

CLASIFICACIÓN AUTOMÁTICA DE FUENTES DE RUIDO DE TRÁFICO

Manuel A. Sobreira Seoane¹, Alfonso Rodríguez Molares²

Sonitum®, E.T.S.I. de Telecomunicación, Universidad de Vigo. Rúa Maxwell s/n, Campus de la Universidad de Vigo. 36310 Vigo-España..

¹msobre@gts.tsc.uvigo.es; ²amolares@gts.tsc.uvigo.es

Resumen

La realización de mapas de ruido de tráfico requiere conocer la Intensidad media diaria, su velocidad media y la composición (porcentaje de vehículos pesados y ligeros). Así mismo, la última revisión de la norma internacional ISO 1996-2:2007 “Determination of Environmental Noise Levels” en el apartado 6.2 especifica que cuando se procede a la medición del nivel equivalente continuo emitido por una carretera se debe proceder a la clasificación de tráfico en al menos dos clases: “ligeros” y “pesados”. Muy a menudo, en calles y carreteras secundarias se carece de esta información, a pesar que algunas de ellas el volumen de tráfico es importante. En estas situaciones se suele recurrir a un muestreo, recurriendo a conteos manuales de tráfico. En este artículo se presenta una herramienta que a partir de la señal de audio del ruido al paso, permite de forma automática contar y clasificar el tráfico en tres clases: vehículos pesados, vehículos ligeros y motocicletas. En el artículo se presentarán los resultados obtenidos mediante la utilización de distintas técnicas de clasificación (esencialmente FLD, y k-NN).

Palabras-clave: reconocimiento de patrones, discriminador lineal de Fischer, k-NN.

Abstract

The last review international standard ISO 1996-2:2007, Determination of Environmental Noise Levels” [1], in its 6.2 section states that if the Leq of road traffic is measured and the results are going to be used to calculate to other traffic conditions, the classification of at least two categories of vehicles: “light” and “heavy” should be registered. This information is also needed to calculate a Noise Map. Quite often, in secondary roads and streets there is no information on the traffic characteristics, so a manual sampling is used to get the average number of vehicles. In this paper, a first approach to get an automatic classification of vehicles is presented. Some basic classifiers have been tested (k-nearest neighbours, FLD (Fischer Linear Discriminator) and Principal Components. As first approach, the aim of the job was to determine if the different classes (trucks, cars and motorbikes) could be separable using different time and frequency characteristics. The results shows that for some of the characteristics the signals are separable, so a continuous traffic noise signal could be processed to get the information of the number of heavy trucks, cars and motorbikes that passed by during the measurement period. Information of a stereo recording could be used to get information of the direction of the vehicle.

Keywords: pattern recognition, FLD, k-NN.

1 Introducción

Las características en tiempo y frecuencia de una señal, proporcionan conjuntamente información relevante sobre la fuente sonora, de tal forma que podríamos afirmar que el sonido emitido contiene la firma individual del foco que lo ha emitido. Si las características temporales y espectrales se eligen apropiadamente, podríamos afirmar que esta firma es única, permitiendo identificar el foco entre otros focos emisores interferentes. A modo de ejemplo, si se reproduce de forma inversa el sonido de un piano, de forma tal que el tiempo de caída de la señal pase a ser el tiempo de ataque en la señal reproducida, uno no podría afirmar que el foco emisor de la nota escuchada es un piano, si se considera como característica discriminante únicamente la frecuencia fundamental. Aunque el contenido espectral de la señal es el mismo, la envolvente temporal ha variado de tal forma que se parece al sonido emitido por un instrumento de cuerda “frotada”. Ambas características, temporales y espectrales, son muy importantes a la hora de distinguir y clasificar diferentes fuentes de señal. A priori, se puede esperar buenos resultados cuando un conjunto adecuadamente seleccionado de características actúan conjuntamente.

Si la complejidad del problema aumenta (clasificación de fuentes del mismo tipo), el número de características que deben ser consideradas va a crecer. El ruido emitido por un coche con motor diesel y el de un camión ligero no son muy diferentes y a pesar de ello una persona puede distinguir el paso de distintos tipos de vehículos. Las características en tiempo y frecuencia que permiten a una persona identificar un vehículo de otro, deben entonces ser utilizadas para establecer los conjuntos (clases) que permitirán a un clasificador detectar qué tipo de vehículo ha pasado en determinado momento. Está claro que el resultado final, la probabilidad de detección correcta de un determinado tipo de vehículo, está fuertemente condicionado a la correcta selección de una combinación adecuada de un conjunto de características y un método de clasificación. Con alguna experiencia y conocimiento en técnicas de clasificación, algunos métodos van a ser automáticamente descartados, mientras que otros van a ser candidatos. El resultado óptimo se obtiene después de una fase de “pruebas y errores”, donde los se prueban los distintos métodos candidatos con diferentes conjuntos de características.

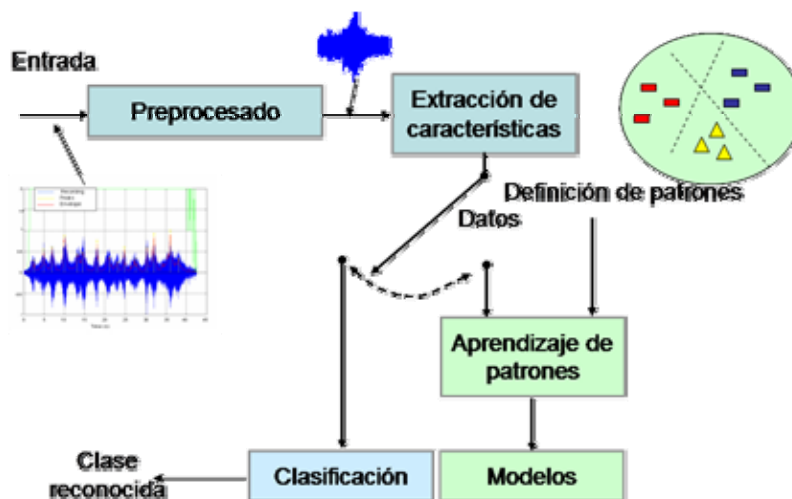


Figura 1 – Esquema de funcionamiento de un clasificador automático

El proceso de clasificación en general, y el de fuentes de ruido en particular, incluye varias etapas: en primer lugar, el sonido debe ser preprocesado (eliminación de ruidos interferentes y supresión de la señal continua en sucesos simples). Una vez se ha preprocesado, se procede a extraer el conjunto de características que situará al evento en un punto del espacio n-dimensional de clasificación. El suceso es representado por un vector de n dimensiones que contiene la información de las n características seleccionadas para representare la firma de la fuente. Este vector de características es enviado al algoritmo de clasificación que detectará la clase a la que pertenece el evento. En una etapa previa, la fase de entrenamiento, se procede a definir las clases y a entrenar el modelo con un conjunto de señales conocidas. La figura 1 muestra la estructura básica de un sistema de clasificación.

La identificación de la firma de fuentes de ruido en general y la clasificación automática de tráfico en particular han sido muy poco estudiadas en la literatura, comparado con los estudios de reconocimiento de voz o música, o clasificación de géneros de música. La teoría sobre extracción de características y técnicas de clasificación puede encontrarse en la literatura básica común sobre el tema [1,2,3].

Para el desarrollo de este trabajo se ha grabado una base de datos de 300 señales, 100 para cada una de las clases consideradas: coches, motocicletas y camiones. Para este propósito se seleccionó una carretera llana, ver figura 2, con una intensidad media de tráfico. Las grabaciones con alto ruido de fondo o viento se rechazaron a priori y como primera aproximación únicamente se ha tratado con pasos simples (un único vehículo) dejando la detección y clasificación del paso de varios vehículos simultáneamente para etapas futuras del trabajo. Se han grabado señales estéreo (con dos micrófonos), en formato PCM, 48 kbps y 16 bits por muestra. La grabación estéreo permite, además de la detección del paso y la identificación del vehículo, la estimación de la velocidad y sentido del paso de vehículo.



Figura 2 – Emplazamiento seleccionado para la grabación de la base de datos.

2 Detección de vehículos

En esta sección se proporciona una breve descripción de la técnica adoptada para la detección de vehículos. Esta es la etapa crítica del sistema, cuyo papel es detectar si un vehículo ha pasado y enviar el segmento de señal al bloque de extracción de características. El detector del paso de vehículos únicamente especifica la presencia de ruido de tráfico, extrayendo la señal de tráfico del ruido de

fondo. La señal de tráfico puede proceder de un único vehículo o de una combinación de vehículos (paso simultáneo). El tipo de suceso (clase) será decidido en la etapa de clasificación. Para detectar y segmentar la señal de tráfico se ha seleccionado un método muy sencillo: la ecuación (1) define la energía a corto plazo – Short Time Energy o STE – de una trama t .

$$STE_t = \sum_{n=0}^{N-1} |x_t[n]| = \frac{1}{N} \sum_{k=0}^{N-1} |X_t[k]|^2 \quad (1)$$

Una determinada trama será catalogada como ruido ambiental o ruido de tráfico dependiendo del calor de la STE comparado con un umbral dado. La mejor aproximación para fijar los umbrales que se ha probado es la basada en las estadísticas del nivel de ruido expresada en términos de percentiles L_N (indicadores que expresan el nivel de presión sonora que se excede durante el N% del tiempo del tiempo de medida). Se considera que el L_{90} representa el nivel de ruido de fondo. EL valor del L_{90} variará en función de la densidad media de tráfico y debe ser reajustado. Se ha elegido el $TH=L_{90}$ como umbral, de tal manera que si el nivel de señal excede $TH + 3$ dB o $TH + 6$ dB se considera que la señal es debido al paso de un vehículo. En función de las características del tráfico se elige $TH+3$ o $TH+6$ dB como criterio de decisión. La figura 3 muestra un ejemplo de detección y segmentación de señal continua de tráfico basada en la STE.

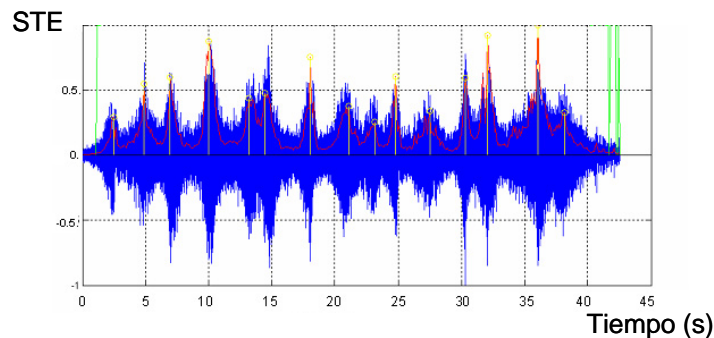


Figura 3– Detección del paso de vehículos mediante la utilización de la Energía a Corto Plazo.

Una vez se han detectado los intervalos de tráfico, el siguiente paso será aislar cada evento. Una vez se ha cubierto esta etapa, el sistema estará preparado para proceder al siguiente paso: la clasificación de las muestras.

La forma más simple de detectar si un vehículo ha pasado o no es analizar la evolución temporal de la envolvente de la señal, localizando los picos máximos. Como se están procesando bloques con N muestras, se puede obtener una estimación grosera de la envolvente mediante la STE de cada trama. Para este trabajo, la STE nos proporciona una estimación suficientemente precisa de la envolvente si se selecciona un tamaño de trama lo suficientemente corto. Si la trama es suficientemente corta, pueden detectarse vehículos con un intervalo de paso muy corto.

La figura (4) muestra la detección de vehículos con paso muy próximo (alto grado de solapamiento entre las señales). La señal de tráfico una vez segmentada puede proceder a ser limpiada mediante técnicas de sustracción de ruido [4,5].

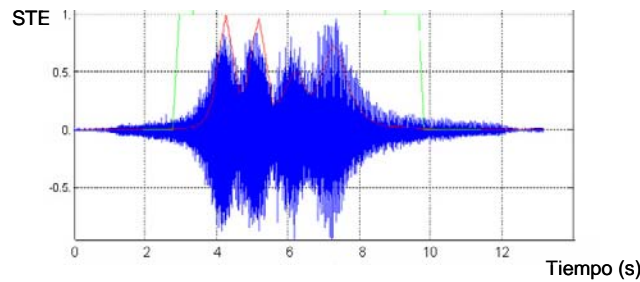


Figura 4– Detección del paso de vehículos muy próximos

3 Extracción de características.

La elección de características es clave para el sistema de clasificación y determinará la respuesta final del sistema (tasa de error). El conjunto de características seleccionadas constituirán un espacio de características de n dimensiones en el que la señal a detectar será representada. Para obtener un alto grado de precisión en la clasificación debe seleccionarse un conjunto adecuado de características y a menudo la única forma de llegar al conjunto óptimo es mediante prueba y error. A continuación se exponen el conjunto de características que han sido seleccionadas como candidatas a ser probadas. Todas las características expuestas están definidas en la bibliografía clásica de procesado de señal reconocimiento de patrones [1,3].

3.1 Características Temporales.

3.1.1 Tasa de Cruces por cero – ZCR.

Este parámetro se define como el número de cruces por cero que ha tenido lugar dentro de una trama de análisis. Aunque es un parámetro calculado en el dominio del tiempo proporciona una idea del contenido frecuencial de la señal. Se puede calcular a partir de la ecuación (2).

$$ZCR_t = \frac{1}{2} \sum_{n=0}^{N-1} |\text{sign}(x_t[n]) - \text{sign}(x_t[n-1])| \quad (2)$$

Donde $\text{sign}()$ representa la función signo con valor 1 para los valores positivos (incluyendo el cero) y -1 para los valores negativos.

3.2 Características frecuenciales

3.2.1 Centroide Espectral – SC.

Representa el centro de gravedad de la distribución de la densidad espectral de potencia. Está relacionada con la brillantez del sonido (mayor contenido en altas frecuencias que en medias y graves). Cuanto mayor es el valor del centroide espectral, más desplazada está la energía del espectro hacia las altas frecuencias. Se calcula mediante la expresión (3).

$$Centroid_t = \frac{\sum_{k=0}^{N-1} |X_t[k]| \cdot k}{\sum_{k=0}^{N-1} |X_t[k]|} \quad (3)$$

3.2.2 Punto de Caída Espectral: Spectral Rolloff point:

Esta característica mide la frecuencia por debajo de la cual se concentra la mayor parte de la energía del espectro. Se puede calcular mediante la ecuación (4)

$$SR = \max_m \left\{ \sum_{k=0}^m |X_t[k]| \leq TH \cdot \sum_{k=0}^{N-1} |X_t[k]| \right\} \quad (4)$$

donde el umbral TH toma valores entre 0.85 y 0.99.

3.2.3 Relación de energía en subbandas – SBER.

La SBER define la relación de energía entre una subbanda determinada y la energía total de la trama. Para la i -ésima banda, se calcula a partir de la expresión (5).

$$SBER_t = \frac{\sum_{k \in S_i} |X_t[k]|^2}{\sum_{k=0}^{N-1} |X_t[k]|^2} \quad (5)$$

EL espectro se divide típicamente en 4 intervalos no uniformes, definiendo las subbandas $S_1=[0, f_0/8]$, $S_2=[f_0/8, f_0/4,]$, $S_3=[f_0/4, f_0/2]$ y $S_4=[f_0/2, f_0]$ donde f_0 es la mitad de la frecuencia de muestreo. La figura 5 muestra la SBER en la cuarta subbanda. Puede observarse como hay claras diferencias entre las tres clases y a priori, esta característica se muestra con un alto poder discriminante y en apartados posteriores se comprobará como es la que permite una menor tasa de error en la clasificación.

La figura 5 permite ya anticipar buenos resultados a partir de la representación de la SBER de la cuarta banda. Se observa cómo claramente las motocicletas presentan una SBER desplazada hacia zonas altas de la banda, mientras que la energía total en la banda es superior en los vehículos ligeros que en los pesados.

3.3 Características perceptuales: parametrización MEL.

Durante el proceso de prueba, se determinó utilizar los coeficientes MEL (MFCC). Los MFCC se consideran coeficientes que recogen las características perceptuales de las señales. debido a que en su cálculo se utiliza un filtrado en el dominio de la frecuencia mediante un banco de filtros en escala Mel. Si una persona es capaz de distinguir si una determinada señal procede de una motocicleta o un coche, debería ser posible reproducir las características subjetivas que hacen que el oído humano sea capaz de apreciar las diferencias. Para representar la voz humana se utilizan 13 coeficientes cepstrales MFCC. Para el presente trabajo, se han utilizado 5, y comprobado que son suficientes para producir buenos resultados en la clasificación [6]. Para obtener estos coeficientes, la señal se filtra en el dominio de la frecuencia mediante un banco de filtros en escala Mel y a continuación se realiza la transformada inversa de Fourier del logaritmo del espectro.

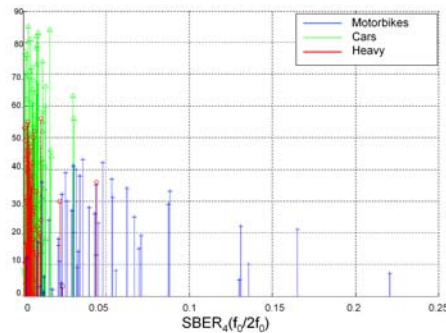


Figura 5 – Representación de la SBER de la cuarta subbanda.

4 Algoritmos de Clasificación.

En este apartado se detallan la serie de métodos de clasificación que han sido probados durante este trabajo: *k*-vecinos más próximos (*k*-NN) y el discriminador lineal de Fischer.

4.1 K-Vecinos más próximos (k-NN).

El método de los *k*-vecinos más próximos, sitúa los puntos de entrenamiento en el espacio de características. Para determinar a qué clase pertenece un nuevo evento se toman los *k*-puntos más cercanos y se determina cuál es la clase dominante en el entorno del nuevo suceso.

Para calcular la distancia a los puntos de entrenamiento en un espacio de *N* características (*n* dimensiones por tanto) se utiliza la distancia de Mahalanobis, definida en la ecuación (6).

$$d_M(x, y) = \sqrt{(x - y)^T C^{-1} (x - y)} \quad (5)$$

donde *C* es la matriz de covarianza del conjunto de datos de entrenamiento. La utilización de esta distancia en lugar de la distancia Euclídea presenta algunas ventajas importantes, la principal de ellas es que es independiente de la escala de las características del conjunto de entrada ya que realiza automáticamente un escalado de los ejes del espacio de características.

4.2 Discriminador Lineal de Fischer – FLD.

Los clasificadores que se basan en el análisis con un discriminante lineal son métodos supervisados que utilizan la información del conjunto de entrenamiento para establecer las fronteras entre clases. El análisis busca la proyección de los datos de un espacio de *d* dimensiones a lo largo de una línea en la dirección del discriminante. Dando una interpretación geométrica a este proceso, la superficie de decisión correspondiente será un hiperplano, *H_s*, que separa las zonas de decisión.

El discriminador lineal de Fischer únicamente permite separar entre dos categorías (*C1* y *C2*) aunque puede ser fácilmente generalizado a un número arbitrario de clases. La dirección del discriminante será la solución que maximice una función criterio determinada. El discriminador lineal de Fischer – FLD – propone la proyección sobre el vector *w* que maximiza la separación de los datos, en términos

del error cuadrático medio ponderado por la dispersión total intraclase [7]. En la referencia [7] se puede encontrar una descripción completa del FLD.

Si existen más clases implicadas en el proceso de separación, se puede extender el método de Fischer para trabajar con más clases. En este trabajo se ha optado por la división del problema en múltiples problemas de dos clases. Para ello se pueden adoptar dos estrategias diferentes:

1. *Uno frente a todos.* En el entrenamiento se considera que una clase es la positiva mientras que todas las demás constituyen la clase negativa. Por tanto, es necesario entrenar tantos clasificadores como clases. Al finalizar el proceso, la clase asignada será la que haya obtenido un mayor margen (en el caso en el que se haya resultado positiva más de una clase).
2. *Uno frente a uno.* En esta estrategia, se implementan $c(c-1)/2$ clasificadores lineales de tal forma que se cubren todas las posibles combinaciones. Se sigue entonces un procedimiento de votación: cada clasificador genera un voto a una clase, y la clase estimada será la que obtiene al final un mayor número de votos.

La figura (6) evidencia por qué la estrategia *uno contra uno* es la más utilizada: presenta una mayor precisión en la clasificación al tener una menor región ambigua (región donde no se puede determinar a qué clase pertenece un evento).

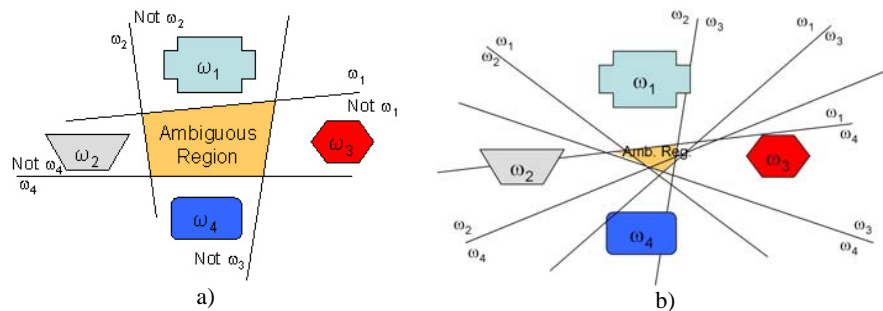


Figura 6 – Fronteras de decisión establecidas por las estrategias: a) uno contra todos, b) uno contra uno

5 Resultados

Para poder probar las posibilidades de clasificación automática de ruido de tráfico, se ha procedido a registrar una base de datos con 100 señales de cada clase. Las señales se han grabado en formato PCM con una frecuencia de muestreo de 44100 bps y 16 bits por muestra. En el proceso de clasificación cada señal ha sido diezmada hasta los 11025 bps con lo que el ancho de banda efectivo para el análisis (extracción de características) es $f_s/2 = 5512\text{Hz}$. Se seleccionaron 40 señales de cada clase como conjunto de entrenamiento y el resto de la base de datos se utilizó para probar el sistema.

La tasa de cruces por cero, mostró un buen comportamiento discriminante entre los vehículos pesados y las motocicletas, pero no así ante la clasificación entre vehículos ligeros y pesados. El centroide espectral mostró un comportamiento similar al ZCR. Sin embargo, las SBER han presentado un gran poder discriminante: los camiones pesados muestran una mayor concentración de energía en frecuencias bajas, mientras que la densidad espectral de potencia es superior en altas frecuencias en el caso de las motocicletas. Las bandas con más capacidad de discriminación han sido:

1. $S_3 = [f_0/4, f_0/2] = [1.4, 2.8] \text{ kHz}$
2. $S_4 = [f_0/2, f_0] = [2.8, 5.6] \text{ kHz}$

La SBER en la cuarta subbanda se ha mostrado en la figura 5 y se han comentado sus características discriminantes. Otras características en el dominio de la frecuencia muestran un buen comportamiento, como el *Spectral Rolloff* –SR– con valores para el umbral entre 0.55 y 0.70. Los coeficientes MEL también presentaron buenas características para la clasificación.

La última revisión de la norma ISO 1996:2007 exige el conteo del número de vehículos y su clasificación en *al menos dos clases*, durante el proceso de medida de la emisión de ruido de un ramo de carretera. Para probar el sistema de clasificación se ha optado por seguir dos estrategias. En primer lugar se ha utilizado únicamente dos clases: *vehículos pesados y vehículos ligeros*, agrupando en esta última a motocicletas y coches. La tabla 1 muestra los resultados en este caso. Para las pruebas se utilizó un k_NN con k=3 y el FLD. Puede observarse en la tabla, como en ambos casos la SBER presenta los mejores resultados, dándose la menor tasa de error para FLD (10.17 %).

<i>Parámetros</i>	<i>Probabilidad de Error (%)</i>	
	<i>3-NN</i>	<i>FLD</i>
ZCR	30.34	35.45
Centroide Espectral	51.49	38.25
Punto de Caída Espectral	28.68	26.12
SBER	12.18	10.17
MFCC	15.73	13.57

Tabla 1 – Probabilidad de error para dos clases utilizando únicamente una característica para clasificación.

La tabla 2 muestra el resultado obtenido cuando se utiliza sólo una característica e intentamos clasificar en tres clases. Se observa que el comportamiento es ligeramente peor que en el caso anterior, y que el mejor resultado se obtiene de nuevo para el FLD, utilizando una estrategia uno contra uno y los coeficientes cepstrales Mel como característica. (15.76 %)

<i>Parámetros</i>	<i>Probabilidad de error (%)</i>		
	<i>3-NN</i>	<i>FLD</i>	
		<i>Uno contra todos</i>	<i>Uno contra uno</i>
ZCR	39.98	-	43.97
Centroide Espectral	38.43	-	27.04
Punto de Caída Espectral	31.94	-	22.01
SBER	16.24	36.21	27.59
MFCC	18.42	17.19	15.76

Tabla 2 – Probabilidad de error para tres clases utilizando una característica.

Finalmente la tabla 3 muestra los resultados obtenidos cuando se utilizan tres características simultáneamente. Se muestra la mejor combinación obtenida. Se observa como se llega a conseguir resultados similares a los obtenidos para la separación en dos clases. El mejor comportamiento lo presenta en este caso el 3-NN. Para interpretar si el orden del error que estamos cometiendo en la clasificación es o no elevado, se debe considerar el objetivo de la clasificación: primero, que un conteo manual no consigue mejor representación del tráfico ya que suelen ser meramente puntuales y muestrales. En segundo lugar, la emisión de ruido de una carretera es función del $10 \log(N)$, donde N es el número de vehículos circulando. Un error de conteo del 10 % conduce a un error en la estimación de la emisión de unos 0.5 dB (dependiendo del rango del tráfico).

Parámetros	Probabilidad de error (%)		
	3-NN	FLD	
		Uno contra todos	Uno contra uno
MFCC, SBER, SR	10.07	13.36	11.82

Tabla 3 – Probabilidad de error para tres clases combinando MFCC, SBER y SR (espacio de clasificación de 3 dimensiones)

6 Conclusiones

En este artículo se ha presentado las pruebas preliminares de un trabajo cuyo objetivo es la determinación automática de toda la información requerida para la caracterización de tráfico con el objetivo de calcular la emisión acústica de una carretera. Se han obtenido unos resultados preliminares de un 10 % de error utilizando una combinación de características (MFCC, SBER y SR) y el método de k-vecinos más próximos con $k=3$. Los resultados son lo suficientemente buenos como para pensar en que merece la pena profundizar y ampliar el trabajo. De entrada, son posibles muchas mejoras, entre ellas:

- Ampliar la base de datos de entrenamiento. El número de señales utilizadas para entrenar el sistema es muy pequeño.
- Utilizar otros sistemas de clasificación (redes neuronales por ejemplo)
- Incrementar el número de clases, subdividiendo más el tipo de vehículos. Precisamente el mayor error se da entre coches y camiones, ya que esta clasificación es demasiado “gruesa”: una subdivisión en camiones ligeros y camiones pesados podría mejorar la tasa de clasificación.
- Se puede plantear la clasificación de pasos simultáneos de vehículos, incrementando el número de clases, una nueva clase para cada par de eventos posibles.

Agradecimientos

Este trabajo ha sido financiado parcialmente por el proyecto ref. TEC2006-13883-C04-02: “Separación De Fuentes Sonoras Para Su Aplicación En Medidas Acústicas” del Ministerio de Educación y Ciencia.

Referencias

- [1] John G. Proakis, Dimitris Manolakis. *Digital Signal Processing. Principles, Algorithms and Applications*. Prentice Hall, Febrero 2004.
- [2] Enrique A. Cortizo, Manuel Rosa-Zurera and F. López Ferreras. “Application of Fischer Linear Analysis to Speech/Music Classification”. *Proceedings of EUROCON*, Belgrado 2005, pp. 1666-1669.
- [3] Dietrich W. R. , J. Hornegger. *Applied Pattern Recognition*. Fourth Edition. Ed. Vieweg, Febrero 2004.
- [4] Vary, P. ”Noise suppression by spectral magnitude estimation – mechanism and theoretical limits”. *Signal Processing* 8(4), 387-400 (1985)

- [5] Kamath, S. and P. Loizou (2002). "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise". *In Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing (ICASSP'02) Orlando 2002.*
- [6] Enrique A. Cortizo, Manuel Rosa-Zurera and F. López Ferreras."Application of Fischer Linear Analysis to Speech/Music Classification". *Proceedings of EUROCON, Belgrado 2005, pp. 1666-1669.*
- [7] Max Welling. *Fischer Linear Discriminant Analysis*, en <http://www.cs.huji.ac.il/~csip/Fisher-LDA.pdf>