

EDITORIAL

Lector, otra vez tienes entre tus manos un ejemplar de Buran y con este ya van 19, que no son pocos. Habrás notado que pesa más que de costumbre, es porque estamos de celebración: hemos alcanzado el decenio. Pero como dicen que más importante que llegar es saber llegar, queremos aprovechar para disculparnos por la supresión de la edición de Diciembre pasado, que a buen seguro esperabas. Debemos achacar este hecho a la prioridad que para nosotros representa la calidad de esta publicación sobre otros factores.

En esta sociedad tan competitiva y a la vez pragmática parece imposible que alguien edite Buran desinteresadamente. Más aún si quienes lo hacemos somos estudiantes. Sin embargo, sólo nos mueve el afán de conocer y divulgar las tecnologías de la telecomunicación. Por ello los artículos de este número tratan de abarcar un espectro lo más amplio posible dentro del marco anterior, tocando temas sobre electrónica, diseño e implantación de redes, comunicaciones digitales, procesado de la señal, sistemas ópticos, satélites... Esperamos que sean de tu agrado.

Como curiosidad para aquellos que nos sois fieles desde el principio, hemos intentado que todos aquellos pioneros que ya colaboraron con Buran en el primer número repitieran la experiencia en este. Nos gustaría agradecer la atención prestada y el trabajo a todos los repetidores y animar para próximas ediciones a aquellos que, por diferentes causas, no han llegado a tiempo.

Aunque el contenido de Buran es fundamentalmente de cariz técnico, no quisiéramos desaprovechar la oportunidad que nos brindan estas líneas para hacer uso de la libertad de expresión que nuestra independencia nos proporciona. Por ello, y hoy más que nunca, deseamos expresar nuestro completo rechazo a toda acción bélica injustificada. La vida es el bien más preciado que poseemos y debemos luchar por preservarlo.

Eduard Calvo Page
Vicepresidente de la Rama
Coordinador de Buran

COORDINACIÓN BARCELONA

Eduard Calvo Page

EDICIÓN BARCELONA

Felipe Calderero Patino
Eduard Calvo Page
José Antonio Casas
Jordi Gilabert Berenguer
Jose Luis Hernández
Ignasi Pérez Riverola
José A. López Salcedo
Alfredo C. López Salcedo
José Cástor Vallés Martínez
Jaume Ribot Caner
Rosa Mª Tarruella Castells

REVISIÓN

Carles Gómez Montenegro
José A. López Salcedo
Miguel Angel Sastre
Jose Cástor Vallés Martínez

DISEÑO PORTADA

Alfredo C. López Salcedo

AGRADECIMIENTOS

Il. Dir. Juan A. Fernández Rubio,
Ángel Cardama, Jose A. Delgado-Peníñ, Jorge Luis
Sánchez-Ponz y a los puntos de distribución en la
UPC: Abacus, CPET, CPDA y Kiosk Campus Nord.

We would also like to thank Ms. Laura Durrett
(IEEE Student Services Manager), and IEEE
International for their helpful support, encouragement
and financial funding for distributing Buran across
south american Region 9 IEEE Student Branches.

IMPRESIÓN

RET, s.a.l.

FOTOMECAÑICA

Sistemes d'Edició

DEPÓSITO LEGAL

B-19.950-96

SIIUPC, 2003 (6834)

La organización se reserva el derecho de publicar los artículos. La opinión expresada en los artículos no tiene por qué coincidir con la de la organización.

Agradecemos las colaboraciones hechas desinteresadamente, y a causa de la falta de espacio, pedimos disculpas a todas aquellas personas a las cuales no se les ha publicado su colaboración. Esperamos que en un próximo número tengan cabida.



UNA DÉCADA DE TELEDETECCIÓN EN EL GRUPO DE INGENIERÍA ELECTROMAGNÉTICA Y FOTÓNICA

*Departamento de Teoría de la Señal y Comunicaciones,
Universitat Politècnica de Catalunya*

0. INTRODUCCIÓN

El Grupo de Ingeniería Electromagnética y Fotónica (EEF) pertenece al Departamento de Teoría de la Señal y Comunicaciones de la Universitat Politècnica de Catalunya. Dicho grupo está formado por 35 profesores, unos 15 estudiantes de doctorado, y unos 60 estudiantes que realizan su proyecto final de carrera. Las líneas de investigación se centran en el diseño y análisis de antenas, circuitos de microondas, teledetección (activa o radar, pasiva o radiometría, y óptica) y comunicaciones ópticas. El grupo EEF dispone de las siguientes instalaciones:

- un laboratorio de radiofrecuencia, completamente equipado hasta 75 GHz,
- una cámara anecoica de 10 m x 7.5 m x 7.5 m para la medida de antenas en campo próximo hasta 40 GHz,
- una cámara anecoica media con un campo compacto para la medida de sección recta radar,
- un laboratorio de óptica integrada,
- una laboratorio de software con una plataforma UNIX compuesta por un servidor multiprocesador RISC, 15 estaciones de trabajo UNIX/LINUX y varios terminales X; y una plataforma MS-Windows con tres servidores Windows 2000 y más de 100 clientes Windows. El backbone de comunicaciones está basado en una red Gigabit Ethernet, con acceso al backbone del Campus por medio de enlaces ATM redundantes.
- Además se dispone de acceso a las capacidades de supercomputación del propio campus (CESCA y CEPBA).

Las actividades de investigación del grupo EEF se llevan a cabo en cooperación con la industria e instituciones nacionales y europeas, de las que se obtiene la financiación. Las actividades de teledetección del grupo EEF se pueden dividir en tres grandes líneas: pasiva en microondas (radiometría y utilizando señales de oportunidad GPS reflejadas), activa en microondas (radar de apertura real y sintética) y óptica, cada una de las cuales cubre tanto desarrollos teóricos como experimentales [1]. A continuación, un miembro de cada una de estas líneas de investigación, describe brevemente las principales actividades en los últimos 10 años.

1. TELEDETECCIÓN PASIVA EN MICROONDAS

por A. Camps y Mercé Vall-llossera

En 1993 se inicia una serie de proyectos financiados tanto por la Agencia Europea del Espacio (ESA), como por la industria nacional y europea y proyectos de investigación de la CICYT en el campo de la teledetección pasiva en microondas, y más concretamente, en radiometría de microondas, o medida de la potencia de ruido térmico captada por una antena a una frecuencia y polarización determinadas.

1.1. Radiometría por síntesis de apertura

La investigación dedicada a la radiometría por síntesis de apertura ha venido principalmente asociada a la Misión Earth Explorer de la ESA llamada SMOS/MIRAS (Soil Moisture and Ocean Salinity/Microwave Imaging Radiometer by Aperture Síntesis, Figura 1a) [2]. En esta misión se aplican por primera vez estas técnicas, heredadas de la radioastronomía (Figura 1b) [3], a la observación de la Tierra. El objetivo es conseguir una resolución angular relativamente alta a partir de una agrupación de antenas de pequeñas dimensiones. Asimismo, esta misión será la primera dedicada a la medida desde el espacio de la humedad del terreno y la salinidad del mar. La medida de estos parámetros a escala global ayudarán a comprender mejor el cambio climático, la desertización y fenómenos meteorológicos como el Niño y la Niña.

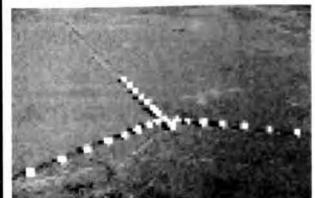
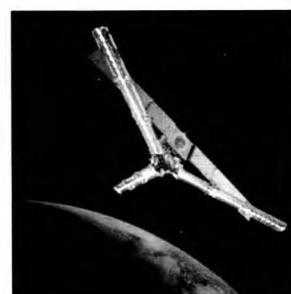


Figura 1a. Visión artística del instrumento MIRAS para la misión SMOS de la ESA. Nótese el parecido con el VLA (véase Figura 1b).

Figura 1b. El Very Large Array (VLA) en Socorro, Nuevo México, USA.

La participación del grupo de la UPC en este proyecto ha abarcado el análisis de las prestaciones del instrumento en términos de resolución angular y radiométrica [4,5], y el compromiso que existe entre ambas, el análisis de las diferentes fuentes de error (receptores, antenas, correladores etc.) [6,7,8] y la propuesta de nuevos métodos de calibración [9] y de reconstrucción de imágenes [10] a partir de los



Figura 2a. Prototipo en banda X formado por 2 antenas móviles que se desplazan a lo largo de unos brazos en forma de Y.



Figura 2b. Composición de 4 imágenes sintéticas de temperatura de brillo. Los absorbentes de microondas aparecen a temperatura ambiente. Las placas metálicas reflejan la radiación del cielo y aparecen fríos.

observables, así como el análisis de fallos e interferencias etc. Las prestaciones de un instrumento tan complejo como éste, unido al hecho de que sea el primero de su serie, ha requerido la implementación de un simulador “end-to-end” [11], desarrollado por la UPC en colaboración con las empresas EADS-CASA y GMV [12]. Asimismo, para validar este nuevo diseño de radiómetro se construyó un demostrador de laboratorio del mismo en banda X (Figura 2a) [13], con el que se estudiaron algunos de los problemas reales que aparecen en este nuevo tipo de instrumentos. Dicho demostrador estaba formado por dos antenas que se desplazaban a lo largo de los tres brazos en forma de Y, como se muestra en la Figura 2a. La Figura 2b muestra 4 imágenes sintetizadas con dicho radiómetro a partir de las medidas de la temperatura de brillo de las placas metálicas situadas sobre el absorbente que se muestran debajo de dichas imágenes.

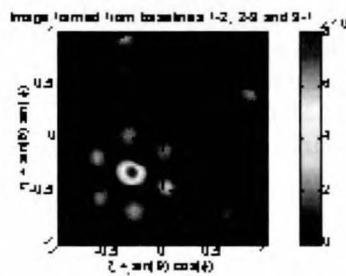


Figura 3b. Imagen “sintética” del Sol formada a partir de las correlaciones mostradas en la Figura 3a.

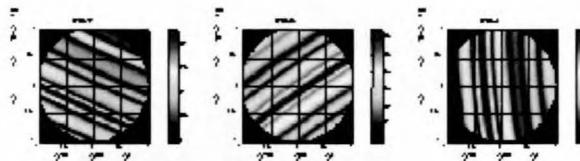


Figura 3a. Pares de correlaciones obtenidos entre las antenas 1 y 2, 2 y 3, y 3 y 1 (L'Albi, 18/12/2002).

Además, también en el campo de la radiometría por síntesis de apertura, en este último año, se están estudiando nuevas configuraciones de radiómetros que permitan formar imágenes a partir de las correlaciones de las señales recibidas por sólo tres antenas, pero introduciendo distintos retardos. La Figura 3 muestra un ejemplo del proceso de formación (sin compensación de fase) de una imagen del Sol a partir de cada uno de los tres pares de correlaciones, con un sistema en banda L (1.413 MHz) con 8 MHz de ancho de banda y una separación entre antenas de 170 m [14].

1.2. Radiometría interferométrica

La radiometría interferométrica utilizando una única línea de base se ha estudiado en teledetección de hidrometeoros (lluvia, hielo...). Esta técnica aprovecha las diferentes propiedades que éstos presentan frente al fondo (superficie de la tierra o mar, mucho más uniforme). La Figura 4 muestra uno de los experimentos llevados a cabo en el European Microwave Signature Laboratory, Joint Research Center/Space Applications Institute (Ispra, Italy), donde un generador de lluvia artificial en el interior de una cámara anecóica se dispuso frente de un radiómetro interferométrico. El generador fue diseñado con 2000 poros y podía generar intensidades de hasta 1500 mm/h desde una altura de 3 m. El radiómetro interferométrico está formado por dos antenas en el que se varía la separación entre ellas [15].

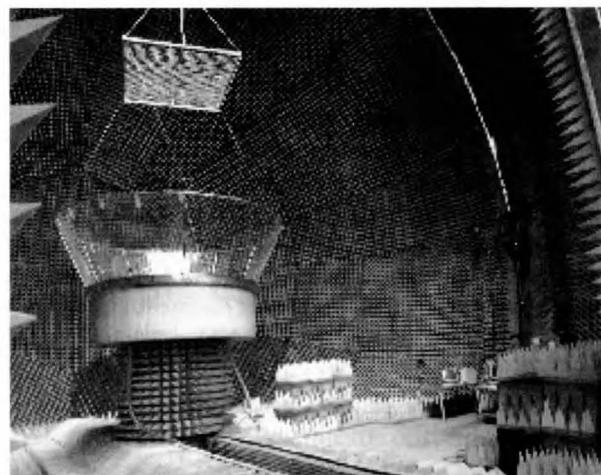


Figura 4. A la izquierda, generador de lluvia (JRC) para generar intensidades de lluvia de hasta 1500 mm/h desde 3 m de altura. A la derecha, radiómetro interferométrico en banda X (UPC).

1.3. Radiometría Polarimétrica

La radiometría polarimétrica mide, además de la potencia de ruido captada por la antena en polarizaciones vertical y horizontal, la correlación cruzada compleja de los campos eléctricos recibidos en ambas polarizaciones.

Numerosos estudios teóricos y experimentales han demostrado en la última década que la aplicación de estas técnicas permite determinar la dirección y la velocidad del viento sobre la superficie del mar, con prestaciones simi-

lares a los dispersómetros de viento radar. Sin ir más lejos, este mismo año ha sido lanzada al espacio la misión Coriolis/Windsat, que es la primera en incorporar radiómetros polarimétricos con este fin.



Figura 5a. Radiómetro polarimétrico en banda X (Laboratorio del Módulo D3, UPC-Campus Nord, Barcelona).

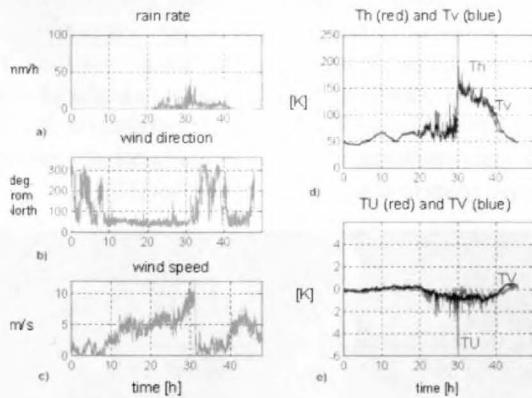


Figura 5b. La suma de las temperaturas de brillo en polarización vertical y horizontal sigue aproximadamente la intensidad de lluvia. La diferencia indica la asimetría de las gotas de lluvia (mayor a mayor intensidad). El tercer parámetro (U) depende de la orientación producida por el viento y el cuarto (V) por la depolarización producida fundamentalmente por los cristales de hielo.

Sin embargo, todavía se conoce poco sobre la emisión polarimétrica del suelo cubierto por vegetación, de los hidrometeoros, etc. Para avanzar en estos campos, en la UPC se han llevado a cabo algunos estudios teóricos que incluyen el modelado discreto del suelo y la vegetación [16], de las gotas de lluvia y de los cristales de nieve y hielo, en función de su tamaño y forma. La Figura 5a muestra el radiómetro polarimétrico en banda X montado en la UPC, situado frente a la ventana del laboratorio para medir la temperatura de brillo en distintas condiciones de lluvia. Las temperaturas de brillo en polarizaciones vertical y horizontal se miden con un radiómetro de Dicke, y el tercer y cuarto parámetros de Stokes

$$U = 29te(\langle E_v E_h^* \rangle), \quad V = 25m(\langle E_v E_h^* \rangle)$$

se miden con un correlador digital complejo. La Figura 5b muestra 48 h de medidas durante los días 2 y 3 de Diciembre

de 1998, que muestran la dependencia de los 4 parámetros de Stokes (gráficas de la derecha en Figura 5b) comparadas con la intensidad de lluvia, y también la velocidad y dirección del viento (gráficas izquierda en Figura 5b), medidos por las estaciones meteorológicas de Barcelona situadas dentro del haz de la antena del radiómetro [17].



Figura 6a. Cara Norte de la Plataforma Casablanca donde se instaló el radiómetro en banda L LAURA durante las campañas WISE 2000 y 2001.

Por otra parte, en el marco de la misión SMOS, se han llevado a cabo las campañas de medida WISE 2000 y 2001, destinadas a la mejora de los modelos de emisión del mar en banda L [18] y, de esta forma, poder desarrollar algoritmos de recuperación de salinidad precisos a partir de medidas radiométricas [19]. Estas campañas se llevaron a cabo en la plataforma petrolífera Casablanca de Repsol, en las costas de Tarragona (Figura 6) en otoño del 2000 y del 2001, y fueron fruto de la cooperación internacional con el Instituto de Ciencias del Mar de Barcelona, las Universidades de Valencia y Massachusetts (USA), y el Laboratoire d'Océanographie Dynamique et Climatologie y el Centre d'Études Terrestres et Planétaires (Francia).

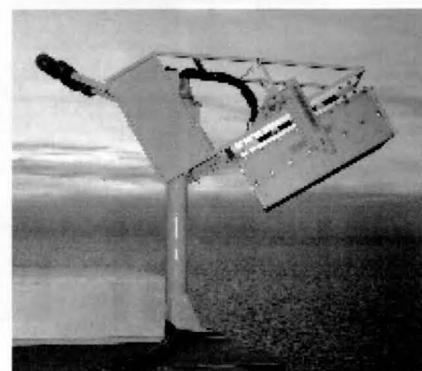


Figura 6b. Radiómetro LAURA en la plataforma Casablanca.

Próximamente, también en el marco de la misión SMOS, se llevarán a cabo nuevas campañas de medida. Una de ellas será específica para la caracterización de la espuma del mar y se realizará en Poble Nou del Delta (Tarragona). En la recuperación de salinidad es necesario determinar la emisividad de las zonas de mar cubiertas por espuma en banda L. Además, dentro del marco de la misión SMOS, es importante realizar campañas de medidas que estudien la emisividad del suelo descubierto y cubierto por vegetación. Está prevista una campaña de medidas en unos campos con humedades controladas en Requena (Valencia) para el próximo verano.

1.4. Altimetría y monitorización del estado del mar utilizando señales de oportunidad GPS

La altimetría y estado del mar utilizando señales de oportunidad GPS es la línea de investigación abierta más recientemente, en el 2001, como fruto de la colaboración con el Institut d'Estudis Espacial de Catalunya, para quien se construyó el receptor GPS de 3 cadenas DODEREC (DOppler DElay REceiver) mostrado en la Figura 7a [20]. Este sistema permite captar hasta tres señales GPS procedentes de antenas diferentes, demodularlas en fase y cuadratura, muestreárlas a 20 MHz y almacenarlas en un disco duro de 80 Gbytes de manera sostenida.

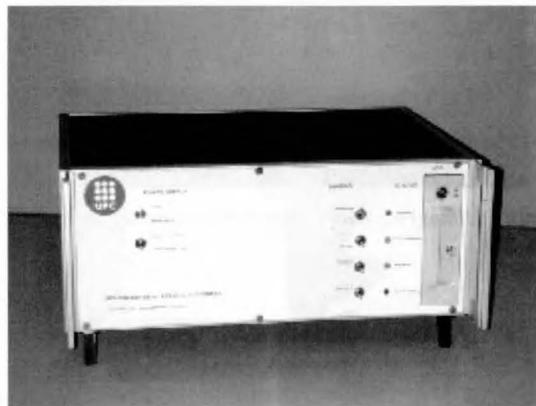


Figura 7a. DO-DEREC receiver desarrollado para el IEEC.

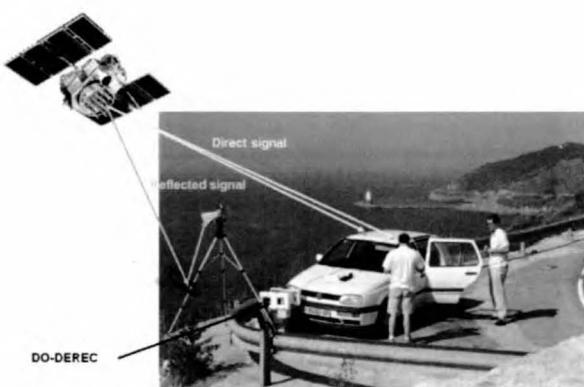


Figura 7b. Test del DODEREC en las costas del Garraf.

Este receptor permite captar la señal directa y reflejada procedente de un mismo satélite (Figura 7b), y constituye un altímetro si se conoce la geometría de observación. Si además, se estudian las propiedades de la señal reflejada, es posible obtener información sobre el estado del mar: el fenómeno es similar al de la reflexión del Sol sobre el mar: la reflexión es distinta para un mar en calma que cuando está perturbado por el viento (Figura 8).



Figura 8a. Reflexión del Sol en el mar en calma.



Figura 8b. Reflexión del Sol en el mar con una ligera brisa. Notar que la reflexión ya no es tan especular.2.

2. TELEDETECCIÓN ACTIVA EN MICROONDAS

por J.J. Mallorquí

En aplicaciones de observación de la Tierra las ventajas de los radares de apertura sintética (SAR) respecto a los sistemas ópticos se conocen desde los años 70: alta resolución espacial, capacidad de trabajar a cualquier hora del día, independencia de las condiciones atmosféricas y meteorológicas, gran sensibilidad a la morfología de las superficies y objetos observados, etc. Las actividades en este campo se iniciaron en 1992 con el procesado interferométrico de los primeros datos del satélite ERS en el marco de un proyecto con la ESA. Desde entonces el grupo ha trabajado en todos los aspectos relacionados con el SAR: construcción de sensores terrestres, procesado de datos crudos (*raw data*) de sensores tanto orbitales, aerotransportados y terrestres, interferometría e interferometría diferencial para aplicaciones cartográficas y de control de riesgos, polarimetría orientada a la clasificación y caracterización de cultivos, detección de barcos y contaminantes en el mar, simuladores de sistemas SAR, etc. El grupo ha colaborado en el marco de varios proyectos con distintas instituciones y empresas, entre las que hay que destacar: ESA, DLR, Indra Espacio, INTA, Institut Cartogràfic de Catalunya, IREA-CNR, Joint Research Center (JRC-Ispra) y Thales Airborne Systems.

2.1. Procesado SAR

Un SAR es básicamente un radar coherente montado en una plataforma móvil que se va desplazando a medida que manda los pulsos. Dado que el haz de la antena ilumina una zona grande de la superficie terrestre, conocida como *swath*, todas las contribuciones situadas a la misma distancia del sensor llegan simultáneamente, de manera que los datos crudos forman una imagen totalmente desenfocada.

Los ecos medidos por el radar coherente que se desplaza sobre la escena contienen la información holográfica de su distribución espacial de reflectividad radar. La formación de una apertura o agrupación sintética de antenas mediante el procesado SAR de los ecos registrados permite obtener imágenes de reflectividad de la escena con resoluciones espaciales que pueden llegar a ser inferiores al metro dependiendo del ancho de banda del pulso y de las características de la antena radar. Esta reflectividad posee un valor complejo que contiene la fase del camino eléctrico recorrido por la onda en su viaje emisor-superficie-receptor, información básica para las aplicaciones de interferometría que se presentarán a continuación.

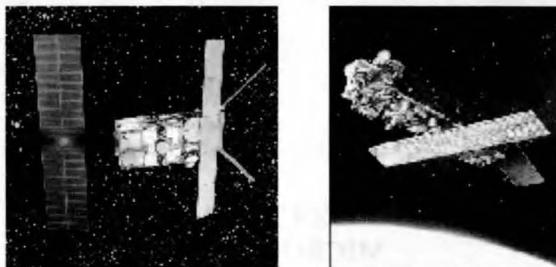


Figura 9a. Satélites ERS1 (izquierda) y ENVISAT (derecha) de la ESA.

El grupo tiene experiencia en el desarrollo de procesadores para distintas plataformas tanto orbitales, como el caso de ERS, como aerotransportadas, sistema E-SAR del DLR (Figura 9). Aunque los algoritmos de procesado para datos orbitales y aerotransportados son muy parecidos, en el segundo caso la complejidad es mayor al tener que incluirse técnicas de compensación de movimiento que permitan corregir la trayectoria inestable de la plataforma [21]. Existen versiones del procesador para PC, estación de trabajo e incluso para el ordenador paralelo SGI O2000 del CEPBA [22], versión realizada en el marcos del proyecto europeo PARSAR en colaboración con la empresa INDRA Espacio.



Figura 9b. SAR aerotransportado E-SAR del Deutschen Zentrum für Luft- und Raumfahrt (DLR).

Se han desarrollado algoritmos para los distintos modos de funcionamiento de los sensores: *stripmap*, en el que la antena apunta de forma fija y ortogonal a la trayectoria de la plataforma, y *spotlight*, en el que la antena apunta siempre a la misma escena durante el movimiento de la plataforma. Actualmente estamos trabajando en la ampliación del procesador para trabajar con datos del satélite RADARSAT en modo *scansar*, en este modo el haz de la antena se va desplazando cada cierto tiempo entre *subswaths* vecinos para aumentar el área total explorada. También está previsto el procesado de datos del nuevo sensor ENVISAT de la ESA. La Figura 10 muestra un ejemplo de imagen



Figura 10. Imagen SAR del área de Barcelona del satélite ERS2 de la ESA.

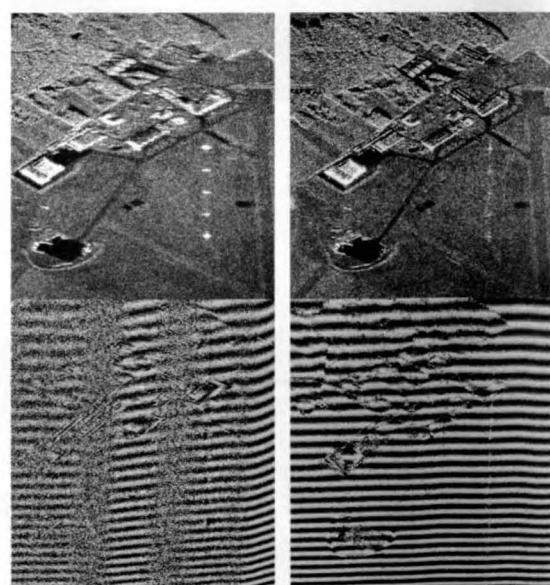


Figura 11. Imagen e interferograma con datos del E-SAR (DLR) procesados sin compensación de moviendo (izquierda) y con compensación de movimiento (derecha).

del satélite ERS y la Figura 11 la importancia de la compensación de movimiento en el caso aerotransportado.

2.2. Interferometría e interferometría diferencial

En interferometría SAR se trabaja con pares de imágenes obtenidas desde posiciones separadas por una distancia, o línea de base, adecuada. Mediante un alineado y corrección compleja se obtiene la diferencia de fase para cada punto de la imagen que está relacionado con la elevación del terreno, permitiendo la generación de modelos digitales de elevación (DEM) con precisión submétrica. La cadena interferométrica desarrollada en el grupo permite trabajar tanto con sensores orbitales como aerotransportados. Partiendo de un par de imágenes procesadas, se procede a su corregistrado para formar el interferograma. La información útil se encuentra en la fase, pero para poder extraer topografía es necesario desenrollarla (Figura 11a). Una vez la fase está desenrollada se pueden geolocalizar todos los puntos y, finalmente, elaborar el mapa topográfico (Figura 11b). En el caso aerotransportado, además, se han desarrollado algoritmos de calibración y corrección de errores en las imágenes interferométricas [23, 24, 25]. Se han realizado convenios de transferencia tecnológica con Indra Espacio, la cadena interferométrica, y el DLR, algoritmos de georeferenciación precisa.

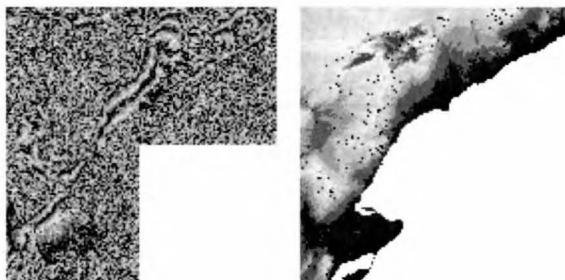


Figura 12a. Interferograma obtenido a partir de dos imágenes tandem de los satélites ERS1 y ERS2 de la ESA. A la izquierda en interferograma enrollado y a la derecha desenrollado.

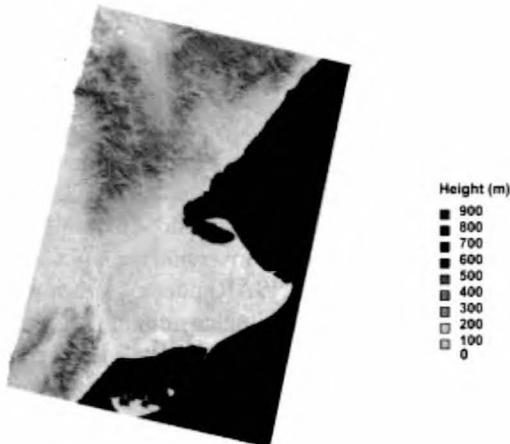


Figura 12b. Modelo Digital de Elevación (DEM) de la zona del Delta de Ebro en coordenadas UTM.

Si se trabaja con imágenes interferométricas en modo diferencial, pueden medirse alteraciones en altura de las superficies con precisiones del orden de los milímetros, permitiendo la monitorización de cambios de terrenos con indudables aplicaciones en estudios de erosión, dinámica costera, explotación petrolífera, geodinámica, sismología y vulcanología. La interferometría diferencial clásica se limitaba a utilizar un par de imágenes de línea de base corta, para minimizar el efecto de la topografía, a la que se eliminaba la componente topográfica de la fase mediante la utilización de un DEM de la zona. En la práctica las limitaciones son múltiples debido a la decorrelación temporal o pérdida de coherencia de fase de las imágenes, los artefactos atmosféricos presentes en las imágenes y los errores residuales del DEM utilizado. Para solventar estas limitaciones se han desarrollado algoritmos avanzados que se basan en la utilización de un gran número de imágenes adquiridas a lo largo del período de observación, generalmente varios años. Mediante el ajuste de un modelo a los datos se pueden determinar las componentes de movimiento para cada píxel de la imagen tanto lineal como no lineal, el error en el DEM utilizado para eliminar la componente topográfica y recuperar la atmósfera para cada imagen, eliminando así su contribución [26, 27, 28]. Se han realizado convenios con Indra Espacio y colaboraciones con el Institut Cartogràfic de Catalunya y el IREA italiano. Actualmente colaboramos con el DLR en la extensión de estas técnicas al caso aerotransportado.

La Figura 13 muestra la capacidad del sistema para medir pequeñas deformaciones en edificios, en este caso una nave industrial metálica de la ciudad de Nápoles. El patrón de deformación se ajusta perfectamente a las variaciones térmicas registradas.

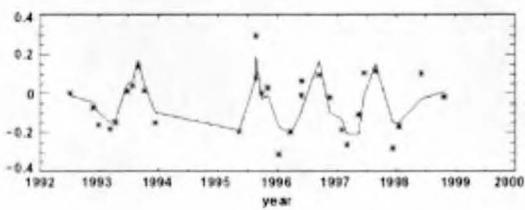


Figura 13. Correlación entre la deformación (cm) en una estructura metálica (*) y las variaciones térmicas (línea continua, +5°-35°) en la ciudad de Nápoles.

2.3. Simulación de sistemas SAR

En el marco del proyecto europeo IMPAST [29] se está desarrollando un simulador de sistemas SAR polarimétricos que permite simular de forma realista el raw data que registraría un sistema SAR orbital al iluminar un barco pesquero. El simulador se basa en el código de predicción de sección recta radar (RCS) mediante algoritmos de alta frecuencia desarrollado en el Departamento llamado GRECO al que se le han añadido módulos para el cálculo de la órbita del satélite, su actitud, modelado de la señal radar, etc. Los modelos de barcos se construyen utilizando programas de CAD a partir de

información de barcos reales proporcionada por los astilleros. El simulador permite conocer las limitaciones de los sistemas actuales para la detección y clasificación de barcos, y estudiar futuras configuraciones para esta aplicación que incluyan polarimetría y distintos modos de funcionamiento [30]. La Figura 15 muestra el diagrama de bloques del simulador y la Figura 16 una pantalla de GRECO con un pesquero.

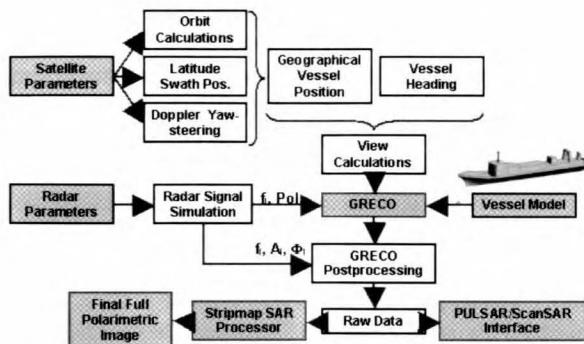


Figura 15. Diagrama de bloques del simulador SAR orbital.

El simulador se está validando a partir de las medidas adquiridas en distintas campañas sobre barcos conocidos y monitorizados. En el marco del mismo proyecto también se están desarrollando algoritmos para la detección de barcos en imágenes SAR.

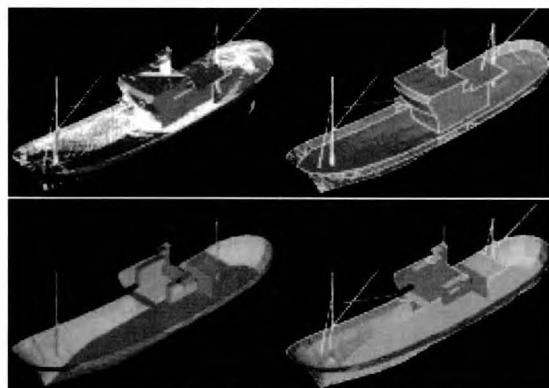


Figura 16. Modelo de barcos pesquero en una pantalla de GRECO. La ventana superior izquierda muestra los impactos de rayos en las superficies visibles del objeto (puntos blancos), la ventana inferior izquierda muestra las caras ocultas. La ventana superior derecha muestra las aristas del objeto y la inferior izquierda las aristas que pertenecen a un solo objeto.

2.4. Dispersómetro polarimétrico

Las imágenes proporcionadas por radares de apertura sintética (SAR) embarcados en satélites o aerotransportados aportan una importante información sobre diferentes parámetros geofísicos y biológicos de la superficie terrestre. Para la correcta interpretación de estas imágenes, debe de establecerse un nexo de unión entre los mecanismos que producen la dispersión electromagnética del eco radar y los parámetros geofísicos y

biológicos que caracterizan la reflectividad del blanco. Para este fin, existen unos sistemas radar, de baja potencia y corto alcance, denominados dispersómetros, capaces de caracterizar la reflectividad de superficies terrestres y marinas en condiciones controladas.

En el grupo se ha desarrollado un dispersómetro capaz de efectuar medidas en las cuatro combinaciones polarimétricas posibles (HH,VV,HV y VH) y en un margen de frecuencias que cubre de los 2 GHz hasta los 18 GHz (Figura 17). El conocimiento de la reflectividad de un blanco en todas las polarizaciones, en un gran margen frecuencial y en diversos ángulos de incidencia, permite una profunda caracterización de éste. Además, este sistema, mediante unas mínimas modificaciones mecánicas, es capaz de generar imágenes SAR e INSAR de resolución centimétrica [31].



Figura 17. Dispersómetro polarimétrico desplegado en la campaña de medidas en Castellterçol.

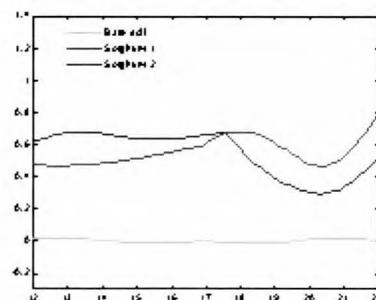


Figura 18. Estimación del valor de altura de vegetación.

En el marco del proyecto Europeo: “Polarimetría Radar: Teoría y Aplicaciones”, en colaboración con otros centros europeos como el Centro Aeroespacial Alemán (DLR), el Centro Europeo de Investigación (JRC)-Italia y otros grupos, y utilizando los datos proporcionados por nuestro dispersómetro, se han desarrollado un conjunto de algoritmos polarimétricos de inversión que aplicados sobre imágenes SAR e INSAR pueden proporcionar parámetros geofísicos y biológicos como: rugosidad y humedad del terreno, y biomasa vegetal. La Figura 17 muestra la campaña realizada con el dispersómetro en el pueblo de Castellterçol, donde se midieron dos campos cubiertos de sorgo. Como la máxima respuesta de la cubierta vegetal (sorgo) y el terreno está asociada a polarizaciones diferentes, pueden utilizarse técnicas

polarimétricas-interferométricas con la finalidad de separar los centros de fase de cada blanco y así poder realizar una estima del valor medio de la altura vegetación (Figura 18).

El dispersómetro también se ha aplicado al problema de detección de minas enterradas [32]. En el marco de un nuevo proyecto CICYT se está desarrollando un nuevo sistema radar para monitorización de costas, control de corrimientos de tierras y movimientos estructurales de edificios.

3. TELEDETECCIÓN ÓPTICA (LIDAR)

por F. Rocadenbosch

Las actividades del grupo EEF en sistemas lidar atmosféricos se iniciaron en 1993 y han dado lugar al desarrollo de dos grandes sistemas de radar láser pulsado, un sistema lidar elástico 1D (Sec. 3.1) y otro Raman elástico 3D (Secs. 3.2 y 3.3), así como a un considerable know-how en sistemas láser Doppler y subsistemas de baja potencia (Sec. 3.4). Las principales claves tecnológicas y potencialidades asociadas a estos nuevos sistemas de teledetección se describen, a modo de introducción, en el artículo de Rocadenbosch et al., "Remote Sensing mediante LIDAR en la UPC", que aparecerá en el próximo número de esta revista. Su lectura es recomendable de cara a profundizar los resultados aplicados que se describen a continuación y que pretenden ilustrar parte del itinerario de proyectos a través de estos diez años de tecnología lidar en el grupo EEF de la UPC:

3.1 Sistemas lidar elásticos 1D

En 1993 se inicia el desarrollo de la estación lidar elástica (misma longitud de onda incidente que retrodispersada) de exploración atmosférica vertical (de ahí, la denominación de 1D), pionera en España, dentro del proyecto del Plan Nacional de I+D TIC93-043 [33]. El sistema fijo (Figura



Figura 19. Estación lidar elástica de exploración vertical (1993-1996). (Izq.) La estación lidar en operación nocturna; (centro) Fuente emisora basada en un láser de Nd:YAG, operando a 532 nm (2 f0) con potencias de pico de > 50MW; (der.) Equipos optoelectrónicos de recepción en el interior de la cúpula en la azotea del módulo D3, Campus Nord, UPC.

19) operaba con pulsos de unos 400 mJ y 10 ns/10Hz de duración a 532 nm (VIS, verde).

Ya con este primer sistema monolongitud de onda, cuyo desarrollo se centró el período 1993-96, se llevaron a cabo diversas campañas de medidas semicuantitativas de los parámetros ópticos atmosféricos de extinción y retrodispersión, coordinadas con el grupo del Instituto de Modelización Ambiental (ITEMA) de la UPC y con el Centro Territorial de Cataluña del Instituto Nacional de Meteorología (INM). Se dispuso así, de una primera base de datos relacional de 237 medidas (1996-97), algunas de las cuales comprendían el muestreo (aunque irregular) de ciclos diurnos y nocturnos a lo largo de períodos continuados de más de 16 h y, también, de una avanzada librería software de inversión y simulación denominada *link-lidar* (Figura 20).

En la actualidad *link-lidar* comprende tres paquetes principales: 1) *link-atmos*, que permite realizar estudios de la dispersión (scattering) elástica atmosférica (Rayleigh y Mie) para mono y polidispersiones, estudios de radiancia solar y lunar y, simulación de trayectos láser sobre el área metropolitana de Barcelona; 2) *link-budget*, orientado al diseño, especificación y simulación extremo a extremo de enlaces lidar; y 3) *link-detect*, el más extenso, que integra múltiples algoritmos de procesado e inversión de señales lidar así como numerosas herramientas de simulación estocástica de dichos algoritmos.

Por lo que respecta al procesado de las señales lidar, desde 1998 el grupo ha contribuido interesantes resultados a la comunidad Optica y *Lidar Remote Sensing*. Uno de ellos ha sido la evaluación y tipificación de los algoritmos de inversión de los parámetros ópticos atmosféricos conocidos en el estado del arte, en términos de error versus relación señal a ruido y margen de visibilidad atmosférico, así como en términos de las hipótesis calibradoras que manejan [34-36]. Otro de los avances realizados, ha sido la aplicación de la experiencia en técnicas de procesado de la señal en el seno del Dep. de Teoría de la Señal y Comunicaciones, a la inversión adaptativa de señales lidar, tanto mediante algoritmos basados en aproximaciones sucesivas (algoritmos KFS en *link-detect*) como mediante filtrado de Kalman [37]. De manera continua, *link-lidar* ha ido integrando desde 1993 todas las herramientas de simulación e inversión que se han ido desarrollando hasta nuestros días e incluye, no sólo herramientas ad-hoc orientadas a sistemas elásticos, sino también a sistemas inelásticos de tipo Raman [38], estas últimas desarrolladas en el marco del proyecto europeo EARLINET (2000-2003), que se presentará más adelante.

3.2 Sistemas elásticos 3D

Partiendo de la experiencia adquirida con la estación lidar fija 1D y de la voluntad de explotar las capacidades de medida de estos sistemas en el campo medioambiental (calidad del aire), a finales de 1997 se inicia el desarrollo

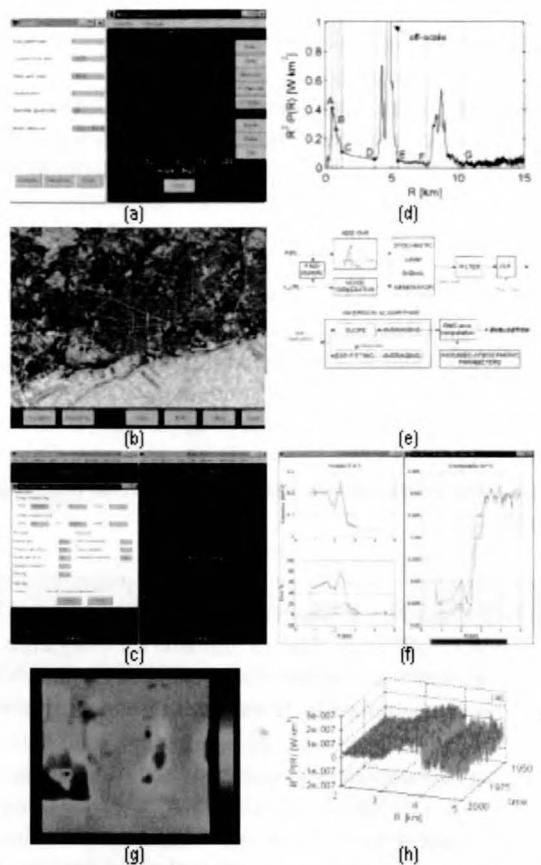


Figura 20. Presentación de macro-paquete de procesado y simulación lidar, link-lidar.

(a) Análisis polarimétrico de la dispersión Mie (parámetros de Mueller/Stokes) de un aerosol tipo (haze M) dados sus parámetros de tamaño, longitud de onda, densidad de partículas e índice de refracción complejo mediante link-atmos.

(b) Estimación de la trayectoria 3D de un haz láser sobre el área metropolitana de Barcelona mediante link-atmos. Acceso a los parámetros de contaminación del Ayuntamiento de Barcelona.

(c) Simulación paramétrica de alcance lidar (elástico) vs. energía, margen de visibilidad atmosférico y radiancia de fondo para diferentes cabezales detectores (en términos de ganancia, NEP y umbrales de saturación) mediante link-budget.

(d) Herramienta de procesado de señales lidar mediante segmentación morfológica en link-detect. Las rayas verticales, que indican intervalos de distancia con características óptico-atmosféricas similares (p. ej. DE y FG identifican nubes), son definidas automáticamente por el algoritmo de procesado de imagen.

(e) Simulador del paquete link-detect para la evaluación estocástica de algoritmos de inversión no adaptativos, en base a sesgo y error cuadrático medio vs. SNR y margen de visibilidad atmosférico.

(f) Inversión iterativa de los parámetros de extinción y retrodispersión Mie, mediante el algoritmo KFS (Klett-Fernald-Sasano) de link-detect.

(g) Reconstrucción de una escena de nubes medianas (entre 3.2 y 3.7 km de altura) detectadas mediante sondeo lidar elástico utilizando procesado adaptativo de Kalman (paquete link-detect).

(h) Cancelación de ruido en el retorno rango-corregido de una señal lidar ($R^2 P(R)$) entre 1950 y 2000 m de altura) mediante filtrado de Kalman no-lineal en link-detect.

de un lidar elástico portátil con capacidad de apuntamiento 3D (barridos en elevación y acimut [39]). Como criterios ingenieriles básicos se fijan la portabilidad, modularidad y escalabilidad del equipo. El crecimiento continuo de este sistema iba a continuar hasta nuestros días merced a múltiples proyectos nacionales de la CICYT (AMB96-1144-C02, TIC99-1050-C03-01 y REN2000-1754-C02), varios de ellos coordinados, fundamentalmente al nivel de campañas de medidas, con el Instituto de Modelización Ambiental (ITEMA, actualmente Dep. de Ingeniería y Proyectos, UPC), acciones especiales y, de la financiación del V Programa Marco de Investigación y Desarrollo Tecnológico de la Unión Europea (EVR1-CT-1999-40003). Este último proyecto ha permitido consolidarse en la red europea de lidares EARLINET [40] (actualmente formada por más de veinte estaciones), la cuál, sobre la base de operación regular y coordinada, tiene como objetivo el establecimiento de bases de datos estadísticas a escala continental que cuantifiquen las emisiones biogénicas y antropogénicas de aerosoles y su impacto climático y ambiental.

Actualmente, el sistema dispone de dos canales elásticos, que le permiten operar a 1064 nm (infrarrojo cercano, NIR) y 532 nm (visible, VIS) con pulsos de unos 120 mJ y 6 ns de duración y una capacidad de barrido 3D de 300°x120° (acimut elevación). Además de estas capacidades de medida, se están integrando ópticas polarimétricas y también un tercer canal inelástico en recepción a 607 nm de tipo Raman (V. Sec.3.3 y Figura 21), elementos todos ellos, que lo hacen pionero dentro de red europea EARLINET.

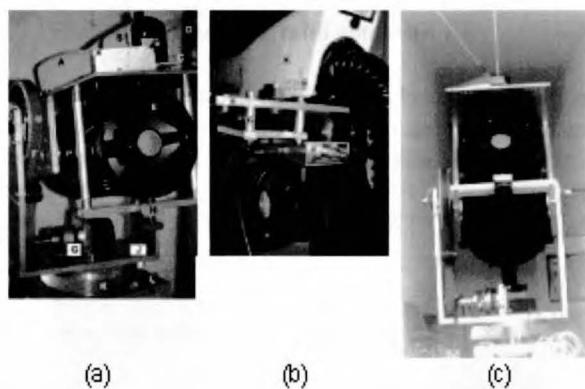


Figura 21. Cabezal de emisión/recepción del lidar 3D, 1064/532 nm transportable (actualmente sistema Scanning Raman Lidar, SRL). (Izq.) El láser de Nd:YAG y el doblador de frecuencia (A, B) se encuentran montados sobre una plataforma (C) solidaria al telescopio de 20 cm de apertura (E). Se aprecia el motor de elevación (G) dispuesto transversalmente sobre la base de la horquilla giratoria en azimut (J) y la correa dentada (F), que transmite el movimiento de elevación. (Centro) Vista posterior ilustrando el receptor optoelectrónico acoplado a la rótula porta-objetivos (E); (B, C) corresponden a la mecanización micrométrica de ajuste del factor de solapamiento/apuntamiento óptico. (Der.) El sistema lidar 3D en operación diurna simultánea a 532 nm y 1064 nm (en la foto, tan sólo es visible el haz a 532 nm).

Desde el punto de vista de campañas de medida, la sinergia entre el grupo EEF y el del Ing. de Proyectos tanto en diversos de los proyectos previos comentados del Plan Nacional (AMB96-1144-C02 y REN2000-1754-C02), el proyecto EARLINET de la Unión Europea y, principalmente, en el proyecto IMMPACTE de la CIRIT/Departamento de Medio Ambiente del Gobierno Autónomo Catalán, ha permitido una alta capacidad de interpretación y análisis de resultados. Ello ha hecho que se esté en una situación especial, que permite avanzar notablemente en el estudio de un ciclo anual que cubra el conjunto de situaciones meteorológicas típicas que se dan, durante un período de tres años (2001-2003), con el feedback que esto supone desde el punto de vista del desarrollo y verificación de nuevas metodologías y modelos meteorológicos [41].

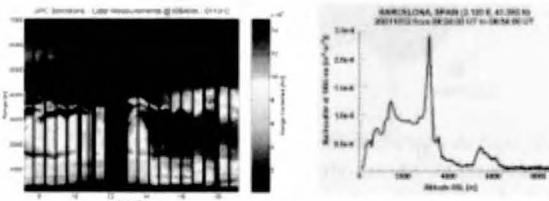


Figura 22a. Señal lidar a 1064 nm corregida en distancia ($R2P(R)$), en función de la altura y del tiempo el día 12 de octubre de 2001 (escala de color, unidades arbitrarias).

Figura 22b. Inversión del coeficiente de retrodispersión ($m-1sr-1$) mediante integración temporal de las medidas de Fig. 4a en el intervalo de media hora entre las 8:24 y las 8:54 UTC.

A título de ejemplo, la Figura 22a muestra la evolución temporal de los retornos lidar rango-corregidos a 1064 nm, con una resolución temporal de 1 minuto, obtenida mediante el lidar 3D de la Figura 21 desde el edificio D3 del Campus Nord de la UPC, el día 12 de octubre de 2001, durante una alerta coordinada medioambiental de invasión de polvo procedente del Sáhara [42]. La imagen es aproximadamente proporcional al coeficiente de retrodispersión de la atmósfera en función de la altura con una resolución espacial de 7.5 m. Se observan claramente

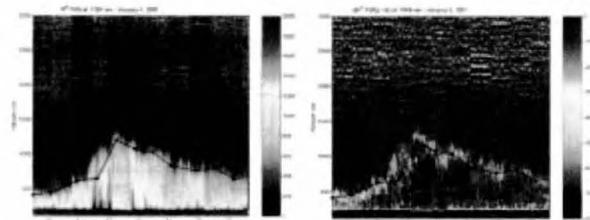


Figura 23a. Determinación de la altura de la capa de mezcla a partir de las medidas del 8 de enero de 2001 obtenidas con el lidar 3D de Figura 21a 1064 nm. Señal lidar corregida en distancia ($R2P(R)$), en función de la altura y del tiempo (resolución temporal 1 minuto, escala de color en unidades arbitrarias).

Figura 23b. Procesado de la señal lidar mediante derivación y filtrado del retorno corregido en distancia ($R2P(R)$) de Figura 23a, en función de la altura y del tiempo (escala de color, unidades arbitrarias). Puntos superpuestos: altura de la capa de mezcla (PBL).

las estratificaciones de aerosoles hasta unos 4000 m de altura así como un incremento en su espesor y densidad a partir de las 14:00 UTC. La Figura 22b muestra la inversión del coeficiente de retrodispersión a 1064 nm mediante link-detect para la medida de Figura 22a integrada en el intervalo de media hora comprendido entre las 8:24 y las 8:54 UTC.

La Figura 23 muestra la estimación de la evolución temporal de la altura de la capa de mezcla (PBL, una capa formada por aerosoles bien mezclados que se extiende desde el suelo hasta unos 2-3 km de altura y muy influenciada por los ciclos diurnos/nocturnos, vientos y fenómenos de transporte como, por ejemplo, las inyecciones de masas debidas a incendios y a invasiones de polvo en suspensión) a partir de las medidas efectuadas el 8 de Junio de 2001 con el lidar 3D de Fig.21, a 1064 nm. Se combinan algoritmos de inversión de link-detect con modelos interpretativos del Dep. de Ing. de Proyectos.

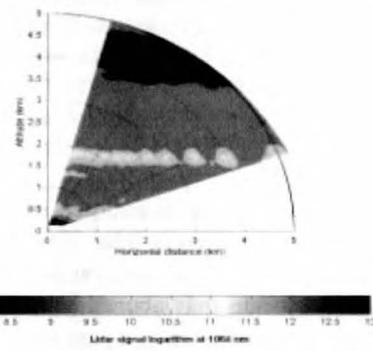


Figura 24. Barrido en elevación (RHI) a 1064 nm mediante el sistema lidar 3D de la Fig. 3. Alcance 5 km, escala radial a 1km/división, mapa de color en unidades arbitrarias ($\ln(R2P(R))$) en a.u.)

Finalmente, la Figura 24 es un ejemplo de las posibilidades de información que permite obtener un barrido 2D en elevación (RHI). El barrido entre 30° y 75° de elevación pone de manifiesto la presencia de una estratificación entre 1500 y 2000 m de altura [43].

3.3 Lidar Raman

Independientemente de las posibilidades de los lidares elásticos 1D y 3D, este tipo de lidares tan sólo permite invertir información de los coeficientes de retrodispersión y extinción que producen los aerosoles a las longitudes de onda de trabajo con una calidad que depende de hipótesis correladoras no siempre contrastables, como es la relación aproximada entre ambos coeficientes ópticos (lidar ratio) y los valores de calibración que el usuario debe introducir en los algoritmos de inversión [35,44].

A diferencia de lo que ocurre con los sistemas elásticos, en que la longitud de onda en recepción es la misma que en emisión, los sistemas lidar Raman utilizan la radiación dispersada por efecto Raman (inelástica) con los constitu-

yentes atmosféricos, que se produce a una longitud de onda desplazada con respecto de la emitida y que es propia de cada molécula o átomo, para autocalibrar los retornos lidar procedentes de los canales elásticos a 1064 nm y 532 nm.



Figura 25. Layout óptico del cabezal Raman del sistema lidar 3D (Scanning Raman Lidar, SRL), que permite separar las tres longitudes de onda en recepción (las dos elásticas 1064/532 nm y la inelástica Raman a 607 nm). Se observa la entrada por fibra óptica (A) procedente del telescopio (no visible en la foto), lentes colimadoras (B), filtros dicroicos separadores (C,D), receptor PMT (tubo fotomultiplicador) (E) y los dos receptores elásticos con control digital (G,F).

Desde el 2001, se está trabajando en integrar un canal Raman en el citado sistema lidar 3D [45]. Este canal Raman, aprovecha la interacción con el nitrógeno atmosférico N₂ (la especie más abundante y de proporción perfectamente conocida (78%)), para autocalibrar el sistema e invertir así, de forma totalmente independiente [46], los coeficientes de extinción y retrodispersión atmosféricos. La Figura 25 muestra un prototipo del cabezal óptico separador de las tres longitudes de onda en recepción, y que dará paso al sistema SRL (Scanning Raman Lidar), pionero en Europa por sus capacidades de medida integradas. La Figura 26 presenta un ejemplo de inversión de señales Raman en *link-detect* con recuperación de la *lidar ratio* (ya no una hipótesis correladora) en el marco de una intercomparación software [47] dentro de EARLINET.

3.4 Actividades de lidar Doppler y baja potencia

El grupo de investigación ha desarrollado además un trabajo simultáneo en técnicas de teledetección óptica mediante láseres de baja potencia, que ha dado lugar a: 1) la realización de radares Doppler láser para la medida de velocidades en blancos sólidos [48] (Figura 27) y 2) aplicaciones específicas de medida de velocidades hidrodinámicas [49-50] (Figura 28). Parte de sus resultados se han utilizando en el proyecto de transferencia tecnológica PETRI 95-0249-OP con la empresa española Monocrom, S.L.

La primera actividad, se ha organizado entorno a maquetas de blancos sólidos y un interferómetro Fabry-Pérot –

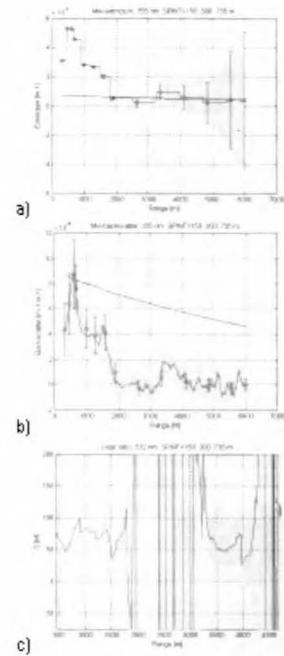


Figura 26. Inversión señales lidar Raman mediante link-detect en el marco de la intercomparación con el resto de equipos investigadores del proyecto europeo EARLINET (2000-2003). El procesado de señal utiliza herramientas apropiadas de cancelación de ruido y promediado espacial con resolución variable para adecuar la resolución de inversión a la SNR disponible para cada tramo de distancias. (a) Recuperación de la extinción Mie atmosférica (en rojo), errorbars asociados (amarillo) y comparación con el retorno Rayleigh (azul). (b) Idem que (a) para la variable atmosférica de retrodispersión Mie (backscatter). (c) Recuperación de la lidar ratio, que puede relacionarse con el tipo de aerosol atmosférico monitorizado. Obsérvese como el algoritmo estima perfectamente este parámetro (con baja varianza) en el primer tramo (500-1800 m), correspondiente a la capa de mezcla atmosférica (PBL), y en un segundo intervalo (3200-4000 m) correspondiente a una estratificación por inyección de aerosoles a esa altura.

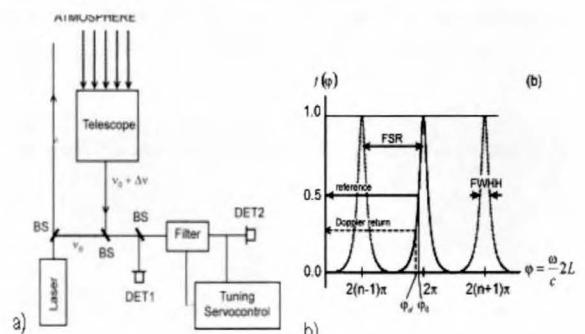


Figura 27. (a) Layout óptico correspondiente a un prototipo de laboratorio de lidar Doppler basado en la técnica de detección por flanco (edge-technique). En recepción, el retorno Doppler (w_d) se compara periódicamente con una muestra láser de la fuente emisora (w_0) y, mediante un interferómetro Fabry-Perot, se mide el desplazamiento Doppler y su signo. (b) Principio de la detección por flanco. El interferómetro óptico (equivalente a un filtro paso banda de gran pureza espectral) actúa como transductor frecuencia-tensión que cuantifica el desplazamiento Doppler (w_d).

como corazón del sistema- como primer paso para abordar un sistema laser Doppler atmosférico de medida de campos de vientos. El sistema utiliza la técnica de detección por flanco (o edge-tecnique), como técnica de detección directa del desplazamiento Doppler y un sistema específico de servocontrol y sincronización, para mantener la estabilidad y repetibilidad de la curva de respuesta del filtro interferencial.

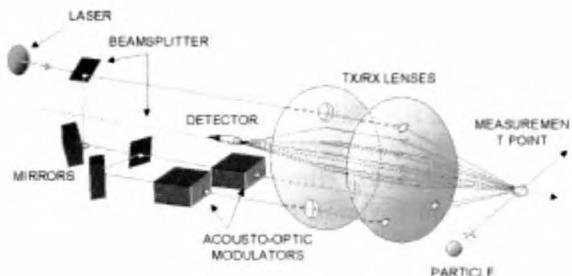


Figura 28. Teledetección óptica semiconductor de baja potencia. Ejemplo de un prototipo de laboratorio de anemometría láser Doppler (LDA) para la medida de velocidades de fluidos mediante interferometría de franjas.

Como actividad complementaria orientada a establecer una línea estratégica de investigación y transferencia tecnológica medioambiental e industrial, el grupo se está introduciendo en el desarrollo de técnicas de detección de compuestos químicos mediante sistemas lidar de bajo coste que empleen como fuentes diodos láser y, eventualmente, con resolución en distancia utilizando esquemas de modulación apropiados, sintonizables [51].

CONCLUSIONES

Este artículo describe brevemente las actividades realizadas en los últimos 10 años en el campo de la Teledetección por el grupo de Ingeniería Electromagnética y Fotónica del Departamento de Teoría de la Señal y Comunicaciones de la Universitat Politècnica de Catalunya [1]. Además de los otros profesores que trabajan en Teledetección, estas actividades no se hubieran podido llevar a cabo sin la colaboración de los técnicos de laboratorio y taller, y los demás profesores del Grupo EEF que trabajan en los campos afines como las Antenas, los Circuitos de Microondas, los Métodos Numéricos etc.

BIBLIOGRAFÍA

- [1] UPC remote sensing: <http://www-tsc.upc.es/rs>
- [2] SMOS Earth Explorer web page: http://www.esa.int/export/esaLP/ESAMBA2VMOC_smos_0.html (Febrero 2003)
- [3] Thompson, A. R., J. M. Moran, G. W. Swenson, «Interferometry and Synthesis in Radio Astronomy», John Wiley and Sons, 1986.
- [4] Bará, J., A. Camps, F. Torres, I. Corbella, «Angular Resolution of Two-dimensional, Hexagonally Sampled Interferometric Radiometers», Radio Science, Vol 33, No 5, pp. 1459-1473, Septiembre-Octubre 1998.
- [5] Camps, A., I. Corbella, J. Bará, F. Torres, «Radiometric Sensitivity Computation in Aperture Synthesis Interferometric Radiometry», IEEE Transactions on Geoscience and Remote Sensing, Vol 35, No 2, pp. 680-685, Marzo 1998.
- [6] Torres, F., A. Camps, J. Bará, I. Corbella, «Impact of Receiver Errors on the Radiometric Resolution of Large 2D Aperture Synthesis Radiometers. Study Applied to MIRAS», Radio Science, Vol. 32, No 2, pp. 629-642, Marzo-Abril, 1997.
- [7] Camps, A., J. Bará, F. Torres, I. Corbella, J. Romeu, «Impact of Antenna Errors on the Radiometric Accuracy of Large Aperture Synthesis Radiometers. Study Applied to MIRAS», Radio Science, Vol. 32, No 2, pp. 657-668, Marzo-Abril 1997
- [8] Camps, A., F. Torres, I. Corbella, J. Bará, J. A. Lluch, «Threshold and Timing Errors of 1 bit/2 level Digital Correlators in Earth Observation Synthetic Aperture Radiometry», Electronics Letters, Vol. 33, No 9, pp. 821-823, 24 Abril 1997.
- [9] Torres, F., A. Camps, J. Bará, I. Corbella, R. Ferrero, «On-Board Phase and Modulus Calibration of Large Aperture Synthesis Radiometers: Study Applied to MIRAS», IEEE Trans on Geosci. and Rem. Sensing, GRS-34, No 4, pp 1000-1009.
- [10] Camps, A., J. Bará, F. Torres, I. Corbella, «Extension of the CLEAN Technique to the Microwave Imaging of Continuous Thermal Sources by Means of Aperture Synthesis Radiometers», Progress In Electromagnetics Research, Vol. PIER 18, pp. 67-83, Enero 1998.
- [11] Camps, A., F. Torres, J. Bará, I. Corbella, M. Pino, M. Martín Neira, «Evaluation of MIRAS spaceborne instrument performance: snap shot radiometric accuracy and its improvement by means of pixel averaging», European Symp. on Aerosp. Rem. Sens., Conf. on Sensors, Systems and Next Generation Satellites III, Proc. of the Europto 1997, London, UK, Sept. 22-26,1997, Vol 3221, pp 43-52, Septiembre 1997.
- [12] Camps, A., M. Zapata, I. Corbella, «SMOS Performance Simulation Analysis», International Geoscience and Remote Sensing Symposium IGARSS 2002, Toronto, Canada, Junio 2002.
- [13] Camps, A., F. Torres, I. Corbella, J. Bará, X. Soler, «Calibration and experimental results of a two-dimensional interferometric radiometer laboratory prototype», Radio Science, Vol. 32, No 5, pp 1821-1832, Septiembre-Octubre 1997.

- [14] Sumpsi, A., «Design and test of a Delay-Doppler Radiometer», Proyecto Final de Carrera, ETSETB, 2003 (en prensa).
- [15] J. Bará, I. Corbella, F. Torres, A. Camps, M. Martín-Neira, G. Nesti, E. Olhmer, «Remote Sensing of Hydrometeors by means of interferometric radiometry: theory and experimental results», Radio Science, Vol 35, No 3, pp. 799-812, Mayo-Junio 2000.
- [16] Martínez-Vázquez, A., A. Camps, N. Duffo, M. Vall-llossera, J.M. López-Sánchez, «Full-Polarimetric Emissivity of Vegetation-Covered Soils: Vegetation Structure Effects», International Geoscience and Remote Sensing Symposium IGARSS 2002, Toronto, Canada, Junio 2002.
- [17] Camps, A., M. Vall-llossera, N. Duffo, F. Torres, J. Bará, I. Corbella, J. Capdevila, «Polarimetric Radiometry of Rain Events: Theoretical Prediction and Experimental Results», 6th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment, Florence, Italy, March 15-18, 1999.
- [18] A. Camps, J. Font, J. Etchetto, V. Caselles, A. Weill, I. Corbella, M. Vall-llossera, N. Duffo, F. Torres, R. Villarino, L. Enrique, A. Julià, C. Gabarró, J. Boutin, E. Rubio, S.C. Reising, P. Wursteisen, M. Berger, y M. Martín-Neira, «Sea Surface Emissivity Observations at L-band: First Results of the Wind and Salinity Experiment WISE-2000», IEEE Trans. on Geoscience and Remote Sensing, Vol 40, No 10, pp 2117-2130, October 2002.
- [19] A. Camps, I. Corbella, M. Vall-llossera, N. Duffo, F. Torres, R. Villarino, L. Enrique, F. Julbé, J. Font, A. Julià, C. Gabarró, J. Etchetto, J. Boutin, A. Weill, V. Caselles, E. Rubio, P. Wursteisen, M. Martín-Neira, «L-Band Sea Surface Emissivity: Preliminary Results Of The WISE-2000 Campaign And Its Application To Salinity Retrieval In The SMOS Mission», Radio Science (in press).
- [20] Nogués, O., «Disseny, muntatge i validació d'un receptor/gravador de senyals GPS per a la demostració del concepte PARIS», Proyecto Final de Carrera, ETSETB, 2002.
- [21] M. Bara, «Airborne SAR Interferometric Techniques for mapping applications», Tesis Doctoral, Septiembre 2000.
- [22] J. Mallorquí, M. Bara, A. Broquetas, M. Wis, A. Martínez, L. Nogueira, V. Moreno, «Parallel Algorithms for High-Speed SAR Processing», EUROPTO'98, September 1998.
- [23] Bara, M.; Prats, P.; Mallorqui, J.J.; Lopez, J.; «Validation of SAR interferometric phase corrections in the squinted case for DEM quality enhancement» Electronics Letters, Volume: 38 Issue: 22, Oct 2002 Page(s): 1380 -1381
- [24] D. Carrasco, «SAR Interferometry for Digital Elevation Model Generation and Differential Applications», PhD Dissertation, March 1998.
- [25] D. Carrasco, D. Esteban, F. López, J. Tena, O. Rodríguez, M. Bara, A. Broquetas, «Precise Geometry Techniques for ERS SAR Interferometry», European Conference on Synthetic Aperture Radar, EUSAR'98, May 1998.
- [26] Mora, O.; Mallorqui, J.J.; Duro, J., «Generation of deformation maps at low resolution using differential interferometric SAR data», Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International, Volume: 5 , 2002, Page(s): 2696 -2698.
- [27] Mora, O.; Lanari, R.; Mallorqui, J.J.; Berardino, P.; Sansosti, E., «A new algorithm for monitoring localized deformation phenomena based on small baseline differential SAR interferograms», Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International, Volume: 2, 2002, Page(s): 1237 -1239.
- [28] O. Mora, J.J. Mallorquí, A. Broquetas, «Linear and Non-Linear Terrain Deformation Maps from a Reduced Set of Interferometric Images», en proceso de revisión en IEEE Transactions on Geoscience and Remote Sensing.
- [29] IMPAST web page <http://intelligence.jrc.cec.eu.int/fish/index.htm>
- [30] Mallorqui, J.J.; Rius, J.M.; Bara, M., «Simulation of polarimetric SAR vessel signatures for satellite fisheries monitoring», Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International, Volume: 5, 2002, Page(s): 2711 -2713
- [31] L. Sagués, M. Bara, A. Broquetas, X. Fàbregas, «An Interferometric Circular Synthetic Aperture Radar (InCSAR) System for Ground-Based Applications», NATO RTO-SET Symposium on High Resolution Radar Techniques, March 1999.
- [32] L. Sagués, J. M. López, J. Fortuny, X. Fàbregas, A. Broquetas, Sieber, «Wide-Band Polarimetric SAR Interferometry for Buried Mine Detection», IEEE Transactions on Geoscience and Remote Sensing, Vol. 39, No. 6, pp.1271-1278, June 2001.
- [33] Rocadenbosch, F. (1996). «Lidar Sensing of the Atmosphere: Receiver Design and Inversion Algorithms for an Elastic System.» Tesis doctoral. Dep. Teoría de la Señal y Comunicaciones, Universidad Politécnica de Catalunya, UPC.
- [34] Rocadenbosch, F., Comerón, A., Pineda, D. (1998), «Assessment of lidar inversion errors for homogeneous atmospheres,» *Appl. Opt.* 37 (12), 2199-2206.
- [35] Rocadenbosch, F., Comerón, A., (1999), «Error Analysis For The Lidar Backward Inversion Algorithm,» *Appl. Opt.* 38(21), 4461-4474.
- [36] Rocadenbosch, F., Comerón, A., Albiol, L., (2000), «Statistics Of The Slope-Method Estimator,» *Appl. Opt.* 39(33), 6049-6057.
- [37] Rocadenbosch, F., Vázquez, G., Comerón, A. (1998), «Adaptive Filter Solution For Processing Lidar Returns: Optical Parameter Estimation,» *Appl. Opt.* 37 (30), 7019-7034.
- [38] Matthias, V.; Böckmann, C.; Freudenthaler, V.; Pappalardo, G.; Bösenberg, J.; Amiridis, V.; Amodeo, A.; Ansmann, A.; Balis, D.; Boselli, A.; Chaykovski, A.; Chourdakis, G.; Comerón,

- A.; Delaval, A.; De Tomasi, F.; Eixmann, R.; Frioud, M.; H⁻g⁻rd, A.; Iarlori, M.; Komguem, L.; Kreipl, S.; Larcheveque, G.; Linné, H.; Matthey, R.; Mattis, I.; Papayannis, A.; Pelon, J.; Perrone, R. M.; Persson, R.; Resendes, D. P.; Rizi, V.; Rocadenbosch, F.; Rodriguez, J. A.; Sauvage, L.; Schneider, J.; Schumacher, R.; Shcherbakov, V.; Simeonov, V.; Wandinger, U.; Wang, X.; Wiegner, M.; Zerefos, C. (2002), «Lidar Intercomparisons on algorithm and system level in the frame of EARLINET,» Max-Planck-Institut für Meteorologie, Hamburg, Germany, ISSN 0937-1060, 67 pp.
- [39] Rocadenbosch, F.; Soriano, C.; Comerón, A.; Baldasano, J. M.; Rodríguez, A.; Muñoz, C.; García-Vizcaíno, D. (2001), «3D Scanning Portable Backscatter Lidar Platform for Atmospheric Remote Sensing: Performance and Architecture Overview.» *Proc. EUROPTO 2000*, Barcelona, Sept. 26-28, 2000; Russell, J.E., Schäfer, K., Lado-Bordowsky, O., Eds.; SPIE, Bellingham, Washington, 2001; 158-169. Proc. 4168.
- [40] Bösenberg, J.; Ansmann, A.; Baldasano, J. M.; Balis, D.; Böckmann, C.; Calpini, B.; Chaikovsky, A.; Flamant, P.; Hägård., A.; Mitev, V.; Papayannis, A.; Pelon, J.; Resendes, D.; Schneider, J.; Spinelli, N.; Trickl, T.; Vaughan, G.; Visconti, G.; Wiegner, M. (2000), «EARLINET: A European aerosol research lidar network.» *Proc. 20th International Laser Radar Conference (ILRC)*, Vichy (France), July 10-14, 2000; Pelon, J., Dabas, A., Loth, C., Eds.; IPSL (Institut Pierre Simon Laplace), 2000; Session 2 (Aerosol Properties and Radiation Budget (ARB)), paper 4.
- [41] Soriano, C., Rocadenbosch, F., Rodríguez, A., Muñoz, C., García-Vizcaíno, D., Baldasano, J.M., Comerón, A., (2001), «Barcelona Atmospheric Monitoring with Lidar: first measurements with the UPC's scanning portable lidar» *Proc. EUROPTO 2001*, Barcelona, Sept. 26-28, 2000; Russell, J. E., Schäfer, K., Lado-Bordowsky, O., Eds.; SPIE, Bellingham, Washington, 2001; 158-169. Proc. 4168.
- [42] Papayannis, A., Boselli, A., Calpini, B., Chaikovsky, A., Chourdakis, G., Cuomo, V., Frioux, M., Iarlori, M., Kreipl, S., Larcheveque, G., Matthey, R., Pappalardo, G., Pelon, J., Perrone, M., Rizi, V., Rocadenbosch, F., Sauvage, V., Sobolewski, P., Soriano, C., Spinelli, N., Tomasi, F.D., Amoiridis, V., Balis, D., Trickl, T., (2001), «Simultaneous observations of free tropospheric Saharan dust layers over Europe monitored by a co-ordinated ground-based lidar network in frame of the EARLINET project», *Journal of Aerosol Science* 32, 389-390.
- [43] Sicard, M., Rocadenbosch, F., López, M.A., Comerón, A., Rodríguez, A., Muñoz, C., García-Vizcaíno, D. (2003), «Characterization of aerosol backscatter-to-extinction ratio from multi-wavelength and multi-angular lidar profiles.» *Proc. SPIE*, European Symposium on Remote Sensing. Remote Sensing of Clouds and the Atmosphere VII (EUROPTO 2002), In press.
- [44] Klett, J.D. (1986), «Extinction boundary value algorithm for lidar inversion», *Appl. Opt.* 25(15), 2462-2464.
- [45] Rocadenbosch, F., Sicard, M., Comerón, A., Baldasano, J.M., Rodríguez, A., Agishev, R., Muñoz, C., López, M.A.. García-Vizcaíno, D. (2002), «The UPC Scanning Raman Lidar: An engineering overview.» *Proc. ILRC21*, International Laser Radar Conference, Québec (Canada), Jul. 8-12, 2002; Luc R. Bissonnette, Gilles Roy, and Gilles Vallée, Eds.; Library Services, Defence R&D Canada –Valcartier; Session on New Lidar Technologies and Methods, pp. 353-356.
- [46] Ansmann, A.; Wandinger, U.; Riebesell, M.; Weitkamp, C.; Michaelis, W. (1992), «Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar,» *Appl. Opt.* 31(33), 7113-7131.
- [47] Böckmann, C., Wandinger, U., Ansmann, A., Bösenberg, J., Amiridis, V., Boselli, A., Delaval, A., Tomasi, F.D., Frioud, M., Iarlori, M., Komguem, L., Kreipl, S., Larcheveque, G., Matthias, V., Papayannis, A., Rocadenbosch, F., Schneider, J., Shcherbakov, V., Wiegner, M., (2001), «EARLINET-Lidar Algorithm Intercomparison,» *J. Aerosol Science* 32, 433–434.
- [48] Rodríguez, A., Comerón, A., González, E., Elías, A., (1995), «Traffic monitoring: The coherent laser radar approach», *Microwave Eng. Europe*, pp. 29-34.
- [49] Rodríguez, A., Comerón, A., García, D., (2001), «Homodyne laser radar system for surface displacement monitoring», *Opt. Eng.* 40(3), 398-405.
- [50] Rodríguez, A., García, D., Comerón, A., Dios, F., Rocadenbosch, F., (2001), «Self-aligned heterodyne laser radar system for surface displacement monitoring». *Proc. EUROPTO 2001*, Barcelona, Sept. 25, 2000; Ginglewski, J. D., Kamerman, G., Kohnle, A., Schreiber, U., Werner, C., Eds.; SPIE, 2001, 221-229. Proc. 4167.
- [51] Marx, B.R. (1994), «New devices close in on environmental pollution,» *Laser Focus World*, 30(8), S11-S14.[33] Papayannis, A., Boselli, A., Calpini, B., Chaikovsky, A., Chourdakis, G., Cuomo, V., Frioux, M., Iarlori, M., Kreipl, S., Larcheveque, G., Matthey, R., Pappalardo, G., Pelon, J., Perrone, M., Rizi, V., Rocadenbosch, F., Sauvage, V., Sobolewski, P., Soriano, C., Spinelli, N., Tomasi, F.D., Amoiridis, V., Balis, D., Trickl, T., (2001), «Simultaneous observations of free tropospheric Saharan dust layers over Europe monitored by a co-ordinated ground-based lidar network in frame of the EARLINET project», *Journal of Aerosol Science* 32, 389–390.



AUTORES



Adriano Camps nació en Barcelona en 1969. Es Ingeniero en Telecomunicaciones (1992) y Doctor Ingeniero en Telecomunicaciones (1996) por la Universitat Politècnica de Catalunya, Barcelona. En 1991-1992 estuvo en la ENS des Télécommunications de Bretagne, France con una beca Erasmus, realizando el Proyecto Final de Carrera en la empresa

Eurowaves-Sorep (hoy perteneciente a Thalkes), Rennes. En 1992-1993 trabajó para el Badalona Centre International de Negocios planificando la infraestructura de comunicaciones. En 1993 inició su carrera académica en el Departamento de Teoría de la Señal y Comunicaciones de la Universitat Politècnica de Catalunya (UPC) como Profesor Asociado (1993-1997) y como Profesor Titular (1997-actualidad). Durante 1999 estuvo en el Microwave Remote Sensing Laboratory, de la Universidad de Massachusetts. Su investigación se ha centrado principalmente en la teledetección pasiva en microondas, y más concretamente en el proyecto SMOS/MIRAS. Es miembro Senior del IEEE Geoscience and Remote Sensing Society (GRS), Presidente del Capítulo Español del IEEE GRS-S, desde Enero de 2003 Editor del IEEE GRS-S Newsletter y desde Enero de 2000 Editor Asociado de la Revista Radio Science. En 1993, recibió el Segundo premio Nacional de Terminación de Estudios Universitarios, en 1997 el Premio INDRA del COIT a la mejor Tesis doctoral en Teledetección, en 1999 el Premio Extraordinario de Doctorado de la UPC, y junto con el resto de los miembros del grupo de radiometría de microondas, en 2000 el Primer Premio Duran Farell de Investigación Tecnológica, y en 2001 el Primer Premio Ciutat de Barcelona d'Investigació Tecnològica, y en 2002 la «Distinció de Recerca de la Generalitat de Catalunya».



Mercè Vall-llossera nació en Lleida en 1966. Recibió los títulos de ingeniera en telecomunicaciones y Doctora en ingeniería de telecomunicaciones por la Universidad Politécnica de Cataluña, en los años 1990 y 1994, respectivamente. Es profesora la Escuela Técnica Superior de Telecomunicaciones de Barcelona desde noviembre

de 1990 primero como profesora asociada, en 1996 obtuvo la plaza de Profesora Titular de Escuela Universitaria y en 1997 la de Profesora Titular de Universidad. En 1996 obtuvo la beca postdoctoral Programme Québécoise de Bourses d'Excellence: Stages de Formation Postdoctorale au Québec pour jeunes diplômés étranger, por la que estuvo trabajando en la Universidad de Concordia (Montreal, Canadá), durante 9 meses. Su investigación estuvo dedicada inicialmente en los métodos numéricos aplicados al análisis de la sección recta radar y el diseño de antenas. En la actualidad está centrada en la radiometría por microondas. Junto con

todos los demás miembros del grupo de radiometría de microondas, en el año 2000 recibió el primer Premio Duran Farell de Investigación Tecnológica, y en 2001 el Primer Premio Ciutat de Barcelona d'Investigació Tecnológica.



Jordi J. Mallorquí nació en Tarragona en 1966. Ingeniero de Telecomunicación (1990) y Doctor Ingeniero (1995) por la Universitat Politècnica de Catalunya inició en 1993 su carrera académica en el Departamento de Teoría de la Señal y Comunicaciones de la UPC como Profesor Asociado (1993-1997) y posteriormente como Profesor Titular (1997-actualidad) asignado a la Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona. Inicialmente su investigación se centró en técnicas de formación de imágenes para aplicaciones biomédicas mediante tomografía de microondas. En 1997 recibió el Premio Extraordinario de Doctorado en Ingeniería de Telecomunicación de la UPC. Durante 1999 realizó una estancia de 11 meses en el Jet Propulsion Laboratory (JPL-NASA) trabajando en calibración de sistemas SAR aerotransportados. Actualmente su investigación se centra en el desarrollo de algoritmos de procesado SAR, interferometría diferencial para el control de subsidencias y simulación de sistemas SAR. Miembro del IEEE Geoscience and Remote Sensing Society (GRS) es autor de numerosas publicaciones internacionales y revisor del IEEE y del IEE.



Francesc Rocadenbosch nació en Barcelona en 1967. Ingeniero Superior en Telecomunicación (1991) y Doctor Ingeniero (1996) por la Universidad Politécnica de Cataluña inició su carrera profesional en 1991, en la Universidad de Las Palmas de Gran Canaria, donde colaboró en el diseño de sistemas de alta frecuencia y microondas, y posteriormente, en 1993, la continuó como profesor del Departamento de Teoría de la Señal y Comunicaciones (UPC), de la que actualmente es profesor Titular de Universidad desde 1997. Premio al mejor Proyecto de Investigación en Cataluña (ACET, 1999) y máster en Administración y Dirección de Empresas (MBA) por la Universidad de Barcelona (2001), desde 1993 ha desarrollado una intensa actividad investigadora en el campo de los radares láser y la optoelectrónica a nivel internacional, donde ha participado en más de una decena de proyectos. Miembro de la Optical Society of America, es autor de numerosas publicaciones internacionales y revisor de diversas revistas de ingeniería óptica y procesado de la señal. Sus intereses actuales comprenden el diseño de sistemas de radar láser (lidar) de baja energía, receptores de altas prestaciones y técnicas de procesado en teledetección láser.



NAVEGANDO CON SATÉLITES

(O CÓMO POSICIONARSE CON GPS Y GALILEO CON ERRORES DE POCOS CENTÍMETROS, EN TIEMPO-REAL Y A ESCALA CONTINENTAL)

Manuel Hernández-Pajares, J.Miguel Juan, Jaume Sanz

Dept. de Matemática Aplicada IV y Física Aplicada, Grupo de Astronomía y Geomática,
Univ. Politécnica de Catalunya (gAGE/UPC)
E08034-Barcelona
(e-mail de contacto: manuel@mat.upc.es)

Los sistemas de navegación por satélite constituyen un avance tecnológico que sin duda no tardará en sernos muy familiar. En particular el «Sistema de Posicionamiento Global», conocido corrientemente como GPS, ha llegado por ejemplo a expandir el concepto de reloj, de modo que en estos momentos ya es posible adquirir relojes de pulsera que junto a la hora proporcionan de forma continua las coordenadas geográficas tridimensionales. Estos «relojes», constituidos como núcleos de mapas inteligentes, están llamados a facilitar y guiar el recorrido en automóvil por rutas óptimas.

El sistema GPS fue concebido hace cosa de treinta años por el Pentágono para uso militar. Dispone de una constelación de más de veinticuatro satélites que emiten señales a modo de radiofaros y que permiten navegar en todo momento, y circunstancia metereológica con un error del orden de varios metros. Cabe decir que estos satélites están orbitando alrededor de la Tierra a una altura media de 20200 kilómetros sobre la superficie terrestre y con un periodo de unas 12 horas (dando dos vueltas cada día). Por otro lado, el cálculo de la posición se obtiene resolviendo un sencillo problema geométrico de triangulación, donde a partir de las distancias de al menos cuatro satélites GPS, medidas por el receptor (a partir de la señal emitida por los mismos) y de sus órbitas, se puede determinar la posición tridimensional del usuario en cualquier lugar del planeta.

En los años ochenta, la ciencia «civil» empezó a desarrollar con éxito nuevas estrategias de procesado de la señal GPS que permitieron un posicionamiento con errores centimétricos, y que se han aplicado de inmediato en ingeniería y topografía, así como en la monitorización de los movimientos de placas tectónicas (esto es, el movimiento o deriva de los continentes). Desde entonces son muchos y significativos los avances que se vienen obteniendo gracias a este sistema. Entre ellos está su uso como un gran «scanner» planetario que permite reconstruir tomográficamente la atmósfera terrestre, en concreto la parte superior llamada Ionosfera, que es la que más afecta a la señal GPS.

En el congreso del 'Institute of Navigation' (ION-GPS99), uno de los más referenciados en su campo, celebrado en 1999 en Nashville (Estados Unidos), demostramos que con GPS es factible conocer con un error inferior a los diez centímetros la posición en tiempo real (es decir, inmediatamente) de cualquier móvil situado a centenares de kilómetros de la estación de referencia más cercana. Todo ello incluso en el caso de tormentas geomagnéticas, cuando la Ionosfera es perturbada notablemente por la irrupción de partículas muy energéticas del viento solar.

Este avance no se podría comprender plenamente si se ignora la tensión de intereses contrapuestos entre militares y civiles en los Estados Unidos. Mientras los primeros procuran restringir la precisión a los usuarios no autorizados, los segundos intentan lo contrario. Paradójicamente, todos se nutren de la misma fuente de fondos públicos.

En este contexto se desarrolló hace más de diez años el concepto de 'Differential GPS' (DGPS), que consiste en posicionar un receptor GPS respecto a otro que se toma como referencia y del que se conocen perfectamente sus coordenadas. De esta forma se consigue cancelar buena parte de los errores, básicamente la parte común de propagación, errores en las órbitas y relojes de los satélites.

Con la introducción del DGPS, y mediante la técnica denominada 'Real Time Kinematics' (RTK) o alternativamente, dependiendo del contexto, 'Local Area DGPS' (LADGPS), se consiguen precisiones de unos cuantos centímetros para distancias inferiores a 10 kilómetros de la estación de referencia. A escalas mucho más amplias, de

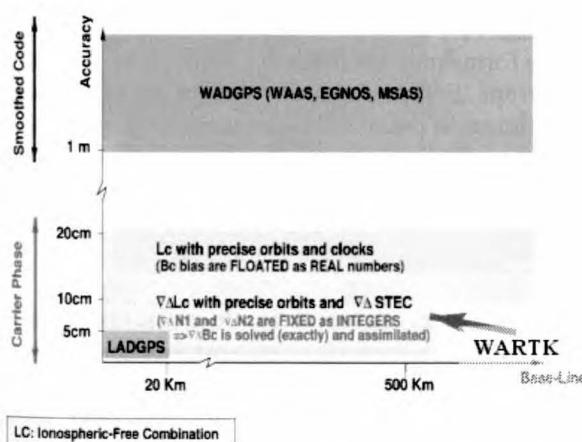


Figura 1: Exactitud en el posicionamiento frente a distancia a la estación de referencia más cercana, para distintas



unos 1000 kilómetros, se ha establecido el sistema ‘Wide Area DGPS’ (WADGPS) que permite un posicionamiento con errores del orden del metro para usuarios equipados con receptores adecuados.

La estrategia citada anteriormente —elaborada por el grupo Astronomía y Geomática de la Universidad Politécnica de Cataluña (gAGE/UPC), en colaboración con el Dr. Oscar Colombo (NASA)— permite extender las precisiones a escala local (LADGPS) a escalas continentales (WADGPS), tal como se puede ver en la figura 1.

El éxito de la nueva técnica (WARTK, “Wide Area Real Time Kinematics”, patentada por UPC) radica en una predicción muy exacta de los efectos ionosféricos (alta atmósfera) sobre la propagación de la señal GPS. Mediante técnicas tomográficas similares a las usadas en medicina, en conjunción con el modelado muy preciso de los observables geodésicos, se reconstruye la distribución tridimensional de electrones libres. Los receptores de tierra, junto a la constelación de satélites GPS, forman un enorme escáner planetario que sondea continuamente la atmósfera y permite reconstruir, en tiempo real (inmediatamente), su distribución electrónica tridimensional.

Así, a partir de un conjunto de estaciones de referencia que recogen los haces de rayos procedentes de los satélites, y que pueden estar separadas cente-

posicionamiento, esta cantidad constituye un valor límite para modelar con una exactitud superior al centímetro las pseudodistancias receptor-satélite con las que se resuelve el problema de navegación.

Una de las virtudes de la estrategia de navegación presentada es su buen comportamiento en condiciones de alta actividad solar. Actualmente nos encontramos todavía en esta situación, después del máximo alcanzado en los años 2000-2001. La variación de esta actividad, que se repite cíclicamente cada once años —el llamado ciclo solar— está jugando un papel importante en la puesta al día de los resultados en posicionamiento GPS de precisión. La época de mayor efervescencia en el desarrollo de este tipo de aplicaciones ha coincidido con un periodo de baja actividad solar. A medida que ésta ha ido creciendo, buena parte de los modelos y algoritmos que involucraban a la Ionosfera y en su momento funcionaban, están dejando de funcionar.

El modelo tomográfico se ha puesto a prueba bajo diferentes condiciones de actividad geomagnética y perturbaciones (que junto con la el máximo del ciclo solar constituyen un auténtico talón de Aquiles de los modelos ionosféricos). Para ello se han utilizado datos procedentes de estaciones permanentes y móviles GPS en diferentes latitudes. Buena parte de los receptores permanentes corresponden a la red de estaciones que desde principios de los años noventa mantiene el ‘International Geodynamical Service for GPS’ (IGS) (se trata de una extensa base de datos con receptores distribuidos a lo largo de todo el planeta, y que son de libre distribución para la comunidad científica y técnica).

En el ámbito de las aplicaciones civiles, un posicionamiento en tiempo real con exactitudes de unos centímetros, a escala continental, puede simplificar las tareas y abaratar costes en actividades de ingeniería y topografía, además de abrir nuevas aplicaciones insospechadas en el campo del transporte. Por ejemplo para la aviación civil, genera expectativas de cara a la posibilidad de proporcionar de forma muy económica a miles de aeródromos de Europa, Estados Unidos y el Japón, categoría de aproximación precisa III, que supondría, por ejemplo, una exactitud «garantizada» en la coordenada vertical mejor que 60 cm. Queda pendiente sin embargo en esta aplicación cómo garantizar la alta integridad requerida en aviación civil.

Efectivamente, un servicio de posicionamiento subdecimétrico a escala continental puede ser factible utilizando una técnica como WARTK. Baste considerar que para poder proporcionar tales servicios con la técnica anterior RTK, válida para distancias inferiores a unos 10 kilómetro de la estación de referencia más cercana, harían falta en

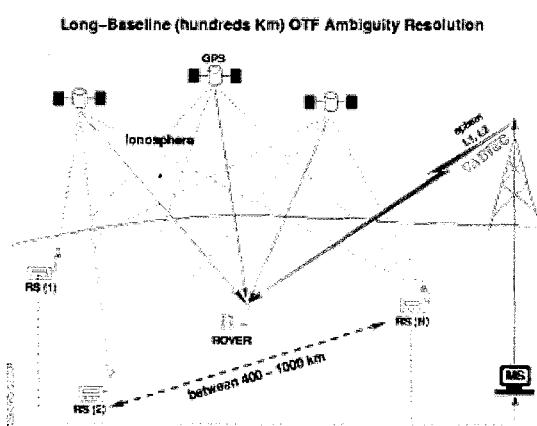


Figura 2: Las correcciones WARTK, en particular las ionosféricas, se han de distribuir a los usuarios junto a las observaciones de los satélites.

nares de kilómetros (figura 2), se consigue modelar el efecto de la ionosfera en cada observación a partir de la estimación aproximada de la distribución tridimensional de los electrones libres que interactúan con las señales emitidas por los satélites GPS. Esto permite estimar en cualquier punto de la región de cobertura el efecto ionosférico con un error inferior a 2.7 cm; dentro del algoritmo de

un país de tamaño medio como España (unos 500000 kilómetro²) un conjunto de más de 1000 estaciones de referencia. Y ésto es poco factible, no únicamente por la inversión requerida (más de 10 millones de Euros solamente en los receptores) sino sobre todo por el esfuerzo de construcción, mantenimiento y coordinación de las más de 1000 estaciones permanentes. Y no digamos si, lógicamente, consideramos la extensión a escala continental. Por ejemplo para la Unión Europea, cerca de 10000 estaciones serían requeridas, aumentando toda la inversión y coordinación asociada en un orden de magnitud.

Ésta es la razón por la que un servicio factible de Navegación por Satélite a escala europea, con errores típicos de pocos centímetros, requiere una determinación ionosférica en tiempo real muy precisa. De esta forma se puede aumentar entre 10 y 100 veces la distancia entre estaciones de referencia, pudiéndose usar, por ejemplo, redes de GPS diferencial de Área amplia (“Wide Area DGPS”), con separaciones típicas de varios centenares de kilómetros o más. En otras palabras, WARTK permite que con sólo 20-30 estaciones (en vez de unas 10000) se pueda dar servicio de navegación

subdecimétrica en toda Europa. Y además estas redes ya se están construyendo con otro objetivo: el de proporcionar una navegación con errores de pocos metros y con integridad, que cumpla los estándares de la aviación civil, su principal aplicación (EGNOS en Europa, WAAS en Estados Unidos, MSAS en Japón, etc...).

Actualmente, la capacidad mencionada de generar correcciones ionosféricas diferenciales precisas a larga distancia, está siendo aplicada por los autores de este artículo en dos proyectos: (1) Desarrollo de una nueva técnica de navegación subdecimétrica instantánea (“single epoch mode”) con las futuros sistemas de 3 frecuencias Galileo y GPS modernizado (WARTK-3, recientemente patentada por ESA). (2) Análisis de la calidad de las correcciones ionosféricas distribuidas para el uso de agencias de aviación civil (como EUROCONTROL en Europa).

En este contexto podemos mencionar asimismo un proyecto internacional que involucra a varias agencias con el objetivo primordial de generar cada 2 horas mapas globales de correcciones ionosféricas. Éste proyecto está realizándose ininterrumpidamen-

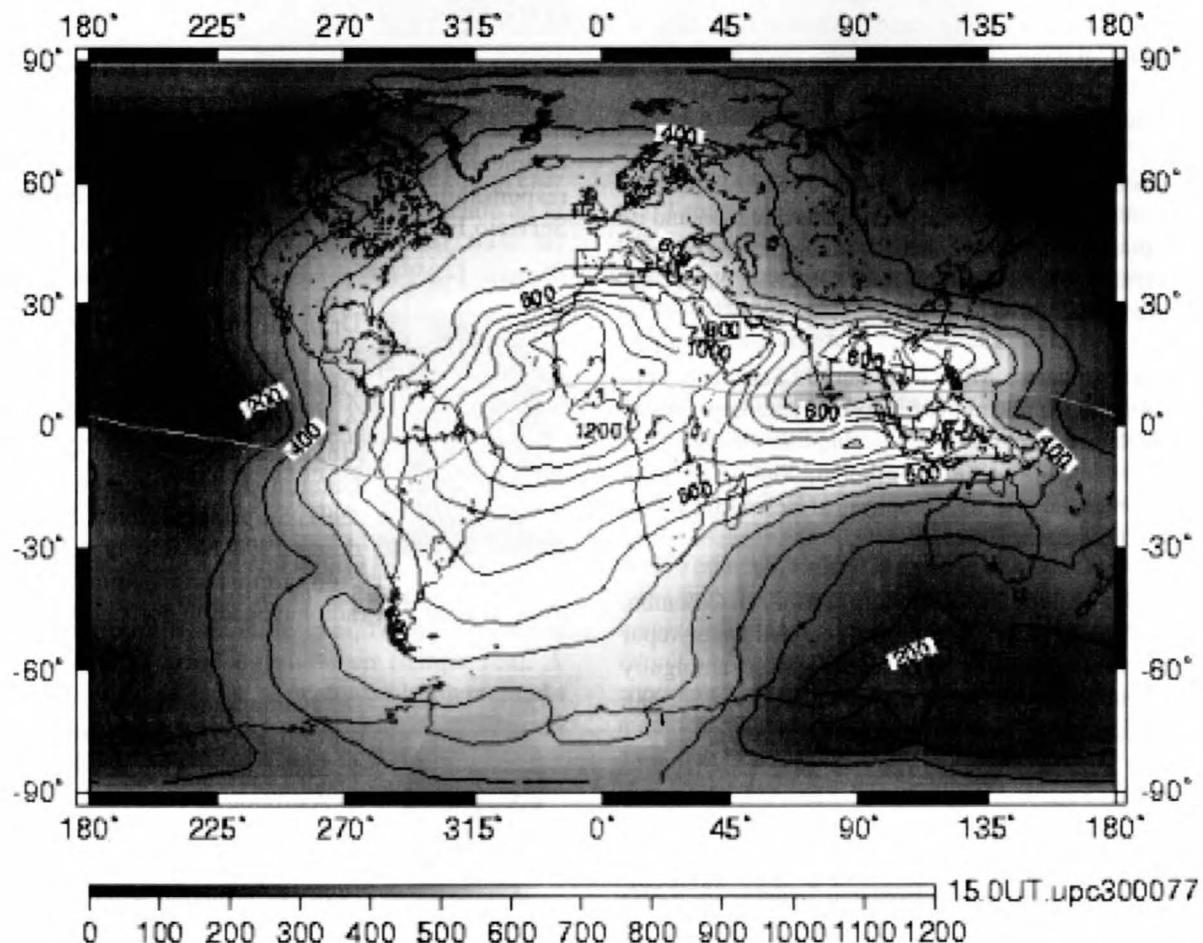


Figura 3: Ejemplo de mapa global de correcciones ionosféricas, calculado por UPC a partir de datos de unas 150 estaciones GPS distribuidas en todo el mundo (17 de Marzo de 2002, 15 UT). Las unidades son décimas de TECU (1 TECU = 0,16 cm en L1, aproximadamente).

te desde 1998, e involucra a cinco centros de análisis del Servicio GPS Internacional (IGS): CODE (Univ. Berna), ESA (Agencia Espacial Europea), JPL (Jet Propulsión Lab), NRCAN (Nacional Research Council of Canada) y UPC (los autores).

Estos mapas (véase un ejemplo en figura 3), y muy pronto su combinación como producto oficial IGS, están disponibles en <ftp://cddisa.gsfc.nasa.gov/pub/gps/products/ionex/YEAR/DOY> (en mayúsculas el año y día del año correspondientes). Uno de los próximos objetivos de las actividades ionosféricas IGS será calcular tales mapas globales de la ionosfera casi instantáneamente, de tal forma que puedan beneficiarse tanto aplicaciones científicas como tecnológicas.

Parece claro que nuestro futuro, en el que Europa está apostando fuertemente con el sistema GALILEO, está cada vez más en manos de la electrónica y de los satélites, los cuales a pesar de orbitar velozmente y a gran distancia, nos dejan saber «dónde estamos» con un precisión inusitada.

MÁS INFORMACIÓN EN:

Colombo, O.L., M. Hernández-Pajares, J.M. Juan, J. Sanz and J. Talaya, Resolving carrier-phase ambiguities on-the fly, at more than 100 kilómetro from nearest site, with the help of ionospheric tomography, ION GPS'99, Nashville, USA, September 1999.

Hernández-Pajares M., J.M. Juan, J. Sanz and O.L. Colombo, Precise ionospheric determination and its application to real-time GPS ambiguity resolution, Institute of Navigation ION GPS'99, Nashville, Tennessee, USA, September 1999.

Hernández-Pajares, M., J.M. Juan, J. Sanz and O.L. Colombo, Application of ionospheric tomography to real-time GPS carrier-phase ambiguities resolution, at scales of 400-1000 kilómetro, and with high geomagnetic activity, Geophysical Research Letters, 27, 2009-2012, 2000a.

Hernández-Pajares, M., J.M. Juan, J. Sanz, O. Colombo, H. Van der Marel, Real-time integrated water vapor determination using OTF carrier-phase ambiguity resolution in WADGPS networks, ION GPS'2000, Salt Lake City, September 2000b.

Hernández-Pajares, M., J.M. Juan, J. Sanz, O.L. Colombo, and H. van der Marel, A new strategy for real-time Integrated Water Vapour determination in WADGPS networks, Geophysical Research Letters, 28, 3267-3270, 2001a.

Hernández-Pajares, M., J.M. Juan, J. Sanz, O.L. Colombo, Tomographic modeling of GNSS ionospheric

corrections: Assessment and real-time applications, ION GPS'2001, Salt Lake, USA, September 2001b.

Hernández-Pajares, M., J.M. Juan, J. Sanz, O.L. Colombo, Improving the real-time ionospheric determination from GPS sites at Very Long Distances over the Equator, Journal of Geophysical Research, 107(A10), 1296, doi:10.1029/2001JA009203, 2002.

AGRADECIMIENTOS:

Los autores desean expresar su agradecimiento al profesor de la UPC, Miquel Escudero, por la revisión estilística realizada en una versión anterior de este trabajo.

BREVES BIOGRAFÍAS:



Dr. Manuel Hernández-Pajares es profesor titular de Universidad del Dept. de Matemática Aplicada IV, grupo de Astronomía y Geomática, en la Universidad Politécnica de Catalunya, en Barcelona, desde 1993. Empezó a trabajar con sistemas de navegación por satélite (GPS) y aplicaciones geodésicas en 1989. Su principal tema de trabajo desde 1994 ha sido la determinación atmosférica (ionosférica y troposférica) y la radionavegación por satélite. Actualmente comparte sus obligaciones docentes con la responsabilidad de las actividades ionosféricas del Servicio Internacional GPS (IGS).



Dr. J. Miguel Juan Zornoza es profesor titular de Universidad del Dept. de Física Aplicada, grupo de Astronomía y Geomática, en la Universidad Politécnica de Catalunya, en Barcelona, desde 1991. Su principal tema de trabajo desde 1994 ha sido la determinación atmosférica (ionosférica y troposférica) y la radionavegación por satélite.



Dr. Jaume Sanz Subirana es profesor titular de Universidad del Dept. de Matemática Aplicada IV, grupo de Astronomía y Geomática, en la Universidad Politécnica de Catalunya, en Barcelona, desde 1988. Su principal tema de trabajo desde 1996 ha sido la determinación atmosférica (ionosférica y troposférica) y la radionavegación por satélite.



COMPRESIÓN ANALÓGICA DE AUDIO EN TIEMPO REAL

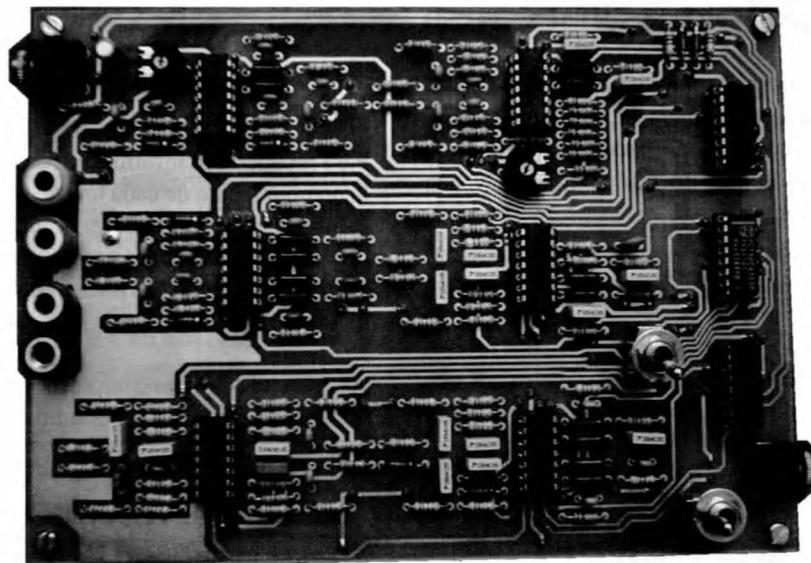
Marcelo Castro Chamorro

Ingeniero Técnico de Telecomunicaciones por la EUP de Vilanova i la Geltrú (UPC)

Proyectista de Ingeniería Electrónica en la ETS d'Enginyeria de Telecommunicació de Barcelona (UPC)

Profesor asociado del Departament de Teoria del Senyal i Comunicacions de la UPC

marcelo@tsc.upc.es <mailto:marcelo@tsc.upc.es>



Resumen- Frente a las realizaciones habituales de compresores de audio digitales, en este trabajo se presenta una propuesta novedosa de un compresor analógico para audio en tiempo real, inspirado en la filosofía de enmascaramiento frecuencial, base del formato de compresión de sonido MPEG-1 layer-3, conocido comúnmente como MP3.

El sistema que se presenta se generó como solución al trabajo propuesto por el profesor Alberto Poveda, del Departamento de Ingeniería Electrónica, para la asignatura “Electrónica para audio, vídeo y comunicaciones”, impartida en la ETSETB.

Se comienza con la descripción de los sistemas propuestos para implementar la codificación/compresión y la decodificación/descompresión de forma analógica de la señal de audio en tiempo real. A continuación se describe el prototipo implementado como proyecto para dicha asignatura, que ha sido una primera aproximación al diseño de todo el sistema, y que tenía como objetivo comprobar la viabilidad de éste en función de la inteligibilidad de la señal resultante.

1. INTRODUCCIÓN: COMPRESOR/DESCOMPRESOR ANALÓGICO DE AUDIO

Son muchos y variados los sistemas de codificación digital de la señal de audio, entre ellos los basados en

efectos psicoacústicos, que tienen en cuenta la capacidad auditiva del oído humano (principalmente las limitaciones de éste) para conseguir tasas de compresión elevadas [1]. Tomando como base de trabajo precisamente estas limitaciones, se puede pensar en la manera de implementar analógicamente y trabajando en tiempo real (esto es, sin necesidad de memorizar información) un sistema que sea capaz de realizar esta compresión.

Así, el sistema implementado se basa en el efecto de enmascaramiento frecuencial: toda componente frecuencial de una señal que se encuentre por debajo de un determinado nivel con respecto al resto de componentes frecuenciales de la misma señal no será percibida por el oído humano, es decir, a todos los efectos es como si no estuviera, de manera que se puede eliminar. De esta manera, el sistema debe trabajar dividiendo el espectro de interés en varias bandas y realizando comparaciones entre sub-bandas, eliminando aquéllas que en cada momento presenten un nivel inferior con respecto a sus bandas más cercanas.

Conceptualmente, lo que se lleva a cabo es la selección de unas bandas en detrimento de otras, y todo ello de modo pseudoinstantáneo (esto se justificará más adelante). Una vez realizada la selección de bandas, todas ellas deben compactarse en un intervalo frecuencial lo más reducido posible (por ejemplo, sería ideal conseguir reducir todo a unos 3 KHz para su transmisión vía cable telefónico). Es importante no perder de vista que cuanto menor sea el



ancho de banda definitivo peor calidad de sonido se obtendrá.

2. COMPRESOR - CODIFICADOR

El diagrama de bloques del sistema compresor es el que se muestra en la figura 1. En éste ya se han tenido en cuenta ciertas decisiones previas al diseño:

- La banda de interés será la comprendida entre 60 Hz y 8 KHz (las componentes superiores a este valor dentro de la banda de audio son habitualmente pequeñas).
- Se ha dividido el espectro en 7 bandas de una octava cada una.
- Cada una de las bandas se ha subdividido en 2 sub-bandas de media octava, de esta manera la lógica de control necesaria para seleccionar la sub-banda de mayor nivel en una banda se puede implementar simplemente mediante un comparador.

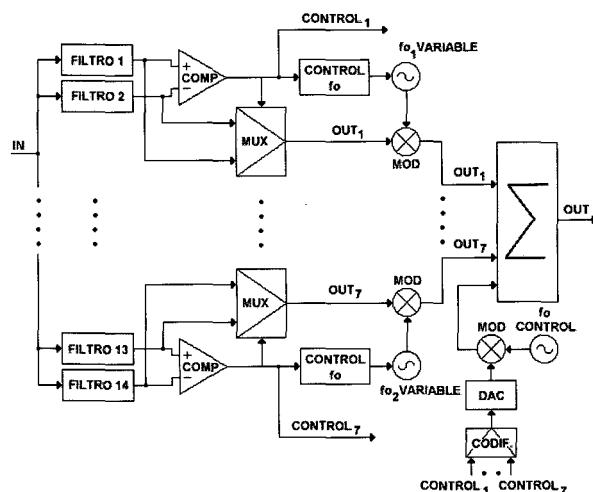


Figura 1: Compresor - codificador

Puede apreciarse como, mediante el comparador y el multiplexor, de cada una de las siete bandas se obtendrá una única sub-banda, que será modulada a una frecuencia determinada mediante un mezclador que recibirá como entrada la señal de un VCO, con objeto de que todas las bandas de la señal comprimida ocupen espectros frecuenciales contiguos. Dependiendo de la sub-banda que se conserve en cada instante y de las posibles frecuencias portadoras del resto de bandas conservadas, la lógica de control determinará un valor u otro para este VCO, que será quién fije la frecuencia portadora para cada banda.

Toda esta información de control debe ser tenida en cuenta en caso de que se desee volver a modular el resultado final de todas las bandas sumadas de cara a una posible transmisión vía radio.

3. DESCOMPRESOR - DECODIFICADOR

El proceso inverso al expuesto es el que se deberá llevar a cabo en el descompresor, un diagrama de bloques del cuál se muestra en la figura 2.

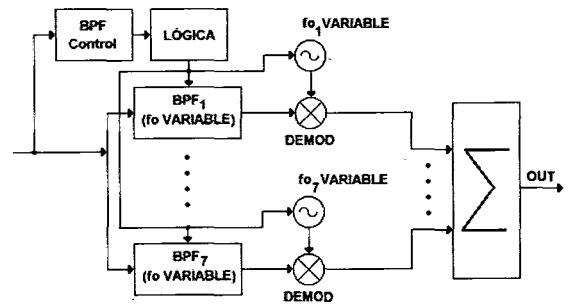


Figura 2: Descompresor - decodificador

Para recuperar la información de audio deberá realizarse la demodulación de cada una de las bandas en función de la información de control procedente del compresor. Una vez extraídas las bandas simplemente se han de sumar para reproducir la señal (con menos componentes frecuenciales que la original debido a la compresión inicial).

Así pues, el conjunto compresor-descompresor permite la transmisión de la señal de audio comprimida y en tiempo real.

4. IMPLEMENTACIÓN FÍSICA DEL PROTOTIPO

4.1. Introducción

Como ya se ha comentado, el prototipo implementado es una primera aproximación al diseño total del sistema, que incluye únicamente la selección de bandas, tal y como puede verse en el diagrama de bloques de la figura 3.

En este diagrama de bloques se realiza la eliminación de bandas frecuenciales sin compactar el espectro resultante, lo cuál permite hacerse una idea de la viabilidad del proyecto, puesto que sirve para comprobar la inteligibilidad de la señal resultante.

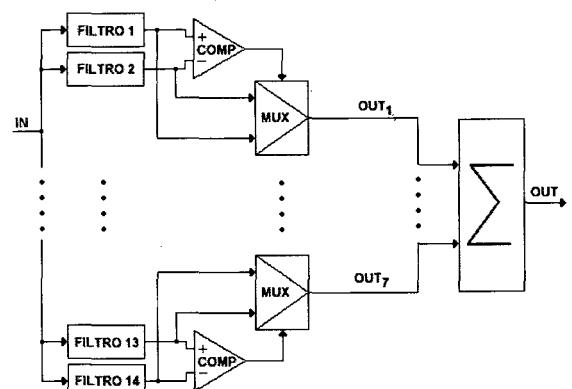


Figura 3: Diagrama de bloques del prototipo

4.2. Banco de filtros

La división frecuencial realizada puede contemplarse en la figura 4: en total hay 14 sub-bandas, lo que equivale a la utilización de 14 filtros pasa-banda.

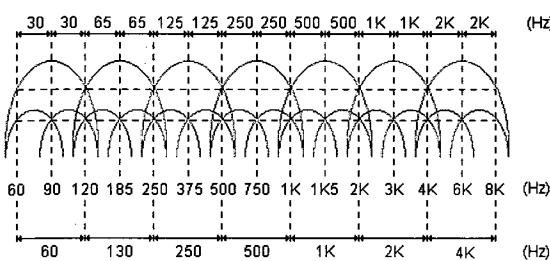


Figura 4: División frecuencial

La estructura utilizada para la implementación de los filtros, que se muestra en la figura 5, es del tipo Sallen-Key (filtros KRC) [2]. Una buena opción para simplificar el dimensionado es dar el mismo valor a los condensadores para todos los filtros y hacer iguales las resistencias de un mismo filtro (a excepción de R_a y R_b , que determinarán la ganancia K).

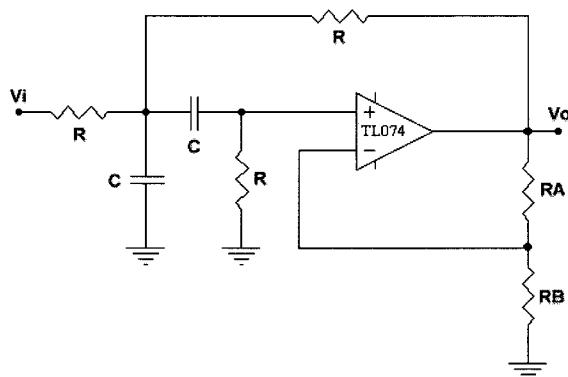


Figura 5: Filtro paso-banda de Sallen-Key o KRC

Siguiendo este criterio, el dimensionado de cada filtro obedece a las siguientes expresiones:

$$Wn = \frac{\sqrt{2}}{R \cdot C} \quad (\text{Ecuación 1: frecuencia central del filtro})$$

$$K = 1 + \frac{R_A}{R_B} \quad (\text{Ecuación 2: ganancia del filtro})$$

$$Q = \frac{\sqrt{2}}{4 - K} \quad (\text{Ecuación 3: factor de calidad del filtro})$$

Fijando el valor de C a 100 nF, cada filtro se ha dimensionado a partir de la expresión dada por la ecuación 1. Una simulación Pspice realizada con los valores calculados de los componentes se muestra en la figura 6. En la tabla 1 se dan los valores elegidos para las bandas y sus frecuencias centrales, así como los valores calculados para las resistencias de cada uno de los filtros, y las frecuencias obtenidas como consecuencia de estos valores. En las dos últimas columnas de dicha tabla se muestran los valores de las resistencias reales utilizadas finalmente en el prototipo para compensar las tolerancias de los condensadores, así como las frecuencias centrales medidas en el mismo.

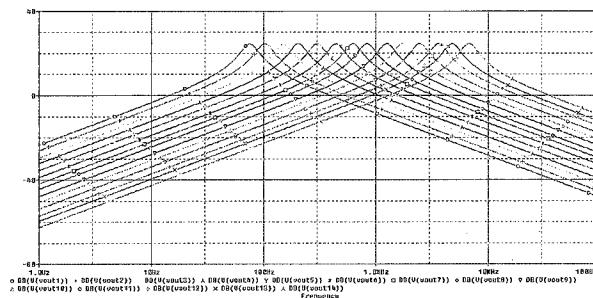


Figura 6: Simulación PSpice del banco de filtros

Nº de filtro	Intervalo frecuencial (Hz)	f_0 central (Hz)	R calculada (Ω)	f_0 simulada (Hz)	R implementada a (Ω)	f_0 medida (Hz)
1	60 - 90	75	30K	75.02	30K	73
2	90 - 120	105	22K	102.31	27K	102
3	120 - 185	152.5	15K	150.05	22K	146
4	185 - 240	212.5	11K	204.62	11K	280
5	250 - 375	312.5	6K8	330.99	4K7	325
6	375 - 500	437.5	7K5	441.33	8K2	420
7	500 - 750	625	3K6	625.22	3K6	700
8	750 - 1K	875	2K7	833.63	3K3	830
9	1K - 1.5K	1250	1K8	1250.44	2K2	1K
10	1.5K - 2K	1750	1K3	1731.37	1K3	2K
11	2K - 3K	2500	910	2473.39	910	2.52K
12	3K - 4K	3500	620	3630.31	620	3.53K
13	4K - 6K	5000	470	4788.92	470	5.2K
14	6K - 8K	7000	330	6820.58	330	7.5K

Tabla 1: Relación frecuencial y valores de resistencias del banco de filtros



El amplificador operacional utilizado para los filtros y los comparadores ha sido el TL074. El multiplexor analógico a través del cual es seleccionada una de las sub-bandas de cada una de las bandas es el 74VHC4053.

4.3. Detectores de amplitud

A la hora de realizar las comparaciones entre las salidas de los filtros se ha de tener en cuenta un aspecto fundamental: al tratarse de una comparación instantánea sin ningún nivel adicional de continua sumado a ninguna de las señales, durante el semiciclo negativo la señal de mayor amplitud será más negativa que la de menor amplitud, ocasionando que el comparador cambie de nivel a su salida de manera que el multiplexor dejará pasar la señal que no corresponde, forzando así un mal funcionamiento del sistema durante dicho semiciclo (obviamente, durante el semiciclo positivo todo es correcto). Este hecho se ilustra en la figura 7.

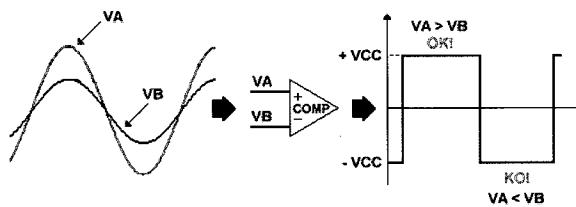


Figura 7: Problema en la comparación instantánea

Debido a esto, se debe realizar una comparación pseudoinstantánea, es decir, las señales que llegan a las entradas del comparador no son las salidas de los filtros directamente, sino que previamente se detecta su amplitud, de manera que siempre se realice la selección correcta, tal y como se muestra en la figura 8.

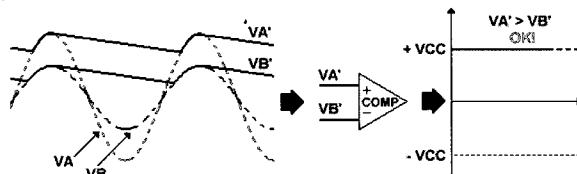


Figura 8: Solución al problema de la comparación instantánea mediante comparación con detectores de amplitud

Circuitalmente, el conjunto filtro-detector-comparador-multiplexor se muestra en la figura 9:

La constante de tiempo RC ha sido calculada en cada caso para un valor 3 veces mayor al periodo correspondiente a la frecuencia central de cada uno de los filtros, para tener así un margen de seguridad, y los diodos utilizados son de germanio, puesto que presentan una menor tensión umbral y entrarán en conducción más fácilmente.

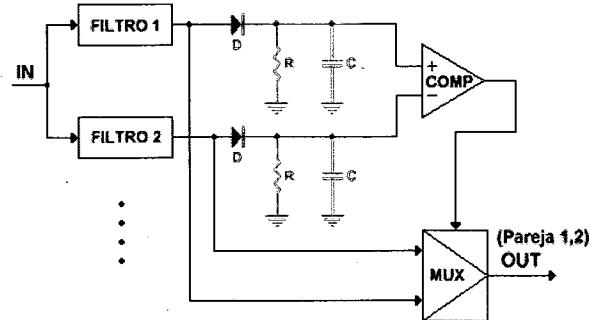


Figura 9: Conjunto filtro-detector-comparador-multiplexor

4.4. Estructura sumadora

Una vez realizada la selección de bandas, deben sumarse todas para obtener así la señal de audio comprimida resultante a reproducir. Esto se ha realizado mediante una estructura sumadora inversora clásica, como se muestra en la figura 10. A la hora del dimensionado se ha asegurado que, incluso en el peor de los casos, cuando el nivel de señal sea máximo, no se pedirá más corriente de salida a los amplificadores operacionales de la que pueden suministrar, 25 mA.

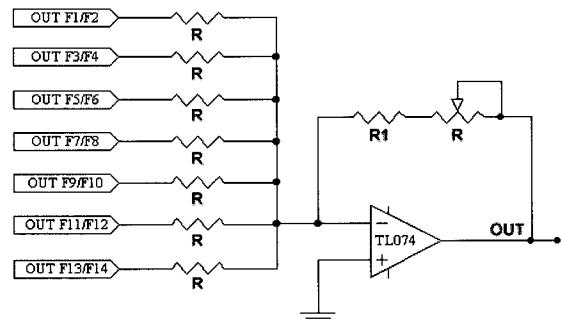


Figura 10: Estructura sumadora

4.5. Entrada y salida al sistema

Por último, solo queda mencionar cómo se ha llevado a cabo la implementación de la entrada y la salida al sistema. La primera se ha realizado mediante un seguidor de tensión con acople en AC, tal y como se muestra en la figura 11.

El condensador presenta una impedancia en módulo muy inferior, para el peor de los casos, a los 47 KW del

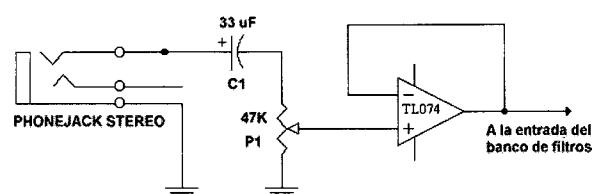


Figura 11: Seguidor de tensión con acople en alterna

potenciómetro. Mediante la manipulación de éste último es posible modificar la ganancia de todo el sistema sin alterar la impedancia de entrada del mismo.

Puede apreciarse también en la figura 11 que sólo se realiza la compresión de uno de los dos canales de una señal estéreo. El prototipo implementado ofrece una salida de doble canal, así como la posibilidad de escuchar por ambos canales de salida la misma señal (monoaural) o realizar una simulación de audición estéreo, mediante el sencillo circuito desfasador mostrado en la figura 12.

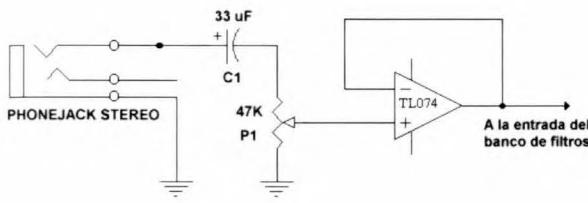


Figura 12: Desfasador para simulación estereofónica

Este circuito recibe como entrada la salida del sistema compresor, que se utilizará como uno de los canales estéreo, y proporciona la señal que deberá ir al otro canal estéreo. Manteniendo siempre una ganancia unitaria, introduce un desfase en la señal de entrada que oscila entre 0 y 180 grados, dependiendo de la frecuencia de ésta.

5. RESULTADOS EXPERIMENTALES Y CONCLUSIONES

El primer prototipo implementado ha ofrecido resultados positivos. La señal que se obtiene, a pesar de resultar desde un punto de vista auditivo más pobre, con menos matices (cosa que ya se esperaba y que resulta evidente debido a la eliminación de bandas frecuenciales), es del todo inteligible, lo cuál pone de manifiesto la viabilidad del proyecto, animando a una implementación futura de las etapas siguientes.

No obstante existe una gran cantidad de ruido a la salida, a pesar de haber eliminado el que procedía de la alimentación mediante condensadores de filtrado. Esto lleva a pensar que el origen de este ruido puede encontrarse en la ausencia de histéresis en los comparadores.

Desde un punto de vista puramente funcional el sistema aquí presentado permite una reducción en un 50% del ancho de banda necesario para transmitir la información de audio. Pueden obtenerse resultados similares partiendo de directrices de diseño totalmente diferentes: número de bandas y sub-bandas y criterio de elección de las mismas, tipo de filtros utilizado, orden y factor de calidad, determi-

nación de un umbral de corte para la discriminación de unas bandas en lugar de otras, etc. Todo esto afectará no sólo a la calidad de la señal obtenida, sino también a la complejidad del sistema.

6. AGRADECIMIENTOS

Quisiera agradecer a mi compañero y amigo Jordi Vidal la realización de todas las figuras incluidas en este artículo, y al Doctor Alberto Poveda por proponer y guiar con sus consejos la realización de este proyecto.

REFERENCIAS

- [1] David Rincón Rivera, «MP3, sonido digital al alcance de todos», Buran nº14, diciembre 1999
- [2] S. Franco, «Design with operational amplifiers and analog integrated circuits», McGraw-Hill, 1998

AUTOR



Marcelo Castro es Ingeniero Técnico de Telecomunicaciones por la Escola Universitària Politècnica de Vilanova i la Geltrú (EUPVG), de la Universitat Politècnica de Catalunya (UPC). En estos momentos trabaja sobre temas relacionados con la estimación espectral, para la realización de su Proyecto Final de Carrera de Ingeniería Electrónica, estudios que ha cursado en la Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona (ETSETB), también de la UPC.

Después de un corto paso por la industria, lleva varios años dedicado a la docencia técnica, siendo el actual el segundo a nivel universitario dentro de la propia UPC. Tras pasar el último año académico como profesor asociado del Departamento de Ingeniería Eléctrica, destinado en la Escola Universitària d'Enginyeria Tècnica Industrial de Barcelona (EUETIB), actualmente es profesor asociado del Departamento de Teoría de la Señal y Comunicaciones en la EUPVG, donde imparte la asignatura “Análisis de Circuitos II” de Ingeniería Técnica de Telecomunicaciones.



DISEÑO DE UN ROBOT RASTREADOR

José Antonio Casas Bueno

Estudiante del ETSETB y Coordinador de la sección de robótica de IEEE
Barcelona

E-mail: ihouses@iespana.es

INTRODUCCIÓN

Este es el primero de una serie de artículos de la nueva sección de robótica de la revista. Intentaremos mostrar una visión práctica de la robótica a nivel de aficionado o principiante introduciendo los temas básicos de la materia. En este primer artículo se planteará la problemática del diseño y construcción de un robot rastreador. Es ésta una modalidad muy extendida entre los concursos de robótica estudiantiles de ámbito local y estatal, debido a la sencillez de su diseño.

ROBOT RASTREADOR

Consiste en un robot que es capaz de seguir una línea blanca sobre fondo negro, o viceversa, con curvas y bifurcaciones. El mérito está en no salirse de la línea y en recorrerla lo más rápidamente posible.

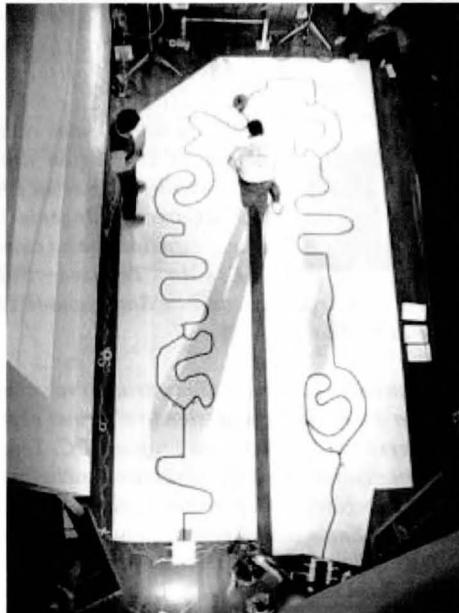


Figura 1. Concurso de robots Alcabot

PARTES DEL ROBOT

Abordaremos el diseño del rastreador dividiéndolo en tres apartados: la mecánica, la electrónica y por último el control. La parte mecánica incluye todo lo referente al chasis, los motores, las ruedas y las cajas reductoras. La parte electrónica estará basada en una o más placas de circuito impreso con la circuitería para controlar los sensores y tomar las decisiones. Por último, la

parte de control es la encargada de gobernar los movimientos del robot, de tomar las decisiones. Normalmente se basará en el programa que hayamos diseñado para algún dispositivo programable de la parte electrónica.

MECÁNICA

La estructura mecánica que se empleará en este artículo será la de "estructura diferencial", compuesta por dos ruedas tractoras y una rueda libre (rueda loca u omnidireccional).

Por tanto necesitaremos dos motores con reducción idénticos, la elección de los cuales dependerá de la velocidad máxima que queramos alcanzar, de la precisión del movimiento y del consumo máximo.

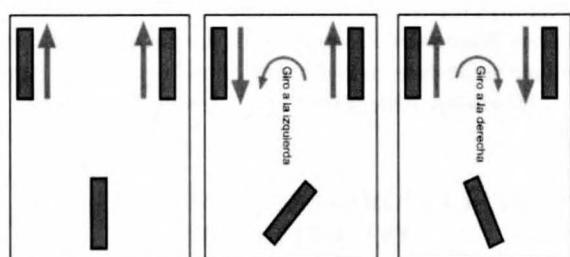


Figura 2. Estructura diferencial.

Los datos que necesitamos para saber si cierto motor con reducción nos sirve son las RPM (revoluciones por minuto) del eje al que acoplaremos la rueda, el par y la tensión de alimentación del motor. Si tenemos un valor de RPM muy bajo el robot tendrá una velocidad máxima pequeña, pero podrá hacer movimientos más lentos y precisos, por lo que aparece un compromiso en la elección de este valor. Esto plantea un problema de diseño importante, ya que no podemos hacer un robot que tenga

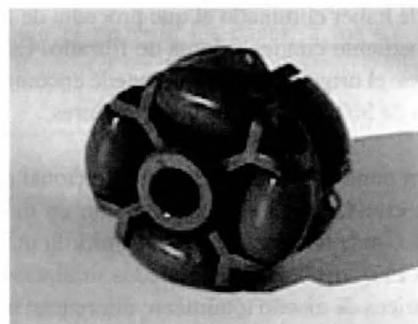


Figura 3. Rueda omnidireccional.

un rango de velocidades grande con una caja reductora determinada. El mismo problema que en los automóviles se resuelve mediante el uso de una caja reductora modificable, *el cambio de marchas*. En nuestro caso la elección adecuada es 2 a 4 revoluciones por segundo, puesto que en las competiciones de rastreadores se premia la velocidad.

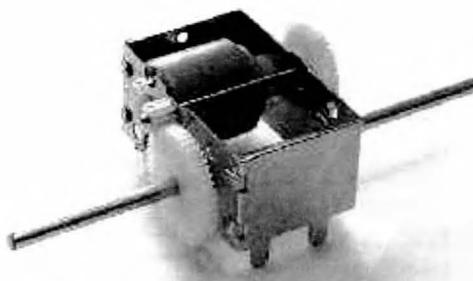


Figura 4. Motor con reducción 23:1

El par es la fuerza con la que el motor hace girar su eje. Está relacionado con la corriente que circula por los bobinados del motor y sus características propias. El par no será muy importante si el robot no es muy pesado (el peso máximo suele estar limitado a 2Kg) y tenemos una reducción razonable (un factor de 20 aprox.) pues al incluir la reducción dividimos la velocidad en vacío del motor (el motor solo, sin reducción) a la vez que multiplicamos el par. El otro valor a tener en cuenta es la tensión de alimentación del motor. No nos sirve cualquier tipo de valor. Si damos un vistazo podremos ver que los motores empleados en juguetería empiezan a funcionar a una tensión de 0.5V y alcanza su régimen máximo de RPM a los 2V (aprox.). Ello es debido a que están diseñados para ser alimentados con 1 o 2 pilas normales, que tienen una tensión nominal de 1.5V y que normalmente funcionan en un régimen de ON/OFF, es decir, que no se utilizan para sistemas que requieran una velocidad variable. También es importante apreciar que casi siempre se conectan a las baterías directamente, no hay un circuito de control intermedio.

En el otro extremo, tenemos los motores industriales (siempre me refiero a motores de pequeño tamaño) de prestaciones mucho más altas, que funcionan con tensiones entre 12V y 24V. Permiten un control de velocidad más preciso y tienen una eficiencia mayor. El precio de estos últimos es unas 10 veces mayor.

Lo más razonable es un término medio, un motor que alcance su máximo régimen de RPM entre los 6-12V, dependiendo de la tensión que proporcione la batería que queremos usar.

ELECTRÓNICA

La electrónica del robot consistirá en un determinado número de sensores y un microcontrolador o microporcesador.

La diferencia básica entre microcontrolador y microporcesador es que el microcontrolador integra en un mismo encapsulado un microporcesador y periféricos básicos, es decir la RAM y la ROM o EEPROM (la memoria de programa) además de otros periféricos adicionales como conversores A/D, UART... etc.

Dado que nuestros requerimientos son muy modestos, un microcontrolador de 8 bits nos será más que suficiente. En el mercado hay un gran número de ellos, siendo los más famosos los de la casa Microchip (PIC) y la casa Motorola.

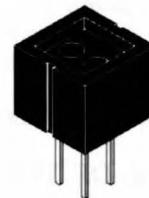


Figura 5. Aspecto externo de un CNY70

Puesto que queremos construir un robot rastreador, el robot deberá tener como mínimo dos sensores para detectar la línea. Los sensores más habituales son los CNY70. Un CNY70 es un sensor infrarrojo que funciona por reflexión. Emite una luz IR (infrarroja) y un fototransistor recibe la luz IR que le llega. Esta luz puede ser la que emite el propio sensor (que es lo deseable) o luz presente en el ambiente.

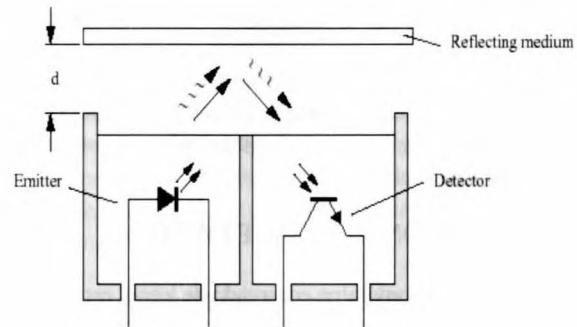


Figura 6. Esquema de funcionamiento.

Supondremos que el sensor se encuentra a una distancia del suelo suficientemente pequeña (de 0.5cm a 1cm) y que está de alguna manera "tapado" de manera que la luz ambiente sea despreciable. De no ser así se podrían utilizar otras técnicas para modular la luz emitida y así poder eliminar el efecto de la luz ambiente, complicando el diseño final.

Es importante destacar que este tipo de sensor no detecta color, detecta si hay reflexión o no. Por norma general las superficies oscuras reflejan menos que las claras, pero sería difícil distinguir

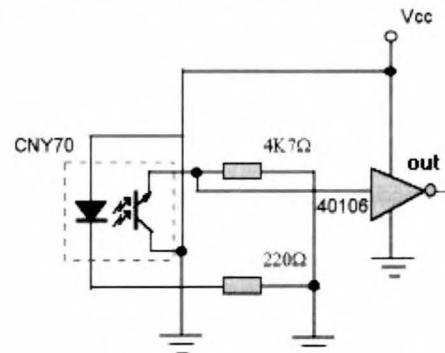


Figura 7. Posible conexionado de un CNY70

con un CNY70 entre una superficie negra brillante y una superficie clara, por lo tanto la superficie negra de la pista tiene que ser lo más mate posible.

El número y disposición de estos sensores determinará el grado de complejidad del robot. Como mínimo son necesarios dos pero hay robots presentados a concurso que disponen de hasta 23 sensores. En este artículo supondremos que el número de sensores es 5 y están dispuestos en línea recta.

Sensor1 Sensor2 Sensor3 Sensor4 Sensor5

CONTROL DE VELOCIDAD

Si queremos que nuestro robot disponga de este recurso tendremos que emplear una técnica llamada PWM, (Pulse Width Modulation) encaminada a poder controlar la tensión de alimentación del motor para así controlar su velocidad.

La generación de la señal PWM y su uso se salen de los propósitos de este artículo y se abordarán en futuros artículos más específicos sobre el control de velocidad de motores y servos, como también se tratarán temas relacionados con los *encoders* que nos aportan información sobre la velocidad de la rueda, su posición y su aceleración.

Para alimentar los motores se necesita un circuito adicional situado entre el microcontrolador y el motor, el denominado *driver*. Hay en el mercado circuitos integrados con esta función, como pueden ser el l298, l293B... etc.

CONTROL (SOFTWARE)

El control será el programa encargado de leer el estado de los sensores y actuar en consecuencia. Las actuaciones se limitan a mover hacia adelante o hacia atrás los motores o mantenerlos parados. Supondremos que no podemos controlar la velocidad de los motores, por lo tanto su velocidad viene determinada por la tensión de la batería que vamos a usar.

Así pues un motor tiene 3 estados diferentes:

ADELANTE ATRÁS PARADO

Una posible implementación del programa de control podría ser el siguiente código en C:

```
int dirección = adelante;
for(;;){
if (sensor1) { dirección = giro_derecha;}
else if (sensor2) { dirección = derecha;}
else if (sensor3) { dirección = adelante;}
else if (sensor4) { dirección = izquierda;}
else if (sensor5) { dirección = giro_izquierda;}
Mover_motores(dirección);}
```

DIRECCIÓN	MOTOR_DERECHO	MOTOR_IZQUIERDO
ADELANTE	ADELANTE	ADELANTE
ATRÁS	ATRÁS	ATRÁS
DERECHA	ADELANTE	PARADO
IZQUIERDA	PARADO	ADELANTE
GIRO_DERECHA	ADELANTE	ATRÁS
GIRO_IZQUIERDA	ATRÁS	ADELANTE
PARADO	PARADO	PARADO

Cabe destacar que ante una bifurcación, el robot escogerá el camino de la derecha. Ello se debe a que los sensores de la derecha tienen prioridad sobre los de la izquierda debido al orden en que se encuentran en los "if" ya que evalúan primero.

Otra característica de este código es que ante la posibilidad de que ningún sensor de "1", el robot seguirá en la dirección anterior a dicho suceso.

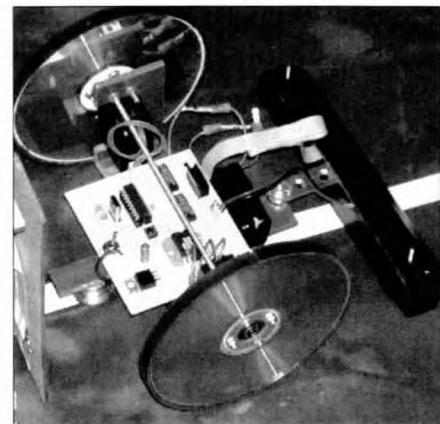


Figura 8. Robot sideral, 2º premio en el concurso de robots AEES 2002. Construido por José Luis Hernández, uno de los integrantes del grupo de robótica de IEEE Barcelona

Así pues hemos presentado a grandes rasgos la estructura de un robot rastreador simple. Para los que alguna vez se habían preguntado la manera de funcionar de estos robots espero que les haya aclarado alguna de sus dudas. Si bien este texto tiene carácter divulgativo y dista mucho de los verdaderos problemas que abordan los ingenieros que diseñan y construyen robots.

REFERENCIAS

- CNY70** Reflective Optical Sensor with Transistor Output
<http://new.vishay.com/docs/83751/83751.pdf>
- L293 Motor Driver**
<http://www.me.umn.edu/courses/me2011/robot/technotes/L293/L293.html>
- Concurso de robótica AEES
http://aess.upc.cat_robot.html
- Concurso de robots de la universidad de Alcalá de Henares
<http://www.alcabot.com/>
- Microcontroladores PIC: www.microchip.com



José Antonio Casas. Estudiante de 4º de Ingeniería Superior de Telecomunicación en la Escuela Técnica Superior de Ingenieros de Telecomunicaciones de Barcelona (ETSETB) de la Universitat Politècnica de Catalunya, UPC. Ganador del primer premio del concurso nacional de robótica AEES los años 2001 y 2002 en la categoría de robots limpiadores de superficies. Miembro reciente de la rama IEEE Barcelona y coordinador de la nueva sección de robótica de la rama.



INVERTIR EN INVESTIGACIÓN CIENTÍFICA NO ES IMPRESCINDIBLE

Lluís Torner

*Director de l'Institut de Ciències Fotòniques y catedrático de la UPC
(una versión reducida de este artículo fue publicada en el diario español La Vanguardia)*

En la década de los 80, el parlamento de Florida decidió comprometer importantes inversiones en la creación de centros de investigación del más alto nivel. Estar entre los mejores del mundo en un negocio muy concurrido requiere motivación, esfuerzo y recursos. En este caso se trata de recursos públicos, así que conviene poner esmero en destinarlos a partidas adecuadas. Los parlamentarios ofrecían un argumento convincente: Querían aumentar el bienestar del contribuyente, mediante mejores infraestructuras, mejores servicios, un medio ambiente más cuidado, etc. Y habían echado cuentas.

Florida (*the sunshine state*) basaba una gran parte de su economía en dos sectores: Cultivo de frutas y verduras, en particular cítricos, y turismo masivo asociado a un clima benigno, playas soleadas, parques de atracciones, etc. Estas actividades generan muchos puestos de trabajo, pero en general no requieren cualificaciones especiales y por tanto están mal remuneradas. Así pues, soportan cargas fiscales bajas. Resumen: El estado recauda pocos impuestos. Y, "on no n'hi ha, no en raja".

Afortunadamente, estamos manejando una ecuación sencilla, que admite dos soluciones: O aumentar los impuestos o aumentar los salarios. La primera solución es ideal para no salir reelegido. Por otro lado, una forma de conseguir que las empresas aumenten los salarios de sus empleados es convertirlas en monopolios de bienes necesarios a precios muy elevados. Esta opción ya la había probado Jaime I de Inglaterra con el jabón, con el resultado de impedir que sus súbditos adquirieran la costumbre de lavarse. Otra posibilidad es crear sectores productivos basados en la tecnología y en la innovación.

En "El señor de los anillos" la aventura no termina cuando Gandalf descubre lo que hay que hacer. Es necesario encontrar el camino, para que Frodo lleve el Anillo único a las Grietas del Destino. ¿Cómo se crea un tejido industrial potente basado en la tecnología? El mejor ejemplo es siempre el Silicon Valley. Los ingredientes de la receta californiana incluyen algunas personalidades clave, individuos emprendedores, muchos abogados, una legislación adecuada, capital riesgo, y legiones de personal altamente cualificado. Todo ello aderezado con un poco de suerte, un clima meteorológico y social adecuado para formar una familia, y una muy activa tolerancia frente al fracaso de las iniciativas ajenas emprendidas con honestidad.

La receta requiere mucho personal con formación media y superior, y una pequeña fracción de individuos bregados en ideas avanzadas a su tiempo, doctorados en las fronteras del conocimiento científico, tecnológico y humanístico. Menos del 5 % de la tecnología que se explota en el Valle proviene de investigación patentada en la Universidad de Stanford durante toda su historia. Ahora bien, sin los investigadores y graduados de Stanford y Berkeley, el Valle no existiría.

Así pues, en Florida resistieron la tentación de invertir en iniciativas apresuradas, ligadas a una posible explotación empresarial inminente de tecnologías de moda. La parte del león se la llevaron la Educación y la creación de centros de investigación básica. Después de dos décadas la semilla empieza a brotar en el Florida High-Tech Corridor. No se conocen atajos. La gestación de las yeguas también dura once meses. Si uno no está dispuesto a esperar, no hay potro.

Cuando decidimos invertir en investigación científica no estamos respondiendo a ninguna necesidad imprescindible del país. Más bien estamos eligiendo qué tipo de país queremos. La historia nos juzgará por nuestra elección. Con hospitales, infraestructuras y servicios regulares, y sin participar en el desarrollo de instrumentos que permitan mejorar la salud, aumentar y redistribuir la riqueza en el mundo, e iniciar estrategias de sostenibilidad, seguramente también se puede vivir.

Esto me recuerda una pegatina heroica que preside muchas oficinas del Valle, y que advierte a los incautos: "Si crees que la Educación es cara, prueba la ignorancia."

AUTOR



Lluís Torner es catedrático de la UPC y Fellow de la Optical Society of America. Fue galardonado por la Generalitat de Catalunya con una distinción a la promoción de la investigación universitaria en el año 2000. Doctor por la UPC en 1989, Torner ha sido investigador postdoctoral en las universidades de Florida, Arizona, Maryland y Stanford, en los Estados Unidos. Actualmente es director de l'Institut de Ciències Fotòniques.





APLICACIONES DEL LÁSER AL ANÁLISIS Y TRATAMIENTO DE OBRAS DE ARTE

Alejandro López-Gil Serra, Sergio Ruiz-Moreno, Amador Gabaldón Borque, Rosanna Pérez-Pueyo, M.J. Soneira Ferrando

Grupo de Comunicaciones Ópticas. Departamento de Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingenieros de Telecomunicación de Barcelona
Universidad Politécnica de Cataluña

laserraman@tsc.upc.es

1.- INTRODUCCIÓN

Dado que el legado artístico constituye una de las más importantes fuentes históricas de las que disponemos, éste debe ser, por encima de todo, conocido y preservado. En este sentido, las tecnologías fotónicas han demostrado su potencialidad en cuanto a su gran resolutividad y no destructividad.

Más concretamente, la espectroscopía Raman, basada en la respuesta por parte de la materia cuando ésta es iluminada por una radiación continua de características concretas, se ha convertido en una herramienta importantísima en cuanto al análisis, autentificación y datación de obras de arte.

Por otro lado, la fotoablación no térmica, consecuencia de la interacción entre la luz UV pulsada y la materia, está comenzando a despuntar en el campo de la conservación y restauración de obra patrimonial.

2.- CONCEPTOS TEÓRICOS

Espectroscopía Raman

Cuando una radiación monocromática intensa pasa a través de un material, la mayor parte de esta radiación se transmite, una parte se absorbe y una pequeña fracción se dispersa en todas las direcciones (efecto Tyndall). Casi toda la luz dispersada tiene la misma frecuencia que la radiación incidente, es lo que se conoce como dispersión elástica o dispersión Rayleigh. El descubrimiento de C.V. Raman se basó en la detección de líneas adicionales de dispersión inelástica e intensidad extremadamente débil que son particulares para cada molécula. Es el llamado efecto Raman.

La espectroscopía Raman es una técnica no destructiva que permite analizar la superficie de la obra sin dañarla, o microdestructiva en el caso de la toma de muestras para la identificación estratigráfica de capas pictóricas. Permite la identificación de muestras *in situ* debido a la transportabilidad de los equipos gracias a la aplicación de la fibra óptica. Otra de las ventajas que presenta esta técnica es la gran especificidad, dado que permite identificar compuestos a nivel molecular y la no ambigüedad en los resultados, ya que cada material

tiene su espectro característico. La espectroscopía Raman ofrece, además, una buena resolución espacial, con la posibilidad de identificar materiales orgánicos e inorgánicos, ya sea en estado sólido, líquido o en disolución acuosa. Esto último es posible ya que el agua presenta un espectro Raman muy débil [1,2]

Foto ablación no térmica

El proceso mediante el cual la radiación UV pulsada es capaz de eliminar de forma controlada los materiales orgánicos no deseables (espúreas) en obras de arte, es conocido como fotoablación no térmica.

Básicamente, el efecto que domina en este proceso recibe el nombre de efecto fotoquímico y viene favorecido por longitudes de onda de la radiación pulsada en el ultravioleta. El efecto fotoquímico describe el proceso de disociación de enlaces carbono-carbono que unen a los monómeros adyacentes dentro de un mismo polímero. Esta despolimerización crea un aumento de volumen de las unidades "libres", y por tanto de presión, que es lo que permite que cierta parte del material salga despedido, consiguiendo de esta forma la ablación del compuesto. En la figura 1 se muestra la dinámica del proceso de fotoablación no térmica en la irradiación de un polímero. Las esferas de la figura representan monómeros dentro de una misma macromolécula [3].

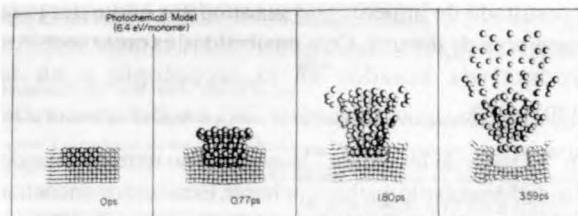


Figura 1. Dinámica del efecto fotoquímico (UV) para la sección transversal de un polímero.

Podemos cuantificar la profundidad de ablación por un pulso, ℓ_f , mediante la expresión

$$\ell_f = \frac{1}{\alpha} \log \frac{F}{F_{th}} \quad \text{para } F > F_{th} \quad (1)$$

donde α es el coeficiente de absorción del material, F es la fluencia o densidad de energía propia de la radiación y F_{th} es la fluencia umbral a partir de la cual se produce la ablación del material. Por ejemplo, según [4], con un láser excímero trabajando a 248 nm, emitiendo pulsos de 15 ns, con una frecuencia de repetición de 1 Hz y fluencia de 1 J/cm², se consigue una ablación del polímero polimida de 0,38 mm por pulso.

3.- DESCRIPCIÓN DEL EQUIPO EXPERIMENTAL

Sistema Raman con fibra óptica

En la figura 2 se presenta el diagrama de bloques del sistema de espectroscopía Raman con tecnología de fibra óptica disponible en el laboratorio de la U.P.C. Consta de dos fuentes monocromáticas (láser He-Ne y Ar), la radiación de una de las cuales es guiada a través de la fibra óptica de excitación hasta el cabezal óptico. En éste se realizan las funciones de filtrado, enfoque y detección.

Un filtro interferencial se encarga de eliminar tanto las bandas Raman introducidas por la propia fibra como las líneas plasma provenientes del láser. A continuación, un objetivo enfoca el haz de excitación sobre la muestra y se encarga de colectar la radiación dispersada por la materia. Mediante un doble filtro notch se elimina la línea Rayleigh, fracción de intensidad óptica reflejada que no aporta información acerca del material analizado, que enmascararía por completo a la señal Raman debido a su mayor intensidad.

La radiación resultante es guiada a través de la fibra de colección hasta el monocromador cuya función es la de separar espacialmente las diferentes componentes espectrales y dejar pasar únicamente un rango frecuencia dado. El conjunto de frecuencias seleccionado por el monocromador incide en el detector CCD que realiza la conversión de fotones en electrones.

Por último, esta corriente eléctrica es digitalizada y remitida a un PC para su visualización, almacenamiento y/o tratamiento. El sistema informático controla, además, los parámetros del monocromador y del CCD.

Láser pulsado de Nd:YAG y generador de armónicos ultravioleta

La estrategia utilizada para la generación de luz ultravioleta pulsada se basa en el uso de un láser muy extendido como es el de Nd:YAG que proporciona radiación infrarroja pulsada (1064 nm) y, mediante la

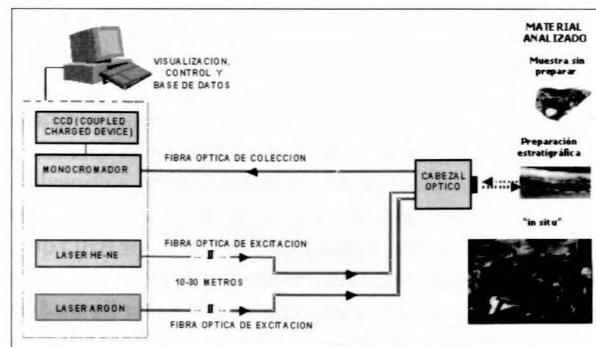


Figura 2. Diagrama de bloques del sistema de espectroscopía Raman con tecnología de fibra óptica.

inclusión de cristales activos no lineales, conseguir radiación pulsada visible y ultravioleta, es decir, el segundo, tercer y cuarto armónico (532, 355 y 266 nm, respectivamente).

El láser disponible en el laboratorio basado en esta estrategia es el modelo Polaris III de la empresa New Wave Research Inc. Sus parámetros característicos en función del armónico de trabajo se presentan en la Figura 3 [5].

	1064 nm	532 nm	355 nm	266 nm
Energía del pulso	90 mJ	50 mJ	11 mJ	10 mJ
Diámetro del spot	3.75 mm	3.5 mm	3.25 mm	3.25 mm
Duración del pulso	5-7 ns	5-6 ns	4-6 ns	4-6 ns
Divergencia del haz	< 4 mrad	< 4 mrad	< 3 mrad	< 3 mrad
Ritmo de repetición			1-10 Hz	
Requerimientos de potencia		100-240 VAC, 50/60 Hz, 250 W		

Figura 3. Características del láser pulsado Polaris III.

El láser está compuesto físicamente por tres unidades claramente diferenciables: la unidad de potencia, el panel de control y el cabezal láser (figura 4).

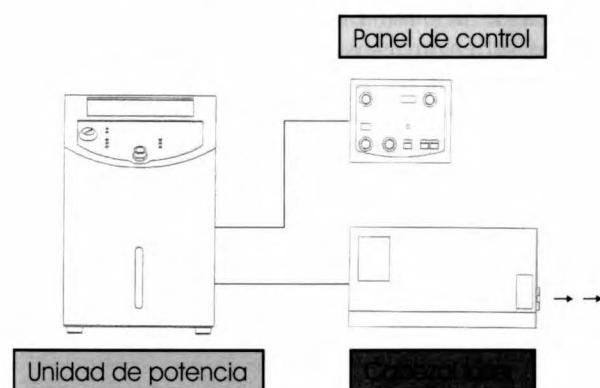


Figura 4. Diagrama de bloques del láser pulsado Polaris III.

La unidad de potencia contiene la electrónica propia del sistema y controla el circuito cerrado de refrigeración por agua. Por su parte, el panel de control, permite controlar todas las variables del láser, es decir, la energía relativa



por pulso, el ritmo de repetición, la modalidad de disparo y la carga de la lámpara de bombeo. Por último, el cabezal láser aloja la cavidad láser propiamente dicha, el atenuador variable, los cristales activos no lineales y varios filtros dicroicos para filtrar y desviar el haz en la dirección correcta. Se dispone de tres cristales activos no lineales: un KTP que permite duplicar la frecuencia natural del láser obteniéndose el segundo armónico y dos BBO que, utilizándolos individualmente junto al KTP, permiten la generación del tercer y cuarto armónico. En la figura 5 se muestra el cabezal láser con una configuración de KTP+BBO para la generación del cuarto armónico.

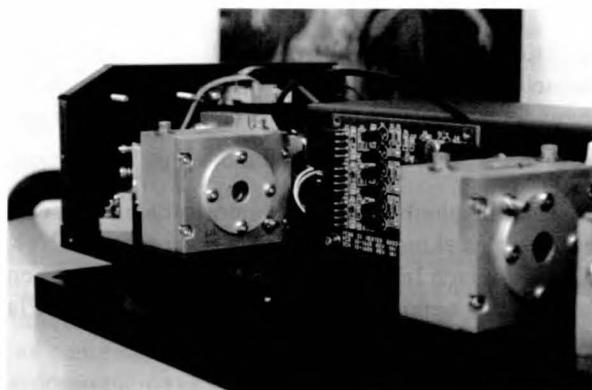


Figura 5. Configuración, KTP + BBO, para la generación del cuarto armónico.

4.- RESULTADOS EXPERIMENTALES

Se ha analizado la aplicación de la luz ultravioleta pulsada en la eliminación de la capa de barniz de obras pictóricas [6]. El interés en la eliminación de esta capa reside tanto en cuestiones de conservación como de estética de la obra. Los barnices, con el paso del tiempo, sufren procesos de degradación debido a agentes externos que provocan que el barniz deje de cumplir los objetivos por los que fue aplicado, es decir, proteger la obra y ensalzar los colores.

Cabe destacar, que una consecuencia de este investigación experimental es la posibilidad de mejorar la relación señal a ruido de un espectro Raman debido a la reducción de fluorescencia que se consigue al eliminar el barniz superficial [6,7].

Eliminación de barniz con láser UV pulsado

Sobre una placa de vidrio de 7,5 cm de largo por 2,5 cm de ancho se depositó una fina capa de pigmento massicot aglutinado con aceite de linaza. La mezcla se dejó secar al sol durante 5 horas. Posteriormente se aplicó una capa de barniz sobre el pigmento aglutinado y se dejó reposar la muestra durante 15 días. En la figura 6 se presenta un esquema estratigráfico de la muestra preparada.

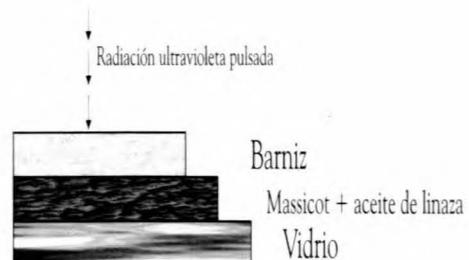


Figura 6. Muestra preparada para la limpieza de barniz con luz ultravioleta pulsada.

Transcurrido este tiempo, se procedió a la limpieza de la capa superficial de barniz. Se experimentó sobre diversas áreas de la muestra para observar la dependencia de los parámetros del láser en el proceso de fotoablación. En concreto, se experimentó con la energía del pulso y el número de pulsos. En cada caso se procedió de la siguiente forma: se trató la muestra mediante luz UV pulsada con unos parámetros fijos del láser pulsado y se observó mediante el microscopio el resultado de la interacción luz-materia. A continuación se obtuvo una fotografía de la zona irradiada mediante el videoprint conectado al microscopio.

a) Energía del pulso

Para estudiar el efecto de la energía por pulso en el proceso de eliminación de barniz sobre la muestra preparada se mantuvo fijo el número de pulsos, en este caso uno, y la frecuencia de repetición, 1 Hz. La longitud de onda de la radiación pulsada fue de 266 nm (UV) y la muestra se colocó a 4 cm de la salida del láser pulsado. Con ello se pretendió calcular empíricamente el umbral de ablación (fluencia o densidad de energía a partir de la cual se produce fotoablación) del barniz utilizado. Los resultados experimentales obtenidos irradiando sobre diferentes zonas de la muestra fueron:

- 1) Para una energía del pulso menor o igual a 6 mJ, es decir, fluencia menor o igual a 72,46 mJ/cm², la muestra no sufrió ningún cambio o transformación.
- 2) Para una energía de 7 mJ por pulso, fluencia de 84,44 mJ/cm², se observa una ligera eliminación del barniz. Este es el umbral de ablación del barniz empleado. En la figura 7 se muestra la imagen obtenida con dichos parámetros. El hecho de que sólo la zona equivalente a la parte central del spot se haya ablacionado se debe al perfil energético gausiano del pulso generado por el Polaris III.
- 3) Para un pulso de energía 8 mJ, es decir, fluencia de 96,50 mJ/cm², se observa como la fotoablación afecta a una zona mayor del área iluminada por el pulso ultravioleta (figura 8). Esto concuerda con los estudios teóricos acerca de la fotoablación que pre-

dicen un aumento logarítmico de la profundidad de ablación al aumentar la fluencia de trabajo (Ec. 1).

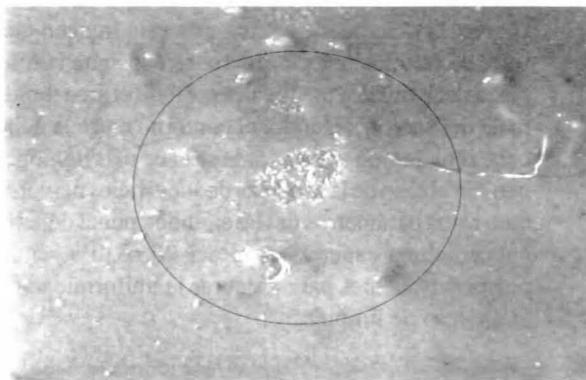


Figura 7. Muestra tratada con pulso de fluencia 84,44 mJ/cm².

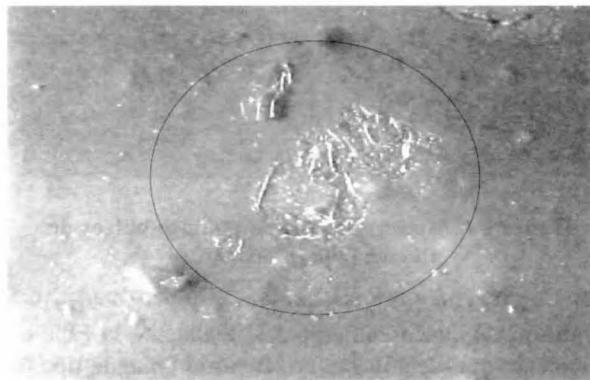


Figura 8. Muestra tratada con pulso de fluencia 96,50 mJ/cm².

4) Procediendo de igual forma que en los casos anteriores, se obtuvieron las imágenes de la superficie de la muestra cuando ésta es irradiada con 9 mJ (108,57 mJ/cm²) y 10 mJ (120,63 mJ/cm²), figuras 9 y 10, respectivamente. Se evidencia como el área y la profundidad de ablación sigue creciendo progresivamente. Cabe destacar, por un lado, que parte de la capa de pigmento más aglutinante ya se encuentra en contacto con el medio exterior y, por otro lado, que en ningún momento se observan cambios ni transformaciones en zonas que no haya sido iluminadas.

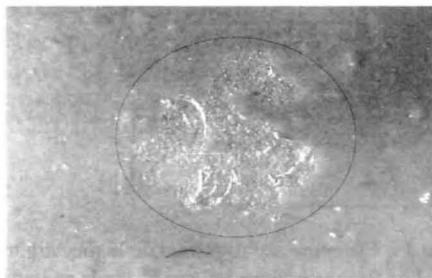


Figura 9. Muestra tratada con pulso de fluencia 108,57 mJ/cm².



Figura 10. Muestra tratada con pulso de fluencia 120,63 mJ/cm².

b) Número de pulsos

Para observar el efecto del número de pulsos en el proceso de fotoablación del barniz se procedió de la siguiente forma. Primero se escogió una energía por pulso tal que estuviera por encima del umbral de ablación que como se vio anteriormente se midió entorno a 7 mJ por pulso. En este caso, se optó por irradiar la muestra con 7,5 mJ por pulso, o lo que es lo mismo, con una fluencia de 90,48 mJ/cm². A continuación se hizo incidir sobre una zona intacta de la muestra, es decir no irradiada, un pulso con dicha fluencia y se obtuvo una fotografía. Posteriormente, y sobre la misma zona, se hizo incidir otro pulso y se obtuvo, de nuevo, una fotografía. Actuando de esta forma se obtuvieron las imágenes de la superficie de la muestra cuando se irradió con 1, 2, 3 y 15 pulsos en total. En todos los casos la fluencia fue de 90,48 mJ/cm², la longitud de onda de 266 nm y se trabajó a una distancia del cabezal láser de 3,5 cm.

El objetivo de este experimento fue observar la evolución de la fotoablación del barniz en función del número de pulsos cuando la fluencia de trabajo se sitúa por encima de la fluencia umbral de ablación. Los resultados de detallan a continuación.

- 1) Con un pulso de fluencia de 90,48 mJ/cm² se obtuvo la imagen que se muestra en la figura 11. Se constata de nuevo como parte de la capa de barniz ha sido eliminada a la vez que la zona más afectada es la que corresponde a la parte central del spot del láser debido a su perfil energético gausiano.

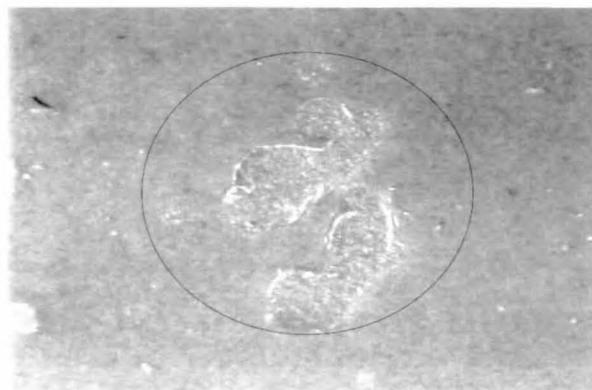


Figura 11. Muestra tratada con pulso de fluencia 90,48 mJ/cm².

2) A continuación se hizo incidir otro pulso con la misma fluencia exactamente sobre la misma zona. De esta forma, el número de pulsos acumulados fue de 2. En la figura 12 se puede ver cómo el área de barniz ablacionado ha crecido al mismo tiempo que la profundidad en la zona central se hace más intensa. Tanto en este caso como en el anterior (1) todavía no se llegó a dejar ninguna parte de la capa de pigmento más aglutinante en contacto con el medio externo.

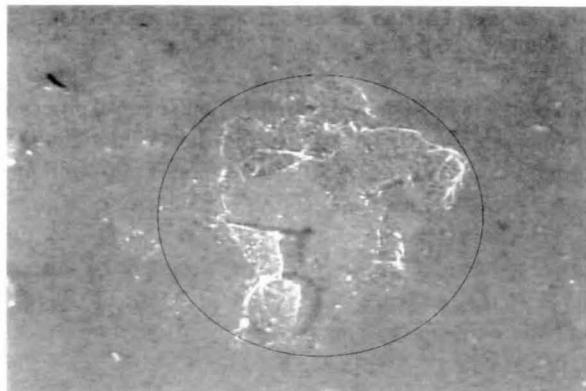


Figura 13. Muestra tratada con tres pulsos de fluencia 90,48 mJ/cm².

3) Sobre la misma zona de los casos anteriores (1 y 2) se hizo incidir otro pulso más. Así, el número de pulsos acumulados fue de 3. La imagen se muestra en la figura 13. En ella se puede ver como una pequeña zona de la capa de pigmento más aglutinante ya está en contacto con la atmósfera exterior. Es decir, que bajo estos parámetros del láser pulsado, ya se logra una cierta eliminación completa de la capa de barniz. A pesar de ello, se hicieron incidir algunos pulsos más para ver la evolución de la superficie de limpieza y comprobar si el pigmento sufría algún tipo de cambio o transformación.

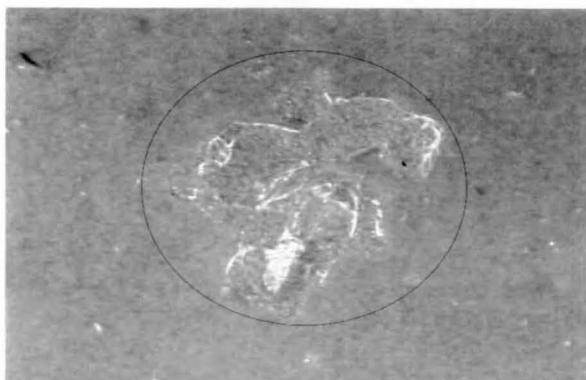


Figura 12. Muestra tratada con dos pulsos de fluencia 90,48 mJ/cm².

4) Procediendo de igual forma que en los casos anteriores, se hizo incidir sobre la muestra 12 pulsos de la misma fluencia, con lo que el número de pulsos acumulados sobre la misma zona ascendió a 15 (figura 14). Se pueden extraer de esta imagen una serie de conclusiones: primero, se consigue la limpieza del barniz en prácticamente todo el área iluminada del spot y, segundo, el contorno entre la zona ablacionada y no ablacionada se hace más homogéneo. En el caso de la limpieza de superficies mayores bajo estos parámetros del láser, únicamente debería aplicarse una superposición del 20 o 30 % entre pulsos adyacentes para asegurar la uniformidad en el proceso de limpieza.

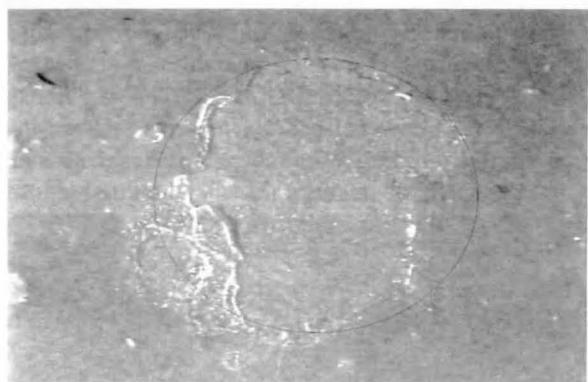


Figura 14. Muestra tratada con quince pulsos de fluencia 90,48 mJ/cm².

Únicamente queda comprobar si realmente la capa de pigmento más aglutinante no ha sufrido ningún tipo de transformación que no pueda ser detectada mediante la inspección microscópica. De esta forma, se procedió a tomar el espectro Raman de la zona ablacionada en el último caso, es decir, cuando se hizo incidir sobre la muestra 15 pulsos a 90,48 mJ/cm² para una longitud de onda de 266 nm. Por otro lado se tomó un espectro Raman de una zona no ablacionada para poder compararlo con el espectro anterior. En la figura 15 se muestra el esquema estratigráfico de la muestra irradiada y los lugares donde se tomaron los espectros Raman.

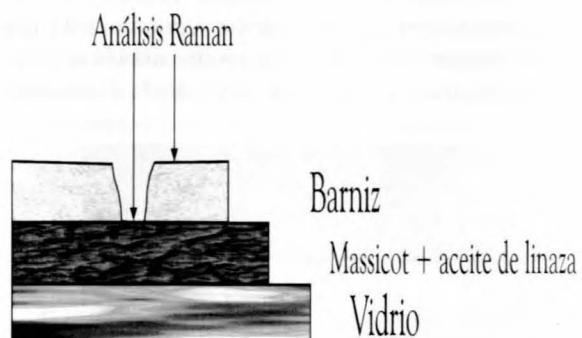


Figura 15. Esquema estratigráfico de la muestra tras ser pulsada. Las flechas indican las zonas donde se tomaron los espectros Raman.

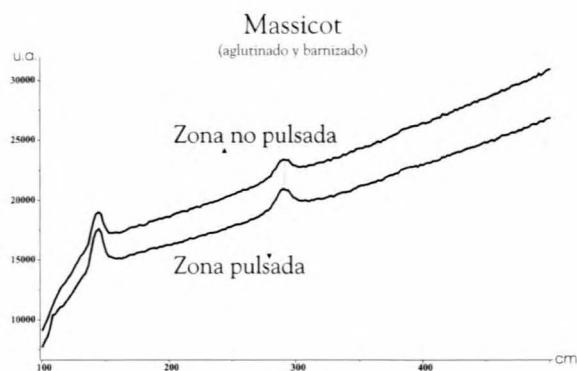


Figura 16. Espectro superior: zona no pulsada de la muestra (tiempo de adquisición de 20 seg); espectro inferior: zona pulsada, barniz ablacionado (tiempo de adquisición de 30 seg).

Como se puede comprobar en la figura 16, no sólo el pigmento más aglutinante ha quedado intacto (misma posición de las bandas Raman) sino que además se ha reducido el nivel de fluorescencia del espectro debido a la eliminación de la capa de barniz, lo que se transforma en una mejora de la relación señal a ruido del espectro.

5.- CONCLUSIONES

La espectroscopía Raman ha demostrado ser una técnica analítica idónea para observar y monitorizar los procesos involucrados en la interacción de la luz UV pulsada y los materiales artísticos. Ello es debido a la no destructividad y gran especificidad que presenta el análisis Raman y a la accesibilidad que ofrece la fibra óptica.

La combinación del láser pulsado Polaris III, de New Wave Research Inc., más óptica no lineal, presenta unas excelentes prestaciones en cuanto a manejabilidad, flexibilidad, elección de rango óptico y energía suficiente, que permiten llevar a cabo una fotoablación no térmica de forma totalmente controlada.

Por primera vez en una universidad española se ha confirmado la efectividad de los pulsos UV en la eliminación de barnices superficiales. El umbral de fotoablación del barniz utilizado se midió entorno a una energía por pulso de 7 mJ, lo que equivale a una fluencia de 84,44 mJ/cm², para una longitud de onda de 266 nm y trabajando a una distancia de 4 cm entre la muestra y la salida del Polaris III.

AGRADECIMIENTOS

Este trabajo se enmarca dentro del proyecto de investigación “Aplicaciones del Láser a la Investigación Analítica Global de Obras de Arte (ALIAGOA)”, ref. TIC-2000-1045 (CICYT)

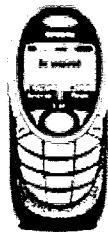
REFERENCIAS BIBLIOGRÁFICAS

- [1] S. Ruiz-Moreno, J.M. Yífera, M.J. Manzaneda, M.J. Soneira, P. Morillo, T. Jawhari, “La Ciencia al Servicio del Arte. La Espectroscopía Raman Aplicada a la Identificación de Pigmentos”, Mundo Electrónico, 265, (1996)
- [2] S. Ruiz-Moreno, C. Sandalinas, R. Pérez, A. Gabaldón, M.J. Soneira, “Raman Laser Fibre Optic Strategy for Non-destructive Pigment Analysis. Identification of a New Yellow Pigment (Pb,Sn,Sb) of the Italian XVII Century”, Journal of Cultural Heritage, (pendiente de publicación)
- [3] B.J. Garrison, R. Srinivasan, “Ablative Photodecomposition of Polymers”, J. Vac. Sci. Technol, A 3, (1985)
- [4] S. Küper, J. Brannon, K. Brannon, “Threshold Behavior in Polyimide Photoablation: Single-Shot Rate Measurements and Surface-Temperature Modeling”, Appl. Phys A, 56, (1993)
- [5] <http://www.new-wave.com/products/polaris.html>
- [6] PFC de Alejandro López-Gil Serra (enero 2003), ETSETB, UPC
- [7] S. Ruiz-Moreno, A. Gabaldón, A. López-Gil, M.J. Soneira, “Signal to Noise Improvement in Raman Spectra with Pulsed UV Laser”, Applied Spectroscopy, (pendiente de aceptación).

AUTOR



Alejandro López-Gil Serra es Ingeniero Técnico Superior de Telecomunicación y doctorando del departamento de Teoría de la Señal y Comunicaciones de la Universidad Politécnica de Catalunya.



INTRODUCCIÓN AL SISTEMA GPRS Y A SU GESTIÓN DE RECURSOS RADIO

Enric Chillaron i Farré

Estudiante de la ETSETB y de la ENST de Paris
y Miembro de la Rama de
estudiantes del IEEE de Barcelona
chillaro@enst.fr

I. INTRODUCCIÓN

En este artículo se pretende introducir al sistema GPRS y presentar una visión global de su gestión de recursos radio.

En una primera parte, se dará una explicación general del sistema GPRS, así como su arquitectura protocolaria, su capa física y su capacidad de codificación de canal adaptable.

Más adelante, se detallarán los canales lógicos que utiliza este sistema de segunda generación y como realiza la compartición de los recursos radio y el acceso al medio.

Finalmente se presenta un breve apartado de simulación para observar el tráfico de diferentes paquetes GPRS cuando comparten canales clientes WEB y WAP [9].

II. BREVE PRESENTACIÓN DE GPRS

GPRS es un sistema de segunda generación basado en la capa física de GSM que utiliza la conmutación de paquetes y la clase *multislot* junto con varias codificaciones de canal para obtener diferentes caudales [3].

El sistema GPRS permite asignar *QoS* diferenciadas a los distintos usuarios móviles, en función del caudal o *throughput* medio/pico del enlace, de los retardos o de la fiabilidad del enlace.

Este sistema es idóneo para aplicaciones donde las transmisiones de datos sean intermitentes (en forma de ráfaga), frecuentes de pequeño volumen e infrecuentes de volumen elevado.

GPRS es una red superpuesta a GSM que puede transportar paquetes IP y que comparte con ella la red de acceso, es decir, todo lo que son las MT (*Mobile Terminal*) y las BSS (*Base Station System*). Además, GPRS introduce tres nuevos nodos funcionales respecto GSM que son el SGSN (*Serving GPRS Support Node*), el GGSN (*Gateway GPRS Support Node*) y el PCU (*Packet Control Unit*). El primero, es el responsable de la entrega de paquetes al terminal móvil en su área de servicio, y el segundo actúa como interfaz lógico hacia las redes de paquetes de datosexternas. En cambio, el PCU es el encargado de gestionar el interfaz aire de la red.

A nivel MAC, las asignaciones de recursos entre el *uplink* y el *downlink* se realizan de manera independiente (al revés de lo que sucede con GSM) y un mismo canal físico o PhyCh (*Physical Channel*) puede ser compartido por varios dientes.

III. ARQUITECTURA PROTOCOLARIA DE GPRS

En este apartado se analizará la arquitectura protocolaria de GPRS y se realizará una breve descripción de los niveles que lo integran. En la figura 1 se observa la pila de protocolos y los interfaces que utiliza GPRS [2][3].

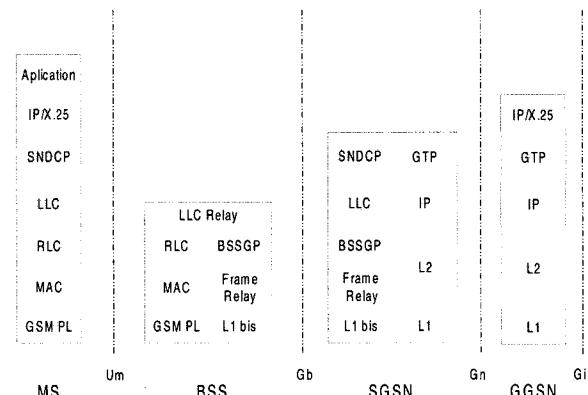


Figura 1. Arquitectura protocolaria de GPRS[3].

Algunos de los protocolos utilizados en el interface radio Um son los siguientes [1]:

- La capa física (nivel 1) que corresponde a la capa física de GSM.
- La capa MAC que gestiona el acceso al medio compartido entre todos los posibles clientes.
- RLC (*Radio Link Control*) que corresponde, equiparándolo con OSI, a la capa de enlace (al igual que la capa MAC).
- LLC (*Logical Link Control*) que remplaza las capas LAPDm o RLP. Esta capa se utiliza para restablecer un enlace débil utilizando el interface radio.

Otro protocolo utilizado a nivel de BSS es:

- BSSGP (*BSS de GPRS*) que es una BSS de GSM con algunas adaptaciones de GPRS (introducción de PCU).

Finalmente, algunos de los protocolos utilizados en la arquitectura protocolaria de red o *coeur* son:

- Los protocolos del interface radio (LLC)
- El protocolo SNDCP (*Sub-Network Dependence Convergent Protocol*) que corresponde a una capa radio especificada entre los niveles 2 y 3 (subcapa de adaptación).
- Los protocolos más clásicos, de Internet u otros, como TCP-IPGTP (*GPRS Túnel Protocol*) o los protocolos de Frame Relay.

IV. INTERFACE RADIO. CAPA FÍSICA

GPRS utiliza la estructura de canal físico de GSM donde cada PhyCh corresponde a 1 slot por trama TDMA (más o menos a 10Kbps en GSM. Véase figura 3). A partir de ahora, se hablará de PDCh (*Packet Data Channel*) como cada PhyCh correspondiente a GPRS.

Para empezar, GPRS utiliza un formato de multitrama de 52 tramas TDMA (en GSM se consideraban multitramas de 26 y de 51). Igualmente, se trabaja con bloques o *Phy SDU* de 4 bursts ya sea para informaciones de control o para señalización de GSM. La duración total de una multitrama GPRS es de 240 ms y la de un bloque de 20 ms.

En la figura 2 se puede observar la estructura temporal de GPRS.

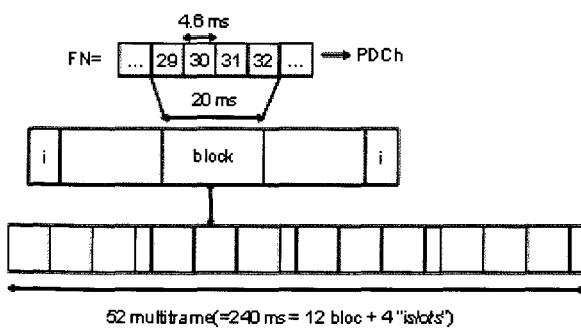


Figura 2. Ilustración de un bloque dentro de una multitrama de 52 de GPRS[1]

Destacar que los *bursts* se utilizan para la gestión del TA (*Time Advance*), es decir, para la sincronización del terminal móvil, y que la cantidad de bits enviados por bloque MAC es de 456 (=4x114, *protected bits+overhead*).

En GPRS se habla de FN (*Frame Number*) como el contador de tramas TDMA. Cada trama TDMA tiene una duración de 4.6 ms y se define como:

$$\text{ContadorFN} = \frac{12}{13} * 5\text{mseg} + 1$$

FN se incrementa una unidad cada trama TDMA. El concepto de FN se puede observar en la figura 3 [1].

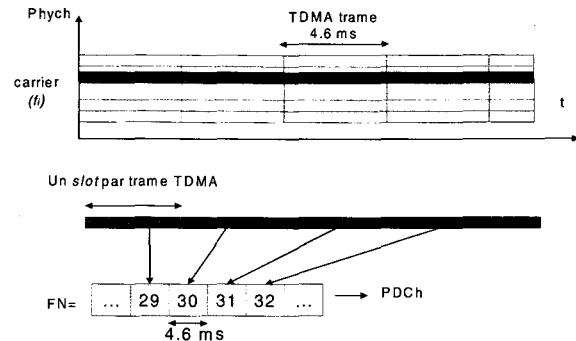


Figura 3. Ilustración del funcionamiento de FN [1]

V. CODIFICACIÓN DE CANAL ADAPTA- BLE

GPRS presenta 4 posibles esquemas de codificación de canal (*Coding Schemes*) o tasas de transmisión de datos variables por multislot [3]:

- CS-1: 9.05 Kbps.
- CS-2: 13.4 Kbps.
- CS-3: 15.6 Kbps.
- CS-4: 21.4 Kbps.

La información de CS se transmite vía los *stealing flags*, los cuales tienen la ventaja de no necesitar señalización suplementaria. No obstante, presentan el posible problema de desadaptación dentro de la capa RLC. La figura 4 representa la utilización de los *stealing flags*.

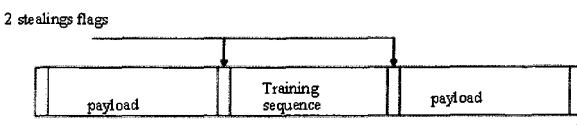


Figura 4. Fomato de un burst GSM normal [1]

Referente a los esquemas de codificación, se puede decir que por los entornos con bajo CIR (*Carrier-to-Interference Ratio*) es mejor trabajar con el CS-1. En cambio, por entornos con elevado CIR es mejor trabajar con CS-4 puesto que ofrece más velocidad y menor protección de los datos a transmitir.

VI. CLASES DE EQUIPOS MOBILES

En este apartado se hablará especialmente de dos puntos: de los equipos móviles con funcionalidades diversas y de las clases multislots de un móvil. Por lo que se refiere al primer punto, se presentan las dos clases siguientes [1]:

- Clase A/B/C
 - Clase A: GSM y GPRS al mismo tiempo y simultáneamente.
 - Clase B: GSM o GPRS secuencialmente.
 - Clase C: elección a priori de GSM o GPRS.
- Clase multislots: con diferentes caudales.

Por lo que respecta al segundo punto, se puede hablar de:

- Capacidades de tratamiento.
- Trafico asimétrico, donde un móvil GPRS puede recibir hasta i bursts y transmitir hasta j por trama TDMA pero con la condición que $i+j=k$.

Destacar que hay 29 clases multislots que han estado especificadas (información codificada en 5 bits).

VII. COMPARTICIÓN DE RECURSOS RADIO

Dentro del sistema GPRS, aparece el concepto de compartición de recursos radio y éste puede darse en el caso de un solo o varios PDCHs. En el caso de un solo PDCh, se puede observar en la figura 5 donde dos clientes comparten un mismo PDCH en el uplink y en el downlink. En cambio, en el caso de varios PDCHs, se puede ver en la figura 6 donde cuatro clientes comparten 3 PDCHs.

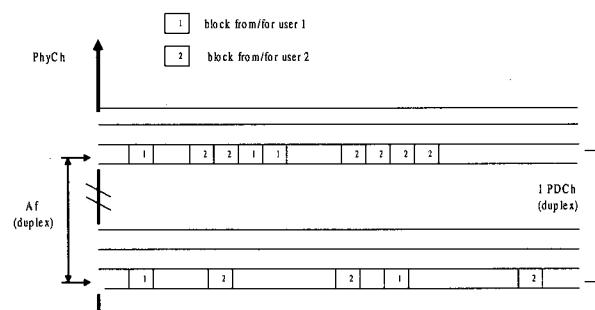


Figura 5. Compartición de recursos radio en el caso de un solo PDCh y dos clientes

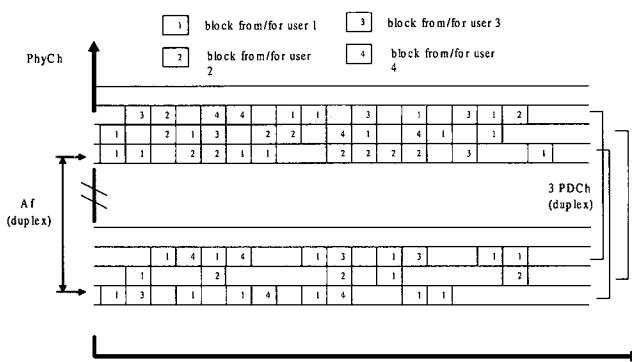


Figura 6. Compartición de recursos radio en el caso de tres PDCh y cuatro clientes

Remarcando que en el caso de la figura 6 hay 4 clientes que comparten 3 PDCH donde el cuarto cliente tiene una clase multislots ($i=3$, $j=2$, 4) (Véase el apartado VI).

Destacar que la compartición de canales PDCh se realiza siguiendo la clase multislots de los MT de los clientes. Finalmente, decir que se puede efectuar una compartición de recursos radio entre clientes GSM-c y clientes GPRS (PhyCh y PDCh).

VIII. CANALES LÓGICOS

En la tabla 1 se puede observar buena parte de los canales lógicos sobre los que trabaja GPRS.

Grupo	Nombre	Dirección	Función
PBCCH	PBCCH	Downlink	Broadcast
PCCC H	PRACH	Uplink	Random Access
	PPCH	Downlink	Paging
	PAGCH	Downlink	Access grant
	PNCH	Downlink	Multicast
PTCH	PDTCH	Both	Data
	PACCH	Both	Control
	PTCCH	Both	Time Advance

Tabla 1. Tabla de canales lógicos que utiliza GPRS [3]

La clasificación “canónica” GSM de los canales lógicos corresponde a la siguiente:

- Canales de difusión (*Broadcast Channels*). Se encargan de dar información del sistema.
- Canales de control común (*Common Control Channels*). Son utilizados en la fase de acceso: para “despertar” al móvil y/o para establecer una transmisión.
- Canales dedicados a un cliente (*Dedicated*) como los canales de tráfico (*TCH*).
 - Canales de control asociado (*Associated Control Channels*). Corresponden a los canales de control lógico y físico de un canal dedicado.

En el sistema GPRS se utiliza el distintivo P- (acrónimo del canal lógico de GSM) donde P significa *packet* y se destaca que los canales PAGCH se sitúan dinámicamente en los recursos, a priori, dedicados al tráfico.

En la figura 7 se observa el escenario de transferencia dentro de un contexto radiomóvil celular.

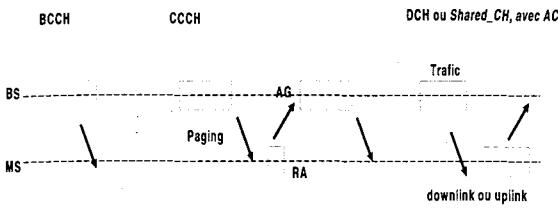


Figura 7. Esquema de un contexto de conexión en una red radiomóvil celular [1]

IX. CONTROL DE ACCESO AL MEDIO

En el sistema GPRS se definen los identificadores siguientes para gestionar correctamente el acceso al medio compartido [3]:

- TBF (*Temporary Block Flow*). Es el encargado de identificar una o varias tramas LLC de un mismo cliente. Un TBF puede comportarse como *open ended* (nombre indefinido de octetos) o *closed ended* (nombre de octetos fijados al principio del TBF).
- TFI (*Temporary Flot Identity*). Permite el multiplexado en el *downlink*. Un único TFI se asigna a cada usuario dentro de la celda. El indicador TFI se incluye en la cabecera de los paquetes RLC para permitir la implantación del protocolo ARQ selectivo.
- USF (*Uplink State Flag*). Permite el multiplexado en el *uplink*. Cada bloque RLC del *downlink* lleva este indicador. Como sucede en el caso del TFI, un único USF se asigna a cada cliente. Si la estación base envía un mensaje con el mismo USF que el del cliente, éste podrá hacer uso del siguiente bloque *uplink* para transmitir sus datos a la BTS. Si USF=FREE, indica que el bloque siguiente se destina al proceso de acceso (PRACH). Existen un total de 8 valores para el indicador USF. Si USF=1 indica que el slot está *free* y puede utilizarse para el acceso. En cambio, cualquiera de los otros siete valores identifica a uno de los móviles activos en esta portadora.

La gestión de acceso al medio se realiza de manera independiente en función del sentido de la transmisión. En el *downlink*, se trata de una mera difusión. Es el caso más simple donde el móvil escucha todos los bloques de los canales PDCh que le han estado indicados. En cambio, en el *uplink* las estaciones bases deben organizar las emisiones de los móviles de una manera dinámica. Es decir, las estaciones bases deben conocer los instantes precisos (y frecuencias) donde ellas deben escuchar.

Por consiguiente, como en el *uplink* se da el problema de acceso múltiple, se plantean diferentes modelos para solventarlo:

· Asignación dinámica. La BS coloca de manera dinámica en el *downlink* los permisos de emisión (*flag USF*) de los TBlock del *uplink*

· Asignación dinámica extendida. Extensión del caso anterior donde se permite trabajar con caudales más elevados.

· Asignación fija. La BS indica durante el establecimiento del TBF los recursos dedicados al terminal móvil (*bipmap*).

X. SIMULACIÓN DE LA GESTIÓN DE RECURSOS GPRS

En este último apartado, se realizará una simulación con una compartición dinámica de los recursos radio entre diferentes clientes GPRS en el *downlink* [9]. La finalidad de la simulación no es otra que intentar ilustrar, de la manera más real posible, un contexto común en GPRS y ver como soporta el sistema el tráfico para cada uno de los usuarios aplicando una política de *scheduling* concreta. Tómese éste apartado a modo de información puntual.

En la simulación, se especifican dos tipos de clientes o demandas GPRS que son el WEB y el WAP. Además, el mecanismo de *scheduling* utilizado corresponde al EDF (*Early Deathline First*) [5][6][7][8], es decir, los paquetes que excedan un cierto tiempo serán descartados. Las representaciones gráficas que se ilustraran son funciones CDF (*Cumulated Distribution Function*) y los parámetros de *QoS* estudiados son la tasa de perdidas de paquetes y el tiempo medio de servicio.

Finalmente decir que el sistema se simulará con una carga del 60% (15 clientes web+30 wap) y otra del 80% (20 clientes web+40 wap), y con diferentes clases multislot (6/4, 4/2 y 2/1). El nombre de PDCh utilizados en la simulación corresponde a 6 y la codificación utilizada es la CS-2.

En la tabla 2 se especifican los resultados de la simulación en función de los parámetros de *QoS* correspondientes a la política de *scheduling* escogida.

		2/1	4/2	6/4
WEB	Charge 60%	Loss :4.68% Service:0.95s	Loss :4% Service:0.85s	Loss :3% Service:0.73s
	Charge 80%	Loss :6.88% Service:1.2s	Loss :4.83% Service:1.1s	Loss :4.35% Service:1s
WAP	Charge 60%	Loss :0.3% Service:0.1s	Loss :0% Service:0.07s	Loss :0% Service:0.47s
	Charge 80%	Loss :1.22% Service:0.15s	Loss :0.8% Service:0.13s	Loss :~0% Service:0.09s

Tabla 2. QoS para cada cliente y clase multislot

En las figuras 8 y 9 se puede observar el resultado del tráfico Wap y Web, utilizando la función normalizada CDF, para una carga del 60% y una clase multislot de 4/2.



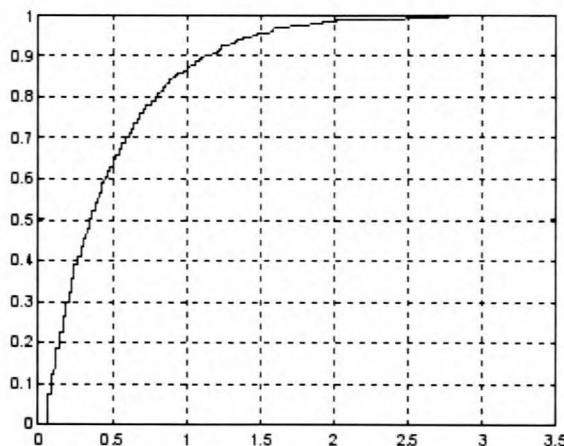


Figura 8. Tráfico Web (60%, 4/2)

La conclusión que se pueden obtener de la simulación es que la probabilidad de pérdidas de clientes Web es superior a la de Wap debido a que el número de paquetes por ráfaga entrantes en el sistema es más elevado en el caso Web que en el Wap (a pesar de que se gestionen más clientes Wap que Web). Además, se observa que el tiempo de servicio disminuye cuando más aumenta la clase multislot.

En las dos gráficas adjuntas se pude ver como el tráfico Wap aumenta de manera más abrupta que el Web (y los dos exponencialmente) debido a la carga de usuarios y, sobretodo, al bajo número de paquetes perdidos respecto al Web.

En el caso del Web se observa que los paquetes entrantes tienen un *deadline* más bajo y se descartan pronto del sistema.

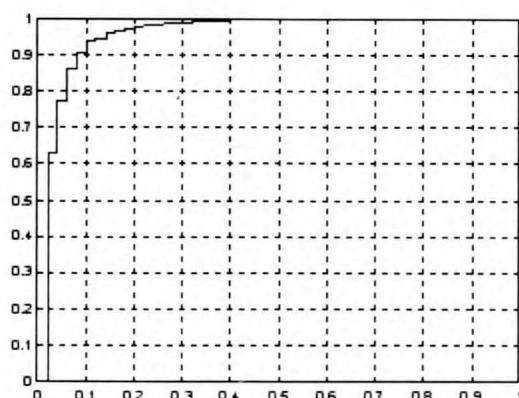


Figura 9. Tráfico Wap (60%, 4/2)

XI. CONCLUSIONES

El sistema de segunda generación GPRS es una red de datos basada en la transmisión de paquetes que soporta tanto

Internet, X.25 y otras redes públicas y privadas. Además, permite la transmisión asimétrica entre el *uplink* y el *downlink* que se gestionan de manera independiente [3].

GPRS realiza una optimización de los recursos radio mediante la asignación dinámica e independiente y la multiplexación de usuarios. Asimismo, GPRS introduce el concepto de conmutación de paquetes de datos extremo a extremo.

Este sistema ofrece una tarificación por volumen de datos intercambiados, *QoS* y tipo de servicio. Igualmente, es un sistema completamente compatible con la transmisión de voz vía GSM.

XII. REFERENCIAS

- [1] Philippe Godlewski, Xavier Lagrange y Philippe Martins. “L'accès paquet dans GPRS (General Packet Radio Services : L'accès Paquet da GSM“. ENST Paris. Versión B4. Octubre 2002.
- [2] Philippe Godlewski y Alexis Blavette. “Scheduling et Qualité de Services dans GPRS“. ENST Paris. 28 Junio 2002.
- [3] Ferran Casadevall y Oriol Sallent. “Sistema GPRS“. UPC Barcelona. Mayo 2002.
- [4] Ferran Casadevall y Joseph Bada. “Algoritmos del Scheduling para el sistema GPRS“. UPC Barcelona. Mayo 2002.
- [5] Peixiang Gong, Myung J. Lee y Tarek Saadawi. “A novel delay differentiation QoS Architecture”. Electrical Engineering Department. The City University of New York, Graduate Center and the City College. New York 2001 IEEE.
- [6] ECPE 4984 Real-Time Systems. Periodic Task Scheduling (dynamic priority). “Earliest Deadline First Scheduling Algorithm”.
- [7] Chengzhi Li y Edward W. Knightly. “Schedulability Criterion and Performance Analysis of Coordinated Schedulers”. Rice University. <http://www.ece.rice.edu/networks>.
- [8] Samih Hijwel, Naveed Hussain, Don Ngo y Avishkar Misra. “Advanced Packet Scheduling for Wireless Networks”. Group 15, Topic 4. 25 Mayo 2002.
- [9] Philippe Godlewski, Alexis Blavette. “GPRS Scheduling Software“. Versión 3.0. ENST. Modifiée en janvier. 2003 par Victor Martinez et Enric Chillaron.

AUTOR



Enric Chillaron i Farré nació en Lleida el 23 de agosto de 1979. Estudió Ingeniería Técnica de Telecomunicación, especialidad en Sistemas de Telecomunicaciones, en la Universidad de Vic. Titulado en el año 2000, realizó su Proyecto de Final de Carrera sobre GPS. Actualmente es estudiante de Ingeniería de Telecomunicación en la Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona. Desde el año 2001 pertenece a la Rama de Estudiantes del IEEE de Barcelona.



ESTUDIO DEL DESPLIEGUE DE RED UMTS EN INFRAESTRUCTURAS COMPARTIDAS

Josep Maria Esquius Toledano, Antoni Liria Righetti

I. INTRODUCCIÓN

En la última década del siglo XX se ha producido un gran cambio en todos los aspectos relacionados con las telecomunicaciones. De la misma manera que el telégrafo y el teléfono marcaron en su época un inicio en las comunicaciones, podemos considerar que la aparición de Internet y el desarrollo de las comunicaciones móviles han sido los precursores de un nuevo concepto de telecomunicación. Un concepto que busca integrar todos los tipos de telecomunicaciones y que ha de ser capaz de crear la llamada Sociedad de la Información, en donde Telecomunicación, Informática y Telemática formen el tejido de la sociedad.

El acceso a esta Sociedad de la Información debe adaptarse a los requisitos impuestos por los usuarios, en donde los sistemas de comunicaciones móviles juegan un papel muy importante. Estas comunicaciones deben proporcionar una gran capacidad de transmisión de datos para poder conseguir que éstas sean realmente unas verdaderas comunicaciones multimedia. Y la tecnología que permitirá este mercado masivo de servicios y aplicaciones con gran contenido será la UMTS.

La tecnología UMTS (Universal Mobile Telecommunication System), será la apropiada para ser usada por una gran variedad de usuarios y será capaz de ofrecer multitud de nuevos servicios que coexistirán con los actuales; es el caso de los servicios de voz, que mantendrán una posición privilegiada durante varios años.

Las grandes posibilidades tecnológicas que ofrece UMTS se reflejan directamente en el gran movimiento que está creando este sistema en el mercado. Así, los principales operadores de telefonía móvil ya han iniciado la implantación de las infraestructuras necesarias para la extensión y cobertura de sus redes UMTS. Este despliegue de red puede suponer unos elevados costes para los operadores y, además, teniendo en cuenta que generalmente no hay sólo un operador en el ámbito nacional con licencia de telefonía móvil UMTS, puede comportar problemas frente a autoridades locales, autonómicas o público en general.

Así, por tanto, la extensión de la nueva red de telefonía multimedia debería realizarse con el mínimo impacto urbanístico y el máximo respeto por el medio ambiente,

pero resulta una tarea complicada cuando el número de operadores con licencia es elevado.

El contenido de este artículo [1] va analizar las soluciones que pueden adoptar los operadores ante los problemas comentados. En primer lugar, se va a realizar un análisis del mercado de la telefonía móvil, con el objetivo de conocer si realmente es necesaria una nueva tecnología como es el sistema UMTS. Así, después de la pregunta ¿es necesaria una nueva tecnología?, se analizarán las preguntas ¿cuál es la situación actual del UMTS?, ¿cuál es su futuro?, ¿qué nos puede ofrecer? y se analizará un nuevo modelo de negocio, basado en la compartición de infraestructuras. En segundo lugar, se va a analizar el despliegue de una red UMTS; se va a realizar una descripción de los emplazamientos y su proceso de instalación y se va a analizar los costes del despliegue, comparando el despliegue realizado según el modelo tradicional y el despliegue según el modelo de la compartición de infraestructuras.

II. ANÁLISIS DEL MERCADO DE LA TELEFONÍA MÓVIL

Situación actual de la telefonía móvil

Un buen indicador para conocer el estado de la telefonía móvil en un país es el del nivel de penetración, que es el porcentaje de población del país que hace uso de la telefonía móvil. Tal y como se puede observar en el gráfico de la figura 1, el aumento de la penetración en España de la telefonía móvil ha sido espectacular. En 1997 era de casi el 11%, y en tan solo cuatro años ha alcanzado una cifra de más del 72% [2].

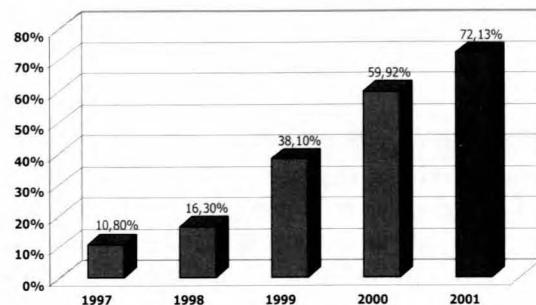


Figura 1. Evolución de la tasa de penetración en España



La evolución en la Unión Europea ha sido similar a la española. En la figura 2, se puede observar que en los principales países europeos la tasa de penetración se sitúa en torno al 75%, lo que implica que tres de cada cuatro europeos son usuarios de la telefonía móvil [3].

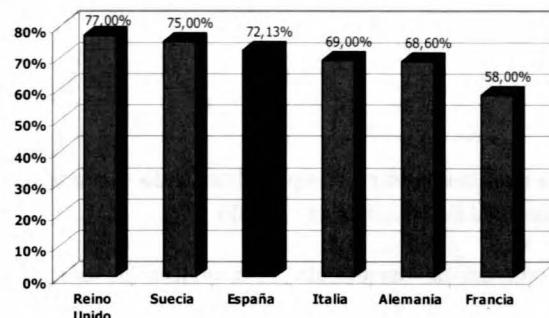


Figura2. Penetración de la telefonía móvil en Europa (2001)

Así pues, se puede afirmar que en Europa la telefonía móvil ha alcanzado el estado denominado como saturación del mercado, en el que los teléfonos móviles se han convertido en un elemento imprescindible para la población.

Evidentemente, el nivel de penetración va directamente ligado al tráfico cursado. En el gráfico de la figura 3, se observa como el tráfico medio por cliente y año ha ido descendiendo progresivamente [2][4]. En líneas generales, la tendencia ha sido que el tráfico total ha aumentado espectacularmente en los últimos años, pero que el tráfico medio por usuario ha descendido. ¿Cuáles son las causas de este descenso?

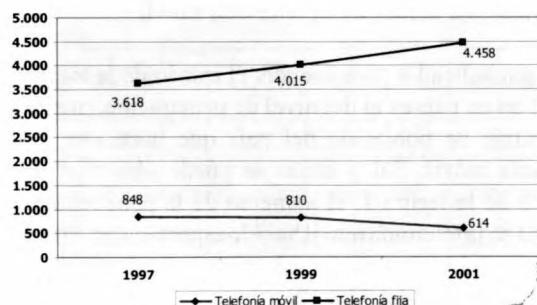


Figura3. Evolución tráfico medio por cliente y año (en minutos)

Se puede responder a esta pregunta analizando la relación existente entre la tasa de penetración y el ARPU (Average Rate Per User), que es el nivel medio de ingresos por usuario que recibe un operador de telefonía. Tal y como se puede observar en la figura 4 [3], cuanto más alto es el nivel de penetración más bajo es el ARPU. Esto es debido a que con el paso de los años, la telefonía móvil accede a

nuevos segmentos de población (generalmente cada segmento nuevo tiene menos poder adquisitivo), por lo que el uso medio del servicio disminuye.

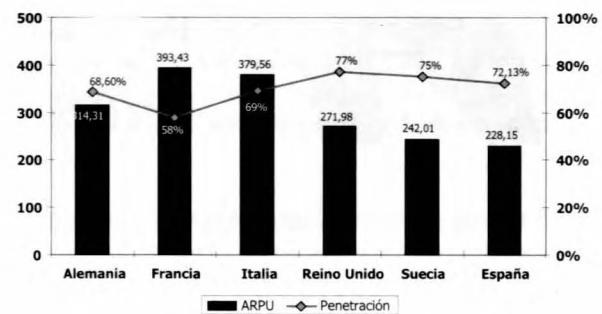


Figura4. Penetración del servicio y ARPU (euros)

Así, un aumento de la penetración implica una reducción del ingreso medio por cliente, lo que significa también un descenso del tráfico medio por cliente.

Si comparamos la evolución del tráfico de la telefonía fija con el de la móvil (figura 3), se puede apreciar como en el sector de la telefonía fija ésta es ascendente debido a que se encuentra ya en plena etapa de madurez y posee las características propias de esta fase: reducción de los precios y aumentos de tráfico cursado. Esta reducción del precio del servicio es consecuencia directa de la competencia existente en esta fase del ciclo de vida.

Situación y perspectivas del sistema UMTS

Existen varios motivos por los que el UMTS está creando una cierta situación de incertidumbre. De entre todos estos motivos, hay tres que son los más significativos.

En primer lugar, el alto coste que ha supuesto a los operadores de telefonía la adquisición de las licencias UMTS, ya que los gobiernos europeos, conscientes de la inmensa valoración que se da en estos momentos a la tecnología 3G, las han otorgado por grandes cantidades. En segundo lugar el retraso en el desarrollo de equipos y terminales UMTS, ya que la promoción del UMTS se produjo cuando esta tecnología aún estaba a medio desarrollar. Y en tercer lugar la dura competencia existente entre las empresas de telecomunicaciones, tanto de fabricantes de equipos como de operadores de telefonía.

Estos tres motivos infunden pocos alicientes para que los operadores de telefonía móvil realicen el despliegue de la tecnología UMTS. No obstante, existen otros motivos que indican unas buenas perspectivas para este sistema.

Volviendo de nuevo al concepto del ARPU, se puede observar la progresiva disminución del ARPU ocasionado por el tráfico de voz (figura 5).

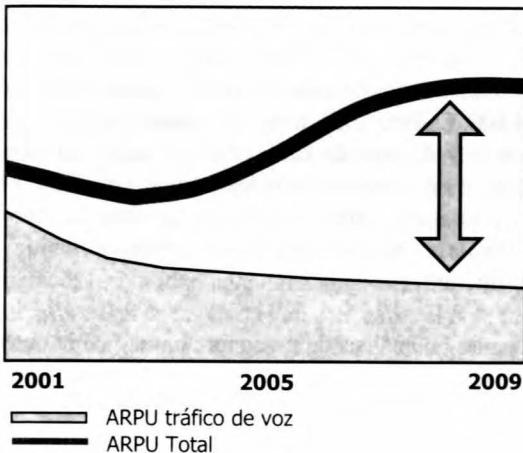


Figura 5. Proyección del ARPU para un operador de 3G en la UE

Aún así, el ARPU ocasionado por el tráfico de datos tiene un aumento espectacular que implica un aumento considerable del ARPU total.

Por lo tanto nos encontramos que en el futuro se espera un aumento del ARPU de datos y un descenso del ARPU de voz [5]. Este hecho producirá un gran aumento del tráfico de datos, por lo que es totalmente necesario una tecnología capaz de proporcionar con una buena calidad de servicio el comentado aumento de este tipo de tráfico.

La tecnología idónea actualmente para soportar este aumento del tráfico de datos es la tecnología GPRS, que al estar basada en GSM hace que su despliegue y puesta en marcha de equipos suponga una pequeña inversión. Pero el GPRS ha de ser una tecnología de transición, ya que en cuanto aumente la cantidad y calidad de los servicios de datos hará falta una tecnología como el UMTS.

Así pues, el GPRS será el responsable de infundir la cultura de uso de los nuevos servicios y ayudará a los operadores de telefonía a establecer los modelos de negocio para el UMTS.

Describir estos nuevos servicios que UMTS puede ofrecer es una tarea realmente complicada, ya que en primer lugar, muchos de estos servicios aún no están definidos, y en segundo lugar pueden presentar una variedad potencial muy grande. Lo que si está claro es que el UMTS podrá ofrecer servicios multimedia o servicios basados en el acceso a bases de datos o Internet. Y lo que es totalmente seguro es que algunos de los servicios que se están prestando actualmente se continuarán prestando con el UMTS. No obstante, el éxito de todos estos servicios pasa por que las operadoras de telefonía adopten nuevos modelos de negocio (figura 6) [6].

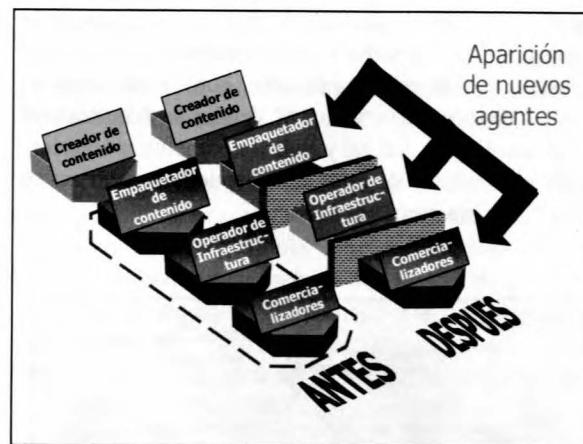


Figura 6. Nueva cadena de valor

Hasta el momento el operador de telefonía móvil era el responsable de gran parte de la cadena de valor: creaba contenidos o los reutilizaba de otras tecnologías adaptándolos al entorno de la telefonía móvil, poseía la red de infraestructuras y comercializaba todos los servicios. Pero el nuevo modelo de negocio propone la creación de mercados nichos de contenidos, infraestructuras y venta de servicios, con lo que se deben crear nuevas empresas especializadas en cada uno de estos aspectos. Así, el operador de telefonía puede centrarse exclusivamente en el nuevo “core” del negocio, es decir, el nuevo objetivo del negocio: el servicio; obtener nuevos clientes, fidelizarlos y proporcionarles numerosos contenidos pasa a ser su principal actividad.

III. ANÁLISIS DEL DESPLIEGUE DE RED UMTS

Proceso de instalación de emplazamientos

En la figura 7 se pueden observar todos los pasos a realizar en el despliegue de una red UMTS. El procedimiento general que se efectuaba hasta el momento es que el operador de telefonía realizara todos los pasos, excepto el de instalación de equipos e integración a la red, que lo realizaba el Vendor (fabricante de equipos).

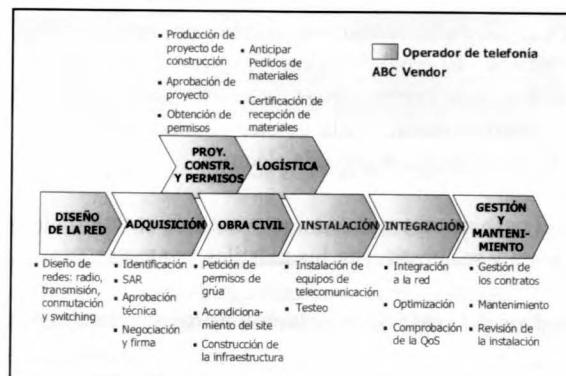


Figura 7. Procesos de instalación de emplazamientos (modelo tradicional)



Con el nuevo modelo de negocio (figura 8) se propone la desaparición total del operador de telefonía en el proceso de instalación de emplazamientos, con lo que entre el operador de infraestructuras y el Vendor se debe realizar todo el despliegue. Así, tal y como se comentaba anteriormente, el operador de telefonía se puede centrar en su nuevo “core business”, el servicio.

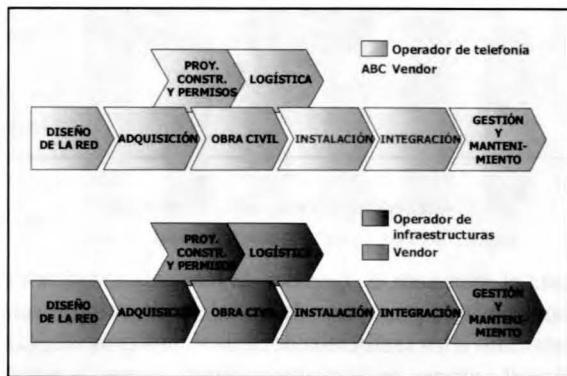


Figura8. Procesos de instalación de emplazamientos (modelo operador de infraestructuras)

Descripción de un emplazamiento

Generalmente, los emplazamientos se clasifican en dos tipos: urbanos y rurales. Los emplazamientos urbanos son todos aquellos situados en azoteas de viviendas, de oficinas o naves industriales de zonas urbanas, mientras que los emplazamientos rurales son aquellos situados en zonas rurales o semirurales en proximidades de autopistas y carreteras o zonas industriales (polígonos...).

Los emplazamientos urbanos pueden ser del tipo caseta y mástil, indoor y mástil u outdoor y mástil, mientras que los rurales pueden ser del tipo caseta y torre u outdoor y torre.

Los emplazamientos urbanos del tipo caseta y mástil constan de una caseta prefabricada de 5,5 m², que está anclada sobre una bancada de protección de la estructura del edificio en el cual está ubicado el emplazamiento. La caseta suele contener un cuadro eléctrico, sensores de alarma, aire acondicionado, cuadro de distribución digital, canaletas y conductos para el cableado, sistemas de tierras, luminarias, cerrojos, extintores... El emplazamiento se completa con un mástil para la instalación de antenas, diferentes tipos de seguridad de acceso al mástil y a la caseta, instalación de tierras y realización de una acometida eléctrica.

En el caso de un emplazamiento rural, la caseta suele ser de obra y se substituyen los mástiles por torres.

Análisis del coste de instalación de un emplazamiento

Para realizar el análisis de los costes del despliegue de una red UMTS, se empezará por analizar los costes de instalación de un emplazamiento.

Suponiendo que el emplazamiento a instalar es de tipo urbano, con una caseta y un mástil, se obtiene que los costes totales de instalación de dicho emplazamiento son de 54.100 €. Este total tiene en cuenta los costes de ingeniería y diseño, de obra civil, de adquisiciones y licencias y de “management fee”, cuyo significado se refiere a los gastos producidos por la gestión de todo el proyecto de instalación por parte del promotor de la infraestructura.

Realizando un análisis de inversión, basado en la obtención de una T.I.R. de, por ejemplo, el 12%, en el que se supone que el emplazamiento estará en servicio durante 15 años, que la vida útil de las instalaciones es también de 15 años, y que la tasa de devaluación del VNC (Valor Neto Contable) de los equipos es del 25% en el caso de su posible reutilización, se obtiene que los cobros mínimos que se deben obtener para la inversión de 54.100 €, a la T.I.R. considerada, y sin tener presente la renta del emplazamiento u otros recursos, son de 9.283 €. Si a esta cifra se le añade la renta de alquiler o compra del emplazamiento, que se supone de 7.500 € en el caso de un emplazamiento urbano, se obtiene que los cobros anuales mínimos que deben obtenerse son de 16.783 €.

En consecuencia, si el despliegue lo realiza el operador de telefonía, en la zona en donde se instala el emplazamiento se debe obtener de los usuarios en concepto de prestación del servicio unos ingresos anuales mínimos iguales a los acabados de calcular. En zonas de gran demanda de servicio se alcanzarán más fácilmente estos ingresos, pero en otras zonas existe la posibilidad de no alcanzarlos. No obstante, aunque se obtengan dichos ingresos, puede ser más recomendable la elección del operador de infraestructuras, ya que, tal y como se explica a continuación, puede ofrecer al operador tarifas más económicas.

Si el despliegue lo realiza el operador de infraestructuras, y en el emplazamiento en cuestión únicamente está instalado un operador, será éste quien deba satisfacer los ingresos a obtener por el operador de infraestructuras. Realizando una comparación con el caso anterior, en esta situación, el operador de telefonía sería el equivalente a los usuarios, y el operador de infraestructuras sería el equivalente al operador de telefonía.

Sin embargo, en los emplazamientos construidos por el operador de infraestructuras pueden coexistir varios operadores de telefonía, por lo que se puede repartir entre ellos el importe de los cobros que el operador de infraestructuras necesita ingresar. Así, cuantos más operadores de telefonía comparten emplazamiento, más bajas serán las tarifas que el operador de infraestructuras deberá aplicarles a cada uno de ellos.

Y no sólo los motivos económicos pueden hacer escoger esta última opción, ya que realizando el operador de infraestructuras el despliegue, el operador de telefonía puede centrarse exclusivamente en su “roll-out”, es decir,

en su principal objetivo, que es el de obtener usuarios y proporcionarles servicio.

Puede darse el caso que un operador necesite toda la caseta del emplazamiento para la instalación de sus equipos. A primera vista parece que en esta situación, dejando de lado el “roll-out” del operador de telefonía, no importe la opción escogida, ya que en ambos casos se debe satisfacer la cantidad calculada anteriormente. Pero esta reflexión es errónea, ya que en este caso, el operador de infraestructuras puede construir en el mismo emplazamiento dos o más casetas, tantas como sean necesarias para ubicar a los operadores de telefonía en ellas.

Evidentemente, la construcción de varias casetas (y más de un mástil) en el emplazamiento hace aumentar la cantidad a invertir. Sin embargo, el capital necesario para la construcción de varias casetas no es el mismo que el necesario para una multiplicado por el número de casetas a construir, ya que en este caso se obtienen unos costes totales de instalación de 77.881 €. Así, los costes de instalación no aumentan proporcionalmente al aumento del número de casetas, por lo que cuantos más operadores de telefonía y más casetas haya instaladas en un mismo emplazamiento, menores serán las tarifas que éstos deberán satisfacer al operador de infraestructuras.

Hay que tener presente que el único concepto que ha aumentado es el de obra civil. Este aumento se produce debido a que varias acciones a realizar en la obra civil han duplicado su precio, como el suministro de dos casetas en vez de una; pero cabe destacar que el concepto obra civil no se ha duplicado, puesto que tareas como la instalación de la red de tierras o de la acometida eléctrica, entre otras, no aumentan considerablemente su coste por instalarse dos casetas en el emplazamiento.

El coste del resto de conceptos no varía, puesto que el hecho de instalar más casetas o más mástiles en el emplazamiento, no hace que su coste aumente, como es el caso del management fee o la adquisición de licencias, en el que hay que realizar las mismas gestiones sean una, dos o varias casetas las que se tengan que instalar.

En esta nueva situación, la renta por el alquiler (o adquisición) de los terrenos en los que está ubicado el emplazamiento cambia. Obviamente, el espacio requerido para la instalación de dos casetas es superior al requerido para una. Aún así, esta renta no se duplica, por lo que consideramos que la renta por estos emplazamientos es de 12.000 € anuales. No obstante, en muchas ocasiones, la renta mantiene el valor del caso anterior, disminuyendo aún más los costes globales de instalación.

Así pues, con estos nuevos costes de instalación, si se realiza un análisis de inversión, con las mismas condiciones que el anterior análisis, se obtiene que los cobros mínimos que se deben obtener para la inversión de 77.881 €, a la T.I.R. considerada, y sin tener presente la renta del

emplazamiento u otros recursos, son de 13.364 €. Si a esta cifra se le añade la renta de 12.000 € comentada anteriormente, se obtiene que los cobros anuales mínimos que deben obtenerse son de 25.364 €.

Si se compara este último valor, con el obtenido para la instalación de una caseta (16.783 €) podemos observar que no ha aumentado el doble, sino aproximadamente 1,5 veces, con lo que se corrobora que los costes de instalación no aumentan proporcionalmente al aumento del número de casetas, por lo que, consecuentemente, cuantos más operadores de telefonía y más casetas haya instaladas en un mismo emplazamiento, menores serán las tarifas que éstos deberán satisfacer al operador de infraestructuras.

Análisis del coste de despliegue de una red UMTS en Cataluña

A continuación se va a evaluar los costes del despliegue de infraestructuras de una red UMTS en Cataluña. Ya no se va a tener únicamente en cuenta un emplazamiento genérico, sino que ahora se tendrán presentes Nodos B, HUBs o POCs. En este análisis no se incluirán los RNCs, ya que éstos emplazamientos se consideran puntos estratégicos del operador de telefonía, por lo que muy difícilmente se pueden compartir con otros operadores. Por este motivo, la inclusión de los RNCs en el estudio no aportaría ninguna ayuda en la comparación de los costes del despliegue entre el operador de infraestructuras y el de telefonía.

En la tabla 1 se muestra el número de elementos de red necesarios para realizar la primera fase del despliegue de la red UMTS. Están distribuidos según la provincia y el tipo de elemento de red (Nodo B, HUB o POC). Cabe destacar, que el 60% de los POCs y el 95% de los HUBs pueden compartir el emplazamiento con un Nodo B, por lo que el número de emplazamientos necesarios en cada una de las provincias se muestra en la tabla 2.

	NODO B	HUB	POC
BARCELONA	550	75	25
TARRAGONA	150	20	6
LLEIDA	75	10	3
GIRONA	105	15	4
TOTAL CATALUÑA	880	120	38

Tabla 1 Número y tipos de elementos de red requeridos

La instalación de estos emplazamientos se corresponde con la primera fase de despliegue. Es por este motivo, que se pretende proporcionar cobertura en primer lugar a zonas urbanas. Así, se considerará que el 60% de los emplazamientos son urbanos (según definición comentada anteriormente), mientras que el 40% se supondrán rurales (incluye emplazamientos con torre ubicados en polígonos industriales).



	BARCELONA	TARRAGONA	LLEIDA	GIRONA	CATALUNYA
Emplazamientos tipo Nodo B	464	127	64	89	744
Emplazamientos tipo HUB	4	1	1	1	6
Emplazamientos tipo POC	10	2	1	2	15
Emplazamientos tipo Nodo B + HUB	71	19	9	14	113
Emplazamientos tipo Nodo B + POC	15	4	2	2	23
TOTAL EMPLAZAMIENTOS	564	153	77	107	901

Tabla2. Número real y tipos de emplazamientos

En la tabla 3 se muestran los costes de instalación de un emplazamiento urbano y de uno rural; los costes de instalación del emplazamiento rural se han obtenido realizando un análisis de inversión con las mismas condiciones que los anteriores análisis, procediendo de similar manera que para el urbano.

	URBANO (1 caseta - 1 mástil)	URBANO (2 casetas - 2 mástiles)	RURAL
INVERSIÓN TOTAL	54.100	77.881	72.333
RENTA	7.500	12.000	3.750
TARIFA ANUAL	16.783	25.364	16.161

Tabla3. Coste de un emplazamiento urbano y rural (euros)

Con los todos los datos obtenidos hasta el momento, se está en condiciones de realizar el cálculo de los costes del despliegue de emplazamientos de una red UMTS en Cataluña.

En dicho cálculo, se considera que la inversión a realizar en un emplazamiento tipo Nodo B es similar a la de uno tipo HUB o tipo POC, por lo que los cobros anuales que deben obtenerse son los mismos en cada uno de los anteriores tipos de emplazamientos. No obstante, sí que se distinguirá entre emplazamientos rurales y urbanos, pudiendo ser estos últimos de una caseta y mástil o de dos casetas y dos mástiles.

La tabla 4 muestra el coste del despliegue de infraestructuras, suponiendo que todos los emplazamientos urbanos son del tipo una caseta y un mástil. Este tipo de despliegue (emplazamientos con únicamente una caseta), lo pueden realizar tanto el operador de telefonía como el operador de infraestructuras.

No hay que olvidar, que el operador de infraestructuras también tiene la posibilidad de construir emplazamientos con más de una caseta; en las siguientes tablas (tablas 5-7) se muestran los costes de realizar el despliegue de

emplazamientos del tipo una caseta y del tipo dos casetas, en función del número de emplazamientos de cada tipo.

	BARCELONA	TARRAGONA	LLEIDA	GIRONA	CATALUÑA
TOTAL EMPLAZAMIENTOS	564	153	77	107	901
EMPLAZAMIENTOS URBANOS	338	92	46	64	541
EMPLAZAMIENTOS RURALES	226	61	31	43	360
COSTE INSTALACIÓN EMPLAZAMIENTOS (euros)	9.325.289	2.529.733	1.273.133	1.769.159	14.897.314

Tabla4. Costes del despliegue de emplazamientos I (euros)

	BARCELONA	TARRAGONA	LLEIDA	GIRONA	CATALUÑA
TOTAL EMPLAZAMIENTOS	564	153	77	107	901
EMPLAZAMIENTOS URBANOS	70%	URBANOS (1 caseta - 1 mástil)	237	64	32
	30%	URBANOS (2 casetas - 2 mástiles)	102	28	14
EMPLAZAMIENTOS RURALES			226	61	31
COSTE INSTALACIÓN EMPLAZAMIENTOS (euros)	10.196.432	2.766.053	1.392.066	1.934.429	16.288.981
INCREMENTO RESPECTO DESPLIEGUE DE EMPLAZAMIENTOS CON UNA ÚNICA CASETA					9%

Tabla5. Costes del despliegue de emplazamientos II (euros)
70% emplazamientos de 1 caseta; 30% emplazamientos de 2 casetas

	BARCELONA	TARRAGONA	LLEIDA	GIRONA	CATALUÑA
TOTAL EMPLAZAMIENTOS	564	153	77	107	901
EMPLAZAMIENTOS URBANOS	50%	URBANOS (1 caseta - 1 mástil)	169	46	23
	50%	URBANOS (2 casetas - 2 mástiles)	169	46	23
EMPLAZAMIENTOS RURALES			226	61	31
COSTE INSTALACIÓN EMPLAZAMIENTOS (euros)	10.777.194	2.923.601	1.471.355	2.044.610	17.216.759
INCREMENTO RESPECTO DESPLIEGUE DE EMPLAZAMIENTOS CON UNA ÚNICA CASETA					13%

Tabla6. Costes del despliegue de emplazamientos II (euros)
50% emplazamientos de 1 caseta; 50% emplazamientos de 2 casetas

	BARCELONA	TARRAGONA	LLEIDA	GIRONA	CATALUÑA
TOTAL EMPLAZAMIENTOS	564	153	77	107	901
EMPLAZAMIENTOS URBANOS	0%	URBANOS (1 caseta - 1 mástil)	0	0	0
	100%	URBANOS (2 casetas - 2 mástiles)	338	92	46
EMPLAZAMIENTOS RURALES			226	61	31
COSTE INSTALACIÓN EMPLAZAMIENTOS (euros)	12.225.099	3.317.468	1.669.576	2.320.060	19.536.203
INCREMENTO RESPECTO DESPLIEGUE DE EMPLAZAMIENTOS CON UNA ÚNICA CASETA					24%

Tabla7. Costes del despliegue de emplazamientos II (euros)
100% emplazamientos de 2 casetas

A primera vista, parece más conveniente realizar un despliegue de emplazamientos con únicamente una caseta, ya que es sustancialmente más económico; así, como que el despliegue implica la misma inversión tanto para el operador de telefonía como para el operador de infraestructuras, no importaría quién realizará dicho despliegue. Pero lo acabado de comentar es doblemente erróneo.

En primer lugar, aunque la inversión a realizar sea la misma para los dos operadores, hay que tener en cuenta que el operador de infraestructuras puede promover la compartición de emplazamientos, por lo que, tal y como ya se ha comentado anteriormente, resultaría ventajoso para ambos.

Y en segundo lugar, aunque el coste de despliegue de emplazamientos con una o dos casetas sea superior, no implica necesariamente que dicho despliegue sea menos rentable que el anterior, ya que en este tipo de emplazamientos el operador de infraestructuras tiene la posibilidad de ubicar a un mayor número de operadores, con todo lo que ello significa.

Además, es importante destacar, que el aumento de los costes debido a incrementar el número de casetas por emplazamiento no es proporcional a dicho incremento. Efectivamente, tal y como se ha comentado anteriormente, los costes debido a la instalación de dos casetas son 1,5 veces mayor que los costes de instalación de una caseta. Este hecho provoca que el aumento de los costes totales de despliegue de emplazamientos solo se incrementen en un 23% en el caso de que todos los emplazamientos urbanos tengan dos casetas (tabla 9), corroborando que cuantos más operadores de telefonía y más casetas haya instaladas en un mismo emplazamiento, menores serán las tarifas que éstos deberán satisfacer al operador de infraestructuras.

Por los motivos acabados de comentar, se expone a continuación el ahorro que significa para un operador de telefonía que sea un operador de infraestructuras quien realice el despliegue de red, sea con emplazamientos con una única caseta o con emplazamientos con más de una.

Ahorro por alquiler de infraestructuras

Conocidos los costes del despliegue de emplazamientos, es interesante saber cuantitativamente el ahorro que supone escoger la opción del alquiler de infraestructuras.

No obstante, este ahorro depende del número de operadores que se instalen en el emplazamiento, porque, tal y como ya se ha comentado, cuantos más operadores y más casetas haya instaladas en un emplazamiento mejores tarifas podrá ofrecer el operador de infraestructuras.

Para evaluar el ahorro que le supone a un operador de telefonía encargar el despliegue de emplazamientos a un operador de infraestructuras se han desarrollado seis casos diferentes, en función del grado de expectativa de ocupación de los emplazamientos. En todos los casos se considera la situación analizada en la tabla 5, en la que el 70% de los

emplazamientos urbanos son de 1 caseta, mientras que el 30% restante son de dos casetas.

- Caso 1

Este primer caso se corresponde con un grado de expectativa de ocupación muy buena. Así, todos los emplazamientos con una caseta se suponen ocupados por dos operadores, mientras que todos los emplazamientos con dos casetas se suponen ocupados también por dos operadores por caseta. El ahorro de cada operador en esta situación se muestra en la tabla 8.

	URBANO (1 caseta)	URBANO (2 casetas)	RURAL
COSTE ANUAL DEL EMBLAZAMIENTO	16.783	25.364	16.161
NÚMERO DE OPERADORES INSTALADOS	2	4	2
AHORRO POR EMBLAZAMIENTO	50%	62%	50%
COSTE ANUAL POR EMBLAZAMIENTO CON AHORRO	8.392	6.341	8.081
NÚMERO DE EMBLAZAMIENTOS (CATALUÑA)	379	162	360
COSTE TOTAL DESPLIEGUE CON AHORRO	3.180.379	1.027.242	2.908.980
COSTE TOTAL DESPLIEGUE PARA OPERADOR DE TELEFONÍA			14.897.314
AHORRO TOTAL OPERADOR DE TELEFONÍA POR ALQUILER DE INFRAESTRUCTURAS			52%

Tabla8. Ahorro con muy buena expectativa de ocupación

- Caso 2

El segundo caso se corresponde con un grado de expectativa de ocupación buena. Así, todos los emplazamientos con una caseta se suponen ocupados por dos operadores, mientras que todos los emplazamientos con dos casetas se suponen ocupados por un operador por caseta. El ahorro de cada operador en esta situación se muestra en la tabla 9.

	URBANO (1 caseta)	URBANO (2 casetas)	RURAL
COSTE ANUAL DEL EMBLAZAMIENTO	16.783	25.364	16.161
NÚMERO DE OPERADORES INSTALADOS	2	2	2
AHORRO POR EMBLAZAMIENTO	50%	24%	50%
COSTE ANUAL POR EMBLAZAMIENTO CON AHORRO	8.392	12.682	8.081
NÚMERO DE EMBLAZAMIENTOS (CATALUÑA)	379	162	360
COSTE TOTAL DESPLIEGUE CON AHORRO	3.180.379	2.054.484	2.908.980
COSTE TOTAL DESPLIEGUE PARA OPERADOR DE TELEFONÍA			14.897.314
AHORRO TOTAL OPERADOR DE TELEFONÍA POR ALQUILER DE INFRAESTRUCTURAS			45%

Tabla9. Ahorro con buena expectativa de ocupación

- Caso 3

Este caso se corresponde con un grado de expectativa de ocupación medio-alto. Así, el 50% de emplazamientos con una caseta se suponen ocupados únicamente por un operador, estando ocupados por dos operadores el otro 50%, mientras que los emplazamientos con dos casetas se suponen ocupados por dos operadores por caseta. El ahorro de cada operador en esta situación se muestra en la tabla 10.

	URBANO (1 caseta)	URBANO (2 casetas)	RURAL
COSTE ANUAL DEL EMBLACAMIENTO	16.783	25.364	16.161
NÚMERO DE OPERADORES INSTALADOS	50% 1	50% 2	50% 1
AHORRO POR EMBLACAMIENTO	0% 0%	50% 24%	0% 0%
COSTE ANUAL POR EMBLACAMIENTO CON AHORRO	16.783 8.392	25.364 12.682	16.161 8.081
NÚMERO DE EMBLACAMIENTOS (CATALUÑA)	190 190	162 180	180 901
COSTE TOTAL DESPLIEGUE CON AHORRO	3.180.379 1.590.189	1.027.242 2.908.980	1.454.490 10.161.280
COSTE TOTAL DESPLIEGUE PARA OPERADOR DE TELEFONÍA			14.897.314
AHORRO TOTAL OPERADOR DE TELEFONÍA POR ALQUILER DE INFRAESTRUCTURAS			32%

Tabla10. Ahorro con expectativa de ocupación medio-alta

- Caso 4

El cuarto caso se corresponde con un grado de expectativa de ocupación medio. Así, el 50% de emplazamientos con una caseta se suponen ocupados únicamente por un operador, estando ocupados por dos operadores el otro 50%, mientras que los emplazamientos con dos casetas se suponen ocupados por un operador por caseta. El ahorro de cada operador en esta situación se muestra en la tabla 11.

	URBANO (1 caseta)	URBANO (2 casetas)	RURAL
COSTE ANUAL DEL EMBLACAMIENTO	16.783	25.364	16.161
NÚMERO DE OPERADORES INSTALADOS	50% 1	50% 2	50% 1
AHORRO POR EMBLACAMIENTO	0% 0%	50% 24%	0% 0%
COSTE ANUAL POR EMBLACAMIENTO CON AHORRO	16.783 8.392	25.364 12.682	16.161 8.081
NÚMERO DE EMBLACAMIENTOS (CATALUÑA)	190 190	162 180	180 901
COSTE TOTAL DESPLIEGUE CON AHORRO	3.180.379 1.590.189	1.027.242 2.054.484	1.454.490 11.188.522
COSTE TOTAL DESPLIEGUE PARA OPERADOR DE TELEFONÍA			14.897.314
AHORRO TOTAL OPERADOR DE TELEFONÍA POR ALQUILER DE INFRAESTRUCTURAS			25%

Tabla11. Ahorro con expectativa de ocupación media

- Caso 5

El grado de expectativa de ocupación de este caso es medio-bajo. Así, el 75% de emplazamientos con una caseta se suponen ocupados únicamente por un operador, estando ocupados por dos operadores el otro 25%, mientras que los emplazamientos con dos casetas se suponen ocupados por un operador por caseta. El ahorro de cada operador en esta situación se muestra en la tabla 12.

	URBANO (1 caseta)	URBANO (2 casetas)	RURAL
COSTE ANUAL DEL EMBLACAMIENTO	16.783	25.364	16.161
NÚMERO DE OPERADORES INSTALADOS	75% 1	25% 2	75% 1
AHORRO POR EMBLACAMIENTO	0% 0%	50% 24%	0% 0%
COSTE ANUAL POR EMBLACAMIENTO CON AHORRO	16.783 8.392	25.364 12.682	16.161 8.081
NÚMERO DE EMBLACAMIENTOS (CATALUÑA)	284 95	162 270	90 901
COSTE TOTAL DESPLIEGUE CON AHORRO	4.770.568 795.095	2.054.484 4.363.470	1.727.245 12.710.861
COSTE TOTAL DESPLIEGUE PARA OPERADOR DE TELEFONÍA			14.897.314
AHORRO TOTAL OPERADOR DE TELEFONÍA POR ALQUILER DE INFRAESTRUCTURAS			15%

Tabla12. Ahorro con expectativa de ocupación medio-baja

- Caso 6

Este último caso se corresponde con un grado de expectativa de ocupación baja. Así, todos los emplazamientos con una caseta se suponen ocupados por un operador, mientras que todos los emplazamientos con dos casetas se suponen ocupados por un operador por caseta. El ahorro de cada operador en esta situación se muestra en la tabla 13.

	URBANO (1 caseta)	URBANO (2 casetas)	RURAL
COSTE ANUAL DEL EMBLACAMIENTO	16.783	25.364	16.161
NÚMERO DE OPERADORES INSTALADOS	1	2	1
AHORRO POR EMBLACAMIENTO	0%	24%	0%
COSTE ANUAL POR EMBLACAMIENTO CON AHORRO	16.783	12.682	16.161
NÚMERO DE EMBLACAMIENTOS (CATALUÑA)	379	162	360
COSTE TOTAL DESPLIEGUE CON AHORRO	6.360.757	2.054.484	5.817.960
COSTE TOTAL DESPLIEGUE PARA OPERADOR DE TELEFONÍA			14.233.201
AHORRO TOTAL OPERADOR DE TELEFONÍA POR ALQUILER DE INFRAESTRUCTURAS			4%

Tabla13. Ahorro con expectativa de ocupación baja

Puede parecer que este ahorro es muy bajo. Pero hay que tener presente, que en todos los casos presentados, el operador de infraestructuras está asumiendo un riesgo nulo, ya que en las casetas en las que únicamente hay instalado un operador, aunque ocupe tan solo una pequeña parte del espacio disponible, se le aplica la tarifa del coste total anual del emplazamiento. Incluso en este caso, al operador de telefonía le interesa decantarse por esta opción, ya que sigue teniendo un ahorro que no tendría si fuera él quien realizara el despliegue. No obstante, el operador de infraestructuras tiene también la posibilidad de asumir algún riesgo, rebajando en estos casos las tarifas, y esperando la instalación de algún otro operador en el emplazamiento.

Pese a todo, no hay que olvidar, que en el caso de que el operador ocupe sólo una pequeña parte de la caseta, existe la posibilidad de instalar otro operador en ella, con lo que el ahorro de cada uno de ellos aumentaría sustancialmente.

IV. CONCLUSIÓN

Tal y como se acaba de exponer, el ahorro que le supone a un operador de telefonía encargar a un operador de infraestructuras el despliegue de red puede llegar a ser muy importante. Así, cuantos más operadores estén instalados en un emplazamiento, más alto será el ahorro de cada uno de ellos.

Por tanto, es importante que el operador de infraestructuras consiga ocupar sus emplazamientos con el máximo número de operadores posibles. Pero hay situaciones en que en determinados emplazamientos, el operador de infraestructuras solo consigue que se instale un único operador, a veces ocupando únicamente una pequeña parte de la caseta, e incluso se puede dar el caso de que no consiga instalar a ningún operador en el emplazamiento.

Generalmente, los ingresos que el operador de infraestructuras obtiene del operador de telefonía por el alquiler de un emplazamiento son proporcionales a la cantidad de espacio que éste ocupe en la caseta. Así, idealmente, si el operador ocupase el 50% de la sala, el operador de infraestructuras debería obtener en concepto de alquiler el 50% del coste anual del emplazamiento, ya que el otro 50% lo obtendría de la instalación de otros operadores en la caseta. Pero tal y como se acaba de comentar en el anterior párrafo, hay casos en que el otro 50% de la sala no llega a ocuparse por ningún operador.

Es por este motivo, que aunque el operador de telefonía ocupe la mitad de la caseta, el operador de infraestructuras debe ingresar más del 50% del coste anual del emplazamiento, ya que existe el riesgo de que la otra mitad de la caseta no se ocupe nunca por ningún otro operador.

Así, el operador de infraestructuras debe aplicar un incremento de sus tarifas en todos sus emplazamientos, ya que de esta manera puede compensar el riesgo de que varias cestas no lleguen a ocuparse totalmente por varios operadores e incluso compensar el riesgo de tener cestas vacías.

Aunque se aplique este aumento, los operadores de telefonía siguen teniendo un ahorro importante, obviamente no tan alto como en las situaciones anteriormente descritas, pero sí lo suficientemente representativo como para escoger la opción de alquiler de infraestructuras. Además, este aumento de tarifas puede significar para el operador de infraestructuras unos ingresos extras, íntimamente ligados al riesgo que asume, si finalmente consigue instalar a uno o varios operadores en los emplazamientos con el riesgo de quedar parcialmente ocupados o totalmente vacíos.

Se ha puesto de manifiesto, que la compartición de emplazamientos es económicamente ventajosa para los operadores de telefonía que adopten esta solución. A parte de reducir significativamente la inversión en infraestructuras, la solución de alquiler de emplazamientos permite a los operadores móviles que distribuyan sus necesidades financieras en el tiempo, ya que es el operador de infraestructuras quien realiza el despliegue.

Pero la compartición de infraestructuras no sólo comporta ventajas económicas para los operadores. A parte del ahorro sustancial que significa la compartición, esta solución ofrece otras ventajas como son la reducción del impacto urbanístico gracias a mimetizaciones, disminución de problemas de los operadores de telefonía frente a autoridades, administraciones públicas o la población e incluso mejorar su imagen corporativa al desentenderse de la propiedad de los emplazamientos, con lo que no se les asocia con posibles problemas que puedan ocurrir en ellos.

Muchos detractores de la solución de infraestructuras compartidas argumentan que esta solución no es la más conveniente puesto que existen operadores de telefonía que no la adoptan. No obstante, no hay que olvidar que el modelo de integración vertical que hasta ahora había sido característico del sector de las

comunicaciones móviles se está sustituyendo por un nuevo modelo, en el que el transporte de la señal, y por lo tanto la red de infraestructuras, se convierte en un eslabón más de la cadena de valor.

Así, hasta ahora, la red era el “core” del negocio de los operadores. Poseer y desplegar nuevas infraestructuras era su principal actividad. Pero actualmente, este modelo deja de tener valor, ya que el nuevo “core business” de los operadores móviles es el “servicio”; la obtención de nuevos clientes, la fidelización y el proporcionarles contenidos pasa a ser su principal actividad.

Por este motivo, ya en países como Estados Unidos o Australia, los operadores móviles han confiado el despliegue a operadores de infraestructuras, y es en estos momentos cuando también en Europa se empieza a apostar por esta solución.

Y es que con el transcurso de los años, cada vez más operadores de telefonía u otros operadores de servicios con tecnología radio irán adoptando la opción de instalarse en infraestructuras compartidas, puesto que ello les proporciona realmente muchas ventajas y facilidades que les permite centrarse en su principal objetivo y “olvidarse” del despliegue y gestión de la red de infraestructuras.

REFERENCIAS

- [1] J. M. Esquius. *Estudio del despliegue de red UMTS en infraestructuras compartidas*. Proyecto Final de Carrera ETSETB UPC. Diciembre 2002.
- [2] *Informe anual 2001 - Capítulo 2: situación de los mercados*. CMT. 25 de Julio de 2002.
http://www.cmt.es/cmt/centro_info/publicaciones/Inf%20Anual%202001/2000-capitulo-II.pdf
- [3] *El mercado de la telefonía móvil en la unión europea*. CMT. Mayo 2002.
http://www.cmt.es/cmt/centro_info/publicaciones/pdf/movil_europa.pdf
- [4] *La CMT analiza a fondo la competencia en el mercado de las comunicaciones móviles*. CMT. 16 de Enero de 2001.
http://www.cmt.es/cmt/centro_info/publicaciones/pdf/moviles.pdf
- [5] *mCommerce. Hacia un mundo sin hilos*. Arthur Andersen. Abril 2001.
- [6] B. Duarte. *UMTS: retos y perspectivas*. Revista de Telecomunicaciones de Alcatel. 1er Trimestre 2001.
<http://www.alcatel.es/Review/ReviewList.asp?id=18>

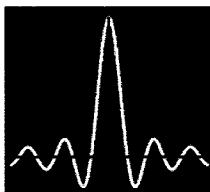
AUTOR



Josep M^a Esquius es ingeniero de telecomunicación y forma parte de la unidad de negocio de operadores del departamento comercial de TRADIA (L'Hospitalet del Llobregat, Barcelona).

Antoni Liria es ingeniero de telecomunicación, director de la unidad de negocio de operadores del departamento comercial de TRADIA y es profesor asociado en la ETSETB.





TECNOLOGÍA ULTRA-WIDEBAND (UWB)

LA REVOLUCIÓN A CORTO ALCANCE

Jordi Diaz

*Center for Communications and Signal Processing Research
New Jersey Institute of Technology*

jordi.diaz@njit.edu

1. INTRODUCCIÓN

El creciente agotamiento del espectro electromagnético, debido a la abundancia de aplicaciones radio, está fomentando el desarrollo de una nueva tecnología que promete reutilizar el espectro para proporcionar comunicaciones inalámbricas de corto alcance con capacidades nunca vistas hasta ahora. Ultra-Wideband, también conocido como "Impulse Radio", intenta escapar de las restricciones clásicas de ancho de banda para comunicaciones radio, explotando al máximo el concepto de espectro ensanchado. Sus impulsos afirman que UWB es capaz de utilizar bandas de frecuencia asignadas a otros servicios y funcionar con ellos sin causar interferencias perjudiciales.

El interés que ha despertado esta tecnología ha impulsado un proceso de revisión de las normativas que regulan las emisiones radio, tanto en los Estados Unidos como en Europa, para permitir que Ultra-Wideband pueda operar legalmente. Basándose en los recientes avances de las normativas en América, se puede esperar que la tecnología UWB abra las puertas al desarrollo de infinidad de aplicaciones de radar y comunicaciones de corto alcance. En un futuro, todas las cámaras podrían tener dispositivos auto-enfoque UWB, los coches podrían tener radares UWB para asistir en maniobras de aparcamiento. UWB podría convertirse en la tecnología dominante para comunicaciones inalámbricas multimedia en los hogares.

2. ¿QUÉ ES UWB?

Ultra-Wideband se define como cualquier transmisión radio que ocupe un ancho de banda mayor al 25% de la frecuencia central o más de 1.5Ghz. Aunque esta definición no sea muy descriptiva a primera vista, sí sirve para diferenciar UWB de sistemas tradicionales de banda estrecha, así como de los nuevos sistemas celulares de banda ancha, también conocidos como sistemas celulares de tercera generación (3G).

Existen dos diferencias fundamentales entre UWB y los mencionados sistemas de banda-estrecha y banda ancha. La primera y más obvia, es que los sistemas UWB utilizan un ancho de banda mucho mayor del que utiliza cualquier tecnología radio celular actual. La segunda es que UWB transmite pulsos en lugar de señales sinusoidales. Los

sistemas tradicionales utilizan portadoras de radio-frecuencia (RF) para desplazar la señal en el dominio frecuencial desde la banda base hasta la frecuencia de la portadora, donde el sistema tiene licencia para operar. Por el contrario, en los sistemas UWB la información modula directamente impulsos, muy cortos en tiempo, que generan una señal cuyo espectro puede ocupar desde pocos Hertzios hasta varios GHz. En este sentido, UWB es Banda Base.

Uno de los pulsos propuestos para transmisiones UWB es el monociclo Gaussiano, cuya representación en el dominio temporal y frecuencial se muestran en la Figura 1 y Figura 2 respectivamente. La modulación más comúnmente utilizada en los sistemas UWB es Pulse Position Modulation (PPM), principalmente por su capacidad de acceso multi-usuario y su resistencia a la propagación multi-camino. La sección 4 de este artículo describe la utilización de PPM para UWB.

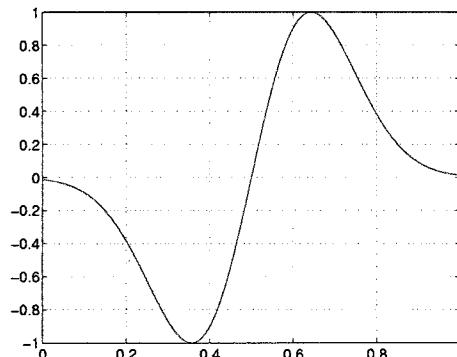


Figura 1. Monociclo Gaussiano en el dominio temporal.

Sin embargo, la característica que constituye una ruptura con todas las tecnologías anteriores, es que la señal UWB está diseñada para ocupar bandas de frecuencia asignadas a otros servicios sin causar interferencias ostensibles. Esto es posible porque los sistemas UWB emiten a una potencia muy baja que, repartida por un ancho de banda tan grande, produce una densidad espectral de potencia de pocos mW/Mhz. Estos niveles de señal son comparables a los niveles de ruido que soportan los receptores de otros sistemas. Se podría decir que la estrategia de UWB consiste en repartir su capacidad de interferencia entre

todos los demás servicios y por tanto, interfiriendo muy poco a cada uno de ellos.

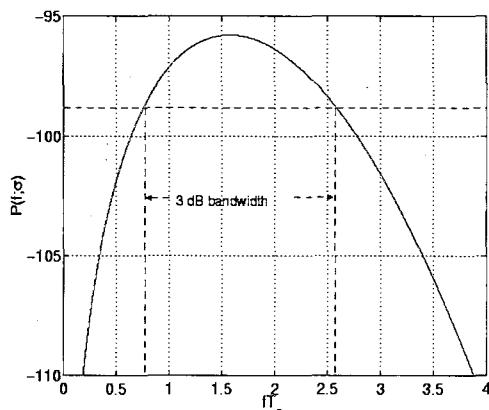


Figura 2. Monociclo Gaussiano en el dominio frecuencial

Sin duda alguna, el factor clave para determinar la inocuidad de la interferencia causada a otros sistemas, es el nivel de potencia en el que se permite operar a UWB. La sección 3 de este artículo presenta el espacio que ocupa UWB dentro de las normativas, tanto europeas como americanas.

Las restricciones de potencia en las que forzosamente tiene que operar la tecnología UWB, la hacen inviable para aplicaciones de largo alcance. Esto no supone ningún contratiempo, puesto que UWB está concebido para comunicaciones de corto alcance, principalmente para dar servicio indoor en oficinas, bibliotecas, hogares, aeropuertos, universidades, etc. En estos entornos la propagación multi-camino es un problema importante, ya que la señal rebota en muchos objetos antes de llegar al receptor. Algunos estudios demuestran que las señales UWB con modulación PPM soportan mucho mejor la propagación multi-camino que las señales sinusoidales [1][2]. Otro problema en ambientes indoor es que emisor y receptor difícilmente tienen visión directa y por tanto la señal tiene que atravesar objetos para llegar a su destino. UWB utiliza las frecuencias más bajas posibles; de esta forma mejora su capacidad de penetración en materiales que tienden a ser más opacos a frecuencias altas. Esta cualidad hace a UWB apropiada para otros propósitos aparte de las comunicaciones; tales como radares personales para detectar colisiones, sistemas de imagen que puedan ver a través de paredes, dispositivos precisos de localización, etc...

UWB ha despertado también el interés del gobierno norteamericano por su posible uso para aplicaciones militares. Tal como ocurriera con CDMA, las características que hacen de UWB una tecnología útil para el sector militar son su indetectabilidad[3] y su resistencia a interferencias de banda estrecha[4].

Así pues, tenemos una tecnología con dos cualidades fundamentales: gran ancho de banda y baja potencia. Estos atributos la habilitan a cohabitar con otros sistemas

de telecomunicaciones reutilizando el espectro; pero también la limitan para aplicaciones de corto alcance. La siguiente sección sitúa la tecnología UWB dentro de las tendencias actuales en comunicaciones radio para corto alcance.

3. EL CORTO ALCANCE.

Si el medio radio-eléctrico fuera ideal, podríamos usarlo para enviar mucha información, muy lejos, muy rápido, para muchos usuarios, todo al mismo tiempo. Desafortunadamente, las leyes físicas hacen imposible alcanzar estos cinco atributos simultáneamente. Debemos comprometer uno de ellos como mínimo si queremos alcanzar otro.

En los inicios de las comunicaciones radio, los usuarios decidieron que la habilidad de enviar información muy lejos era el atributo más importante. Marconi gustosamente comprometió los otros cuatro atributos cuando mandó la primera transmisión radio transatlántica de la historia en Diciembre de 1901. Sin embargo, en los 100 años transcurridos desde entonces, se ha producido una clara tendencia enfocada a mejorar los otros cuatro atributos a expensas de la distancia. El ejemplo más claro es la telefonía celular, que puede cubrir distancias desde 30 kilómetros hasta tan pequeñas como 300 metros. Estas distancias tan cortas son útiles sólo cuando son soportadas por estructuras cableadas; en este caso la red telefónica existente.

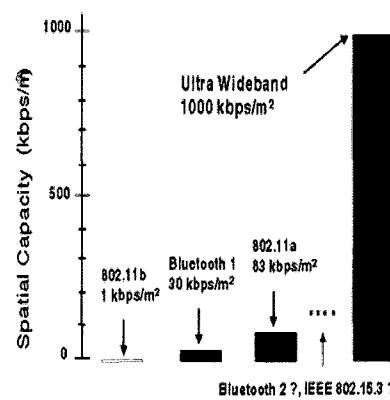


Figura 3. Comparación de capacidad espacial para sistemas radio de corto alcance

En los últimos años sistemas de más corto alcance (de 10 a 100 metros) han empezado a aparecer impulsados por aplicaciones de datos. En lugar de la red telefónica, Internet es la infraestructura cableada de soporte para este tipo de aplicaciones. Son muchos los que creen que, la combinación de comunicaciones radio a corto alcance y Internet cableado, puede convertirse rápidamente en el complemento perfecto para la nueva generación de sistemas celulares para datos, voz, audio y vídeo. Cuatro son



los factores que están impulsando las comunicaciones radio de corto alcance en general y UWB en particular:

- 1)La creciente demanda de servicios de datos, para entornos móviles, con mayor ancho de banda pero con menor coste y consumo energético que los existentes actualmente.
- 2)Super-explotación del espectro radioeléctrico, que se encuentra fragmentado por las autoridades normativas de la manera tradicional.
- 3)La popularización de acceso a Internet de alta capacidad en empresas, hogares y espacios públicos.
- 4)Continua bajada del coste de semiconductores así como de su consumo de potencia para procesado de señal.

Los factores 1 y 2 favorecen sistemas que ofrecen, no sólo altas velocidades de pico, sino también capacidad espacial; donde capacidad espacial se define como bits/seg/metro-cuadrado . De la misma forma que la red telefónica permitió la telefonía celular, el factor 3 hace posible la provisión de gran ancho de banda para dispositivos móviles que usen estándares para comunicaciones sin cables a corto alcance tales como Bluetooth o IEEE 802.11. Finalmente, el factor 4 hace posible el uso de técnicas de procesado de señal que eran impracticables hace tan sólo unos años. Es precisamente este último factor el que hace posible la tecnología Ultra-Wideband.

Los emergentes estándares para comunicaciones radio para corto o medio alcance, llevan implícitas unas ciertas capacidades espaciales. Por ejemplo:

- **IEEE 802.11b** tiene un rango de operación de 100 metros. En la banda ISM 2.4Ghz, hay cerca de 80Mhz de espectro libre. Por lo tanto, en un círculo de radio 100 metros, tres IEEE 802.11b a 22Mhz pueden operar sin interferencias, cada uno de ellos ofreciendo una velocidad máxima de 11Mbps. La suma total de 33Mbps, dividido por el área del círculo, resulta en una capacidad espacial de 1000bits/seg/metro-cuadrado.

- **Bluetooth**, en su versión de baja potencia, opera en un rango de 10 metros y tiene una velocidad máxima de 1 Mbps. Algunos estudios han mostrado que aproximadamente 10 «piconets» Bluetooth pueden operar simultáneamente en el mismo círculo de radio 10 metros, con mínima degradación. Esto resulta en una velocidad de 10Mbps, que dividido por el área del círculo produce una capacidad espacial de aproximadamente 30.000 bits/seg/metro-cuadrado.

- **IEEE 802.11a** está concebido para tener un rango de operación de 50 metros y una velocidad máxima de 54Mbps. Considerando los 200Mhz de espectro libre en

la parte baja de la banda U-NII a 5Ghz, 12 sistemas pueden operar simultáneamente con una aceptable degradación para conseguir una velocidad total de 648Mbps. Por tanto, la capacidad espacial prevista para este sistema es aproximadamente 83.000 bits/seg/metro-cuadrado.

Diversas son las capacidades proyectadas para sistemas UWB. Por ejemplo, un desarrollador de UWB ha medido velocidades máximas de 50Mbps en un rango de 10 metros. Seis sistemas pueden operar en la misma área de 10 metros de radio con mínima degradación. Siguiendo el mismo procedimiento, la capacidad espacial para el sistema UWB resulta aproximadamente de 1.000.000 bits/seg/metro-cuadrado.

Otros estándares aún en desarrollo en el “Bluetooth Special Interest Group” y el “IEEE 802 Working Group”, pueden elevar considerablemente las capacidades de sus respectivos sistemas, pero no parece probable que puedan alcanzar la capacidad espacial de UWB, tal como muestra la Figura 3.

$$C = B \log_2 \left(1 + \frac{S}{N} \right)$$

Figura 4: Teorema de la capacidad de Shannon

Una justificación para ello es que todos los sistemas están limitados por el teorema de la capacidad del canal de Shannon descrito en la Figura 4. Debido a que el límite superior de la capacidad de un canal crece linealmente con el ancho de banda disponible; los sistemas UWB, que ocupan 2Ghz o más, tienen más posibilidades de expansión que otros sistemas más limitados en ancho de banda.

Para aprovechar al máximo el ancho de banda, UWB utiliza la modulación PPM, cuyos fundamentos se presentan en la siguiente sección.

4. MODULACIÓN PPM-UWB

Las modulaciones de amplitud o frecuencia/fase no son apropiadas para este tipo de comunicaciones por impulsos. Por este motivo la modulación más utilizada en comunicaciones UWB es PPM (Pulse Position Modulation). PPM es una forma de modulación temporal en la que cada símbolo de información modula la posición en tiempo de un pulso. Los sistemas PPM-UWB utilizan trenes de pulsos, no solo un único pulso, para comunicar un símbolo. El intervalo entre pulsos se controla pulso por pulso en función de 2 componentes: una componente de información y un código pseudo-aleatorio. La componente pseudo-aleatoria se utiliza para repartir la energía en el dominio frecuencial, habilitar el acceso multi-usuario y para hacer el sistema más resistente a interferencias externas.

Como ejemplo véase la Figura 5 en la que se muestran 3 posibles señales UWB. En las tres señales se ha elegido transmitir 2 pulsos por bit. La primera corresponde a la transmisión de la secuencia de información [101] por parte

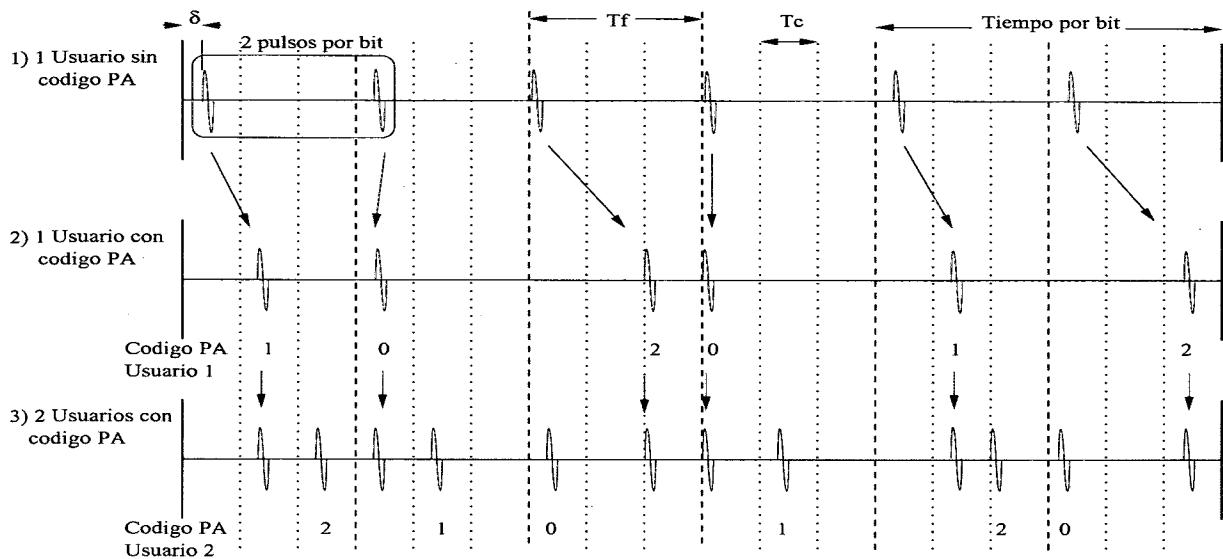


Figura 5. Ejemplo de 3 señales Ultra-Wideband con modulación PPM

	Información	Código
Usuario 1	101	102012
Usuario 2	110	210120

Tabla 1. Secuencia de información y códigos pseudo-aleatorios de los usuarios de la señal 3

de un sólo usuario. Se puede observar que los pulsos correspondientes al bit '1' están desplazados d respecto a la posición que tendrían si fueran un cero. En esta señal se utiliza la componente pseudo-aleatoria. La segunda señal de la Figura 5 contiene la misma información que la señal 1, pero en este caso se utiliza la componente pseudo-aleatoria. La utilización de la secuencia PA permite el acceso multiusuario tal como se muestra en la señal numero 3. La información y secuencias PA utilizadas en esta señal se incluyen en la Tabla 1. A partir de este ejemplo se pueden definir algunos de los parámetros básicos de una señal PPM-UWB.

- **Ns** : Número de pulsos por bit. En el ejemplo Ns=2. En casos reales se utilizan valores desde 200 a 2000 pulsos por bit.
- **d** : Tiempo de desplazamiento por información. Valor típico 0.156ns.
- **Tf** : Tiempo de repetición de pulso. En casos reales suele ser entre 100 y 1000 veces mayor que la duración del pulso.
- **Tc** : Tiempo de desplazamiento por código.

Con estos parámetros resulta evidente calcular la velocidad de transmisión $R_s = 1/T_f \cdot N_s$. Una descripción más exhaustiva de la señal UWB generada por la modulación PPM puede encontrarse en [5].

La detección óptima en entornos multi-usuario conlleva un aumento en la complejidad del diseño de los receptores. Sin embargo, si el número de usuarios es grande y la detección multiusuario es irrealizable, entonces es razonable aproximar el efecto combinado de los otros usuarios por un proceso aleatorio

Gaussiano. En este caso, el receptor óptimo para UWB es el receptor de correlación.

El detector de correlación multiplica la señal de entrada por un pulso "plantilla". La señal que sale del multiplicador se integra y el valor de la integración, de todos los pulsos que forman un bit, es el que determina la decisión entre 0 o 1. Si el pulso recibido es $r(t)$, el pulso plantilla acostumbra a ser $v(t) = r(t) - r(t-d)$. En la Figura 6 se muestra un pulso recibido. La Figura 7 muestra el pulso plantilla correspondiente, eligiendo $d=0.156\text{ns}$.

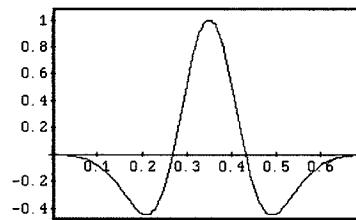


Figura 6. Típico pulso recibido en función del tiempo en nanosegundos

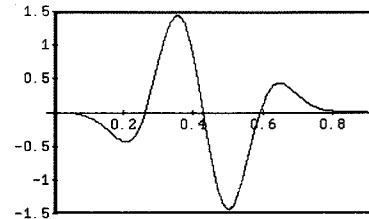


Figura 7: Pulso plantilla $v(t) = r(t) - r(t-d)$ diseñado para recibir señales UWB con parámetro $d = 0.156$ nanosegundos

Tal como es de esperar, la correlación de estos dos pulsos tiene un valor máximo cuando no existe retardo entre ellos. Por el contrario, el valor de la correlación entre ambos pulsos es mínimo cuando el pulso recibido llega retardado d. El valor umbral de decisión es 0. Esto significa que para valores positivos de la correlación se decide que el bit recibido es 0 y para valores negativos de la correlación se decide que el bit recibido es 1.

Una de las ventajas de los sistemas UWB con modulación por impulsos es su inherente robustez al fading multi-camino^[1]. Intuitivamente, esto puede explicarse de la siguiente manera. El fading multi-camino se produce por la interferencia destructiva causada por la suma de señales que, al llegar por diferentes caminos, pueden estar desfasadas unas de otras. Los pulsos extremadamente estrechos de UWB que resultan de las reflexiones en la propagación multi-camino, son tratados de forma independiente en el receptor, en lugar de combinarse destructivamente. Como resultado, la atenuación varía con el tiempo, que tanto daño hace a los sistemas de banda estrecha, se reduce significativamente por la naturaleza de la forma de onda UWB.

Los sistemas UWB parecen tener un gran potencial para soportar futuros sistemas sin cables de alta capacidad. Sin embargo, existen aún importantes desafíos que afrontar antes de que esta tecnología sea una realidad. Uno de ellos es encontrar la forma de hacer que esta tecnología sea legal sin causar inaceptables interferencias a otros usuarios compartiendo el espectro. Sobre este tema trata la próxima sección.

5. NORMATIVA

La regulación será el factor crítico que determine qué tipo de dispositivos UWB se desarrollarán en el futuro. Las autoridades normativas Europeas aún no han afrontado la regulación de sistemas UWB. Todo lo contrario está ocurriendo en los Estados Unidos, donde la FCC (Federal Communications Commission) ha estado activa en el área de UWB desde hace varios años. En setiembre de 1998, la FCC inició un procedimiento llamado «Notice of Inquiry» para investigar la operatividad de sistemas radio Ultra-Wideband sin licencia bajo el apartado 15 de su normativa. El apartado 15 de la normativa de la FCC establece límites para emisiones, tanto intencionadas como no intencionadas, en bandas sin licencia.

Más recientemente, el 11 de Mayo de 2000, la FCC publicó una «Notice of Proposed Rule Making» (NPRM) con la intención de revisar el apartado 15 de su normativa para incluir a sistemas UWB. En su NPRM, la FCC proponía que provisionalmente las emisiones UWB por encima de 2 GHz deberían cumplir con los límites del apartado 15, y que las emisiones UWB por debajo de 2 GHz deberían atenuarse 12dB por debajo de los límites de dicha norma. Este último límite proporcionaría protección adicional a servicios en el congestionado espectro por debajo de 2 GHz.

Una de las preocupaciones de las autoridades reguladoras son las interferencias que pueden causarse por la proliferación de dispositivos UWB [6]. La FCC se ha comprometido a asegurar que los servicios de seguridad y asistencia, en particular el sistema de posicionamiento global (GPS), estén protegidos de interferencias UWB. GPS es un componente esencial de muchos sistemas, tales como sistemas de navegación para aviación, y es utilizado como referencia de tiempo para varios sistemas de telecomunicaciones. De hecho, una organización

americana, la National Telecommunications and Information Administration, está realizando tests para analizar la interferencia provocada por dispositivos UWB y de esta forma entender los efectos que esta interferencia provoca en los receptores GPS.

Como resultado de estos y otros estudios la FCC emitió, el pasado 14 de febrero de 2002, su primer «Report and Order for UWB technology», que autoriza el desarrollo comercial de la tecnología UWB para usos limitados[7]. También se ha comprometido a revisar los estándares en los próximos 6-12 meses para explorar la posibilidad de estándares más flexibles que permitan el desarrollo de otro tipo de dispositivos UWB. Por ahora, los dispositivos para comunicaciones UWB están restringidos a operar bajo la normativa para operaciones intencionadas entre 3.1 y 10.6GHz. La normativa de la FCC constituye sólo el primer paso en el camino que Ultra-Wideband tiene que recorrer para hacer realidad todo su potencial.

REFERENCIAS

- 1.-M. Z. Win and R. A. Scholtz, "On the robustness of Ultra-Wide Bandwidth Signals in Dense Multipath Environments," IEEE communications letters, vol. 2, no. 2, february 1998.
- 2.-M. Z. Win and R. A. Scholtz, "Ultra-Wide Bandwidth Signal Propagation for indoor Wireless Communications," in Proc. IEEE Int. Conf. Communications, vol. 1, Montréal, Canada; June 1997, pp. 56-60.
- 3.-G.D. Weeks, J.K. Townsend, and J.A. Freebersyser, "Quantifying the covertness of pulse radio," 1999 International UWB Conference, Sep. 1999.
- 4.-L. Zhao and A.M. Haimovich, "Interference suppression in ultra-wideband communications" Proceedings of 35th Annual Conference on Information Sciences and Systems (CISS '01), pp. 759-763, Mar. 21-23 2001, Baltimore, Maryland, USA.
- 5.-R.A. Scholtz, "Multiple access with time-hopping impulse modulation" Proceeding of IEEE MILCOM'93, vol. 3, Oct. 1993.
- 6.-Multiple Access Communications Ltd.: 'An Investigation into the Potential Impact of Ultra-wideband Transmission Systems', August 1999, www.macltd.com, report commissioned by the UK Radiocommunications Agency.
- 7.-El Vídeo de la reunión de la FCC se encuentra disponible en:<http://www.fcc.gov/realaudio/agendameetings.html>

AUTOR



Jordi Diaz es ingeniero de telecomunicación por la Universitat Politècnica de Catalunya. Durante el año 2002 realizó su proyecto final de carrera sobre comunicaciones Ultra-Wideband en el Center for Communications and Signal Processing Research (New Jersey, EE.UU). Anteriormente trabajó para el Col.legi d'arquitectes de Catalunya y al-pi Telecomunicacions. Actualmente es estudiante de doctorado en el New Jersey Institute of Technology.



SISTEMA DE DETECCIÓN DE GRIETAS BASADO EN EL ANÁLISIS ESPECTRAL DE LA SEÑAL DE RESONANCIA

Xavier Castell^(*), Margarita Cabrera^(*), Rafael Montoliu^(**)

(*) Signal Theory and Communications Department. Universitat Politècnica de Catalunya.

Campus Nord D5. Jordi Girona 1-3. 08034-Barcelona-SPAIN.

(**) Aleaciones de Metales sinterizados: ALME-AMES. Sant Vicenç dels Horts. Barcelona Spain.

Contact mail: marga@gps.tsc.upc.es

Resumen.- En este artículo se presenta un sistema de detección de grietas sobre piezas metálicas sinterizadas, diseñado como estrategia de control de calidad, para su introducción en tiempo real dentro de la cadena de producción de dichas piezas. El algoritmo de procesado de señal se halla basado en el análisispectral de la señal medida a través de técnicas de resonancia ultrásónica. El espectro se obtiene mediante la aplicación del algoritmo Chirp. A partir de los máximos del mismo, coincidentes con las frecuencias de resonancia de las piezas se calcula la distancia euclídea respecto al vector de frecuencias de resonancia esperado. Mediante la probabilidad de falsa alarma esperada se apartan finalmente las piezas clasificadas como defectuosas. El sistema ha sido testeado en diferentes entornos de producción y con diferentes tipos de piezas, dando en todas las situaciones resultados satisfactorios.

1. INTRODUCCIÓN

La inspección por resonancia ultrasónica es una técnica ampliamente utilizada para la detección de piezas metálicas (especialmente sinterizadas) defectuosas. En un proceso de producción en serie, resulta un proceso de gran interés, debido a que consiste en una estrategia no destructiva. Cada pieza producida es excitada a frecuencia de ultrasonidos para vibrar en un determinado ancho de banda. La señal de resonancia queda de este modo modulada por las frecuencias de resonancia naturales de la pieza producida, de modo que el espectro de dicha señal constituye una especie de firmapectral de la pieza o de señal de identificación. Cuando se analiza una pieza diferente a la serie de piezas en cuestión, (pieza con diferente densidad de materiales, pieza agrietada, etc.), se produce un cambio en el conjunto de frecuencias de resonancia obtenidas. El éxito de la inspección por resonancia se basa en el hecho de que el espectro de resonancia es muy sensible a los cambios producidos en la composición o en la estructura. Así pues, como paso previo a la clasificación de piezas propiamente dichas se debe obtener un modelo para las frecuencias de resonancia que caracterice a una clase o serie de piezas no defectuosas, admitiendo cierta variación. De este modo, una pieza defectuosa podrá ser identificada mediante un cambio en sus frecuencias de resonancia. En general el cambio producido es proporcional al tamaño del defecto.

El presente trabajo, ha sido desarrollado para su implantación como sub-sistema de control de calidad, dentro de un sistema de fabricación en serie de piezas metálicas. Las técnicas algorítmicas seleccionadas para el análisispectral, la detección de picos espectrales y el algoritmo de clasificación, constituyen un compromiso entre un sistema de detección de defectos eficiente y unas etapas de procesado de señal computacionalmente rápidas.

En la sección 2 presentamos el esquema general del sistema diseñado. En la sección 3 se describen brevemente las etapas de procesado de señal. En la sección 4 se desarrolla el cálculo de la probabilidad de falsa alarma (es decir, la probabilidad de clasificar una pieza defectuosa cuando no lo es), y finalmente en la sección 5 se presentan algunos resultados interesantes de análisis de las piezas.

2. ESQUEMA GENERAL

En la figura 1 se muestra un esquema general del sistema desarrollado. Como señal de excitación se genera digitalmente una señal de ruido blanco en un determinado ancho de banda y posteriormente a través de un conversor D/A, se obtiene la señal analógica de ruido de espectro plano. El máximo ancho de banda permitido es de 200 KHz. Dicha señal se filtra paso bajo, con propósitos de reconstrucción y se amplifica con el objetivo de excitar directamente a un transductor emisor piezoelectrónico que hace vibrar la pieza en cuestión. A partir de la vibración dicha pieza, se obtienen dos señales eléctricas de resonancia, las cuales se miden mediante dos transductores receptores, situados correctamente bajo la pieza a testear. Las señales se digitalizan a una frecuencia de conversión de 625 KHz cada una de ellas.

Sobre las dos señales de resonancia capturadas a través de sendos conversores A/D, se aplica sobre ellas un algoritmo de análisispectral con la finalidad de obtener a partir de sus máximos las frecuencias de resonancia. El algoritmo seleccionado, consiste en la transformada Chirp, cuya implementación se realiza basada en la FFT. La transformada Chirp permite seleccionar un determinado ancho de banda en el cual se concentran todos los puntos del espectro, mejorando de este modo la precisión en la obtención de muestras del espectro. Como resultado de aplicar la transformada Chirp sobre las dos señales de resonancia, se obtienen dos vectores en el dominio de la



frecuencia. A partir de los mismos se han de calcular las frecuencias de resonancia. La etapa de detección de máximos o picos espectrales puede considerarse especialmente crítica, debido al carácter aleatorio y ruidoso del espectro obtenido para cada una de las dos señales de resonancia. Es decir, en la práctica aparecen numerosos máximos locales y con gran variedad y variación en cuanto a sus amplitudes. Por tanto y como paso previo a la detección de los máximos espectrales o frecuencias de resonancia, se aplica un algoritmo de “limpieza” de picos espectrales esporádicos, de baja potencia en general. Dicho algoritmo se conoce con el nombre de “Top Hat”. A partir de las señales procesadas mediante “Top Hat”, se estiman directamente las frecuencias de resonancia las cuales se introducen directamente a la etapa de clasificación propiamente dicha, donde finalmente se decide si la pieza se clasifica como defectuosa o no.

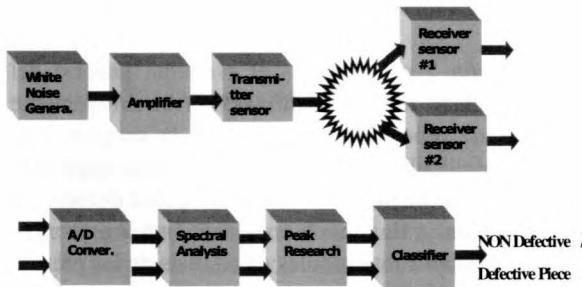


Figura 1. Diagrama General de Bloques

3. ETAPAS DE PROCESADO DE SEÑAL

Durante la primera etapa de investigación del proyecto se probaron diferentes técnicas de estimación espectral, tanto paramétricas como no paramétricas [1], con el fin de obtener las frecuencias de resonancia. Para la presente aplicación, la técnica que produjo mejores resultados en cuanto a un compromiso de resolución, precisión y eficiencia computacional fue el uso de la transformada Chirp [2], promediada según la técnica de estimación espectral no paramétrica del periodograma. La etapa de clasificación propiamente dicha (ver figura 1), se tuvo que resolver con mayores dificultades, debido a la imposibilidad de modelar el comportamiento en cuanto a frecuencias de resonancia de piezas defectuosas. Se observó gran variabilidad que el reducido conjunto disponible de dichas piezas, al ser analizado rigurosamente.

3.1 Análisis Espectral.

La transformada Chirp se aplica sobre cada una de las dos señales digitalizadas mediante los conversores A/D: $x_1[n]; x_2[n]$. Los vectores capturados a partir de una única pieza se dividen en L segmentos de N_B muestras cada uno.

$$x_{i-l}[n] = x_i[n + (l-1)N_B]; \quad i=1, 2; l=1..L; 0 \leq n \leq N_B - 1 \quad (1)$$

La transformada Z se aplica sobre cada segmento de N_B muestras:

$$S_{i-l}(z) = \sum_{n=0}^{N_B-1} x_{i-l}[n] z^{-n}; \quad i=1, 2 \quad (2)$$

Fijando un punto inicial en el dominio transformado:

$z_0 = r_0 e^{jw_0}$ y un término de incremento frecuencial $W = e^{j\Delta w}$, la ecuación (2) se puede re-escribir como (3) cuando se evalúa para una frecuencia simple: $z_k = z_0 W^k$.

$$S_{i-l}(z_k) = \sum_{n=0}^{N_B-1} x_{i-l}[n] (z_0)^{-n} W^{-kn} \quad (3)$$

$k=1..M-1$. M designa el número de puntos del espectro o frecuencias dentro del ancho de banda seleccionado, para los cuales se calcula el espectro. La ecuación (3) puede obtenerse de modo eficiente considerando:

$$nk = (n^2 + k^2 - (k-n)^2)/2$$

$$S_{i-l}(z_k) = \sum_{n=0}^{N_B-1} y_{i-l}[k] W^{-\frac{k^2}{2}} \quad (4)$$

$$\text{donde } y_{i-l}[k] = \sum_{n=0}^{N_B-1} g_{i-l}[n] \cdot h[k-n],$$

$$h[n] = W^{\frac{n^2}{2}} \quad y \quad g_{i-l}[n] = x_{i-l}[n] \cdot (z_0)^{-n} \cdot W^{-\frac{n^2}{2}}$$

De la ecuación (4) se interpretan las señales “ y_{i-l} ” como una convolución de las señales “ g_{i-l} ” con “ h ” y por tanto se puede procesar a partir de i) transformada de Fourier mediante algoritmo FFT, ii) producto directo de señales en el dominio transformado y iii) transformada inversa de Fourier mediante algoritmo IFFT. Posteriormente se utiliza el estimador de periodograma para promediar los L segmentos de espectro obtenidos.

$$\hat{S}_i(f) = \frac{1}{L} \sum_{l=0}^{L-1} S_{i-l}(f) \quad (5)$$

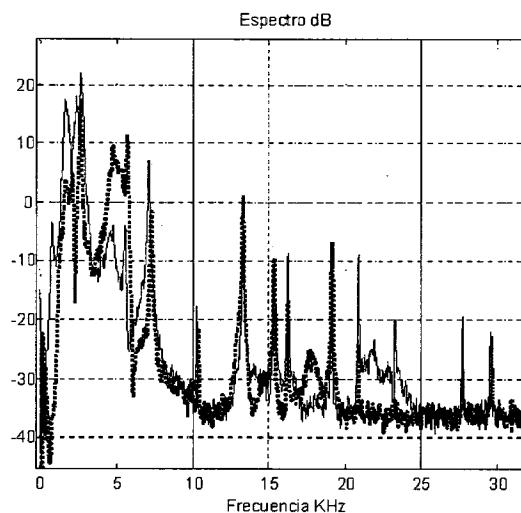


Figura 2. Espectro Chirp obtenido a partir de una pieza no defectuosa. Línea continua para la señal 1 y discontinua para la señal 2.

En la figura 2 se muestra el espectro obtenido a partir de una pieza no defectuosa. En este ejemplo se trabajó con $M=4096$ puntos espaciados uniformemente en un ancho de banda de 40 KHz y promediando $L=150$ segmentos de longitud $N_B=28673$ muestras pre-enventanadas mediante una ventana de Hanning. Las dos señales especreales corresponden a una pieza circular y por tanto con simetría axial. Aunque ambas señales, deberían presentar idénticas frecuencias de resonancia, la forma del espectro presenta diferencias debido a la presencia de ruido de medida y a la diferente respuesta impulsional de los dos sensores receptores utilizados en la grabación de las señales.

3.2 Algoritmo de Detección de picos especreales

Tal como se comenta en la sección 2, el proceso de detección de picos especreales se facilitó mediante la aplicación del algoritmo “Top Hat” [4] [5], algoritmo basado en principios de morfología matemática mediante el cual se obtiene información relativa a la geometría del espectro. El objetivo final de la aplicación de este algoritmo es conseguir que la señal espectral no presente picos ruidosos que dificulten la detección de los picos principales o frecuencias de resonancia. El algoritmo top hat, se basa en dos operadores básicos en morfología: dilatación y erosión. Con la dilatación y la erosión, se construyen operadores más complejos, como son la apertura y el cierre, que también son utilizados en el algoritmo.

Tanto el operador de dilatación como el de erosión, se basan en el deslizamiento de un elemento estructurante a través de la señal a procesar. En nuestro caso, el elemento estructurante es una ventana temporal que se centra en el punto que se quiere procesar y que engloba los puntos de la señal que intervienen en el procesado. En el caso de la erosión, se realiza un estudio local para cada punto a procesar y se escoge el mínimo que se encuentra dentro de la ventana que conforma el elemento estructurante. En

el caso de la dilatación, el efecto es el contrario, pues se escoge el punto de mayor amplitud.

Los resultados obtenidos mediante estas transformaciones morfológicas, dependen directamente del tamaño seleccionado para el elemento estructurante. A priori, es difícil conocer cual es el tamaño que implica mejores resultados. En nuestro caso, la anchura de los lóbulos principales de las frecuencias de resonancia se encuentra directamente relacionada con el tipo de enventanamiento utilizado en el proceso de estimación espectral y el número de muestras utilizado para este. Es por esto que la anchura del elemento estructurante es un factor que debe ser modificado cuando cambian los parámetros de configuración del sistema, en cuanto a análisis espectral se refiere.

El operador de cierre, se define como la erosión de una señal que previamente ha sido dilatada. El operador de apertura en cambio, realiza el efecto contrario, pues realiza primero una dilatación de la señal, y posteriormente realiza una erosión. La aplicación del algoritmo “Top Hat” es el resultado de sustraer a la señal original la señal procesada mediante la apertura del cierre de la señal original y forzar a cero todos los puntos para los cuales se obtiene un resultado negativo.

$$x - \wedge \{ \gamma(\varphi(x)), x \} \quad (6)$$

En (6), $\wedge \{ \}$ se refiere al operador mínimo. $\gamma()$ y $\varphi()$ son respectivamente los operadores de apertura y de cierre.

Si se procesa una señal con el operador de cierre, se consigue una señal envolvente que elimina los picos inferiores manteniendo los picos principales y los picos superiores debidos al ruido. En la figura 3 se observa juntamente con la señal original, un fragmento de señal procesada con el operador de cierre.

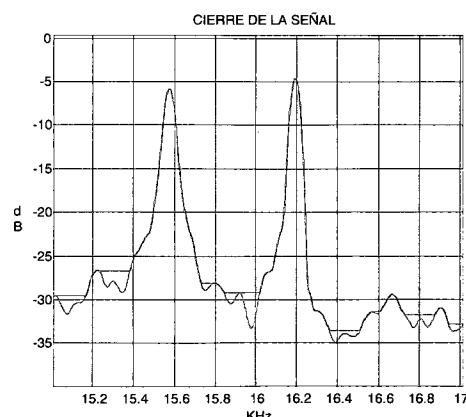


Figura 3. Fragmento de espectro procesado mediante el operador de cierre, junto con la señal original. Línea continua para la señal procesada y línea discontinua para la señal original.

Si la señal obtenida mediante la aplicación del cierre es procesada mediante el operador de apertura, se consigue una señal como la que se presenta en la figura 4. Dicha señal, recorta los picos principales, que son los que

queremos detectar y mantiene la propiedad de señal envolvente eliminando los picos inferiores.

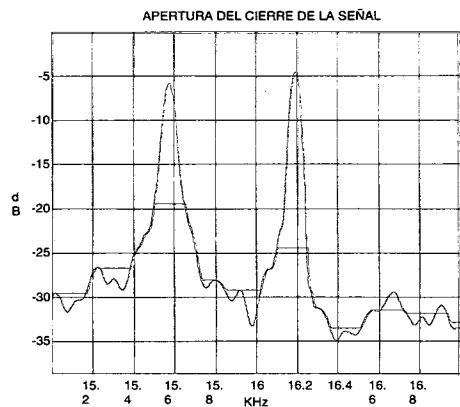


Figura 4. Fragmento de espectro procesado mediante la apertura del cierre, junto con la señal original. En línea continua se presenta la señal procesada y en línea discontinua la señal original.

Para recuperar una señal que solo presente los picos significativos, se tiene que substraer la señal procesada mediante la aplicación del cierre y apertura a la señal original y poner a cero los puntos cuyo valor sea negativo. Como puede suceder que no se eliminan todos los picos de ruido, en vez de inicializar a cero los puntos negativos, se pueden poner a cero los puntos que no superen un umbral determinado.

En la figura 5 se muestra el resultado de procesar mediante el algoritmo “top hat” una señal de espectro completo, como la mostrada en la figura 2.

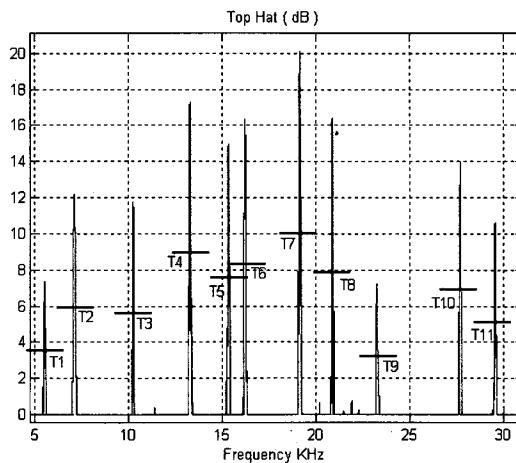


Figura 5. Espectros procesados mediante el algoritmo Top Hat. En este ejemplo se seleccionaron 11 picos con sus respectivos umbrales.

A partir de los cruces por cero de la derivada de la señal “Top Hat” se detectan las frecuencias de resonancia. Mediante un umbral en amplitud se limita inferiormente la potencia que debe presentar un máximo espectral para ser considerado frecuencia de resonancia de la pieza. Previa-

mente a la etapa de clasificación propiamente dicha, se desarrolla una etapa de entrenamiento, en la cual se aprende o graba el comportamiento esperado para las piezas buenas en cuanto a: i) Número de frecuencias de resonancia, ii) Vector de frecuencias medias, iii) Varianzas del vector de frecuencias. La clasificación de cada nueva pieza se realiza en base a los parámetros anteriores.

3.3 Algoritmo de Clasificación

Desde una perspectiva de clasificación [3], el algoritmo de detección debería distinguir o clasificar entre dos modelos: Piezas defectuosas y piezas no defectuosas. En la presente aplicación se ha aplicado el criterio de distancia euclídea mínima, entre cada nuevo vector de frecuencias a clasificar y el valor esperado para el mismo en caso de que se trate de una pieza no defectuosa. Debido a la dificultad para obtener un modelo simplificado para las piezas defectuosas, se ha modelado únicamente el comportamiento estadístico de las piezas no defectuosas. Así, de cada pieza medida se graban dos vectores que constituyen su firma espectral o señal identificativa. En (7) N_f representa el número de frecuencias de resonancia. Es decir, el número de componentes de los vectores $\mathbf{f}_1, \mathbf{f}_2$:

$$\mathbf{f}_i = (f_{i-1} : f_{i-N_f})^T; \quad i = 1, 2 \quad (7)$$

Durante el periodo de entrenamiento se miden los dos vectores para el máximo número de piezas no defectuosas posible y se graba el siguiente conjunto de parámetros:

- Vectores media de los canales 1 y 2: $\hat{\mathbf{f}}_1; \hat{\mathbf{f}}_2$
- Matrices de covarianza de los canales 1 y 2: $\mathbf{C}_1; \mathbf{C}_2$

Tras el periodo de entrenamiento se desarrolla la estrategia de clasificación aplicando el siguiente criterio. Para la pieza número “n” :

- Se miden los vectores de frecuencia $\mathbf{f}_1[n]; \mathbf{f}_2[n]$
- Si una componente frecuencial no se halla en el margen esperado: $(f_{i-k} - K_T \sigma_{i-k}, f_{i-k} + K_T \sigma_{i-k})$ la pieza se clasifica como defectuosa. En la expresión previa i denota el canal ($i=1,2$), k es el número de componente ($k=1..N_f$) y K_T es un umbral que se selecciona empíricamente a partir de la etapa de entrenamiento.
- Cada vector de frecuencias se normaliza a partir de la media y la varianza esperadas.
- Se calculan las dos distancias euclídeas: $d_1 = d(\mathbf{f}_1, \mathbf{f}_1[n])$, $d_2 = d(\mathbf{f}_2, \mathbf{f}_2[n])$ y para piezas

con simetría axial también la distancia cruzada $d_{1-2} = d(\mathbf{f}_1[n], \mathbf{f}_2[n])$. Debido a la normalización previa, el valor esperado para estos parámetros es:

$E[d_1^2] = 1$, $E[d_2^2] = 1$, $E[d_{1-2}^2] = 2(1 - \rho)$. r representa la correlación cruzada entre los dos canales. Para un conjunto de N_p piezas el valor esperado es de:

$$\frac{1}{N_f} \frac{1}{N_p-1} \sum_{n=1}^{N_p} \sum_{i=1}^{N_f} f_{1-i}[n] f_{2-i}[n]$$

- Tras el cálculo de las tres distancias normalizadas, se testea si cada una de las tres métricas medidas se halla en

el correspondiente intervalo: $(0, K_T \sqrt{E[d_x^2]})$, donde x denota distancia ($x=1,2$ o $1-2$), y K_T representa un umbral seleccionado empíricamente durante la etapa de entrenamiento.

- Finalmente, si las medidas de la pieza aparecen en todos los márgenes esperados, tanto a nivel de vector de frecuencias como a nivel de distancias medidas, la pieza se clasifica como no defectuosa y en caso contrario como defectuosa.

Hay que destacar, que el conjunto de parámetros estadísticos (medias y varianzas de los vectores de frecuencias y de las distancias euclídeas) medidos en etapa de entrenamiento, se adaptan continuamente durante la etapa de clasificación, con cada nueva pieza clasificada como no defectuosa. De este modo, el sistema se va adaptando a pequeños cambios dentro de la serie de piezas fabricadas.

En la figura 6 se muestra el histograma obtenido para una de las frecuencias (15.58KHz) medido a partir de una serie de 150 piezas no defectuosas y en la figura 7

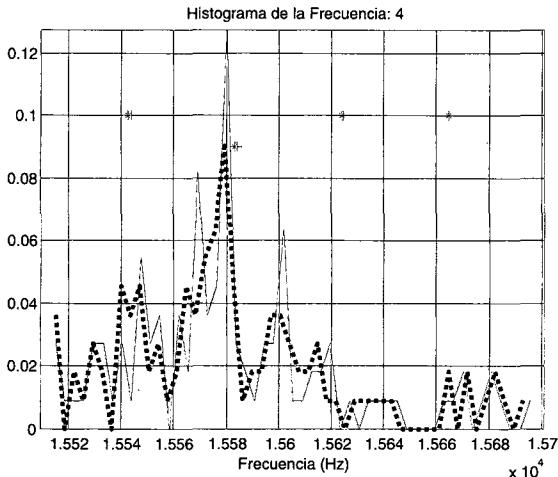


Figura 6. Histograma obtenido para la frecuencia de resonancia: 15.58KHz Línea continua para la señal 1 y discontinua para la señal 2. Los asteriscos muestran los saltos en varianza: $f_{i-k} \pm n\sigma_{i-k}$, $n = -2 : +2$.

se muestra el histograma obtenido para las 3 distancias euclídeas normalizadas.

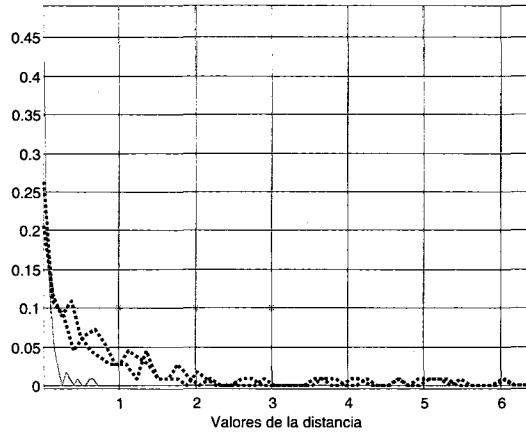


Figura 7. Histogramas obtenidos para las tres distancias euclídeas normalizadas. Línea discontinua para d_1 y d_2 y continua para d_{1-2} . Los asteriscos muestran los múltiplos de

$$los valores esperados: $nE[\hat{d}_x^2]$; $n = 1 : 3$$$

4. PROBABILIDAD DE FALSA ALARMA

Con el objeto de estimar la probabilidad de falsa alarma como medida de calidad del sistema, se han asumido las siguientes consideraciones:

- Los dos vectores de frecuencias presentan una distribución gaussiana.
- La probabilidad de falsa alarma se calcula en el dominio del subespacio de señal.

Los vectores de frecuencia se proyectan sobre el subespacio de las N_R componentes o autovectores principales obtenidos a partir de la descomposición de la matriz de covarianza medida:

$$C = E \left[\begin{pmatrix} \mathbf{f}_1 - \hat{\mathbf{f}}_1 \\ \mathbf{f}_2 - \hat{\mathbf{f}}_2 \end{pmatrix} \begin{pmatrix} \mathbf{f}_1^T - \hat{\mathbf{f}}_1^T & \mathbf{f}_2^T - \hat{\mathbf{f}}_2^T \end{pmatrix} \right] \quad (8)$$

El nuevo vector proyectado es de N_R componentes y se distribuye como:

$$\mathbf{x} : \mathcal{N}(0, \text{diag}(\sigma_1^2, \dots, \sigma_{N_R}^2)) \quad (9)$$

El umbral previo K_T se transforma en K_N . En estas condiciones la probabilidad de falsa alarma se calcula como:

$$P_{FA} = \sum_{n=1}^{N_R} 2 \int_{K_N \sigma_n}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{x^2}{2\sigma_n^2}\right) dx = N_R \text{erfc}(K_N) \quad (10)$$

K_N	1	2	3	4
$\text{erfc}(K_N)$	0,16	4,6 e(-3)	2,2 e(-5)	1,5 e(-8)

Tabla 1. Valores estimados para la probabilidad de falsa alarma.



5. CONCLUSIONES Y RESULTADOS

A partir de piezas defectuosas, se observa como las frecuencias de resonancia sufren cambios (figura 8), o incluso para algunas piezas desaparecen algunas de las frecuencias de resonancia (figura 9). Cuando ocurre cualquiera de estos dos fenómenos, automáticamente el sistema clasifica la pieza como defectuosa.

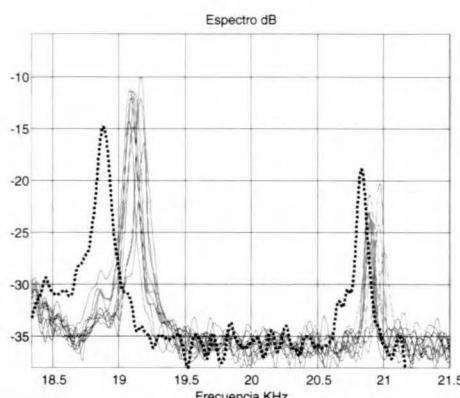


Figura 8. Espectro parcial en dB y frecuencia en KHz. Pieza defectuosa en línea discontinua

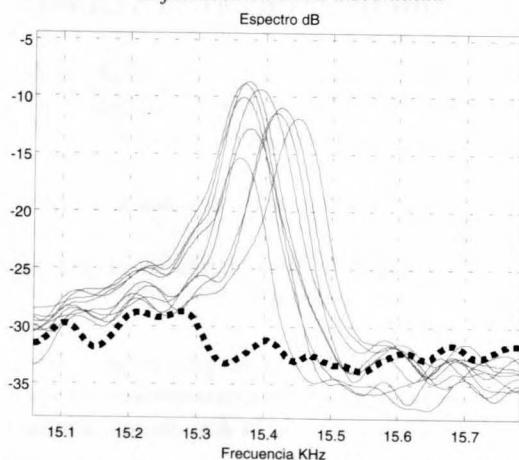


Figura 9. Espectro parcial en dB y frecuencia en KHz. Pieza defectuosa en línea discontinua.

Como conclusión, se ha desarrollado un sistema de detección de piezas defectuosas, para su aplicación como sistema de control de calidad dentro de una cadena de fabricación en serie. El sistema es lo suficientemente flexible como para trabajar con anchos de banda de hasta 200 KHz y para testear diferentes tipos de piezas metálicas sinterizadas.



Figura 10. Sistema de excitación y grabación de señales de resonancia por ultrasonidos. Un sensor transmisor y dos sensores receptores se colocan simétricamente alrededor de la pieza metálica sinterizada circular y con simetría axial.

6. REFERENCIAS

- [1] S. M. Kay. "Modern Spectral Estimation - Theory and Application". Prentice Hall, 1988.
- [2] J. G. Proakis, D. G. Manolakis. "Introduction to Digital Signal processing". Macmillan, 1988.
- [3] R. O. Duda, P. E. Hart, D. G. Stork. "Pattern Classification", Ed. Wiley Interscience, 2001.
- [4] F. Meyer, A. Van Driel "The application of mathematical morphology in cervical cancer screening" Elsevier Publishers 1985.
- [5] Jean Serra "Image analysis and mathematical morphology" vol1. Ed. Academic Press, 1982.
- [6] L. L. Sharf, "Statistical Signal Processing". Addison Wesley, 1991.
- [7] M. Cabrera, X. Castell, R. Montoliu. "Crack Detection System Based On Spectral Analysis of Ultrasonic Resonance Signals". IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Hon Kong 2003.

AUTORES



Xavier Castell Castells nació en Vic el 29 de Marzo de 1978. Obtuvo la titulación de Ingeniero Superior de Telecomunicaciones en el año 2002 en la Escuela Técnica Superior de Ingeniería de Telecomunicaciones de Barcelona (Universidad Politécnica de Cataluña). Su proyecto final de carrera, consistió en el diseño e implementación de un sistema de control de calidad basado en la técnica de inspección por resonancia ultrasónica. Actualmente trabaja como investigador en Formación en el IRI (Instituto de Robótica e Informática Industrial UPC-CSIC) donde colabora con el grupo de visión por computador.

Margarita Cabrera es Profesora Titular de Universidad del Departamento de "Teoria del Senyal i Comunicacions" de la "Universitat Politècnica de Catalunya". Desarrolla sus actividades de docencia en la "Escola Tècnica d'Enginyers de Telecomunicació de Barcelona". Participa en proyectos de investigación relacionados con UMTS, modelado de canales de propagación, procesado de antenas inteligentes, y en general aplicaciones de procesado de señal. Es miembro del IEEE.

Rafael Montoliu es Ingeniero de Telecomunicaciones y actualmente Jefe de Planificación en ALME (Aleaciones de Metales Sinterizados), AMES. Sant Vicenç dels Horts (Barcelona).



SEPARACIÓN CIEGA DE FUENTES

Irene Aldecoa Bilbao¹, Noemí Carné Serrano², Enric Monte Moreno³

^{1,2} Estudiantes de la E.T.S. Enginyeria de Telecomunicació de Barcelona (UPC)

³ Profesor Titular de la E.T.S. Enginyeria de Telecomunicació de Barcelona (UPC)

Departamento de Teoría de Señal y Comunicaciones, Grupo de Procesado de Voz

{alirene, alnoemi, enric}@gps.tsc.upc.es

Resumen- La separación ciega de fuentes es un problema que consiste en obtener señales procedentes de n fuentes a partir de las mezclas que llegan a m sensores, disponiendo sólo de estas últimas. En éste artículo presentamos algoritmos de separación de señales de voz. En primer lugar resumiremos los fundamentos teóricos de los métodos de separación ciega de fuentes, y luego procederemos a la deducción y simulación de dos algoritmos de gradiente basados en el método denominado Independent Component Analysis (ICA), que realizan la separación en los dominios temporal y frecuencial.

1. INTRODUCCIÓN

El problema de la separación de fuentes aparece en diversos campos, principalmente en procesado de señal de audio. También se aplica en arrays de antenas, sensores químicos y en geología. En todos los casos se deben separar varias fuentes linealmente superpuestas, o mezcladas, y captadas por varios sensores. Dado que los datos tienen una estructura lineal, la dificultad reside en estimar la **matriz de mezcla**, es decir el conjunto de coeficientes en la superposición lineal, desconocida y que refleja la geometría del problema. Éste consiste en separar señales de las que a lo sumo tenemos alguna información sobre su estadística, y que han sido mezcladas en una configuración geométrica que desconocemos. De ahí el nombre de **separación ciega de fuentes**.

En audio, se presenta el problema denominado *cocktail party* [1, 2] que se da cuando varios locutores hablan simultáneamente. En ese contexto el problema reside en centrar la atención en un único locutor de entre un número de conversaciones, y el ruido de fondo, y extraer sólo una voz. Este proceso puede modelarse como una mezcla lineal y el posterior filtrado de fuentes de sonido. El método de **Análisis de Componentes Independientes** (*Independent Component Analysis*, ICA) permite recuperar de forma ciega las fuentes desconocidas asumiendo que las señales originales son independientes.

2. SEPARACIÓN CIEGA DE MEZCLAS INSTANTÁNEAS CON ICA

ICA es una técnica estadística que permite hallar variables ocultas en un conjunto de medidas o señales [3, 4, 5, 6]. Dichas variables son componentes de datos estadísticos multivariados. La característica que diferencia ICA de otros métodos de separación de fuentes es que la condición para extraer componentes ocultos es que sean al mismo tiempo *estadísticamente independientes y no gaussianos*.

2.1 Definición del modelo lineal, instantáneo y sin ruido

Sean x_1, \dots, x_n n variables aleatorias recogidas por n sensores, que proceden de la combinación lineal de n variables aleatorias s_1, \dots, s_n , denominadas fuentes y mutuamente independientes por definición. Los **observables** x_i vienen dados por:

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n, \quad (1)$$

$$\forall i = 1, \dots, n$$

donde $\{a_{ij}\}$ con $i, j = 1, \dots, n$ son coeficientes reales que modelan la atenuación relativa entre las fuentes y se denominan **coeficientes de mezcla**.

El modelo básico de ICA es un modelo generativo, en la medida en que los datos observables son generados mediante un proceso de mezcla lineal de las fuentes, siendo imposible observar estas últimas directamente. Los coeficientes de mezcla a_{ij} son desconocidos, y se deben estimar al mismo tiempo que los **componentes independientes** s_i a partir de los observables x_i . Este modelo básico es estático en cuanto a que las variables que intervienen en él son variables aleatorias. Omite también cualquier retardo temporal entre las fuentes que pueda darse durante la mezcla, por lo que se denomina a este modelo básico **modelo de mezcla instantánea**.



Es conveniente utilizar una notación matricial para describir la ecuación (1). Sean \underline{x} y \underline{s} los vectores columna cuyos elementos son las mezclas x_1, \dots, x_n , y los componentes independientes s_1, \dots, s_n respectivamente. El modelo de mezcla instantánea se escribe como:

$$\underline{x} = \underline{A}\underline{s} \quad (2)$$

donde \underline{A} es la denominada matriz de mezcla formada por los coeficientes a_{ij} .

Este modelo básico no contempla la posibilidad de que el número de observables y de componentes independientes sea diferente, pero si así fuese sería importante que el número de sensores fuese superior o igual al número de fuentes para poder aplicar este método.

El objetivo del método es estimar la matriz de mezcla \underline{A} para obtener la **matriz de separación** \underline{W} al invertirla. Así pues se pueden recuperar los componentes independientes mediante:

$$\underline{y} = \underline{W}\underline{x} \quad (3)$$

donde \underline{y} contiene los componentes independientes estimados o recuperados.

Existe un gran número de criterios para estimar la matriz de separación. En cualquier caso hay varios supuestos así como dos ambigüedades sobre los componentes recuperados.

2.2 Restricciones y ambigüedades del modelo

A continuación se expresan las condiciones que debe cumplir un problema de separación ciega y las ambigüedades que se tienen al recuperar las señales.

2.2.1 Restricciones

Las suposiciones que deben hacerse para que el modelo funcione de forma correcta son las siguientes:

- a.** Los componentes se asumen estadísticamente independientes.
- b.** Los componentes, con la posible excepción de un componente, deben tener distribuciones no gaussianas.

c. Se asume el mismo número de componentes independientes que de observables, y que la matriz de mezcla es cuadrada.

2.2.2 Ambigüedades

En la estimación de los componentes independientes aparecen dos ambigüedades:

- a.** No queda determinada la varianza (energía) de los componentes independientes.

Los componentes estimados \underline{y} son proporcionales a los originales, existe un factor de escala α_i que no se puede determinar:

$$\underline{x} = \sum_i \left(\frac{1}{\alpha_i} a_i \right) (s_i \alpha_i) \quad (5)$$

Por ello se suele asumir que los componentes a estimar tienen varianza unidad: $E\{s_i^2\} = 1$.

- b.** No queda determinado el orden de los componentes independientes.

Esto es debido a que se modela la mezcla mediante una matriz, y ésta no refleja el orden espacial de las fuentes originales. De forma analítica, dada una matriz de permutación \underline{P} y su inversa pueden dar

$\underline{x} = \underline{A} \underline{P}^{-1} \underline{P} \underline{s}$, donde los elementos de \underline{P} son las variables independientes originales s_j , pero en otro orden.

2.3 Preprocesado de los observables: centrado y blanqueo

Antes de realizar la separación, es conveniente realizar dos operaciones sobre los observables, que no alteran la forma de las señales, y sin embargo simplifican en gran medida el algoritmo. Este preprocesado consiste en el centrado y posterior blanqueo (incorelación) de los datos.

2.3.1 Centrado

Con el objetivo de simplificar los algoritmos, se asume que los observables tienen media cero. En caso de no ser así, es necesario realizar un preprocesado sobre ellos para que esta suposición sea cierta. Ello es posible mediante el centrado de los observables, proceso que consiste en restarles su media $E\{\underline{x}\}$, tal como sigue:

$$\underline{x} = \underline{x}' - E\{\underline{x}'\} \quad (6)$$

Los componentes independientes estimados también tendrán media cero:

$$E\{\underline{y}\} = \underline{W}E\{\underline{x}\} \quad (7)$$

La matriz de mezcla no cambia después de este preprocessado, por lo que siempre se podrá realizar sin que afecte a la estimación de la misma. Finalmente la estimación de los componentes independientes obtenidos a partir de los observables con media cero, deben reconstruirse sumándoles la media sustraída, $\underline{W}E\{\underline{x}'\}$.

2.3.2 Blanqueo

El blanqueo de un vector aleatorio de media cero, \underline{x} , supone imponer que sus componentes estén incorrelados y que sus varianzas sean uno. En este caso la matriz de covarianza será igual a la matriz identidad:

$$\underline{C}_x = E\{\underline{x}\underline{x}^T\} = \underline{I} \quad (8)$$

Por tanto el blanqueado consistirá en una transformación lineal sobre los observables, multiplicándolos por una matriz \underline{V} , tal que:

$$\underline{z} = \underline{V}\underline{x} \quad (9)$$

$$\underline{V} = \underline{E}\underline{D}^{-1/2}\underline{E}^T = \underline{C}_x^{-1/2} \quad (10)$$

donde el vector \underline{z} tendrá covariancia diagonal, \underline{E} es la matriz ortogonal de autovectores de la matriz de covarianza y \underline{D} es la matriz diagonal de sus autovalores.

2.4 Exclusión de las variables gaussianas como fuentes

Para demostrar porqué las variables gaussianas no pueden ser componentes independientes separables se analiza qué transformación sufre la densidad de probabilidad conjunta $p_s(s_1, s_2)$ de dos variables gaussianas s_1 y s_2 cuando se realiza la mezcla.

La densidad de probabilidad conjunta es:

$$p_s(s_1, s_2) = \frac{1}{2\pi} \exp\left(-\frac{s_1^2 + s_2^2}{2}\right) = \frac{1}{2\pi} \exp\left(-\frac{\|\underline{s}\|^2}{2}\right) \quad (11)$$

Se asume que la matriz de mezcla \underline{A} es ortogonal, dado que los datos han sido blanqueados, y por tanto $\underline{A}^{-1} = \underline{A}^T$, y $\underline{s} = f^{-1}(\underline{x}) = \underline{A}^{-1}\underline{x} = \underline{A}^T\underline{x}$. Dadas las propiedades en cuanto a las densidades de probabilidad de variables que son fruto de una transformación [7], la densidad de probabilidad conjunta de los observables x_1, x_2 viene dada por la ecuación:

$$p_{\underline{x}}(\underline{x}) = p_{\underline{x}}(x_1, x_2) = \frac{p_s(f^{-1}(\underline{x}))}{|\det J_f(f^{-1}(\underline{x}))|} \quad (12)$$

$$p_{\underline{x}}(\underline{x}) = p_{\underline{x}}(x_1, x_2) = \frac{p_s(\underline{A}^T\underline{x})}{|\det \underline{A}^T|} \quad (13)$$

$$p_{\underline{x}}(\underline{x}) = \frac{\frac{1}{2\pi} \exp\left(-\frac{\|\underline{A}^T\underline{x}\|^2}{2}\right)}{|\det \underline{A}^T|} \quad (14)$$

Sin pérdida de generalidad se supone que \underline{A} es ortonormal, se cumple $\|\underline{A}^T\underline{x}\| = \|\underline{x}\|$, y puesto que $|\det \underline{A}| = 1$, la densidad de probabilidad conjunta de los observables queda de la siguiente forma:

$$p_{\underline{x}}(\underline{x}) = \frac{1}{2\pi} \exp\left(-\frac{\|\underline{x}\|^2}{2}\right) \quad (15)$$

Se ve claramente que la transformación ortogonal no cambia la densidad de probabilidad de los datos, las distribuciones de las mezclas y de los componentes originales son idénticas. La matriz de mezcla \underline{A} no es identifiable si las fuentes son gaussianas, porque en el caso de variables gaussianas conjuntas la condición de incorrelación implica necesariamente independencia.

Sin embargo sí existe la posibilidad de incluir una sola variable gaussiana en el conjunto de las fuentes. Es el caso límite en el que se pueden separar todas las fuentes si ninguna otra fuente tiene una componente gaussiana con la que pudiese estar mezclada.



3. ALGORITMOS DE GRADIENTE NATURAL EN TIEMPO Y FRECUENCIA

A continuación presentamos varios algoritmos de separación de mezclas instantáneas basados en el *algoritmo de gradiente natural*.

En primer lugar se explica el algoritmo de máxima verosimilitud (*Maximum Likelihood ML*) [8] puesto que es una componente del algoritmo de gradiente natural, que permite calcular los coeficientes de $\underline{\underline{W}}$, es decir de la inversa de la matriz de mezcla. En segundo lugar se desarrollan dos algoritmos de gradiente, el primero de los cuales realiza la separación en el dominio temporal, mientras que el segundo lo hace en el frecuencial.

3.1 Estimación por máxima verosimilitud (ML)

Una aproximación para la estimación del modelo ICA es la estimación por máxima verosimilitud (ML). Una interpretación del estimador ML es que selecciona los valores de los parámetros que dan la probabilidad más alta para las observaciones.

La densidad $p_{\underline{x}}$ del vector de mezclas $\underline{x} = \underline{\underline{A}}\underline{s}$ se puede formular como:

$$p_{\underline{x}}(\underline{x}) = |\det \underline{\underline{W}}| p_s(\underline{s}) = |\det \underline{\underline{W}}| \prod_i p_i(s_i) \quad (16)$$

donde $\underline{\underline{W}} = \underline{\underline{A}}^{-1}$, y p_i son las densidades de los componentes independientes s_i [7]. Se puede expresar en función de $\underline{\underline{W}} = (\underline{w}_1, \dots, \underline{w}_n)^T$ y de \underline{x} , obteniendo:

$$p_{\underline{x}}(\underline{x}) = |\det \underline{\underline{W}}| \prod_i p_i(\underline{w}_i^T \underline{x}) \quad (17)$$

Se asume que $\underline{x}(1), \underline{x}(2), \dots, \underline{x}(T)$ son T observaciones de \underline{x} . La verosimilitud se puede obtener como el producto de esta densidad evaluada en los T puntos. Esto se denota por L y se considera una función de $\underline{\underline{W}}$:

$$L(\underline{\underline{W}}) = \prod_{t=1}^T \prod_{i=1}^n p_i(\underline{w}_i^T \underline{x}(t)) |\det \underline{\underline{W}}| \quad (18)$$

Con frecuencia es más práctico trabajar con el logaritmo de la verosimilitud porque su álgebra es más sencilla y el máximo del logaritmo se obtiene en el mismo punto que el máximo de la verosimilitud. El log-verosimilitud es:

$$\log L(\underline{\underline{W}}) = \sum_{t=1}^T \sum_{i=1}^n \log p_i(\underline{w}_i^T \underline{x}(t)) + T \log |\det \underline{\underline{W}}| \quad (19)$$

La base del logaritmo no afecta, por lo que en lo sucesivo se hará referencia al logaritmo natural. Para simplificar la notación, se denota el sumatorio con índice t por el operador esperanza, y se divide la log-verosimilitud por T obteniendo:

$$\frac{1}{T} \log L(\underline{\underline{W}}) = E \left\{ \sum_{i=1}^n \log p_i(\underline{w}_i^T \underline{x}) \right\} + \log |\det \underline{\underline{W}}| \quad (20)$$

donde la esperanza se calcula como un promedio de las muestras observadas.

Existe un nuevo parámetro a estimar en el modelo ICA, las densidades de los componentes independientes, ya que la log-verosimilitud es función de ellas. Se puede usar una parametrización de p_i extremadamente simple que consiste en tomar una de las dos densidades correspondientes a dos no linealidades, detalladas más adelante, que se aplican en el método del gradiente.

3.2 Del gradiente al gradiente natural

Los algoritmos más simples para maximizar la verosimilitud se basan en calcular el gradiente de la función de coste. En este apartado se detalla como se deriva el algoritmo del gradiente natural a partir del *algoritmo de Bell-Sejnowski* [9].

3.2.1 Algoritmo de Bell-Sejnowski

A partir de la expresión (20), donde se realiza un promediado entre todas las muestras, se puede deducir el gradiente estocástico de la función log-verosimilitud, como:

$$\frac{1}{T} \frac{\partial \log L(\underline{\underline{W}})}{\partial \underline{\underline{W}}} = (\underline{\underline{W}}^T)^{-1} + E \left\{ g(\underline{\underline{W}} \underline{x}) \underline{x}^T \right\} \quad (21)$$

donde $\underline{\underline{W}} \underline{x} = \underline{y}$ y $\underline{g}(\underline{y}) = \begin{pmatrix} g_1(y_1) \\ \vdots \\ g_i(y_i) \\ \vdots \\ g_n(y_n) \end{pmatrix}$ es un vector de funciones $g_i(.)$, denominadas funciones *score* de las

distribuciones p_i de los componentes independientes s_i , definidas como:

$$g_i = (\log p_i)' = \frac{p_i'}{p_i} \quad (22)$$

De (21) se obtiene el siguiente algoritmo para estimación ML:

$$\underline{\Delta W} \propto (\underline{\underline{W}}^T)^{-1} + E\{g(\underline{\underline{W}}\underline{x})\underline{x}^T\} \quad (23)$$

donde el símbolo \propto indica proporcionalidad y $\underline{x} = (x_1, x_2, \dots, x_n)^T$ es el vector de observables instantáneos. La versión estocástica del anterior

$$\underline{\underline{W}}(I + E\{g(\underline{\underline{W}}\underline{x})\underline{y}^T\}) = \underline{\underline{W}}(I + E\{g(\underline{y})\underline{y}^T\})$$

algoritmo es el denominado algoritmo de Bell-Sejnowski:

$$\underline{\underline{\Delta W}} \propto (\underline{\underline{W}}^T)^{-1} + g(\underline{\underline{W}}\underline{x})\underline{x}^T \quad (24)$$

Dado que este algoritmo converge lentamente, es necesario, además de blanquear los datos, emplear una versión mejorada del mismo, es decir más rápida, conocida como gradiente natural o relativo. El gradiente natural además de tener mejores prestaciones en cuanto a convergencia, evita tener que invertir la matriz $\underline{\underline{W}}$.

3.2.2 Algoritmo del gradiente natural

La eficacia del gradiente natural reside en multiplicar ambos lados de la ecuación (23) por la derecha con $\underline{\underline{W}}^T \underline{\underline{W}}$, con lo que se obtiene:

$$\underline{\underline{\Delta W}} \propto ((\underline{\underline{W}}^T)^{-1} + E\{g(\underline{\underline{W}}\underline{x})\underline{x}^T\}) \underline{\underline{W}}^T \underline{\underline{W}} = \quad (25)$$

El algoritmo converge cuando $E\{g(\underline{\underline{W}}\underline{x})\underline{y}^T\} = -I$, es decir, cuando todos los elementos y_i y $g(y_j)$ están decorrelados para $i \neq j$.

En general las no linealidades empleadas dependen de la función de densidad de probabilidad de los componentes independientes s_i , y son:

$$f(y)^+ = -2 \tanh(y) \quad (26)$$

$$f(y)^- = \tanh(y) - y \quad (27)$$

para densidades supragaussianas y subgaussianas respectivamente [3].

3.3 Algoritmos de gradiente natural para separación ciega de mezclas instantáneas

A continuación se expone la formulación teórica del algoritmo de gradiente natural tanto en el dominio temporal como en el frecuencial.

3.3.1 Algoritmo de gradiente natural en tiempo

Este primer algoritmo realiza la separación temporal de los observables blanqueados.

Sean $\underline{s}_i = (s_{i1}, s_{i2}, \dots, s_{im})^T$, $i = 1, \dots, n$ los vectores de n componentes independientes de m muestras. Se define la matriz de fuentes $\underline{\underline{S}}$ tal que:

$$\underline{\underline{S}} = (\underline{s}_1, \underline{s}_2, \dots, \underline{s}_n)^T \quad (28)$$

La matriz de observables $\underline{\underline{X}} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)^T$, donde $\underline{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})^T$, se obtiene al realizar la mezcla temporal multiplicando matricialmente:

$$\underline{\underline{X}} = \underline{\underline{AS}} \quad (29)$$

Como procesado previo a la separación, se centran y blanquean los observables, obteniendo la matriz de datos blanqueados $\underline{\underline{Z}} = \underline{\underline{VX}}$. Es conveniente trabajar con la matriz $\underline{\underline{Z}}_p$, obtenida al permutar las columnas de la matriz $\underline{\underline{Z}}$, para romper la estacionariedad de las señales.

Para un número dado de iteraciones, se calcula la matriz de separación $\underline{\underline{W}}$ inicializada a la matriz identidad antes de la primera iteración, $\underline{\underline{W}}^0 = I$.

En cada iteración k se estima la nueva matriz de separación con la siguiente regla de adaptación:

$$\underline{\underline{W}}^{k+1} = \underline{\underline{W}}^k + \mu \partial \underline{\underline{W}}^k \quad (30)$$

donde μ es el denominado paso de adaptación, y $\partial \underline{\underline{W}}^k$ es el incremento matricial de la presente iteración. Este incremento se calcula mediante la expresión:



3.3.2 Algoritmo de gradiente natural en frecuencia

$$\underline{\partial W}^k = \left(I + f(\underline{U}^k) \cdot \frac{(\underline{U}^k)^T}{\underline{\sigma}_{U^k}} \right) \underline{W}^k \quad (31)$$

donde la matriz de componentes estimados intermedios \underline{U}^k se obtiene como:

$$\underline{U}^k = \underline{W}^k \underline{Z}_p = \begin{pmatrix} \underline{u}_1^k \\ \vdots \\ \underline{u}_i^k \\ \vdots \\ \vdots \\ \underline{u}_n^k \end{pmatrix} \quad (32)$$

y $\underline{\sigma}_{U^k} = \begin{pmatrix} \sigma_1^k \\ \vdots \\ \sigma_i^k \\ \vdots \\ \vdots \\ \sigma_n^k \end{pmatrix}$ es el vector de desviaciones estándares

de cada fila \underline{u}_i^k , y $f(\underline{U}^k)$ es el resultado de aplicar la no linealidad $f(\cdot)$ a la matriz \underline{U}^k , en función de la naturaleza subgaussiana o supragaussiana de las señales. La normalización es necesaria para que los componentes estimados tengan varianza unidad. En caso de no tener información a priori de la naturaleza de las fuentes, es necesario evaluar en cada iteración la kurtosis de cada observable, y decidir qué no linealidad aplicar. Para las variables no gaussianas la kurtosis generalmente es diferente de cero, es negativa para las subgaussianas y positiva para las supergaussianas.

Dado que las señales de voz tienen un carácter supragaussiano (siguen una distribución próxima a la de Laplace), la no linealidad empleada puede ser $f(\underline{U}^k) = -2 \tanh(\underline{U}^k)$.

Transcurridas todas las iteraciones, se calcula finalmente la matriz de componentes recuperados,

$$\underline{Y} = (\underline{y}_1, \underline{y}_2, \dots, \underline{y}_n)^T$$

donde,

$$\underline{y}_i = (y_{i1}, y_{i2}, \dots, y_{im})^T$$

mediante:

$$\underline{Y} = \underline{W} \underline{Z} \quad (33)$$

donde \underline{W} es la matriz estimada en la última iteración.

Este segundo algoritmo realiza la separación frecuencial de los observables blanqueados.

Se calcula la matriz de fuentes en frecuencia $\underline{\underline{S}}^f$ aplicando la transformada de Fourier a cada uno de los vectores \underline{s}_i :

$$\underline{\underline{S}}^f = \begin{pmatrix} FFT(\underline{s}_1) \\ \vdots \\ FFT(\underline{s}_i) \\ \vdots \\ FFT(\underline{s}_n) \end{pmatrix} \quad (34)$$

La matriz de observables frecuenciales,

$$\underline{\underline{X}}^f = (\underline{x}_1^f, \underline{x}_2^f, \dots, \underline{x}_n^f)^T$$

donde,

$$\underline{x}_i^f = (x_{i1}^f, x_{i2}^f, \dots, x_{im}^f)^T$$

se obtiene al multiplicar matricialmente:

$$\underline{\underline{X}}^f = \underline{\underline{A}} \underline{\underline{S}}^f \quad (35)$$

En este segundo caso también se centran y blanquean los observables, obteniendo la matriz de datos blanqueados $\underline{\underline{Z}}^f = \underline{\underline{V}} \underline{\underline{X}}^f$.

Se calcula la matriz de separación $\underline{\underline{W}}$, donde ahora sus elementos son complejos, para un número dado de iteraciones.

En cada iteración k se estima la nueva matriz de separación con la ecuación (30), y el incremento $\underline{\partial W}^k$ mediante la ecuación (31).

Al finalizar todas las iteraciones, se calcula la matriz de componentes frecuenciales recuperados,

$$\underline{\underline{Y}}^f = (\underline{y}_1^f, \underline{y}_2^f, \dots, \underline{y}_n^f)^T$$

donde,

$$\underline{y}_i^f = (y_{i1}^f, y_{i2}^f, \dots, y_{im}^f)^T$$

mediante la multiplicación matricial:

$$\underline{\underline{Y}}^f = \underline{\underline{W}} \underline{\underline{Z}}^f \quad (36)$$

donde $\underline{\underline{W}}$ es la matriz estimada en la última iteración.

Por último, a partir de los componentes frecuenciales recuperados, se obtienen sus análogos temporales como:

$$\underline{\underline{Y}} = \begin{pmatrix} \text{Re}(FFT^{-1}(\underline{\underline{y}}_1^f)) \\ \vdots \\ \text{Re}(FFT^{-1}(\underline{\underline{y}}_i^f)) \\ \vdots \\ \text{Re}(FFT^{-1}(\underline{\underline{y}}_n^f)) \end{pmatrix} = \begin{pmatrix} \underline{\underline{y}}_1 \\ \vdots \\ \underline{\underline{y}}_i \\ \vdots \\ \underline{\underline{y}}_n \end{pmatrix} \quad (37)$$

donde $\underline{\underline{y}}_i = (y_{i1}, y_{i2}, \dots, y_{im})^T$.

4. SEPARACIÓN DE SEÑALES DE VOZ EN AMBOS DOMINIOS

Para comprobar la eficacia del algoritmo de gradiente natural, se simula una mezcla de varias voces en un coche, como si varios locutores hablaran simultáneamente en él, para poder después recuperarlas por canales separados. La separación se ejecuta tanto en tiempo como en frecuencia, mientras que la mezcla siempre se realiza en el dominio temporal.

Se realiza el análisis de la calidad de la separación de las señales de voz primero con el algoritmo en el dominio temporal y luego en el frecuencial.

Se toman como señales a mezclar seis ficheros extraídos de la base de datos SpeechDat Car Spanish propiedad de la UPC [10]. Los ficheros de referencia seleccionados para las mezclas, así como su contenido son:

- a. V10000D1: “siete de octubre del setenta y seis”,
- b. V10010A1: “finalizar la llamada”,
- c. V10021T1: “son las cuatro y cinco”,
- d. V10020O5: “avenida del puerto”,
- e. V10011C4: “0 9 5 0 9 6”, y
- f. V1003051: “lista de destinos”.

4.1. Separación en el dominio temporal

Para hacer una separación de señales temporales, se carga la matriz de mezcla que deseemos aplicar al sistema. Una vez comprobada si es válida, es decir si es invertible, se realiza la mezcla de las fuentes, y se obtienen de forma matricial las fuentes mezcladas, es decir los observables.

Se realizan dos medidas distintas de calidad de la separación, una vez obtenida la matriz de separación. La primera es el cálculo de la evolución de las relaciones señal a error, es decir la relación entre la potencia de cada señal original, con respecto a la potencia del error

entre dicha señal y su recuperada. Este parámetro se denomina *SER*, y se define como:

$$SER_i = \frac{\frac{1}{N} \sum_{j=1}^N (\underline{s}_i^j)^2}{\frac{1}{N} \sum_{j=1}^N (\underline{s}_i^j - \underline{y}_i^j)^2}, \quad i=1,2,3$$

$j=1, \dots, N$ (38)

donde N es el número de muestras de las señales original y recuperada. Se calcula esta medida en decibelios, para cada una de las señales de voz recuperadas. También se recoge la media de las mismas.

La segunda medida de calidad consiste en el cálculo del valor de la kurtosis de las señales de voz originales y recuperadas, así como el error entre ambas. La kurtosis de una variable aleatoria x , que por simplicidad asumiremos de media cero y varianza igual a uno, viene dada por:

$$kurt(x) = E\{x^4\} - 3(E\{x^2\})^2 \quad (39)$$

y dado que $E\{x^2\}=1$ se tiene que $kurt(x) = E\{x^4\} - 3$.

La kurtosis es una medida de parecido entre una distribución dada y la distribución gaussiana para la cual la kurtosis es igual a cero. Así, cuanto mayor es el valor absoluto de la kurtosis, menos se parece la distribución a una gaussiana.

Se muestran resultados para una mezcla de los tres primeros ficheros mencionados, un paso de adaptación de 0.01 y una matriz de mezcla simple

$$\underline{\underline{A}} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Se realiza una simulación de 1000 iteraciones. El algoritmo converge desde la primera iteración a los siguientes valores:

$$SER_1 = 14,2981 \text{ dB}$$

$$SER_2 = 17,5244 \text{ dB}$$

$$SER_3 = 11,2809 \text{ dB}$$

Es evidente que si las señales recuperadas fuesen idénticas a las originales, el valor del parámetro sería infinito. Al realizar la media aritmética del error cuadrático de cada muestra, se obtiene un valor



ponderado del mismo. Los valores obtenidos son por tanto relativamente buenos, puesto que en todo caso la señal recuperada tiene una potencia más de diez veces mayor a la potencia de la señal error. La evolución de este parámetro para cada una de las señales de voz, así como la media de las tres puede verse en la Figura 1.

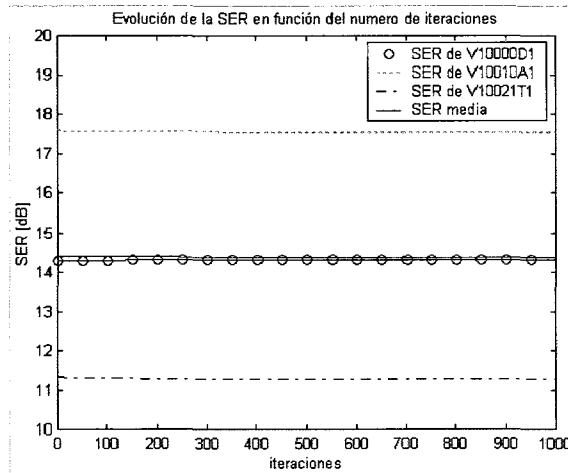


Figura 1: Evolución de las SER en la separación temporal de tres señales de voz.

Se aprecia claramente que la SER varía muy poco desde la primera iteración, reafirmando el hecho de que el algoritmo del gradiente natural converge de forma rápida.

La Tabla 1 recoge los valores de las kurtosis de las señales originales, recuperadas y error. Se observa que las kurtosis de las señales originales y recuperadas son casi idénticas, como lo demuestra el bajo error entre ambas.

Fichero	Kurtosis de la señal original	Kurtosis de la señal recuperada	Error entre las kurtosis
V10000D1	11,1493	11,1504	0,0011
V10010A1	21,5416	21,5601	0,0185
V10021T1	17,3213	17,3572	0,0359

Tabla 1: Comparación de las kurtosis de la separación temporal de tres señales.

4.2 Separación en el dominio frecuencial

Para ver cómo se comporta el algoritmo de gradiente natural al separar señales frecuenciales, se utiliza un procedimiento análogo al anterior. Cabe mencionar que antes de realizarse la mezcla se pasa al dominio frecuencial, para posteriormente hacer la separación, y finalmente volver al dominio temporal, en el que se realizan las medidas de calidad, que son las mismas que anteriormente.

Con el objetivo de ver la evolución del algoritmo en función del número de iteraciones, se realiza una simulación para 1000 iteraciones. La evolución de la SER en este caso se muestra en la Figura 2, donde puede observarse como las evoluciones de las SER son crecientes hasta que alcanzan su valor óptimo y empiezan a decrecer, hasta que sus valores se estabilizan. El número óptimo de iteraciones es aquel que maximiza las SER y minimiza el error entre las kurtosis. Este número es diferente para cada SER, por tanto se considerará el valor óptimo de la SER media para determinarlo. Siguiendo este criterio y tal como puede observarse en la Figura 2, el número óptimo de iteraciones es 85.

Los valores óptimos de las diferentes SER recogidos de la iteración 85 son:

$$SER_1 = 47,4071 \text{ dB}$$

$$SER_2 = 54,4948 \text{ dB}$$

$$SER_3 = 33,5186 \text{ dB}$$

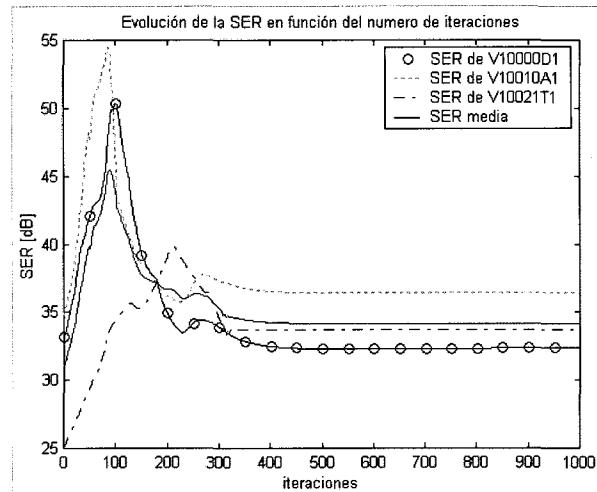


Figura 2: Evolución de las SER en la separación frecuencial de tres señales de voz.

Fichero	Kurtosis de la señal original	Kurtosis de la señal recuperada	Error entre las kurtosis
V10000D1	11,1493	11,1486	0,0008
V10010A1	21,5416	21,5435	0,0019
V10021T1	17,3213	17,3477	0,0265

Tabla 2: Comparación de las kurtosis de la separación frecuencial de tres señales

Los valores obtenidos en frecuencia, para los mismos parámetros de entrada, son ampliamente superiores en

este segundo caso. La segunda señal de voz sigue siendo la que se recupera con mayor calidad de las tres.

La Tabla 2 muestra como las kurtosis de la señal original y recuperada llegan a parecerse mucho en la iteración 85, siendo el error entre ellas muy próximo a cero.

En segundo lugar se mezclan los seis ficheros de referencia con la matriz:

$$A = \begin{bmatrix} 1 & 2 & 0,02 & 0,1 & 1 & 0,01 \\ 1 & 1 & 0,3 & 2 & 1 & 0,5 \\ 1 & 0,2 & 1 & 0,135 & 1 & 1 \\ 0,15 & 0,01 & 2 & 1 & 0,02 & 1 \\ 0,2 & 1 & 0,1 & 1 & 0,02 & 1 \\ 0,03 & 2 & 0,2 & 1 & 1,5 & 0,1 \end{bmatrix}$$

y se emplea el mismo paso de adaptación. Cabe mencionar que la elección del paso de adaptación influye en la velocidad de convergencia, y que implica emplear un número mayor o menor de iteraciones para llegar al punto óptimo. Tras 1300 iteraciones las SER de cada señal ha sufrido una

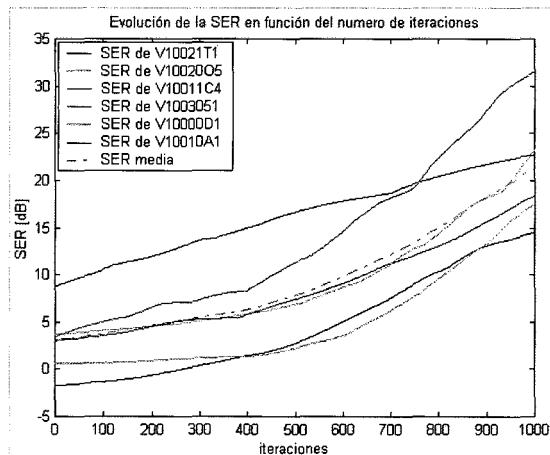


Figura 3: Evolución de las SER en la separación frecuencial de seis señales de voz.

importante evolución hasta estabilizarse como puede verse en la Figura 3.

Los valores óptimos de las diferentes SER recogidos de la iteración 1300 son:

$$SER_1 = 21,9222 \text{ dB}$$

$$SER_2 = 23,5876 \text{ dB}$$

$$SER_3 = 15,2672 \text{ dB}$$

$$SER_4 = 29,9398 \text{ dB}$$

$$SER_5 = 24,2032 \text{ dB}$$

$$SER_6 = 32,7914 \text{ dB}$$

Estos valores son inferiores a los obtenidos en el caso anterior. Con la incorporación a la mezcla de otras tres señales, la calidad de las tres primeras empeora levemente.

La Tabla 3 recoge las medidas de kurtosis para este último caso.

Fichero	Kurtosis de la señal original	Kurtosis de la señal recuperada	Error entre las kurtosis
V10000D1	11,1493	11,0561	0,0932
V10010A1	21,5416	21,5172	0,0245
V10021T1	17,3213	17,1251	0,1962
V10020O5	9,7846	9,5673	0,2173
V10011C4	8,9862	8,9831	0,0031
V1003051	21,2508	21,1885	0,0623

Tabla 3: Comparación de las kurtosis de la separación frecuencial de seis señales.

Los valores de error entre las kurtosis han aumentado, pero siguen siendo muy bajos.

Una mayor complejidad de la mezcla, así como un mayor número de fuentes afectan moderadamente a la calidad de la separación. Sin embargo auditivamente las señales recuperadas se corresponden sin confusión alguna con las originales.

5. CONCLUSIONES

Una vez expuestos los fundamentos del método de Análisis de Componentes Independientes se han presentado las bases del algoritmo de gradiente natural. Se han desarrollado dos algoritmos de gradiente para la separación ciega de fuentes de voz, uno en tiempo y el otro en frecuencia.

Los índices de calidad obtenidos mediante la separación en el dominio frecuencial son superiores a los obtenidos en el temporal. En este caso las señales tienen siempre una potencia de error menor, confirmando así una mayor calidad de la separación.

Además en este caso sí que existe una importante mejora de la relación señal ruido segmental al aumentar el número de iteraciones. Por tanto si la separación se realiza en el dominio frecuencial, el número de iteraciones se convierte en un factor fundamental en la eficacia del algoritmo.



Como conclusión a este estudio, se puede destacar que las dos medidas de calidad de la separación (cálculo de la relación señal a error y comparación de las kurtosis de las señales originales y recuperadas) realizada mediante el algoritmo del gradiente natural, revelan la eficacia del mismo. Cabe mencionar que dicho algoritmo funciona correctamente, es decir que separa auditivamente las voces mezcladas, desde la primera iteración. Con el incremento del número de iteraciones en el caso frecuencial, se dan un aumento de la relación señal a error y la disminución de la diferencia entre las kurtosis de las señales.

Estos hechos confirman la idoneidad de realizar la separación ciega de mezclas instantáneas con algoritmos de gradiente que trabajen en el dominio frecuencial.

BIBLIOGRAFÍA

- [1] A. Prieto, B. Prieto, C.G. Puntonet, A. Cañas, P. Martín-Smith. «*Geometric separation of linear mixtures of sources: Application to speech signals*». International workshop on Independent Component Analysis and Blind Signal Separation (ICA'99), pp. 295-300, Aussois, France, Enero 11-15, 1999.
- [2] A. Westner, V.M. Bove. «*Blind separation of real world audio signals using overdetermined mixtures*». Proc of ICA'99, pp. 251-256, Enero 11-15, Aussois, Francia, 1999.
- [3] Aapo Hyvärinen, Juha Karhunen, Erkki Oja. *Independent Component Analysis*. Wiley Interscience 2001.
- [4] Te-Won Lee. *Independent Component Analysis, theory and applications*. Kluwer Academic Publishers.
- [5] Aapo Hyvärinen. «*Survey on Independent Component Analysis*». Neural Computing Surveys 2, pp. 94-128, 1999.
- [6] Te-Won Lee, Mark Girolami, Anthony J. Bell, Terrence J. Sejnowski. «*A Unifying Information-theoretic Framework for Independent Component Analysis*». International Journal on Mathematical and Computer Modeling.
- [7] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3rd edition, 1991.
- [8] B.A. Pearlmutter, L. C. Parra, «*Maximum Likelihood Blind Source Separation: A Context-Sensitive generalization of ICA*». Advances in Neural Information Processing Systems. 1997.
- [9] A.J. Bell, T.J. Sejnowski. «*An information-maximization approach to blind separation and blind deconvolution*». Neural Computation 7, pp. 1129-1159, 1995.
- [10] A. Moreno. «*Documentación de la base de datos SpeechDat Car Spanish*». V2. Universitat Politècnica de Catalunya, 12 Septiembre 1999.

AUTORES



Irene Aldecoa Bilbao.

Nació en Barcelona el 3 de septiembre de 1979. Estudió Ingeniería Superior de Telecomunicaciones en Barcelona. En la actualidad está realizando el proyecto fin de carrera en el departamento de Teoría de Señal y Comunicaciones de la ETSETB.



Noemí Carné Serrano .

Nació en Barcelona el 12 de junio de 1979. Estudió Ingeniería Superior de Telecomunicaciones en Barcelona. En la actualidad está realizando el proyecto fin de carrera en el departamento de Teoría de Señal y Comunicaciones de la ETSETB.



Enric Monte Moreno.

Se graduó y se doctoró en Ingeniería de Telecomunicación por la Universidad Politécnica de Cataluña (UPC) en 1987 y 1992 respectivamente. Vinculado al departamento de Teoría de Señal y Comunicaciones desde el año 1989, actualmente ejerce el cargo de profesor titular. Su interés actual se centra en el tratamiento de señal de voz.



INTRODUCCIÓN AL WATERMARKING

Elena Martínez Villacampa, Elisa Sayrol Clos

Departamento de Teoría de la Señal y Comunicaciones
E. T. S. Ingeniería de Telecomunicación de Barcelona

e-mail {elmarvi21@yahoo.com , elisa@gps.tsc.upc.es}

ABSTRACT

En los últimos años han aumentado los estudios sobre las técnicas denominadas de *watermarking* de contenidos multimedia. Con este artículo, se pretende dar una visión global del concepto de *watermarking*, explicar las técnicas empleadas para su uso y las principales utilidades que posee. Así mismo, se describe una estructura general del proceso de inserción del *watermark*, o marca de agua, en los datos originales y de la extracción o detección de la marca a partir de una imagen marcada. Pretendemos dar una visión completa y resumida de conceptos más utilizados y destacar algunas de las aplicaciones en las que se emplea como son la protección de copyright.

1. INTRODUCCIÓN

En la actualidad el uso cada vez mayor de Internet como vehículo para la distribución de datos multimedia y aplicaciones, así como los servicios on-line y las aplicaciones de comercio electrónico han provocado un aumento en el empleo de los datos digitales. Y aunque estos datos tienen numerosas ventajas frente a los analógicos, estos últimos son más seguros frente a reproducciones no autorizadas ya que sucesivas copias degradan notablemente el contenido original.

Ejemplos de información digital que no tiene protección ante posibles fraudes incluyen imágenes fijas, secuencias de vídeo, señales de audio, documentos de texto y gráficos.

Las técnicas de watermarking insertan una señal adicional, conocida como watermark o marca de agua, directamente en los datos originales. Esta marca no es más que un mensaje idealmente imperceptible y de difícil extracción por parte de un usuario no autorizado. La información de este mensaje puede representar, por ejemplo, una secuencia binaria que contiene un número de serie, un logo, un número de tarjeta de crédito, una imagen o una firma.

Existen dos tipos de marcaje conocidos como marcaje frágil y marcaje robusto.

- **Marcaje frágil:** es una inserción débil de la marca, pensada para desaparecer tan pronto como modificamos

demasiado el objeto a marcar. Este tipo de marcaje se emplea, por ejemplo, para la autentificación de datos.

- **Marcaje robusto:** este tipo de marcaje es imposible de eliminar sin degradar el objeto a niveles por debajo de los tolerables. Se emplea, por ejemplo, para la protección de copyright.

En la tabla 1 aparecen las propiedades más destacadas y deseables que han de tener los watermarks frágiles y robustos.

Marcaje frágil	Marcaje robusto
El watermark es alterado por la aplicación de las técnicas más comunes de procesado de imágenes.	El watermark permanece inalterado en una imagen, aún después de la aplicación de las técnicas más comunes de procesado de imágenes.
Es difícil para una persona no autorizada insertar un watermark falso.	El watermark es difícil de detectar por una persona no autorizada.
El watermark ha de ser rápidamente extraible por una persona autorizada.	El watermark ha de ser rápidamente extraible por una persona autorizada.
El watermark extraído indica donde han tenido lugar las alteraciones.	El watermark se ha de poder extraer, aún después de que la imagen marcada sufra alteraciones.

Tabla 1. Propiedades principales del marcaje frágil y robusto si la marca es invisible [1]

Existen también las denominadas técnicas semi-frágiles cuyas características consisten en ser robustas frente a transformaciones deseables, por ejemplo a la compresión, y frágiles a las manipulaciones mal intencionadas [11].

2. APLICACIONES DEL WATERMARKING

El watermarking tiene múltiples utilidades entre las que se pueden encontrar: [2, 3]

- **Protección de Copyright (identificación del propietario):** Para la protección de la propiedad intelectual, el propietario de los datos puede introducir una marca de agua representando la información de copyright de esos datos. Esta marca de agua puede probar su propiedad cuando alguien ha infringido sus copyrights. Actualmente la industria discográfica realiza grandes esfuerzos en introducir copyrights fiables en grabaciones en formato digital [12].



- “Fingerprinting” (toma de impresiones digitales – identificación del comprador): El propietario puede emplear una técnica de ‘fingerprinting’ para rastrear la procedencia de copias ilegales. En este caso, el propietario de los datos digitales puede introducir diferentes marcas de agua en las copias de los datos que son suministrados a diversos clientes. Esta técnica es comparable a introducir un número de serie en los datos, que está relacionado con la identidad del cliente. Esto permite al dueño de la propiedad intelectual identificar clientes que han roto su contrato (licencia) suministrando estos datos a terceras partes.

- Protección a la copia: La información almacenada en una marca de agua puede controlar directamente dispositivos de grabación digitales con el fin de evitar copias ilegales. En este caso, el watermark representa un bit de prohibición de copias y los detectores de watermark en las grabadoras determinan si esos datos pueden ser almacenados o no. Actualmente la tecnología DVD incluye técnicas de watermarking en su sistema de protección de copias que contiene posibles mensajes como: “copia no permitida”, “copiar una sola vez”, “agotado el número de copias”, ...[13]

- Vigilancia de la transmisión: Por ejemplo, para introducir marcas de agua en anuncios comerciales, un sistema de vigilancia automatizado puede verificar si los anuncios son transmitidos como han sido contratados.

- Autentificación de datos: Las marcas de agua frágiles pueden ser empleadas para comprobar la autenticidad de los datos. Este tipo de marcas indican si los datos han sido alterados y suministra información de localización como, por ejemplo, donde han sido alterados estos datos.

Las técnicas de watermarking se pueden emplear en otras aplicaciones además de como protección. Entre estas otras aplicaciones destacan:

- Seguridad médica: Puede ser una medida de seguridad útil incrustar la fecha y los datos del paciente en las imágenes médicas.

- Esteganografía (ocultación de los datos): Las técnicas de watermarking pueden ser empleadas para la transmisión de mensajes privados y secretos.

Dentro de las técnicas comentadas previamente, vamos a enfocar este escrito al caso del empleo de técnicas de watermarking como sistema de copyright.

Las marcas empleadas deben ser robustas a cualquier alteración que se produzca en su contenido e incluir información suficiente para autenticar al propietario de estos datos. De este modo, se evita que personas no autorizadas se adjudiquen la propiedad de los datos a tratar o que manipulen esta información, eliminándola o haciéndola ilegible.

Las marcas que se introducen para autenticar al propietario legal de los datos a marcar, son los denominados watermarks robustos, que han de ser invisibles y soportar cualquier manipulación que se produzca en su contenido.

En la Fig. 1 aparece la imagen original ‘lena’, la imagen marcada y la marca introducida, que no es más que la diferencia entre la imagen original y la imagen marcada. Podemos ver cómo en la fig. 1 b), la imagen marcada, no se aprecia diferencias significativas respecto a la fig. 1 a), imagen original. La fig. 1 c) representa la marca introducida. Para marcar esta imagen se han utilizado las técnicas desarrolladas en [16].



Figura. 1. a) Imagen original ‘lena’, b) imagen marcada y c) marca introducida en la imagen original (diferencia entre la imagen original y la imagen marcada)

Estos datos marcados serán posteriormente extraídos mediante un proceso de detección que se puede utilizar para demostrar la propiedad o no de los mismos.

3. PROPIEDADES DEL WATERMARKING EN IMÁGENES DIGITALES

En imágenes digitales, hemos de considerar algunos factores importantes a la hora de introducir una marca de agua. Es deseable un compromiso entre diversas características como son:

- Robustez.
- Invisibilidad.
- Cantidad de información.
- Seguridad.
- Recuperabilidad.

- Rapidez de extracción.
- Exclusión de ambigüedad.

• Robustez

Esta característica se refiere a la habilidad de la marca para permanecer en la imagen independientemente de que la calidad haya sido degradada, intencionada o no intencionadamente. Como ejemplos de degradaciones no intencionadas, podemos destacar, las aplicaciones relacionadas con el almacenamiento o la transmisión de datos, donde las técnicas de compresión son aplicadas a los datos para reducir las tasas de bits e incrementar la eficiencia. También, como degradaciones no intencionadas, encontramos las provocadas por conversiones analógico-digital, digital-analógicas, filtrados, muestreos, Otras manipulaciones incluirían las conocidas como distorsiones geométricas, que incluyen rotaciones, "cropping" (recorte de la imagen), deformaciones y desplazamientos.

Las características que ha de tener un sistema para que sea robusto son:

- La detección de la marca debería exigir el conocimiento de un secreto.
- Múltiples marcas no deben interferirse entre ellas.
- Las marcas deberían sobrevivir a cualquier posible ataque que no degrade la calidad a percibir.

• Invisibilidad

El algoritmo de watermarking debe introducir la marca de agua en la imagen de tal modo que ésta no afecte a la calidad de los datos que se encuentran debajo de ella. El proceso de introducción de la marca es realmente imperceptible si no somos capaces de distinguir los datos originales frente a los datos con la marca insertada.

Existe un compromiso entre la robustez y la invisibilidad. Cuanta más fuerza le demos a una marca, más visible se vuelve. Para ello se estudian máscaras perceptuales que, teniendo en cuenta el sistema visual humano (HVS), modulan el mensaje a insertar para utilizar la máxima potencia posible con la restricción de la imperceptibilidad. Por otro lado se estudian métodos de evaluación que nos den una medida de calidad visual a posteriori y que sean más apropiados que la relación señal a ruido (considerando como señal la imagen y como ruido la marca).

• Cantidad de información

Para lograr introducir la máxima información, ésta se reparte por toda la imagen. En algunas aplicaciones puede ser un factor determinante. Existen estudios teóricos sobre la capacidad máxima que se puede incluir en una imagen [4].

• Seguridad

En las técnicas de watermarking la marca a introducir ha de ser secreta y sólo accesible para las partes autorizadas.

Esta marca ha de ser resistente a posibles manipulaciones de usuarios no autorizados. Las personas que no posean derechos legales para su uso, no han de ser capaces de detectar ni descifrar la información introducida en ella.

• Recuperabilidad

En sistemas de detección ciega, donde no disponemos de la imagen original, la marca introducida ha de tener unas propiedades concretas que permitan su posterior recuperación por los usuarios autorizados que poseen permiso para su empleo.

Además, también hemos de tener en cuenta que la probabilidad de error en detección ha de ser baja para evitar que se produzcan falsas detecciones positivas.

• Rapidez de extracción

Es deseable que el proceso de detección sea rápido para toda persona autorizada o para el propietario original de las imágenes digitales. Sobre todo, en los sistemas que trabajan en tiempo real, es un factor determinante.

• Exclusión de ambigüedad

Para aplicaciones de copyright y fingerprinting, la marca introducida ha de identificar con total seguridad a la persona o empresa propietaria legal de los derechos de la propiedad intelectual de la imagen o al comprador.

4. MODELADO DE UN SISTEMA DE WATERMARKING COMO UN SISTEMA DE COMUNICACIONES

Un sistema de watermarking se puede interpretar como un sistema de comunicaciones convencional, en el que para el proceso de marcaje tenemos un canal de comunicación de datos (el contenido mismo de la imagen) que nos sirve como soporte de la información a introducir (marca de agua). En el proceso de detección, la imagen marcada es la señal recibida, mientras que la marca es la señal transmitida. Además, como en los sistemas ciegos la imagen original no es conocida en recepción, se puede considerar ésta como un ruido aditivo que dificultará la detección de los datos. En la fig. 2 podemos ver un diagrama de bloques que representa un sistema de watermarking con detector ciego (en el que no disponemos de la imagen original) representado como un modelo de comunicación



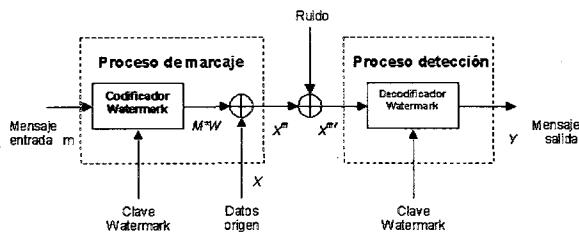


Figura2. Sistema de watermarking con detector ciego representado como un modelo de comunicaciones

El conocimiento del canal en el proceso de marcaje puede utilizarse para intuir y mejorar el proceso de detección. En este caso el modelo es un canal de comunicaciones con información lateral [3].

4. 1 Proceso de marcaje

La marca de agua se introduce de forma aditiva en la imagen, (aunque algunos autores utilizan modelos multiplicativos [17]). Así, en general, la señal marcada será:

$$X^m = X + a \cdot M \cdot W$$

donde X^m es el vector que representa los datos marcados, X representa el vector de los datos originales, W es la marca de agua que se introduce en los datos a tratar, M representa la máscara perceptual, que modifica los datos a marcar según sus características y, finalmente, a es un valor que pondera la marca introducida para controlar la potencia de la marca en los datos.

El proceso de marcaje puede introducir la marca directamente en los datos originales o bien en una versión transformada de los mismos. Entre los distintos dominios empleados podemos destacar, el temporal (para secuencias de audio) o espacial (para imágenes); y otros dominios transformados como son la transformada discreta de Fourier, DFT, la transformada discreta del coseno, DCT, [5, 6] o la transformada discreta *wavelets*, DWT [7, 5].

Las técnicas en dominios transformados son muy utilizadas debido a su estructura natural para incorporar conocimientos perceptuales en los algoritmos de marcaje y porque muchas técnicas de compresión como JPEG, MPEG2 (DCT) o JPEG-2000 (DWT) trabajan en esta misma estructura, lo que permitirá recuperar la marca con una decodificación parcial y no total de los datos. El uso de estos dominios hará posible obtener unos mejores resultados en sistemas donde se produce compresión.

La marca introducida puede ser un valor binario o real. En algunas aplicaciones, a esta marca se le aplica una clave secreta. Sin el conocimiento de ella, resultará difícil

extraer o alterar el mensaje insertado sin destruir los datos originales. Los sistemas que emplean técnicas de espectro ensanchado se usan frecuentemente para aplicaciones de este tipo [8]. Habrá otras, en cambio, donde no será deseable el empleo de claves. Los sistemas QIM (Quantization Index Modulation) serán los aconsejados en estos casos. [9, 10]

4. 2 Proceso de detección

La detección o verificación de la marca recibida se refiere al proceso de realizar una decisión binaria en el detector. Comprobar si una marca concreta se encuentra o no en los datos recibidos.

Existen diferentes esquemas de detección de la marca. Los que incluyen clave secreta y los datos originales, clave secreta pero sin el original (detección ciega) o los que no poseen ni clave ni datos de origen.

Para recuperar la información introducida en los datos marcados, se puede emplear la detección por máxima verosimilitud. También haremos referencia a un caso particular, la detección por correlación dada su simplicidad y buenos resultados [11]. Este método es óptimo en los casos con presencia de ruido blanco Gaussiano aditivo. La correlación normalizada es muy similar a la anterior. En este caso, previo a la realización del producto promedio entre los elementos de dos vectores, estos son normalizados. De este modo, disponemos de un detector más robusto a posibles cambios en los datos marcados.

En el detector, los valores de detección se calculan para cada una de las marcas que forman el mensaje introducido en la imagen. El mensaje más probable será aquel cuyas marcas tengan los valores de detección mayores. Si el valor obtenido es menor a un umbral, el detector no reconocerá la existencia de la marca de agua, mientras que si el valor es mayor a ese mismo umbral se considerará como una detección positiva de la watermark.

Con este método cabe la posibilidad de obtener falsos positivos, se detecte una marca cuando no existe, o de falsos negativos, que no detecte una que existe. Para muchas aplicaciones es más importante tener pocos o ningún falso positivo, aún a costa de aumentar el número de falsos negativos. Por ese motivo el valor umbral escogido, se puede modificar en relación a la probabilidad de detección y la probabilidad de falsos positivos.

Por otro lado, el desarrollo y evolución de los algoritmos de detección se ha beneficiado del estudio de los ataques que se producen en los esquemas de watermarking y en las posibles soluciones (contraataques) que se crean [14, 15]. Un ejemplo de ataque es la pérdida de sincronización que provoca que el cálculo de la correlación no sea efectivo. El contraataque propuesto consiste en usar partes de la marca para introducir una sincronización conocida en los datos a tratar.

5. CONCLUSIONES

El uso cada vez más extendido de Internet ha implicado una revolución en el manejo de información y datos digitales, hecho que ha propiciado el empleo de técnicas de *watermarking*, para asegurar fiabilidad y seguridad en los datos.

Las técnicas de watermarking se emplean para múltiples aplicaciones entre las que podemos destacar la protección de copyright, la protección a la copia o la autenticación de datos. Estas técnicas consisten en la introducción de un mensaje oculto en los datos a tratar. Muchas de ellas se apoyan en los conceptos de sistemas de comunicaciones y del sistema perceptual humano.

El watermarking proporciona robustez, frente a posibles ataques, a todas aquellas aplicaciones que lo emplean. Existirá un compromiso entre cantidad de información insertada en los datos a tratar y robustez. También existirá un compromiso entre robustez y percepción visual.

Existen aún muchas líneas de investigación abiertas tanto por lo que respecta a las actuales limitaciones que presenta el proceso de marcaje y como el de detección. El número de posibles ataques que se pueden producir en los datos marcados, como son las distorsiones geométricas, son incalculables. Todos estos factores crean un campo de investigación en desarrollo para conseguir un aumento de la efectividad de las técnicas de watermarking.

6. REFERENCIAS

- [1] F. C. Mintzer, G. W. Braudaway y M. M. Yeung, "Effective and Ineffective digital watermarks". IEEE in Proceedings, the International Conference on Image Processing, Oct. 1997, vol. 3, pp. 9-12.
- [2] Gerhard C. Langelaar, Iwan Setyaman, y Reginald L. Langendijk, "Watermarking Digital Image and Video Data". IEEE Signal Processing Magazine, Sept.2000, pp. 20-46.
- [3] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom, "Digital Watermarking". Morgan Kaufmann. 2002.
- [4] Pierre Moulin, M. Kivanç Mihçak, "A Framework for Evaluating the Data-Hiding Capacity of Image Sources". IEEE Transactions on Image Processing, Sep. 2002, vol.2, pp. 1029 – 1042.
- [5] International Workshop on Information Hiding, 1996.
- [6] E. Koch, J. Zhao, "Towards robust and hidden image copyright labeling". Nonlinear Signal Processing Workshop, Thessaloniki, Greece, 1995.
- [7] M. Swanson, B. Zhu, A. Tewfik, "Multiresolution scene-based video watermarking using perceptual models". IEEE J. Select. Areas Commun., May. 1998, vol. 16, pp. 525-539.
- [8] I. J. Cox, J. Kilian, T. Leighton, T. Shamoon, "Secure spread spectrum watermarking for multimedia". NEC Research Institute, Princeton, NJ, Technical Report 95-10, 1995.

- [9] B. Chen, G. Wornell, "Dither modulation: A new approach to digital watermarking and information embedding". SPIE, Security and Watermarking of Multimedia Contents, San Jose, CA, Ene. 1999, vol. 3657, pp. 342-353.
- [10] B. Chen, G. Wornell, "Preprocessed and postprocessed quantization index modulation methods for digital watermarking". SPIE, Security and Watermarking of Multimedia Contents II, San Jose, CA, 2000, vol. 3971, pp. 48-59.
- [11] Christine I. Podilchuk, Edward J. Delp, "Digital Watermarking: Algorithms and Applications". IEEE Signal Processing Magazine, Jul. 2001, pp. 33-46.
- [12] L. Boney, A. Tewfik, K. Hamdy, "Digital watermarks for audio signals". IEEE Proc. Multimedia, 1996, pp. 473-480.
- [13] J. Bloom, I. Cox, T. Kalker, J. P. Linnartz, M. Miller, C. Traw, "Copy protection for dvd video". Proc. IEEE, nº 87, Jul. 1999, pp. 12667-12760.
- [14] S. Craver, N. Memon, B.-L. Yeo, M. Yeung, "Can invisible watermarks resolve rightful ownerships?". SPIE Electronic Imaging '97: Storage and Retrieval of Image and Video Databases". 1997, pp. 310-323.
- [15] I. Cox, J.-P. Linnartz, "Some general methods for tampering with watermarks". IEEE J. Select. Areas Commun., May. 1998, vol. 16, pp. 587-593.
- [16] Elena Martínez, "Watermarking de Imágenes en Color en el dominio wavelet". Proyecto Fin de Carrera. ETSETB. UPC. 2003.
- [17] M. Barni, F. Bartolini, A. de Rosa, A. Piva, "A New Decoder for the Optimum Recovery of Nonadditive Watermarks". IEEE Transactions on Image Processing, May. 2001, vol. 10, nº 5, pp. 755-766.

AUTORES



Elena Martínez Villacampa nació en Barcelona el 21 de octubre de 1976. Finalizó los estudios de Ingeniería Técnica de Telecomunicaciones, especialidad en Sistemas de Telecomunicación en la Escuela Universitaria Politécnica del Baix Llobregat (UPC) en 1999. En la actualidad está realizando el proyecto de fin de carrera sobre watermarking de imágenes en color en el dominio wavelet en el departamento de Teoría de Señal y Comunicaciones de la Escuela Técnica Superior de Ingeniería de Telecomunicaciones de Barcelona (UPC) y trabaja como becaria en la empresa Vodafone.



Elisa Sayrol Clols recibió el título de Ingeniera de Telecomunicación y el título de Doctora de la Universitat Politècnica de Catalunya en los años 1989 y 1994 respectivamente. Desde 1990 hasta 1993 hizo cursos de postgrado en la Northeastern University, y fué investigadora visitante en la University of Southern California, donde trabajó con el Profesor C.L. Nikias. Actualmente es profesora titular de la Universitat Politècnica de Catalunya, en la Escola Superior d'Enginyeria de Telecomunicació de Barcelona e imparte el curso Senyals i Sistemes 1, así como el curso de doctorado Processat Digital d'Imatge. Sus intereses investigadores incluyen estadísticas robustas i de orden superior, el análisis de imagen y video y las técnicas robustas de watermarking.





LIMITES COLATERALES

Miguel Escudero

Departament de Matemàtica Aplicada, Universitat Politècnica de Catalunya

escudero@mat.upc.es

“Que errar lo menos no importa quien acertó lo demás”, dice el alcalde de Zalamea en la célebre obra de Calderón. El rey hará suya la frase en una de las últimas escenas. Para estar de acuerdo con ella hay que precisar el contexto y distinguir el orden de magnitud de lo que se desprecia. Todo lo que es humano tiene límites.

Tienen límites nuestra visión, nuestro oído, nuestro gusto, nuestro tacto, nuestro olfato. Pero no sólo nuestros sentidos, también el día y la noche, la felicidad y la tristeza, nuestra altura y nuestro peso, la longitud de un artículo y su nivel. El hombre anda envuelto en umbrales.

En cuanto persona, el hombre puede ser más de lo que es, por eso necesita soñar para rebasar sus propias limitaciones. Es forzosamente libre, pero está fuertemente condicionado. A su alrededor encuentra vallas atravesadas, así sucede con su pasado y su presente, realidades irrevocables aunque se puedan revertir mientras se viva (“ni el pasado ha muerto ni está el mañana –ni el ayer- escrito”, sentenció Antonio Machado). Su futuro está por hacer, está indeterminado pero no puede ser cualquiera. Mi porvenir reside en lo que yo haga ‘mañana’ con lo que he hecho ya y con lo que me va deparando el destino.

En matemáticas existe, esto es, se usa el concepto de límite, cuya formulación actual data de hace un par de siglos. Se habla del límite de una sucesión o de una función en un punto, siempre de una o varias variables. En 1821, el prolífico escritor de matemáticas Augustin Louis Cauchy decía en un trabajo que “se llama variable a una cantidad que se considera que tiene que tomar sucesivamente muchos valores diferentes unos de los otros”. Nacido en 1789, el parisino Cauchy fue un ardiente partidario de los Borbones y por motivos políticos se exilió en Turín, donde durante varios años enseñó latín e italiano. Fundamentó el cálculo infinitesimal sobre bases más sólidas que sus predecesores, fijando de manera precisa el sentido de las notaciones empleadas logró hacer desaparecer las vaguedades de argumentos hasta entonces consentidos.

Cauchy llamó infinitésimo a una cantidad variable que “se hace infinitamente pequeña cuando su valor numérico decrece indefinidamente de modo tal que converge al límite cero”. El sacerdote y matemático bohemio Bernhard Bolzano (1781-1848) dio en 1817 la definición ‘apropiada’ de la continuidad de una función en un punto. Karl

Weierstrass (1815-1897), uno de los personajes más simpáticos de la historia de las matemáticas, interpretó una variable simplemente como una letra que se usa para denotar cualquier valor de los de un conjunto dado que se le puede asignar a la letra. Estas contribuciones fueron capitales para hablar con rigor de límites. (Si existe, el límite es único, es decir, no importa la dirección o el sentido de la aproximación).

Los límites primordiales de los hombres son, no importa cuáles sean sus trayectorias, su nacimiento y su muerte. Y ellos conducen a los mitos. Mircea Eliade recuerda que, en lenguaje coloquial, los mitos son fábulas, pero son la única revelación válida de la realidad para los pueblos arcaicos (las sociedades tradicionales). El mito explica cosas imposibles o improbables. Así, el mito del buen salvaje (o el prestigio del origen) prolonga el mito de la edad de oro, la perfección del comienzo de las cosas; de modo laico revaloriza el mito del paraíso terrenal.

El hombre moderno despreciaba el mito como explicación opuesta a la razón y la ciencia, pero en verdad no sabe desprenderse de él. El mito es refugio de fantasías y nostalgias. Hace cincuenta años Eliade veía que a escala colectiva el mito se manifestaba con gran fuerza bajo forma de mito político y que, fuera cual fuese su estadio de cultura y su forma de organización social, una de las nostalgias básicas el hombre es siempre el deseo de libertad absoluta. La participación, dice, de toda una sociedad en ciertos símbolos ha sido interpretada como prueba de la supervivencia del ‘pensamiento colectivo’, y, por tanto, de cierto comportamiento mítico. Éste puede descifrarse, en medio de camuflajes, en la vertiginosa cantidad de distracciones inventadas por el hombre moderno. Por más que esté desprovisto de una experiencia religiosa auténtica, tiende a descuidar el tiempo presente, el ‘momento histórico’. Lo que importa es salir de la propia historia y vivir un ritmo temporal cualitativamente distinto. “La verdadera ‘caída en el tiempo’ comienza con la desacralización del trabajo; el hombre se siente prisionero de su oficio sólo en las sociedades modernas, pues no puede escapar del tiempo. Y como no puede ‘matar’ su tiempo durante las horas del trabajo –es decir, mientras desempeña su verdadera identidad social- se esfuerza por ‘salir del tiempo’ en sus horas libres”, escribe Mircea Eliade.

Así, nos encontramos el afán sin límite por alargar la noche los fines de semana en recorridos organizados o la pasión por vociferar o buscar pelea en los estadios de fútbol, por dar dos ejemplos que convocan multitudes. Las noticias de los medios de comunicación se concentran en cualquier suceso truculento y perdemos los contornos de la realidad, su frontera queda desfigurada.

No hay fecha sin mito. Todo acaba por desvanecerse en nuestra memoria, incluso lo que no vivimos. El año que viene, se cumplirá el primer medio siglo de, por ejemplo, la muerte de Stalin y el fin de la guerra de Corea, de la concesión del premio Nobel de literatura a Winston Churchill y de la ascensión de Hillary y Norkay a la cima del

Everest, del estreno de la película ‘Bienvenido Mister Marshall’, premiada en el festival de Cannes, y de la ruptura de Chaplin con el cine de Hollywood, yéndose a vivir a Suiza. A estos acontecimientos podríamos agregar muchos más, acaso se conmemoren, pero ¿cuáles serán los límites de nuestro conocimiento de ellos, recuperaremos el aroma con que fueron vividos por gentes tan distintas, su repercusión y significado, sabremos conectarlos con su posterior conexión, la que impone el paso de los años?

Los países europeos se van poblando con numerosas gentes venidas de otros continentes. Cada vez se va haciendo más patente nuestro estado de ignorancia con respecto a las realidades ajenas. Hay gentes instruidas que se dedican a dar oxígeno a necesidades instaladas en los hombres y los pueblos. Leo con disgusto un largo ensayo con un título desmedido y pretencioso, que inevitablemente defrauda al lector, pero que ha estado las últimas semanas entre los libros de ‘no ficción’ más vendidos; a pesar de tal clasificación, se puede leer ahí que Juana la Loca asesinó a Felipe el Hermoso y que Hitler declaró la guerra a los Estados Unidos.

Para no entrar en los viejos tópicos que enjarea de los españoles, recojamos este otro inefable aserto: los alemanes han tenido con Francia hasta 1945 “la misma relación que hoy tienen los árabes o los iraníes con los norteamericanos: los odiaban porque admiraban su clara superioridad”. El estado de error va en muchas direcciones y por eso se hace imprescindible volver la mirada a la realidad personal de los hombres.

Eliade lamentaba que a menudo el occidental se dejase impresionar por la manifestación de una ideología, ignorando lo único que importa conocer: “la ideología en sí misma, es decir, los mitos”. Hace cincuenta años decía que ya no era suficiente, como lo había sido otro medio siglo atrás, descubrir y admirar el arte negro o polinesio; “hay que descubrir las fuentes espirituales de ese arte en nosotros mismos”, afirmaba. Esa era una necesidad para que la cultura occidental no acabase siendo provinciana, dialogar con otras culturas no

europeas, y esforzarse en no errar el sentido de los términos. Proponía tratar de vernos tal y como aparecemos ante los ojos de ‘fuera’: “Es urgente que comprendamos cómo se nos considera y juzga, en tanto que forma cultural, desde una perspectiva extra europea. No hay que olvidar que todas las culturas cuentan con una estructura religiosa, es decir, que han surgido y se han constituido en tanto valoración religiosa del mundo y de la existencia humana”. Hablaba del diálogo con el “verdadero mundo asiático, africano y oceánico” como de un modo para descubrir de nuevo posturas espirituales que puedan ser consideradas como universalmente válidas: “no se trata de fórmulas provincianas, creaciones de un único fragmento de la historia, sino de posiciones – osaríamos decir- ecuménicas”.

No hay límites para saber que hace tres siglos el orientalista y viajero Jean Antoine Galland dio a conocer al público occidental los cuentos de ‘Las mil y una noches’, en una serie de doce volúmenes traducidos por él mismo al francés. En aquél tiempo ‘mítico’ la ciencia islámica tuvo su primera edad de oro. Abdullah al Ma’mum, califa citado en esos relatos, fundó en Bagdad (la actual capital de Irak) la Casa de la Sabiduría y tuvo en su corte como bibliotecario al padre del álgebra.

Éste era asimismo astrónomo en el observatorio de la capital de Mesopotamia. Se llamaba Mohamed ben Musa y era conocido como Al-Khwarizmi. Nació el año 780 de nuestra era, siglo y medio después de morir Mahoma, en Uzbekistán país que fue soviético y que es fronterizo con Afganistán. El nombre Al-Khwarizmi ha originado los de guarismo y algoritmo. Su obra principal fue sumamente divulgada y traducida en Europa, trata de la ciencia de las ecuaciones, también conocida como el ‘arte de la cosa’, la cosa era la incógnita. Se impuso la denominación de álgebra: al-jabr, ‘restauración’, el paso de un miembro a otro en una ecuación. Hay un significado arcaico de álgebra recogido en el Tesoro de la lengua castellana o española (1611), de Sebastián Covarrubias, como el “arte de concertar los huesos desencajados y quebrados”. De ahí que algebrista fuera también un cirujano dedicado a la curación de dislocaciones de huesos, restituyéndolos a su lugar. Es la misión que distingue a un intelectual en una sociedad limitada.

AUTOR



Miguel Escudero es profesor titular de la Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona, adscrito al Departamento de Matemática Aplicada IV de la Universidad Politécnica de Cataluña. También es Doctor en Filosofía y Letras y ha escrito artículos de opinión en publicaciones como ‘Cuenta y Razón’, el diario ‘La Vanguardia’ o ‘Buran’.



DESPLIEGUE DE UNA WLAN EN LA EPSC

Rafael Vidal Ferré y Xavier Bordoy Rodríguez

rafael.vidal@entel.upc.es: Departamento de Ingeniería Telemática

Universitat Politècnica de Catalunya;

xavier.bordoy@upc.es: Becario investigación i2CAT - Mediaticat (tecnologías multimedia)

Universitat Politècnica de Catalunya

1. INTRODUCCIÓN

Durante el pasado curso 2001/02, a iniciativa del grupo de comunicaciones inalámbricas del Departamento de Ingeniería Telemática y como parte del proyecto I2Cat (Internet 2 a Catalunya), se llevó a cabo el despliegue de una red WLAN con tecnología IEEE 802.11b a la EPSC (Escuela Politécnica Superior de Castelldefels).

El objetivo del despliegue era doble. Por un lado ofrecer una plataforma sobre la que desarrollar y probar soluciones para el soporte de la movilidad, como por ejemplo Mobile IP y Cellular IP, y aplicaciones que sacan provecho de ella, como teléfonos sin hilos utilizando VoIP, SIP o PDAs o servicios de localización. Por otro, se pensaba en ofrecer a toda la comunidad que forma la EPSC la posibilidad de un acceso sin hilos de calidad a su red y por extensión a la de la UPC y a Internet.

El presente artículo se centra en explicar como se ha llevado a cabo este segundo propósito. En primer lugar haremos un breve resumen de las principales características de la tecnología WLAN utilizada, el estándar IEEE 802.11b, como paso previo a la descripción de los equipos de esta tecnología que se probaron y de los criterios seguidos para seleccionar los que finalmente se utilizaron. A continuación se definirán los criterios y objetivos que guiaron el despliegue de la red y posteriormente el proceso seguido para realizarlo. Se seguirá, comentando la integración de este acceso sin hilos a la red de la EPSC indicando los problemas más significativos que esta suponía, en especial de seguridad, y como se resolvieron. Para acabar se detallará las conclusiones.

2. EL PROTOCOLO 802.11B

El protocolo 802.11b pertenece a la familia de protocolos 802.11 desarrollado por el IEEE. Puede considerarse una adaptación del estándar Ethernet al medio radio. El cambio de medio de transmisión nos introduce dos modos de funcionamiento: Managed y Ad-Hoc. La diferencia radica en el hecho en que en una red Ad-Hoc la comunicación se realiza entre los terminales directamente y en una red en modo Managed hay un elemento llamado Access Point (AP) que es el que gestiona todas las comunicaciones, y haciendo un símil con las redes cableadas tendría, inicialmente, un

papel similar al de un hub. La red desarrollada en la EPSC se corresponde con este segundo tipo.

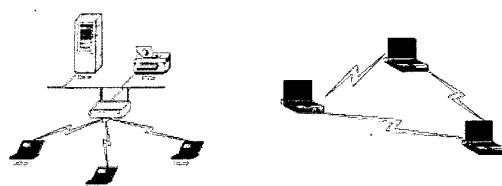


Figura 1. Sistema Managed y Ad-Hoc

Como en el caso del 802.3, el 802.11b sólo especifica el funcionamiento de las capas física y de control de acceso al medio (dentro del esquema OSI).

2.1. Capa física

La capa física se encarga de la codificación y de la modulación, básicamente. Cada AP utiliza una banda de frecuencias y las distribuye entre los usuarios. Este reparto se puede llevar a cabo de dos maneras, dependiendo de la tecnología usada. La primera es el FHSS (Frequency Hopping Spread Spectrum) donde se divide el medio en 75 canales no solapados que van saltando según un patrón definido que conocen los AP y los dispositivos conectados a ellos y permite tener diversas redes en una misma área. En segundo lugar tenemos el DSSS (Direct Sequence Spread Spectrum) donde la división se produce por código y no por frecuencia. El estándard define un total de 14 canales con un ancho de banda de 30 MHz por canal y una separación entre ellos de 5 MHz que tal y como se observa en la figura 2 se traduce en un solapamiento. Debido a esto es recomendable, y siempre que sea posible, dejar un espacio libre de cinco canales. Esta tecnología es actualmente la más usada, ya que permite llegar a tasas de hasta 11Mbps, mientras que con el FHSS no se pasa de 2Mbps.

Hay que remarcar que todos estos canales no están disponibles en todos los países, y dependiendo de la legislación vigente en cada país serán distintos para cada uno de ellos. En España, y según consta en la norma UN-115, dentro de la banda de los 2.4 GHz disponemos de los canales 1 al 13. Además, para su uso libre dentro de esta banda de frecuencias los dispositivos deben estar limitados a una potencia máxima de radiación de 100 mW EIRP en toda la UE.

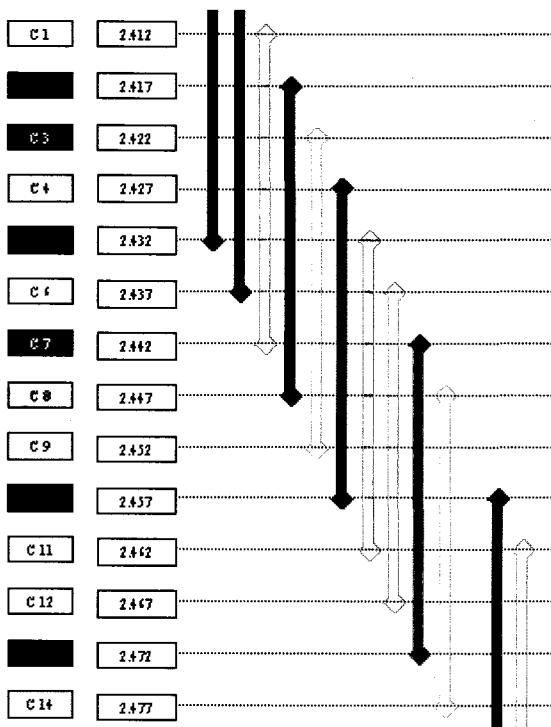


Figura 2. Canales DSSS

2.2. Capa de control de acceso al medio

La capa de control de acceso al medio, como su nombre indica, aparte de controlar el acceso al medio, realiza el control del roaming (cambio de AP de los usuarios), control de autenticación y finalmente también realiza el control de energía. No debemos olvidar que estamos hablando de una tecnología desarrollada para dispositivos móviles alimentados por baterías y consecuentemente, la conservación de la energía es un elemento básico para estos dispositivos.

En referencia a la capa de control de acceso al medio, como ya hemos dicho, efectúa diversas tareas. La primera es la de evitar las colisiones. Por eso ha de coordinar todas las comunicaciones y ha de establecer métodos de detección de medio ocupado. El que se ha definido como sistema obligatorio es el CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance). Con este sistema, tal y como se indica en su nombre, lo que se hace es escuchar el medio para ver si está ocupado (detector de portadora) y en caso de que lo esté no se transmite para evitar colisiones. Junto a esta tecnología se ha añadido la extensión RTS-CTS (Ready To Send - Clear To Send) que evita el problema de los terminales ocultos y sobreexpuestos.

En segundo lugar, junto al señal de sincronización se envía información del área (BSSID) para poder localizar la red wireless y poderse asociar. Esta señal de sincronización lo que hace es regular las comunicaciones, ya que dependiendo de la prioridad de las comunicaciones a realizar se

asignan tiempos de contención distintos a la espera de que se libere el medio. Cada dispositivo genera un tiempo de contención aleatorio para evitar que se lleguen a producir colisiones en el nuevo intento de entrada, ya que si hubiese dos dispositivos que esperasen el mismo tiempo para acceder, al volver a escuchar el medio ambos lo verían libre y empezarían la comunicación colisionando.

En tercer lugar hay el tema de la movilidad de los usuarios. En efecto, los usuarios estarán en movimiento y consecuentemente, no estarán siempre cubiertos por el mismo AP. Así que es necesario algún mecanismo de control que notifique que un usuario ha cambiado de nodo de control (AP) y por tanto, todos sus paquetes ahora irán por otro camino. Para poder llevar a cabo esto el control cada AP tiene una base de datos donde están registrados todos los usuarios y puede ser consultado por los AP's vecinos para poder realizar el envío de paquetes más eficiente posible. El cambio de zona se produce cuando el terminal no tiene suficiente potencia para continuar estableciendo una comunicación de calidad (rápida) y entonces escanea el medio en busca de un nuevo punto de conexión más potente.

En cuarto lugar hay el control de potencia. Como que la energía es limitada (duración limitada de las baterías) los dispositivos entran en reposo cuando no se están usando, y miran el medio periódicamente para saber si hay algún paquete en espera de ser entregado. Por este motivo la sincronización es básica, ya que si un AP recibe un paquete para un dispositivo que está en reposo sabe el tiempo que tardará este en volver a mirar el medio y por tanto podrá entregarlo. En caso contrario, si no estuviesen sincronizados, el paquete podría demorarse excesivamente, y incluso podría llegar a ser eliminado del buffer por acumulación de paquetes en cola.

Finalmente, hablar del protocolo de cifrado, llamado WEP, Wired Equivalent Privacy es decir, nivel de privacidad equivalente a una red cableada. Puede ser de 64 o 128 bits y en principio parecía un sistema bastante robusto, pero se ha demostrado que romperlo es muy fácil, y que el paso de 64 a 128 bits simplemente aumenta linealmente el tiempo necesario para hallar el código. Además, este sistema obliga al usuario y al personal de administración de la red a gestionar las claves de cifrado que se usarán, cosa que añade inconvenientes a su correcta utilización.

2.3. El certificado Wi-Fi

Para concluir este apartado, comentaremos qué es Wi-Fi (Wireless Fidelity) y cuál es su relación con el IEEE 802.11.

Wi-Fi no es nada más que un certificado que acredita que los dispositivos de distintos fabricantes podrán interactuar entre ellos sin ningún tipo de problema. Esta certificación se originó debido al hecho de que en sus inicios, el protocolo del IEEE 802.11b no estaba totalmente definido



y dejaba mucha libertad a los fabricantes, de modo que si no comprabas todos los dispositivos a un mismo fabricante te podías encontrar con el problema de que no se comunicasen todos los dispositivos debido a errores en el formato de las transmisiones. Esto se traducía en un freno en la expansión del 802.11b. Para superar este problema se creó una asociación de fabricantes, la WECA (Wireless Ethernet Compatibility Alliance), con el objetivo de crear un certificado de interoperabilidad, el certificado Wi-Fi.

El éxito de Wi-Fi y el hecho que el término Wireless Ethernet no fuese usado demasiado, normalmente se utiliza el término WLAN, ha hecho que la WECA pasase a denominarse Wi-Fi Alliance y que utilizase Wi-Fi como sinónimo de todas las redes 802.11 y no solamente de las 802.11b. Así actualmente oímos hablar de redes Wi-Fi, es decir 802.11, formadas por productos con o sin certificación Wi-Fi. Para consultar una lista de productos con certificado Wi-Fi se puede visitar la web de la Wi-Fi Alliance, www.wi-fi.org <<http://www.wi-fi.org>>.

A día de hoy esta certificación aún tiene sentido debido a que los protocolos dejan a los fabricantes posibilidades para que incluyan mejoras, algunas en fase de estandarización, pero también han definido muy claramente cuáles han de ser las funcionalidades básicas que se han de cumplir para que las comunicaciones se puedan llevar a cabo. Estos mínimos son los que asegura el certificado Wi-Fi. El posible soporte de cualquier función adicional queda fuera del ámbito de garantía de Wi-Fi.

3. EQUIPOS PROBADOS PARA EL DESPLIEGUE

Para poder llevar a cabo el despliegue de la infraestructura wireless de la EPSC se provaron diversos equipos para poder ver su rendimiento, así como las opciones que los diferenciaban del resto.

En primer lugar, se probaron dos AP's. Un Cisco Aironet 340 y un 3Com AirConnect. En todas las medidas que se hicieron, no se apreciaron grandes diferencias ni en lo referente a cobertura como a la calidad de las comunicaciones. Primero, hay que recordar que estamos en la banda ISM y que estamos limitados en potencia, y que por tanto no eran de esperar muchas diferencias en las áreas de cobertura debido a que los dos AP's trabajan a la potencia máxima permitida y con un diagrama de radiación omnidireccional. Sí que habría habido diferencias si se hubiesen usado antenas exteriores adicionales, pero este no era nuestro caso de interés, ya que lo que buscábamos era una radiación isotrópica que nos cubriera el máximo volumen posible. Así que la elección del AP se hizo más por cuestiones de adaptación a nuestras necesidades más que por cuestiones de rendimiento.

En segundo lugar se evaluaron tres tarjetas PCMCIA, Lucent WaveLAN Silver, Cisco Aironet 340 y 3Com AirConnect. Aquí sí que apreciaron diferencias de calidad

entre ellas. En primer lugar, las tarjetas 3Com son las que obtuvieron el peor resultado en las pruebas que se realizaron, ya que eran las que peor negociaban la velocidad de comunicación con el AP, y en segundo lugar porque eran las que menos potencia recibían en igualdad de condiciones. Las otras dos tarjetas probadas demostraron una alta calidad de funcionamiento, ya que se mantenían en los 11 Mbps casi hasta las zonas límite de cobertura. I para destacar una diferencia tenemos que la Lucent permite acoplarle una antena exterior que mejoraría las prestaciones, ya que con una antena de mayor calidad mejoraríamos la calidad de la señal recibida, aumentando el área de cobertura o la calidad de la señal, cosa que con la tarjeta Cisco no es posible. El siguiente aspecto que se evalúa es el soporte que tienen las tarjetas para los diferentes sistemas operativos. Aquí 3Com vuelve a quedar en último lugar, ya que sólo ofrece drivers para Windows, mientras que Cisco y Lucent, a parte del ya mencionado, ofrecen drivers y software para Linux, Unix y un largo etcétera de sistemas operativos. Este es un aspecto bastante importante, ya que las redes wireless son bastante heterogéneas en cuanto a los usuarios y es importante que no estén obligados a cambiar sus sistemas operativos tan solo para poder un hardware determinado.

Una vez finalizadas las pruebas, se decidió el AP usado sería el 3Com AirConnect y que la tarjeta PCMCIA que se usaría para las mediciones sería la Lucent. La decisión se tomó por los siguientes motivos. En primer lugar, el rendimiento que mostraron los dos AP's era muy similar, así que descartar alguno por este motivo no era fácil. Así que el elemento determinante fue la adaptación del AP al que sería nuestro entorno. Los AP's estarán conectados directamente a la red de la escuela para poder acceder a los recursos que esta nos ofrece, y por tanto detrás habría un conjunto de servidores que se encargarían del control de acceso a los recursos, así que en este sentido lo que se necesitaba era un punto de access que no tuviese demasiadas herramientas adicionales, y en esto el AP de 3Com se presentaba como un equipo más acorde a lo requerido, ya que el AP Cisco presentaba un conjunto de funcionalidades que en nuestro entorno no eran necesarias (com por ejemplo servidores DHCP, BOOTP y opciones de filtrado avanzado entre otras, que en nuestra infraestructura no eran requeridas).

En lo referente a la tarjeta de acceso, la elección entre la Lucent y la Cisco fue bastante a gusto personal, ya que ambas presentaban unas características de funcionamiento muy similares y softwares de medición muy completos.

4. DISEÑO DE LAS COBERTURAS

4.1. Requerimientos

En primer lugar, hay los requerimientos de lo que hay que cubrir. Este despliegue representaba la primera fase de lo que se espera sea la red en un futuro, y por tanto no se quería cubrir todo el edificio en el que se ubica la Escuela.

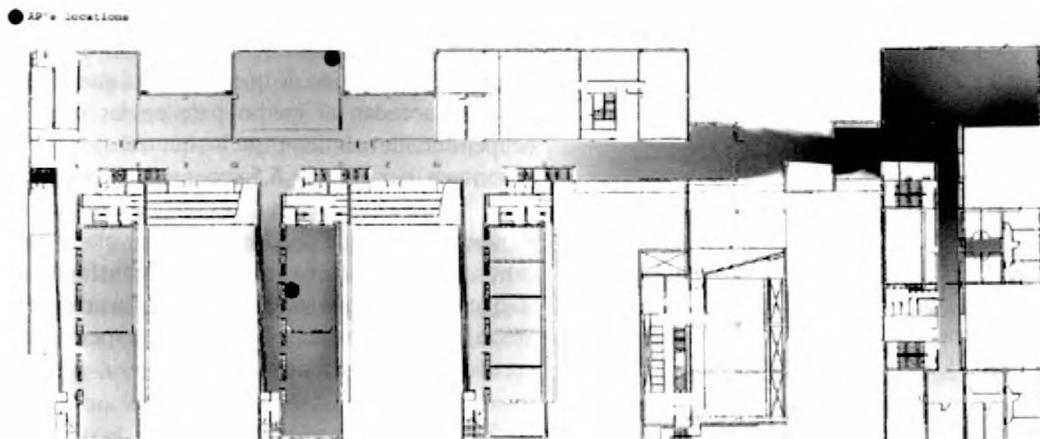


Figura 3. Plano coberturas y localización de los AP's

Las áreas que se tenían que cubrir eran todas zonas comunes de la planta baja del edificio, que son salas de estudio, salones de actos y sala de reuniones, la biblioteca y el bar. En segundo lugar, se deseaba ofrecer una continuidad en el servicio, o sea, que los usuarios que se moviesen entre las áreas comentadas antes no perdiesen la comunicación en ningún momento. El tercer requisito era el de minimizar el número de AP's, por dos motivos. El primero y más obvio, es el factor económico. Cuantos más dispositivos se usen, más caro será el despliegue. El segundo, y no menos importante, es el tema de la ocupación de canales, ya que a más puntos de acceso, más canales ocupados, y debemos recordar que los canales se solapan y por tanto no todos los canales son útiles en todos los lugares, ya que si usamos canales cercanos deberemos mantener una distancia física suficiente para que no se creen interferencias, o deberemos reducir la potencia de transmisión.

Y como última condición al despliegue había que garantizar la calidad de la comunicación, entendiendo como tal la máxima velocidad de acceso posible, o sea garantizar los 11 Mbps en todas las zonas a cubrir. Y con estas premisas se inició el despliegue.

4.2. Proceso de ubicación de los AP's

Lo primero que se hizo fue un cálculo estimativo de las atenuaciones que se originaban por distancia y por los distintos materiales que constituyan la estructura del edificio a partir de medidas hechas con el AP y la tarjeta seleccionada. Estos valores servirán de orientación para realizar una primera estimación de cuál podría ser la ubicación ideal de los equipos.

En cuanto a las medidas previas destacar la importancia que en el desarrollo de este proyecto han jugado las zonas acristaladas del edificio. Los vidrios no suponen ningún tipo de atenuación adicional para la señal y por tanto es conveniente aprovechar dichas superficies el máximo posible con tal de mejorar la calidad de la señal en las áreas a cubrir.

Una vez decididas las ubicaciones, se sitúan los AP's de forma provisional en la que será su ubicación definitiva y se miden las coberturas ofrecidas. Si son satisfactorias, es decir, cumplen con los requisitos de SNR y cobertura, se decide que esa será su ubicación. Entonces se pasa a determinar la localización del siguiente AP, y así se va reiterando el proceso, hasta obtener el resultado definitivo. Para minimizar el número de movimientos es muy importante que las mediciones iniciales, a pesar de ser ser aproximadas, sean lo más correctas posibles, ya que un mal cálculo nos puede hacer esperar unos resultados que posteriormente será imposible obtener, y como consecuencia, esto nos llevará a un número elevado de mediciones y cálculos que nunca coincidirán con lo esperado.

En la figura 3 se puede observar el plano de la planta baja del edificio y la ubicación de los AP's y el área que cubre cada uno donde están garantizados los 11Mbps.

5. EXPLOTACIÓN DE LA RED

Actualmente la WLAN actúa como una extensión de la red de la EPSC y en concreto de una de sus VLANs (Virtual LAN). Un usuario cualquiera con un sistema operativo plug&play tan solo necesita insertar la tarjeta WLAN en su ordenador y mediante DHCP obtiene todos los parámetros de configuración TCP/IP necesarios para comunicarse con la red. Esta situación que en principio puede parecer la idónea puede provocar gran cantidad de problemas, algunos asociados a la seguridad y otros de escalabilidad, derivados principalmente de las particularidades de la tecnología 802.11b. A continuación pasamos a comentar los más importantes y cuales han sido las soluciones aplicadas.

5.1. Problemas que se plantean

La VLAN en la que se encuentran los APs está asociada a una subred IP con un total de 254 IPs públicas libres a repartir entre los diferentes equipos que la participan, algunos de ellos usuarios de la WLAN. Esta situación

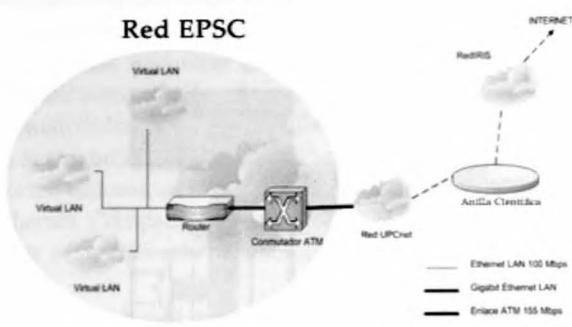


Figura 4. Esquema simplificado de la red EPSC y su salida a Internet

puede provocar la aparición de dos problemas. El primero una posible insuficiencia de direcciones IP que actuará como limitador del número de usuarios posibles de la WLAN. El segundo viene dado por el hecho de que el access radio es más inseguro por naturaleza, y en concreto la tecnología 802.11 se ha demostrado bastante vulnerable. Esto podría suponer un agujero de seguridad potencial para todos los equipos que forman parte de la VLAN, ya sean usuarios de la WLAN o no. Anteriormente ya se ha comentado que el cifrado WEP no es una buena solución para garantizar la confidencialidad de los datos de los usuarios de la WLAN. Además el suyo presenta algunos inconvenientes prácticos asociados: necesidad de tarjetas y AP's compatibles, distribución de claves y caída de la velocidad de transmisión.

Per otro lado el mecanismo de acceso del 802.11b puede ofrecer como mucho 11Mbps que se pueden convertir en el mejor de los casos en poco más de 6Mbps de datos de aplicación. Estos recursos són a repartir para todos los usuarios conectados a un mismo AP, y si su número es muy elevado se ha comprobado que la velocidad que obtiene cada usuario está por debajo del resultado de dividir la velocidad total de acceso por el número de usuarios activos.

Otra situación que puede degradar las prestaciones de la red es el hecho que algunos usuarios accedan a baja velocidad (por ejemplo 2Mbps) debido a que se encuentren en zonas límite de cobertura. A menor velocidad, más tiempo de transmisión, lo que provocará que cuando estos usuarios accedan al medio para enviar una trama lo ocupen durante más tiempo que los que trabajen a velocidades superiores, por ejemplo 5,5 veces más que los que van a 11Mbps. La coexistencia de estos dos tipos de usuarios provoca que los que pueden trabajar a velocidades superiores vean reducida su velocidad de transferencia global ya que a pesar de que su velocidad de transmisión no se ve afecta sí que aumenta el tiempo de espera para poder acceder al medio y transmitir.

5.2. Soluciones usadas

Como primera medida para preservar la seguridad del resto de equipos de la red de la EPSC se decidió crear una nueva VLAN formada únicamente por los AP's de la WLAN. Además esta VLAN se conectaría al resto de la red de la EPSC a través de un PC Linux que actuaría como Firewall. De esta manera se consigue aislar y controlar todo el tráfico de la red WLAN. Aprovechando esta arquitectura de red se decidió usar IPs privadas y que el PC Linux Firewall realizase las funciones de NAT y servidor DHCP, convirtiendo el PC Linux en lo que se conoce como *Wireless Firewall Gateway* (WFG).

Con esta propuesta se rompía el límite impuesto por número de IP's públicas disponibles y gracias al servidor DHCP propio se podía realizar una asignación de IP's a medida en la que, por ejemplo, a diferentes colectivos (alumnos, PAC, PAS o otros) se les asignan grupos de IP's diferentes que faciliten su diferenciación, incluso en términos de QoS utilizando Diffserv, y pudiesen dar a cabo políticas de filtrado distintas en el Firewall.

Respecto al problema de la confidencialidad de los datos enviados por el medio radio se ha desestimado la utiliza-

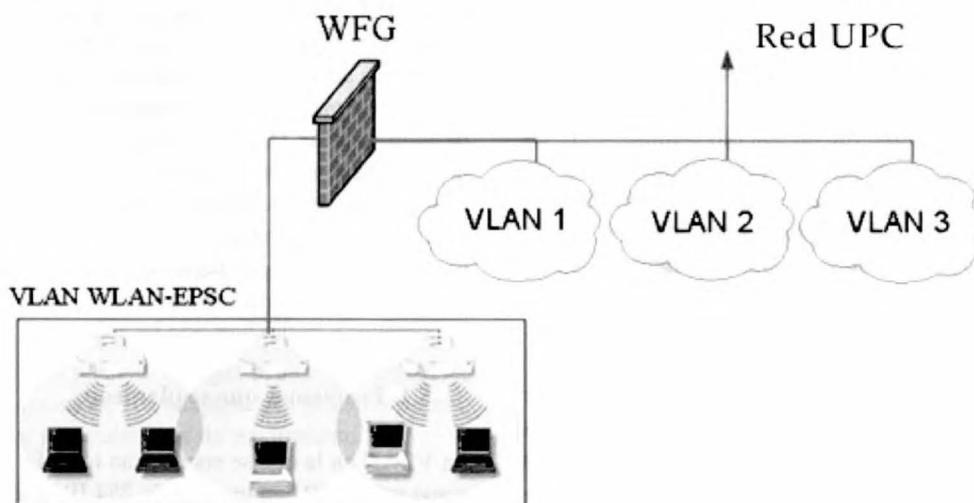


Figura 5. Integración de la WLAN en la red de la EPSC

ción de WEP porque hemos creido que sus inconvenientes pesaban más que sus virtudes. El usuario ha de ser consciente de que utiliza un canal inseguro y si lo cree conveniente puede utilizar herramientas alternativas para proteger sus datos: cifrado del correo, acceso a web o correo seguro o redes privadas virtuales.

Para evitar que usuarios no deseados utilicen la red WLAN y provoquen una degeneración del rendimiento de la red se limita el acceso a ésta usando una función presente en la mayoría de los AP's: el filtrado por dirección física (MAC). Esto tiene un inconveniente: para poder utilizar la WLAN es necesario que el usuario de alta la MAC de su tarjeta al gestor de la red, y que este la de alta en los AP's. Otra funcionalidad de los AP's que nos puede ayudar a evitar el efecto de los usuarios de baja velocidad es el fijar una velocidad mínima de asociación por debajo de la cual un usuario no puede utilizar un AP. Esto tiene como inconveniente que se limita el área de servicio a las zonas donde la calidad de la señal permite conseguir la velocidad fijada. Por lo que hay un compromiso entre ofrecer la máxima cobertura posible o limitarla para asegurar unas mejores prestaciones.

6. CONCLUSIONES

Después de todas las pruebas, medidas y cálculos se consiguió montar la red con solo tres APs. Esto fue posible gracias a la curiosa forma del edificio, que ha permitido que grandes zonas tuvieran visión directa desde un mismo AP, pudiendo así minimizar su número. Además, todas las zonas están solapadas entre sí, y por tanto no existe pérdida de conectividad en ningún momento, y finalmente, se ha garantizado que en todos los puntos requeridos inicialmente la calidad de la conexión es tal que la comunicación sea a 11 Mbps. La integración a la red de la EPSC y su explotación como servicio se ha resuelto con creación de una nueva VLAN y la introducción de un PC Linux con las funciones de cortafuegos, NAT y servidor DHCP, y la utilización de control de acceso por MAC en los APs.

El servicio se encuentra en funcionamiento desde el cuatrimestre de otoño 2002/03 con un préstamo de tarjetas y soporte técnico a las personas interesadas en utilizar la red gestionada por Servicios Técnicos de la EPSC. Además, gracias a un acuerdo con el CRSD (Centro de Recursos de Soporte a la Docencia) enmarcado dentro de un convenio con entre l'ICE (Instituto de Ciencias de la Educación) y la EPSC para llevar a cabo experiencias de innovación docente utilizando WLAN se ha conseguido incrementar de manera considerable el número de APs y tarjetas disponibles. Esto permitirá aumentar la cobertura del acceso wireless a prácticamente la totalidad de la EPSC convirtiendo cualquier aula de teoría, reuniones o trabajo en grupo en un potencial laboratorio informático.

Como líneas futuras de este proyecto nos hemos marcado dos caminos a seguir. Por un lado se pretende extender la cobertura al exterior de la EPSC y por otro desarrollar

herramientas basadas en software de código abierto y libre distribución que permitan una gestión integral de la red wireless y faciliten su utilización. Respecto a este último objetivo se está trabajando para pasar de una política de control de acceso basada en la MAC a otra basada en el login y password utilizando el software NoCatAuth.

7. BIBLIOGRAFÍA

- [1] Despliegue de una red WLAN a la EPSC. Xavier Bordoy (autor), David Remondo i Rafael Vidal (director y subdirector) Trabajo fin de carrera, EPSC, julio 2002
- [2] Explotación de una red WLAN a la EPSC. Daniel Martínez (autor) i Rafael Vidal (director) Trabajo fin de carrera, EPSC, julio 2002
- [3] Mobile Communications. Jochen Schiller. Editorial Addison-Wesley, 2000
- [4] Wireless Lans. Jim Geier. Editorial MacMillan technical publishing, 1999
- [5] Weaknesses in the Key Scheduling Algorithm of RC4. Scott Fluher, Itsik Mantin, and Adi Shamir. 8th Annual Workshop Selected Areas in Cryptography, Agosto 2001
- [6] An Empirical Characterization of Instantaneous Throughput in 802.11b WLANs A. Vasani A.U. Shankar. University of Maryland, report tècnic, CS-TR-4389, UMIACS-TR-2002-69. <<http://www.cs.umd.edu/~shankar/Papers/802-11b-profile-1.pdf>>
- [7] Wireless Firewall Gateway White Paper. Nichole K. Boscia, Derek G. Shaw. NASA Advanced Supercomputing Division. <<http://www.nas.nasa.gov/Groups/Networks/Projects/Wireless/>>
- [8] NoCatAuth. The NoCat Community Wireless Network Project. <<http://nocat.net/nocatrfc.txt>>

AUTORES



Rafael Vidal, Ingeniero de Telecomunicaciones por la ETSETB y profesor del Departamento de Ingeniería Telemática desde el año 2000, con docencia en la EPSC. Forma parte del grupo de investigación de comunicaciones inalámbricas desde el año 1998, su ámbito de trabajo es el soporte a la movilidad en redes fijas.



Xavier Bordoy, Ingeniero Técnico de Telecomunicaciones por la EPSC desde el año 2002 y becario del departamento de Ingeniería Telemática de la misma facultad. Vinculado actualmente al proyecto i2CAT, investiga sobre tecnologías multimedia para las redes de Internet 2, y más específicamente en la retransmisión de televisión de alta definición sobre IP.



QOS EN REDES WIRELESS LAN IEEE-802.11

Alex Berdonces (proyectista), Anna Calveras

{aberdonc, acalveras}@entel.upc.es

Grup de Comunicacions Mòbils Departament d'Enginyeria Telemàtica (ENTEL)
Universitat Politècnica de Catalunya (UPC) BARCELONA, ESPAÑA

Resumen - El mundo de las comunicaciones móviles ha ido evolucionando rápidamente junto con Internet y otras nuevas tecnologías que proporcionan calidad de servicio en la red. No es de extrañar pues que quiera introducirse el concepto de QoS en las comunicaciones inalámbricas, de forma que los usuarios finales móviles puedan tener prestaciones similares a los que utilizan redes fijas. En este artículo mostraremos como conseguir esta QoS en redes 802.11. Ofrecer QoS a nivel IP no es efectivo sin la modificación de la capa MAC del 802.11, por tanto veremos las diferentes propuestas que existen al respecto.

INTRODUCCIÓN

Actualmente el mundo de las comunicaciones móviles y el de Internet están sufriendo un crecimiento espectacular. Por tanto no es de extrañar que se estén integrando los dos para ofrecer al usuario final las ventajas de ambos mundos. La tecnología "wireless" está teniendo mucho empuje en el mercado. Cada vez es más común ver todo tipo de terminales que acceden a servicios de Internet de forma inalámbrica tales como *smart phones* y terminales GPRS o ordenadores portátiles y PDA's mediante WLAN o Bluetooth.

Por otro lado el mundo de Internet ya no se centra únicamente en la World Wide Web, sino que continuamente aparecen nuevas aplicaciones como la telefonía IP, videoconferencias, vídeo bajo demanda, etc. Por eso está en auge el mundo de la calidad de servicio en las redes IP, ya que las nuevas aplicaciones tienen restricciones estrictas en cuanto a retardos y ancho de banda. Es por este motivo que han surgido métodos para ofrecer QoS en IP como IntServ [11] y DiffServ [12].

IntServ es una técnica basada en la reserva de recursos que un determinado tipo de tráfico hace a lo largo de la ruta que van a seguir sus paquetes. En cambio DiffServ sigue un esquema de prioridades mediante el cual clasifica cada flujo a la entrada de la red y posteriormente le asigna los recursos adecuados a su prioridad.

Es lógico pensar pues, que haya un creciente interés en reunir estas tecnologías dentro de los estándares para redes inalámbricas. Si nos centramos en el IEEE 802.11 [2], que es el más comercializado, nos damos cuenta que ofrecer QoS a nivel IP [1] (o superior) no es posible sin hacer modificaciones en las técnicas de acceso al medio. Se ha comprobado que, en las redes inalámbricas, los servicios diferenciados de IP son subóptimos sin el soporte de las capas inferiores [8]. Por tanto el Task Group E (TGe) del IEEE 802.11 centra sus esfuerzos en modificaciones de la capa MAC que permitan proporcionar un cierto soporte para QoS a los niveles superiores, pero que a la vez no creen incompatibilidades con lo desarrollado hasta el momento.

WIRELESS LAN IEEE 802.11

En 1990 se formó el grupo de trabajo del IEEE 802.11 para la estandarización de las redes inalámbricas de área local. El 802.11 llevó a cabo un estándar global que inicialmente tenía tres especificaciones de nivel físico: una de infrarrojos y otras dos RF (Radio Frecuencia) que trabajan en la banda de los 2,4 GHz, no requieren licencias, y funcionan a una velocidad de 1 o 2 Mbps.

Actualmente tenemos velocidades mayores: con el 802.11b se llega a los 11 Mbps (es el que más se está implementando en las tarjetas comerciales) y con el 802.11a se consiguen velocidades de transmisión de 36 Mbps.

Capa MAC del 802.11

La capa MAC (Medium Access Control) [2] la encontramos sobre el nivel físico de las WLANs. Proporciona dos tipos básicos de funcionamiento: DCF (Distributed Coordination Function) y PCF (Point Coordination Function).

El DCF es de implementación obligada y es el que se utiliza en redes adhoc. Sobre éste se implementa el PCF que permite el acceso mediante infraestructura y proporciona un intervalo libre de contenciones. El PCF es opcional.

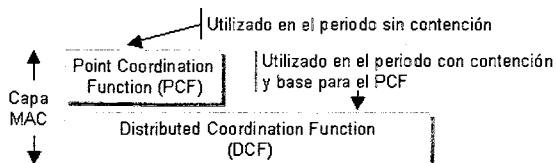


Figura 1 - Capa MAC del 802.11

DCF (Distributed Coordination Function): DCF es el mecanismo básico de acceso al medio del 802.11. En este modo de funcionamiento cada estación puede conectarse con otras que estén visibles y en su proximidad sin que haya un control centralizado (redes adhoc).

ACK tienen prioridad frente a nuevas transmisiones.

Como el canal wireless es *half duplex* (debido a la diferencia entre las potencias de la señal recibida y la transmitida) no podemos detectar colisiones. Por tanto consideramos que se ha producido una colisión cuando no se recibe el ACK del receptor. En este caso se dobla el valor de CW y se vuelve a intentar la retransmisión de forma que el nuevo backoff que se escoge tiene más posibilidades de ser diferente al de la otra estación con la que se ha colisionado. Con las sucesivas colisiones se va aumentando CW y con las retransmisiones correctas se disminuye. Tenemos pues un backoff aleatorio exponencial gracias al cual se minimizan el número de colisiones entre estaciones que previamente hayan colisionado.

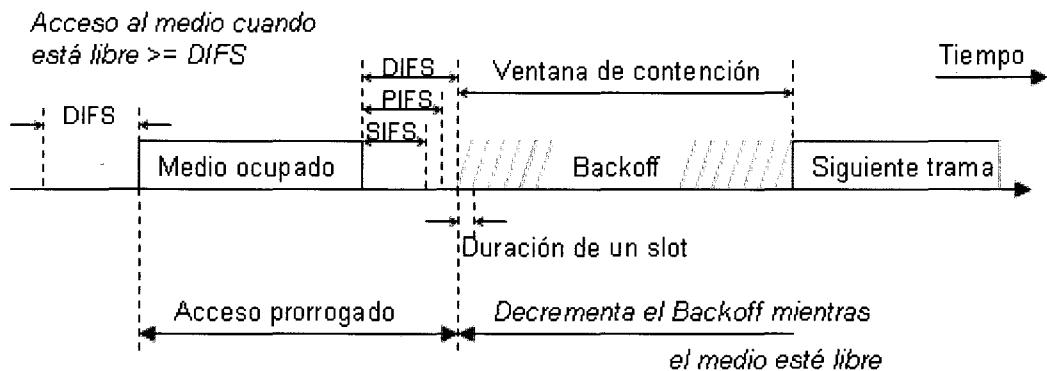


Figura 2 – Acceso al medio en DCF

El CSMA/CA en el que se basa consiste en escuchar el canal antes de la transmisión. Si el canal está libre durante un tiempo DIFS (DCF InterFrame Spacing) entonces se procede a transmitir los datos. Si el canal está ocupado se espera hasta que quede libre y después de DIFS se inicia un mecanismo de backoff aleatorio. El temporizador de backoff va descontándose mientras el canal esté vacío y se detiene cuando haya otra estación transmitiendo. En este caso, cuando el canal vuelva a estar libre, después de esperar DIFS de nuevo continuará la cuenta atrás donde se había quedado. Cuando el contador expira se pueden enviar los datos.

El valor del tiempo de espera de backoff se escoge dentro de la ventana de contención (Contention Window). Este valor estará entre $[0, CW]$. Inicialmente CW es 31 por defecto en el 802.11b.

Cada trama tiene que ser validada mediante la transmisión de un ACK por parte del receptor. Los paquetes de ACK esperan un tiempo SIFS (Short InterFrame Spacing) para acceder al canal. SIFS siempre es menor que DIFS, de esta manera los

Detección de portadora virtual: En sistemas de acceso por radio puede darse el caso de tener estaciones que no se escuchen entre ellas e interpretan que el canal está libre originando una colisión en una tercera que sí pueda escucharlas a ambas.

Para solucionarlo se utiliza un esquema adicional de RTS/CTS [2] (Request To Send / Clear To Send) que es opcional y no tiene por qué estar implementado en todas las WLAN. Los paquetes de RTS compiten por el canal esperando un tiempo DIFS como los paquetes normales. La estación emisora envía un RTS antes de transmitir para realizar la reserva y el receptor contesta con un CTS,

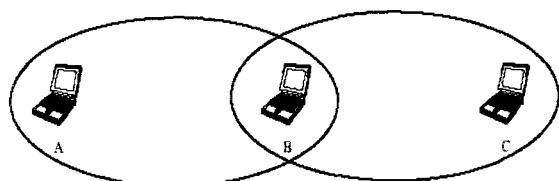


Figura 3 - Problema de estaciones ocultas



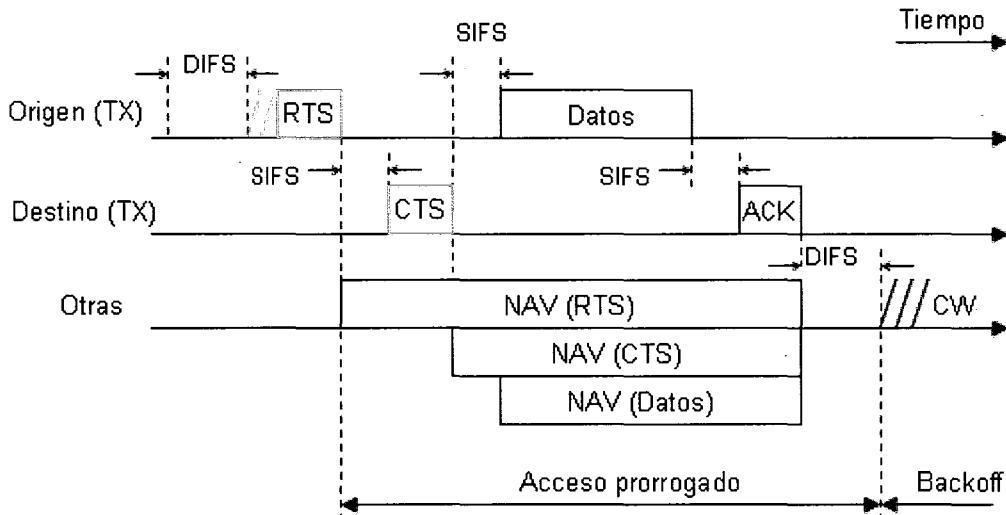


Figura 4 – Esquema de RTS/CTS

quedando reservado el canal. El receptor envía su CTS después de esperar un tiempo SIFS a que el canal esté libre para evitar que otra transmisión pueda acceder al canal antes que éste.

El resto de estaciones que hayan escuchado el RTS, el CTS o los datos, actualizan su NAV (Network Allocation Vector) con el tiempo de reserva y no compiten por el canal hasta que éste haya transcurrido. El NAV se utiliza como mecanismo de detección virtual de portadora.

Este esquema también se utiliza cuando se envían tramas muy largas ya que si no, se reduce la eficiencia de la transmisión debido a errores introducidos por el canal que hacen que tenga que retransmitirse la trama.

PCF (Point Coordination Function): PCF funciona sobre DCF en redes con infraestructura en las que un punto de coordinación (PC) centraliza el control de acceso al canal mediante el *polling* de las estaciones. Cuando se utiliza este método se divide el tiempo en supertramas con un periodo libre de contención para PCF y otro con contención en el que se usa DCF.

Una supertrama empieza con una trama de *beacon* enviada por una estación utilizando DCF, esté activo el modo PCF o no [3]. Esta trama se utiliza para sincronizar a las estaciones y como trama de gestión. El PC las genera a intervalos regulares conocidos como *Target Beacon Transition Time* (TBTT), que se anuncian en la propia trama de *beacon*. Para entrar en modo PCF el AP (que suele ser el PC) debe competir por el canal esperando un tiempo PIFS (PCF InterFrame Spacing) para acceder al

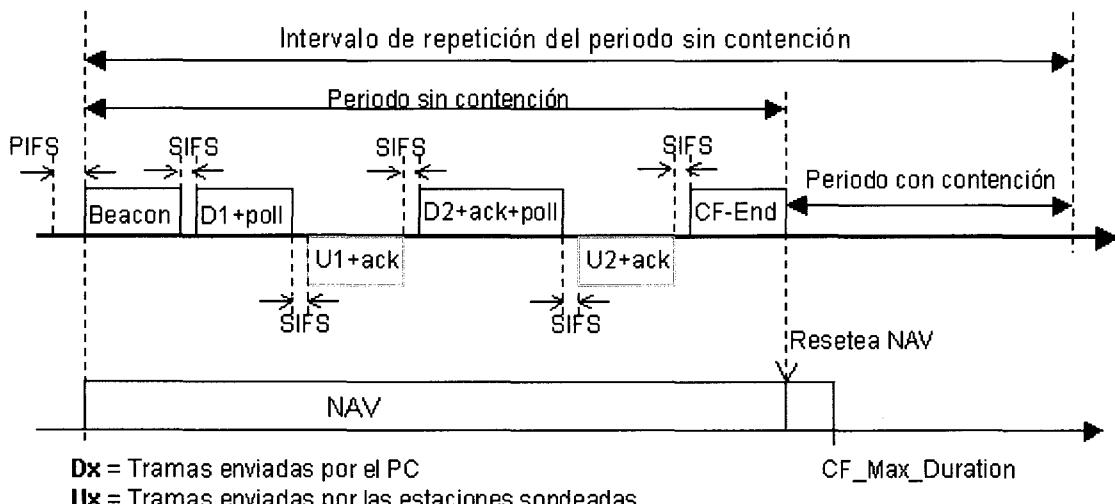


Figura 5 – Acceso al medio mediante PCF

medio. Se cumple SIFS<PIFS<DIFS. Durante el periodo libre de contención, el AP mantiene una lista de todas las estaciones sondeables y las interroga de forma predeterminada aunque no disponga de datos para ellas. Utiliza un algoritmo Round Robin para hacerlo y por eso tenemos garantizado un tiempo de latencia máximo, pero no podemos dar prioridad a las diferentes estaciones.

Todas las estaciones contestan al *poll* y envían sus datos si los tienen. Para aumentar la eficiencia se puede realizar en la misma trama el envío de datos, el *polling* y el ACK. Es lo que se conoce como *piggybacking*.

Cuando el AP gana el canal todas las estaciones actualizan su NAV al valor CFPmaxduration que contiene un valor máximo. No se entrará en el modo DCF hasta que expire el NAV o hasta que el AP envíe un paquete de final del periodo libre de contención (CF-End)

El PCF inicialmente era el método pensado para ofrecer unas ciertas garantías a los servicios con requisitos temporales, pero se ha visto que son del todo insuficientes [13]. Los principales problemas que presenta son el retardo impredecible de los *beacons* y el hecho de no conocer la duración de las tramas que enviarán las estaciones sondeadas.

Además este modo de funcionamiento no es escalable, ya que el AP debe sondear todas las estaciones que entren en el PCF, y puede ser muy ineficiente para redes grandes con muchos terminales.

QoS EN WLAN 802.11

Para poder ofrecer QoS a los niveles superiores se hace necesaria una modificación de la capa MAC. Si no se asegura un acceso al medio condicionado por las prioridades, los esfuerzos que se realicen a nivel superior serán infructuosos. Como resultado del trabajo realizado por el TGe y del análisis de las diferentes propuestas que se están haciendo, han surgido el EDCF (*Enhanced Distributed Coordination Function*) y el HCF (*Hybrid Coordination Function*), que son técnicas orientadas a DiffServ [12].

Existen otros métodos para proporcionar QoS en el 802.11 a nivel MAC [4] a parte de los del 802.11e. Propuestas como la limitación de la longitud de la trama máxima, el escalado de la ventana de contención o la diferenciación mediante IFS se basan en modificaciones de parámetros del DCF y pueden considerarse sustituidas por el EDCF. Funcionando sobre DCF también, tenemos el esquema de DENG, DFS y Blackburst. Sobre PCF encontramos el

Distributed TDM. Finalizaremos con AEDCF que es una mejora para que EDCF sea adaptativo.

A continuación veremos la idea principal de cada uno de ellos junto a las propuestas del TGe.

Limitar la longitud de trama máxima

Uno de los métodos para conseguir esta QoS consiste en limitar la longitud máxima de las tramas transmitidas por cada estación [8]. Es decir: permitiremos enviar tramas mayores a las estaciones de más prioridad. En este aspecto distinguimos dos posibilidades:

a) Descartar paquetes que excedan la longitud máxima permitida a cada estación (o configurarlo para que esto nunca pase).

b) Permitir fragmentar los paquetes que superen la longitud de trama máxima a las estaciones de mayor prioridad.

Como se ha visto anteriormente, entre los paquetes que se fragmentan no hay interrupción y por tanto no hace falta competir de nuevo por el canal, de forma que lo tenemos reservado hasta que se produzca un error debido al ruido o se acabe la transmisión. Es lo que se conoce por *packet bursting*.

La transmisión se acabará cuando no queden paquetes o cuando se haya llegado al tiempo máximo permitido a cada estación. Este tiempo máximo debe limitarse ya que puede afectar negativamente en el *jitter*.

Escalado de la ventana de contención

Otro de los métodos propuesto en [8] se basa en el tamaño de la ventana de contención para proporcionar una cierta diferenciación en los mecanismos de acceso al medio.

Asignando una ventana de contención menor a las estaciones de mayor prioridad conseguiremos una cierta calidad estadística. En la mayoría de los casos las estaciones de alta prioridad transmitirán antes que las de prioridad baja ya que escogen el tiempo de *backoff* en una ventana más pequeña. La fórmula que utiliza es:

$$\text{Nueva_CW} = 2^{i+Pj} * \text{Antigua_CW}$$

Donde Pj es el factor que controla la diferenciación e i es el número de colisiones. Cuanto mayor sea Pj, la ventana será más grande y la estación tendrá menor prioridad.



Diferenciación basada en el IFS

Una nueva forma de dar prioridades a la capa MAC mediante la modificación del tiempo de espera a que el canal quede libre (el IFS) se expone también en [8]. Se consigue prioridad asignándoles un tiempo de espera DIFS_j menor a las estaciones más prioritarias (siendo j la prioridad de cada estación). Se persigue la misma idea del mecanismo original por el que los paquetes ACK tenían prioridad respecto a los paquetes normales de datos.

Prioridad	IFS	Algoritmo de backoff
0	DIFS	$CW_b = \frac{2^{2+i}}{2} + \left[\rho * \frac{2^{2+i}}{2} \right]$
1	DIFS	$CW_a = \left[\rho * \frac{2^{2+i}}{2} \right]$
2	PIFS	$CW_b = \frac{2^{2+i}}{2} + \left[\rho * \frac{2^{2+i}}{2} \right]$
3	PIFS	$CW_a = \left[\rho * \frac{2^{2+i}}{2} \right]$

r es una variable aleatoria en el intervalo (0,1), e i es el número de colisiones

Tabla 1 – Esquema de prioridades propuesto

Cada prioridad j tiene asociado un tiempo de espera DIFS_j=DIFS. Esto significa que será DIFS más un número determinado de slots (pudiendo ser 0). Tiene que cumplirse DIFS_{j+1}<DIFS_j para que se produzca la diferenciación. Además se fija el intervalo máximo que puede añadirse a la ventana de contención de cada estación, denominado *Random Range* (RR_j). Si hacemos que RR_j=DIFS_j-1-DIFS_j entonces tenemos una diferenciación total sin solapamientos. Así conseguimos que ninguna estación de prioridad $j+1$ empiece a transmitir hasta que se hayan enviado todos los paquetes de la estación con mayor prioridad j . Pero de esta forma los tráficos de baja prioridad sufren desigualdades muy severas en el acceso al canal. Si queremos que la diferenciación sea menos estricta tan solo tenemos que hacer que RR_j>DIFS_j-1-DIFS_j.

Esquema de Deng

Deng y Chang propusieron en [6] un esquema basado en DCF que utiliza tanto IFS como el intervalo de *backoff*. Mediante IFS asigna a las estaciones de mayor prioridad un tiempo de espera PIFS y DIFS a las de menor. Vemos que con este esquema no podremos trabajar de forma correcta con PCF ya que no se garantizaría que el AP pueda tomar el

canal cuando sea necesario. Por tanto en redes con infraestructura tendremos que optar por otra técnica.

Por otro lado el algoritmo de backoff se calcula mediante las fórmulas de la Tabla 1, quedando la ventana de contención dividida en dos partes diferentes no solapadas [0,CWa] y [CWa,CWb]. Combinando ambas técnicas obtenemos un esquema de 4 prioridades de acceso al canal como el mostrado en la siguiente tabla.

DFS (Distributed Fair Scheduling)

DFS [7] utiliza el esquema de backoff del 802.11 para determinar qué estación transmite primero, pero le añade ideas de *fair queueing*. Antes de transmitir se inicia siempre un backoff, que será directamente proporcional a la longitud del paquete que desee transmitirse y a la vez inversamente proporcional a la prioridad que se haya asignado al flujo.

Las estaciones de mayor prioridad generarán backoffs más pequeños y por tanto podrán transmitir antes. Para evitar situaciones injustas en las que solamente se envíe tráfico de alta prioridad se ha incluido el tamaño del paquete en el cálculo del backoff. De esta manera los paquetes pequeños se envían más a menudo.

Si se produce una colisión se calcula el nuevo intervalo de backoff utilizando el algoritmo estándar del 802.11.

Blackburst

Otra solución pasa por utilizar la técnica de *blackburst* [5] que minimiza el retardo del tráfico en tiempo real. Para ello se envía un paquete que bloquea el canal, denominado paquete *blackburst*. La longitud de este paquete es proporcional al tiempo que una estación haya estado esperando por el canal con el método DCF normal.

Esta técnica exige unos requisitos que deben cumplir las estaciones de *mayor prioridad*:

a) Tienen que intentar acceder al medio a intervalos constantes de tiempo tsch.

b) Tienen que poder bloquear el canal durante un cierto periodo de tiempo.

Cuando estas estaciones tienen una trama lista para ser transmitida escuchan el canal y, si está libre, la envían tras esperar un tiempo PIFS. Si está ocupado esperan a que quede libre, esperan PIFS y después entran en un periodo de contención por *blackburst*. En este periodo envían el paquete de *blackburst*

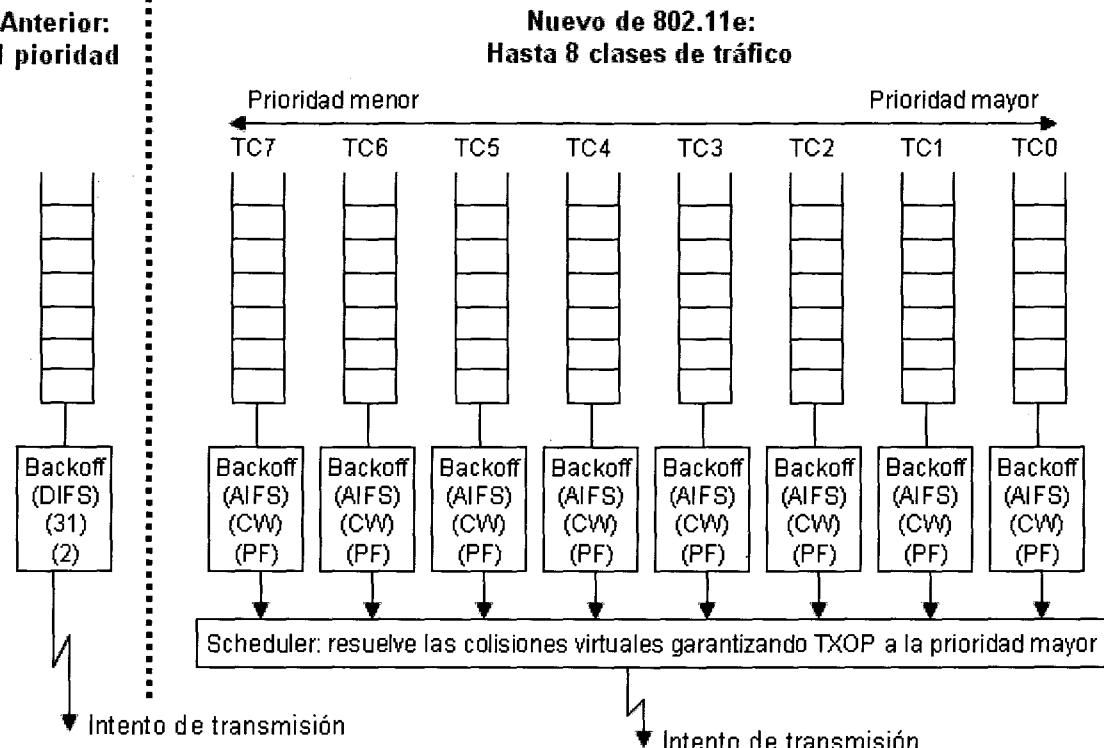


Figura 6 – Implementación interna de los nodos EDCF

para bloquear el canal y escuchan para ver que no haya otra estación que hubiera estado esperando más tiempo y por tanto haya enviado un *blackburst* más largo. Si no es así envían su trama y planifican su próxima transmisión para tsch segundos después.

Imponiendo un valor mínimo de trama se asegura que cada intervalo de contención mediante *blackburst* tenga un único ganador. Así que una vez sincronizados los tráficos de tiempo real, tenemos como un acceso TDM que no requerirá de más períodos de contención, a no ser que alguna trama de baja prioridad consiga hacerse con el acceso al medio.

TDM distribuido

En [4] se habla acerca de la propuesta de TDM distribuido. A diferencia de los anteriores, este método se propone dentro del modo de funcionamiento PCF.

Mediante esta técnica se realiza el sondeo como en PCF normal, pero este nuevo mecanismo nos permite configurar períodos de acceso como si tuviéramos TDM y especificar qué “slot” corresponde a cada estación. Una vez que cada estación sepa en qué “slot” tiene que transmitir el AP, a diferencia de

en PCF, prácticamente no tiene que intervenir más en la comunicación.

Mecanismos propuestos por el 802.11e

A continuación veremos los mecanismos que está incluyendo actualmente el TGe en el estándar [9]. Las estaciones que trabajan bajo el 802.11e se llaman *enhanced stations* y el punto coordinador, que suele estar en el AP, *Hybrid Coordinator (HC)*. Continúan existiendo dos períodos diferenciados en la supertrama, uno con contención y otro sin ella controlado por el HC. EDCF se usa solo en el CP y HCF se usa tanto en CP como en CFP, por eso se le denomina híbrido.

EDCF (Enhanced DCF): EDCF es la base para HCF e introduce el soporte de QoS, mediante las denominadas Categorías de Tráfico o *Traffic Categories (TCs)*, al esquema utilizado en el funcionamiento normal. Se puede ofrecer servicios de diferenciación de hasta ocho TC diferentes.

Una de las nuevas características más importantes del nivel MAC del 802.11e es la de las TXOP (Transmission Opportunities). Para disminuir el retardo, el *jitter* y conseguir una mayor utilización del medio, se permite el *packet bursting* mediante estas oportunidades. Una TXOP es un intervalo de tiempo durante el cual se permite iniciar la transmisión a una estación en particular. Está definida por

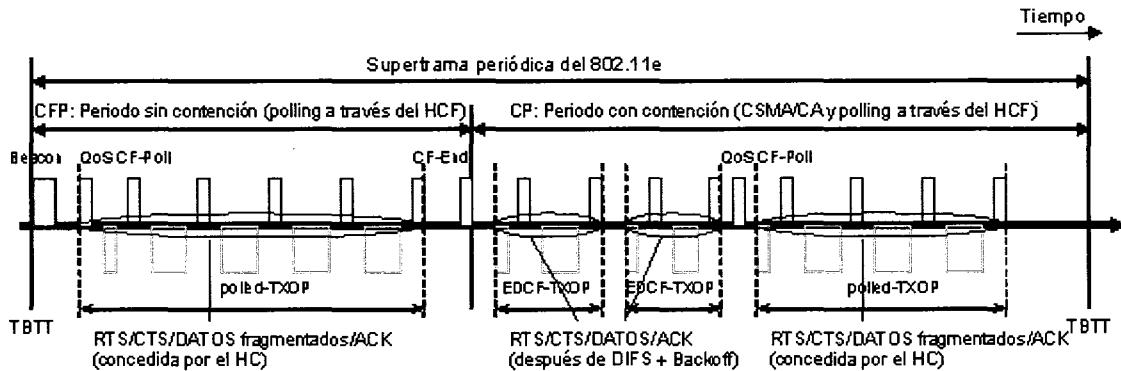


Figura 7 – Acceso al medio durante el HCF

un tiempo de inicio y una duración máxima (TXOPLimit). Las TXOP se pueden conseguir mediante contención, las denominadas *EDCF-TXOP*, o a través de HCF, denominadas entonces *polled-TXOP*, y pueden ser para una única vez o persistentes. El límite para las EDCF-TXOP se indica en las tramas de *beacon*, mientras que para las *polled-TXOP* se hace en las tramas de *poll*. Aunque estas tramas son nuevas del 802.11e, las estaciones “antiguas” también ajustan sus NAV.

Para lograr QoS, cada TC inicia un backoff aleatorio propio y compite por las TXOPs de forma independiente al resto. Para ello espera a que el canal esté libre un tiempo AIFS (*Arbitration InterFrame Space*) diferente para cada TC.

El backoff se elige asignando a cada prioridad una determinada CWMin. La diferenciación se consigue asignando valores menores de CWMin para las prioridades altas. Varios DCF's corren en paralelo con su contador de backoff independiente de forma que las colas con menor CWMin ganen el acceso al canal más fácilmente. Es posible que se produzcan colisiones virtuales (en una misma estación) cuando dos de los contadores lleguen a cero, pero en este caso el scheduler le garantiza la TXOP a la TC de mayor prioridad. El hecho de que el backoff se escoja de forma aleatoria facilita que las colas de menor prioridad también puedan transmitir algún paquete de vez en cuando pese a no estar vacías las de mayor prioridad. Por otro lado, también se limita el valor máximo al que puede llegar la ventana de contención (CWMax) de cada TC. Así las TCs de mayor prioridad tendrán valores de CWMax más pequeños.

Después de una colisión se calcula un nuevo valor para CW con la ayuda del factor de persistencia PF. Este PF puede ser diferente para cada TC en función

$$\text{Nueva_CW[TC]} \geq ((\text{Antigua_CW}+1) \cdot \text{PF[TC]}) - 1$$

de la prioridad que quiera otorgarse. Para el caso de DCF normal, PF era 2.

Mediante AIFS=DIFS y CWMin[TC]<31 (valor por defecto en 802.11b) se consigue prioridad sobre las estaciones que funcionan con el DCF normal. El problema es que AIFS tiene que ser mayor o igual que DIFS, y esto complica la interoperabilidad con estaciones antiguas.

EDCF proporciona una mejora en el servicio a las clases de mayor prioridad, mientras que ofrece un servicio mínimo a las de baja prioridad. El problema es que los parámetros no se adaptan a las condiciones de la red. Por tanto, como cada TC está implementada como una estación virtual, la tasa de colisiones crece cuando las contenciones para acceder al medio son elevadas y se pierde así parte de la mejora inicial.

Se puede observar que el tráfico con comportamiento a ráfagas se ubica mejor en el EDCF, mientras que el que tenga altos requerimientos temporales o de jitter tendrá mejor trato bajo el HCF que se explica a continuación.

HCF (Hybrid Coordination Function): HCF dispone de nuevos servicios y formatos de trama para ofrecer QoS extremo a extremo. HCF provee mecanismos para transferencias libres de contención y con contención controlada en cualquier instante (CFP o CP) permitiendo que el HC genere ráfagas de CFPs en lugar de un único intervalo seguido (como ocurría con PCF). Por tanto se pueden crear un conjunto de “mini-CFPs” durante el CP para ajustarse a las expectativas del tráfico. Esto beneficia a las TCs que necesiten intervalos periódicos de transmisión, ya que podrán extenderse durante el CP.

Siempre que el HC lo requiera puede conceder TXOP después de detectar el canal libre durante PIFS (esto le da prioridad a la hora de acceder al canal). Durante el CP cada TXOP empieza cuando el medio está disponible bajo los criterios de EDCF, o cuando una estación recibe una trama especial de

poll (QoS CF-Poll) del HC. Durante el CFP solamente el HC puede asignar esas TXOP entre las estaciones utilizando las mismas tramas *QoS CF-Poll*.

Como parte del 802.11e se ha definido un protocolo adicional de acceso aleatorio que permite la resolución rápida de colisiones: la contención controlada. El HC va sondeando a las estaciones que tengan MSDUs disponibles para saber qué estaciones necesitan ser sondeadas, en qué instante y durante cuánto tiempo.

Este mecanismo permite a las estaciones reservar las TXOPs enviando peticiones de recursos sin tener que competir por el canal con el resto del tráfico. Esto ocurre durante el intervalo de contención controlada que inicia el HC con una trama especial de control, forzando a las estaciones normales a actualizar su NAV para que no participen durante este tiempo. Esta trama define un número de oportunidades de contención controlada (intervalos cortos separados por SIFS) y un filtro de máscara indicando las TCs permitidas. Las estaciones con datos que pertenezcan a esas TCs enviarán sus tramas de petición de recursos indicando su TC específica y la duración de la TXOP, o el tamaño de la cola de la TC requerida.

Para resolver las colisiones más rápidamente, el HC confirma la recepción de la petición generando una trama de control con un campo de *feedback* para que las estaciones que están solicitando sus TXOP puedan detectar colisiones durante el periodo de contención controlada.

Por otra parte las estaciones sin QoS pueden competir por el canal en los CP y pueden ser interrogadas por el HC en los CFP, del modo que lo hacían hasta ahora.

AEDCF (Adaptive EDCF)

Recientemente también han aparecido propuestas que se basan en el EDCF del 802.11e, pero que intentan que la elección de los parámetros que condicionan el grado de calidad de servicio ofrecido pueda ser fijado de forma adaptativa [10]. En concreto se adapta el parámetro del tamaño de la ventana de contención CWMmin. De este modo si tenemos un conocimiento general del estado de la red podemos reaccionar de forma adecuada frente a diferentes condiciones de carga. Los estudios realizados mediante simulaciones en el mismo artículo [10] muestran su buen comportamiento.

CONCLUSIONES

Hemos visto que la actual demanda de determinadas aplicaciones (como voz sobre IP, video,...) con requerimientos de ancho de banda garantizado, retraso controlado,... hace necesario ofrecer una ca-

lidad de servicio. Para aplicar estos requisitos en entornos de redes 802.11 necesitamos ciertas modificaciones en la capa MAC.

En un inicio se pensó en proporcionarla con PCF pero no garantiza ningún tipo de reserva y las garantías temporales que ofrece son del todo insuficientes debido a factores como el retraso impredecible de los *beacons* y el desconocimiento de la duración de las tramas que enviarán las estaciones sondeadas. Por eso es un campo de gran investigación en los últimos años y en el que han surgido diferentes propuestas.

El TGe del 802.11 está trabajando actualmente en ello pero todavía no hay un estándar cerrado. Sus propuestas son el EDCF y HCF que proporcionan una diferenciación en base a las TC's. EDCF tiene un buen comportamiento con el tráfico a ráfagas y es la base para el HCF. El tráfico que tenga requisitos en cuanto a retardos o jitter tendrá mejor trato bajo el HCF ya que puede garantizar las TXOP tanto durante el periodo libre de contención, como durante el periodo en que se trabaja bajo EDCF.

Propuestas como las de DENG, escalado de la ventana de contención, la diferenciación mediante IFS y el *packet bursting* quedan englobadas y redefinidas dentro del EDCF. Blackburst requiere que las estaciones de alta prioridad puedan transmitir a intervalos regulares, si no, se degrada mucho su funcionamiento. Distributed TDM está indicado para voz pero no tiene un buen comportamiento para datos. DFS soluciona problemas de *fairness* en el reparto del throughput entre estaciones, pero no queda claro lo que pasa con el retraso.

Gracias a los dos nuevos métodos de acceso al medio del TGe podremos ofrecer QoS y seguir permitiendo tráfico *Best Effort* en las WLAN. Pero uno de los problemas no resueltos todavía es qué hacer con las estaciones que no implementen estos métodos, ya que tendrán prioridad en EDCF debido a los parámetros que utilizan por defecto. La propuesta es encaminar todo su tráfico al AP y que éste se encargue de gestionarlo como *Best Effort*.

AGRADECIMIENTOS

Parte de este trabajo ha estado financiado por el proyecto TIC 2000-1041-C03-01.

REFERENCIAS

- [1] Grenville Armitage, "Quality of Service in IP Networks". MTP, Abril 2000.
- [2] IEEE 802.11, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications". IEEE.



- [3] Stefan Mangold et al., "IEEE 80.11e Wireless LAN for Quality of Service", European Wireless 2002 Proceedings, 26-28 de Febrero 2002.
- [4] Yasir Drabu, "A survey of QoS techniques in 802.11", Febrero 2002. Disponible en: <http://trident.mcs.kent.edu/~ydrabu/research/wmac/mac.pdf>
- [5] J.L. Sobrinho et al., "Real-time traffic over the IEEE 802.11 medium access control layer", Bell Labs Technical Journal, pages 172-187, Otoño 1996.
- [6] Dr-Jiunn DENG et al., "A priority Scheme for IEEE 802.11 DCF Access Method", IEICE TRANS. COMMUN., VOL: E82-B, NO.1 Enero 1999.
- [7] Nitin H. Vaiday et al., "Distributed Fair Scheduling in a Wireless LAN", Sixth Annual International Conference on Mobile Computing and Networking, Boston, Agosto 2000.
- [8] Imad Aad et al., "Differentiation mechanism for IEEE 802.11", IEEE Infocom 2001, Anchorage - Alaska, 22-26 de Abril, 2001.
- [9] IEEE 802.11 WG, "Draft Supplement to International Standard [for] Information Technology-Telecommunications and Information Exchange between systems LAN/MAN Specific Requirements." IEEE 802.11e/D2.0, Nov. 2001
- [10] Lamia Romdhani et al., "AEDCF: enhanced service differentiation for IEEE 802.11 wireless ad-hoc networks", INRIA Research Report No. 4544, Sept. 2002.
- [11] R. Braden et al., "Integrated Services in the Internet Architecture: an Overview", RFC-1633, Jun. 1994.
- [12] S. Blake et al., "An Architecture for Differentiated Service", RFC-2475, Dic. 1998.
- [13] Anders Lindgren et al., "Evaluation of Quality of Service Schemes for IEEE 802.11 Wireless LANs", en "Proceedings of the 26th Annual IEEE Conference on Local Computer Networks (LCN 2001)", 15-16 Nov. 2001 Tampa, Florida, USA.

LISTA DE ACRÓNIMOS

AEDCF - Adaptative EDCF
 AIFS - Arbitration Inter Frame Space
 AP - Access Point
 CA - Collision Avoidance
 CFP - Contention Free Period
 CF-End - Contention Free - End
 CF-Poll - Contention Free - Poll
 CP - Contention Period
 CSMA - Carrier Sense Multiple Access
 CTS - Clear To Send
 CW - Contention Window
 CWMax - Contention Window Maximum
 CWMin - Contention Window Minimum
 DCF - Distributed Coordination Function
 DFS - Distributed Fair Scheduling
 DiffServ - Differentiated Services
 DIFS - DCF Inter Frame Space

EDCF - Enhanced DCF
 HC - Hybrid Coordinator
 HCF - Hybrid Coordination Function
 IEEE - Institute of Electrical and Electronics Engineers
 IFS - Inter Frame Space
 IntServ - Integrated Services
 MAC - Medium Access Control
 MSDU - MAC Service Data Unit
 NAV - Network Allocation Vector
 PC - Point Coordinator
 PCF - Point Coordination Function
 PF - Persistence Factor
 PIIS - PCF Inter Frame Space
 QoS - Quality of Service
 RR - Random Range
 RTS - Request To Send
 SIFS - Short Inter Frame Space
 TBT - TTTarget Beacon Transmision Time
 TC - Traffic Category
 TDM - Time Division Multiplex
 TGe - Task Group e
 TXOP - Transmission Opportunity
 WLAN - Wireless Local Area Network

BIOGRAFÍAS



Alex Berdonces Fernandez. Estudiante de Ingeniería Superior de Telecomunicaciones, especialidad en Telemática en la ETSETB (UPC). 2 años como becario en el C.M.U. Ramon Llull (U.B.) en la Comisión de Informática. Beca de colaboración en la asignatura de Laboratorio de Telemática 3 durante 2001-2002. Actualmente realizando el proyecto final de carrera.



Anna Calveras-Augé. Dr. Ingeniero de Telecomunicaciones por la Universidad Politécnica de Cataluña (UPC). Actualmente es profesor Titular de Universidad en dicha universidad, en el Departamento de Ingeniería Telemática. Es experta en optimización de los protocolos de Internet en entornos de red heterogéneos, especialmente el protocolo TCP. Ha estado involucrada en diversos proyectos. Actualmente está trabajando en el proyecto europeo ELIN (The Electronic Newspaper Initiative), el proyecto europeo RIU253 (Recommendations for Internet Usage over 2.5 and 3G networks), y el proyecto español Estudio de las Alternativas para ofrecer servicios multimedia con calidad de servicio extremo a extremo en una red IP integrada con servicios móviles (CICYT TIC2000-1041-C03-01). Ha publicado diversos trabajos en conferencias nacionales e internacionales.



CRPTOGRAFÍA CLÁSICA.

¿CÓMO ROMPER CIFRADOS MONOALFABÉTICOS Y POLIALFABÉTICOS?

ANÁLISIS DE FRECUENCIAS Y MÉTODO KASISKI.

Marc García Arnau

Estudiante de Telecom Paris (ENST) y de la Facultad de Informática de la UPM.

¿QUÉ ES LA CRPTOGRAFÍA? ¿CUÁNDO SURGIÓ?

Es fascinante ver como, a lo largo de los tiempos, el hombre ha ido progresando en el arte de controlar los secretos, y como su empeño por dominar la información, ha ido enmarañando los mecanismos para garantizar la confidencialidad la de la misma. La información y su conocimiento son consustanciales al hombre y a la vida y, a veces, pueden tornarse comprometidas y por ello, adquirir un gran valor.

La criptografía nació, entonces, de la necesidad de salvaguardar la confidencialidad de la información. En realidad, la propia raíz etimológica de la palabra criptografía nos da una idea de su utilidad. Del griego (*kryptos*), «oculto» y (*grapto*), «escrito», actualmente su definición podría perfectamente ser la de “arte de cifrar mensajes” o la de “ciencia que estudia los procesos de cifrado y descifrado de los mensajes”. El resultado inmediato de cifrar un texto o documento es lo que conocemos por criptograma, y el criptoanálisis, la ciencia, proceso, o arte, encargado del análisis de dichos criptogramas para descubrir su clave o su texto original.

Debemos remontarnos al antiguo Egipto y a Mesopotamia, es decir, a los orígenes de la civilización humana, para encontrar los primeros indicios de protocriptografía. Allí se dan ciertos hechos ligados a las escrituras jeroglífica y cuneiforme que pretendían transformar deliberadamente la escritura. A partir de entonces, y a lo largo de los distintos períodos de la historia, se guardan referencias de curiosos e ingeniosos métodos de comunicación secreta con fines, habitualmente, militares o políticos. Como es el caso de la antigua China, donde se usaron métodos más bien esteganográficos (ocultación de la información), para mantener la privacidad de la misma. Así, se enviaban mensajeros que memorizaban los mensajes, o bien éstos se escribían en papel o seda y, tras cubrirllos con una bola de cera, se invitaba a los siempre dispuestos mensajeros a esconderlos en alguna parte de su cuerpo (...). O como cuando Histiaeus envió un mensaje desde la corte persa a su yerno el tirano Aristágoras de Mileto (Grecia), para que se sublevara contra el emperador Ciro de Persia antes de que éste les atacase. Para ello, afeitó la cabeza de un siervo leal y le tatuó en ella un mensaje. Eso sí, tuvo que esperar a que le creciera el pelo antes de dejarlo partir hacia Mileto.

En el siglo V a.C., durante las guerras entre las polis griegas de Esparta y Atenas, ya se usaron ciertos dispositivos de cifrado. Como es el caso de la “*escítala de los Lacedemonios*”, que podría clasificarse dentro de los llamados métodos de transposición. Consistía en una cinta de papiro enrollada en un cilindro o bastón (cuyo diámetro determina la clave), sobre la cual escribían el texto en claro horizontalmente. Al desenrollar la cinta del bastón, las letras aparecían permutadas y constituyan un mensaje cifrado que, posteriormente, era enviado al receptor. Éste, que disponía de una copia idéntica del bastón (es decir, conocía la clave) volvía a colocar la cinta y era capaz de leer el mensaje en claro.

EL CIFRADO DE CESAR Y SU CRPTOANÁLISIS.

Por supuesto, no podíamos olvidar el primer de los métodos de sustitución monoalfabética, el conocido método de Cesar. Este método fue el empleado por Julio Cesar en sus campañas durante el siglo I a.C. para transmitir información en secreto. Era, como decimos, una sustitución que consistía en cifrar un mensaje empleando un alfabeto equivalente al original, pero desplazado en 3 letras. Es decir, al cifrar por ejemplo las palabras *GALOS IRREDUCTIBLES* con este algoritmo, Cesar obtendría:

JDÑRV LUUHGXFWLEÑHV

El primer criptoanálisis aplicable sería el método de complementación al componente original, es decir, probar las 27 permutaciones posibles obtenidas como resultado de desplazar una posición cada vez las letras de una palabra del criptograma. Por ejemplo, para *JDÑRV*, analizando los 27 resultados es más que probable que encontrremos una única palabra que tenga significado en español, con lo cual solo tendríamos que utilizar el mismo desplazamiento obtenido para descifrar el resto del texto.

El segundo criptoanálisis aplicable a este tipo de cifrado es el análisis de frecuencias. De una forma un tanto más elegante que el anterior, mediante este método, estudiamos la frecuencia relativa de aparición de las diferentes letras del texto cifrado, para compararlas con una descripción estadística que hayamos obtenido del lenguaje, en el que sospechamos o sepamos que se encuentra el mensaje original. También se puede realizar un estudio sobre las

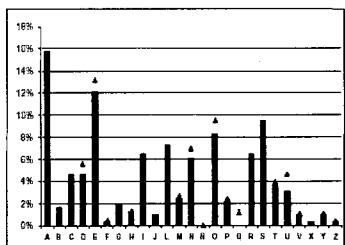


palabras más usadas, o sobre los digramas y trigramas que constituyen el inicio y terminación más frecuente de las palabras de un lenguaje.

Por ejemplo, para un texto en español lo suficientemente representativo, se ha obtenido la siguiente distribución de frecuencias:

A	52	15,76%
B	5	1,52%
C	15	4,55%
D	15	4,55%
E	40	12,12%
F	1	0,30%
G	6	1,82%
H	4	1,21%
I	21	6,36%
J	3	0,91%
L	24	7,27%
M	8	2,42%
N	20	6,06%
N	0	0,00%
O	27	8,18%
P	7	2,12%
Q	0	0,00%
R	21	6,36%
S	31	9,39%
T	12	3,64%
U	10	3,03%
V	3	0,91%
X	1	0,30%
Y	3	0,91%
Z	1	0,30%

Figura 1.- Análisis de Frecuencias



Figuras 1 y 2.- Análisis de frecuencias del criptograma

Este tipo de criptoanálisis es terriblemente efectivo para cifrado monoalfabético. Al emplearlo, se pone en evidencia la principal vulnerabilidad de este método pues, el hecho de sustituir unas letras por otras siguiendo siempre la misma congruencia lineal, hace que las propiedades estadísticas del criptograma y del texto en claro sean exactamente las mismas. Simplemente hay que llevar a cabo un análisis estadístico de los símbolos del criptograma e intentar solaparlo o encajarlo con la distribución de los símbolos de nuestro idioma. De esta forma, hallaremos el desplazamiento que fue aplicado al cifrar el texto original y podremos descifrar, inmediatamente, el resto del mensaje. Por ejemplo, si interceptamos el siguiente mensaje cifrado:

DQR FLPFXHPWD DPWHV GH FULVWR WRGD ND JDÑLD HVWD RXSDGA OHPRV XPD SHTXHQD DÑGHD GH LUUHGXFWEÑHV JDÑRV TXH UHVLVWH DKRUD B VLHOSUH DÑ LPYDVRU

El análisis de frecuencias de los símbolos de este criptograma es:

1	1	_	18	1	5	6	16	_	2	18	_	6	2	6	2	8	3	2	7	10	7	6	1	
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	XYZ

Tabla 1.- Frecuencia de aparición de cada símbolo en el criptograma

Si se observa atentamente esta tabla, que refleja el número de apariciones de cada letra en el texto cifrado, y se compara con el gráfico de frecuencias, se puede hallar rápidamente un encaje con otro alfabeto equivalente desplazado. Únicamente hay que tratar de hacer concordar

las letras más frecuentes en castellano con las más frecuentes en el criptograma. Así, sabiendo que una de las letras más utilizadas en español es la *a*, se le puede hacer corresponder la *D*, que es la más frecuente en nuestro texto cifrado. Cuatro posiciones a la derecha de la *D*, encontramos la *H*, también con un altísimo número de apariciones. Se puede entonces pensar que dicha *H* puede ser la sustituta de la letra *e*, también muy frecuente en nuestro idioma. En efecto, en el alfabeto español la *a* y la *e* distan exactamente cuatro lugares, con lo que se intuye que vamos por el buen camino. Se puede corroborar nuestra hipótesis procediendo de la misma manera con las letras menos frecuentes del lenguaje. Finalmente, concluimos que se trata de un cifrado monoalfabético con la siguiente correspondencia entre letras:

Criptograma	Frecuencia	Texto claro
A	1	X
B	1	Y
C	-	Z
D	18	A
E	1	B
F	5	C
G	6	D
H	16	E
I	-	F
J	2	G
K	1	H
L	8	I
M	-	J
N	-	K
Ñ	6	L
O	2	M
P	6	N
Q	2	Ñ
R	8	O
S	3	P
T	2	Q
U	7	R
V	10	S
W	7	T
X	6	U
Y	1	V
Z	-	W

Tabla 2.- Correspondencia entre las letras del criptograma y las del texto en claro

Es decir, el desplazamiento de este cifrado monoalfabético es 3. Para comprobarlo, aplicamos este cambio al mensaje cifrado y obtenemos:

AÑO CINCUENTA ANTES DE CRISTO TODA LA GALIA ESTA OCUPADA MENOS UNA PEQUEÑA ALDEA DE IRREDUCTIBLES GALOS QUE RESISTE AHORA Y SIEMPRE AL INVASOR

MÉTODO KASISKI. CRIPTOANÁLISIS DE CIFRADOS POLIALFABÉTICOS.

La sustitución polialfabética es una generalización de los sistemas de sustitución monoalfabeto. Este tipo de sustitución consiste en cifrar empleando una clave compuesta, es decir, de dos símbolos o más, que se usa cíclicamente.

Un buen ejemplo de cifrado polialfabético es el cifrado de Vigenére, que se sirve de una tabla para facilitar las operaciones de cifrado y descifrado. Es interesante resaltar el hecho de que cada una de las filas de esta tabla no son más que un cifrado de Cesar. La primera tiene un desplazamiento de 0, la segunda de 1, y así sucesivamente:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z		
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z			
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z				
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z					
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z						
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z							
I	I	J	K	L	M	N	O	P	Q	R	S	U	V	W	X	Y	Z									
J	J	K	L	M	N	O	P	Q	R	S	U	V	W	X	Y	Z										
K	K	L	M	N	O	P	Q	R	S	U	V	W	X	Y	Z											
L	L	M	N	O	P	Q	R	S	U	V	W	X	Y	Z												
M	M	N	O	P	Q	R	S	U	V	W	X	Y	Z													
N	N	O	P	Q	R	S	U	V	W	X	Y	Z														
O	O	P	Q	R	S	U	V	W	X	Y	Z															
P	P	Q	R	S	U	V	W	X	Y	Z																
Q	Q	R	S	U	V	W	X	Y	Z																	
R	R	S	U	V	W	X	Y	Z																		
S	S	U	V	W	X	Y	Z																			
T	T	U	V	W	X	Y	Z																			
U	U	V	W	X	Y	Z																				
V	V	W	X	Y	Z																					
W	W	X	Y	Z																						
X	X	Y	Z																							
Y	Y	Z																								
Z	Z																									

Figura 3.- Tabla de Vigenere

A diferencia de los códigos monoalfabéticos, los polialfabéticos no conservan la misma distribución de frecuencias del texto original. Son más próximos a un cifrador ideal (como el de Vernam) ya que su distribución de símbolos se acerca más a una Uniforme. Sin embargo, pese a ser más segura que la monoalfabética, la sustitución polialfabética no es inmune al criptoanálisis.

Efectivamente, un oficial prusiano llamado Kasiski (1805 - 1881), elaboró un método para hallar el número de alfabetos en una sustitución polialfabética. El método se basa en la idea de que en todos los idiomas aparecen grupos de caracteres con más frecuencia que otros. Es decir, por ejemplo en castellano, existen ciertos digramas y trigramas que tienen mucha más probabilidad de darse que otros, en un texto, como son: es, de, os, en, la, con, etc. Si un texto se cifra con un número x de alfabetos de forma cíclica y, si un grupo de caracteres aparece un número y de veces en un texto, éste será cifrado aproximadamente y/n veces con el mismo alfabeto. En resumen, en un criptograma lo suficientemente extenso encontraremos, ineludiblemente, repeticiones. Por ejemplo, supongamos que ciframos el siguiente alfabeto con la clave *GIN* (3 alfabetos):

D	E	S	C	ON	F	I	A	N	Z	A	C	O	N	E	L	...	
G	I	N	G	I	N	G	I	N	G	I	N	G	I	N	G	I	...
J	M	F	I	W	Z	L	P	N	S	H	N	I	W	Z	K	T	...

Se puede ver como, separada por 9 espacios, aparece repetida la cadena **IWZ**. Así pues, se puede ya deducir que

los posibles períodos de nuestra clave serán 1, 3 ó 9. Hay que decidirse por una de las alternativas, así que escogemos el 3 (la correcta, por otro lado) para avanzar con el método. A continuación, lo que se debe hacer es reordenar el texto cifrado en un número de columnas igual al período de clave supuesto correcto. Esta es la estrategia ganadora pues, tras hacer esto, el problema podrá ser tratado por columnas independientes. Cada una de esas columnas será

J	M	F
I	W	Z
L	P	N
S	H	N
I	W	Z
K	T	...

un simple cifrado monoalfabético, de criptoanálisis simple, como ya se ha visto en el apartado anterior.

El problema de descifrar un criptograma que presentaba una distribución de frecuencias considerablemente uniforme ha quedado reducido a tres análisis frecuenciales independientes, uno para cada columna.

Es interesante resaltar que, algunas veces, pueden aparecer repeticiones que son fruto de la casualidad. Pero juega a nuestro favor el hecho de que estas repeticiones casuales están distribuidas al azar, mientras que las repeticiones útiles a nuestra causa se dan tenazmente según la misma pauta.

Paradójicamente, el documento de 95 páginas que publicó Frederick W. Kasiski no causó ningún interés en su época y decidió abandonar el criptoanálisis para dedicarse a la antropología.

REFERENCIAS

- [1] <http://starbase.cs.trincoll.edu/~crypto/historical/vigenere.html>
- [2] Seguridad y Protección de la Información. J o s é Luís Morant Ramon,... ED. Centro de Estudios Ramón Areces, S.A.
- [3] <http://rinconquevedo.iespana.es/rinconquevedo/Criptografia/introduccion.htm>



AUTOR

García Arnau, Marc. Ingenierio en Informática por la UPM. Actualmente estudia un máster en sistemas informáticos en la École Supérieure Nacional de Télécommunications.

