

# PRIMERAS JORNADAS REGIONALES DE ACÚSTICA AdAA 2009

19 y 20 de noviembre de 2009, Rosario, Argentina



AdAA2009-A053R

## Re-descubriendo la ventana de integración temporal del sistema auditivo: consecuencias en la ciencia Acústica

Alejandro Bidondo <sup>(a)</sup>

(a) Ingeniero Electrónico UBA. Coordinador de la carrera Ingeniería de Sonido de la Universidad Nacional de Tres De Febrero, Valentín Gómez 4752, Buenos Aires, Argentina. abidondo@untref.edu.ar

**Abstract:** On this research the variability of the integration window duration of the auditory system in function of the perceived acoustic signal is verified. The obtained results validate the perception model based on the specialization of the brain's hemispheres and the existence of the autocorrelation and cross-correlation functions achieved at the brainstem. It is essayed a formulae of the duration of the temporal window in function of the effective duration of the autocorrelation's (ACF's) envelope of the sound stimuli.

**Resumen:** En el presente trabajo se comprueba la variabilidad de la duración de la ventana de integración temporal del sistema auditivo en función de señal acústica percibida. Los resultados obtenidos validan el modelo de percepción basado en la especialización de los hemisferios cerebrales y la existencia de las funciones matemáticas de autocorrelación y correlación cruzada realizados en el tronco cerebral. Se ensaya una fórmula para predecir la duración de la ventana de integración temporal en función de la duración efectiva de la envolvente de la función de autocorrelación (ACF) del estímulo sonoro.

## 1. Introducción

### 1.1 Análisis de los modelos actuales de ventanas de integración

Las líneas de investigación relativas a la duración de la ventana de integración del sistema auditivo hasta ahora conocidas pueden dividirse claramente en dos:

Aquellas que establecen una ventana temporal fija (Rossing, 2007) y por lo tanto someten los estudios de las variables subjetivas a parámetros dependientes solamente de la acústica de la sala bajo análisis, con límites temporales fijos, sin suficiente sustento científico.

Las otras que, partiendo de noveles estudios de la fisiología de la audición, someten todo análisis a dos tipos de procesamientos cerebrales: temporal y espacial. Esto hace que los parámetros subjetivos por describir ya no sólo dependan de la acústica, sino también de capacidades probadas de procesamiento del sonido del cerebro humano. Indudablemente, esto introduce escalas subjetivas, las que muy eficientemente utilizando la Ley del juicio comparativo, aplicado sobre normoyentes sin entrenamiento auditivo previo, unen claramente los fenómenos físicos con las sensaciones humanas de la audición.

Se pretende demostrar qué línea de estudio tiene el sustento más sólido y a partir de allí tratar diversos temas en términos de la aplicación de la línea científica corroborada por los ensayos, entre ellos la espacialidad (ASW) y el involucramiento (LEV).

### 1.2 El fenómeno de la audición: de la cóclea al cerebro

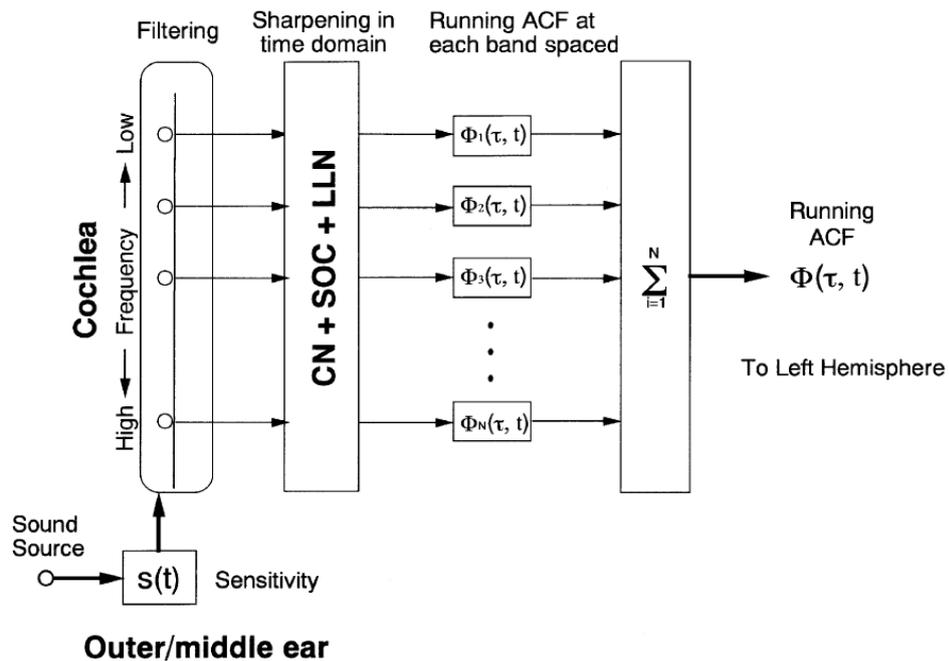
En Tohyama et al., 1995 y Ando, 1994, luego de años de investigación acerca de la respuesta cerebral a diversos sonidos percibidos, mediante SVR's (*Slow Vertex Responses*) y ABR's (*Auditory Brainstem Responses*), métodos de análisis electro-neuro-fisiológicos que permiten registrar las señales electro-químicas que atraviesan el tronco cerebral, plantean un modelo de la audición basado en la especialización del procesamiento de la información sonora por parte de cada hemisfero.

La densidad espectral de potencia de la cóclea puede ser mapeada en cierta posición neuronal como una actividad temporal. Estas actividades neuronales poseen la información suficiente como para lograr procesar la autocorrelación (ACF) de la información sonora. De hecho, la tasa de disparo del nervio auditivo de un gato revela un patrón de autocorrelación en vez de un análisis en frecuencias (Secker-Walker et al., 1990).

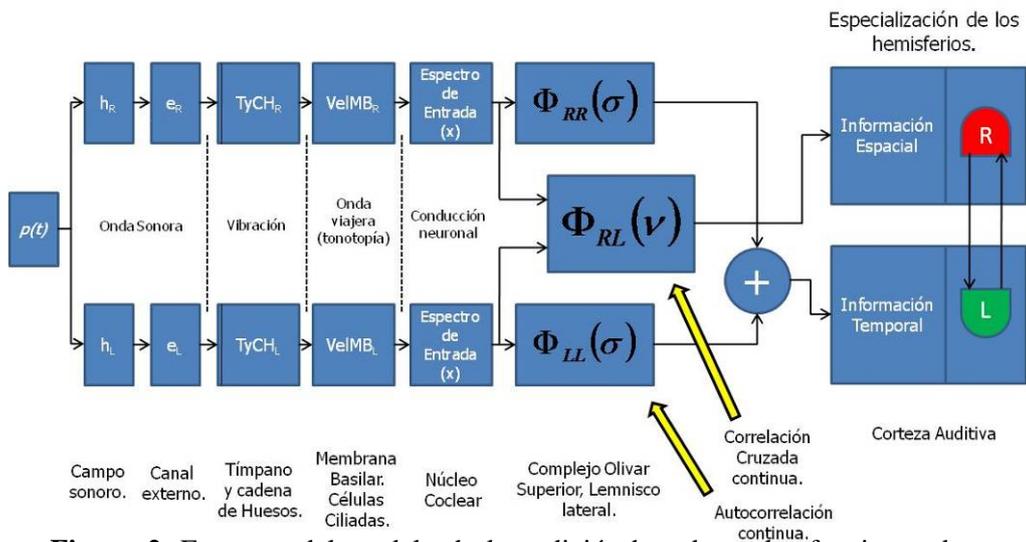
Las distribuciones de los intervalos de disparo en un conjunto de neuronas constituyen un método de codificación, transmisión y representación general y distribuida en el tiempo de la información sonora. Considerando esto es que la unión (por bandas de frecuencias) de estas distribuciones de disparos se asemeja a la envolvente de una función de autocorrelación continua (Cariani, et al., 1996). El diagrama en bloques de lo descrito se muestra en la figura 1.

En la figura 2 se presenta el esquema, el que incluye las funciones matemáticas superiores de cada centro de procesamiento; el lemnisco lateral para la función autocorrelación (ACF) y el colículo inferior y complejo olivar superior para la función correlación cruzada (CCF).

Si el enmascaramiento temporal (*forward masking*) fuese constante para toda señal acústica, haría referencia a una incapacidad de resolución mecánica de la membrana basilar, sosteniendo la teoría del lugar y de patrones espectrales. Por otro lado, si la duración de la ventana de integración fuese variable en función de ciertas características de los estímulos sonoros, haría referencia directa al modelo de resolución temporal del "pitch" y demás características de los estímulos sonoros.



**Figura 1.** Modelo de procesamiento neuronal de la ACF continua (Ando, 2002).



**Figura 2.** Esquema del modelo de la audición basado en las funciones de autocorrelación y correlación cruzada propuesto por Ando (Ando, 1998).

Siendo:

$\Phi_{LL}, \Phi_{RR}$  : Las funciones de autocorrelación izquierda y derecha respectivamente.

$\Phi_{RL}$  : La función de correlación cruzada entre los canales izquierdo y derecho.

Entonces, en el esquema anterior se observa que el hemisferio izquierdo del cerebro recibe la información correspondiente a la autocorrelación de cada señal sonora (Izquierda y Derecha), mientras que el hemisferio derecho recibe la correlación cruzada de las informaciones Izquierda y Derecha. Este modelo refleja la especialización en el procesamiento de señales de cada hemisferio cerebral: el izquierdo procesa información

temporal (palabra hablada e identificaciones secuenciales en el tiempo) y el derecho, información espacial (Ando, 1998).

### 1.3 Codificación temporal de formas auditivas

Genéricamente, los códigos temporales son códigos compuestos por pulsos neuronales en los cuales los tiempos relativos entre los disparos transportan información. Esto es, cómo las neuronas disparan, respecto de los patrones espaciales de la activación de las mismas, o sea qué neuronas disparan más en función de la tonotopía presentada por la membrana basilar.

La codificación temporal de información sensorial es posible cuando hay alguna correlación entre la forma de onda del estímulo y la probabilidad de descarga de neuronas.

Esta correlación puede ser producida por receptores que siguen algún aspecto de la onda estímulo de forma tal que ésta imprime su estructura temporal sobre el disparo de las mismas.

Virtualmente en toda modalidad sensorial del ser humano existe algún aspecto cuya percepción podría corresponder a una codificación temporal de la información (Cariani, 1997 1995; 2001).

En particular, en el sistema auditivo es especialmente evidente la existencia de una estructura temporal producida por el estímulo sonoro. Hay evidencias psicofísicas y neurofisiológicas que sugieren que dicha información temporal contiene la información de importantes cualidades musicales, como ser pitch, timbre y ritmo [8].

Del estudio de las respuestas del ser humano al tiempo de reverberación y al IACC en electroencefalogramas (EEG) y del nivel sonoro e IACC en magnetoencefalogramas (MEG) sobre los hemisferios izquierdo y derecho, se halla que la escala de preferencias subjetivas corresponden a valores relacionados con la duración efectiva de la envolvente de la autocorrelación,  $\tau_e$ , el cual es extraído de la ACF de la onda alfa ( $\alpha$ ) del EEG (del hemisferio izquierdo) y del MEG. En particular, para el EEG, cuando se varía uno de los factores temporales y espaciales del campo sonoro, en aquella combinación preferida, el cerebro tiende a “disfrutarlo” por más tiempo, correspondiendo esto al  $\tau_e$  de la onda alfa ( $\alpha$ ) (Ando, 2008).

### 1.4 Función de autocorrelación (ACF) de largo término

Se define por medio de la siguiente fórmula:

$$\Phi_p(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} p'(t) p'(t + \tau) dt \quad (1)$$

$$p'(t) = p(t) * s(t) \quad (2)$$

Siendo  $s(t)$  la sensibilidad del oído, dada por la ponderación “A” y  $p(t)$ , la presión sonora instantánea.

Esta función, calculada sobre un período de integración ( $2T$ ) de un archivo de audio (de allí su nombre “de largo término”), da por resultado su *auto-similaridad en función del tiempo*, o sea otra función que refleja cuánto se parece la señal bajo análisis a sí misma, a medida que pasa el tiempo. La ACF posee varios parámetros microscópicos significativos, a saber (Ando, 1998b):

- $\Phi_p(0)$  es la energía contenida en la señal, representada por la magnitud de la señal en  $t=0$ . Es posible conocer el nivel sonoro de una señal midiendo el valor de la función de

autocorrelación en  $t = 0$ , o sea,  $\Phi_p(0)$ , aplicando el tiempo de integración  $2T$  apropiado, correspondiente a la duración de la ventana de integración temporal del sistema auditivo.

- La “duración efectiva de la envolvente de la ACF normalizada”, esto es  $\tau_e$ , el cual está definido por aquel instante temporal en que la amplitud de la ACF normalizada es -10dB (respecto de su máximo en  $t=0$ ). Este valor representa una propiedad repetitiva o “reverberación” contenida dentro de la señal.  $\tau_e$  Es la unidad de tiempo fundamental de todo campo sonoro en una sala de conciertos y de aquel, su valor mínimo es la parte más activa de la señal, existiendo muchos  $\tau_{e,min}$  “locales” durante toda porción de sonido analizada.
- El retardo de tiempo y la amplitud del primer máximo de la ACF,  $\tau_1$  y  $\phi_1$ . Estos descriptores son aquellos que utiliza el modelo de la audición moderno de Ando para la detección de la altura tonal (*pitch*) del sonido.
- La estructura fina de la ACF, tanto en amplitudes como en instantes de tiempo posteriores a  $\tau_1$ , que corresponden a las amplitudes positivas y sus respectivos tiempos de aparición respecto del máximo en  $t=0$ .

Cabe mencionar que  $\tau_e$  varía en función del intervalo  $2T$  seleccionado para realizar la integración sobre la señal bajo análisis.

El tiempo de integración apropiado, según Ando, 2009, va de 30ms a 1000ms, dependiendo de la fuente de señal, para una resolución apropiada de la ACF de la señal bajo estudio. Para ser precisos, recomienda  $[2T] = 30 \cdot \tau_{e,minimo}$

El  $\tau_{e,min}$  representa la densidad de información de las fuentes sonoras. Una fuente sonora con un valor corto de  $\tau_{e,min}$  contiene información más densa (Kato et al., 2004). Esto es, mayor cantidad de información<sup>1</sup> que el promedio en la unidad de tiempo, igual información en menor tiempo o ambas.

Este valor  $2T$ , duración de la ventana de integración temporal del sistema auditivo humano, depende de la fuente sonora bajo escucha; en particular depende de la “duración efectiva de su ACF” local. Desde el punto de vista psicológico, según Fraisse, 1967, el período de integración sería el presente psicológico, o sea,  $2T \cong 0.1s \sim 5s$ .

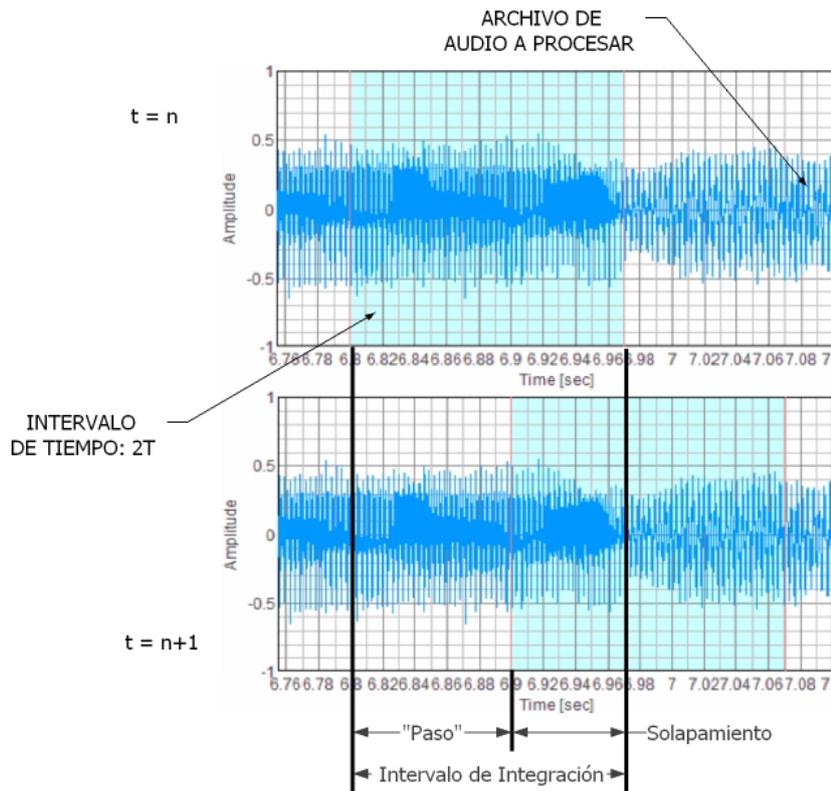
### 1.5 Función de autocorrelación (ACF) de corto término o móvil

$$\phi_p(\tau) = \phi_p(\tau; t, T) = \frac{\frac{1}{2T} \int_{t-T}^{t+T} p'(s) \cdot p'(s + \tau) ds}{\sqrt{[\Phi_p(0; t, T) \cdot \Phi_p(0; t + \tau, T)]}} \quad (3)$$

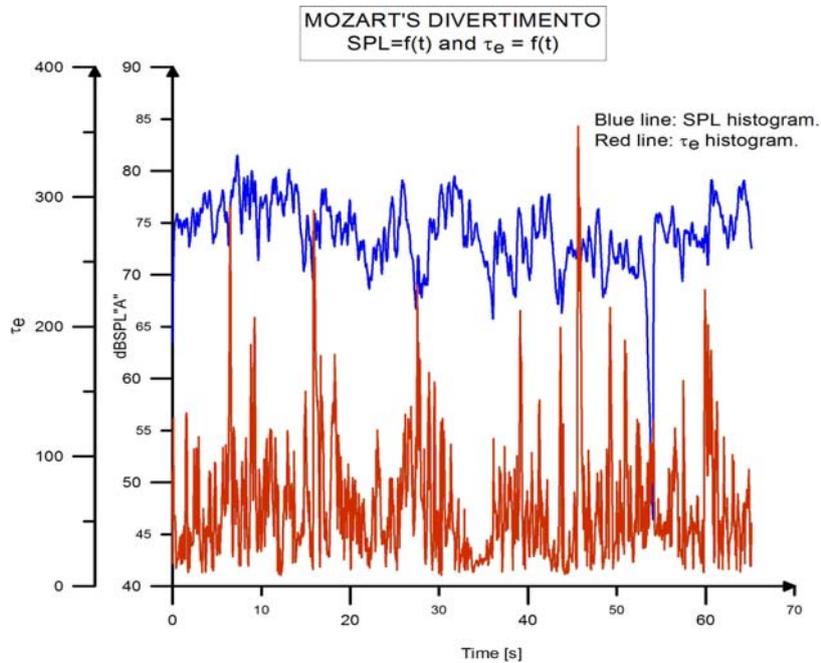
En este caso, la *auto-similaridad* de la señal en función del tiempo es calculada durante el intervalo de integración, el cual es corrido en el tiempo para todo el archivo sonoro bajo análisis, según un valor establecido por el *running step* (equivalente al concepto de solapamiento), según se observa en la Figura 3. Este análisis permite, entre otras cosas,

<sup>1</sup> El término cantidad de información hace referencia a la valoración de la misma, donde el valor más alto se le asigna al mensaje que menos probabilidades tiene de ser recibido o de suceder (ver [http://es.encarta.msn.com/encyclopedia\\_761577650/Teor%C3%ADa\\_de\\_la\\_informaci%C3%B3n.html](http://es.encarta.msn.com/encyclopedia_761577650/Teor%C3%ADa_de_la_informaci%C3%B3n.html))

obtener los valores de ACF y del  $\tau_e$  en forma de histograma, según se puede observar en las Figura 4, para el motivo musical “Mozart’s Divertimento”.



**Figura 3.** Esquema del funcionamiento del cálculo de la ACF móvil en un archivo de audio



**Figura 4.**  $SPL=f(t)$  y  $\tau_e=f(t)$ , para el motivo musical Mozart’s Divertimento

Cabe mencionar que utilizar un mayor tiempo de integración para el cálculo de la ACF oficiaría de filtro pasa bajos de la información (en un modelo eléctrico equivaldría a incrementar el valor de un capacitor en un filtro pasa bajos RC, por lo que la constante de carga se incrementaría y su frecuencia de corte se reduciría), por lo tanto suavizaría los resultados, arrojando diferentes valores de  $\tau_{e,min}$ , que a veces son sustancialmente diferentes de aquellos correspondientes a los valores de  $2T$  reales.

El estudio en forma de histograma del  $\tau_e$ , permite identificar aquellos instantes en que se presentan los varios  $\tau_{e,min}$  “locales”, momentos en que la atención auditiva es máxima, instantes que determinarían fuertemente la duración de la ventana de integración del sistema auditivo Ando, 1998 y 2008.

En estudios previos, (Kato et al., 2002; Mouri et al., 2001; Kato et al., 2004), sólo se ha computado un único  $\tau_{e,min}$  dado que la duración de los archivos analizados ha sido corta, es decir entre 2 y 20 segundos y/o las fuentes sonoras fueron mayoritariamente simples (no eran la suma de varias fuentes sonoras de gran dinámica y variadas expresiones), por ejemplo: diversos vocales, timbre de teléfono, equipo de aire acondicionado funcionando, quinteto de cuerdas, etc. Esto no permitió observar varios  $\tau_{e,min}$  locales dentro de una pieza musical (en este caso la duración aproximada de los motivos musicales es de 1 minuto), lo que resultaría en valores extremadamente bajos, instantes de gran sensibilidad al periodo de integración  $2T$  y sugerirían la constante adaptatividad de la duración de la ventana de integración del sistema auditivo.

¿Cuántos  $\tau_{e,min}$  existen en una pieza musical prolongada?, ¿Cuál es el nivel a partir del cual el cerebro considera un  $\tau_e$  como mínimo local, a modo de dominar la duración de la ventana de integración durante cierto periodo de tiempo? ¿Dependerá esto del deseo de escuchar?, El procesamiento dinámico de señales de audio, ¿modifica los  $\tau_{e,min}$  locales?, de hacerlo, ¿modifica su ubicación dentro del motivo musical, sus valores o ambos?

De lo enunciado anteriormente, se extrae que (Ando, 2002):

Los parámetros acústicos temporales son el tiempo de reverberación (RT60) y la percepción del *delay* (Ej: ITDG o  $\Delta t_1$ ).

Los parámetros espaciales son el IACC y el nivel sonoro (Ando, 2008).

Los parámetros definidos en términos de especialización cerebral son ortogonales entre sí y dependen fuertemente de la duración efectiva de la envolvente de la ACF de la fuente sonora,  $\tau_e$ .

Se hallaron valores de preferencia para RT60 e ITDG en función del  $\tau_e$ .

La audición se adapta al  $\tau_e$  del programa sonoro, por lo que el efecto precedencia debe ser re-definido.

Establecido esto y a modo de sustentar el desarrollo de la teoría de redes neuronales con disparos adaptables en el tiempo basadas en señales codificadas temporalmente (Cariani, 1997), sólo resta verificar de la forma más directa posible el modelo presentado.

## 2 Un simple experimento

Una de las conclusiones de la teoría de la codificación temporal es la adaptatividad de la audición a la duración efectiva mínima de la envolvente de la ACF. Por la negación, si de alguna forma se probase la no adaptatividad de la audición al  $\tau_{e,min}$  de diversos programas sonoros, esto es observar una ventana de integración constante, se negaría automáticamente la hipótesis.

Se procedió entonces a relevar la ventana temporal de la ley del primer frente de onda o efecto precedencia, con 4 temas musicales.

Se desarrolló un protocolo para la evaluación de la duración de la ventana de audición humana. Las variables ortogonales entre sí tomadas en cuenta para el mismo fueron el  $\tau_e$  (mediante el uso de diferentes motivos musicales) y la lateralización de la reflexión.

## 2.1 Experimento

Se relevaron las respuestas de 8 sujetos, normoyentes, con audición semi-entrenada, capaces de comprender las instrucciones.

Fueron sometidos a cuatro temas musicales, todos con diferentes  $\tau_{e,min}$ , con sólo una reflexión. Ésta primeramente se ubicó al centro de la imagen sonora y posteriormente se repitió todo el experimento ubicándola a 45% del paneo sobre un sistema estereofónico (equivalente a 40.5° respecto del eje central del oyente).

Los programas sonoros utilizados fueron un tema de Jazz anecoico monofónico sin voz cantada, una parte para cuarteto de cuerdas de Mozart monofónica llamada “Divertimento”, un vocal femenino anecoico de música pop y una parte de un concierto de órgano de Orlando Gibbons llamado “Fantasía”. Estas condiciones imponen diversos histogramas de  $\tau_e$ .

El monitoreo fue realizado por medio de auriculares (*headphones*) marca Audio Technica modelo ATH D40.

Se mantuvo el nivel sonoro constante entre ensayos; el nivel sonoro continuo equivalente de los mismos fue de 75dBA. Este nivel sonoro fue registrado dentro de la cavidad conformada por los auriculares, con un sonómetro Svantek SV959, clase 1. La interface de audio utilizada fue marca M-Audio, modelo 410, con controles de nivel de auriculares.

Se reprodujo un campo sonoro compuesto por el sonido directo y una única reflexión, con su tiempo de retraso variable. Los retardos establecidos fueron de 20 ms, 50 ms, 100 ms, 200 ms y 500 ms.

Se realizaron 4 tests sobre cada uno de los sujetos para cada una de las reflexiones generadas, con el objetivo de finalmente promediar los resultados y obtener así una media lo más precisa posible. Se analizaron las varianzas de los resultados producidos por cada oyente, para cada retraso utilizado. Se descartaron aquellos registros con gran variancia dado que mostraban poco entrenamiento auditivo del oyente, la no comprensión de las premisas o el agotamiento físico.

Se reprodujo la señal directa con 0 dB de amplitud relativa.

Se reprodujo la señal retrasada iniciando con una amplitud nula e incrementándose en pasos de 3 dB y de 0.1 dB. El software utilizado fue de marca Adobe, modelo Audition.

A los oyentes se les solicitó establecer el nivel, respecto del sonido directo, al cual “apenas percibían” la señal retardada, pudiendo ellos revertir el proceso en vez de incrementar paulatinamente la señal retrasada. También podían desactivar la misma siempre con el objetivo de juzgar el nivel al cual “apenas percibían” la señal retrasada. Sobre esta instrucción debió aclararse que el límite de detección o decisión fuese mantenido constante durante todo el ensayo.

El orden de las pruebas fue aleatorio, tanto en temas musicales elegidos como en el retraso de la reflexión a evaluar.

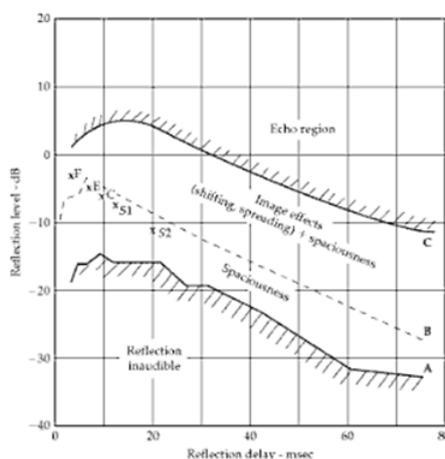
El nivel relativo final de la señal retardada detectada se tomó como resultado en cada experimento.

Cabe mencionar que en el cálculo de  $\tau_{e,min}$  de los temas musicales utilizados no se descartaron aquellos valores resultantes de intervalos de respiraciones o de cuasi-silencios, dado que precisamente en dichos momentos los oyentes también podían percibir los “cues”<sup>2</sup> que les permitieron realizar las valoraciones requeridas para el experimento.

---

<sup>2</sup> El término “cue” o “music cue” se origina en las producciones teatrales y está relacionado con la inserción de música o de un instante en particular de la misma el que ordena a un actor o auxiliar de escenario llevar a cabo una acción o recitar una línea de diálogo o texto (ver <http://www.mediamusicnow.co.uk/Glossary->

Se debió instruir previamente a los sujetos acerca de los diferentes límites de percepción de una única reflexión, límite de detección por nivel sonoro, espacialidad o corrimiento de la imagen y fuentes separadas, según lo hallado por Olive y Toole en Olive et al., 1988, resumizado en la figura 5.



**Figura 5.** Límites de percepción de una reflexión (Olive et al., 1988).

Esta instrucción inicial permitió que los sujetos mantuvieran el tipo de percepción constante en cada ensayo, y así poder obtener resultados válidos.

Es importante tener en cuenta que los temas musicales sólo ofician a modo de señales acústicas, en este caso particular identificables como “música”, con diferentes  $\tau_e$  y por consiguiente con diferentes  $\tau_{e,min}$ . Cabe recordar que el objetivo es poder relacionar la memoria mínima de una señal con la duración de la ventana de integración.

Las instrucciones entregadas a los oyentes fueron:

1. Calibrar el nivel de los auriculares a  $L_{eq} = 75$  dBA para el tema bajo análisis.
2. Luego no tocar el nivel sonoro de los mismos durante el experimento.
3. Establecer el ensayo particular encendiendo y apagando los canales correspondientes.
4. Verificar que el nivel del potenciómetro de la señal no retrasada esté en 0 dB.
5. Oír cuidadosamente y recordar la pieza musical sin retraso alguno.
6. Establecer el retardo particular del ensayo a realizar.
7. Establecer el paneo del canal a retrasar del ensayo particular.
8. Verificar que el nivel del canal retrasado esté por debajo de los -80 dB, antes de iniciar el ensayo particular.
9. Incrementar el nivel del canal retrasado hasta apenas percibirlo, por medio de un ajuste de nivel sonoro “grueso” (pasos de +/-3 dB).
10. Variar el nivel del canal retrasado hasta apenas percibirlo, por medio de un ajuste fino de nivel (+/-0.1 dB).
11. Registrar el nivel del potenciómetro del canal retrasado en a tabla de excell del experimento particular.

---

music-production-terms/music-cue/). En este caso el “cue” es un instante que le otorga ciertos datos al oyente acerca de la fuente sonora que está percibiendo.

12. Una vez finalizadas las 4 detecciones del canal retrasado por retardo elegido, cambiar aleatoriamente al próximo ensayo.

Además:

- No observar el nivel relativo del canal retrasado. Esto podría llevar a errores de resultado.
- Está permitido utilizar el encendido y apagado del canal retrasado para detectar la presencia del mismo.
- Mantener siempre el mismo tipo de detección durante todos los ensayos del experimento.

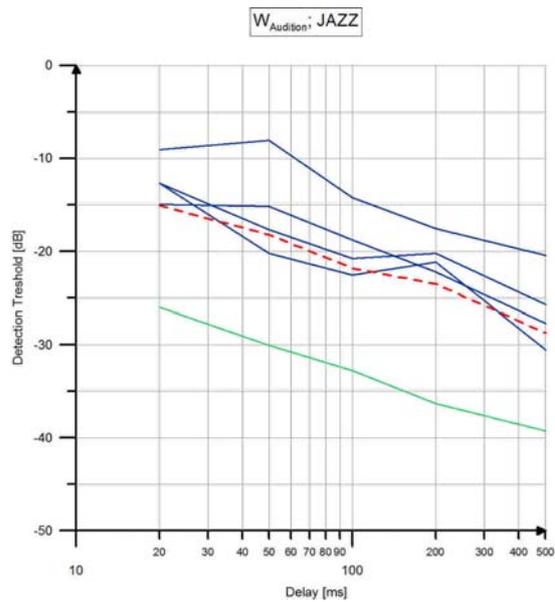
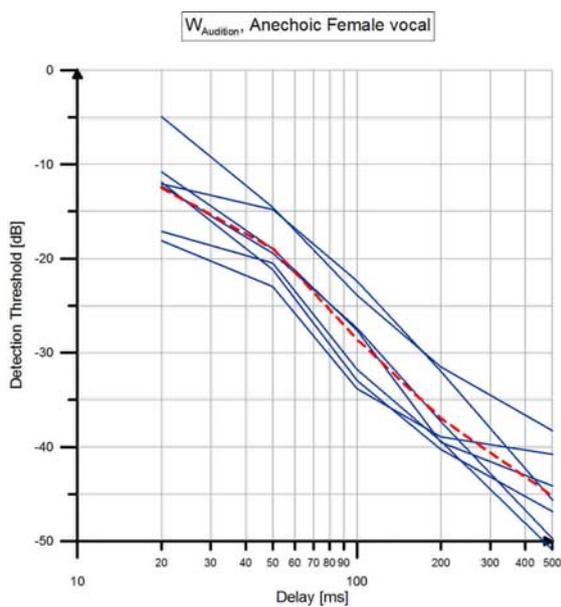
La tabla que debió ser completada por los oyentes fue la mostrada en la tabla 1.

**Tabla 1.** Tabla a completar por los oyentes sometidos al experimento.

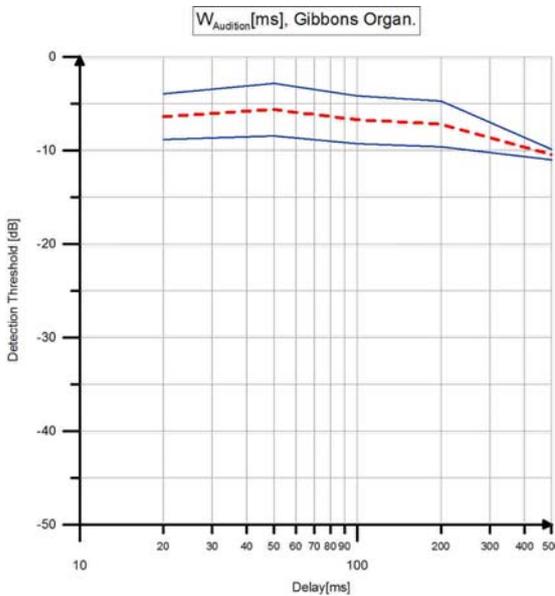
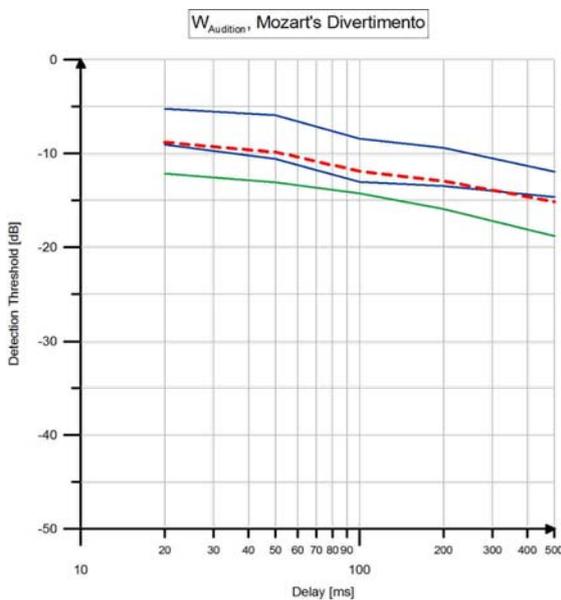
SOUND LEVEL	Music motif 1						
	Le mín [ms]		Organ Gibbons				
	Mean RMS value [dBA]	-12	Subject's Name	John Sciamarella			
75Db	Panning	0%	Test 1	Test 2	Test 3	Test 4	AVERAGE
	<input type="checkbox"/> t1	20ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-2	-3.1	-1.5	-3.3	-2.5
	<input type="checkbox"/> t1	50ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-3.7	-2.9	-3.5	-4	-3.5
	<input type="checkbox"/> t1	100ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-4.1	-5.2	-3.6	-3.5	-4.1
	<input type="checkbox"/> t1	200ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-5.7	-6	-5.4	-6.5	-5.9
	<input type="checkbox"/> t1	500ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-22	-21	-25	-23	-22.8
	Panning	45%	Organ Gibbons				
	<input type="checkbox"/> t1	20ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-2.9	-3.1	-2.4	-3.6	-3.0
	<input type="checkbox"/> t1	50ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-5.6	-4.3	-5.1	-7.2	-5.6
	<input type="checkbox"/> t1	100ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
	Echo perception		-8.8	-8.2	-7	-9	-8.3
	<input type="checkbox"/> t1	200ms	Test 1	Test 2	Test 3	Test 4	AVERAGE
Echo perception		-11	-9.3	-8.9	-10.6	-10.0	
<input type="checkbox"/> t1	500ms	Test 1	Test 2	Test 3	Test 4	AVERAGE	
Echo perception		-17.1	-15.7	-11.1	-16.9	-15.2	

### 3 Resultados

En las figuras 6, 7, 8 y 9 se observan los resultados obtenidos de los ensayos sobre diversos oyentes con las fuentes sonoras “Anechoic Female Vocal”, “Jazz”, “Mozart’s Divertimento” y “Gibbon’s Organ”. Los resultados se muestran para una única reflexión proveniente de la misma dirección que el sonido directo. En la línea roja punteada se muestra el promedio de los resultados de todos los sujetos.

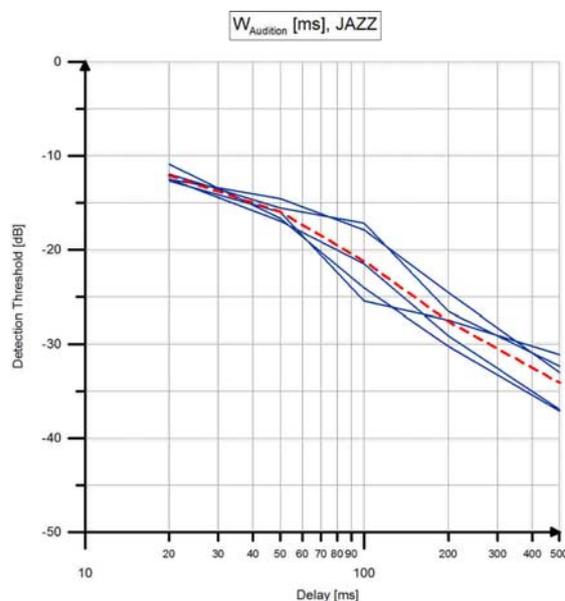
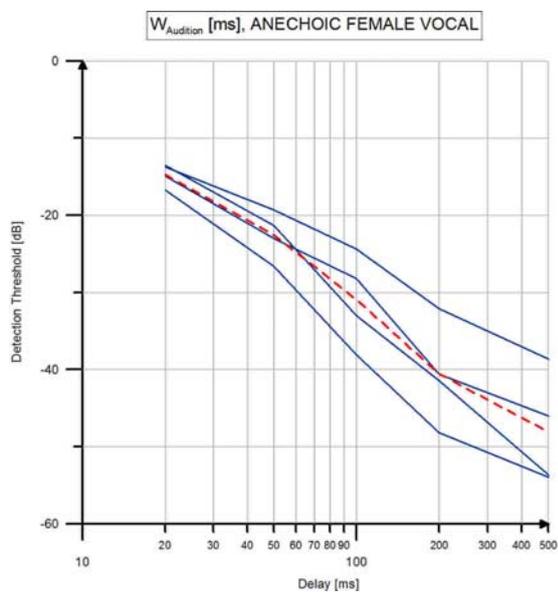


**Figuras 6 y 7.** Resultados de la percepción de una única reflexión proveniente del centro en los motivos musicales “Vocal femenino anecoico” y “Jazz”. (Ver definición de  $W_{Audition}$  en pág. 14)

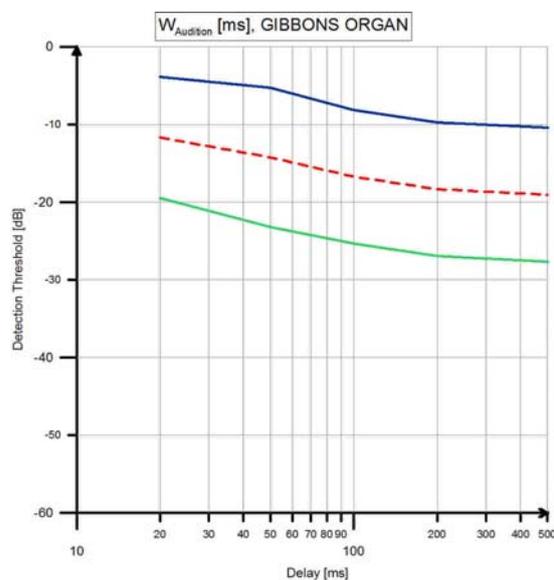
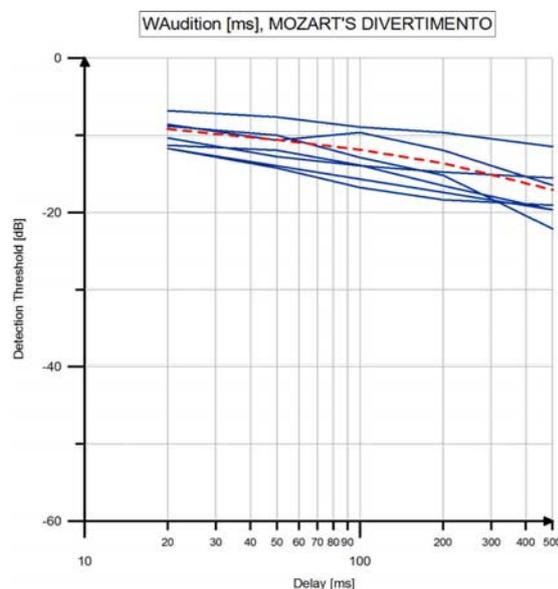


**Figuras 8 y 9.** Resultados de la percepción de una única reflexión proveniente del centro en los motivos musicales “Mozart’s Divertimento” y “Gibbon’s Organ”. (Ver definición de  $W_{Audition}$  en pág. 14)

En las figuras 10, 11, 12 y 13 se observan los resultados obtenidos de los ensayos sobre diversos oyentes con las fuentes sonoras “Anechoic Female Vocal”, “La Cumparsita Tango”, “Jazz”, “Mozart’s Divertimento” y “Gibbon’s Organ”. Los resultados se muestran para una única reflexión de proveniencia lateralizada a un 45% (40.5°) respecto del centro. En la línea roja punteada se muestra el promedio de los resultados de todos los sujetos.



**Figuras 10 y 11.** Resultados de la percepción de una única reflexión lateralizada en los motivos musicales “Vocal femenino anecoico” y “Jazz”. (Ver definición de  $W_{Audition}$  en pág. 14)



**Figuras 12 y 13.** Resultados de la percepción de una única reflexión proveniente del centro en los motivos musicales “Mozart’s Divertimento” y “Gibbon’s Organ”. (Ver definición de  $W_{Audition}$  en pág. 14)

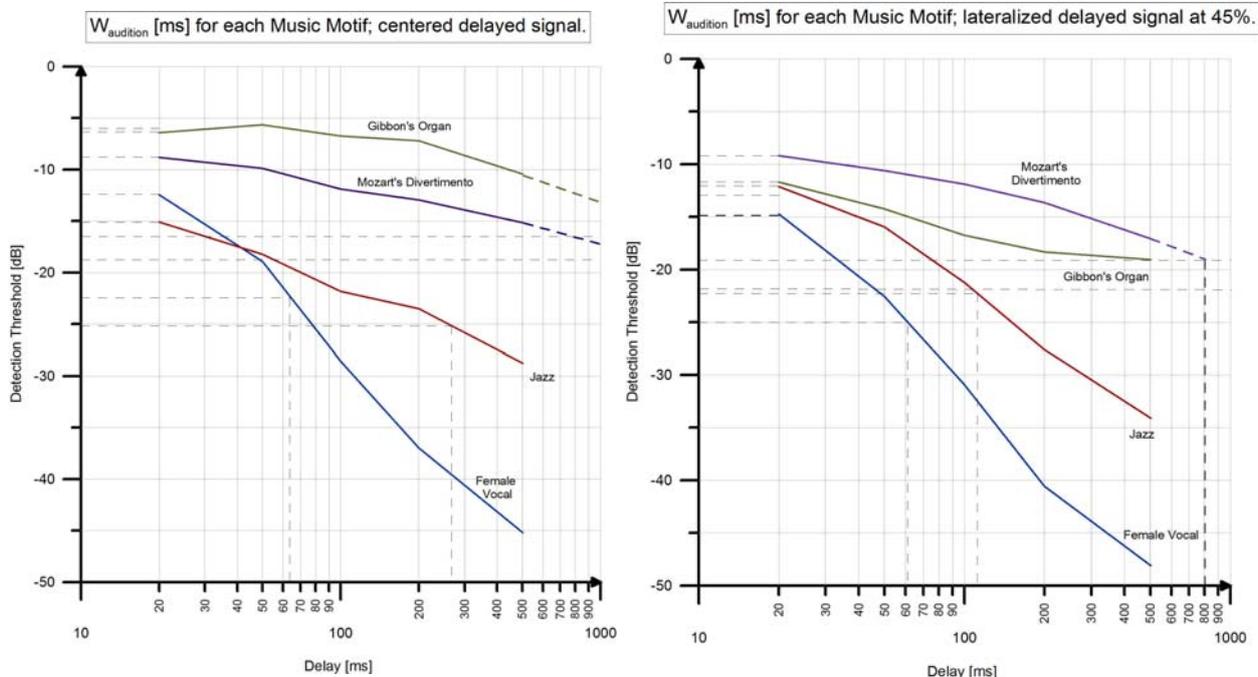
Además, en la Tabla 2 se pueden observar los instantes de detección proporcionados por uno de los sujetos que realizaron el experimento sobre el motivo musical ‘Mozart’s Divertimento’.

**Tabla 2.** Instantes de detección de los ‘cues’.

Subject Result's				
Music Motif	MOZART'S DIVERTIMENTO			
	Detection Threshold		Detection instant [s]	
Delay [ms]	Test 1	Test 2	Test 1	Test 2
20	-12.6	-11.7	12	35
50	-13.4	-12.8	62	20
100	-14.6	-13.9	10	48
200	-16.2	-15.7	8	55
500	-18.8	-18.9	44	32

Los datos encontrados en la tabla 1 sólo tienen por fin confirmar la existencia de varios  $\tau_{e,min}$  locales dentro de una pieza musical, los que proporcionan los llamados ‘cues’, lo que motiva el uso y aplicación de un descriptor estadístico: el percentil. Mediante la iteración se obtuvo aquel percentil relevante para los resultados obtenidos (ver figuras 16 y 17).

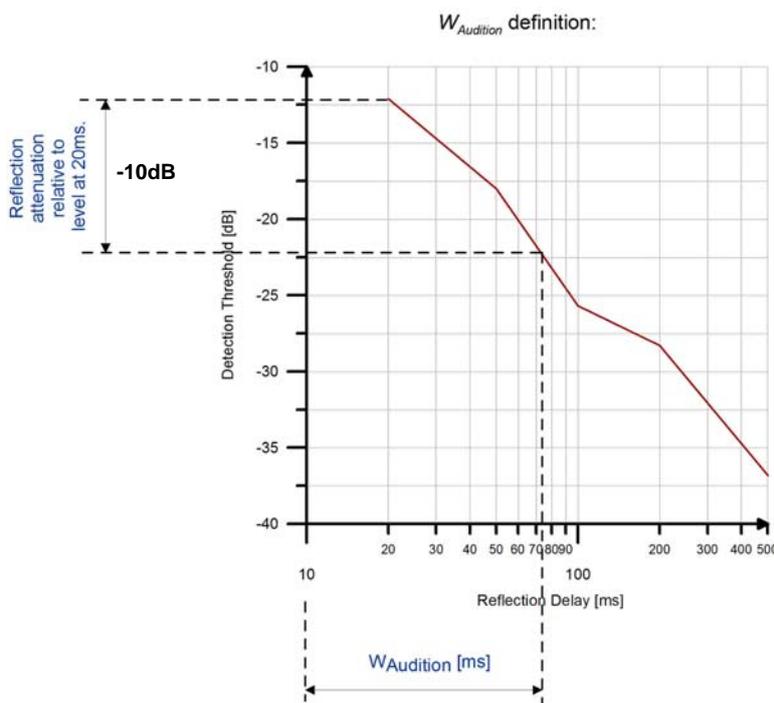
En la figura 14 se observa el conjunto de resultados para la información retrasada ubicada al centro. En la figura 15 se observa el conjunto de resultados para la reflexión retrasada con proveniencia lateralizada a un 45% (40.5°) respecto del centro.



**Figuras 14 y 15.** Resultados promedios de la totalidad de motivos musicales analizados, para una reflexión centrada (14) y lateralizada (15). (Ver definición de  $W_{Audition}$  en pág. 14)

Lo importante de las curvas arriba presentadas son las pendientes y sus formas, no los límites relativos entre ellas, ya que los oyentes tuvieron la libertad de elegir el tipo de detección de la reflexión retrasada (ver figura 5).

Se define la duración de la ventana de audición,  $W_{Audition}$  [ms], como el intervalo entre el instante de percepción de la señal directa y aquel en que la percepción de la reflexión es -10 dB respecto del nivel del retraso de 20 ms. Esto se observa en la figura 16.



**Figura 16.** Definición de la ventana de audición  $W_{Audition}$

Según lo observado, las duraciones de las ventanas de audición,  $W_{Audition}$ , y demás datos relevantes, para los diferentes motivos musicales y ambas proveniencias de la señal retrasada (al centro y lateralizada), se muestran en la tabla 3.

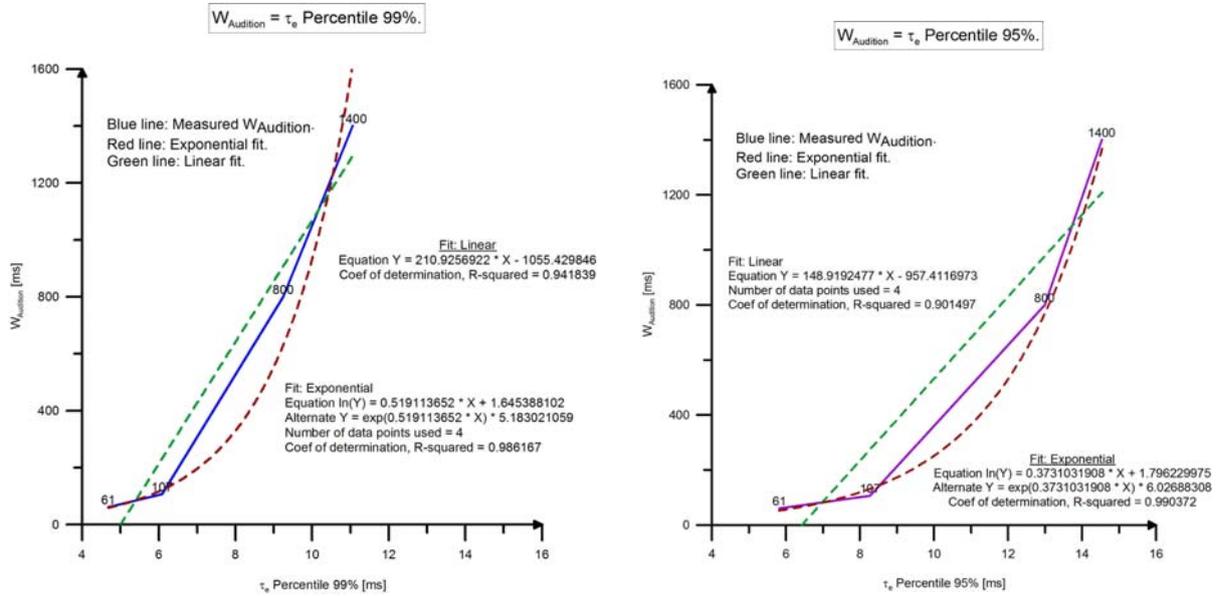
**Tabla 3.** Resultados experimentales de la duración de la ventana de audición  $W_{Audition}$  [ms] (Ver definición de  $\tau_e$  Percentile en pág. 15)

Music Motif	Panning = 0%		Panning = 45%					
	$W_{audition}$ [ms] = 2T	$W_{audition}$ [ms] = 2T	$\tau_e$ Average [ms]	$\tau_{emin}$ [ms]	$\tau_e$ Percentile 99% [ms]	$\tau_e$ Percentile 95% [ms]	$\tau_e$ Percentile 90% [ms]	$\tau_{l1}$ Average [ms]
Anechoic Female Vocal	63	61	224.307	3.680	4.682	5.820	6.830	0.940
Jazz	250	107	87.936	5.050	6.086	8.272	9.276	1.833
Mozart's "Divertimento"	800	800	40.563	7.570	9.247	13.045	15.043	1.077
Gibbons Organ	1400	1400	83.416	8.520	11.059	14.438	16.590	1.366

Visto que en toda situación real la información reflejada no proviene estrictamente de la misma que aquella directa, se realizaron los cálculos en las ACFs con las duraciones correspondientes a las reflexiones lateralizadas.

Los datos de la tabla 2 fueron obtenidos haciendo  $2T = W_{Audition}$ , relevados del experimento.

Realizando ajustes con ecuaciones tanto lineales como exponenciales, para  $\tau_e$  Percentile 99% y  $\tau_e$  Percentile 95%, se observa que el mejor ajuste se obtiene a partir de este último, según se observa en las figuras 17 y 18.



**Figuras 17 y 18.** En trazo azul y fucsia,  $W_{Audition}$  en función del  $\tau_e$  Percentile 99% y 95%. Con trazos punteados verde y rojo, los ajustes lineal y exponencial.

La ecuación que presenta el menor coeficiente de determinación ( $R^2 = 0.99$ ) es aquella cuya variable independiente es  $\tau_e$  Percentile 95%. La misma se muestra en la ecuación 4.

$$W_{Audition} [ms] = 6 \cdot e^{(0.373 \cdot \tau_e \text{ Percentile } 95\%)} \tag{4}$$

Siendo  $\tau_e$  Percentile 95% aquel valor de  $\tau_e$  que es superado el 95% del tiempo.

Según lo obtenido anteriormente, se recomienda utilizar inicialmente (para el cálculo del histograma de  $\tau_e$ ) los periodos de integración,  $2T$ , de la tabla 4, y así poder hallar luego la duración de la ventana de audición  $W_{Audition}$ .

**Tabla 4.** Duraciones  $2T$  recomendadas para iniciar el cálculo de la ACF

ACF	Recommended $2T$ [ms]
Vocal	40 to 80
Jazz	80 to 150
Orchestral	600 to 1000
Organ	1000 to 1500

## 4 Conclusiones

La duración de la ventana de integración del sistema auditivo varía en función de ciertas propiedades del estímulo sonoro.

Conocer la variabilidad del intervalo de integración del sistema auditivo y sus correspondientes valores, debería implicar modificar todo parámetro que haga uso del mismo, desde la forma de medir los niveles sonoros hasta el desarrollo de un nuevo conjunto de parámetros acústicos que incluyan esta variable, siendo entonces subjetivos y procurando sean ortogonales entre sí.

La duración obtenida como resultado del experimento con un “Vocal femenino anecoico” es muy similar al obtenido en Olive et al., 1988, de 53 ms.

Dada una pieza musical, al confirmar la existencia de varios  $\tau_{e,min}$  locales y consecuentemente realizar los cálculos con el valor estadístico  $\tau_e$  Percentile 95%, se observó en éstos mucha mayor consistencia e invariancia frente a las variaciones de período de integración 2T y del solapamiento en el cálculo del  $\tau_{e,min}$ .

Se obtuvo información que confirma el modelo de codificación de la información sonora por medio de patrones temporales de disparos neuronales.

Se observó que la ventana de integración,  $W_{Audition}$ , tiende a reducirse para mayores lateralizaciones. Ensayos para diferentes lateralizaciones y diferentes niveles sonoros deberán ser realizados en el futuro.

Es posible observar el grado de entrenamiento auditivo del sujeto bajo análisis, observando las varianzas obtenidas como resultado del proceso de evaluación de lo percibido.

Se logra observar que la respuesta de diferentes sujetos puede variar arrojando diferentes límites de detección, aunque las formas y pendientes de las curvas relevadas conservan la información buscada en el presente trabajo. Es por esto que lo importante es la forma de las curvas, con sus límites de detección relativos, no absolutos.

Es posible simplificar el protocolo, realizar una menor cantidad de ensayos, reducir considerablemente el tiempo del ensayo y hallar una duración de la ventana de integración “aceptable” realizando una interpolación lineal entre el primer punto de medición (delay = 20 ms) y el último punto de medición (delay = 500 ms), siempre utilizando un eje de abscisas (*Delay time*) en escala logarítmica según se observa en las figuras 14, 15 y 16.

Los oyentes sometidos a evaluación detectaron datos perceptuales o “cues” correspondientes al mismo retraso, en diferentes instantes de los archivos bajo análisis. Esto valida la idea de que el cerebro halla aquellos en más de un único momento, o sea, existen varios  $\tau_{e,min}$  locales donde el cerebro encuentra datos.

Los resultados obtenidos sí dependen de mantener constante el tipo de detección aplicado en el experimento, es decir emulando una situación real, para un mundo binaural, éste sería el de percepción de ensanchamiento de la imagen sonora. Aún no es posible afirmar si es adaptable a una fuente sonora que se desea escuchar inmersa en un entorno mixto y complejo de fuentes sonoras.

## 5 Trabajo a futuro

La duración de la ventana de integración hallada en el presente trabajo implica una situación del oyente de atención baja a media, es decir, dada una señal compleja, no se concentra en una fuente particular. Esta situación es similar a la de un transeúnte en una ciudad expuesto a ruido urbano.

Para el caso de atención alta, es decir, tratar de percibir una señal sonora en particular dentro de un entorno de ruido en forma continua, es posible que la ventana de audición se

altere a voluntad. Esto explicaría el efecto “*cocktail party*”, es decir la capacidad de percibir, reconocer señales particulares con relación señal a ruido negativas.

Además, para un estímulo sonoro continuo real, de gran duración (varios minutos, horas), donde el histograma de  $\tau_e$  varíe considerablemente, será necesario hallar también la velocidad de variación de la duración de la ventana de audición.

Es en estos sentidos continuará el programa de la investigación.

## 6 Agradecimientos

Al Dr. Yoichi Ando, por su colaboración.

A aquellos estudiantes de la carrera Ingeniería de Sonido de la Universidad Nacional de Tres de Febrero que participaron de los ensayos, por su colaboración y pasión por el conocimiento del sonido y sus fenómenos asociados.

## Referencias

- Ando, Yoichi. (1994). "Proceedings of the international conference on Acoustic Quality of Concert Halls. State of the art and research trends". Fundación Ramón Areces. Capítulo. 3, pags 63 - 71. Madrid, october 24 - 25.
- Ando, Yoichi. (1998). "Architectural Acoustics. Blending sound sources, sound fields, and listeners". Ch. 3, page 8 - 23. Modern Acoustics and Signal Processing, AIP Press.
- Ando, Yoichi. (1998b). "Architectural Acoustics. Blending sound sources, sound fields, and listeners". Modern Acoustics and Signal Processing, AIP Press.
- Ando, Yoichi. (2002). "Correlation factors describing primary spatial sensations of sound fields". Journal of Sound and Vibration. 258(3), 405-417, doi:10.1006/jsvi.5264.
- Ando, Yoichi. (2008). Theory of Auditory Temporal and Spatial Primary Sensations", Journal of Temporal Design in Architecture and the Environment, Japan.
- Ando, Yoichi. (2009). Recomendaciones propias del Dr. Ando asentadas en varias comunicaciones personales con él intercambiadas durante el año 2009.
- Cariani, P. A.; Delgutte, B. (1996). Journal of Neurophysiology 76, 1698-1716. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience.
- Cariani, Peter. (1995). "As if time really mattered: Temporal strategies for neural coding of sensory information". Communication and Cognition-Artificial Intelligence (CC-AI), March, 1995, Vol 12, nos. 1-2, pp. 157-219. Special Issue on Self-Reference in Biological and Cognitive Systems, L. Rocha (ed).
- Cariani, Peter. (2001). "Temporal codes, timing nets, and music perception". Journal of New Music Perception, vol. 30, issue 2. Special issue on Rhythm Perception, Periodicity, and Timing Nets.
- Cariani, Peter; Tramo, Mark. (1997). "Neural representation of pitch through temporal autocorrelation" Proceedings, Audio Engineering Society Meeting (AES), New York, September, 1997, Preprint #4583 (L-3).
- Fraisse, P. (1967). Psychologie du temps. París. PUF.
- Kato, Kosuke; Fujii, Kenji; Kawai, Keiji; Ando, Yoichi; Yano, Takashi. (2004). "Blending vocal music with the sound field -the effective duration of autocorrelation function of Western professional singing voices with different vowels and pitches". Proceedings of the International Symposium on Musical Acoustics, March 31st to April 3rd, (ISMA2004), Nara, Japan.
- Kato, Kosuke; Noson, Dennis; Ando, Yoichi. (2002). "Blending Vocal Music With Sound Fields by Variations in Lyrics and Music Tempo". S.E.A.
- Mouri, K., Akiyama, K., Ando, Y. (2001). "Preliminary study on recommended time duration of source signals to be analyzed, in relation to its effective duration of the autocorrelation function". Journal of Sound and Vibration, 241, pages 87 - 95.

- Olive, S. E.; Toole , F. E. (1988). "The Detection of Reflections in Typical Rooms," presented at the 85th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol. 36, p. 1029, preprint 2719.
- Rossing, Thomas. (2007). "Handbook of acoustics", Chapter 10, page 360. Springer.
- Secker-Walker, H. E.; Searle, C. L. (1990). Journal of the Acoustical Society of America 88, 1427-1436. Time domain analysis of auditory-nerve-fiber firing rates.
- Tohyama, M, Suzuki, H & Ando, Y. (1995). "The Nature and Tecnology of Acoustic Space". Academic Press. Capítulo 5.