

**UNIVERSIDAD POLITECNICA DE VALENCIA**

**ESCUELA POLITECNICA SUPERIOR DE GANDIA**

**Master en Ingeniería Acústica**

---



**UNIVERSIDAD  
POLITECNICA  
DE VALENCIA**



**ESCUELA POLITECNICA  
SUPERIOR DE GANDIA**

# **Efecto de los difusores acústicos en el timbre de los sonidos**

## **Cálculo del Tiempo de Ataque Perceptual**

***TESIS DE MASTER***

Autor:  
**Fernando Bejarano Durán**

Director/es:  
***D. Rubén Picó Vila***  
***D. F. Javier Redondo Pastor***

***GANDÍA, 2009***

# **Efecto de los difusores acústicos en el timbre de los sonidos**

## **Cálculo del Tiempo de Ataque Perceptual**

*Autor:* Fernando Bejarano Durán

*Director1:* D. Rubén Picó Vila

*Director2:* D. Francisco Javier Redondo Pastor

*Resumen* — El Tiempo de Ataque Perceptual (PAT) es uno de los parámetros que tiene más influencia sobre el timbre de los sonidos. La mayoría de estudios que se han centrado en esta característica han utilizado métodos perceptuales para calcularla. En este proyecto se desarrolla un método para calcular dicho parámetro a partir de la envolvente temporal de las señales sonoras, y estudiar el efecto que produce sobre dicho parámetro el uso de difusores acústicos, en un primer intento de observar el comportamiento temporal de éstos.

*Abstract* — The Perceptual Attack Time (PAT), is one of the parameters that has more influence on the timbre of the sounds. The majority of studies that have centred on this feature, have used perceptual methods to calculate it. In this paper a method was developed to calculate this parameter using the temporal envelope of sound signals, and in this way, we go to study the effect that produces on the above mentioned parameter, the use of acoustics diffusers, in a first attempt of observing the temporary behavior of these.

## INDICE

<b>I. Introducción y Objetivos</b> .....	2
<b>II. Introducción teórica</b> .....	4
II.1. El timbre.....	4
II.1.1. Parámetros basados en aspectos frecuenciales.....	5
II.1.2. Parámetros basados en aspectos temporales.....	7
II.2. Tiempo de ataque Perceptual.....	7
II.2.1. Introducción.....	7
II.2.2. Cálculo de la Envolvente temporal y Tiempo de Ataque Perceptual.....	10
A. Proceso de detección de ataque.....	12
B. Modelos de detección y síntesis.....	14
B.1. Modelos basados en características de la señal.....	14
B.1.1. Características temporales.....	14
B.1.2. Características espectrales.....	15
B.1.3. Características utilizando la fase.....	16
B.2. Modelos basados en métodos probabilísticos.....	17
II.3. Difusores acústicos.....	18
II.3.1. Efecto del cambio de fase en el dominio frecuencial.....	20
II.3.2. Efecto del cambio de fase en el dominio temporal.....	22
II.3.3. Parámetros que caracterizan a los difusores.....	23
<b>III. Método Experimental</b> .....	24
III.1. Cálculo de la envolvente temporal.....	24
III.2. Calculo del Tiempo de Ataque Perceptual.....	27
<b>IV. Análisis de las señales</b> .....	29
IV.1. Señales de Audio.....	29
IV.2. Respuesta al impulso de difusores.....	30
IV.3. Tiempo de Ataque Perceptual .....	31
IV.3.1. Tiempo de Ataque Perceptual de tonos musicales .....	32
IV.3.2. Tiempo de Ataque Perceptual de tonos musicales en entornos difusivos.....	36
<b>V. Conclusiones y líneas futuras</b> .....	45
<b>Agradecimientos</b> .....	46
<b>Bibliografía</b> .....	47

## I. INTRODUCCIÓN Y OBJETIVOS.

Desde los primeros comienzos de la acústica como ciencia, se han venido empleando diferentes dispositivos que favorecieran la audición de mensajes sonoros en aquellos recintos destinados a su uso. Si bien muchos de ellos tenían un gran peso estético, con el paso del tiempo se ha descubierto que esa “casualidad” o no, tenía una función acústica mucho más importante.

Desde las vasijas de bronce afinadas empleadas en los teatros romanos que actuaban como resonadores, que en realidad redirigían el sonido hacia un lugar diferente del que provenían, pasando por los techos y paredes artesonados de los teatros de ópera barrocos, hasta los más sofisticados sistemas de absorción y dispersión del sonido de la actualidad, todos tenían un mismo objetivo, asegurar dentro de la estancia unas condiciones óptimas para una recepción correcta del sonido.

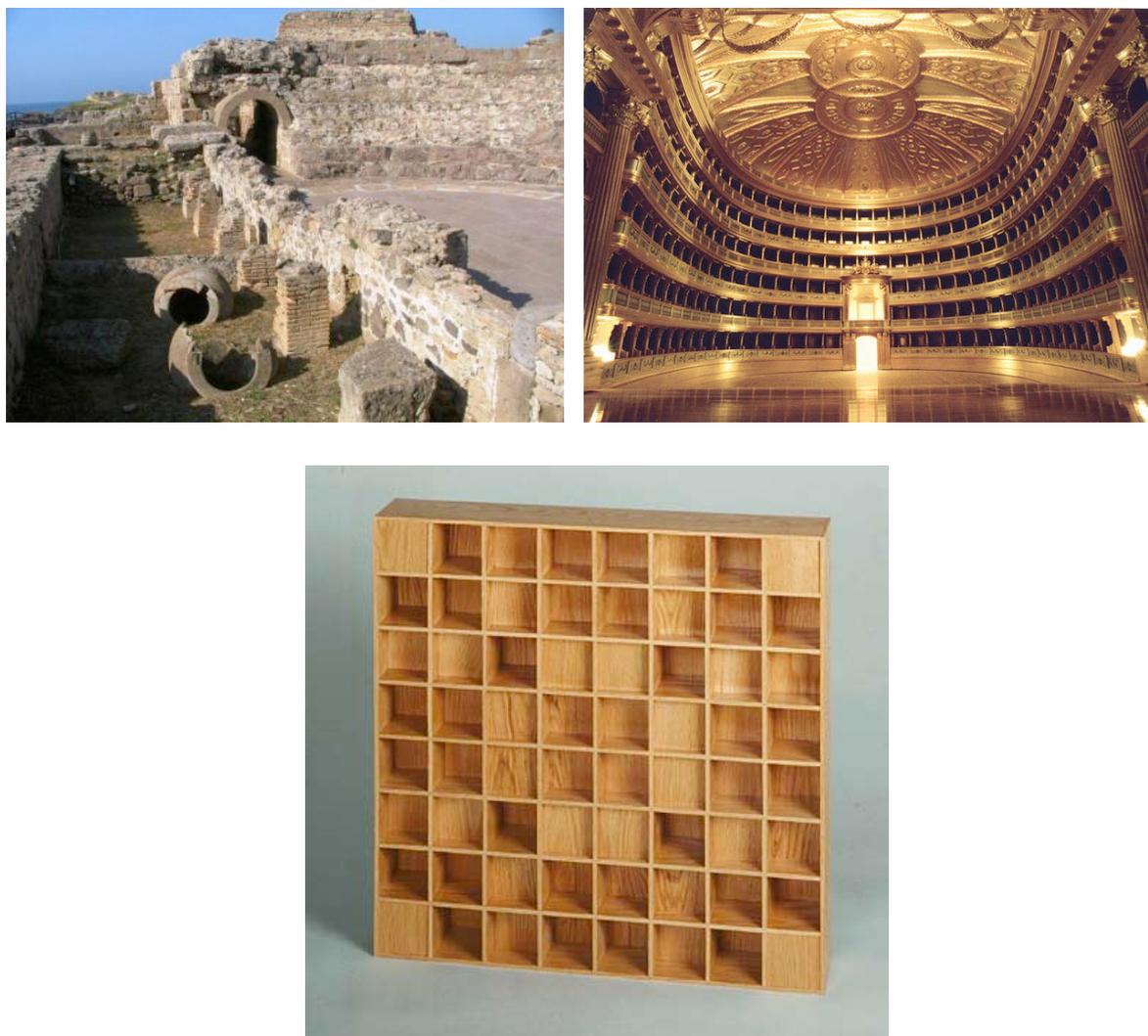


Fig.1. Arriba a la izquierda imagen del teatro romano de Nora en Cerdeña con imágenes de algunas de las vasijas encontradas. Arriba a la derecha, detalle del artesanado del teatro “alla Scala” de Milán. Abajo imagen de un difusor *QRD*.

Son los difusores acústicos uno de estos elementos que hoy día está siendo objeto de numerosos estudios para poder comprender de manera exhaustiva su comportamiento, y poder así diseñar dispositivos que ayuden a construir recintos con unas cualidades acústicas adecuadas a su uso.

Hasta ahora la mayoría de dichos estudios se han basado en las propiedades frecuenciales de los difusores y en la manera que éstos dispersaban espacialmente la energía sonora incidente, pero no se tiene un profundo conocimiento de lo que ocurre con la señal sonora en el dominio del tiempo después de que sea reflejada por este dispositivo.

Conocer cómo afecta temporalmente un difusor a una señal sonora es de vital importancia, porque podríamos estar ante un difusor que espacialmente distribuyera la energía de un sonido de manera eficiente, pero modificara algunos parámetros que afectaran a propiedades temporales del mismo.

El timbre es la cualidad perceptual que permite la distinción entre diferentes sonidos, es la marca de identidad de cualquier señal sonora. Esta cualidad tiene un carácter multidimensional abarcando tanto propiedades frecuenciales como temporales. Es en este último dominio, en el temporal, donde encontramos un parámetro clave dentro del timbre, el “tiempo de ataque perceptual”.

El objetivo de este presente proyecto es obtener el tiempo de ataque perceptual de señales sonoras utilizando la envolvente de la misma y comprobar si la utilización de difusores acústicos modifica de alguna manera dicho parámetro después de que el sonido haya sido reflejado en la superficie de éstos.

## II. INTRODUCCIÓN TEÓRICA.

### II.1. *EL TIMBRE.*

El timbre, según la definición propuesta por la ASA en 1960, es la cualidad para distinguir dos sonidos con la misma frecuencia fundamental, loudness y duración [1], aunque esta definición es muy inexacta puesto que define lo que el timbre no es, no lo que el timbre es.

Posteriormente el Instituto Nacional de Estándares Americano (ANSI) elevó la cualidad anterior a un atributo perceptual, considerándolo como una “herramienta” para la distinción entre dos sonidos diferentes sea cual sea su naturaleza [2].

A lo largo de los años se han formulado numerosas definiciones del timbre sin llegar a obtener una que pueda realmente englobar a todos y cada uno de los aspectos que forman parte de esta cualidad sonora.

Muchas de éstas hacen hincapié en los aspectos frecuenciales de los sonidos que tienen influencia en el timbre, como es el número, amplitud y espaciado de las componentes espectrales a modo de “marca de identidad” exclusiva de un sonido, dejando un poco de lado la envolvente y transitorios temporales de la señal. Pero el más esclarecedor ejemplo viene cuando se reproduce un sonido de atrás hacia adelante en el tiempo. El espectro frecuencial del sonido es el mismo ya sea reproducido en una dirección o en otra, pero a pesar de ello los sonidos cambian drásticamente obteniéndose una percepción muy diferente a la que debería ser. En este caso la envolvente temporal ha cambiado y por ello el sonido pierde su identidad original demostrando la importancia de la envolvente temporal en la determinación del timbre.

Todo esto nos da una idea de la relación que existe entre el timbre y la identificación de una fuente sonora, relación que ha sido objeto de numerosos estudios sobre el reconocimiento de instrumentos basados en el timbre después de varias modificaciones de las señales acústicas originales, como los de Berger [3] y Saldanha [4] en 1964 y Hadja en 1999 [5]. Este último mostró que para tonos de una duración breve (para instrumentos impulsivos) se necesita la estructura temporal total del tono para identificar el instrumento, mientras que para tonos de mayor duración, la parte sostenida (estacionaria) del tono es suficiente para un reconocimiento perceptual del origen del mismo.

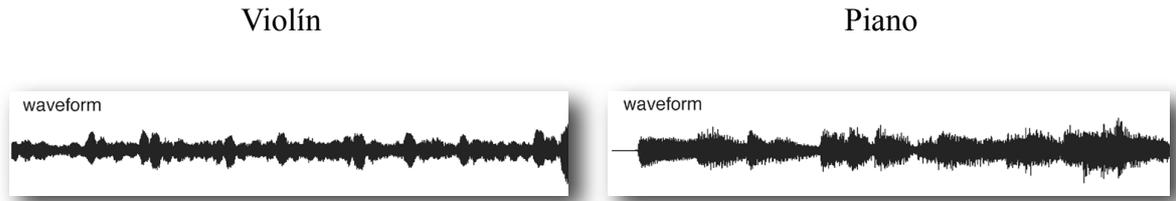


Fig.2. Formas de onda del sonido emitido por un violín y un piano.

Otros estudios se centraron en la primera parte de la envolvente temporal, desarrollándose numerosas experiencias en las que se mostraban a músicos profesionales sonidos de instrumentos a los que se les había eliminado el primer medio segundo de sonido. En estos casos hubo resultados sorprendentes donde los propios músicos confundían sonidos de instrumentos con una gran facilidad.

Strong y Clark en 1967 [6] realizaron una serie de experimentos con músicos donde generaban sonidos con el espectro frecuencial de un instrumento y la envolvente temporal de otro. En algunos casos (oboe, tuba, fagot y clarinete) encontraron que el espectro frecuencial era el más importante para la identificación del instrumento, mientras que en otros casos como la flauta, la envolvente temporal era el principal factor para el reconocimiento del instrumento.

Es por esto que está ampliamente asumido que el timbre es una cualidad o característica de los sonidos que tiene una naturaleza multidimensional, donde algunas de dichas dimensiones tienen su espacio en la envolvente espectral, envolvente temporal, etc. La dificultad a la hora de definir de manera exacta la identidad del timbre es debido a que muchos de los parámetros del mismo son más similares procediendo de sonidos de diferentes instrumentos con la misma frecuencia fundamental, que de sonidos de un mismo instrumento con diferente frecuencia fundamental.

Sobre el timbre, que se considera una cualidad perceptual, se han definido una serie de parámetros físicos, de los cuales se han comprobado experimentalmente el grado de relación con esta cualidad sonora. Estos parámetros los podemos dividir en dos grupos según el dominio donde se engloban: frecuencial o temporal.

### II.1.1. *PARÁMETROS BASADOS EN ASPECTOS FRECUENCIALES.*

Estos parámetros se basan en el estudio de las componentes frecuenciales de la señal (armónicos) y en cierta medida su relación con la frecuencia fundamental de la onda sonora.

**Brightness** .- se calcula como el *spectral centroid* [7], y tiene correlación con la cualidad subjetiva del brillo [8]. Viene desarrollado como:

$$brightness = \left( \sum_{k=1}^N ka_k \right) / \sum_{k=1}^N a_k \quad (1)$$

donde  $k$  es el índice del parcial  $k$ -ésimo y  $a_k$  es la amplitud del parcial  $k$ .

**Tristímulus** .- Fue desarrollado por Pollar y Jansson en 1982 [9] y se define en el timbre de manera equivalente a los atributos del color en la visión. Determinadas aplicaciones lo utilizan para observar el comportamiento transitorio de los sonidos musicales aunque en general se utiliza para analizar la fuente espectral y componentes principales de los sonidos.

$$tristímulus1 = \frac{a_1}{\sum_{k=1}^N a_k} \quad tristímulus2 = \frac{a_2 + a_3 + a_4}{\sum_{k=1}^N a_k} \quad tristímulus3 = \frac{\sum_{k=5}^N a_k}{\sum_{k=1}^N a_k} \quad (2)$$

siendo  $a_k$  el nivel del armónico  $k$ -ésimo.

**Relación Odd/Even** .- Algunos estudios como Benade y Kouzoupis [10] se han servido de esta relación para observar la carencia de energía en determinadas secciones del espectro de diferentes instrumentos.

$$odd = \left( \sum_{k=2}^{N/2} a_{2k-1} \right) / \sum_{k=1}^N a_k \quad (3)$$

$$even = \left( \sum_{k=1}^{N/2} a_{2k} \right) / \sum_{k=1}^N a_k \quad (4)$$

**Irregularidad**.- Estudia la mayor o menor regularidad en los espectros de los sonidos analizados. Fue definida por Krimphoff *et al.* en 1994 [11] como la suma de la amplitud de un armónico menos la media entre el armónico anterior, el mismo y el posterior.

$$irregularity = \sum_{k=2}^{N-1} \left| a_k - \frac{a_{k-1} + a_k + a_{k+1}}{3} \right| \quad (5)$$

### II.1.2. *PARÁMETROS BASADOS EN ASPECTOS TEMPORALES.*

La envolvente de amplitud de sonidos musicales ha sido objeto de muchos estudios, obteniéndose el “tiempo de ataque perceptual” del sonido como el parámetro más significativo en el dominio temporal, con una relación clara con el timbre del mismo. A lo largo de los años, algunos investigadores han llegado a proponer una clasificación de ataques según su género.

Podríamos definir el tiempo de ataque perceptual de un sonido como el tiempo que pasa desde que se produce la perturbación física producida por la onda sonora en el medio (en este caso en el aire), hasta el momento en el que es percibido por el oyente.

Debido a su fuerte correlación con la percepción del sonido, es uno de los parámetros con mayor peso dentro del timbre puesto que la llegada a nuestro sistema auditivo de la señal y el comienzo de esta hace que podamos diferenciar, de manera clara el timbre de dos instrumentos distintos.

## II.2. *TIEMPO DE ATAQUE PERCEPTUAL.*

### II.2.1. *INTRODUCCIÓN.*

La percepción del ritmo musical implica alguna clase de mecanismo auditivo para detectar eventos rítmicos. Ya que el inicio y la parte de ataque de un sonido son sus rasgos rítmicos más importantes, la base de este mecanismo es presumiblemente la llegada y detección de transitorios.

El tiempo de ataque perceptual, o PAT (*Perceptual Attack Time*), se define como el tiempo en el que ocurre el ataque de un tono, o lo que es lo mismo, el tiempo que transcurre desde que se produce el comienzo físico de la señal hasta que es percibido por el oyente.

El PAT es dependiente tanto del tiempo de subida como del nivel de amplitud, indicando que la pendiente de la función de subida es un factor clave. La duración y el espectro pueden actuar también como indicadores del PAT, pero con menos influencia que la amplitud y la intensidad.

Experimentalmente se han medido PATs de grabaciones de instrumentos de orquesta mediante la utilización de ordenadores y sintetizadores de música. Con estos estudios se lograron testear algunos modelos de predicción de PAT comparándolos con los datos empíricos. El modelo más acertado tiene una correlación entre los datos medidos y los datos predichos de 0.995 [12].

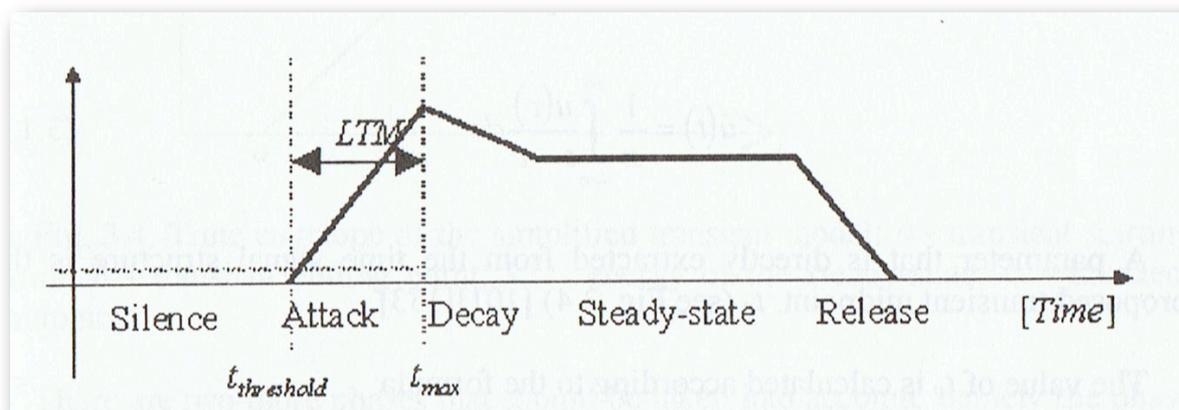


Fig.3. Diagrama de las fases temporales de un tono, observándose el tiempo de ataque perceptual.

Por otro lado se puede hacer una distinción entre el PAT y *tiempo de llegada perceptual*. El *tiempo de llegada perceptual* se podría definir como el tiempo en el que se percibe la llegada del sonido pero no se corresponde con el nivel máximo de la señal, es decir, posteriormente se percibe un “ataque” de la señal definido por un aumento leve pero significativo del nivel sonoro. Estos dos fenómenos normalmente aparecen simultáneamente, pero para algunos tonos dentro de un contexto musical es posible escuchar ambos desfadados, primero percibir la llegada del sonido y posteriormente el “ataque”, como puede suceder en instrumentos con lengüeta, metales con sordina e instrumentos de cuerda con arco.

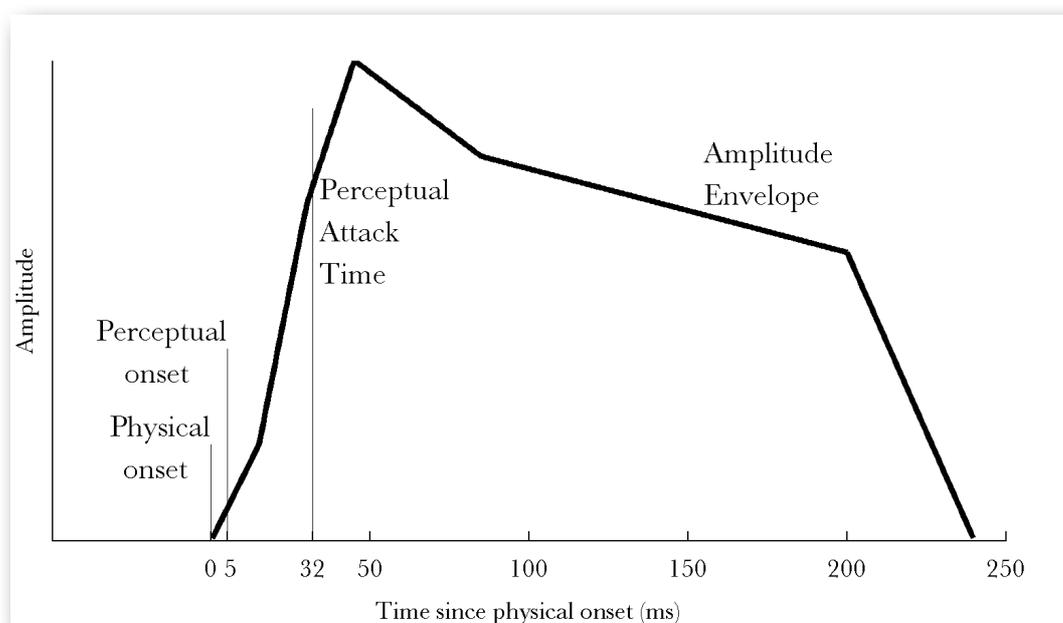


Fig.4. Descripción temporal de la llegada física, llegada física perceptual y el PAT.

En la práctica el PAT sólo es útil como referencia respecto al tiempo de llegada física del sonido o en relación con PATs de otros sonidos. Así por lo tanto se pueden introducir dos nuevos términos, el primero se definiría como el intervalo temporal entre la llegada física del sonido y el PAT, denominado *tiempo de ataque perceptual relativo* (RPAT) [12] y el segundo término sería el  $\Delta$ PAT que es la cantidad de tiempo necesario para cambiar un tono desde un sincronismo/isocronismo físico (con algún tono estándar) hasta obtener un sincronismo/isocronismo perceptual (con ese estándar).

El isocronismo y sincronismo son dos métodos para medir el PAT. En ambos casos el sujeto escucha un sonido en un bucle y tiene que ajustar el tiempo de llegada del sonido con respecto a un sonido de referencia. En el caso isócrono el objetivo es obtener una alternancia rítmica constante de los dos sonidos. En el caso síncrono el sujeto ajusta el tiempo para percibir los dos sonidos como si se ejecutaran al mismo tiempo. El sincronismo/isocronismo físico tiene relación con el tiempo de llegada físico del sonido, mientras que el sincronismo/isocronismo perceptual depende en su totalidad del PAT.

El RPAAT de un tono debe ser un número positivo, por ello el ataque perceptual nunca precede a la llegada física del sonido. Por otro lado el  $\Delta$ PAT de un tono puede ser negativo, positivo o cero, dependiendo de lo que se elija como estándar.

No todos los sonidos musicales son suficientemente rítmicos; por esto el PAT no siempre queda claramente definido, o no se llega a percibir. En la ejecución rítmica se llegan a encontrar desviaciones no intencionadas sobre la isocronía absoluta, las cuales no sólo son debidas a las inherentes limitaciones en la interpretación de las notas, si no también a las limitaciones perceptuales o tolerancias. Se pueden observar también que algunas clases de instrumentos (por ejemplo los instrumentos de metal), se pueden sincronizar mejor que otros (los instrumentos de cuerda). Rasch en 1979 [13] especuló que los músicos que tocaban instrumentos de cuerda escalonaban intencionadamente sus ataques para así clarificar mejor determinadas partes polifónicas; por lo tanto, si el PAT no está definido de manera nítida, una pequeña imprecisión no será suficientemente importante como para impedir que se sientan los ataques como síncronos. La consecuencia es que el PAT es esencialmente instantáneo para los instrumentos de metal y más duradero para los instrumentos de cuerda.

Con esto se puede ver una relación inversa entre el tiempo de subida y cómo de claro está definido el PAT, siendo el tiempo de subida un parámetro bastante relevante dentro del tiempo de ataque perceptual.

Se podría también esperar que entrara en juego una integración temporal desde la llegada del sonido, donde la intensidad es baja [14], [15]. Esto implica que el cambio de energía por unidad de tiempo, o la pendiente de subida, es importante, la cual es a su vez dependiente del nivel de presión sonora total. Esto fue apoyado por Wright en 1960 [16] que observó que el umbral de ruido o no ruido durante el tiempo de subida era dependiente del nivel de presión sonora total. En realidad si el PAT queda claramente definido con ataques impulsivos, se puede esperar que con tiempos de subida más largos que los umbrales de intensidad relacionados, el PAT se vuelve poco claro o no significativo. En estos casos se tiene percepción de la llegada del sonido pero no del ataque.

Cutting y Rosner [17] propusieron que ese tiempo de subida continuo inducía percepciones categóricas, con la distinción del “pluck-bow” ocurriendo en torno a los 40 ms, aunque estudios posteriores desmintieron estas aseveraciones.

## II.2.2. CALCULO DE LA ENVOLVENTE TEMPORAL Y TIEMPO DE ATAQUE PERCEPTUAL.

Derivar una fórmula para predecir o modelizar el PAT conlleva seleccionar una apropiada representación del sonido, y tenerla en cuenta a la hora de elegir aquellos factores que pueden influir en el PAT para obtener una aproximación realista.

La representación del sonido puede ser la propia forma de onda, o alguna transformada de la misma. Las características a tener en cuenta como umbrales u otros parámetros cuyos valores no pueden ser deducidos pueden ser determinados empíricamente.

Hay numerosos modelos que han sido probados, muchos de los cuales son consistentes de una manera u otra desde una perspectiva auditiva. La representación del sonido utilizada era la envolvente en amplitud, intensidad o loudness.

La envolvente en amplitud se basó simplemente en tomar el máximo valor de 82 segmentos sucesivos procedentes de la forma de onda original muestreada. (La longitud de cada segmento se corresponde con el periodo fundamental) Ya que el valor más positivo y el más negativo de cada segmento no siempre son idénticos en amplitud, se usaba la media de ambos valores. La envolvente resultante tenía una tasa de muestreo igual a la frecuencia fundamental, o alrededor 312. Posteriormente Schafer y Rabiner en 1973 [18] utilizaron una técnica de remuestreo para obtener una envolvente con una muestra por cada milisegundo. Por su parte, la envolvente de intensidad se obtuvo elevando al cuadrado la función envolvente de amplitud.

Zwicker desarrollo unos modelos para obtener la envolvente del loudness [19] [20].

Existen varios modelos de detección del tiempo de ataque:

- **Time of max.** El PAT se obtiene midiendo en la función envolvente cuando se alcanza el valor máximo, por lo tanto se busca el máximo. Para tener un valor de tiempo de subida RPAT independiente del nivel sonoro global, se suele utilizar el logarítmico de dicho tiempo.

$$\text{LRPAT} = \log_{10} (T_{\text{max}} - T_{\text{llegada}}) \quad (6)$$

- **Absolute threshold.** Este modelo prueba la teoría de que el ataque perceptual es el mismo que la llegada perceptual. El PAT se modela como el tiempo en el que la envolvente cruza un umbral absoluto (usando una escala normalizada).

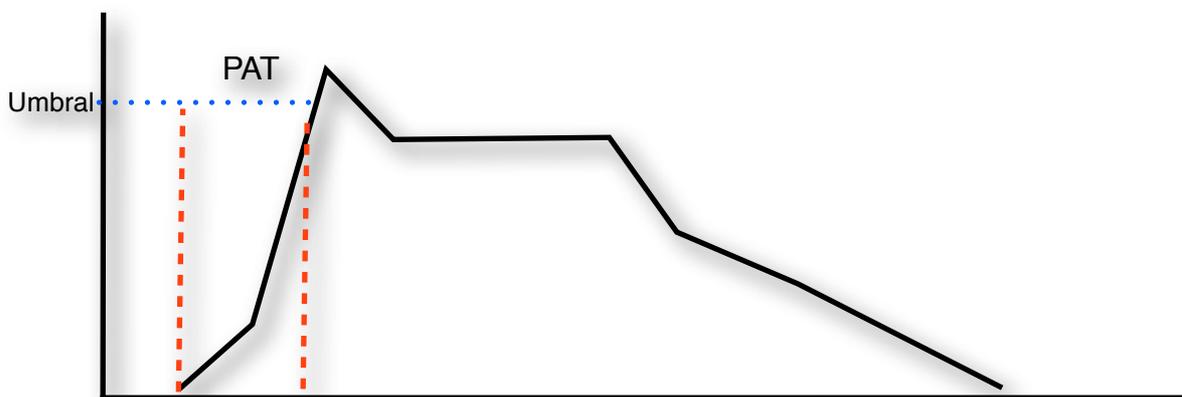


Fig.5. Ejemplo gráfico del método Absolute threshold.

- **Percent of Max.** Vos y Rasch (1981) [21] desarrollaron un modelo que predice el tiempo de ataque perceptual como el momento en el que la amplitud de la envolvente cruzaba un umbral relativo a su máximo valor, expresado como un porcentaje del máximo o como una cantidad de decibelios por debajo del mismo.
- **Energy.** Este modelo considera la distinción entre la llegada perceptual y el ataque perceptual y está basado en la energía y no la amplitud o la intensidad, modelando el PAT como el tiempo de integración de la envolvente cruzando algún umbral.

### A. PROCESO DE DETECCIÓN DE ATAQUES.

La manera habitual de detectar la llegada de una señal consiste en buscar regiones transitorias en la misma, una noción que lleva a muchas definiciones; un repentino arranque de energía, un cambio en el espectro de la señal en un corto periodo de tiempo o en las propiedades estadísticas, etc.

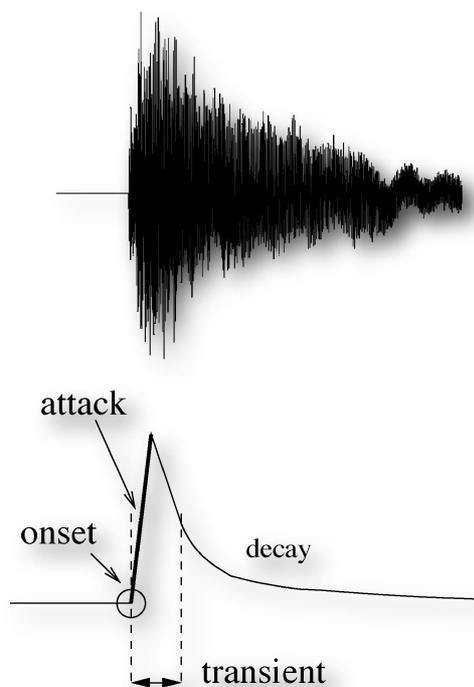


Fig.6. Señal y gráfica de la envolvente para observar la llegada física y ataque.

Un aspecto importante es hacer una clara distinción entre conceptos relacionados entre sí como son los transitorios, llegadas de señal y ataques. La razón por la que es importante es porque diferentes aplicaciones tienen distintas necesidades y cada una requiere de un conocimiento más profundo de uno de estos parámetros.

En el caso de una señal polifónica real hay múltiples objetos sonoros presentes al mismo tiempo en un instante dado, por lo tanto la distinción anterior es en muchos casos complicada e incluso inalcanzable. En general no es posible detectar ataques de señal sin cuantificar primero las variaciones de tiempo de la señal “transitorios”.

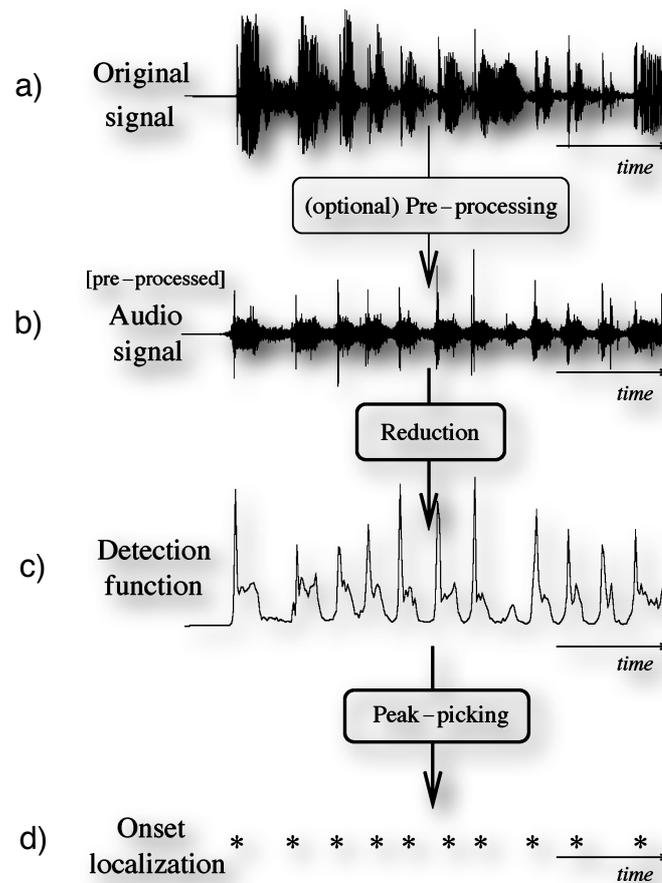


Fig.7. Diagrama de un algoritmo de detección de llegadas estándar. a) señal original. b) Señal después de un preprocesado. c) Señal después de una reducción. d) señal después de un algoritmo peak-picking.

La Fig.7 ilustra el procedimiento empleado en la mayoría de algoritmos de detección de ataques: de la señal original de audio, que puede ser preprocesada para mejorar el funcionamiento de etapas subsecuentes, se aplica una función de detección de picos para obtener así los valores temporales de los ataques.

Diversos estudios de detección de ataques de señal han encontrado útil el análisis independiente de información en varias bandas de frecuencia. En algunos casos ese preprocesado es necesario para satisfacer las necesidades de algunas aplicaciones específicas que requieren la detección individual en subbandas para complementar la estimación global, en otros se puede justificar dicha aproximación como un camino para incrementar la robustez del método de detección.

## B. *MODELOS DE DETECCIÓN-SÍNTESIS.*

El proceso de separación entre la parte transitoria y estacionaria de la señal se asocia normalmente al modelado de señales musicales. Hay una fina línea entre modelado y detección, por lo tanto, algunos modelos dirigidos a la representación de transitorios pueden servir para la detección de llegadas.

En el contexto de detección del tiempo de ataque, podemos dividir los métodos existentes en la actualidad en dos grupos: los métodos basados en el empleo de características de la señal explícitamente especificadas y los métodos basados en modelos de señal probabilísticos.

### B.1. *MODELOS BASADOS EN CARACTERÍSTICAS DE LA SEÑAL.*

#### B.1.1. *CARACTERÍSTICAS TEMPORALES.*

Cuando se observa la evolución temporal de una simple señal musical, es notable que la aparición de un ataque esté usualmente acompañada por un incremento de la amplitud de la misma. Los primeros métodos de detección utilizaban una función que seguía la envolvente de amplitud de la señal [22]. Así tenemos el “*envelope follower*” que se puede construir fácilmente rectificando y suavizando (por ejemplo con un filtro paso bajo) la señal:

$$E_0(n) = \frac{1}{N} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} |x(n+m)| w(m) \quad (7)$$

donde  $E_0$  es la función de envolvente de amplitud y  $w(m)$  es una ventana de  $N$  puntos o un suavizado Kernel, centrado en  $m=0$ . Esto produce resultados satisfactorios para ciertas aplicaciones donde existe unos transitorios percusivos fuertes con un background bastante discreto.

Una variación del método anterior permite obtener la energía local, en lugar de la amplitud, elevando al cuadrado, en vez de rectificar, cada muestra.

$$E(n) = \frac{1}{N} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} [x(n+m)]^2 w(m). \quad (8)$$

A pesar del suavizado, esta señal reducida en su forma sin refinar no es por lo general conveniente para la detección de ataques mediante la detección de picos (peak picking).

Un mayor refinamiento de este método, incluido en un gran número de algoritmos de detección de ataques, consiste en trabajar con la derivada temporal de la energía (o para señales en tiempo discreto la diferencia finita de primer orden), por lo que incrementos repentinos en la energía son transformados en picos estrechos. La energía y su derivada son comúnmente utilizadas en combinación con el preprocesado, ambos con bancos de filtros y separación entre transitorios y el estado estacionario.

### B.1.2. CARACTERÍSTICAS ESPECTRALES.

Aunque se reduce la necesidad de preprocesado (por ejemplo de la parte tonal) estos métodos son también exitosos en un gran número de casos, incluyendo la detección de ataques en señales polifónicas con múltiples instrumentos.

Consideremos la *short-time Fourier transform* (STFT) de la señal  $x(n)$ .

$$X_k(n) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(nh + m)w(m)e^{-\frac{2j\pi mk}{N}} \quad (9)$$

donde  $w(m)$  es otra vez una ventana de  $N$  puntos y  $h$  es el tamaño de salto, o tiempo de cambio entre ventanas adyacentes.

En el dominio espectral, los incrementos de energía debidos a transitorios tienden a aparecer como eventos de banda ancha. La energía de la señal normalmente se concentra a bajas frecuencias, por lo que cambios debido a los transitorios son más evidentes a altas frecuencias. Para enfatizar esto, el espectro puede ser ponderado preferiblemente hacia altas frecuencias antes de la suma para obtener una medida de la energía ponderada.

$$\tilde{E}(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} W_k |X_k(n)|^2 \quad (10)$$

donde  $W_k$  es la ponderación frecuencial. Según el teorema de Parseval, si  $W_k = 1 \forall k$ ,  $E(n)$  es simplemente la energía local como se ha definido previamente. Por otro lado la elección de  $W_k = k^2 \forall k$ , daría la energía local de la derivada de la señal.

Estas medidas espectralmente ponderadas están basadas en el espectro instantáneo a corto plazo de la señal, así se omite cualquier consideración explícita de su evolución temporal. Alternativamente, otras aproximaciones sí consideran dichos cambios, usando variaciones en el contenido espectral entre distintas ventanas para generar funciones de detección más informativas.

### B.1.3. CARACTERÍSTICAS UTILIZANDO LA FASE.

Recientes propuestas hacen uso de la fase del espectro para determinar el comportamiento de los ataques de señal. Esto tiene una gran relevancia ya que la mayor parte de la estructura temporal de una señal viene codificada en su fase.

Definimos  $\varphi_k(n)$  la  $2\pi$  - *unwrapped* fase del coeficiente  $X_k(n)$  de la STFT. Para ondas sinusoidales estacionarias, la fase  $\varphi_k(n)$ , así como la fase en las ventanas previas  $\varphi_k(n-1)$ , se utilizan para calcular un valor de la frecuencia instantánea. Una estimación de la frecuencia actual de la componente  $k$ -ésima dentro de esta ventana vendría definida como:

$$f_k(n) = \left( \frac{\varphi_k(n) - \varphi_k(n-1)}{2\pi h} \right) f_s \quad (11)$$

Donde  $h$  es el salto entre ventanas y  $f_s$  es la frecuencia de muestreo.

Es de esperar que para zonas locales estacionarias de señales sinusoidales, la frecuencia instantánea sea aproximadamente constante en ventanas adyacentes. Así, acorde con la ecuación anterior esto equivale a un incremento de fase constante de una ventana a otra ya que la fase es la misma.

$$\varphi_k(n) - \varphi_k(n-1) \simeq \varphi_k(n-1) - \varphi_k(n-2) \quad (12)$$

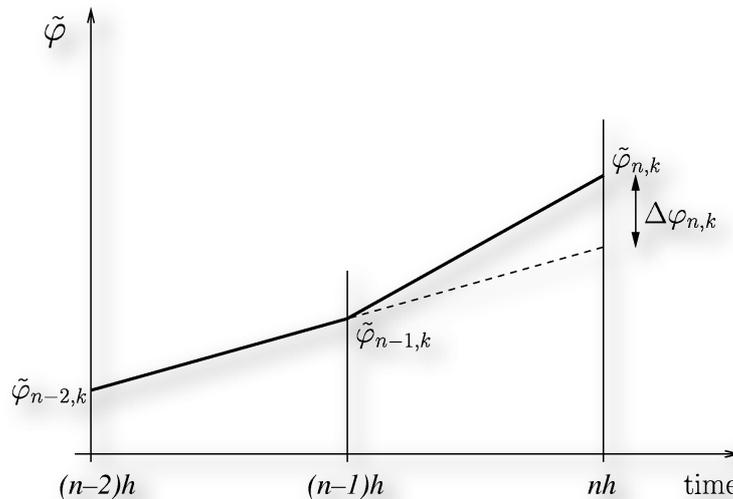


Fig.8. Gráfico que muestra la fase obtenida en dos ventanas adyacentes. Si la señal fuera estacionaria se obtendría un incremento de la misma constante (línea de puntos). Si hay un transitorio aparece un incremento no constante en la ventana adyacente.

Equivalentemente la desviación de fase puede ser definida como la segunda diferencia de la fase:

$$\Delta\varphi_k(n) = \varphi_k(n) - 2\varphi_k(n-1) + \varphi_k(n-2) \simeq 0 \quad (13)$$

Durante la región transitoria, la frecuencia instantánea no queda bien definida y por lo tanto  $\Delta\varphi_k(n)$  tenderá a ser mayor, como se puede ver en la Fig. 7.

## B.2. MODELOS BASADOS EN MÉTODOS PROBABILÍSTICOS.

Los métodos estadísticos para la detección de ataques se basan en la asunción de que la señal puede ser descrita por algún modelo probabilístico. Se puede entonces construir un sistema para que haga inferencias probabilísticas sobre posibles cambios abruptos de tiempo en la señal, dadas las observaciones disponibles. El éxito de esta propuesta radica en el gran parecido entre el modelo asumido, por ejemplo la distribución de probabilidad descrita por el modelo, y la “verdadera” distribución de los datos, y permitiendo ser cuantificado usando medidas de probabilidad y criterios de modelos Bayesianos.

Como ejemplo podemos introducir los modelos basados en métodos de detección de cambio de punto. Una propuesta conocida se basa en el test de proporción de probabilidad secuencial [23]. En él se presupone que muestras de la señal  $x(n)$  son generadas por uno de los dos modelos estadísticos,  $A$  o  $B$ .

$$s = \log \frac{p_B(x)}{p_A(x)} \quad (14)$$

donde  $p_{A}(x)$  y  $p_{B}(x)$  son las funciones de densidad de probabilidad asociadas a los dos modelos. La esperanza de  $s$  depende de qué modelo de señal se esté utilizando. Con el modelo  $A$ , la esperanza es:

$$E_A(s) = - \int p_A(x) \log \frac{p_A(x)}{p_B(x)} dx = -D(p_A \parallel p_B) < 0 \quad (15)$$

donde  $D$  denota la divergencia Kullback-Leibler entre el modelo y la distribución observada. Con el modelo  $B$  la esperanza es:

$$E_B(s) = \int p_B(x) \log \frac{p_B(x)}{p_A(x)} dx = D(p_B \parallel p_A) > 0 \quad (16)$$

### II.3 DIFUSORES ACÚSTICOS.

Los difusores acústicos son elementos arquitectónicos que tienen como función principal reflejar la energía incidente de ondas sonoras en todas direcciones, formando un campo acústico reflejado uniforme en el espacio. Una de las funciones principales para su uso es la eliminación de ecos provenientes de determinados puntos del auditorio, donde una reflexión especular llega al oyente dentro del conjunto reverberante con suficiente energía como para que pueda ser percibida individualmente, o también ecos flotantes provocadas por repetitivas reflexiones entre dos superficies paralelas.

Al contrario que los absorbentes, los difusores no eliminan energía del campo sonoro, simplemente la distribuyen lo más uniformemente posible para evitar que reflexiones especulares provoquen efectos no deseados.

A pesar de esto, el uso de difusores dentro de lugares que no sean auditorios, se basa en ciertas aplicaciones más en una intuición que en hechos científicos.

Una teoría estándar que explica los efectos de la difracción en procesos físicos fue propuesta en el siglo XVII por Huygen. Esta teoría no aproxima de manera fehaciente el fenómeno de *scattering* de la onda sonora en difusores acústicos, pero ayuda a explicar algunos de los procesos que ocurren en los mismos.

Este principio expone básicamente que todos los puntos de la superficie donde incide la onda se convierten a su vez en fuentes secundarias que emiten ondas con la misma frecuencia que la onda incidente. En este proceso hay que hablar de interferencia entre fuentes ya que cada fuente al ser puntual emite de manera esférica, pero la contribución de cada punto de la superficie conforma una interferencia que genera un frente de ondas común.

En el caso de una superficie totalmente plana, la onda reflejada tendrá el mismo ángulo que la onda incidente, es decir espacialmente no se verá alterada y no existirá dispersión.

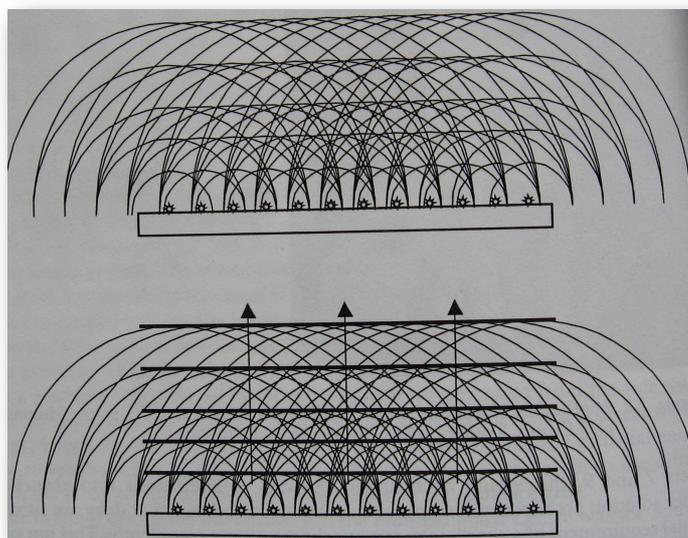


Fig.9. Reflexión en un difusor plano.

Por el contrario en los difusores que tienen una forma no tan regular, el proceso de reflexión especular va desapareciendo, dando lugar a una distribución espacial de la energía recibida. Como ejemplo tenemos el siguiente difusor de sección cilíndrica.

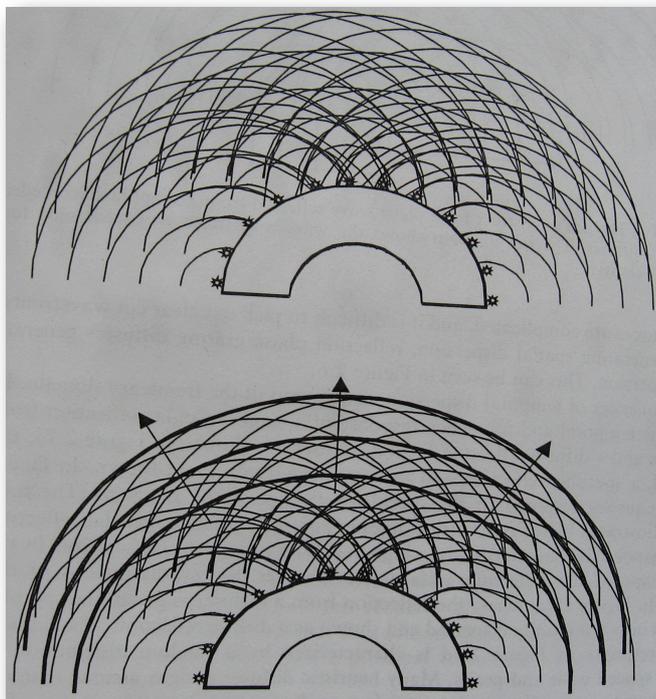


Fig.10. Reflexión en un difusor cilíndrico.

No existe el difusor perfecto, ya que aunque se pueda tener una distribución espacial uniforme, todavía no se ha logrado tener una dispersión temporal controlada.

En difusores como los *Schroeder*, se cambia la fase del frente de ondas reflejado. En este tipo de difusores la previsión de la dispersión espacial no es tan trivial como en el caso anterior y el cambio de fase provoca a su vez una dispersión temporal.

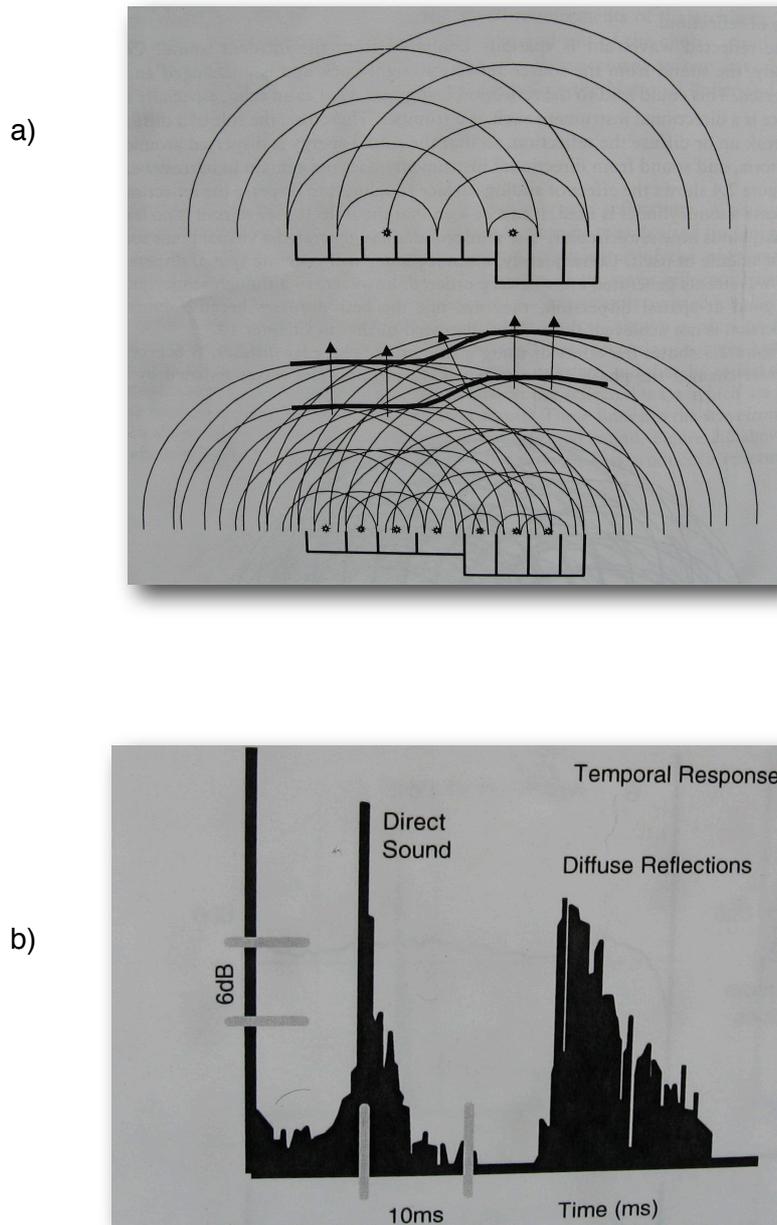


Fig.11. Difusor tipo Schroeder. a) dispersión espacial. b) dispersión temporal.

### II.3.1. EFECTO DEL CAMBIO DE FASE EN EL DOMINIO FRECUENCIAL.

El efecto de la dispersión puede ser visto en primer lugar en el dominio frecuencial. Cuando se tiene una superficie plana, por ejemplo una pared, al incidir una onda sonora tendremos una reflexión especular en una única dirección. En el dominio temporal, tendremos una “réplica” de la señal incidente retardada un cierto espacio de tiempo.

Si se analiza en el dominio frecuencial la reflexión especular, tendremos una respuesta característica de un filtro paso alto caracterizado por el tamaño y la forma de la superficie reflectora.

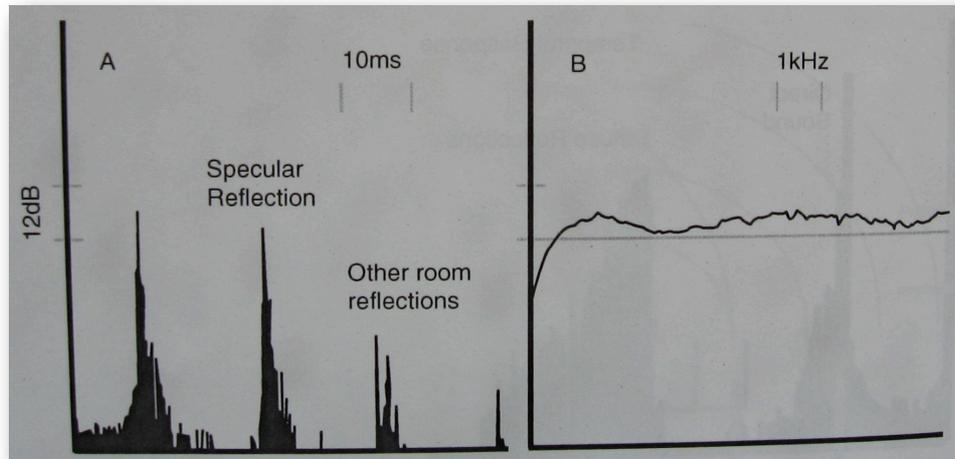


Fig.12. A) Dispersión temporal de una reflexión especular. B) Respuesta en frecuencia de una reflexión especular.

Cuando estamos delante de un difusor la onda especular reflejada es dispersada en el tiempo, mostrando tanto una reflexión espacial uniforme como una reflexión temporalmente ancha. Esto repercute en el espectro frecuencial ya que empiezan a aparecer funciones con máximos y mínimos debidos a la interferencia de ondas.

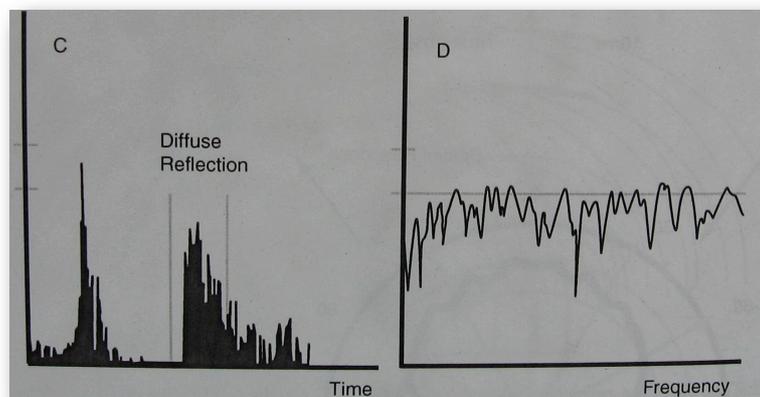


Fig.13. C) Dispersión temporal de una reflexión difusa. D) Respuesta en frecuencia de una reflexión difusa.

Erróneamente se tiene asumido que cualquier dispersión temporal tiene cabida dentro de una buena respuesta frecuencial, cuando en realidad existe una inherente coloración de la señal reflejada.

La ventaja del uso de difusores recae en la combinación de ambas señales, incidente y reflejada. Si tenemos una superficie plana, por lo tanto se produce una reflexión especular, la combinación de ambas señales tendrá una respuesta frecuencial tipo *comb-filter* (filtro peine). Tendremos ciertas frecuencias que serán aumentadas y otras anuladas, pudiendo tener efectos muy evidentes, y por tanto nocivos, en determinadas zonas del auditorio.

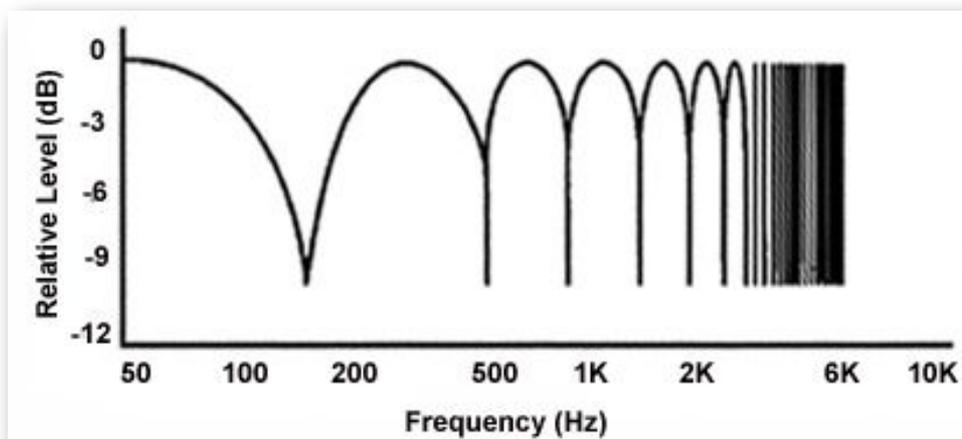


Fig.14. Interferencia tipo filtro peine debido a la combinación entre la señal incidente y reflejada en una reflexión especular.

El uso de difusores provoca que las diferentes interferencias entre onda incidente y ondas dispersadas aparezcan en el espectro frecuencial como un filtro peine muy irregular y con variaciones de niveles menores, obteniéndose unas respuestas más uniformes.

### II.3.2. EFECTO DEL CAMBIO DE FASE EN EL DOMINIO TEMPORAL.

Como se ha explicado en el apartado anterior, existe un efecto bastante notable en la señal global, suma de señal incidente y reflejada, en el dominio frecuencial. Las diferentes cancelaciones e interferencias constructivas provocan una coloración que en los casos más extremos puede llegar a ser perceptible.

Sin embargo no sólo debemos tener en cuenta los efectos en el espectro de la señal resultante, si no observar qué ocurre con la señal reflejada o dispersada en el dominio temporal.

En este caso no existen estudios que recojan con profundidad dichos efectos y no se sabe a ciencia cierta si al salir de un difusor la señal se comprime o por el contrario se expande temporalmente. Esto tendría importantes efectos sobre parámetros de la señal que son cruciales para una correcta interpretación por parte del oyente del sonido emitido.

El timbre de un sonido, como hemos explicado en apartados anteriores, tiene parámetros tanto temporales como espectrales que determinan las características del mismo. En concreto uno de los parámetros de mayor importancia es el tiempo de ataque, que es como sabemos un parámetro exclusivamente temporal.

Si en un difusor tenemos una variación temporal de la señal que afecte a dicho parámetro podremos estar introduciendo una variación irreversible en el timbre y por tanto estaríamos cambiando la percepción que tiene un oyente de los sonidos.

### II.3.3. *PARÁMETROS QUE CARACTERIZAN A LOS DIFUSORES.*

Actualmente existen dos normas de referencia para la caracterización de difusores acústicos basadas en la comparación entre la reflexión difusa y la reflexión especular y la observación del campo sonoro reflejado por el difusor.

La finalidad es obtener parámetros físicamente medibles sobre la dirección y forma de difusión, que ofrezcan una medida objetiva a la hora de diseñar y construir difusores acústicos.

Una de las normas a la que nos referimos es la **ISO/DIS 17497-1** [24] “*Acoustics — Measurement of the sound scattering properties of surfaces - Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room*”.

En esta primera parte de la norma se describe un método de medida para obtener el coeficiente de *scattering* del difusor, que describe el grado de sonido que el difusor es capaz de disgregar en direcciones distintas a la especular. Para esta medida es necesario disponer de una cámara reverberante o un laboratorio con instrumental adecuado para poder llevar a cabo la medida.

La otra parte de la norma es la **ISO/CD 17497-2** [25] “*Acoustics — Measurement of sound scattering properties of surfaces - Part 2: Measurement of the directional diffusion coefficient in a free field*”. Aquí se expone un nuevo parámetro, el coeficiente de difusión, que tiene incidencia en la uniformidad direccional del scattering y tiene relación con la calidad de la superficie difusora.

Al igual que en el caso anterior para obtener dicho parámetro hay que realizar un método riguroso en laboratorio que permita caracterizar de forma individual cada difusor.

Estas dos normas complementarias son las únicas que hoy día se han desarrollado para poder caracterizar de manera objetiva el comportamiento de los difusores aunque sólo hacen incidencia en la forma y grado de difusión del sonido, pero en ningún momento analizan qué ocurre con las características intrínsecas de la señal como puede ser el timbre, loudness, etc.

### III. MÉTODO EXPERIMENTAL.

#### III.1. CÁLCULO DE LA ENVOLVENTE TEMPORAL.

En los últimos años se ha estudiado el tiempo de ataque de sonidos musicales mediante pruebas perceptuales. En este proyecto trataremos de obtener este parámetro a partir de señales sonoras previamente grabadas calculando de manera experimental la envolvente temporal de las mismas y a partir de ésta aproximar el tiempo de ataque.

Como se ha expuesto en el capítulo anterior, la función envolvente se basará en la obtención de la envolvente de amplitud de la señal, por lo tanto dividiremos la señal original en segmentos tomando el valor máximo de amplitud de cada uno, reconstruyendo así una función equivalente y aproximada de la envolvente temporal de la misma.

Como es poco probable que tanto el nivel positivo como el nivel negativo de amplitud de la señal en cada tramo tengan el mismo valor, tomaremos como valor representativo de cada segmento la media (en valor absoluto) del nivel más alto y el nivel más bajo.

Analíticamente el método vendría representado por:

(17)

$$y[n] = \omega[n] \sum_{i=1}^m [media(|x[i]_{max}|, |x[i]_{min}|)]$$

siendo  $x[i]$  la señal sonora a analizar y  $\omega[n]$  la ventana de longitud  $m$ .

Este método tendría una fiabilidad del 100% en tonos puros, donde tomando como enventanado una ventana de tamaño igual al periodo de la señal, podríamos obtener la envolvente de amplitud de la señal de manera exacta.

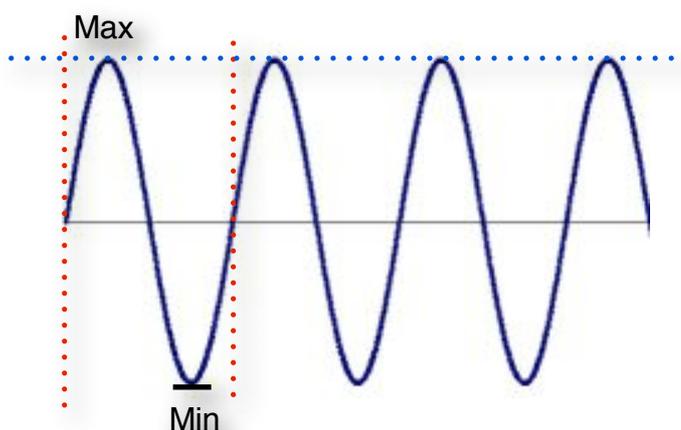


Fig.15. Gráfico del cálculo de la envolvente de amplitud de una señal senoidal. Las líneas verticales punteadas en rojo denotan la ventana de longitud igual a un periodo de la señal y la línea punteada azul marcan la envolvente de la señal.

En nuestro caso vamos a trabajar con señales sonoras que tienen un gran ancho de banda, por lo que intentar tomar como tamaño de ventana un periodo de la señal, introduciría un error demasiado alto como para poder considerar dicha función envolvente como fiable. Para intentar solventar este problema realizaremos un enventanado adaptado dividido en dos zonas bien diferenciadas.

Intentamos obtener el tiempo de ataque de señales musicales, por lo tanto necesitamos que la señal envolvente que tengamos posea una gran precisión en la parte inicial de la señal, es decir necesitamos una envolvente que realmente siga la señal original sin introducir un gran error. En el resto de la señal, con tener una aproximación de la envolvente es suficiente.

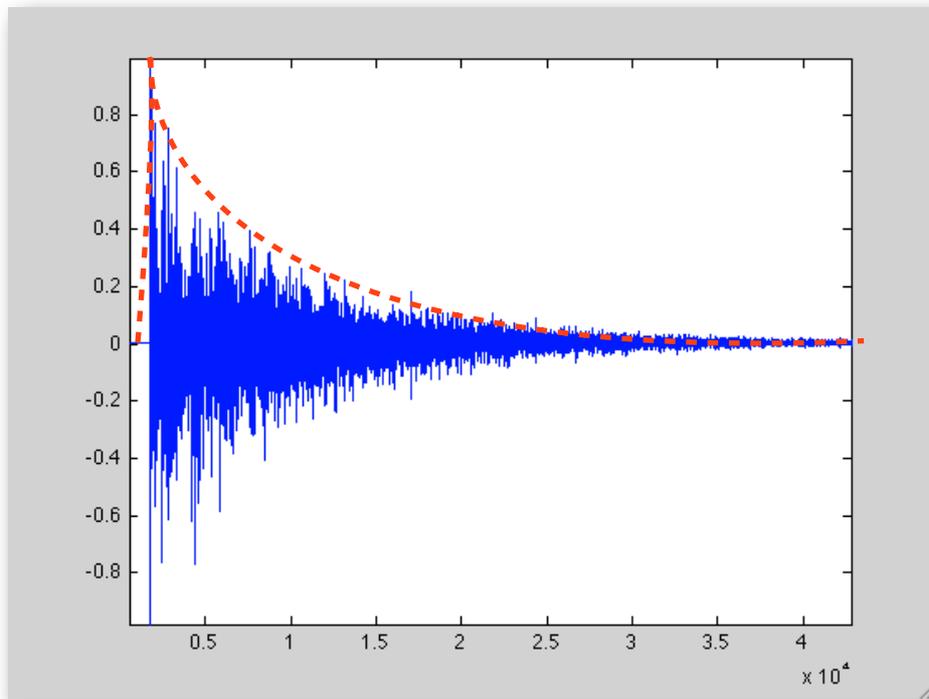


Fig.16. Señal impulsiva y representación de la envolvente mediante la línea de puntos roja.

Para conseguir esto utilizaremos dos ventanas diferentes. La primera se empleará en la parte correspondiente al ataque de la señal, es decir, hasta la llegada del primer máximo. Dicha ventana tendrá una longitud de dos muestras, por lo que utilizando el algoritmo definido en la ecuación 17, obtendríamos la envolvente de la señal con una gran precisión. A partir de este punto utilizaríamos otra ventana con una longitud mayor para obtener una envolvente aproximada que no siga de manera tan exhaustiva la forma de onda de la señal.

Después de realizar pruebas de este algoritmo en señales de banda ancha, la mejor aproximación de la envolvente viene dada cuando se toma como longitud de la ventana el periodo correspondiente a la frecuencia fundamental del ancho de banda de la señal. Por lo tanto en la segunda ventana utilizaremos el periodo asociado a esta frecuencia para obtener una envolvente que siga, de manera bastante aproximada a la señal.

Una vez ejecutado el algoritmo obtenemos una señal envolvente con un número de muestras inferior al de la señal original. Por este motivo tenemos que realizar dos interpolaciones diferentes, cada una en la parte de señal correspondiente a las ventanas utilizadas, para así obtener una función envolvente con un número de muestras igual a la onda original.

Para implementar el algoritmo mencionado nos servimos del paquete MATLAB, desarrollando la función envolvente:

```
function [yi]= envolvente(x,fo,fs) %x es la señal de entrada fo es la frecuencia fundamental de la
señal y fs la frecuencia de muestreo
period=round((1/fo)*44100);
e=length(x);
y=[]; y1=[]; y3=[]; %reservamos memoria para las variables de salida

M=min(find(x>=0.95*max(x))); %obtenemos la posición del primer máximo de la
m=min(find(x<=0.95*min(x))); %señal para definir las zonas de enventanado

maximo=min(M,m);

for i=1:2:maximo %realizamos el cálculo de la envolvente en la parte de
                    % ataque de la señal
    T=x(i:2+i);
    Max=max(abs(T));Min=min(abs(T));

    TT=[Max Min];
    s=median(TT);
    y=[y s];
end

q=length(y);Interpolado=(q-1)/maximo; g = 0:q-1; gi = 0:Interpolado:q-1;

Y = interp1(g,y,gi) %interpolamos para obtener el número de muestras igual
                    % al de la señal original
y1=Y(end);

for i=maximo+2:period:e-period %realizamos el cálculo de la envolvente en
                                %la parte posterior al ataque de la señal
    T=x(i:period+i);
    Max=max(abs(T));Min=min(abs(T));

    TT=[Max Min];
    s=median(TT);
    y1=[y1 s];
end

q=length(y1);Interpolado=(q-1)/(e-maximo);g = 0:q-1; gi = 0:Interpolado:q-1;

Y1 = interp1(g,y1,gi) %interpolamos para obtener el número de muestras
                    %igual al de la señal original
yi=[Y Y1];
attack= min(find(yi==max(yi))) %Calcula el momento del ataque
onset= min(find(yi>=0.1*max(yi)))%Calcula el momento de la llegada física del sonido
PAT= ((attack-onset)/fs)*1000 %Calcula el PAT

end
```

Una vez ejecutado el algoritmo con una señal impulso, el resultado obtenido se puede visualizar en la Fig.17:

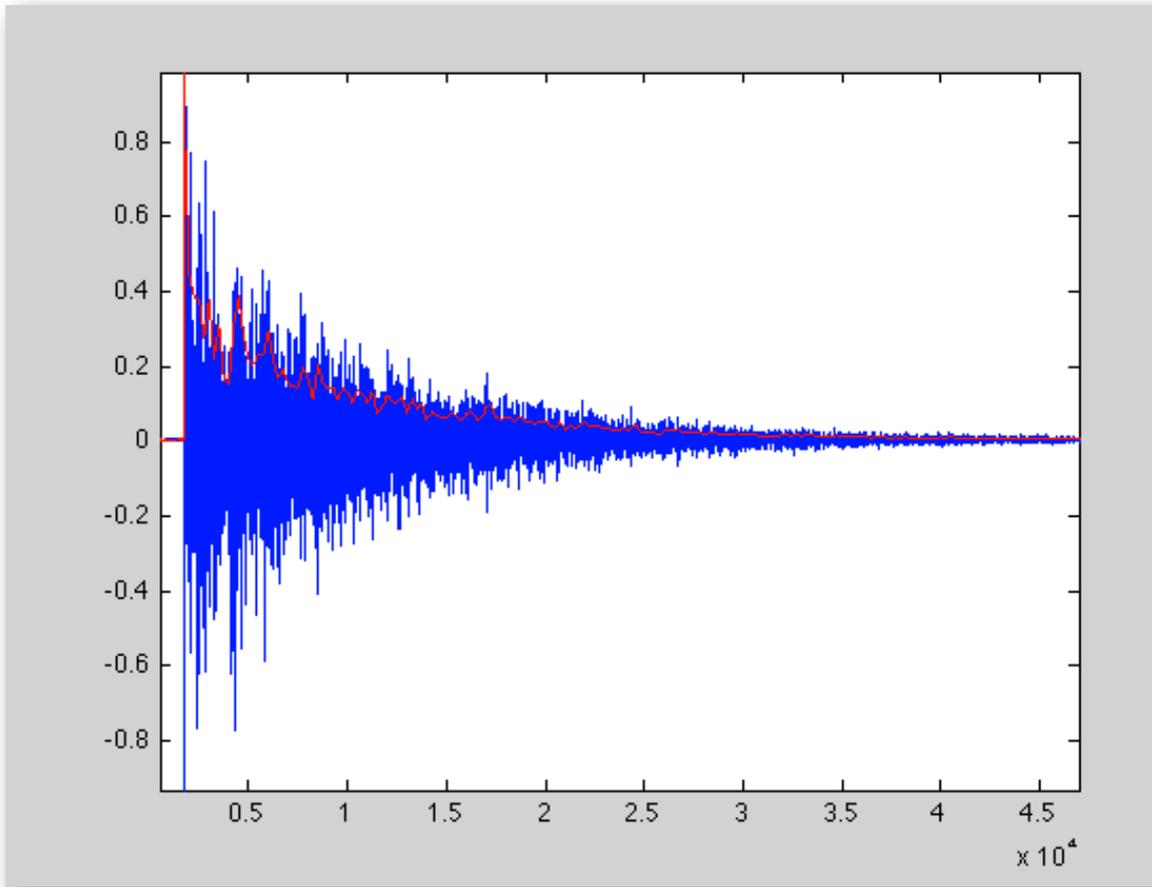


Fig.17. Cálculo de la envolvente de amplitud (línea roja) de una señal impulso utilizando la función “envolvente”.

### III.2 CÁLCULO DEL TIEMPO DE ATAQUE PERCEPTUAL.

En la bibliografía existen numerosos modelos para obtener el tiempo de ataque perceptual a partir de la envolvente temporal. De los que se han podido ver en el capítulo 2, vamos a aplicar en nuestro proyecto el modelo “Percent of Max” definido por Vos y Rasch [VOS1981].

Este modelo predice el PAT como el momento en el que la amplitud de la envolvente cruza un umbral relativo a su valor máximo, expresado como un porcentaje del mismo. Dicho umbral es un parámetro experimental que ayuda a determinar el momento en el que ocurre el ataque de la señal minimizando lo máximo posible el error.

A la hora de ejecutar una nota musical es usual que existan variaciones de la amplitud debidas exclusivamente al músico, como puede ser la aparición de “vibrato” o subidas o descensos de la amplitud de manera involuntaria. Por lo tanto si buscamos exclusivamente el máximo de la función (correspondiente según la teoría a la zona de ataque), podríamos estar introduciendo un gran error.

Es por ello que se establece un umbral correspondiente al 95% del valor máximo que tiene como fin obviar dichas alteraciones externas y determinar de manera más precisa el momento en el que aparece el ataque.

Para poder obtener el PAT deberíamos determinar por otro lado el momento en el que se produce la llegada física del sonido, pero no es algo que se obtenga de manera trivial, ya que las señales con las que contamos contienen de manera inherente un ruido y un retardo que dificultan la detección del comienzo físico del evento sonoro. Es por esto que tomaremos como comienzo físico de la señal la primera muestra que tiene un nivel del 5% con respecto al nivel máximo.

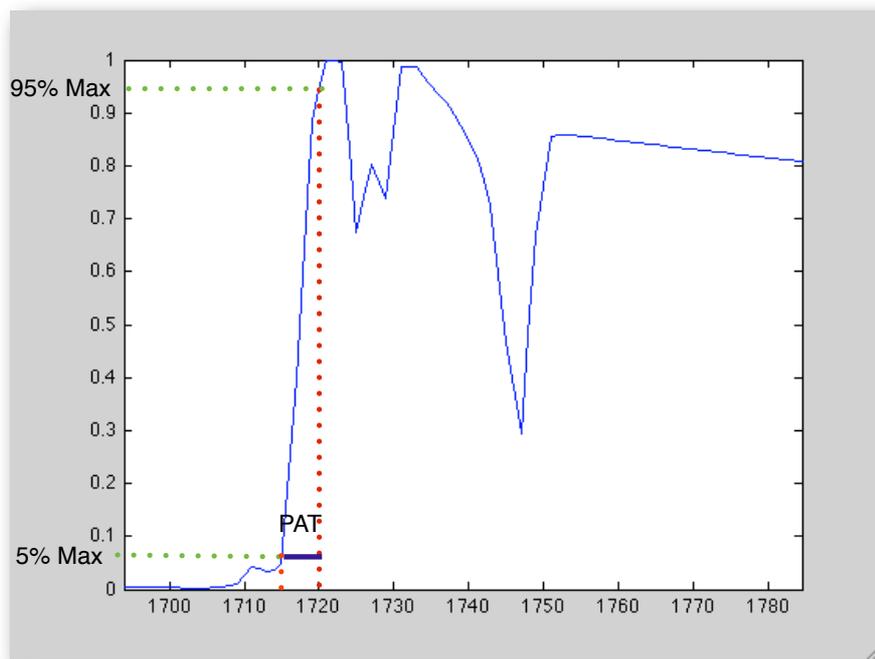


Fig.18. Obtención del PAT mediante el modelo “Percent of max” y tomando la llegada a 5% del nivel más alto.

Con esto eliminamos la parte de señal correspondiente al ruido de fondo y retardo y aproximamos, asegurando la mayor precisión posible, el cálculo del tiempo de ataque perceptual.

#### IV. ANÁLISIS DE LAS SEÑALES.

En este presente capítulo vamos a intentar obtener el tiempo de ataque de varias notas musicales utilizando la función “envolvente”. Esta función devuelve automáticamente el tiempo de ataque perceptual (PAT) de la señal en cuestión.

Posteriormente obtendremos las respuestas al impulso de diferentes difusores que nos permitirán, convolucionándolas con los tonos musicales, analizar las señales acústicas una vez que han sido reflejadas por el difusor, observando si existe o no una influencia medible en el PAT del tono incidente.

##### IV.1. SEÑALES DE AUDIO.

El análisis antes mencionado lo vamos a realizar sobre notas reales de instrumentos musicales grabados previamente. Al tratarse de un estudio empírico y no perceptual debemos tener en cuenta determinadas consideraciones que acoten lo máximo posible el error.

Para el cálculo del PAT necesitamos notas aisladas y no incluidas en un contexto musical. Esto es debido a la propia definición del tiempo de ataque. Éste nos dice que cuando ocurre un sonido, la intensidad del mismo aumenta hasta un punto donde aparece perceptualmente un “acento” (el ataque) y posteriormente en las siguientes etapas el sonido disminuye de nivel hasta su desaparición.

A la hora de ejecutar una nota musical dentro de una frase musical, el intérprete puede de manera intencionada variar el nivel de intensidad de la nota aumentando o decreciendo su nivel (“crescendo”, “diminuendo”, “sforzando”, etc.) enfatizar o no la manera de ejecutar la nota (ataque simple, “staccato”, etc.), o modificar su intensidad de manera circular con un “vibratto”. Todos estos procesos de ejecución pueden llegar a enmascarar el ataque de la señal haciendo que sea casi imposible analizar de manera objetiva (no hablamos de un método perceptual), el PAT guiándonos por la definición del mismo.

La función diseñada para este proyecto se basa en ese principio, por lo tanto necesitamos analizar notas aisladas interpretadas fuera de un contexto musical y ejecutadas de manera lo más estable posible. El método elegido, “Percent of max”, asegura que aquellas variaciones como el “vibratto” o alteraciones involuntarias sean ignoradas obteniéndose una aproximación del tiempo de ataque del instrumento lo más precisa posible.

Se han tomado 7 señales sonoras procedentes de instrumentos pertenecientes a cada una de las familias existente: viento metal (trompeta piccolo), viento madera (clarinete), cuerda frotada (violín), cuerda pulsada (guitarra), percusión (caja y campana) y voz (barítono).

Instrumento	Duración del archivo (segundos)
Trompeta “piccolo”	1,00
Clarinete	4,00
Violín	6,00
Guitarra	1,00
Caja	0,15
Campana	4,00
Voz	2,00

Tabla 1: Señales de audio utilizadas para el análisis.

Todas las señales tienen una frecuencia de muestreo de 44100 muestras/segundo.

#### IV.2. RESPUESTA AL IMPULSO DE DIFUSORES.

Este primer proyecto se basa en un estudio del efecto que tiene sobre el tiempo de ataque de señales sonoras los difusores acústicos, por lo tanto utilizaremos unas simulaciones de la respuesta al impulso de recintos con y sin difusores en sus paredes utilizando el método de diferencias finitas en el dominio de tiempos FDTD, acrónimo de “finite- difference time-domain”.

Este método discretiza el área de simulación creando una malla de puntos. Internamente esta zona de integración consiste en tres mallas al tresbolillo (una malla para la presión sonora, otra malla para la velocidad de partícula en la dirección x, y finalmente una malla para la velocidad de partícula en la dirección y). La distancia entre puntos de cada malla es de 1 cm, por lo que cabe esperar resultados validos hasta los 4000 Hz.

Se simulará una sala en 2D donde se colocará en sus paredes independientemente cilindros y difusores. Para ello se calculará la respuesta al impulso de la sala sin difusores, colocando una pared de difusores, dos, tres y todas. En total tendremos 9 respuestas al impulso que convolucionaremos con las señales de audio anteriormente descritas para tener las ondas sonoras con la respuesta de la sala.

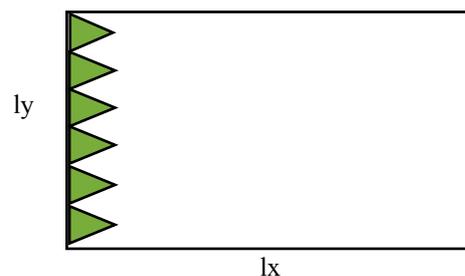


Fig.19. Ejemplo del diseño de la sala en 2D a analizar con una pared de difusores.

Nombre Respuesta al Impulso	Descripción
c0	Respuesta al impulso de la sala vacía
c1	Respuesta al impulso de la sala con una pared de cilindros
c2	Respuesta al impulso de la sala con dos paredes de cilindros
c3	Respuesta al impulso de la sala con tres paredes de cilindros
c4	Respuesta al impulso de la sala con todas las paredes de cilindros
d1	Respuesta al impulso de la sala con una pared de difusores
d2	Respuesta al impulso de la sala con dos paredes de difusores
d3	Respuesta al impulso de la sala con tres paredes de difusores
d4	Respuesta al impulso de la sala con todas las paredes de difusores

Tabla 2: Descripción de las respuestas al impulso utilizadas para el análisis.

### IV.3. TIEMPO DE ATAQUE PERCEPTUAL.

En este apartado, vamos a presentar los resultados obtenidos después de las diferentes simulaciones y análisis de las señales. Los dividiremos en dos subapartados, en el primero de ellos probaremos la funcionalidad de la función “envolvente” calculando el tiempo de ataque perceptual de los diferentes archivos sonoros disponibles, y posteriormente obtendremos el mismo parámetro, pero de las señales simuladas en el interior de la sala con las diferentes configuraciones arquitectónicas descritas anteriormente.

### IV.3.1. TIEMPO DE ATAQUE PERCEPTUAL DE TONOS MUSICALES.

Una vez analizados los diferentes tonos musicales mediante el programa “envolvente”, los resultados obtenidos se pueden observar seguidamente:

- Voz

Ataque = 148.7 ms

Llegada física = 38.3 ms

PAT = 110.4 ms

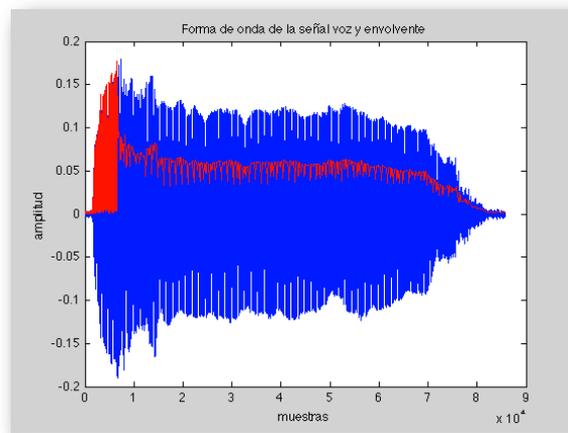


Fig.20. Forma de onda de la señal voz (azul) y forma de onda de su envolvente calculada (rojo).

- Clarinete

Ataque = 167.1 ms

Llegada física = 96.5 ms

PAT = 70.6 ms

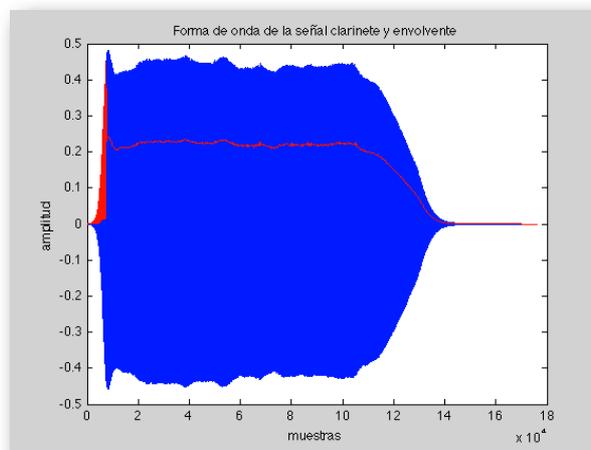


Fig.21. Forma de onda de la señal clarinete (azul) y forma de onda de su envolvente calculada (rojo).

- Trompeta

Ataque = 25.5 ms

Llegada física = 10.6 ms

PAT = 14.9 ms

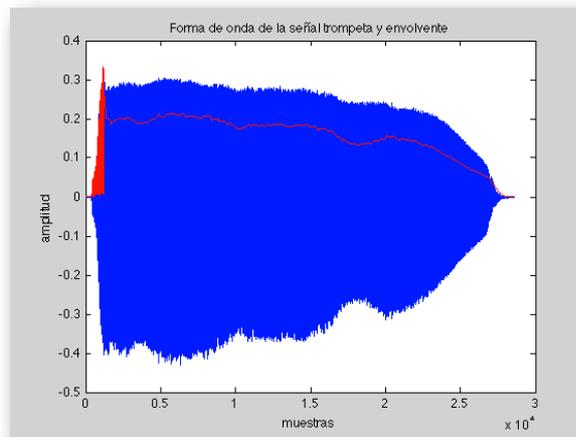


Fig.22. Forma de onda de la señal trompeta (azul) y forma de onda de su envolvente calculada (rojo).

- Violín

Ataque = 343.9 ms

Llegada física = 339.0 ms

PAT = 4.9 ms

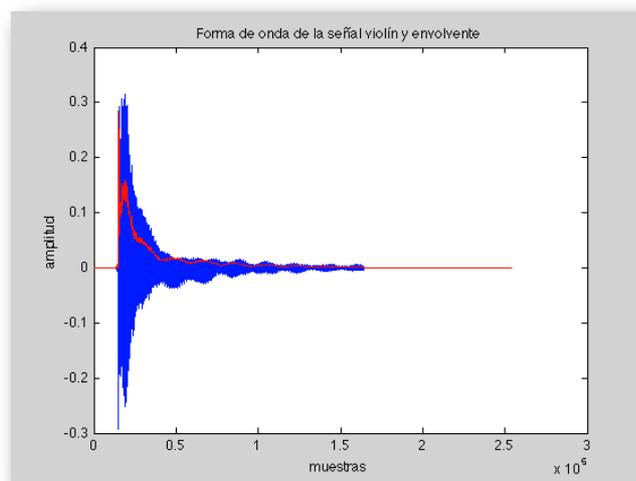


Fig.23. Forma de onda de la señal violín (azul) y forma de onda de su envolvente calculada (rojo).

- Guitarra

Ataque = 556.5 ms

Llegada física = 543.3 ms

PAT = 13.2 ms

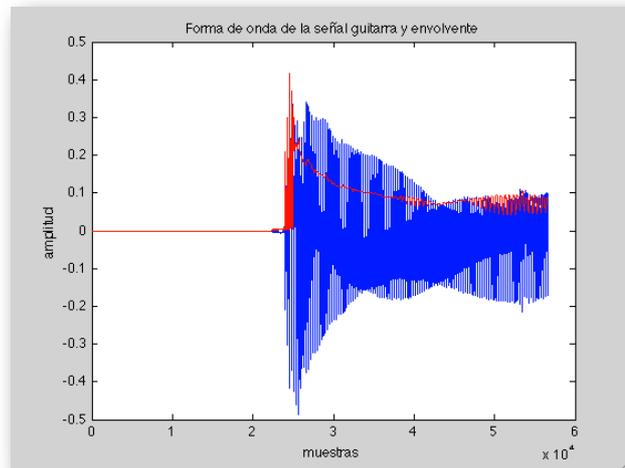


Fig.24. Forma de onda de la señal guitarra (azul) y forma de onda de su envolvente calculada (rojo).

- Tambor

Ataque = 6.6 ms

Llegada física = 0.7 ms

PAT = 5.9 ms

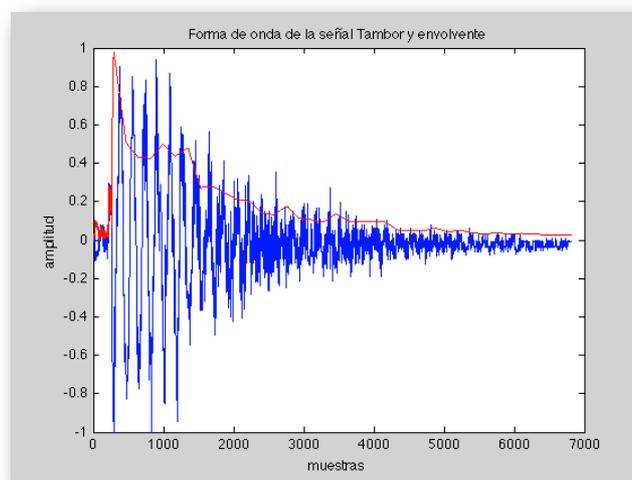


Fig.25. Forma de onda de la señal Tambor (azul) y forma de onda de su envolvente calculada (rojo).

- Campana

Ataque = 305.5 ms

Llegada física = 293.4 ms

PAT = 12.1 ms

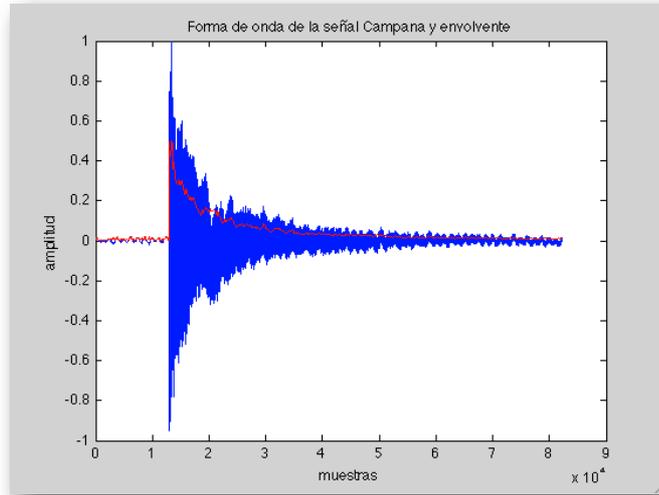


Fig.26. Forma de onda de la señal Campana (azul) y forma de onda de su envolvente calculada (rojo).

En la siguiente tabla podemos ver un resumen de todos los resultados obtenidos:

Instrumento	PAT (ms)
Trompeta "piccolo"	14,9
Clarinete	70,6
Violín	4,9
Guitarra	13,2
Caja	5,9
Campana	12,1
Voz	110,4

Tabla 3: Tiempo de Ataque Perceptual (PAT) de los instrumentos analizados.

Observando los resultados de cada una de las señales analizadas podemos comprobar visualmente que la envolvente sigue en todos los casos la forma de onda del tono original. Por tanto los resultados correspondiente al cálculo del PAT muestran valores coherentes con las formas de onda sobre los que se calcula.

En los diferentes estudios que se han ido realizando a lo largo de los años se han logrado obtener valores de tiempos de ataque perceptual de numerosos instrumentos, aunque de algunos de los que en este presente proyecto se analizan no se tiene constancia de ellos.

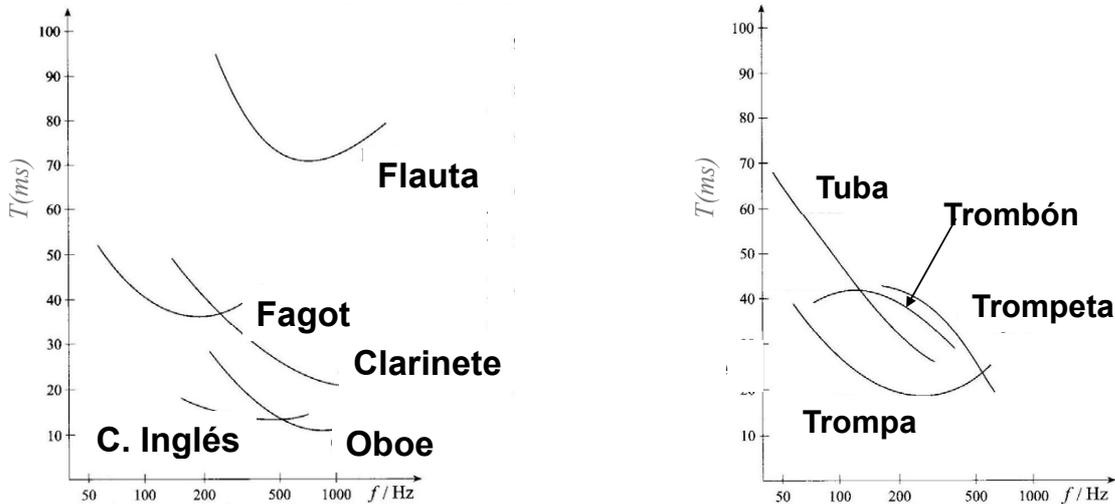


Fig.27. Rangos de PAT para algunos instrumentos en función de la frecuencia.

A pesar de esto no se puede llegar a obtener un valor de referencia bibliográfico de confianza, ya que muchos de estos estudios plantean unos amplios rangos de PAT en función de la frecuencia para cada instrumento y basan su análisis en métodos perceptuales, que tienen una incertidumbre elevada.

Es por esto que basamos la bondad de los resultados obtenidos en la observación de las formas de onda de los tonos musicales analizados, y la aplicación de la definición teórica del tiempo de ataque perceptual a las mismas mediante la función desarrollada.

#### IV.3.2. TIEMPO DE ATAQUE PERCEPTUAL DE TONOS MUSICALES EN ENTORNOS DIFUSIVOS.

Numéricamente podemos obtener una señal sonora con la respuesta de una sala convolucionando dicha señal con la respuesta al impulso del recinto en cuestión.

En nuestro caso tenemos unas respuestas al impulso de una sala de 2D que, debido al tamaño de la malla de integración, tienen una validez numérica hasta los 4000 Hz. Es por esto que aquellas señales con un gran contenido espectral por encima de esta frecuencia no serán analizadas, ya que sufrirían una “coloración” excesiva alterando de manera significativa los resultados.

Contamos con las respuestas de salas con dos tipos diferentes de difusores, y diferentes grados de implementación en las mismas, desde una sala con una pared de difusores hasta las 4 paredes con difusores.

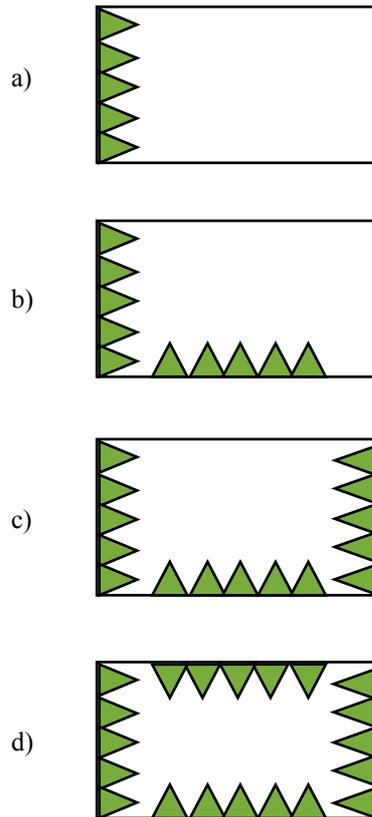


Fig.28. Diferentes configuraciones de los difusores dentro de la sala.

Antes de realizar el análisis hay que tener en cuenta un par de consideraciones importantes. La sala que estamos simulando es una sala rectangular con las paredes paralelas. Por lo tanto aparecerán modos propios tanto longitudinales como transversales que van a “colorear” la señal. Es por eso que en este punto, la localización geométrica exacta de la respuesta al impulso va a ser determinante en dicha respuesta debida a estos modos.

A medida que vamos introduciendo difusores en las paredes vamos eliminando dichos modos, pero es necesario la colocación de difusores en dos paredes para hacer desaparecer dicho efecto, puesto que con una sola pared de difusores seguimos teniendo dos paredes paralelas lisas como podemos observar en la Fig.27 a). Es por esto que el dato obtenido en la sala con una sola pared de difusores hay que tomarla en consideración pero teniendo siempre en cuenta la premisa anterior.

El objetivo es ver si existe una influencia tangible sobre el tiempo de ataque de tonos musicales el empleo de difusores acústicos en recintos.

Una vez obtenidas las señales sonoras convolucionadas con las respuestas al impulso de las diferentes configuraciones, las analizamos con la función “envolvente” para obtener su tiempo de ataque perceptual de la misma forma que en el apartado anterior. Los resultados se pueden ver a continuación.

- Voz en la sala con cilindros

Configuración	c0	c1	c2	c3	c4
PAT (ms)	170,45	131,97	151,99	168,73	171,54

Tabla 4: Evolución del PAT de la voz en la sala con cilindros.

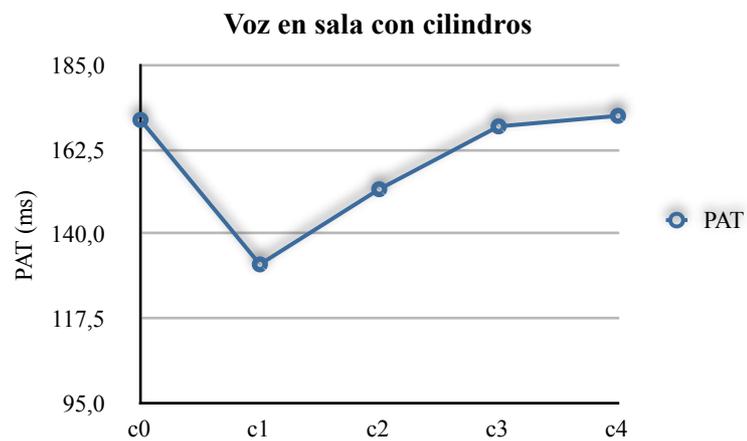


Fig.29. Evolución del PAT de la señal Voz dentro de la sala con cilindros.

- Voz en la sala con difusores

Configuración	c0	d1	d2	d3	d4
PAT (ms)	170,45	175,48	129,04	126,64	106,12

Tabla 5: Evolución del PAT de la voz en la sala con difusores.

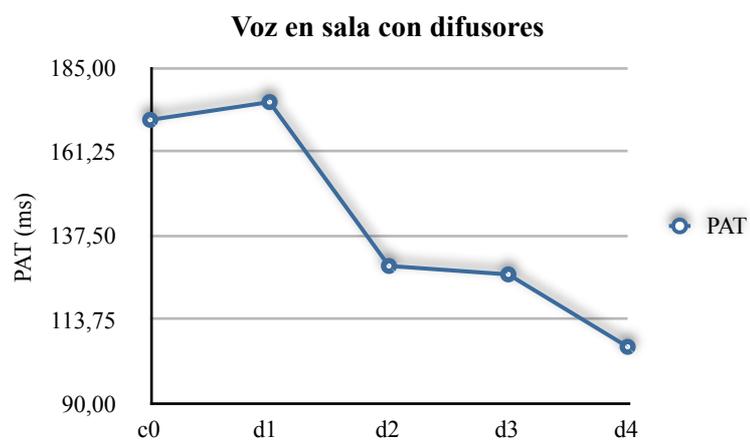


Fig.30. Evolución del PAT de la señal Voz dentro de la sala con difusores.

- Clarinete en la sala con cilindros

Configuración	c0	c1	c2	c3	c4
PAT (ms)	143,94	77,95	73,92	55,58	143,35

Tabla 6: Evolución del PAT del clarinete en la sala con cilindros.

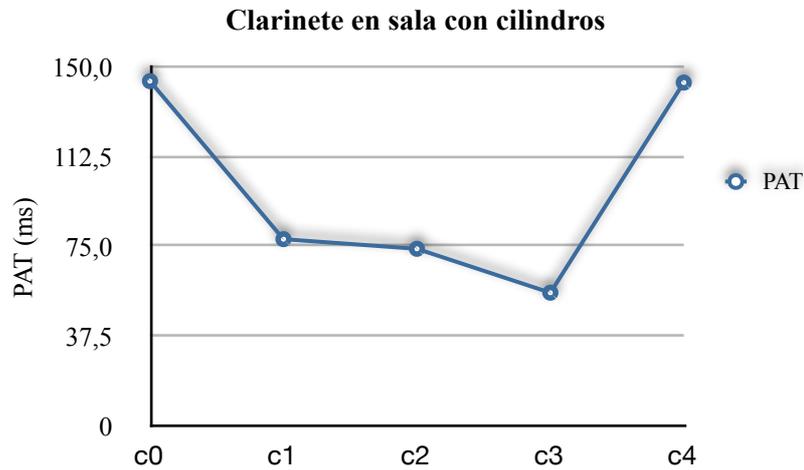


Fig.31. Evolución del PAT de la señal Clarinete dentro de la sala con cilindros.

- Clarinete en la sala con difusores

Configuración	c0	d1	d2	d3	d4
PAT (ms)	143,94	73,94	100,43	76,89	97,30

Tabla 7: Evolución del PAT del clarinete en la sala con difusores.

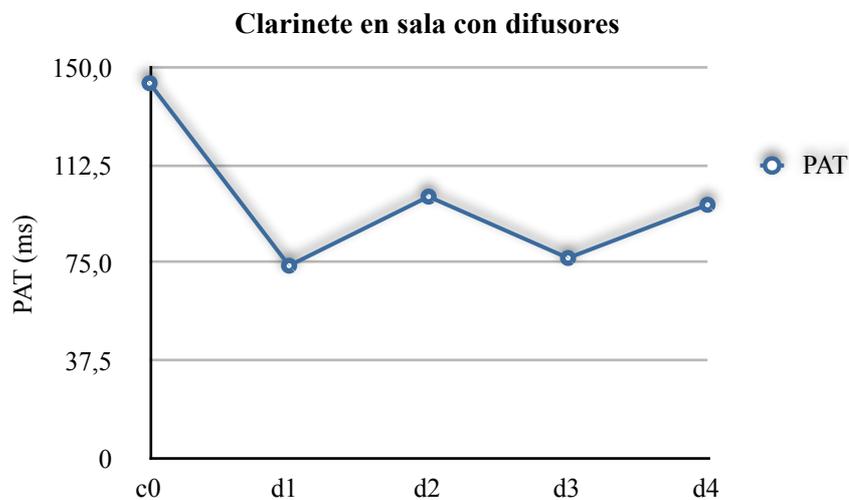


Fig.32. Evolución del PAT de la señal Clarinete dentro de la sala con difusores.

- Guitarra en la sala con cilindros

Configuración	c0	c1	c2	c3	c4
PAT (ms)	41,38	33,60	56,85	87,05	28,79

Tabla 8: Evolución del PAT de la guitarra en la sala con cilindros.

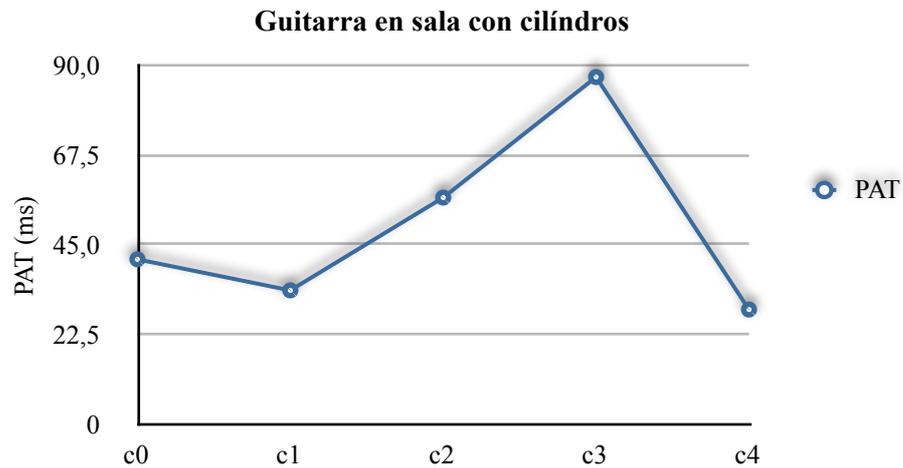


Fig.33. Evolución del PAT de la señal Guitarra dentro de la sala con cilindros.

- Clarinete en la sala con difusores

Configuración	c0	d1	d2	d3	d4
PAT (ms)	41,83	12,01	18,17	23,94	24,51

Tabla 9: Evolución del PAT de la guitarra en la sala con difusores.

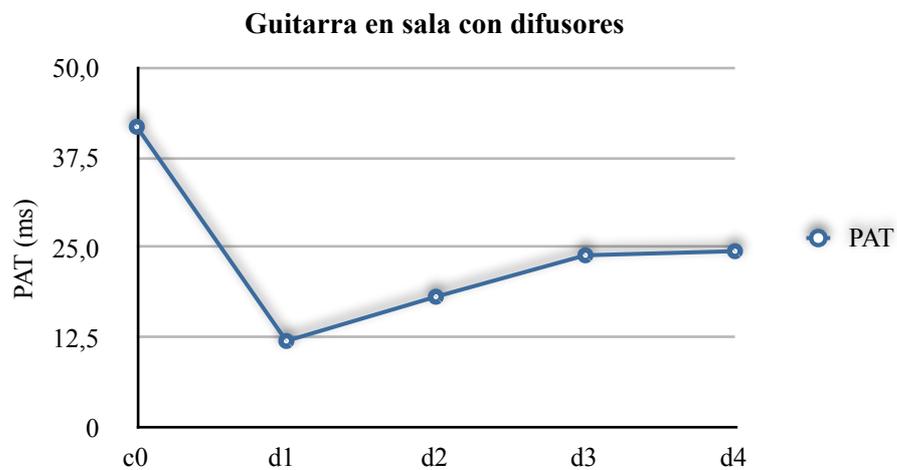


Fig.34. Evolución del PAT de la señal Guitarra dentro de la sala con difusores.

- Violín en la sala con cilindros

Configuración	c0	c1	c2	c3	c4
PAT (ms)	14,04	30,77	14,10	48,43	35,51

Tabla 10: Evolución del PAT del violín en la sala con cilindros.

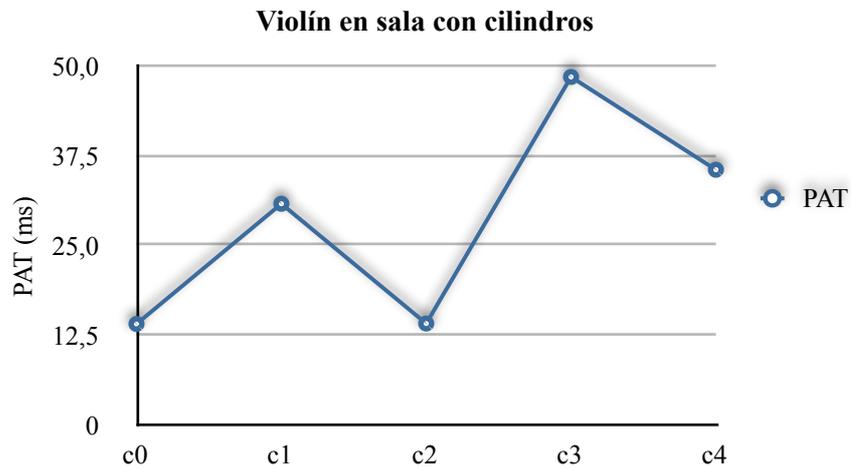


Fig.35. Evolución del PAT de la señal Violín dentro de la sala con cilindros.

- Violín en la sala con difusores

Configuración	c0	d1	d2	d3	d4
PAT (ms)	14,04	10,95	19,59	39,27	42,35

Tabla 11: Evolución del PAT del violín en la sala con difusores.

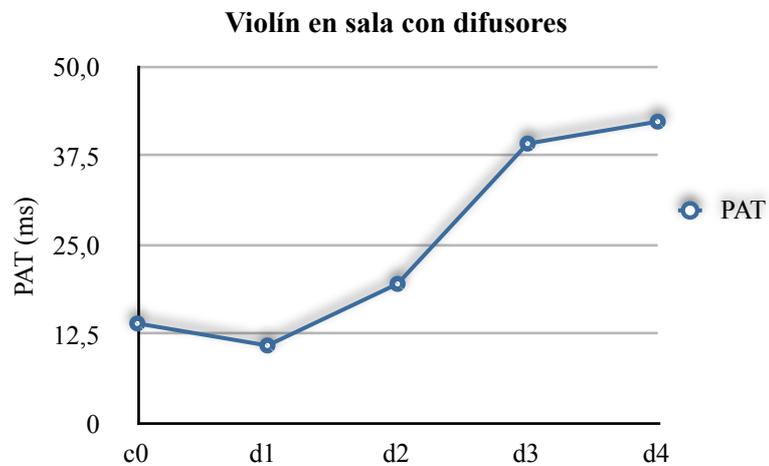


Fig.36. Evolución del PAT de la señal Violín dentro de la sala con difusores.

- Campana en la sala con cilindros

Configuración	c0	c1	c2	c3	c4
PAT (ms)	15,98	6,21	16,41	15,21	10,09

Tabla 12: Evolución del PAT de la campana en la sala con cilindros.

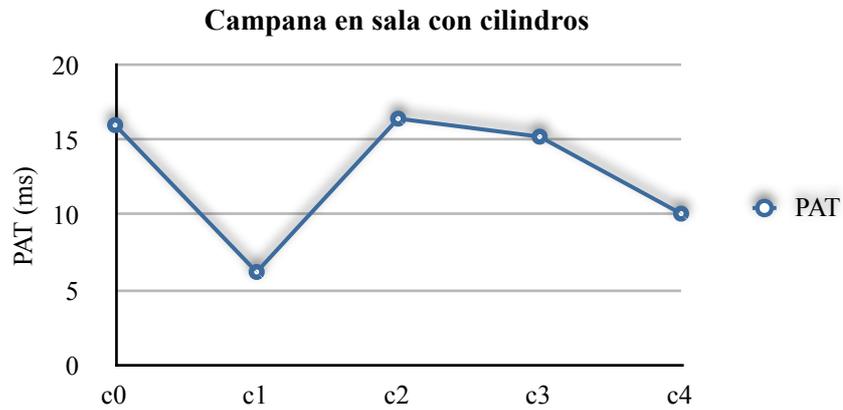


Fig.37. Evolución del PAT de la señal Campana dentro de la sala con cilindros.

- Campana en la sala con difusores

Configuración	c0	d1	d2	d3	d4
PAT (ms)	15,98	5,85	6,23	10,06	9,95

Tabla 13: Evolución del PAT de la campana en la sala con difusores.

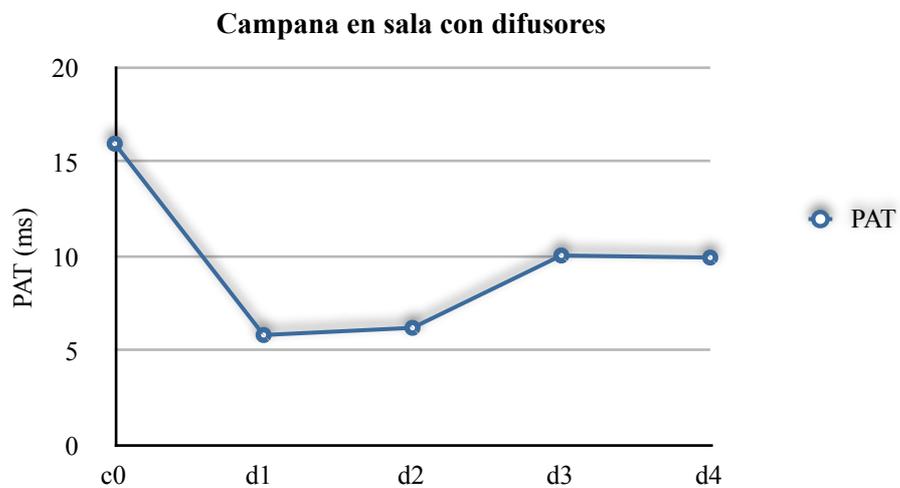


Fig.38. Evolución del PAT de la señal Campana dentro de la sala con difusores.

En todos los análisis se observa una tendencia común. Al aumentar el número de superficies difusoras en la sala tenemos un aumento en el tiempo de ataque perceptual. Si bien es verdad que en algunos casos existen datos o tendencias que no apoyan esta teoría, podemos hablar de una “correlación” general entre el número de superficies difusoras y el PAT.

En las siguientes tablas y figuras podemos ver un resumen de los datos obtenidos:

- Sala con cilindros

	c0	c1	c2	c3	c4
Voz	170,45	131,97	151,99	168,73	171,54
Clarinete	143,94	77,95	73,92	55,58	143,35
Guitarra	41,38	33,60	56,85	87,05	28,79
Violín	14,04	30,77	14,10	48,43	35,51
Campana	15,98	6,21	16,41	15,21	10,09

Tabla 14: Evolución del PAT de todos los instrumentos en la sala con cilindros.

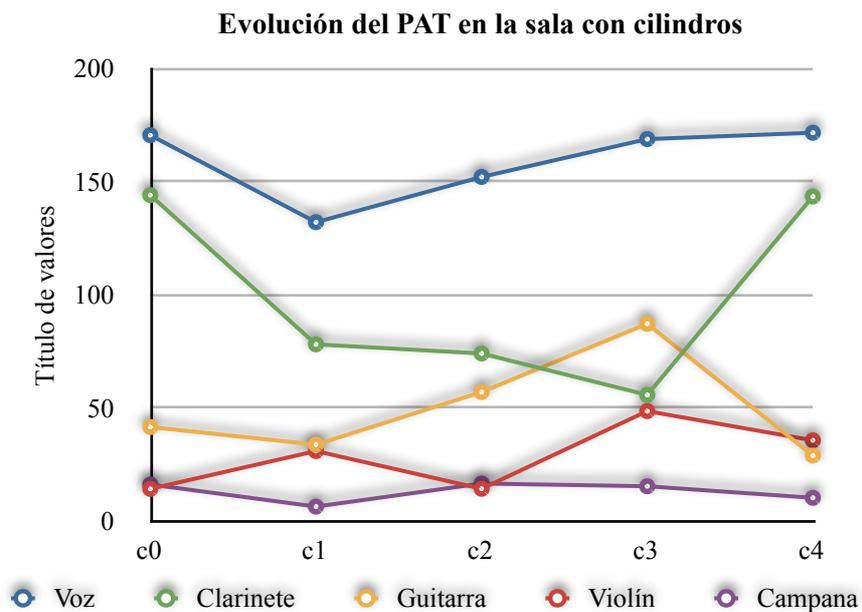


Fig.39. Evolución del PAT de todas las señales dentro de la sala con cilindros.

- Sala con difusores

	c0	c1	c2	c3	c4
Voz	170,45	175,48	129,04	126,64	106,12
Clarinete	143,94	73,94	100,43	76,89	97,30
Guitarra	41,83	12,01	18,17	23,94	24,51
Violín	14,04	10,95	19,59	39,27	42,35
Campana	15,98	5,85	6,23	10,06	9,95

Tabla 15: Evolución del PAT de todos los instrumentos en la sala con difusores.

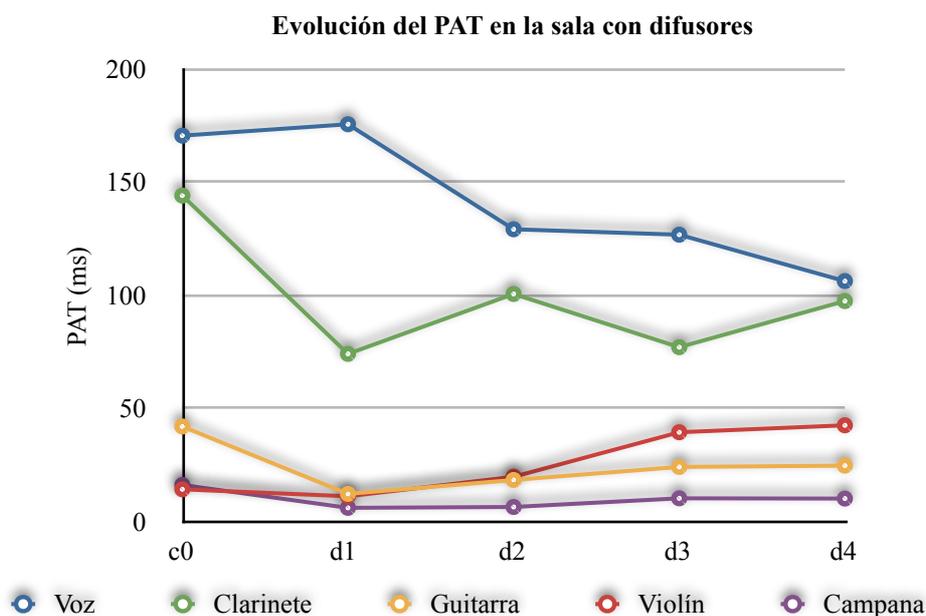


Fig.40. Evolución del PAT de todas las señales dentro de la sala con difusores.

Se ha comprobado, analítica y experimentalmente, que la dispersión espacial que sufre una onda cuando incide en un difusor conlleva a su vez una dispersión temporal, como así lo explican Cox y D'Antonio [28]; una forma de expansión en el tiempo. Es por tanto razonable pensar que el PAT de dicha onda incidente, que es un parámetro exclusivamente temporal, también sufra una dispersión y se expanda en el tiempo.

Aunque hemos podido observar que hay una “cierta tendencia” a dicha hipótesis, no tiene porqué ocurrir así en todos los difusores ya que estos dispositivos están “sintonizados” para funcionar de manera óptima en un rango determinado de frecuencias.

Algunos estudios sobre el PAT en instrumentos han llegado a la conclusión de que dicho parámetro es dependiente también de la frecuencia, siendo inversamente proporcional a la misma, por lo tanto si un difusor dispersa espacialmente (y por tanto temporalmente) el sonido de manera diferente según la frecuencia de la onda incidente, es razonable pensar que el tiempo de ataque perceptual también estará afectado por dicha característica.

## V. CONCLUSIONES Y LÍNEAS FUTURAS.

Son pocos los estudios que se centran en los aspectos temporales de los difusores acústicos y el efecto que producen sobre las señales sonoras incidentes. En este proyecto hemos intentado introducir el cálculo de un parámetro temporal característico de todos los sonidos, el Tiempo de Ataque Perceptual, PAT, que tiene una gran relación con el timbre de los sonidos.

Utilizando uno de los métodos descritos en [12], se ha implementado en MATLAB una función que calcula de manera bastante fiable el PAT a partir de la envolvente de amplitud de las señales sonoras.

Este cálculo se ha inscrito dentro de unas restricciones necesarias, y por otro lado evidentes desde el punto de vista teórico, para el buen funcionamiento del método. Para obtener unos resultados fiables del tiempo de ataque perceptual se deben analizar señales sonoras, en este caso notas musicales, interpretadas de la manera más lineal posible, sin introducir cambios intencionados de intensidad ni *vibrato*, que puedan alterar la percepción del PAT y dificulten o imposibiliten el cálculo de éste, del mismo modo deben ser notas aisladas fuera de un contexto musical, para asegurar la correcta ejecución de las mismas.

Los resultados obtenidos sobre notas de diferentes instrumentos, pertenecientes a diversas familias han ofrecido resultados coherentes desde el punto de vista teórico y experimental, que hacen prever que la eficacia del método desarrollado es significativa.

En un segundo punto, siendo el pilar de este proyecto, se intentó comprobar que el uso de difusores acústicos tiene una influencia en el timbre de los sonidos. Para ello utilizamos un entorno simulado, para generar respuestas al impulso de salas con paredes revestidas con difusores y las convolucionamos con las señales sonoras de los instrumentos.

El resultado obtenido muestra una cierta tendencia a aumentar el PAT de los sonidos a medida que aumenta el entorno difusivo en el que se encuentran. Esta expansión temporal del tiempo de ataque perceptual, directamente proporcional al uso de superficies difusoras, se puede intentar entender con la teoría de dispersión temporal que sufre las señales sonoras una vez que inciden en un difusor debido al cambio de fase.

Este primer estudio debe servir como una introducción en el estudio temporal de los difusores y su influencia con el PAT de los sonidos. Inmediatamente se pueden abrir varias líneas de investigación, que pueden aportar valiosos datos sobre el funcionamiento temporal de los difusores y su efecto práctico y perceptual.

Por un lado se podría profundizar en la relación evidente entre el rango de frecuencias óptimo de funcionamiento de los difusores y el PAT, ya que se ha comprobado que existe una relación entre el tiempo de ataque perceptual y el contenido espectral de la señal sonora. La otra vía sería cuantificar a nivel perceptual, cuál es el efecto que produce esa variación del timbre de los sonidos al verse modificado su PAT, y establecer rangos de valores máximos de variación en función de los distintos difusores analizados, para no llegar a desvirtuar en exceso la percepción sonora de la señal original.

## **AGRADECIMIENTOS**

En primer lugar quiero dar las gracias a Rubén y Javier ya que gracias a su propuesta sobre el tema de esta tesina pude dirigirla hacia un campo que está muy presente en mi vida y es mi pasión, la música, y agradecerles la gran disposición y paciencia que han mostrado a la hora de plantear reuniones para ir viendo el desarrollo del proyecto. El poder estudiar desde una perspectiva científica este gran arte ha sido muy enriquecedor, y espero que sea tan sólo el comienzo de nuevas incursiones en este campo.

También quiero agradecer a el profesor Miguel Ferrer el haber pasado conmigo toda una tarde resolviéndome dudas sobre Matlab y su amabilidad a la hora de permitirme asistir a las tutorías sin estar matriculado.

Del mismo modo hago extensivo mi agradecimiento a todos los profesores del I Máster de Ingeniería Acústica, ya que han despertado en mi inquietudes, al igual que al resto de mis compañeros, que seguro marcarán de manera decisiva nuestro futuro tanto laboral como personal.

No quiero olvidarme de mis compañeros de máster, llevaré siempre conmigo los buenos ratos pasados y las sesiones de estudio sin respiro para llegar a tiempo con todo. Tampoco olvidaré esa noche en cierta casa y cierta piscina... Donde quiera que esté estaremos siempre en contacto.

De manera especial quiero agradecer a Giusy todo lo que me ha dado en este corto pero intenso año. Su apoyo y su ayuda en los malos momentos me han servido para mantener siempre una sonrisa en mi. Me quedo corto al decirle gracias por darle un nuevo rumbo a mi vida en esta nueva etapa. GRACIAS.

Por último quiero dedicar este proyecto a las dos personas que siempre han creído en mi y me apoyan pase lo que pase.

A mis padres.

## BIBLIOGRAFÍA

- [1] ASA, American Standard Association, *Acoustical Terminology*. New York, 1960.
- [2] ANSI, American National Standards Institute, *American National Standard - Psychoacoustical Terminology*. New York, 1973.
- [3] K. W. Berger, *Some factors in the recognition of timbre*. J. Acoust. Soc. Am., 1964, 36, pp. 1888–1891.
- [4] E. L. Saldanha y J. F. Corso, *Timbre cues and the identification of musical instruments*. J. Acoust. Soc. Am., 1964, 36, pp. 2021–2026.
- [5] J. M. Hajda, *The effect of time-variant acoustical properties on orchestral instrument timbres*. Ph.D. thesis, University of California, Los Angeles, 1999.
- [6] W. Strong y M. Clark, *Perturbations of synthetic orchestral wind instrument tones*. J. Acoust. Soc. Am., 1967, 41, pp. 39-52.
- [7] J. Beauchamp, *Synthesis by spectral amplitude and "Brightness" matching of analyzed musical instrument tones*. J. Acoust. Eng. Soc., 1982, vol. 30, nº. 6.
- [8] S. McAdams y E. Bigand, *Thinking in sound, the cognitive psychology of human audition*. Oxford university press, 1993.
- [9] H. F. Pollard y E. V. Jansson, *A tristimulus method for the specification of musical timbre*. Acustica, 1982, vol. 51.
- [10] A. H. Benade y S. N. Kouzoupis, *The clarinet spectrum: Theory and experiment*. J. Acoust. Soc. Am., 1988, 83(1).
- [11] J. Krimphoff, S. McAdams y S. Winsberg, *Caractérisation du timbre des sons complexes. II Analyses acoustiques et quantification psychophysique*. Journal de Physique IV, 1994, Colloque C5, vol. 4.
- [12] J.W. Gordon, *The perceptual attack time of musical tones*. J. Acoust. Soc. Am., 1987, vol 82, nº1, pp. 88-105.
- [13] R. A. Rasch, *Synchronization in performed ensemble music*. Acustica, 1979, 43, pp. 121-131.
- [14] J. Zwislocki, *Theory of temporal auditory summation*. J. Acoust. Soc. Am., 1960, 32, pp. 1046-1060.

- [15] M. J. Penner, *Evidence for two temporal processes in forward masking*. J. Acoust. Soc. Am., 1980, 68, pp. 455-457.
- [16] H. N. Wright, *Audibility of switching transients*. J. Acoust. Soc. Am., 1960, 32, pp.138.
- [17] J. E. Cutting y B. S. Rosner, *Categories and boundaries in speech and music*. Percept. Psychophys., 1974, 16, pp. 564-570.
- [18] R. W. Schafer y L. R. Rabiner, *A digital signal processing approach to interpolation*. Proc. IEEE, 1973, 61, pp. 692-702.
- [19] E. Zwicker y B. Scharf, *A model of loudness summation*. Psychol. Rev., 1965, 72, pp. 3-26.
- [20] E. Zwicker, *Procedure for calculating loudness of temporally variable sounds*. J. Acoust. Soc. Am., 1977, 62, pp. 675-682.
- [21] J. Vos y R. Rasch, *The perceptual onset of musical tones*. Percept. Psychophys., 1981, 29, pp.323-335.
- [22] A. W. Schloss, *On the Automatic Transcription of Percussive Music From Acoustic Signal to High-Level Analysis*. Ph.D. dissertation, Tech. Rep. STAN-M-27, Dept. Hearing and Speech, Stanford Univ. Stanford, CA, 1985.
- [23] M. Basseville y I.V. Nikiforov, *Detection of Abrupt Changes—Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [24] ISO/DIS 17497-1, *Acoustics — Measurement of the sound scattering properties of surfaces — Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room*. 2001.
- [25] ISO/CD 17497-2, *Acoustics — Measurement of sound scattering properties of surfaces — Part 2: Measurement of the directional diffusion coefficient in a free field*. 2002.
- [26] J.P. Bello, L. Daudet, *et al.*, *A Tutorial on Onset Detection in Music Signals*. IEEE Transactions on Speech and Audio Processing, 2005, vol. 13, n°. 5, pp. 1035-1047.
- [27] A. Caclin, S. MacAdams, *et. al*, *Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones*. J. Acoust. Soc. Am., 2005, pp. 471–482.
- [28] T. J. Cox y P. D'Antonio, *Acoustic absorbers and diffusers. theory, design and application*. Spon Press. 2004.
- [29] K. Jensen y G. Marentakis, *Hybrid Perception*. Papers from the 1<sup>st</sup> Seminar on Auditory Models, Lyngby (Denmark), 2001.
- [30] K. Jensen, *Perceptual and physical aspects of musical sounds*. *Journal of Sangeet Research Academy. India*, 2002.

- [31] W.A. Sethares, *Tuning, timbre, spectrum, scale*. Springer. 2005.
- [32] M. Wright, *The shape of an instant: Measuring and modeling perceptual attack time with probability density functions*. Ph.D. Dissertation. Stanford University, 2008.