

ENMASCARAMIENTO POR RÁFAGA: UN NUEVO PRINCIPIO EN PSICOACÚSTICA

Referencia PACS: 43.66.Dc

Oscar Bonello
Facultad de Ingeniería
Universidad de Buenos Aires
3 de Febrero 3254
1429 Buenos Aires. Argentina
Tel/Fax: 54 11 4702 0090
E-mail: info@solidyne1.com

ABSTRACT

The temporal behavior of the auditory system with 2-50 ms signal bursts will be analyzed in order to obtain its discrimination capability. A primary high level tone and a secondary one will be present at the same time. It will be demonstrated that the capability to detect the low level tone depends on the length of the burst. We named this effect : *burst masking*. The lack of this discrimination with short bursts allows us to improve the audio bit compression technologies used today.

RESUMEN

Se analiza el comportamiento temporal del sistema auditivo frente a señales de corta duración, entre 2 y 50 ms, para investigar la capacidad de distinguir la presencia de un segundo tono en presencia de un tono dominante.

El enmascaramiento por ráfagas es presentado en una curva e indica que para ráfagas cortas esta capacidad de discernimiento del oído, se reduce considerablemente. Esto permite la optimización de las nuevas tecnologías de compresión digital de voz y música.

INTRODUCCION

Desde 1982 que estamos trabajando en la tecnología de comprimir señales digitales de audio para manejar música y voz humana almacenándola en discos fijos de computadora. Pese a los enormes avances en capacidad de almacenamiento, aún hoy sigue siendo un tema de importancia la compresión digital para permitir reproductores portátiles con memorias de estado sólido, transmisión de música por Internet, etc. En 1988 se presentó el primer sistema viable de compresión (el ECAM, precursor del MPEG-2 y MP3) que fue utilizado en forma casi inmediata en 35 países, muchos de ellos europeos.

Como parte de este programa de investigación, se profundizó el estudio de las bases científicas que posibilitan la reducción de la cantidad de información. La mayor parte de ellas relativas al enmascaramiento de tonos, actualizando los descubrimientos de Egan-Hake (Ref-1), de Ehmer (Ref-2 y 3). A su vez hemos introducido el tema del enmascaramiento temporal

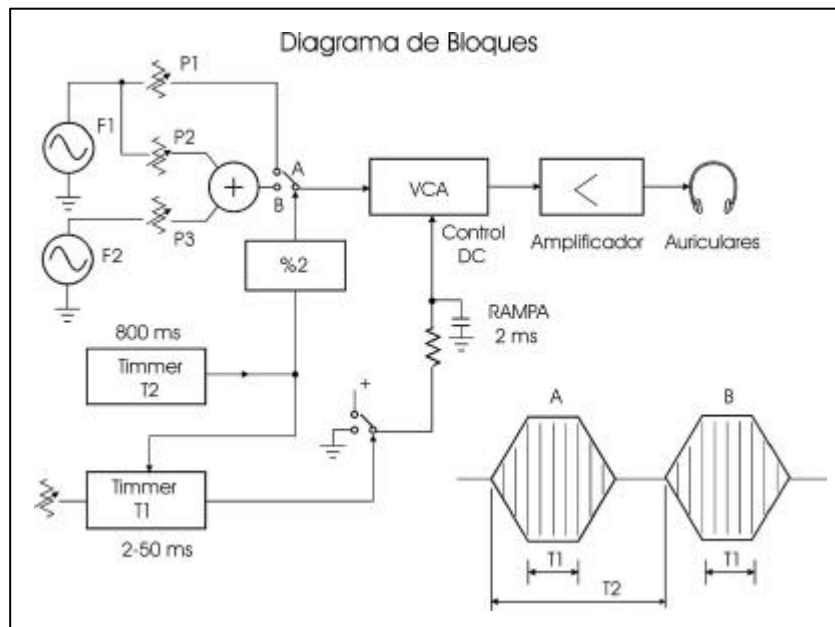
incorporando las ideas de Hirsh (Ref-4), de Zwislocki (Ref-5), de Bacon (Ref-6) y de Moore (Ref-7).

Con la teoría existente hacia 1985 creamos un sistema de compresión digital de audio que resultó consistente y aún en uso.

Sin embargo desde 1995, intentamos crear compresiones digitales mayores, manteniendo una baja audibilidad de la distorsión introducida en este proceso. Esto puede realizarse mediante encoders basados no solamente en el enmascaramiento de tonos, sino en el concepto que denominamos: *cuadros tonales deslizantes*. Esto es algo parecido a lo que ocurre con la persistencia de la retina del ojo humano que ha posibilitado la creación de la cinematografía. Es evidente que el oído no tiene esa propiedad y por lo tanto no puede ser engañado de la misma manera. Pues 24 fotos por segundo son para el ojo "movimiento". Mientras que una sucesión de "cuadros sonoros" son para el oído una confusión tonal que no representa el flujo sonoro.

Sin embargo, persiguiendo a esa idea, analizamos e hicimos numerosas investigaciones acerca de la manera en que el cerebro requiere de tiempo para analizar la información proveniente de la membrana basilar, para discriminar la presencia de tonos. Si la legendaria capacidad del oído para detectar tonos secundarios en presencia de un intenso tono principal (a condición de no ser enmascarado por este), requería de "tiempo de análisis" por parte del cerebro, es evidente que podemos reducir la cantidad de información transferida por el canal de comunicación digital, pues solo enviaremos la cantidad mínima necesaria en cada *cuadro* para evitar que el oído reconozca una diferencia con el sonido original. En cuadros sucesivos podemos ampliar la información a medida que el cerebro la vaya procesando y mejore su resolución.

EXPERIMENTO PARA MEDIR EL EFECTO DE ENMASCARAMIENTO DE RAFAGAS



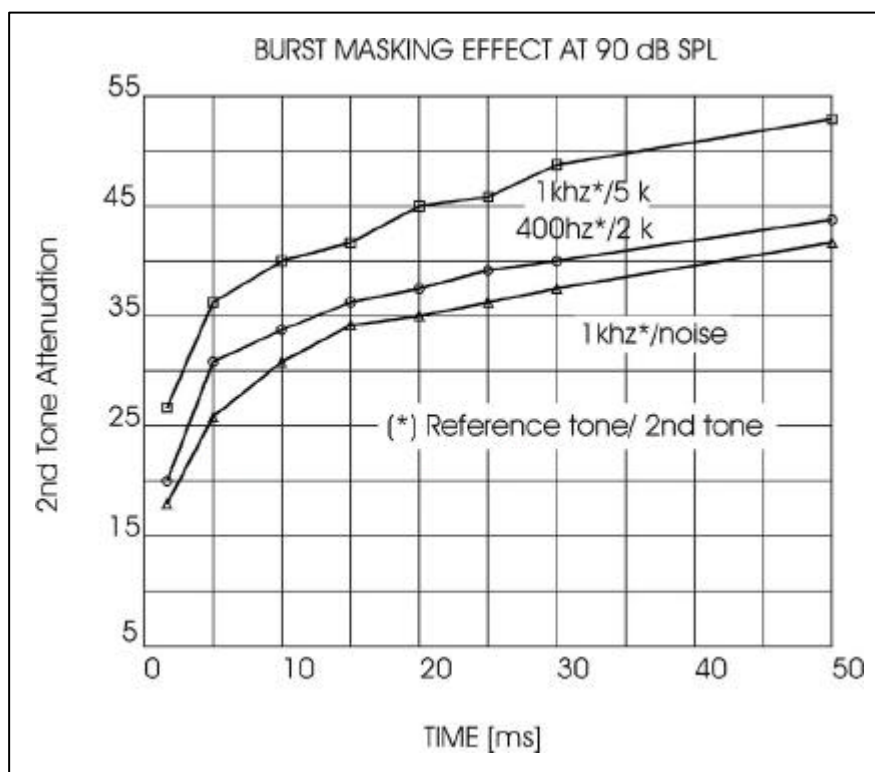
Dado que no existía información teórica que soportara nuestra idea de operar con *cuadros*, debimos recurrir a nuestra propia experimentación para ver si este era un camino viable. Se intentó analizar si durante un cuadro de muy corta duración, la capacidad del oído

para discriminar una suma de tonos se reducía lo suficiente como para permitir transmitir menor cantidad de información.

Podemos ver en la Fig. 1 un diagrama en bloques del sistema diseñado para realizar el experimento. La idea es comparar dos ráfagas de idéntica duración. La primera, denominada "A" contiene solamente el tono dominante f_1 (en nuestra experiencia 400 hz y 1 khz). La segunda ráfaga "B" se presenta a continuación de la primera para poder compararlas y contiene además del tono dominante al segundo tono f_2 de menor nivel que f_1 . Para evitar que existan espectros debidos a la conmutación instantánea de las señales, que pueden invalidar el experimento, ambas ráfagas A y B son conmutadas mediante una rampa de 2 ms que comanda a un amplificador VCA (Voltage Controlled Amplifier) que aumenta o reduce su ganancia en forma progresiva.

Las 10 personas que participaron en el experimento, con auriculares Sennheiser, debían mover el atenuador P3 hasta el límite en que ambos pulsos sonaban igual; es decir al máximo valor de f_2 para el cual el pulso B sonaba igual que el A. A su vez el experimentador se aseguraba en cada experiencia (con P1 y P2) que el valor eficaz de ambos pulsos fuera igual.

Los 10 sujetos tenían edades que variaban entre 18 y 52 años. La dispersión entre los 10 sujetos resultó reducida : $\pm 3,5$ dB por lo que se consideró que el efecto estaba claramente determinado y que la edad casi no tenía influencia.



En la Fig. 2 pueden verse los resultados para dos diferentes relaciones de frecuencia y para ruido rosado. En abscisas tenemos el tiempo de la ráfaga en milisegundos. En ordenadas tenemos la atenuación en decibelios, del tono f_2 con respecto al dominante f_1 , necesaria para que el oído no detecte la existencia de f_2 ; o dicho en otros términos para que f_2 sea enmascarado por f_1 . Este enmascaramiento, sin embargo, no es *enmascaramiento de tonos*

(pues ambos están separados por más de dos octavas) sino el nuevo fenómeno de *enmascaramiento por ráfaga (burst masking)*.

La curva "A" tiene 1 khz como tono dominante y 5 khz como segundo tono o tono enmascarado. La curva "B" tiene 400 hz como tono dominante y 2 khz como segundo tono. Finalmente la curva "C" tiene 1 khz como tono dominante y ruido rosado como tono enmascarado. Todos los ensayos se han hecho para 90 dB SPL de nivel sonoro pues consideramos que ese es el nivel aproximado al cual se juzga la calidad sonora de un sistema de audio en el hogar.

CONCLUSIONES

Puede notarse que en la curva de 1 khz / ruido, el oído distingue la presencia de ruido (que bien puede ser ruido de cuantización) que está a menos de 40 dB por debajo del tono principal, si le damos 50 milisegundos para analizarlo. En cambio en 5 milisegundos, solamente es capaz de advertir ruidos que superen los 25 dB. Esto implica que el umbral de detección empeoró unos 15 dB.

En el caso de un tono de 1 khz enmascarando a otro de 5 khz (curva superior) se hace evidente la mayor sensibilidad del oído en 5 khz, pues si le damos 50 mseg podrá distinguir un tono de 5 khz que se encuentre 53 dB por debajo del de 1 khz. Hacemos notar que el *enmascaramiento tonal* antes visto no afecta prácticamente estos resultados, pues tenemos más de 2 octavas de separación entre tonos, con lo cual casi no hay cambio de umbral en 5 khz.

Al reducir el tiempo de análisis del oído a solamente 5 mseg, se hace necesario que el tono de 5 khz alcance niveles de -36 dB para ser reconocido. En este caso el umbral de detección empeora 17 dB debido al enmascaramiento de ráfaga.

Estas conclusiones tienen varios campos de aplicación, además del mencionado en la compresión digital de información de audio. Otro campo es el procesado de señales en comunicaciones. Es sabido que uno de los recursos para mejorar la energía promedio de una señal de audio para su transmisión, implica el recorte de los picos. La curva de la Fig-2 explica los motivos por los cuales es posible lograr grandes niveles de recorte en picos de breve duración, sin que el oído perciba la distorsión armónica o de intermodulación.

REFERENCIAS

- 1) J.P.. Egan - H.W.Hake, J.Acoust.Soc.Am. Vol 22, 622-630, 1950
- 2) Richar Ehmer, J.Acoust.Soc.Am. Vol 31, 1253-1256, 1959
- 3) Richard Ehmer, idem Vol 31, 1115-1120, 1959
- 4) Ira Hirsh, J.Acoust.Soc.Am. Vol 31 759-767, 1959
- 5) J.J.Zwislocki, J.Acoust.Soc.Am. Vol 46, 431-441, 1969
- 6) Sid Bacon and Neal Viemeister, J.Acoust.Soc.Am. Vol 78, 1220-1230, 1985
- 7) Brian Moore, J.Acoust.Soc.Am. Vol 78, 488-494, 1985