

PRESTACIONES DE UN SISTEMA DE REPRODUCCIÓN WAVE FIELD SYNTHESIS (WFS) UTILIZANDO UNIDADES DE PROCESAMIENTO GRÁFICO.

PACS: 43.60.Gk

José A. Belloch^{1,2}; Laura Fuster²; Enrique S. Quintana¹; Antonio M. Vidal³; Alberto González²

1 Dpt of Computer Science and Engineering (ICC), Universidad Jaime I de Castellón, Av. de Vicent Sos Baynat, s/n 12071 Castellón de la Plana, España. jbelloch@uji.es, quintana@uji.es.

2 Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, camino de vera s/n 46022 Valencia, España, jobelrod@iteam.upv.es, lfuster@iteam.upv.es, agonzal@com.upv.es

3 Dpt de Sistemas Informáticos y Computación, Universitat Politècnica de València, camino de vera s/n 46022 Valencia, España, avidal@dsic.upv.es

ABSTRACT

Wave Field Synthesis (WFS) is a spatial audio reproduction system that provides an accurate spatial sound field in a wide area. This sound field is rendered through a high number of loudspeakers to emulate virtual sound sources. WFS systems require high computational capacity since they involve multiple loudspeakers and multiple virtual sources. Furthermore improvements of the spatial audio perception imply even higher processing capacity, mainly to avoid artifacts when the virtual sources move, and compensate the room effects at certain control points within the listening area. Graphics Processing Units (GPUs) are well known for their potential in highly parallel data processing. This work presents a GPU-based implementation of a WFS system with Room Compensation that yields to render maximum number of sound sources in real time.

RESUMEN

El sistema WFS es un sistema de reproducción de audio que recrea con bastante verosimilitud el campo producido por una fuente sonora dentro de una área de escucha. Cuando el número de fuentes a sintetizar y de altavoces en el sistema es amplio, los requisitos computacionales aumentan significativamente. Por otra parte, aprovechando los altavoces de este sistema, es posible la introducción de un filtrado en una etapa previa a la reproducción que pueda compensar los efectos que la sala añade sobre la señal original. Las GPUs son unidades de procesamiento gráfico que están destacando en muchos campos de procesado de señal por su potencial en procesar los datos en paralelo. El presente trabajo analiza las prestaciones que ofrece la implementación de un sistema WFS sobre plataformas de computación GPUs.

1. INTRODUCCIÓN

En las últimas décadas, ha surgido un creciente interés en mejorar la experiencia sonora y más concretamente en la generación del sonido espacial. Hoy día, uno de los sistemas de reproducción de audio espacial más prometedor es Wave Field Synthesis [1][2], en el que un

campo sonoro se sintetiza con alta calidad en un área amplia de escucha por medio de un número elevado de altavoces. Comparado con otros sistemas, WFS proporciona una zona de escucha mucho mayor; teóricamente tan grande como la zona encerrada por los altavoces. El sistema WFS está basado en la creación de un campo sonoro determinado en una zona de escucha (generalmente en 2D a la altura de las cabezas de los oyentes) mediante la interferencia de los sonidos producidos por un número de fuentes sonoras puntuales (fuentes sonoras secundarias) localizadas alrededor de la zona de reproducción que generan el sonido que el campo sonoro deseado (fuente sonora primaria) tendría en esos puntos

Uno de los mayores problemas a la hora de implementar un sistema WFS consiste en la interacción del array de altavoces con las paredes de la sala de escucha. La sala de escucha introduce nuevas señales que se suman a la señal a reproducir, alterando así la síntesis del campo sonoro, y reduciendo el efecto espacial de este sistema. Una aproximación a la compensación de salas en WFS es obtener soluciones directas del problema de filtrado inverso multicanal mediante bancos de filtros multicanal (sistema MIMO, [3]). Este banco de filtros se colocará delante de los altavoces para compensar los efectos de la sala en múltiples puntos de control dentro del área de escucha.

Una situación especial ocurre cuando se requiere mover una fuente sonora. En un sistema WFS que procese las señales en tiempo discreto, puede darse el caso de tener que retrasar una señal un número de muestras que no es un valor entero. Para ello, proponemos en este trabajo el uso de los filtros de retardo fraccionario [4-5] que permiten generar dicho retardo. Este tipo de filtros se utilizan mayoritariamente en otras aplicaciones de audio relacionadas con los efectos de audio digital [6-7] ó síntesis de sonidos [8-9].

En un sistema donde el número de altavoces puede incluir desde decenas a centenas, existen muchas operaciones que se realizan para cada uno de los altavoces del sistema, desde el filtrado y retardo de cada una de las señales de los altavoces para el sistema WFS, hasta la convolución de dichos filtros inversos por el sistema WFS y por la señal a reproducir para cada uno de los altavoces. Sin embargo, para llevar a cabo todas estas operaciones, se necesita una alta capacidad computacional, lo que representa una severa limitación cuando se trata de ejecutar un sistema WFS con compensación de salas en tiempo real.

Actualmente, las mejoras en la capacidad computacional están íntimamente ligadas al número de unidades de proceso que tiene un computador, lo que permite un alto grado de paralelismo en computación. Este es el caso de las Unidades de Procesamiento Gráfico (GPUs, Graphics Processing Units), que poseen actualmente miles de núcleos computacionales. Las GPUs se han relacionado tradicionalmente con la computación gráfica o el tratamiento de imágenes, pero con la aparición de nuevos entornos de programación para GPUs (CUDA o OpenCL) muchas aplicaciones de otros campos científicos han podido ser aceleradas mediante su implementación en las GPUs.

El artículo se estructura de la siguiente forma. La Sección 2 describe brevemente algunos características de las GPUs y cómo se gestiona una aplicación de audio en tiempo real. Una visión general de las características de un sistema WFS con compensación de salas se describe en la Sección 3. La Sección 4 enumera como se han utilizado los recursos de la GPU para la implementación en GPU del sistema WFS y presenta las prestaciones que obtiene dicha implementación variando el tamaño de los filtros utilizados en la compensación de salas y el número de fuentes sonoras a utilizar. Finalmente, la Sección 5 resume las principales conclusiones de este trabajo.

2. UNIDADES DE PROCESAMIENTO GRÁFICO

Actualmente, la computación paralela se ha incorporado a la informática de consumo. Gran parte de este éxito ha sido debido a la industria de los videojuegos, donde la demanda de los usuarios para obtener mejores gráficos y prestaciones ha provocado una gran inversión en las Unidades de Proceso Gráfico, las GPUs. El potencial de cálculo de dichas unidades no ha pasado desapercibido en los principales centros de investigación, y se están utilizando dichas unidades de proceso para acelerar aplicaciones en diferentes campos, entre ellos el procesamiento masivo de audio digital. Estas aplicaciones están relacionadas con la acústica de salas [10], sonido binaural [11], filtrado adaptativo [12], entre otras.

La programación se realiza utilizando CUDA [13], que es un lenguaje C que permite programar las GPUs de la compañía Nvidia. Con CUDA, podemos gestionar los hilos de ejecución que se lanzan en paralelo cuando ejecutamos código en una GPU. Estos hilos de ejecución son conocidos como *threads*. La zona de código donde la GPU realiza sus operaciones, recibe el nombre de *kernel*, donde se indica las instrucciones que realiza un *thread*. Estas instrucciones serán ejecutadas por todos los *threads* lanzados y de manera paralela.

El código CUDA permite lanzar múltiples *threads* y en número mayor a los procesadores reales que tenga la GPU, la cual, tiene un módulo llamado *warp-scheduler* que se encarga de gestión de los *threads* lanzados. Un aspecto importante a tener en cuenta a la hora de la programación es el acceso de cada uno de los *threads* a los datos de manera paralela, y evitar así posibles condiciones de carrera.

2.1 Aplicaciones en tiempo real

Para nuestra implementación, trabajamos con tarjetas de audio MOTU. Estas tarjetas funcionan con bloques de muestras por canal. Así, cada L/f_s segundos entregan y recogen buffers de L muestras por canal. Dependiendo si el canal es de entrada o salida, las muestras de estos buffers serán procesadas (muestras captadas por los micrófonos), o bien reproducidas por los altavoces (ya han sido procesadas). La Fig 1 muestra la gestión de buffers de audio en un sistema multicanal compuesto por 4 señales de entrada y dos de salida.

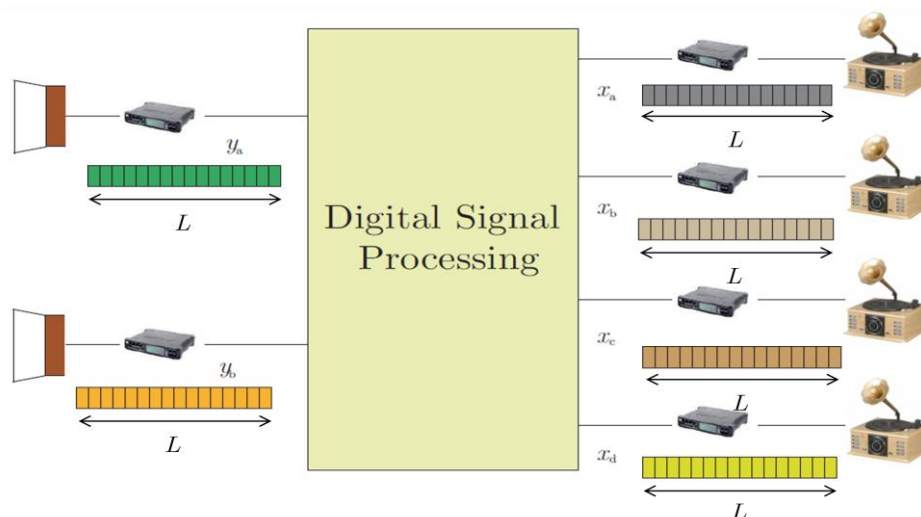


Fig 1. Gestión de buffers de audio en un sistema multicanal compuesto por 4 señales de entrada y dos de salida.

Si llamamos t_{buff} al cociente L/f_s y t_{proc} al tiempo empleado por las operaciones de procesamiento desde que se recogen los buffers de entrada de la tarjeta de audio hasta que se calculan los

buffers de salida, podemos observar que la aplicación multicanal funcionará en tiempo real siempre y cuando t_{proc} sea menor o igual a t_{buff} . Es importante matizar que t_{buff} es independiente del número de canales que tenga el sistema, mientras que el t_{proc} si que depende. Por tanto, para conocer cuáles son las prestaciones de un sistema multicanal, aumentaremos el número de canales (de entrada o salida, según la aplicación) y mediremos t_{proc} hasta que su valor supere el t_{buff} . Será entonces cuando sepamos el máximo número de canales que puede soportar nuestra implementación sobre una GPU funcionando en tiempo real.

3. WAVE FIELD SYNTHESIS Y COMPENSACIÓN DE SALAS.

La base teórica del WFS viene dada por el principio de propagación de Huygens[1-2]. Este principio indica que la propagación de un frente de ondas de una fuente sonora (Fuente primaria) puede ser reconstruido por la intersección de múltiples frentes de ondas, pertenecientes a diversas fuentes sonoras (Fuentes secundarias). La Fig 2 muestra el principio del WFS.

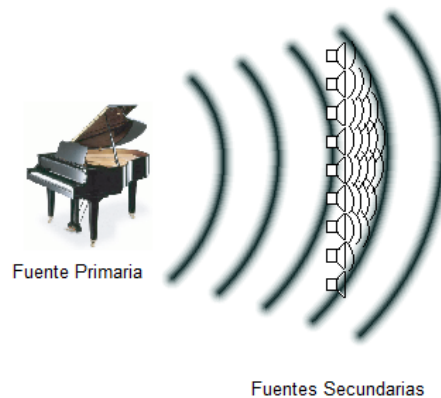


Fig 2. Principio del sistema Wave Field Synthesis.

Cada altavoz dentro de un array de altavoces, funciona como fuente secundaria. La señal q_n que debe ser reproducida por el n -ésimo altavoz para correcta síntesis del campo sonoro viene dada por la ecuación:

$$q_n[k] = \sum_{m=0}^{M-1} a_{mn} \cdot (s_m[k] * \delta[k - \tau_{mn}] * h[k]).$$

donde n representa el n -ésimo altavoz en un array compuesto por N altavoces. La variable m representa la m -ésima fuente sonora en un sistema compuesto por M fuentes sonoras. El coeficiente a_{mn} es un escalar que pondera la fuente primaria s_m cuando es reproducida por el altavoz. La variable k es utilizada como el índice de la muestra dentro un sistema discreto de procesado de señal, mientras que la variable τ_{mn} corresponde al retardo que se debe llevar a cabo (en numero de muestras) sobre la fuente primaria s_m antes de ser reproducida por el altavoz n -ésimo. La teoría sobre el desarrollo de esta ecuación puede ser encontrada en [14-17]. El filtro $h[k]$ es descrito en [18] como un filtro de pre-equalización. Por tanto, la señal que será reproducida por las fuentes secundarias se consigue retrasando y ponderando las muestras correspondientes a las fuentes sonoras primarias.

3.1 Compensación de Salas.

La compensación de salas en WFS consiste en filtrar las señales de cada altavoz q_m por un banco de filtros multicanal, el cual, se colocará delante de los altavoces para compensar los efectos de la sala en múltiples puntos de control dentro del área de escucha. Así pues, la señal que será finalmente reproducida por los altavoces, vendrá dada por la ecuación:

$$y_n[k] = \sum_{j=0}^{N-1} q_j[k] * f_{jn}[k].$$

donde f_{jn} representa uno de los filtros del banco inverso de filtros, tal y como muestra la Fig 3. Por tanto, además del procesado de las señales de cada altavoz, hay que sumarle al sistema WFS N^2 filtrados.

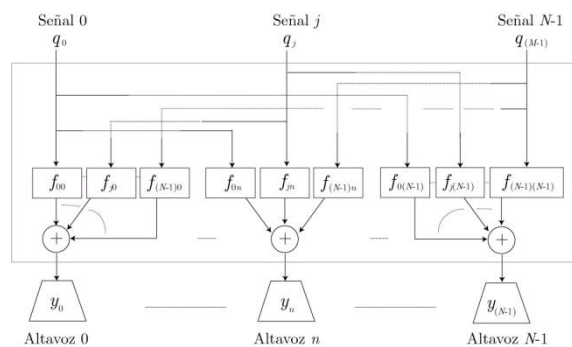


Fig 3. Banco de filtros multicanal.

3.2 Retardos Fraccionarios.

Otra característica a tener en cuenta en un sistema basado en el procesado digital de señal, es cómo generar los retardos de muestras fraccionarios, es decir, casos en los que τ_{mn} no sea un número entero.

En [8] se presentan diferentes técnicas de generar estos retardos fraccionarios a partir de diferentes tipos de interpolaciones: lineal, cúbica ó Lagrange.

Para esta implementación, vamos a usar la interpolación particionada de Lagrange que ofrece excelentes resultados, tal y como se aprecia en [19]. Para ello, se diseña para cada valor de retardo τ_{mn} un filtro FIR de orden 10. Por tanto, MN filtros deberán ser diseñados para posteriormente ser utilizados para filtrar la fuente primaria s_m . Por tanto la cantidad de filtrados del sistema WFS suma un total de $MN + N^2$ filtrados.

4. PRESTACIONES DE UN SISTEMA WFS EN GPU.

A la hora de implementar el sistema WFS en GPU, podemos observar que existe un paralelismo de grano fino. En la GPU implementaremos diferentes *kernels* donde cada *thread* se encargará de calcular el valor de una muestra de salida. Podemos distinguir dos tipos de *kernels*. Por una parte, están los *kernels* que se encargarán del filtrado. Estos *kernels* computan el filtrado en frecuencia y con un solape del 50%. Por tanto, cada *kernel* lanzará $2LNM$ *threads* para el cálculo de cada una de las muestras de las NM señales que corresponden a la pareja Fuente Primaria – Fuente Secundaria.

El segundo tipo de *kernels* se encargará de sumar las señales procesadas para cada uno de los altavoces. Para estos casos, únicamente $2LN$ *threads* serán lanzados y cada *thread* se encargará de realizar M sumas. Detalles en la implementación de estos *kernels* destinados al procesamiento de las señales q_n están descritos en [20].

Por otra parte, la implementación de la etapa de filtrado multicanal correspondiente al banco de filtros inversos está detallada en [21].

4.1 Resultados

Para evaluar las prestaciones de la implementación GPU, se ha utilizado el sistema WFS que se encuentra ubicado en el laboratorio del gtac [22], en la Universitat Politècnica de València, el cual, está compuesto por 96 altavoces ($N=96$). Ver Fig 4.



Fig 4. Sistema WFS compuesto por 96 altavoces del Instituto iteam en la UPV.

La Fig. 5 muestra los tiempos de procesamiento t_{proc} para el caso particular de buffers de muestras $L=256$ y para diferentes tamaños de los filtros compensadores f_{jm} , denotados en la figura como L_f . Debemos tener en cuenta la línea que marca el t_{buff} separa las combinaciones que pueden ser ejecutadas en tiempo real.

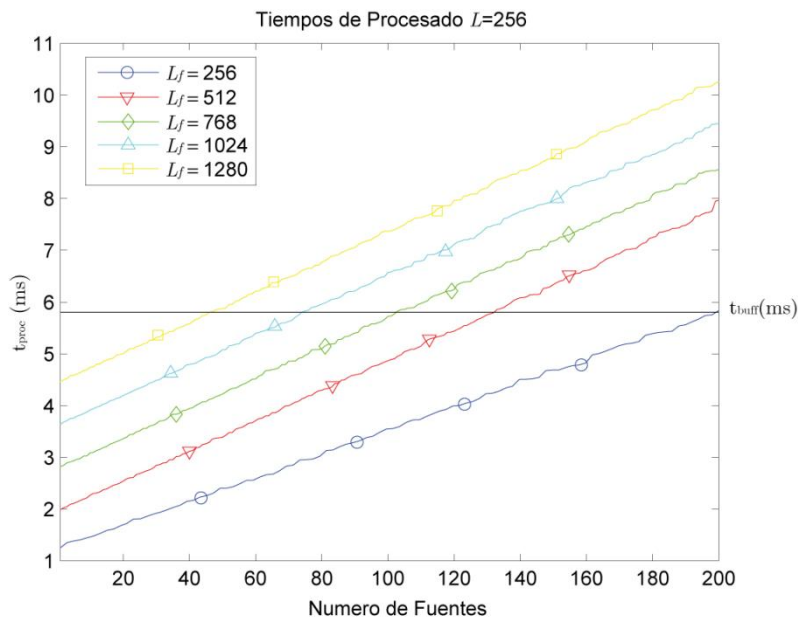



Fig 5. t_{proc} para el caso particular de buffers de muestras $L=256$ y variando el tamaño de los filtros compensadores f_{jm}

Entre otros datos, podemos extraer de la gráfica que utilizando un t_{buff} de 5.805 ($f_s = 44.1$ kHz), podemos implementar sistemas WFS, que utilicen filtros f_{jn} con tamaños de 1280 coeficientes y que pueden llegar a gestionar hasta 45 fuentes en tiempo real (corte de la gráfica , con la línea marcada por t_{buff}). De la misma manera se puede proceder para otros tamaños de filtros.

La Tabla 1 completa los resultados mostrados por la Fig.5 y recoge las prestaciones del sistema WFS variando el tamaño de los buffers de entrada L , y por tanto el t_{buff} , y el tamaño de los filtros utilizados en el banco de filtros inversos f_{jn} . La columna 4 muestra el tiempo de procesado que se requiere para reproducir una única fuente, mientras que la columna 6 muestra el tiempo de procesado para el máximo número de fuentes que pueden ser reproducidas en tiempo real (Columna 5). Nótese que los valores máximos alcanzados para el caso $L=256$ también se pueden obtener a partir de la Fig. 4.

L	t_{buff}	Tamaño de f_{jn}	t_{proc} de una fuente (ms)	Fuentes Tiempo Real	t_{proc} Fuentes T. Real (ms)
256	5.805	256	1.186	198	5.752
		512	1.983	132	5.793
		768	2.821	102	5.770
		1024	3.634	73	5.744
		1280	4.457	45	5.750
		1536	5.303	18	5.791
		1792	6.115	0	0
512	11.610	512	2.267	232	11.606
		1024	3.894	147	11.579
		1536	5.542	117	11.583
		2048	7.215	85	11.569
		2560	8.861	53	11.527
		3072	10.509	21	11.517
		3584	12.174	0	0
1024	23.22	1024	4.343	262	23.175
		2048	7.650	162	23.062
		3072	10.939	130	23.198
		4096	14.234	94	23.092
		5120	17.618	60	23.199
		6144	20.880	25	23.121
		7168	24.154	0	0

Tabla 1. Prestaciones del sistema WFS para tamaños de buffers $L=256$, $L=512$ y $L=1024$ para diferentes tamaños de los filtros compensadores f_{jn} .

Es importante apreciar como las entradas de la tabla de la columna 5 que indican 0 fuentes en tiempo real es porque el t_{proc} de una fuente es ya superior al tiempo t_{buff} y por tanto, la condición de tiempo real, no se cumple.

6. CONCLUSIONES.

En este trabajo, hemos mostrado las prestaciones que pueden dar las Unidades de Procesado Gráfico a un sistema de audio espacial WFS, al que además se le ha añadido un banco de filtros multicanal cuyo objetivo es compensar los efectos de las salas; y un filtrado adicional a

cada combinación fuente-altavoz para general retardos fraccionarios de señal. Este trabajo sirve como modelo y establece los límites del tipo de salas donde se encuentre el sistema. Así, a salas más reverberantes será necesario utilizar filtros compensadores con mayor número de coeficientes, lo cual, limitará el número de fuentes a gestionar en tiempo real y el tamaño de buffer, el cual, a su vez está también relacionado con la latencia del sistema.

7. REFERENCIAS

- [1] A.J. Berkhout, "An holographic approach to acoustic control", Journal of the Audio Engineering Society, vol.36, pp.977–995, December 1988.
- [2] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis", Journal of Acoustic Society of America, vol. 93(5), pp. 2764–2778, May 1993.
- [3] J. J. Lopez, A. Gonzalez, L. Fuster, "Room Compensation in Wave-Field Synthesis by means of multichannel inversion" IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, EEUU, October 2005.
- [4] T. Laakso, V. Välimäki, M. Karjalainen, and U. Laine, "Splitting the unit delay: Tools for fractional delay filter design," IEEE Signal Processing Magazine, vol. 13, no. 1, pp. 30–60, 1996.
- [5] M. Karjalainen, T. Paatero, J. Pakarinen, and V. Välimäki, "Special digital filters for audio reproduction," in Proc. of the 32nd AES Conference, Hillerod, Denmark, September 2007.
- [6] J. Dattorro, "Effect designpart 2: Delay-line modulation and chorus." J. Audio Eng. Soc, vol. 45, no. 10, pp. 764–788, 1997.
- [7] U. Zölzer, "Dafx - digital audio effects (second edition)," Edited by Udo Zölzer, 2011.
- [8] V. Välimäki, "Discrete-time modeling of acoustic tubes using fractional delay filters," Ph.D. dissertation, Helsinki University of Technology, 1995.
- [9] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, "Discrete-time modelling of musical instruments," Reports on Progress in Physics, vol. 69, no. 1, pp. 1–78, 2006.
- [10] L. Savioja, "Real-time 3D finite-difference time-domain simulation of low- and mid-frequency room acoustics," in Proc. Of the Int. Conf. Digital Audio Effects, Graz, Austria, September 2010.
- [11] J. A. Belloch, M. Ferrer, A. Gonzalez, F.J. Martinez-Zaldivar, and A. M. Vidal, "Headphone-based virtual spatialization of sound with a GPU accelerator," J. Audio Eng. Soc, vol. 61, no. 7/8, pp. 546–561, 2013.
- [12] J. Lorente, A. Gonzalez, M. Ferrer, J.A. Belloch, M. De Diego, G. Piñero, and A.M. Vidal, "Active noise control using Graphics Processing Units," in Proc of the International Congress on Sound and Vibration, Vilnius, Lithuania, July 2012.
- [13] NVIDIA CUDA Developer Zone: <https://developer.nvidia.com/cuda-downloads>
- [14] P. Vogel, "Application of Wave Field Synthesis in Room Acoustics," Ph.D. dissertation, Delft University of Technology, 1993.
- [15] E. Start, "Direct sound enhancement by Wave Field Synthesis," Ph.D. dissertation, Delft University of Technology, 1997.
- [16] E. Verheijen, "Sound reproduction by Wave Field Synthesis," Ph.D. dissertation, Delft University of Technology, 1997.
- [17] A. Franck, "Efficient Algorithms for Arbitrary Sample Rate Conversion with Application to Wave Field Synthesis," Ph.D. dissertation, Ilmenau University of Technology, 2012.
- [18] S. Spors and J. Ahrens, "Analysis and Improvement of Pre-equalization in 2.5-Dimensional Wave Field Synthesis," in Proceedings of the 128th AES Convention, London, UK, May 2010.
- [19] V. Välimäki and A. Haghparast, "Fractional Delay Filter Design Based on Truncated Lagrange Interpolation," IEEE Signal Processing Letters, vol. 14, no. 11, p. 816819, November 2007.
- [20] J. Belloch, M. Ferrer, A. Gonzalez, J. Lorente, and A. Vidal, "GPU-based WFS Systems with Mobile Virtual Sound Sources and Room Compensation," in Proc. of the 52nd AES Conference, Guildford, United Kingdom, September 2013.
- [21] J. A. Belloch, A. Gonzalez, F. Martinez-Zaldivar, and A. M. Vidal, "Multichannel massive audio processing for a generalized crosstalk cancellation and equalization application using GPUs," Integrated Computer-Aided Engineering, vol. 20, no. 2, pp. 169–182, 2013.
- [22] Instituto de las Telecomunicaciones y Aplicaciones Multimedia: <http://gtac.upv.es/>