UMBRALES DE DENSIDAD DE PULSOS PARA VELVET NOISE USADO EN REVERBERACIÓN ARTIFICIAL: UN ESTUDIO DE ESCUCHA EN LÍNEA

Juan Almaraz[1], Shin-ichi Sato[1], Mariano Girola[1]

[1]Ingeniería de Sonido, Universidad Nacional de Tres de Febrero

juan.almaraz097@gmail.com

ssato@untref.edu.ar

mgirola@untref.edu.ar

Resumen - El velvet noise es un tipo de ruido disperso compuesto por impulsos unitarios distribuidos aleatoriamente con densidad controlada, lo que produce una señal perceptualmente suave. Este estudio investiga la densidad mínima de impulsos a la que el velvet noise es perceptualmente indistinguible del ruido blanco gaussiano cuando se aplica como reverberación sintética a fragmentos musicales. Se realizaron pruebas de escucha ABX en línea con 28 participantes, utilizando tres fragmentos musicales: percusión, voz y flauta. Los umbrales perceptivos variaron ampliamente en función del fragmento, desde 338,1 pulsos/s para la percusión hasta tan solo 1,3 pulsos/s para la flauta. Estas variaciones mostraron una relación inversa con la duración mínima efectiva $(\tau e)_{min}$ de la función de autocorrelación de cada fragmento, lo que sugiere que las señales con mayor persistencia temporal requieren un ruido menos denso para la equivalencia perceptiva. Aunque se identificaron errores sistemáticos, como clics transitorios, se consideró que no habían sesgado significativamente los valores umbral. Los resultados ponen de relieve el potencial del uso de velvet noise de menor densidad en la síntesis de reverberación, lo que permite una mayor eficiencia computacional adaptada al contenido musical.

Palabras clave - Velvet noise, Reverberación Artificial, Test ABX

1. INTRODUCCIÓN

Grabar instrumentos musicales en espacios acústicamente ideales, como teatros o salas de conciertos, suele ser poco práctico debido a limitaciones logísticas y económicas. Como resultado, la reverberación artificial se ha estudiado ampliamente desde el trabajo seminal de Schroeder, *Natural Sounding Artificial Reverberation* [1]. Existen varios algoritmos que tienen como objetivo replicar la forma en que el sonido interactúa con las superficies de la sala, imprimiendo características temporales y espaciales en la percepción del oyente. Valimaki et al. clasificaron estos algoritmos en tres tipos principales: redes de retardo, que manipulan la señal de entrada mediante retardos, filtrado y retroalimentación; métodos convolucionales, que aplican respuestas al impulso de espacios reales o estimados; y modelos acústicos computacionales, que simulan la propagación del sonido dentro de una geometría virtual [2]. El presente estudio se centra en la reverberación convolucional, en particular utilizando respuestas al impulso de salas (*Room Impulse Response* o RIR) sintetizadas basadas en secuencias de ruido.

Moorer sugirió por primera vez que la parte final de una RIR puede caracterizarse eficazmente como ruido blanco con un decaimiento exponencial [3]. Basándose en esta idea, Rubak y Johansen introdujeron el primer algoritmo de reverberación basado en ruido disperso, utilizando una secuencia pseudoaleatoria conocida como ruido totalmente aleatorio (*Totally Random Noise* o TRN) [4]. En el TRN, cada muestra tiene la misma probabilidad de ser cero o no cero, lo que da lugar a grupos impredecibles de impulsos y largos intervalos de silencio.

Un avance significativo en las secuencias de ruido pseudoaleatorio se produjo con la introducción del *velvet noise* por parte de Järveläinen y Karjalainen [5]. El *velvet noise*, llamado así por su calidad perceptivamente más suave en comparación con otras secuencias de ruido, se genera en tres pasos: primero, se crea una serie de impulsos espaciados uniformemente con una densidad de pulsos específica; segundo, se desplaza aleatoriamente la posición de cada impulso; y tercero, se asigna a cada impulso un valor aleatorio positivo o negativo. El resultado es una señal de ruido aleatorio ternario que consiste principalmente en ceros, con algunas muestras en -1 o 1, y un espectro de potencia plano similar al ruido blanco. Es importante destacar que el *velvet noise* conserva su suavidad característica a densidades de pulso más bajas en comparación con otras secuencias de ruido dispersas [6]. Basándose en el desarrollo del *velvet noise*, Välimäki et al. introdujeron secuencias de *velvet noise* intercaladas para una síntesis eficiente de reverberación tardía [7]. El enfoque consiste en dividir el proceso de reverberación en múltiples secuencias de

velvet noise desplazadas en el tiempo que mantienen las propiedades estocásticas de la reverberación tardía y evitan artefactos periódicos no deseados. Al optimizar la densidad de pulsos y el intercalado de secuencias, el método garantiza una atenuación suave similar al ruido blanco gaussiano de atenuación exponencial.

Más allá de la reverberación, el *velvet noise* se ha aplicado en diversas tareas de procesamiento de señales de audio. Se ha utilizado para la decorrelación de señales de audio, una técnica esencial en la reproducción espacial del sonido para sistemas multicanal [8,9]. En la síntesis del habla, el *velvet noise* ha servido como señal de excitación en la generación del habla basada en *vocoders* [10]. Además, ha encontrado aplicaciones en mediciones acústicas, donde se ha empleado como señal de prueba [11].

La densidad de pulsos es un factor crítico en la percepción del *velvet noise*. Karjalainen y Järveläinen descubrieron que se necesitan al menos 1500 pulsos/s para que el *velvet noise* se perciba tan suave como el ruido blanco gaussiano [5]. Su estudio comparó directamente el *velvet noise* y el gaussiano utilizando una escala de suavidad de 7 puntos. Su prueba de audición se llevó a cabo en condiciones idénticas para todos los participantes en un entorno controlado y contó con la participación de nueve participantes entrenados que evaluaron señales de ruido filtradas y de banda ancha de 2 segundos (paso bajo y paso banda) con densidades de pulso que oscilaban entre 600 y 20.000 pulsos/s. Más recientemente, Roberts et al. examinaron la rugosidad dependiente de la frecuencia del *velvet noise*, demostrando que la densidad de pulso requerida disminuye al aplicar filtros pasa-bajos o de octava [12]. Se pidió a los participantes que encontraran la densidad de pulsos más baja en la que las señales de prueba coincidieran con la suavidad perceptiva de una referencia de *velvet noise* de banda ancha de 2000 impulsos/s. El *velvet noise* filtrado a 125 Hz se calificó como suave con densidades medias tan bajas como 401 impulsos/s (octava) y 725 impulsos/s (pasa-bajos), con umbrales que aumentaban junto con la frecuencia del filtro. Aunque estos estudios evaluaron directamente el *velvet noise*, no abordaron cómo las características de la señal de entrada influyen en la densidad de pulsos necesaria para simular la reverberación. Välimäki et al. examinaron la percepción del *velvet noise* en el contexto de los estímulos musicales [13]. Simularon la reverberación tardía en las RIR utilizando *velvet noise* filtrado con una densidad de pulsos variable en el tiempo, comenzando con 100 pulsos/s y disminuyendo con el tiempo. Una prueba ABX con una escala de 5 puntos comparó las RIR sintetizadas y originales utilizando fragmentos musicales de percusión, acordes y voces. Los sonidos de percusión con transitorios

prominentes fueron los más fáciles de distinguir, mientras que los sonidos de acordes con ataques lentos, más parecidos a señales en estado estacionario, fueron los más difíciles de diferenciar.

Las investigaciones sobre preferencias en las condiciones de reverberación sugieren una fuerte relación entre la naturaleza de la señal fuente y su función de autocorrelación (ACF). Ando y colegas descubrieron que las preferencias de los oyentes por campos sonoros con múltiples reflexiones y reverberación dependen de la ACF a largo plazo de la señal fuente, especialmente en un intervalo de integración de 2T = 33-35 s [14-16]. La duración efectiva de la ACF (τe) se define como el retraso en el que la envolvente de la ACF normalizada cae por debajo de 0,1, lo que sirve como métrica clave para analizar las propiedades temporales de una señal.

Reconociendo que las señales musicales suelen presentar variaciones temporales significativas, Ando et al. introdujeron posteriormente el análisis ACF continuo (*running Autocorrelation Function* o r-ACF) con 2T alrededor de 2 s, que proporciona una representación más dinámica de la ACF de una señal a lo largo del tiempo [17]. Partiendo de esta base, Kuroki et al. realizaron pruebas de preferencia utilizando diez piezas musicales diferentes y descubrieron que el tiempo de reverberación más preferido podía predecirse basándose en la duración mínima efectiva $(\tau e)_{min}$ de cada pieza [18].

Dada esta fuerte relación entre la estructura temporal de la señal y la percepción de la reverberación, este estudio tiene como objetivo investigar más en profundidad la densidad mínima de impulsos necesaria para que el *velvet noise* se perciba como subjetivamente equivalente al ruido blanco utilizado en la generación de reverberación artificial. Dado que este umbral depende en gran medida de las características de los estímulos, es necesario adoptar un enfoque más sistemático. Para abordar esta cuestión, proponemos evaluar los estímulos en función de su $(\tau e)_{min}$, como medio para comprender mejor la relación entre las propiedades del estímulo y la suavidad percibida.

2. METODOLOGÍA

2.1 Señales utilizadas

Las señales de audio seleccionadas son fragmentos de una percusión sintetizada, una voz femenina cantando y una flauta (Figura 1). De ahora en más estos estímulos se denominarán percusión, voz y flauta, respectivamente. Cada fragmento tiene una duración de entre 5 y 6 segundos, con una frecuencia de muestreo de 48 kHz y una profundidad de bits de 16 bits. El espectrograma de la

percusión muestra las características impulsivas, mientras que la voz y la flauta muestran los componentes tonales.

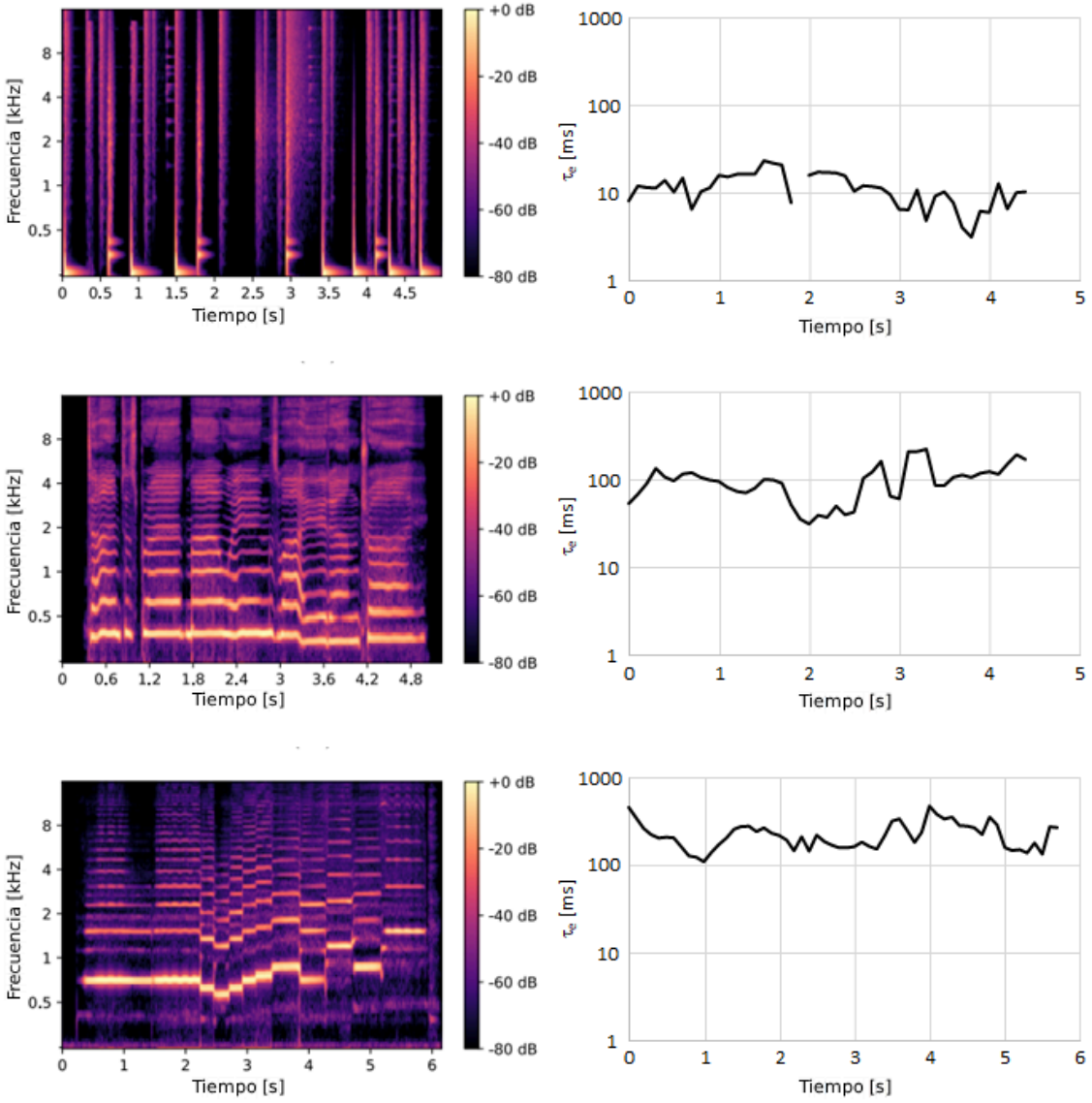


Figura 1: Espectrograma Mel (izquierda) y τ_e de la r-ACF (derecha) para las señales fuente. Percusión (arriba), voz (en medio) y flauta (abajo). El valor perdido de τe para la percusión en 1,9 s se debe a que la ACF es demasiado pequeña para producir un valor τ_e positivo.

El τ_e del ACF para cada estímulo se calculó utilizando un análisis ACF continuo con código MATLAB [19], basado en el método iterativo propuesto por Sato y Wu [20]. El análisis se realizó con un intervalo de integración de 1 s y un paso continuo de 0,1 s. Como también se muestra en la Figura 1, τ_e para los estímulos de percusión y voz fluctúa alrededor de 10 ms y 100 ms, respectivamente, mientras que τ_e para la flauta supera constantemente los 100 ms. La duración

mínima efectiva $(\tau_e)_i$ es de 4 ms para la percusión, 42 ms para la voz y 108 ms para la flauta. Esta tendencia también se refleja en otros valores percentiles, no solo en los mínimos (Tabla 1).

Tabla 1: Valores percentiles τe de las señales utilizadas.

	Valor percentil de τe [ms]				
	$(\tau e)_{min}$	10% τe	50% τe	90% τe	$(\tau e)_{max}$
Percusión	3,1	6,1	10,8	16,7	22,8
Voz	31,2	40,9	100,4	180,7	223,5
Flauta	108,5	141,4	203,9	329,8	458,1

2.2 Generación del *velvet noise*

Se utilizó el siguiente procedimiento para generar *velvet noise*. En primer lugar, se crearon pulsos espaciados uniformemente con amplitud unitaria, basándose en una densidad de pulsos especificada. A continuación, cada pulso se desplazó aleatoriamente dentro de un ±50 % del espaciado medio, utilizando una distribución uniforme. Esto evita el solapamiento al mismo tiempo que introduce una aleatoriedad controlada. Por último, a cada pulso se le asignó aleatoriamente un signo de -1 o +1.

Para determinar el umbral perceptivo, se generó *velvet noise* con densidades de pulsos que aumentaban sistemáticamente. Para el estímulo de percusión, de 50 a 350 pulsos/s en incrementos de 50 pulsos/s; para el estímulo vocal, la densidad de pulsos osciló entre 10 y 70 pulsos/s en pasos de 10 pulsos/s; y para el estímulo de flauta, de 1 a 13 pulsos/s en pasos de 2 pulsos/s. Estos rangos se seleccionaron basándose en las diferencias de umbral esperadas entre los estímulos.

2.2 Respuesta al impulso sintética usando *velvet noise*

Se utilizó el siguiente procedimiento para generar RIRs artificiales utilizando el *velvet noise* descrito anteriormente. El sonido directo se introdujo como un impulso en la primera posición de muestreo, seguido de un retardo de 9 ms antes del inicio de la cola de reverberación. Para garantizar una relación directo-reverberado (DRR) de 10 dB, se ajustó la magnitud del impulso de sonido directo en consecuencia. Estos parámetros se basaron en un estudio subjetivo realizado por Rubak y Johansen (1998). La Figura 2 presenta un diagrama de bloques de este proceso.

También se creó un estímulo de referencia utilizando el mismo procedimiento, pero sustituyendo el *velvet noise* por ruido blanco. Los estímulos de prueba se generaron convolucionando cada extracto con una RIR sintética con un tiempo de reverberación de 1 segundo. Todos los estímulos fueron monofónicos.

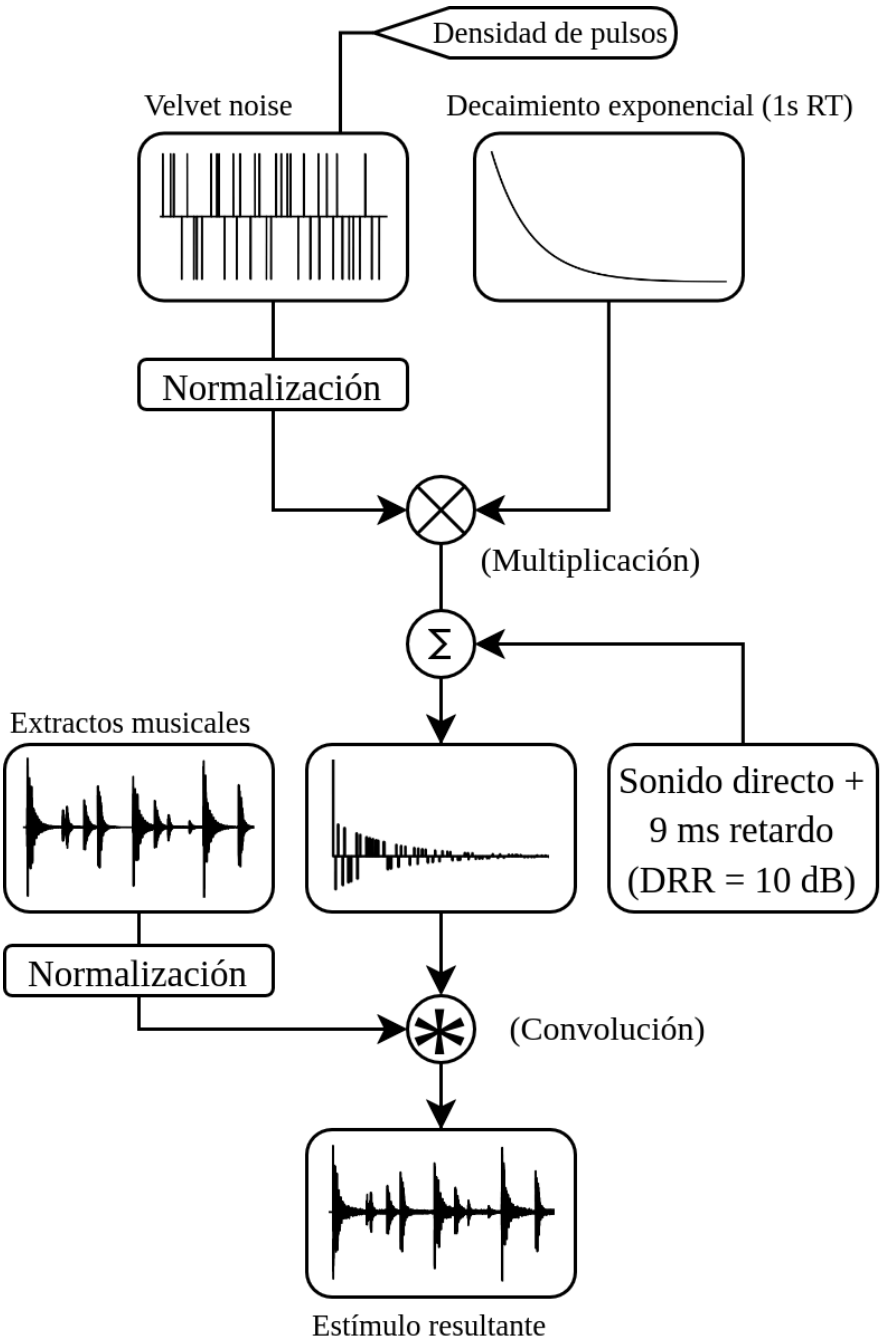


Figura 2: Diagrama de bloques del proceso de generación de las señales estímulo.

2.2 Test subjetivo (ABX forzado)

Para cada fragmento musical y densidad de pulsos, se realizó una prueba ABX en línea con el fin de maximizar la accesibilidad y recopilar de manera eficiente las respuestas subjetivas. Se utilizó la plataforma ABX – Listening Tests as a Service (s.f.). La señal de referencia fue el fragmento musical convolucionado con reverberación de ruido blanco gaussiano. En cada prueba, se presentó a los oyentes tres estímulos: A y B (los estímulos de referencia y de prueba) y X (idéntico a A o B). Los participantes podían cambiar libremente entre los tres y se les pidió que identificaran si X coincidía con A o B. Se les animó a limitar la reproducción a no más de tres veces por estímulo antes de responder. Si no se apreciaba ninguna diferencia perceptiva, se les indicó que eligieran al azar.

Cada tipo de estímulo (percusión, voz o flauta) se probó con siete valores diferentes de densidad de pulsos, lo que dio como resultado siete comparaciones ABX por fragmento. Para reducir la fatiga del oyente y mantener la fiabilidad de la prueba, a cada participante se le asignaron aleatoriamente solo dos de los tres fragmentos. El orden de presentación también fue aleatorio.

Antes de comenzar la prueba, se pidió a los participantes que indicaran su edad, su experiencia auditiva relacionada con la música y el tipo de dispositivo de reproducción utilizado (auriculares intrauditivos, auriculares supraaurales o altavoces). Aunque se recomendó el uso de auriculares, las respuestas de los participantes que utilizaron altavoces se excluyeron del análisis.

Se incluyó una fase de entrenamiento para familiarizar a los oyentes con el procedimiento de la prueba ABX. En esta fase, se presentó un fragmento de percusión convolucionado con ruido blanco gaussiano (estímulo A) y *velvet noise* a 20 pulsos por segundo (estímulo B). Se indicó a los oyentes que ajustaran el volumen de reproducción a un nivel cómodo durante el entrenamiento y que mantuvieran ese volumen durante toda la prueba.

En total, cada participante completó 15 pruebas: una prueba de entrenamiento y 14 pruebas experimentales (siete para cada uno de los dos fragmentos). La duración de la prueba fue de aproximadamente 6 minutos.

2.2 Participantes

La participación fue voluntaria, y solo las personas que deseaban participar completaron la prueba. Un total de 28 participantes completaron la prueba subjetiva. De ellos, 18 respuestas correspondieron al estímulo de la percusión, 19 a los estímulos de voz y 19 a los de flauta. Los participantes tenían entre 17 y 59 años, con una edad media de 30 años y una desviación estándar de 10 años.

En cuanto al equipo de reproducción, el 68 % de los participantes declaró utilizar auriculares *over-ear*, mientras que el 32 % utilizaba auriculares *in-ear*. Ningún participante declaró utilizar altavoces, por lo que no se excluyó ninguna respuesta en función del dispositivo de reproducción.

En cuanto a la experiencia auditiva, el 39 % de los participantes indicaron tener cierto nivel de experiencia auditiva, el 32 % se identificaron como músicos o productores musicales, el 18 % indicaron tener un hobby relacionado con la música y el 10 % seleccionaron «ninguna de las anteriores».

3. RESULTADOS

El número de respuestas correctas en la prueba ABX se modeló como una variable aleatoria que sigue una distribución binomial. A cada respuesta correcta se le asignó un valor de 1, mientras que a las respuestas incorrectas se les asignó un valor de 0. Para determinar si el porcentaje de respuestas correctas indicaba una diferencia audible estadísticamente significativa con un nivel de confianza del 95 %, se calculó el porcentaje mínimo requerido para N ensayos utilizando la Ecuación (3). El porcentaje mínimo de respuestas correctas necesario para alcanzar un nivel de confianza del 95 % es del 73,6 % para N = 18 (percusión) y del 72,9 % para N = 19 (voz y flauta).

$$95\%\ \text{nivel de confianza} = N/2 + \sqrt{N} \tag{3}$$

La Figura 3 muestra el porcentaje de respuestas correctas en función de la densidad de pulsos para cada fragmento musical. Dado que la tendencia general era similar en todos los fragmentos, se aplicó una regresión lineal a cada conjunto de datos. A continuación, se calculó el error cuadrático medio (*Mean Square Error* o MSE) del ajuste lineal para cada fragmento, que se presenta en la Tabla 2. Los valores de MSE aumentaron en el orden siguiente: percusión, voz y flauta.

El umbral de densidad de pulsos para cada fragmento se estimó identificando el punto de intersección entre la línea de regresión y el nivel de confianza del 95 %. Para el fragmento de flauta, esta intersección indicaba una densidad de pulsos negativa, lo que no tiene sentido desde el punto de vista físico. Para solucionar esto, se excluyó el punto de datos de 5 pulsos/s, identificado como la mayor desviación de la línea de regresión, y se volvió a estimar el umbral. Los umbrales resultantes fueron 338,1 pulsos/s para la percusión, 31,0 pulsos/s para la voz y 1,3 pulsos/s para la flauta.

Además, los umbrales se calcularon utilizando un criterio de respuesta correcta del 75 %, que también se utiliza habitualmente en la estimación del umbral perceptivo. Estos valores también se recogen en la tabla 2. Independientemente del tratamiento del valor atípico en los datos de la flauta, la mayoría de los oyentes no percibieron ninguna diferencia clara entre la reverberación generada con *velvet noise* y la generada con ruido gaussiano, incluso con densidades de pulso muy bajas, especialmente en comparación con los otros fragmentos.

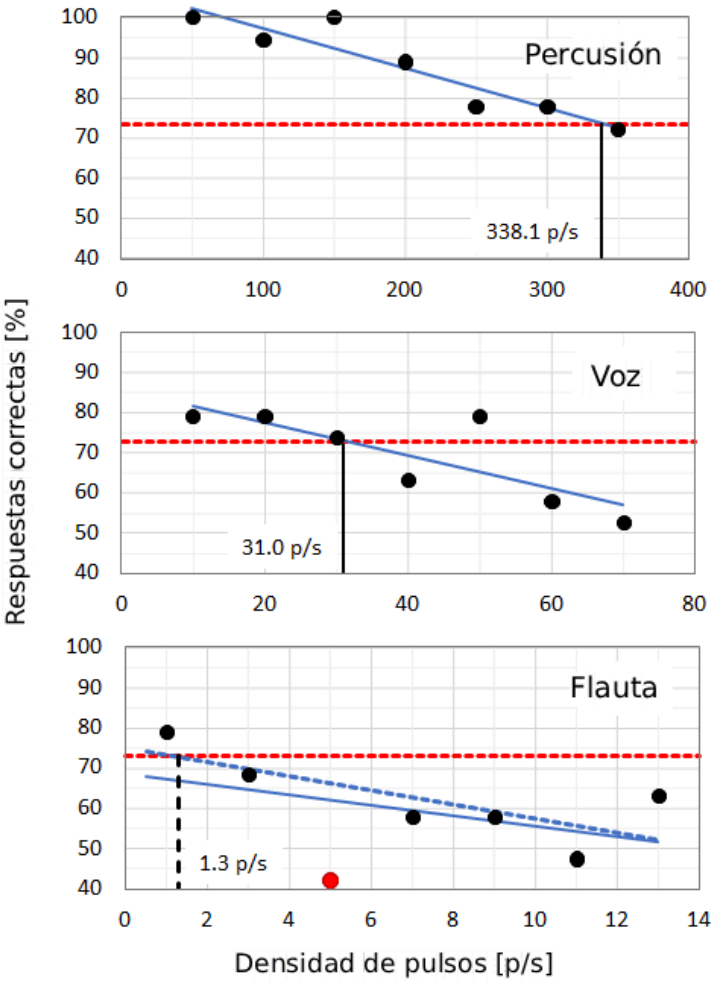


Figura 3. Porcentaje de respuestas correctas en función de la densidad de pulsos. El umbral de densidad de pulsos se indica mediante el punto de intersección entre la línea de regresión (azul) y el nivel de confianza del 95 % (rojo punteado). Para el estímulo de flauta,

se muestra una línea de regresión adicional (azul punteado), calculada tras excluir la muestra de 5 pulsos/s (punto rojo) como valor atípico.

Tabla 2: Estimaciones del umbral a partir del análisis de regresión que se muestra en la Figura 3. También se incluye el error cuadrático medio (MSE) para cada regresión. La flauta 2 se refiere a los resultados tras excluir el valor atípico de 5 pulsos/s.

	Umbral con 95% nivel de confianza (pulse/s)	Umbral en 75% de respuestas correctas (pulse/s)	MSE de la regresión
Drum	338.1	324.0	13.7
Vocal	31.0	25.9	37.7
Flute	–	–	104.5
Flute 2	1.3	0.1	

Para evaluar en profundidad la significación estadística de las diferencias observadas, se realizó una prueba χ^2. Esta prueba ofrece un enfoque alternativo para evaluar si la proporción de respuestas correctas supera significativamente el nivel esperado por casualidad con un nivel de confianza del 95 %. Los resultados para cada densidad de pulsos y condición de fragmento musical se presentan en la Tabla 3. Para el estímulo de percusión, todas las respuestas en densidades de pulsos de hasta 350 pulsos/s indicaron una diferencia audible. En el caso del estímulo vocal, las densidades de pulsos de hasta 30 pulsos/s y 50 pulsos/s mostraron una discriminación significativa. En el caso del estímulo de flauta, solo la condición de 1 pulso/s arrojó una diferencia estadísticamente significativa. A pesar de los resultados irregulares a 50 pulsos/s para el estímulo vocal, estos hallazgos coinciden en general con el análisis gráfico que se muestra en la Figura 4, donde los puntos negros por encima de la línea discontinua roja denotan una discriminación perceptiva estadísticamente significativa.

Table 3. Estadísticas χ^2 y valores p para cada densidad de pulsos y fragmento musical. $p < 0,05$ se indica con *, y $p < 0,01$ con **.

Percusión		Voz		Flauta	
Densidad de pulsos (pulsos/s)	χ^2	Densidad de pulsos (pulsos/s)	χ^2	Densidad de pulsos (pulsos/s)	χ^2
50	0.000**	10	0.012*	1	0.012*
100	0.000**	20	0.012*	3	0.108
150	0.000**	30	0.039*	5	0.491
200	0.001**	40	0.251	7	0.491
250	0.018*	50	0.012*	9	0.491
300	0.018*	60	0.491	11	0.819

350	0.059*	70	0.819	13	0.251

4. DISCUSIÓN

En cada prueba, el rango de densidades de pulsos se diseñó con éxito para determinar los umbrales perceptivos, aunque fue necesario tratar los valores atípicos en el fragmento de flauta. Los umbrales de densidad de pulsos observados en este estudio fueron significativamente más bajos que los descritos en investigaciones anteriores que evaluaban directamente las características perceptivas del *velvet noise* en sí mismo (es decir, sin su aplicación en la simulación de reverberación). Esta diferencia puede atribuirse en parte a la configuración experimental: el presente estudio se llevó a cabo en condiciones de escucha en línea, mientras que los estudios anteriores se realizaron en entornos controlados con equipos e instalaciones uniformes. Las condiciones de escucha en línea pueden reducir la densidad de pulsos necesaria para simular la reverberación convolucionada con señales musicales.

Otro hallazgo clave fue que la densidad de pulsos umbral se correlaciona inversamente con la duración efectiva (τ_e) de la ACF de los fragmentos musicales. Los fragmentos con valores τ_e más largos requieren densidades de pulsos significativamente más bajas para ser distinguidos perceptualmente. Esto concuerda con los hallazgos de Välimäki et al. (2017), quienes demostraron que, con una densidad de pulsos fija, los sonidos de percusión con transitorios prominentes eran los más fáciles de distinguir, mientras que los sonidos de acordes con ataques lentos, similares a señales en estado estacionario, eran los más difíciles de diferenciar.

Esta relación entre el umbral perceptivo y la ACF de la señal fuente amplía los hallazgos de estudios anteriores. Karjalainen y Järveläinen (2007) y Roberts et al. (2023) evaluaron directamente la percepción del *velvet noise*, que puede interpretarse como impulsos de Dirac filtrados o sin filtrar convolucionados con *velvet noise*. A efectos comparativos, se analizaron las ACF de estos impulsos de Dirac filtrados/sin filtrar y se estimaron los valores τ_e. Como se muestra en la Figura 4, se observó una tendencia inversa similar en todos los estudios, a pesar de las diferencias en la metodología y las condiciones de las pruebas. La ACF refleja el grado de persistencia de una señal; un τ_e más largo sugiere una mayor autosimilaridad temporal, lo que puede enmascarar los efectos de la reverberación, reduciendo así la sensibilidad a las diferencias perceptivas.

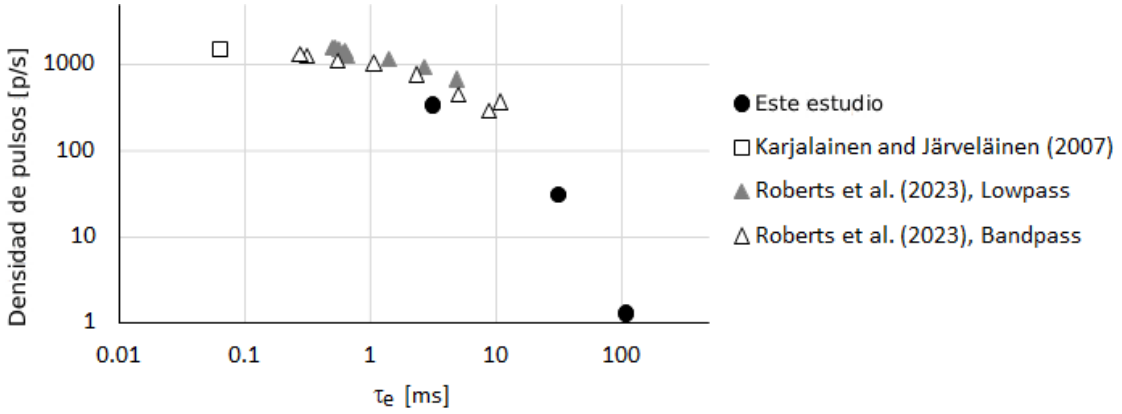


Figura 4. Relación entre τ_e de las señales fuente utilizadas para las pruebas y la densidad del pulso umbral.

Curiosamente, el umbral para el fragmento de flauta fue de casi 0 pulsos/s. Este resultado se debe probablemente a las características del propio estímulo, que tenía un $(\tau_e)_{min}$ de 108 ms. Estudios previos han demostrado que las señales musicales con $(\tau_e)_{min}$ superior a 100 ms suelen consistir en instrumentos solistas que tocan acordes largos o pasajes legato (Kuroki et al., 2004; Sato et al., 2011). Para estos fragmentos, el *velvet noise* se puede sintetizar con densidades de pulso extremadamente bajas, lo que ofrece una oportunidad prometedora para la reducción o compresión eficiente de datos.

Como se muestra en la Figura 3 y la Tabla 2, el modelo de regresión utilizado para estimar los umbrales presentó errores cuadrados medios (MSE) relativamente elevados. La prueba ABX utilizó un método de elección forzada, que teóricamente converge a una precisión del 50 % cuando los participantes no pueden distinguir entre las condiciones. Sin embargo, el número limitado de ensayos puede haber impedido la convergencia ideal, aumentando el MSE. Además, es probable que un error sistemático en el método de prueba haya contribuido al alto MSE. Más tarde se descubrió que, durante el cambio rápido entre ciertas pistas de audio, ocasionalmente se podía escuchar un clic, un *artifact* que no estaba presente al comparar pistas idénticas. Esto puede haber llevado a los participantes a identificar la opción correcta basándose en el clic, en lugar de en su juicio perceptivo real. Por lo tanto, aunque los umbrales obtenidos se consideran válidos, podrían ser ligeramente inferiores si la prueba se repitiera sin este artefacto. Aunque posteriormente se identificó un error sistemático, se cree que tuvo un efecto mínimo en las estimaciones del umbral. No obstante, la cuestión pone de relieve la importancia de la metodología de prueba en los experimentos perceptivos y debe abordarse en futuras réplicas.

5. CONCLUSIONES

Esta investigación proporciona pruebas de que la densidad de pulsos necesaria para que el *velvet noise* se perciba como subjetivamente equivalente al ruido blanco gaussiano disminuye significativamente en el caso de las condiciones de escucha en línea, dependiendo de las características del fragmento musical al que se aplica la reverberación sintética. Concretamente, las señales musicales con mayor autosimilaridad temporal, reflejada en duraciones efectivas más largas (τ_e) en su ACF, se asociaron con umbrales perceptivos más bajos para la densidad de pulsos. Esta relación concuerda con hallazgos previos que indican que los sonidos donde predominan los transitorios requieren un *velvet noise* más denso para preservar la similitud perceptiva, mientras que las señales más estacionarias o legato toleran distribuciones de pulsos mucho más dispersas.

Estos hallazgos sugieren que adaptar la densidad de pulsos del *velvet noise* según las características temporales de la señal de entrada podría permitir algoritmos de reverberación artificial más eficientes, lo que podría reducir el coste computacional o el tamaño de los datos sin una pérdida perceptible de calidad.

Este estudio sirve como punto de partida para seguir explorando la relación entre las características temporales basadas en la autocorrelación y los requisitos perceptivos en la síntesis de reverberación. En futuros trabajos se debería ampliar la gama de fragmentos musicales, perfeccionar el protocolo de pruebas ABX para eliminar artefactos y explorar modelos más precisos que relacionen τ_e y los umbrales de densidad de pulsos.

REFERENCIAS

[1] Schroeder, M. R. (1961). Natural sounding artificial reverberation. Audio Engineering Society Convention 13.

[2] Valimaki, V., Parker, J. D., Savioja, L., Smith, J. O., & Abel, J. S. (2012). Fifty years of artificial reverberation. IEEE Transactions on Audio, Speech, and Language Processing, 20(5), 1421–1448. doi: 10.1109/TASL.2012.2189567

[3] Moorer, J. A. (1979). About this reverberation business. Computer Music Journal, 13–28. doi: 10.2307/3680280

[4] Rubak, P., & Johansen, L. G. (1998). Artificial reverberation based on a pseudo-random impulse response: Part I. Audio Engineering Society Convention 104.

[5] Järveläinen, H., & Karjalainen, M. (2007). Reverberation modeling using velvet noise. Audio Engineering Society 30th International Conference on Intelligent Audio Environments.

[6] Välimäki, V., Lehtonen, H.-M., & Takanen, M. (2013). A perceptual study on velvet noise and its variants at different pulse densities. IEEE Transactions on Audio, Speech, and Language Processing, 21(7), 1481–1488. doi: 10.1109/TASL.2013. 2255281

[7] Välimäki, V., & Prawda, K. (2021). Late-reverberation synthesis using interleaved velvet-noise sequences. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 1149–1160. doi: 10.1109/TASLP.2021.3060165

[8] Alary, B., Politis, A., & Välimäki, V. (2017). Velvet-noise decorrelator. Proceedings of the 20th International Conference on Digital Audio Effects, 405–411.

[9] Schlecht, S. J., Alary, B., Välimäki, V., Habets, E. A., et al. (2018). Optimized velvet-noise decorrelator. Proceedings of International Conference on Digital Audio Effects (DAFx-18), 87–94.

[10] Kawahara, H., Sakakibara, K. I., Morise, M., Banno, H., Toda, T., & Irino, T. (2018). Frequency domain variants of velvet noise and their application to speech processing and synthesis. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2018, 2027–2031.

[11] Kawahara, H., Sakakibara, K.-I., Mizumachi, M., Banno, H., Morise, M., & Irino, T. (2019). Frequency domain variant of velvet noise and its application to acoustic measurements. 2019 AsiaPacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 1523–1532.

[12] Roberts, J., Fagerström, J., Schlecht, S., & Välimäki, V. (2023). How smooth do you think I am: An analysis on the frequency-dependent temporal roughness of velvet noise. International Conference on Digital Audio Effects, 312–318.

[13] Välimäki, V., Holm-Rasmussen, B., Alary, B., & Lehtonen, H.-M. (2017). Late reverberation synthesis using filtered velvet noise. Applied Sciences, 7(5), 483. https://doi.org/10.3390/app7050483

[13] Ando, Y., & Gottlob, D. P. (1979). Effects of early multiple reflections on subjective preference judgments of music sound fields. Journal of the Acoustical Society of America, 65(2), 524–527. https://doi.org/10.1121/1.382308

[14] Ando, Y., & Imamura, M. (1979). Subjective preference tests for sound fields in concert halls simulated by the aid of a computer. Journal of Sound and Vibration, 65(2), 229–239. https://doi.org/10.1016/0022-460X(79)90516-9

[15] Ando, Y. (1983). Calculation of subjective preference at each seat in a concert hall. Journal of the Acoustical Society of America, 74(3), 873–887. https://doi.org/10.1121/1.389874

[16] Ando, Y., Noson, D., & Sakai, H. (1989). The running autocorrelation function of different music signals relating to preferred temporal parameters of sound fields. Journal of the Acoustical Society of America, 86(2), 644–649. https://doi.org/10.1121/1.398196

[17] Kuroki, S., Sato, H., & Ando, Y. (2004). Individual preference in relation to the temporal and spatial factors of the sound field: Factors affecting individual differences in subjective preference judgments. Journal of Temporal Design in Architecture and the Environment, 4(1), 29-40.

[18] Sato, H., & Wu, C. J. (2010). Definition of the effective duration (τe) of the running autocorrelation function of music signals. Acta Acustica united with Acustica, 97(3), 432–440.

[19] Sato, S. (2014). MATLAB program for calculating the parameters of the autocorrelation and interaural cross-correlation functions based on Ando's auditory-brain model. Audio Engineering Society Convention 137. http://www.aes.org/e-lib/browse.cfm?elib=17491

[20] ABX app - Listening Tests as a Service. (n.d.). Retrieved June 20, 2024, from https://abxtests.com/