TECNIACUSTICA
BARCELONA • 1996

Modelado de la señal en reconocimiento de habla ruidosa

*P. Ejarque, J. Hernando, J.B. Mario, G. Hernández**

Departamento de Teoría de Seal y Comunicaciones. Universitat Politcénica de Catalunya.
javier@gps.tsc.upc.es

ABSTRACT
Conventional modelling techniques of speech suffer a very big performance degradation in adverse noisy environments. So, it is necessary to research for more robust representations of speech signal. This paper presents new models that have succeeded in adverse environments. They are hybrid models of the classical parametrizations techniques used so far that have demonstrated being very useful in order to obtain good results in different noisy environments. In order to prove the their performance we have used white and machine noise in our experiments.

INTRODUCCION

Los sistemas actuales de reconocimiento, que dan buenos resultados en ambientes sin ruido, se degradan rápidamente en presencia de señales interferentes ajenas a la señal de voz como los producidos por máquinas o entornos adversos [1]. Este artículo pretende mostrar diferentes métodos para mejorar los resultados de reconocimiento empleando técnicas diversas en la parametrización en dos entornos diferentes, ruido blanco y ruido de máquina.

La etapa de parametrización consiste en estimar la envolvente espectral de la señal, que es la que aporta la información necesaria para el reconocimiento ya que su evolución temporal parece caracterizar las diferentes realizaciones acústicas. En nuestro caso buscamos representaciones de la señal robustas al ruido pero que, a la vez, den buenos resultados en condiciones ambientales poco ruidosas.

En este trabajo se pretende comparar las técnicas clásicas de parametrización LPC-cepstrum (LPCC) y mel-cepstrum (MFCC) con métodos híbridos presentando una visión unificadora del proceso de modelado de la señal mediante un esquema único que engloba todas las paramatrizaciones expuestas en este artículo. Para la evaluación de los diferentes modelados se han usado seales contaminadas con ruido blanco y con ruido de máquina para diferentes relaciones señal a ruido.

El orden de este artículo ser como sigue. El apartado 2 se dedica a las parametrizaciones tratando brevemente las parametrizaciones clásicas y más específicamente las nuevas parametrizaciones propuestas en [2] . En el apartado 3 se detallan las diversas pruebas realizadas y se presentan los diferentes resultados obtenidos con las diferentes parametrizaciones para diferentes relaciones señal a ruido y para cada uno de los dos ruidos considerados. El apartado 4 se dedica a las conclusiones que de estos resultados hemos podido extraer.

METODOS DE PARAMETRIZACIN.

Métodos clásicos: LPCC y MC.

La técnica LPCC [3] equivale a un modelado autorregresivo de la señal de voz relacionado con los modelos de producción del habla por el tracto vocal humano. Con él se pretende obtener una envolvente espectral de la señal de voz mediante la respuesta frecuencial de un filtro todo polos. Esta envolvente se representa mediante los parámetros cepstrales que se obtienen por recursión de los parámetros LPC. Los parámetros LPC se obtienen aplicando el algoritmo de Levinson - Durbin a la autocorrelación de la señal de voz.

*Este trabajo ha sido financiado por los proyectos TIC95-0884-C04-02 y TIC 95-1022-C05-03

El primer proceso necesario para el cálculo de los coeficentes mel-cepstrum (MFCC) [4] es realizar el módulo al cuadrado de la transformada discreta de Fourier de la trama de voz enventanada. Posteriormente, se estima la energía por bandas multiplicando en frecuencia por un banco de filtros triangulares espaciados en una escala perceptual. Para finalizar, se calcula la transformada discreta inversa para obtener los parámetros MFCC. Este modelo intenta imitar la respuesta de la cóclea humana a los estímulos que produce la voz, es decir, intenta emular el tratamiento que el oído humano da a la señal de voz. Es ampliamente aceptado el uso de 20 filtros de forma triangular en el banco de filtros.

Métodos híbridos: LMC y MLC.

Entre los posibles modelos híbridos los autores han desarrollado las técnicas LPC-mel-cepstrum (LMC) y mel-LPC-cepstrum (MLC) [2] y ha estudiado su comportamiento en ambientes que precisan robustez al ruido. Estos modelos se detallan a continuación.

La técnica LMC consiste en aplicar la técnica mel-cepstrum sobre el espectro LPC de la señal de voz. A partir del espectro se realiza un cálculo integrador de la potencia de señal en cada banda de frecuencia. A los valores obtenidos se les aplica la técnica mel-cepstrum para obtener los parámetros LMC. Con ello se pretende con esta parametrización es simular la concatenación de los procesos de articulación y audición de la voz por parte del tracto vocal y el oído humanos.

La técnica MLC realiza el proceso inverso al método LMC aplicando predicción lineal a la estimación del espectro obtenido a partir de un banco de filtros espaciados en la escala perceptual mel. La DFT inversa de la salida del banco de filtros puede interpretarse como autocorrelación y aplicarse a la entrada del algoritmo de Levinson-Durbin.

Esquema unificado de parametrización.

Tal y como se han descrito los métodos de parametrización anteriores, se puede llegar a un esquema general que engloba todas las parametrizaciones de manera que se llega al modelado elegido según el camino que sigamos a través del esquema general de parametrización que se detalla en la Figura 1.

En este esquema se puede observar como las técnicas MLC y LMC son hibridaciones de las técnicas LPCC y MFCC y aprovechan los diferentes procesos utilizados por estos dos parametrización para conseguir modelados de la señal de voz que superan en robustez a sus predecesoras.

PRUEBAS Y RESULTADOS.

Base de datos y sistema de reconocimiento.

Para la realización de las pruebas de reconocimiento se ha utilizado la base de datos TI [5] que contiene realizaciones de los once dígitos en inglés (de 1 a 9, oh y zero) pronunciados por 111 locutores adultos para las señales de entrenamiento y 113 para las señales de test. Cada locutor ha pronunciado dos realizaciones de cada dígito.

Las pruebas se han realizado con dígitos aislados y son independientes de locutor. Se ha utilizado el sistema de reconocimiento HTK v1.5 [6] basado en modelos ocultos de Markov (HMM) de densidad cotinua. Cada dígito se ha caracterizado con un modelo de Markov de 10 estados de izquierda a derecha (5 estados en el caso del silencio) con matriz de covarianza diagonal.

Los modelos se han entrenado con las señales de entrenamiento de la base de datos limpias mientras que el reconocimiento se ha efectuado a partir de las señales de test de la base de datos ensuciadas con ruido para relaciones señal a ruido de 0, 10 y 20 dB y con las señales lim-

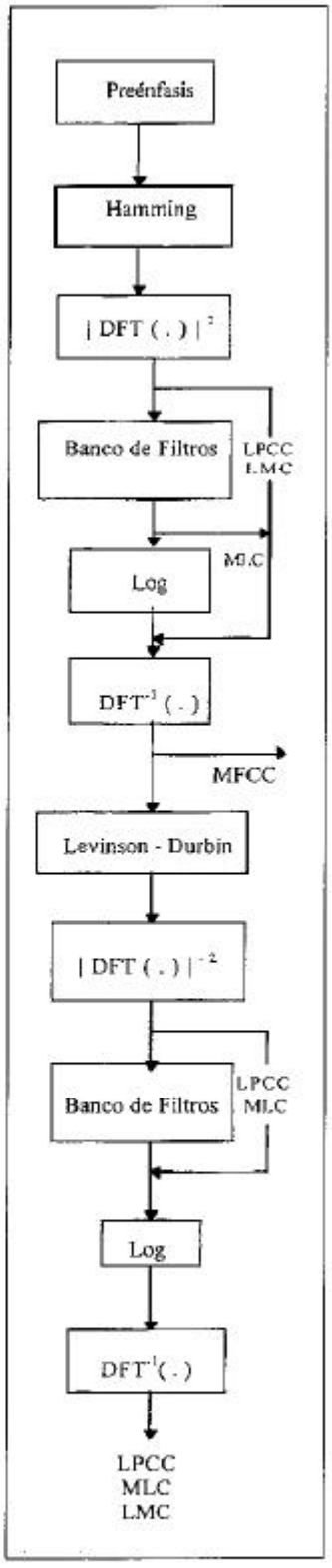

Figura 1. Esquema general de parametrización.

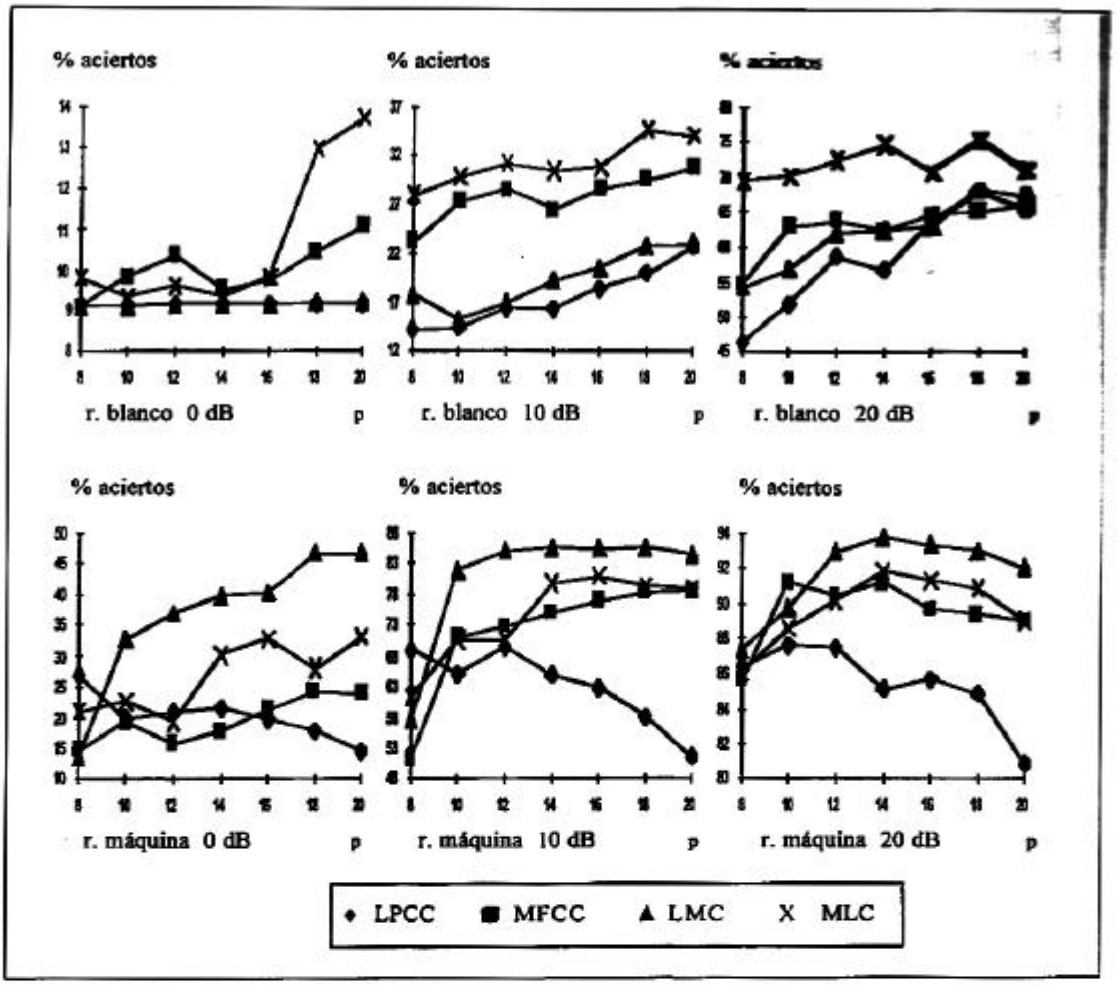

Figura 2. Porcentajes de reconocimiento para las diferentes parametrizaciones. Arriba, con ruido blanco. Abajo con ruido de máquina. De izquierda a derecha, relaciones señal a ruido de 0, 10 y 20 dB.

pias. Se han utilizado dos ruidos diferentes, ruido blanco generado aleatoriamente y ruido de máquina con los cuales se han ensuciado las señales de test con las que se han obtenido los resultados expuestos en el siguiente apartado.

No se ha realizado preénfasis [2]. La señal se ha divido en tramas de 30 ms. con 10 ms. de solapamiento. Sólo se han utilizado los parámetros estáticos, ni parámetros dinámicos ni energía. Se ha realizado un barrido desde los 8 hasta los 20 parámetros para cada parametrización, cada ruido y cada relación señal a ruido. Para las parametrizaciones LPCC, MLC y LMC se han hecho coincidir el orden del modelo de predicción y el número de coeficientes cepstrales por haberse demostrado empíricamente que da mejores resultados que la utilización de valores dispares [5].

Resultados experimentales

En este apartado se exponen los resultados obtenidos en las pruebas realizadas y que se reflejan en las gráficas de la figura 2.

Para ruido blanco, la parametrización MLC es la que proporciona mejores resultados para relaciones señal a ruido de 10 y 20 dB y para relación señal a ruido de 0 dB con 18 y 20 parámetros. Con esta parametrización se obtienen mejores resultados máximos para las tres relaciones señal a ruido: 13.72 % de reconocimiento para 0 dB y 20 parámetros, 34.69 % para 10 dB y 18 parámetros y 75.29 % para 20 dB y 18 parámetros. El modelado MFCC es el que proporciona mejores resultados para relación señal a ruido de 0 dB y pocos coeficientes y, en general, se sitúa por encima de los LMC y LPCC y por debajo del MLC. La técnica LMC suele situarse entre las dos parametrizaciones clásicas y la técnica

LPCC es la que obtiene menores tasas de reconocimiento excepto con poco ruido y un elevado número de parámetros.

Para ruido de máquina, la parametrización LMC obtiene los mejores resultados para relación señal a ruido de 0 dB, 10 dB y 20 dB a partir de 12 coeficientes. Con esta parametrización se obtienen los porcentajes de reconocimiento más altos: para 0 dB, 46.72 % con 20 parámetros; para 10 dB, 85.63 % con 14 parámetros y para 20 dB, 93.84 % también con 14 parámetros. El modelado MLC se comporta mejor, en general, que las dos parametrizaciones clásicas para relaciones señal a ruido bajas y para 20 dB de relación señal a ruido con 14 o ms parámetros. La técnica MFCC se comporta mejor que la LPCC utilizando un número suficiente de parámetros.

Para seales limpias la técnica LPCC obtiene un resultado del 96.26 % con 18 parámetros, la MC del 97.1 % con 8 parámetros y la LMC del 96.26 % con 14 parámetros. El modelado MLC supera a todas las anteriores parametrizaciones con un 97.26 % de reconocimiento con 12 parámetros.

CONCLUSIONES

De los resultados anteriormente expuestos se derivan las siguientes conclusiones:

- Dentro de los mátodos clsicos de parametrización la técnica MFCC demuestra una mayor robustez frente al ruido que la técnica LPCC ya que la supera en la mayora de los casos para relaciones señal a ruido bajas. Esto ocurre siempre en el caso de ruido blanco y para un número alto de parámetros en el caso de ruido de máquina.

- Para ruido blanco, la técnica MLC es la que presenta mejores prestaciones para todas las relaciones señal a ruido, aunque para 0 dB de relación señal a ruido es necesario utilizar un número alto de parámetros. La técnica LMC obtiene unos resultados que la sitúan entre las dos clásicas.

- Para ruido de máquina, el modelado LMC de la señal ofrece los mejores resultados para relaciones 0, 10 y 20 dB de relación señal a ruido, aunque para 20 dB hay que usar un número elevado de parámetros. La parametrización MLC también supera a las dos clásicas en estas situaciones cuando se utilza más de 14 coeficientes.

- Para señales limpias el método MLC el que proporciona mejores resultados.

REFERENCIAS.

[1] B. H. Juang, *Speech Recognition in Adverse Environments, Computer Speech and Language*, vol. 5, 1991, pp. 275 - 294.

[2] J. Mndez, J. Hernando, C. Nadeu, F. Vallverd *Esquema Unificado de Parametrización de la Seal de Voz en Reconocimiento del Habla*, Proc. URSI95, Valladolid, Septiembre 1995, pp. 97-100.

[3] J. Makhoul, *Proceedings of the IEEE*, vol. 63, n 4, Abril 1975, pp. 561 - 580.

[4] S. B. Davis, P. Mermelstein, IEEE Trans. ASSP, vol. 28, 1980, pp. 357 - 366.

[5] R. G. Leonard, Proc. ICASSP84, Marzo 1984, pp. 42.11.1 - 4.

[6] Cambridge University Engineering Departament Speech Group and Entropic Research Laboratories INC. HTK - Hidden Markov Model Toolkit v1.5 Diciembre 1993.