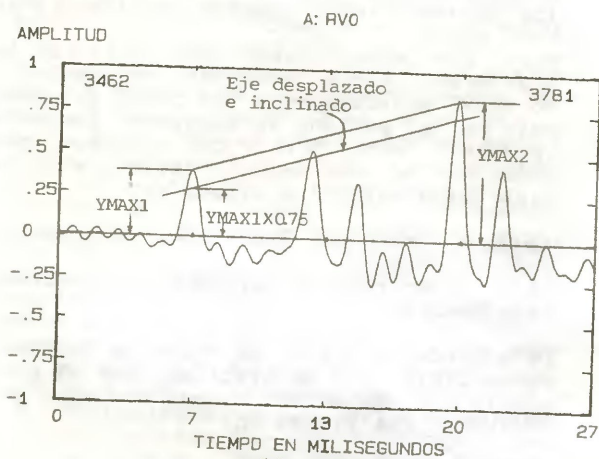


Fig. 11.

Con este procedimiento, que aumenta sólo ligeramente la complejidad del algoritmo, los errores en la detección de periodicidad se reducen en forma notable. La única desventaja de este eje inclinado es que la determinación de la duración de cada período puede contener un pequeño error, ya que para este propósito, ciertamente, un eje horizontal (no inclinado) a nivel

cero sería mejor.

Después de la exposición de las principales ideas que dirigen la realización del algoritmo, se presenta un diagrama de flujo del mismo en la fig. 12.

EXPLICACION DEL DIAGRAMA DE FLUJO

Como un primer paso el programa fija los valores de tres constantes: NUMAX, NUMIN, DIMIN que son diferentes para voces masculinas y femeninas, y cuyo significado se ha explicado anteriormente.

El proceso continúa con la entrada de datos. Los datos se toman de archivos de disco que contienen 32000 muestras. Esto significa 2 y 2/3 segundos de locución muestreados a 12 KHz con doce bits por muestra. Se tienen por tanto 4096 niveles de cuantización. Teniendo en cuenta que el muestreo es bipolar esto implica, como máximo, 2048 niveles de cuantización para las variaciones positivas y negativas de la onda. Una vez que los datos están disponibles el proceso se realiza en bloques de NUMAX muestras a la vez. Cada vez que se llega a una conclusión sobre la periodicidad o no del tramo bajo análisis, el proceso continúa de la siguiente manera: Si el segmento bajo análisis resulta aperiódico el comienzo del nuevo segmento a analizarse se inicia NUMIN/2 muestras hacia adelante.

Si el segmento bajo análisis resulta periódico, entonces el nuevo segmento a analizarse comienza unas pocas muestras antes del comienzo del segundo período. Después del ingreso de datos se determinan dos máximos: el primer máximo de los dos primeros tercios del bloque bajo análisis, el segundo máximo del último tercio del mismo bloque.

Si estos dos máximos: YMAX1 y YMAX2 son menores que 48 niveles de cuantización, entonces se evalúa los cruces con cero respecto de un eje horizontal trazado a 15 niveles de cuantización. Si es que no hay cruces con cero a este nivel, el segmento es una oclusión. Fijar el valor de D (discriminante de periodicidad) en 23 y seguir adelante con el proceso.

Si hay cruces con cero a este nivel, el segmento es aperiódico (posiblemente una fricativa), fijar el valor de D en 23 y continuar con el proceso.

Si YMAX1 o YMAX2 son mayores que 48 niveles de cuantización, obtener la pendiente de un eje que pase a través de estos dos máximos. Luego calcular los cruces con cero de la onda sonora respecto de un eje trazado con esta pendiente y que pase por un punto directamente debajo del primer máximo (YMAX1) e igual a 0.75 veces este valor. Los cruces con cero se determinan como sigue: primero se detectan todos los cruces con cero, luego se eliminan los cruces con cero arracimados, como se explicó anteriormente. El conjunto definitivo de cruces con cero se almacena en un arreglo [X2]. Usando este arreglo la subrutina ACTARR forma los arreglos [X] y [X1] de la siguiente manera:

[X] = X(1), X(2), X(3) [X1] = X1(1), X1(2), X1(3)

para dos periodos tentativos como los que se indican en la fig. 13.

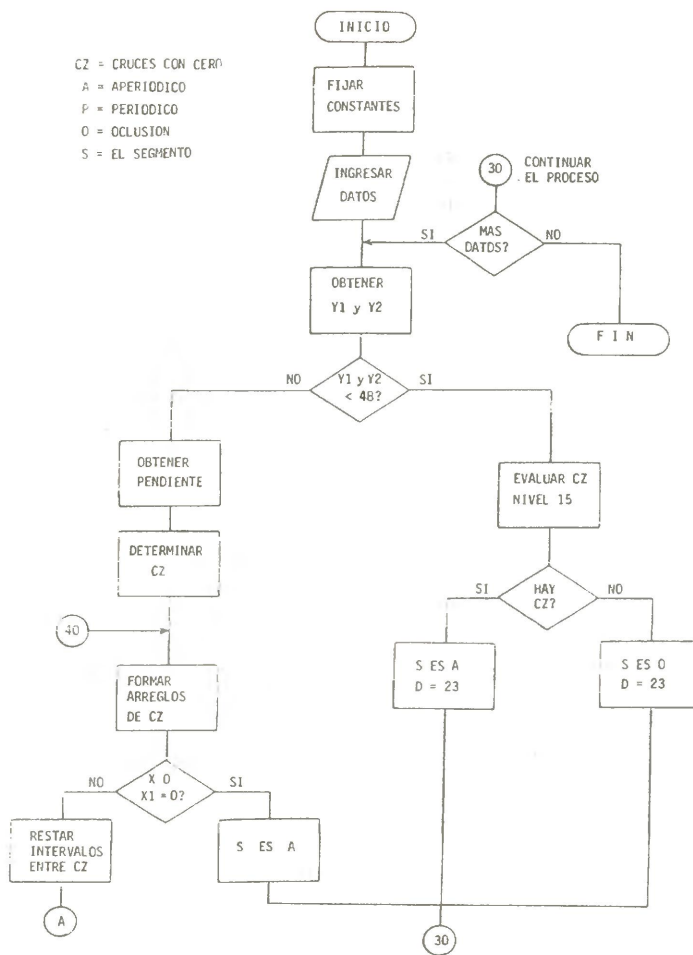


Fig. 12a. Diagrama de flujo del programa para detectar periodicidad.

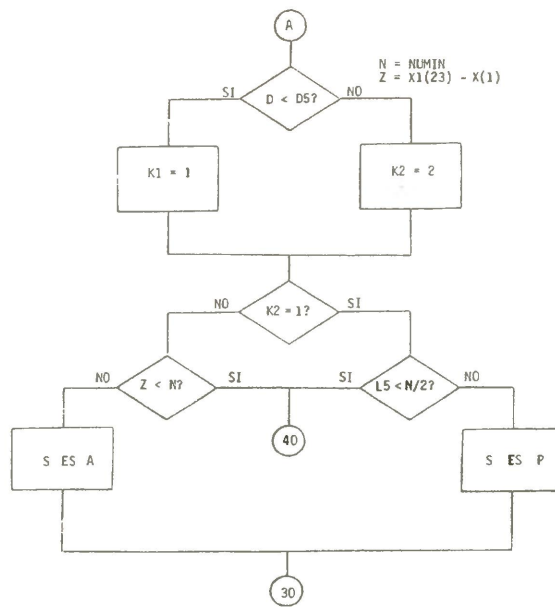


Fig. 12b. Diagrama de flujo del programa para detectar periodicidad.

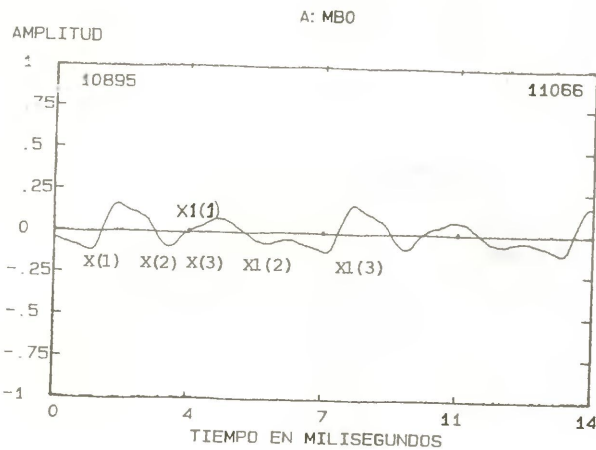


Fig. 13.

Para esta figura:

$$D = \begin{vmatrix} [X(2) - X(1)] & - & [X1(2) - X1(1)] \\ [X(3) - X(2)] & - & [X1(3) - X1(2)] \end{vmatrix} +$$

si esta posibilidad resulta fallida los arreglos [X] y [X1] toman la forma:

$$[X] = X(1), X(2), X(3), X(4), X(5)$$

$$[X1] = X1(1), X1(2), X1(3), X1(4), X1(5)$$

para dos periodos consecutivos como los de la fig. 14

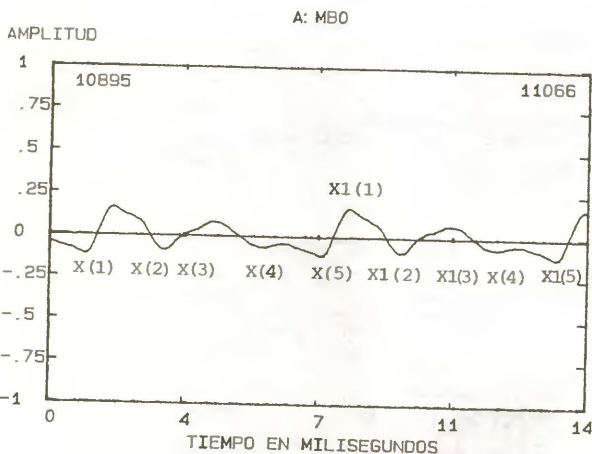


Fig. 14.

para esta figura:

$$D = \begin{vmatrix} [X(2) - X(1)] & - & [X1(2) - X1(1)] \\ [X(3) - X(2)] & - & [X1(3) - X1(2)] \\ [X(4) - X(3)] & - & [X1(4) - X1(3)] \\ [X(5) - X(4)] & - & [X1(5) - X1(4)] \end{vmatrix} +$$

y así en adelante hasta que el intervalo permitido para dos periodos consecutivos resulte demasiado grande.

Antes de cada estructuración de los arreglos [X] y [X1] los correspondientes elementos se fijan en cero. Por tanto el hecho de que algunos de los elementos de estos arreglos tengan el valor cero es una clara indicación de la falta de periodicidad. Específicamente si $X(Z3)$ o $X1(Z3) = 0$. Donde $Z3 = 3, 5, 7$, etc.. Esto significa que uno o ambos arreglos

no se han llenado con cruces con cero. (Estando los cruces con cero indicados por la muestra anterior al cambio de signo de la onda, nunca pueden ser menores que uno). Por lo tanto el segmento debe ser aperiódico. El valor de D se fija en 23 y el proceso continúa con la búsqueda de periodicidad.

Si $X(Z3)$ Y $X1(Z3)$ son diferentes de cero entonces el algoritmo resta término a término los correspondientes intervalos de los supuestos dos periodos consecutivos. Esta substracción da como resultado el valor de D que es el discriminante de periodicidad. En efecto si $D < D5 = 5.9$ el segmento puede ser periódico si el intervalo entre las muestras ubicadas en los puntos representados por los elementos $X(1)$ y $X(Z3)$ tiene un valor apropiado.

Si $D < D5$ entonces hacer $K2 = 1$, de otra manera fijar $K2 = 2$.

Si $K2 = 1$ entonces ver si $L5$ (valor promedio de los dos supuestos periodos consecutivos) es menor que $NUMIN/2$ (duración de un periodo para la voz masculina mas atiplada). Si es que es así se deberá incrementar el número de intervalos entre cruces con cero por periodo. Si $L5 > NUMIN/2$ el segmento es periódico. Continuar el proceso comenzando algunas muestras antes del comienzo del segundo periodo.

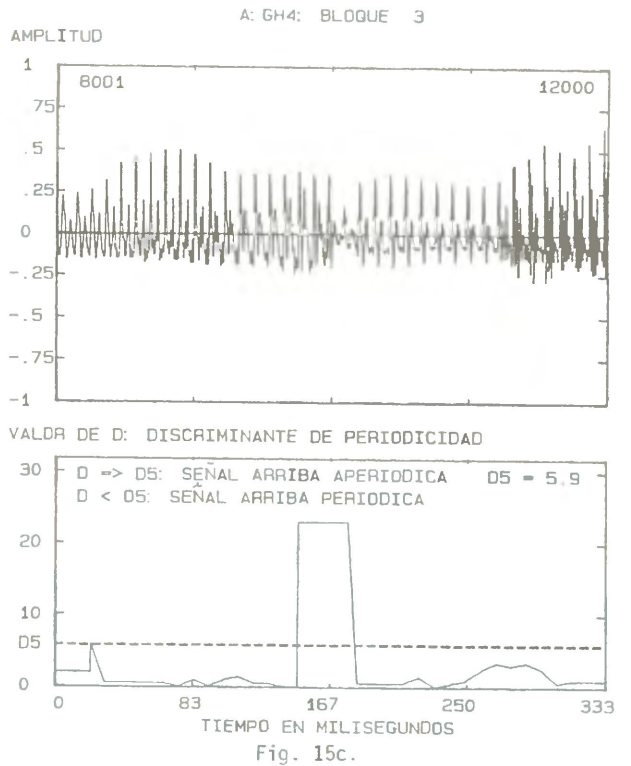
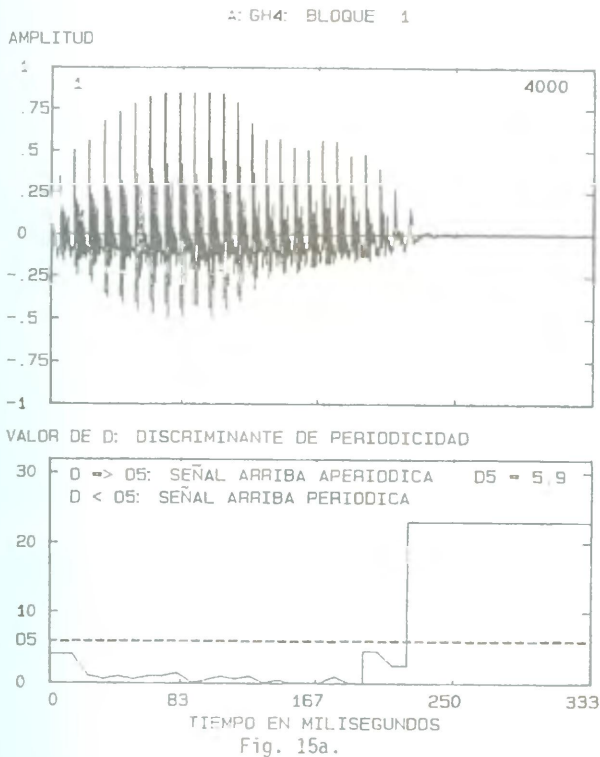
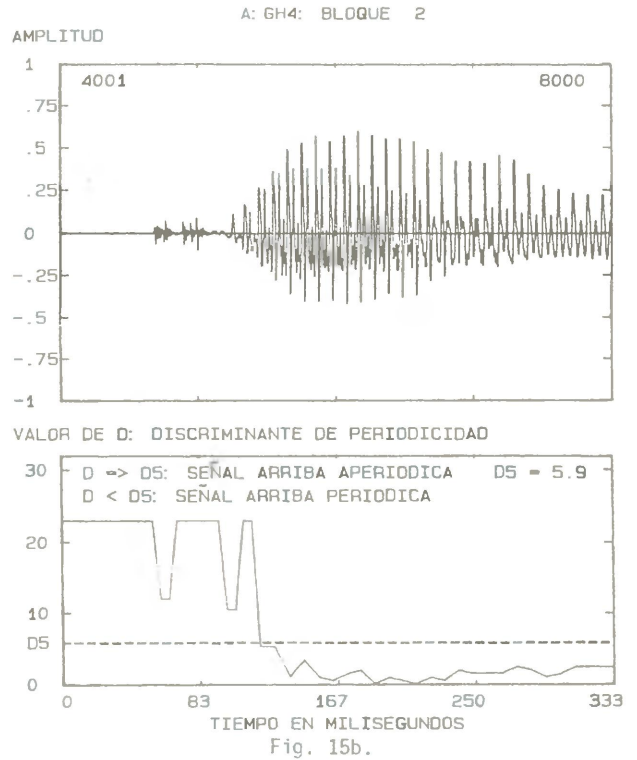
Si $K2$ diferente de 1 entonces comprobar si $X1(Z3) - X(1) < NUMIN$ (intervalo de dos periodos consecutivos de una voz masculina atiplada), si es así, aumentar el número de intervalos entre cruces con cero por periodo.

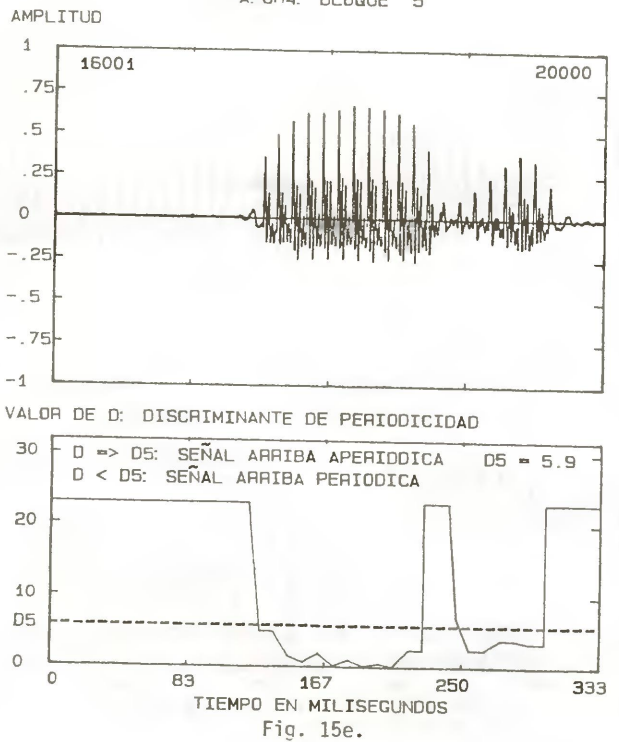
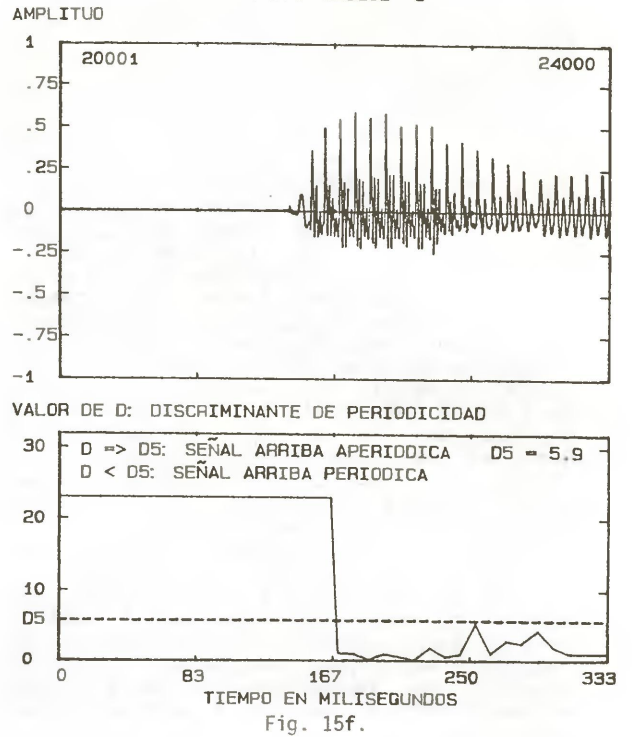
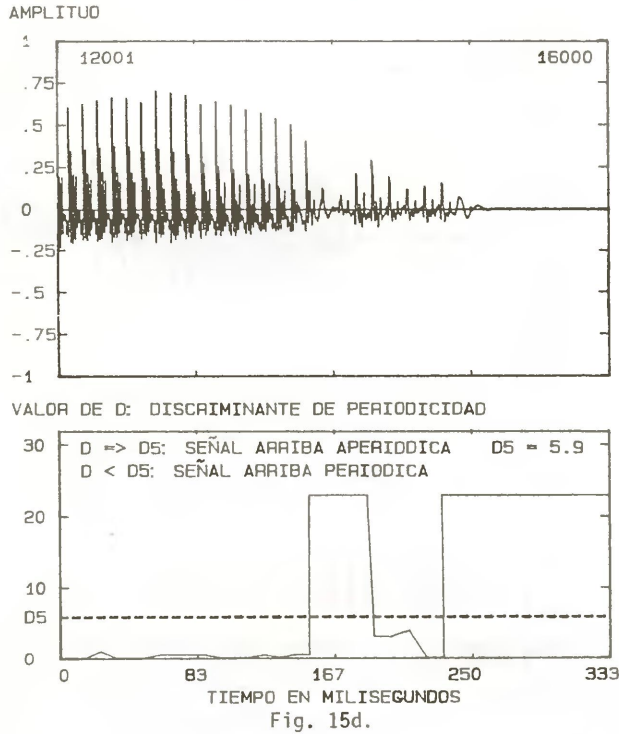
Si $K2 = 2$ y el intervalo $X1(Z3) - X(1) \Rightarrow NUMIN$ entonces el segmento aperiódico. Respetar el valor de D obtenido y continuar con el proceso.

IV. EVALUACION DEL ALGORITMO

Para una verdadera evaluación del algoritmo será deseable compararlo con otros algoritmos y comprobar su desempeño periodo por periodo. Teniendo en cuenta que la verdadera finalidad de este algoritmo no es una determinación precisa de la secuencia de valores de cada periodo, como sería el caso para síntesis o codificación de locución, sino la segmentación de la palabra hablada, aquello no se consideró necesario. Así pues por razones de brevedad, y teniendo en cuenta que una imagen vale mas que mil palabras, se ha utilizado una representación gráfica para valorar los resultados obtenidos. Las figuras que se presentan a continuación con este objeto, presentan dos curvas. La primera curva en la parte superior del gráfico presenta la onda de la locución, el número de muestras para cada figura puede obtenerse restando el número en la esquina superior izquierda del gráfico del número en la esquina superior derecha y sumando 1. Así por ejemplo si los dos números mencionados arriba son 1 y 2000, el numero de muestras será $2000 - 1 + 1 = 2000$. Siendo la frecuencia de muestreo 12 KHz cada 1200 muestras constituye una décima de segundo. La curva en la parte inferior del gráfico presenta el valor de D, el discriminante de periodicidad. La línea segmentada trazada a un nivel $D5 = 5.9$ establece el nivel crítico de periodicidad. Esto es, un valor de D

menor que D_5 implica que la onda en la parte correspondiente superior del gráfico es periódica. Un valor de D igual o mayor que D_5 indica que la curva en la parte superior del gráfico es aperiódica. Del análisis visual de las figuras se puede establecer fácilmente donde el algoritmo trabaja correctamente y donde falla. Para dar los resultados en porcentaje, sería tal vez aconsejable contar el número de periodos donde el algoritmo falla y dividir este número para el número total de periodos en el mismo intervalo de tiempo. Conocidas las partes donde el algoritmo falla, éstas pueden ser analizadas con mayor detenimiento para mejorarlo. Finalmente debe tenerse en cuenta que la última sección de la curva inferior no entrega resultados confiables. En efecto el algoritmo antes de analizar un segmento, primero establece si las muestras disponibles son 330 o más para hombres, y 171 o más para mujeres. Si esta condición no es satisfecha el algoritmo simplemente ignora esta última sección e interpola la curva del valor de D con el último valor previo. La figura 15 muestra el desempeño del algoritmo para detectar periodicidad. Esta figura presenta 24000 muestras de una voz masculina en bloques de 4000 muestras. De la observación de la figura se desprende un desempeño aceptable del algoritmo.





CONCLUSION

Del examen de las voces analizadas es posible concluir que el algoritmo trabaja a veces en forma excelente, y otras veces tiene un número variable de errores. Entre éstos los más relevantes son aquellos en los cuales el algoritmo pierde periodicidad cuando la sección analizada es claramente periódica, y aquellos en los cuales el algoritmo encuentra periodicidad cuando esta es inexistente. Afortunadamente estos errores, siendo de corta duración, podrán seguramente ser eliminados por el contexto. Esto se analizará en un siguiente artículo que tratará de la segmentación automática del lenguaje hablado. Puede también mencionarse que la fijación de las cantidades que guían las búsqueda de periodicidad, como: NUMAX, NUMIN, DIMIN y la pendiente del eje para determinar los cruces con cero se hizo en forma heurística. Es de esperarse que la fijación de estos valores a través de un análisis estadístico riguroso de la locución mejorará los resultados obtenidos.

Referencias

- [1] Speech Analysis, R. W. Schafer and J. D. Markel, Eds., IEEE Press, 1978, pp. 145-215.
- [2] Wolfgang Hess, Pitch determination of speech signals, Springer Verlag, 1983.
- [3] B. Gold and L. R. Rabiner, "Parallel Processing Techniques for Estimating Pitch Periods of Speech in Time Domain," J. Acoust. Soc. Am. Vol. 46, pp. 442-448, Aug. 1969.
- [4] C. S. Chen and Jing Yuan, "A Robust Pitch Boundary Detector," ICASSP 88, New York, S8.17, pp. 366-369.
- [5] C. S. Wu, V. Nguyen, V. Goncharoff, W. Kushner, J. Damoulakis, "Adaptive Pitch Detection Algorithm for Noisy Signals," ICASSP 89, Glasgow, S11.5, pp. 576-579.
- [6] E. Barnard, R. A. Cole, M. P. Veal, and F. A. Alleva, "Pitch Detection with a Neural-Net Classifier," IEEE Trans. Signal Processing, Vol. 39, pp. 298-307, Feb. 1991.
- [7] D. A. Krubsack and J. R. Niederjohn, "An Autocorrelation Pitch Detector and Voicing Decision with Confidence Measures Developed for Noise-Corrupted Speech," IEEE Trans. Signal Processing, Vol. 39, pp. 319-329, Feb. 1991.
- [8] T. Funada, T. Suzuki and L. Yu, "A Pitch Extraction Method using a Bank of Bandpass Filter-Pairs," Speech Communication, Vol. 9, pp. 203-216, June 1990.
- [9] Y. Medan, E. Yair, and D. Chazan, "Superresolution Pitch Determination of Speech Signals," IEEE Trans. Signal Processing, Vol. 39, pp. 40-48, Jan. 1991.
- [10] E. Chilton and B. G. Evans, "The Spectral Autocorrelation Applied to the Linear Prediction Residual of Speech for Robust Pitch Detection," ICAASP 88, New York, S8.15, pp. 358-361.
- [11] S. Seneff, "Real-Time Harmonic Pitch detector," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-26, pp. 358-365, Aug. 1978.
- [12] A. M. Noll, "Cepstrum Pitch Determination," J. Acoust. Soc. Am., Vol. 41, pp. 293-309, Feb. 1967.
- [13] J. D. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation," IEEE Trans. Audio Electroacoustics, Vol. AU-20, pp. 367-377, Dec. 1972.

HIDALGO, GUALBERTO

Ingeniero en Electrónica y Telecomunicaciones graduado en la Escuela Politécnica Nacional en el año de 1974. Obtuvo el título de Master en Ingeniería de Comunicaciones en el Imperial College de Londres, Inglaterra, en 1979. Actualmente se desempeña como profesor a tiempo completo en la Escuela Politécnica Nacional y es estudiante externo de la Universidad de

Londres. Dirige el PROYECTO CONUEP 88-01, RECONOCIMIENTO DE FONEMAS POR COMPUTADOR.

PEREZ, TANIA

Ingeniero en Electrónica y Telecomunicaciones graduada en el Instituto Bónch Bruyevich, Leningrado, 1977. Actualmente se desempeña como profesora principal a tiempo completo en la Escuela Politécnica Nacional y realiza estudios de postgrado en Computación e Informática en la misma Institución. Participa en el PROYECTO CONUEP 88-01, RECONOCIMIENTO DE FONEMAS POR COMPUTADOR.