



FIA 2008- A017

Procesado psicoacústico de audio para la transmisión de alta calidad en FM Estéreo

Oscar Bonello

Fellow Audio Engineering Society, New York. Solidyne Labs, Buenos Aires. E-mail: oscar@solidyne.com.ar

Abstract

The psychoacoustics theory involved in the design of state-of-the-art digital multiband audio processors used for AM and FM broadcasting transmission is described. It is analyzed how the average rms value of the audio signal can be improved without losing audio quality, using the Burst Masking concepts. It is further demonstrated that audio processing increases the coverage of FM stereo transmission.

Resumen

Se describe la teoría psicoacústica vinculada con el diseño de los procesadores multibanda empleados en radiodifusión de AM y FM estéreo. Se analizarán en particular los conceptos de enmascaramiento del recorte de señales y la aplicación del Enmascaramiento por Ráfaga (Burst Masking). Asimismo se verá la manera en que el procesado psicoacústico incrementa el alcance de las transmisiones de FM estéreo.

1 Introducción

Vemos en la figura 1 el diagrama en bloques simplificado, de un procesador de audio multibanda, ya sea digital o analógico. El número de bandas usualmente varía entre 3 y 6 bandas. La señal de audio viene con pre-énfasis y es llevada a filtros de banda, generalmente de 24 dB/octava. Cada banda es comprimida con compresores de pendiente elevada (usualmente mayor de 20:1). Los tiempos de ataque y recuperación son ajustados por el ingeniero de la estación para optimizar el sonido, desde un punto de vista psicoacústico. A primera vista surge que al comprimir en bandas separadas es posible lograr una elevada densidad de sonido pues cada banda es nivelada para máxima señal. Esto se vincula con el concepto de sonoridad de una señal de programa. Es un concepto clásico de la psicoacústica vinculado con la forma en que en el oído interno, más precisamente en la membrana basilar, son excitadas las ciliias y sus correspondientes terminales nerviosas. Dicha excitación se distribuye por bandas de frecuencia sobre la membrana formando una especie de analizador de espectro biológico

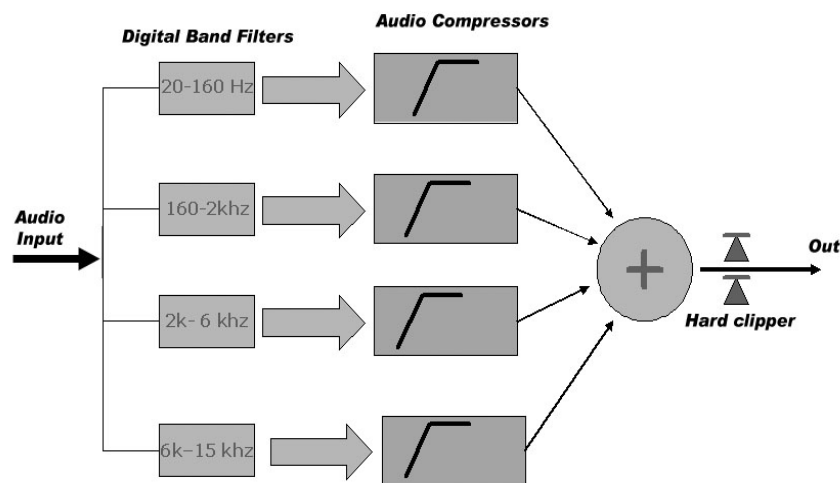


Figura 1. Diagrama en Bloques de un compresor multibanda

Cada una de las notas musicales ocupa una determinada zona de la membrana basilar. Vemos en la figura 2 la excitación que producen dos tonos. Esto representa una típica situación al escuchar un tema musical en el cual existen dos bandas predominantes. Si este material sonoro es comprimido en varias bandas, el nivel sonoro de cada una de ellas se incrementa, como se ve en la parte inferior de la figura 2, logrando una excitación más amplia de la membrana basilar, que el oído asocia con un sonido de mayor sonoridad, aunque su nivel rms sea el mismo.

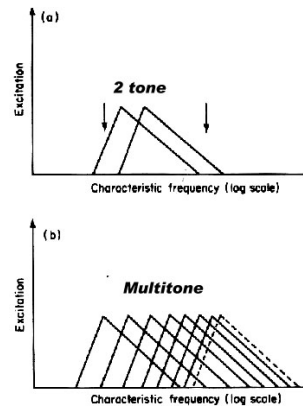


Figura 2. Excitación de la membrana basilar del oído

Esto explica una parte de los resultados que se obtienen con la compresión multibanda. Pero para aumentar el nivel rms hace falta algo más; un recortador de audio. Esto es debido a que el material de programa de una radio tiene una distribución de amplitudes aleatoria en donde los impulsos de muy corta duración pueden triplicar el promedio rms. Como la máxima modulación, aún en picos cortos, no debe sobrepasar el 100%, la forma más efectiva de aumentar el valor rms es eliminar por recorte los picos de muy reducida duración. El concepto de “recorte” es naturalmente asociado a la idea de distorsión y es absolutamente rechazada como parte de cualquier cadena de audio de calidad. Sin embargo, en radiodifusión el recorte de señales tiene una larga tradición, debido a la necesidad de obtener elevados valores de modulación. A lo largo de los años los ingenieros de desarrollo han creados varias técnicas para reducir los efectos indeseables del recorte, manteniendo las ventajas psicoacústicas de obtener una mayor excitación en la membrana basilar. Casi siempre estas técnicas se han desarrollado empíricamente y raramente han tenido un justificativo teórico que las avale. Analizaremos entonces la forma en que el oído percibe la distorsión derivada del recorte de una señal senoidal

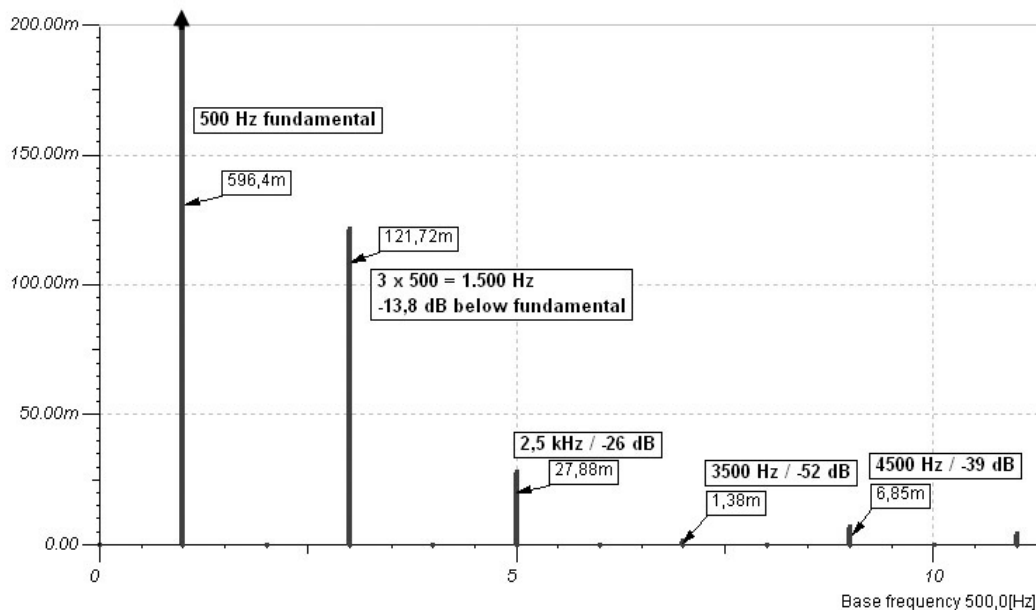


Figura 3. Espectro de una onda senoidal de 500 Hz recortada 6 dB

En la figura 3 vemos el espectro armónico de una onda senoidal de 500 Hz, recortada 6 dB. este es un caso extremo pues el recorte real de un procesador moderno es típicamente menor de 3 dB. Para analizar la audibilidad de la distorsión deberemos tener en cuenta el efecto de enmascaramiento. Para ello dibujamos en el mismo gráfico el espectro armónico de la señal recortada y el enmascaramiento que el tono de 500 Hz produce. Emplearemos para ello el procedimiento de cálculo propuesto por Terhardt et al [2] cuyos resultados se aproximan mucho a las curvas de enmascaramiento reales, medidas por otros autores. El resultado es el gráfico de la figura 4. Puede verse a simple vista que el enmascaramiento no mejora la situación y la distorsión es perfectamente audible. Es decir que una sola nota musical sostenida en el tiempo y recortada 6 dB sería totalmente audible.

Hay, sin embargo, otros factores que mejoran la situación. Al tratarse de material sonoro complejo (música o palabra), la compresión multibanda eleva la energía espectral de las bandas adyacentes. Vemos en la figura 5 el resultado de agregar dos bandas de frecuencias separadas a una y dos octavas con niveles 10 y 20 dB por debajo de la banda principal de 500 Hz. Vemos que los armónicos indeseados generados por el recorte son parcialmente enmascarados. Si elevamos al mismo nivel las 3 bandas (figura 6) mediante mayor compresión, todas las componentes, menos la de 3era armónica, quedan enmascaradas y por lo tanto no son escuchadas. Vemos entonces que en función del tipo de material de programa y de la compresión utilizada es posible reducir la distorsión producida por un recorte severo. Obviamente si el material de programa carece de energía en zonas espectrales próximas al recorte, no será posible usar esta técnica. O dicho de otro modo: con este tipo de material de programa (por ejemplo un piano o clavicordio) un recorte de 6 dB será escuchado. Por este motivo los procesadores más modernos de FM reducen a 2 ó 3 dB el nivel máximo de recorte.

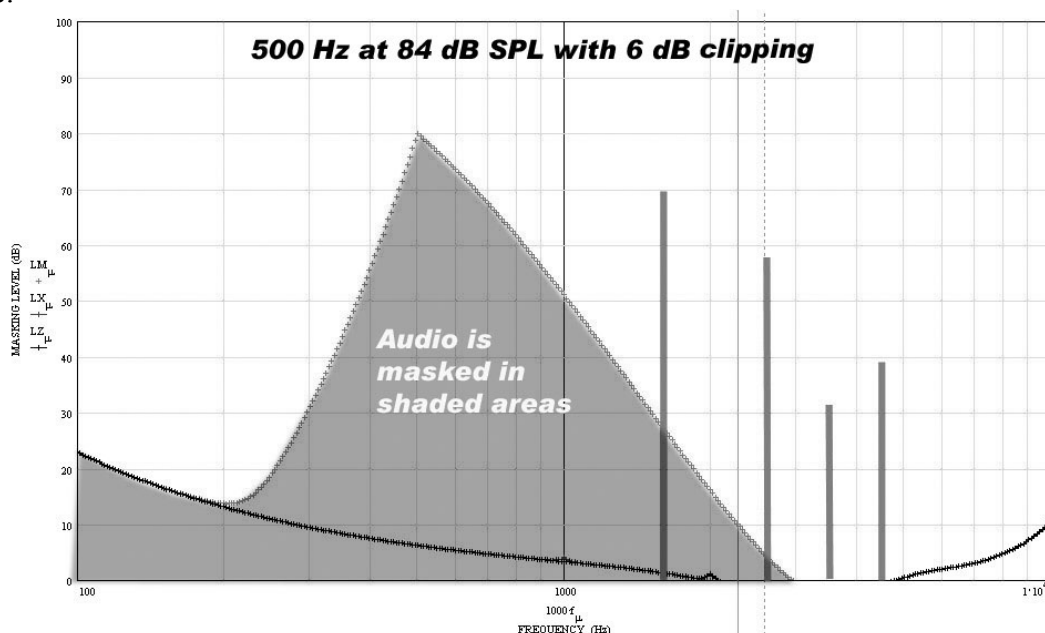


Figura 4 Armónicas de la onda de 500 Hz recortada

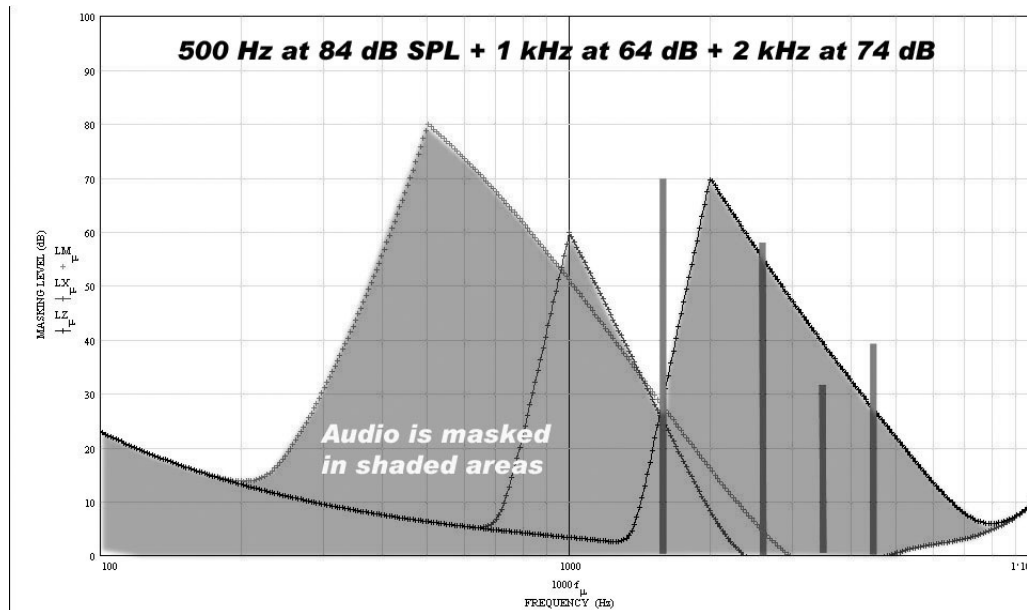


Figura 5. Enmascaramiento cuando 3 bandas están presentes

El análisis precedente, sin embargo, no tiene en cuenta otro de los avances que los modernos procesadores de audio han incorporado: el control del tiempo de recorte. Volvamos a la figura 1 y analicemos la acción de los compresores en función del tiempo. Imaginemos que el umbral de compresión coincide casi exactamente con el umbral del recorte. Si además los tiempos de ataque son instantáneos, el recortador no trabajará nunca. Pero en los procesadores de audio para AM / FM el tiempo de ataque es considerablemente lento y diferente para cada banda y para cada tipo de programa. De esta manera, cuando ensayamos el procesador con onda senoidal nunca veremos el recorte. Pero si lo ensayamos con una ráfaga de señal, por ejemplo un tren de 50 ms de señal de 1 kHz, veremos que los primeros ciclos son recortados y luego, cuando el compresor alcanza su valor final, el recorte desaparece. De esta manera se logran recortar los picos de muy corta duración que deseamos eliminar, evitando que los picos de duración considerable, afecten a la calidad de audio pues la reducción de ganancia del compresor hace que el sistema opere linealmente. Esto último merece un detallado análisis.

La percepción de la distorsión durante cortos intervalos es diferente del análisis en régimen continuo, dado por el efecto enmascaramiento, antes visto. Durante intervalos muy cortos, está en juego la capacidad del cerebro de resolver la información que desde el oído interno llega por 30.000 fibras nerviosas. En el caso de una computadora que resuelve una transformada de Fourier tenemos que computar N muestras dentro de una ventana temporal T_n . El tiempo requerido para hacer los cálculos nos dará la demora necesaria para obtener el espectro. En el caso del cerebro, la "computadora biológica" también requiere un determinado tiempo para transformar en una percepción sensorial el hecho de que un tono venga acompañado por otros que serían interpretados como distorsión. Este fenómeno psicoacústico puede interpretarse como un enmascaramiento denominado Enmascaramiento por Ráfaga (Burst Masking) tal como ha demostrado Bonello [3]. Este enmascaramiento se aplica al reconocimiento de un tono de audio, de corta duración (ráfaga) en presencia de otro tono más intenso y de muy diferente frecuencia (para estar fuera de la zona de enmascaramiento en régimen permanente). También se aplica al caso en que un tono puro enmascara al ruido blanco.

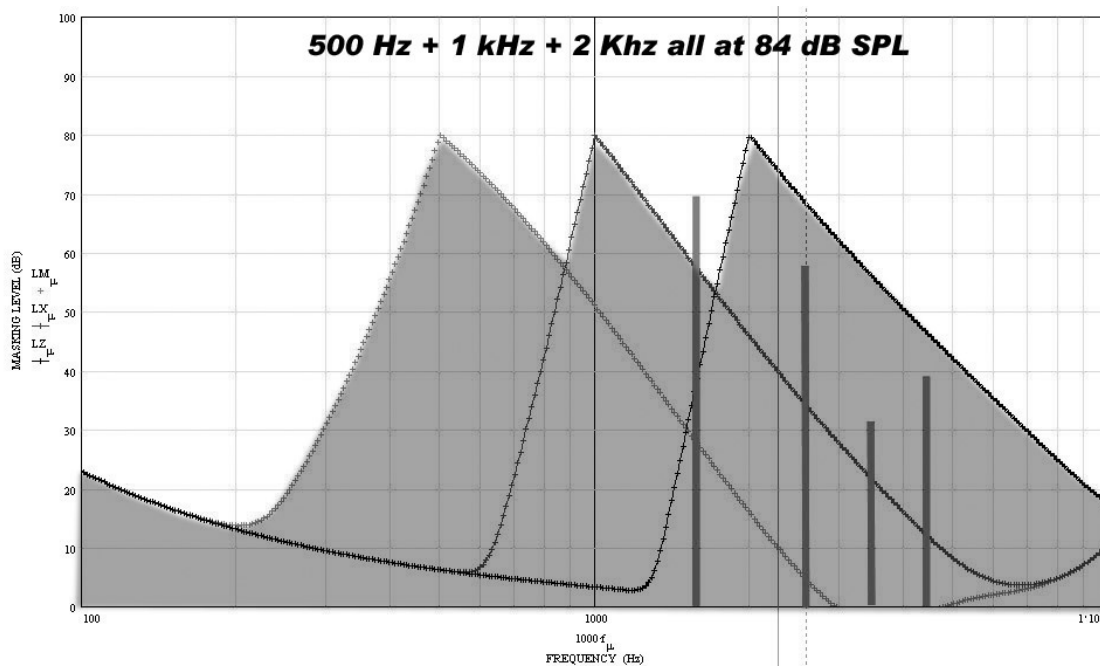


Figura 6. Enmascaramiento con 3 bandas al mismo nivel

Vemos en la figura 7 una curva, tomada de [3], que resume los resultados del estudio estadístico realizado. A los sujetos se les presenta alternativamente una ráfaga de N milisegundos del tono enmascarante y otra del mismo más un segundo tono o ruido. Las curvas de la figura 7 indican el límite por debajo del cual el segundo tono deja de ser reconocido. Por ejemplo, si aplicamos una curva de enmascaramiento (estático) a un tono puro de 1 kHz a 90 dB SPL, encontramos que podemos reconocer el tono de 5 kHz a 65 dB por debajo del tono de 1 kHz. Pero la curva superior de la figura 7 (1kHz / 5 kHz) nos informa que dentro de una ráfaga de 50 ms este valor se reduce a 53 dB y en 5 ms es de solamente 36 dB. El concepto de Enmascaramiento por Ráfaga nos ayuda a entender, de una manera científica, por qué motivo ajustando los tiempos de ataque de los compresores multibanda, es posible reducir y hasta eliminar los efectos negativos de la distorsión por recorte, manteniendo sin embargo la ventaja de aumentar el valor rms de la señal de audio.

1.1 Efectos del enmascaramiento por ráfaga

Debemos observar que la acción del enmascaramiento es reducida para la 3ra armónica, pero es definitivamente importante para la 5ta armónica y superiores, que son particularmente responsables del sonido áspero del recorte. La 3era armónica, en cambio, le agrega un timbre al sonido original que lo modifica levemente, pero de manera natural, sin causar desagrado al oído. Imaginemos por ejemplo una señal que es recortada un 30 %. Mirada al osciloscopio la calificaríamos como un recorte muy severo. Sin embargo el cálculo de Fourier nos indica que la 5ta armónica y las sucesivas están por debajo de 40 dB de la fundamental. En condiciones de escucha permanente este valor es perfectamente detectable (recordemos que nuestro umbral de enmascaramiento estático era de -65 dB) Pero aplicando las curvas del enmascaramiento de ráfaga, vemos que para 10 ms tenemos 40 dB de enmascaramiento que hacen inaudible un recorte tan severo.

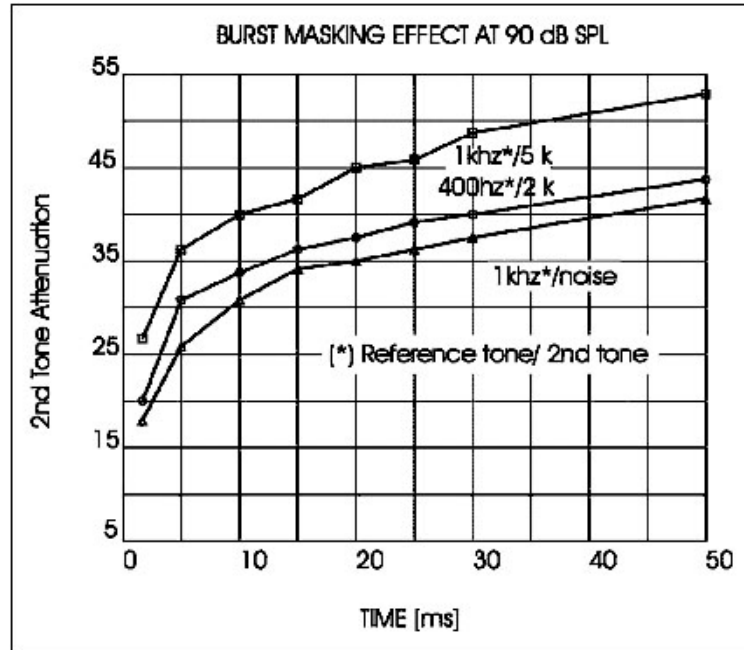


Figura 7. Enmascaramiento por ráfaga según O.Bonello [3]

Esto explica, creemos que por primera vez, el motivo por el cual en numerosos ensayos de audición, los oyentes han definido el sonido procesado como de mayor calidad que el original, pues el oído percibe una mayor sonoridad, sin percibir un aumento de la distorsión.

2 Alcance de una emisora de radio de FM

Una vez analizada la tecnología que gobierna a un procesador de audio para FM, podremos estudiar a fondo el tema del aumento del área de cobertura de una estación debida al procesado de audio. Aquí nos olvidaremos por un momento de las consideraciones psicoacústicas y admitiremos que más allá de la sensación auditiva de mayor sonoridad, hay elementos perfectamente mensurables como el aumento del nivel rms de la señal de audio, debido a la compresión multibanda y al recorte de picos. La técnica de medición propuesta en [1] es hoy empleada en Monitores de Modulación de FM y en procesadores de audio, para que el ingeniero de la estación pueda evaluar exactamente el incremento en el valor rms que produce determinado ajuste de parámetros en el procesador de audio. Este dato es importante pues está directamente relacionado, como luego veremos, con el área de cobertura de la estación.

Ha demostrado Bonello [5] que la sumatoria de las potencias de las infinitas bandas laterales de una transmisión de FM está dada por:

$$P_{lat} = P_o (1 - (J_0(m))^2)$$

Siendo

P_{lat} la potencia total en bandas laterales, P_o la potencia del transmisor

m = índice de modulación = Desviación de frecuencia / Frecuencia de modulación

J_0 = Función de Bessel de Tipo-I, orden cero para el valor m

El próximo paso es graficar esta función para obtener conclusiones. Analizaremos la potencia de salida en bandas laterales Plat, en función del porcentaje de modulación, para una desviación de frecuencia de 75 kHz

En la figura 8 vemos dicho gráfico correspondiente a un transmisor de FM de 100 W de potencia. Para el gráfico se usaron numerosas frecuencias de modulación (fm) dentro de la octava de 1 kHz (707 Hz – 1414 Hz)

Tenemos en el eje X los valores de modulación desde 1 % al 100% y en el eje Y la potencia de salida en bandas laterales, de las cuales depende el alcance de la radio, desde 1 W hasta 100W

Hemos incluido en el mismo gráfico a la potencia en bandas laterales de un transmisor de AM de 200 W, tomado como referencia, por ser un hecho conocido que las señales de AM aumentan su alcance al aumentar la modulación [1].

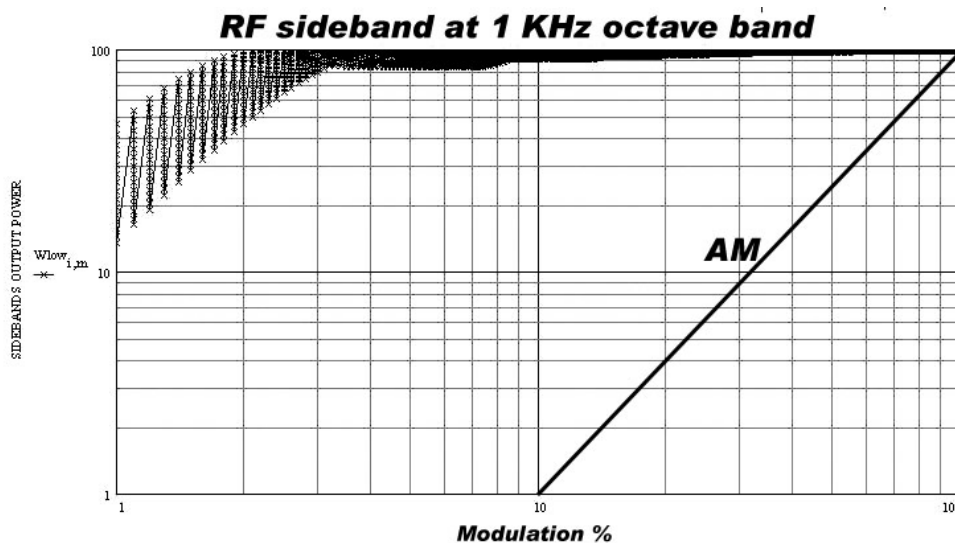


Figura 8 Bandas laterales de modulación en octava de 1 kHz (707 Hz a 1414 Hz) para transmisor de FM de 100 W, comparada con transmisor de AM de 200 W

Puede verse la enorme diferencia entre AM y FM pues en AM necesitamos tener un 45 % de modulación para obtener 20 W de salida. Esto indica la enorme importancia del procesado de audio, que permite alcanzar altos valores de modulación. Contrariamente, en FM con solamente un 1 % de modulación ya obtenemos 20 W de potencia de RF en bandas laterales. ¿Qué significa esto? Muy sencillo: que en una transmisión de FM mono el procesado de audio no ayuda a alcanzar mayores potencias de RF. Es decir que seguramente aumentará la sonoridad del audio pero no aumentará su alcance.

¿Qué pasa en altas frecuencias? Vemos en la figura 9 el comportamiento de una señal de 12 kHz transmitida en FM mono. Si bien la situación ha cambiado, aproximándose un poco hacia la de la modulación de AM. Vemos que aún en esta frecuencia el procesado de audio (y el aumento de la modulación) no brindan ventajas importantes. De hecho con el 10% de modulación ya se obtienen 20 W y con 30 % de modulación se alcanza la máxima potencia de salida

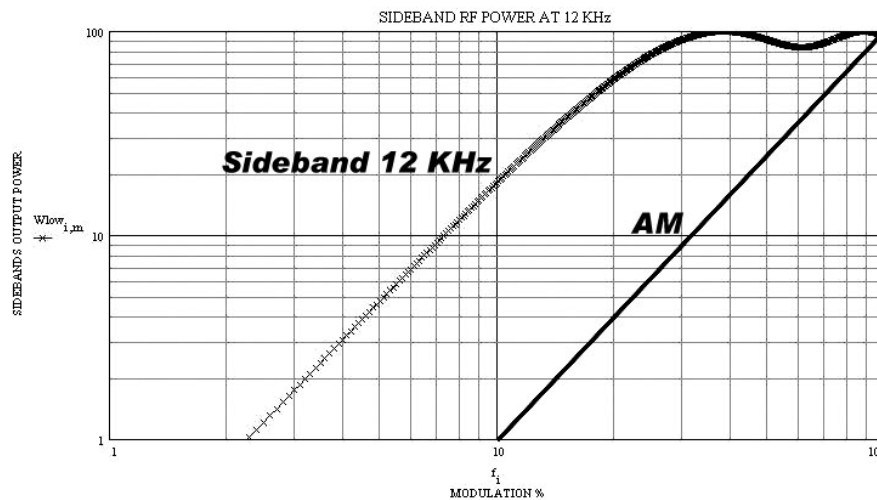


Figura 9. Bandas laterales de FM para 12 kHz de modulación

2.1 Alcance en FM Estéreo

El alcance de la radio es función de la potencia transmitida, es decir la potencia total en bandas laterales. La señal estéreo es lograda mediante una sub-banda de 38 kHz modulada en amplitud por la componente L-R. Si esta banda de 38 kHz llega débil a los receptores, se produce distorsión y ruido de fondo que impide la recepción confortable. En términos generales el ruido en la subportadora de 38 kHz marca el final del área de cobertura de una radio de FM Estéreo.

Veamos entonces en la figura 10 el comportamiento cuando enviamos una subportadora de 38 kHz. Puede verse ahora un notable acercamiento a los valores de AM que indica que el aumento del valor rms de la señal de audio de modulación, también incrementará la potencia en bandas laterales y por lo tanto el alcance de la radio en FM Estéreo. Por primera vez aparece el fenómeno notado en forma empírica por muchos ingenieros de estaciones de FM, pero que hasta el momento fue siempre considerado como una leyenda por muchos ingenieros de comunicaciones y libros de texto universitarios.

Si consideramos ahora que la banda de frecuencias a transmitir en FM es de 30 Hz a 15 000 Hz, resultará que la subportadora de 38 kHz deberá tener un ancho de banda de 38 kHz +/- 15 kHz.

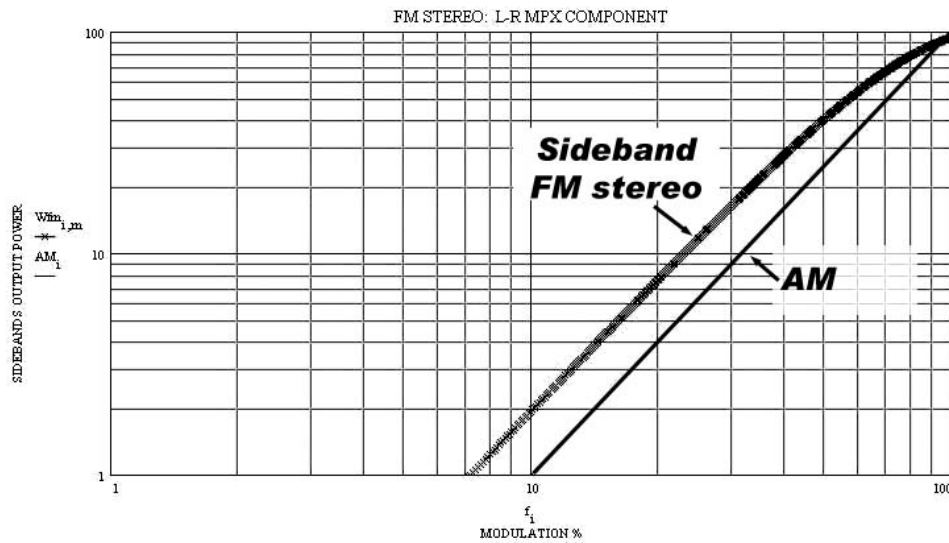


Figura 10. Bandas laterales en transmisión estéreo debidas a subportadora de 38 kHz

Si graficamos nuevamente para toda la banda comprendida entre esos extremos, obtenemos la figura 11. En la misma se advierte el área sombreada que representa el ancho de banda de transmisión hasta 15 kHz. Todo el material sonoro emitido caerá forzosamente dentro de esta banda. Por lo tanto el comportamiento de la transmisión total dependerá de la potencia aplicada al peor punto de esta banda, que corresponde a su borde inferior, donde para la misma modulación, las potencias son menores. Es interesante observar que la transferencia de potencia, en función de la modulación, es casi exactamente igual en AM y en FM estéreo

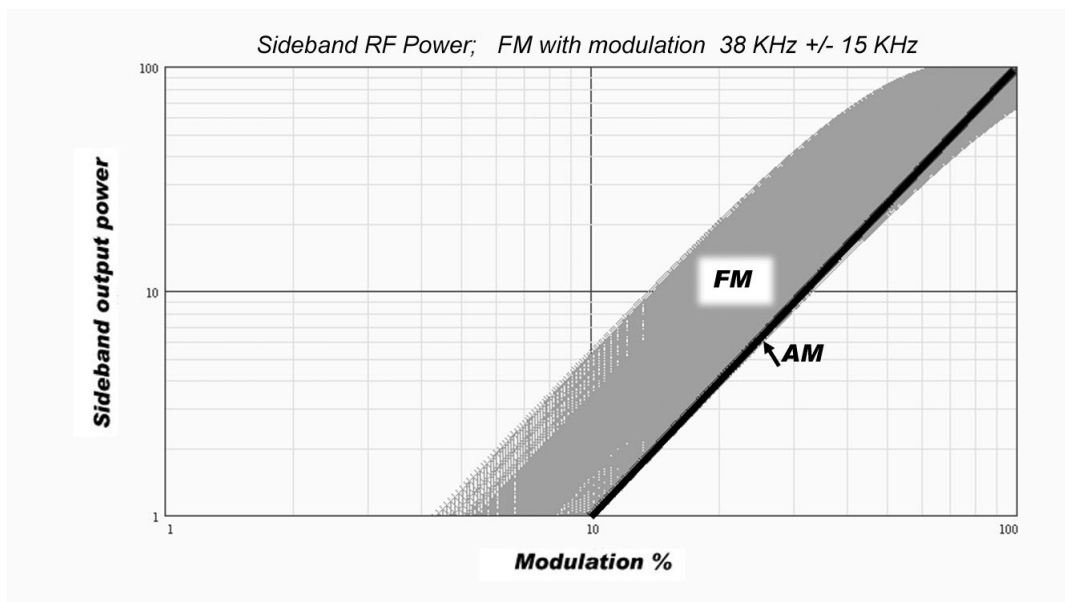


Figura 11. FM estéreo con modulación MPX entre 30 Hz y 15 000 Hz

3 Conclusión

Hemos analizado el estado actual de las teorías que involucran el diseño de los procesadores de audio digitales y analógicos para radiodifusión. Hemos visto las razones por

las que logran un importante incremento en la sonoridad de una señal de audio y en los valores rms de modulación con un mínimo deterioro audible de la calidad del sonido. Este aparente deterioro es, sin embargo, compensado por la sensación de alta sonoridad y graves de alto impacto que hace que una alta proporción de la audiencia considere este sonido como *mejor que el original*.

Finalmente, analizando las bandas laterales de la transmisión, considerando la propiedad ortogonal de las funciones de Bessel, hemos podido demostrar que en FM mono solamente nos beneficiamos con el aumento de sonoridad del procesado de audio, mientras que en FM Estéreo ocurre un importante aumento de la potencia radiada, dado en la figura 11 lo que nos permite predecir con exactitud el incremento real del alcance de una transmisión conociendo el aumento del valor rms de la modulación debida al procesado de audio multibanda.

Referencias

- [1] Oscar Bonello “New improvements in audio signal processing for AM Broadcasting”, AES Journal Vol 24 #5, June 1976
- [2] E.Terhardt, Gerhard Stoll and M.Seewann, “Algorithm for extraction of pitch and pitch salience from complex tonal signals” Journal of Acoust.Soc.Am. Vol 71 pag 679, March 1982
- [3] Oscar Bonello, Burst Masking, “Enmascaramiento por Rafaga”, EAA Symposium of Architectural Acoustics and Iberian Congress on Acoustics, Madrid, Spain, October 2000, ISBN 84-87985-03-3
- [4] R.Cabot, R.Genter II and T.Lucke, “Sound Levels and Spectra of Rock Music”, AES Journal Vol 27, # 4, April 1979
- [5] Oscar Bonello, “Multiband audio processing and its influence on the coverage area of FM stereo transmission” AES Journal, March 2007