PC Based Real-Time Multichannel Convolver for Ambiophonic Reproduction

Jose Javier Lopez and Alberto Gonzalez
Departamento de Comunicaciones, Universidad Politecnica de Valencia
Camino de Vera s/n, 46022 Valencia, Spain
Email: jjlopez@dcom.upv.es, agonzal@dcom.upv.es

This paper describes a very efficient software tool for real-time multichannel convolution. This software is applied to the reproduction of Ambiophonic surround sound. The program runs on a Windows based PC using compatible sound cards. The binaural input signal is acquired by the sound card and then processed by the software producing outputs for the stereo-dipole and for the surround loudspeakers. Using state of the art PCs is possible to achieve set-ups with more than 10 loudspeakers and simulate room impulse responses larger than three seconds.

INTRODUCTION

At the present time a strong tendency exists to increase the realism of the sound reproduction systems. It is looked for as much spatial sensation as re-creation of acoustic environments, [1][2].

The multichannel sound systems, well established in the industry of the cinema, try to re-create this type of acoustic sensations. From the first systems of rendering sound until the current 5.1, 6.1 and 7.1 systems, a great evolution of signal processing techniques has taken place. These signal processing techniques have been used mainly in digital compression of the multichannel sound, giving place to different standards, which became property of different companies. However, these systems still lack for providing a true periphonic space sensation and sound continues seeming to come from the speakers. In order to improve present systems, it is planed to increase the number of channels to 10.2 or even more. As consequence of the increase of the number of channels two main problems appear:

- New difficulties are added in the production and mixing of multichannel sound and also new special techniques of mixing are needed.
- An increase of the memory necessary to store many channels and /or to transmit them.

On the other hand, the systems based on the HRTF, binaural systems and transaural systems using cross-talk cancellation filters, have also experienced a considerable evolution. These systems have the advantage that only two channels are required for their transmission. However their application field is more limited because they force the listener to be located at a fixed point. As well the listeners can undergo confusions in the location of sound sources, especially front/back confusion due mainly to the diversity of pinnae between different persons, making this system not suitable for commercial purposes.

Wavefield reconstruction methods, as for instance Ambisonics, require more than two channels for transmitting the spatial information. Special recording techniques and microphones must be employed in the recording phase [3]. Furthermore, improvement of the reproduction needs a large number of loudspeakers.

Recently a novel system of sound reproduction called Ambiophonics has been proposed [4]. This system combines the well-known techniques of cross-talk cancellation using closely-spaced loudspeakers to reproduce the front signal [5], and a group of loudspeakers surrounding the listener in order to reconstruct hall ambience signals. These signals are derived from the left and right channels coming from the CD, convolved with a set of real hall impulse responses.

Figure 1 shows a possible ambiophonic set-up. The listener is located in front of a stereo-dipole [5] and a set of loudspeakers surrounds the listener.

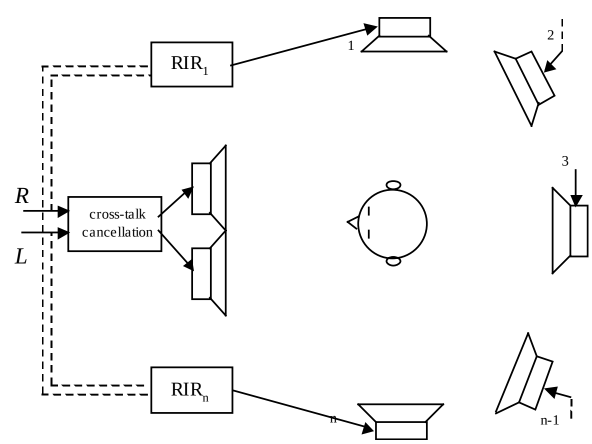

Figure 1. Scheme of Ambiophonic Sound System

Ambiophonic surround sound system presents important advantages compared with other surround sound systems (binaural based, Ambisonics or proprietary 5.1 systems):

- Only two channels are stored or transmitted.
- Sensation of 3D-space is good.
- The system is capable of reproducing any kind of pre-recorded live music material not initially prepared for Ambiophonics [4].

However, powerful computational resources are needed in order to generate the signals that feed the different loudspeakers in a typical Ambiophonic set-up. A cross-talk cancellation system is required to obtain the stereo-dipole signal. The cross-talk filters are usually not too long. On the other hand a set of quite long impulse responses that simulate the original room impulse response (RIR) are convolved with the source signals.

As an example, a RIR of 1500 ms becomes a very long filter (66050 samples) at a sampling rate of 44100 Hz. When the system feeds N surround loudspeakers, 2*N convolutions of this size must be performed in reproduction-time. These high computational requirements represent the main drawback of this reproduction system.

HOST SIGNAL PROCESSING

The new techniques in the sound field demand more and more processing power to implement the digital signal processing algorithms. The manufacturing companies of Digital Signal Processors (DSP) are aware of this fact and bring periodically faster and more powerful DSP to the market. Different companies coexist that manufacture different DSP models, all them incompatible to each other. Also the development tools for these DSP are not in many cases easy to manage. All these difficulties, together with the difficulty of designing the complementary hardware to produce a complete system of audio processing using DSP, made us to think about other ways of carrying out digital audio processing.

The personal computer has experienced an impressive evolution in the last years, as much in the process speed as in the multimedia devices that can be connected to it. The computational power of PCs double each year, whilst the computational power of DSPs doubles each two years. Ten years ago, DSPs were much faster than the PCs of that time, but nowadays it is true the opposite. As a result of the aforementioned statement, the personal computer can be used for real-time digital signal processing applications.

Host signal processing presents some advantages compared to DSP systems:

- The development tools for PC (compilers, debuggers,...) are much more powerful and easier to use than those of the DSP.
- Friendly user interfaces for the final application can be developed.
- There are more people qualified to program PC applications than to do it over DSPs.
- The cost of a basic computer (motherboard, processor, memory) is in most of the cases cheaper that a DSP development board.
- The current processors, PIII-PIV-K7, have much more processing power than the most powerful DSP.
- There is a great offer in the market of high-quality multichannel audio cards.

However, there are drawbacks:

- Impossibility of working sample by sample. The transfer with the sound card is carried out with blocks of samples.
- Unavoidable processing latency.
- Use of non real-time operating systems (Windows, Linux,…) for a real-time application.

The software optimization represents a crucial aspect of the algorithms running on a PC. We use fast convolution methods (overlap-save) based on Fast Fourier Transforms (FFT) [6]. Therefore optimization of the FFT algorithm is an essential software need. There are mainly two public domain software libraries that supply FFT routines for software developers: the FFTW [7] and the Intel Signal Processing Library (SPL) [8]. The first one provides an open code of C language and can be compiled to run on different platforms and operating systems. The second one is distributed already compiled in two formats: static libraries (.LIB) and dynamic libraries (.DLL). This software library works only for Intel (and compatible) processors over Win32 environments.

The Intel SPL has been used in this project because it yields lower execution time than FFTW over Intel processors and also because we had some experience in its use, gained from previous developments [9][10].

INTEL SIGNAL PROCESSING LIBRARY

Introduction

The Intel Signal Processing Library provides a set of signal processing routines optimized for general-purpose Intel (and compatible) processors rather than specialized DSP processors. It is targeted to non-real-time applications. The computing power of the latest generations of processors enables the use of many signal processing functions which previously were done by add-in DSPs. The library includes functions for finite impulse response (FIR) and infinite impulse response (IIR) filters, fast Fourier transforms (FFTs), wavelet transforms, tone generation, and many vector operations. It also allows the CPU to process audio, video, and communications data using software only, rather than down-loading the data to fixed-function, dedicated digital signal processing hardware.

Hardware/Software Requirements

The Signal Processing Library is designed for use on 32-bit processors (Intel386™ processors and higher). Performance of the library is optimized on processors of the latest generations, such as Pentium ® III processors. The library includes a DLL which detects the processor on which it is running and loads an appropriate processor-specific DLL. The processor-specific DLLs are also provided. The user's system must be running the Windows (95/98/Me/NT/2000) operating system.

Performance Analysis

In order to test the performance of the library for the FFT algorithm, an exhaustive test has been carried out. In this test, three different personal computers with different features and running different operating system versions have been compared. The next table shows a summary of the PCs features.

	Processor	Speed	RAM	O.S.
PC 1	Intel P-II	350 MHz	64 MB	Win Me
PC 2	Intel P-II Mobile	400 MHz	180 MB	Win 98 SE
PC 3	AMD K7	1000 MHz	256 MB	Win 2000

FFTs of real vectors in single precision floating point, whose sizes were between 1024 and 262144 (2^{10} and 2^{18} respectively), have been carried out. In order to avoid unpredictable measure errors an average between a large number of simulations is presented. No other program was running on the PC while the simulations were carried out. Figures 2, 3 and 4 show simulations results in solid line. It can be observed that results tend to separate from the expected '*n* log *n*' line (discontinuous line) as the vector size increases.

Algorithm performance degradation can be due to cache memory failures that cause processor idle cycles while data is been transferred between the cache and the main memory. Therefore degradation of the system performance is low for short vectors because the whole data is loaded in the cache. Ratio between the measured time and the expected (theoretical) time is presented in figures 5, 6 and 7 (1 represents no performance degradation).

Computation times for the FFTs are very good (compared with commercial DSP), especially for the faster PC tested. For instance a 1024 points single precision real FFT takes only 28 μs in this processor and a 16384 points FFT takes 0.71 ms.

SYSTEM ARCHITECTURE

This software tool has been programmed to carry out several simultaneous (multichannel) convolutions. Audio signal is obtained from the sound card in real time. Then the samples are converted to floating point single precision arithmetic, and all the signal processing is done in floating point.

The software runs over Windows (95/98/Me/NT/2000) and a standard Personal Computer, although state-of-the-art PC is recommended in order to increase filter lengths and number of surround channels.

System has been developed employing standard operating system sound drivers allowing to use any compatible sound card, although a high-quality multichannel sound card is recommended for scientific or professional applications. Actually there are several multichannel sound cards offering eight or more analog output channels, and also digital inputs/outputs (SPDIF).

Figure 8 shows how the different parts of the system interact. The application software drives the sound card through the standard OS and sound card drivers. Therefore, the application software can be used with any sound card. The number of output channels that can be commanded by the application software depends on the hardware capabilities and the available computational resources.

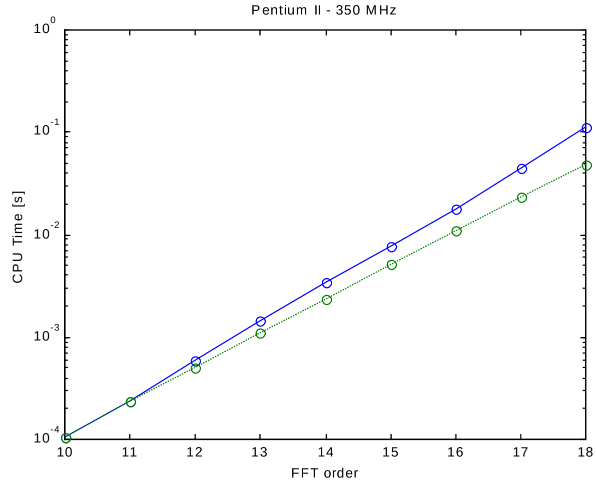

Figure 2. CPU Times for real FFT (PC-1)

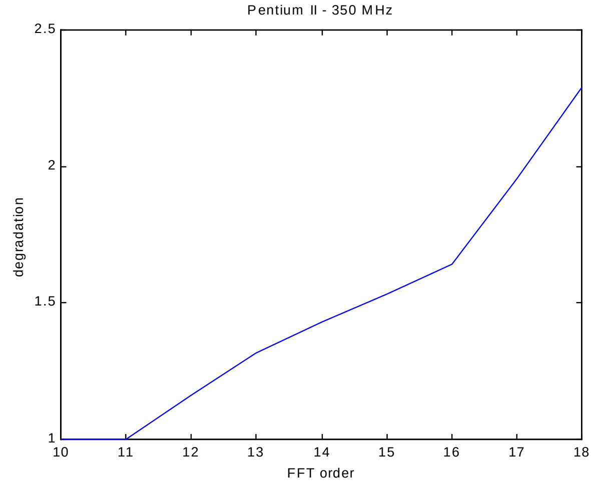

Figure 5. Performance degradation (PC-1)

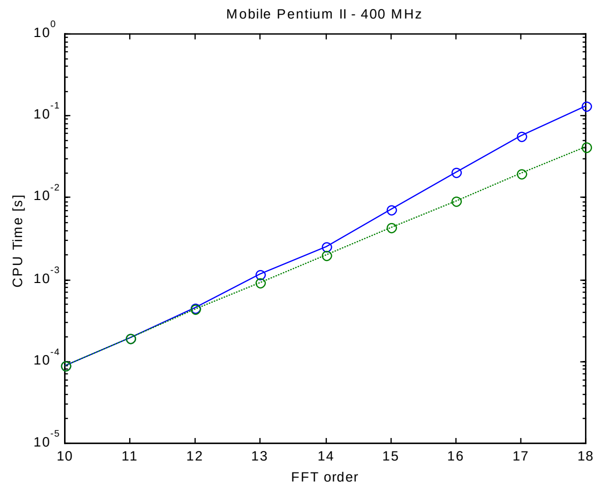

Figure 3. CPU Times for real FFT (PC-2)

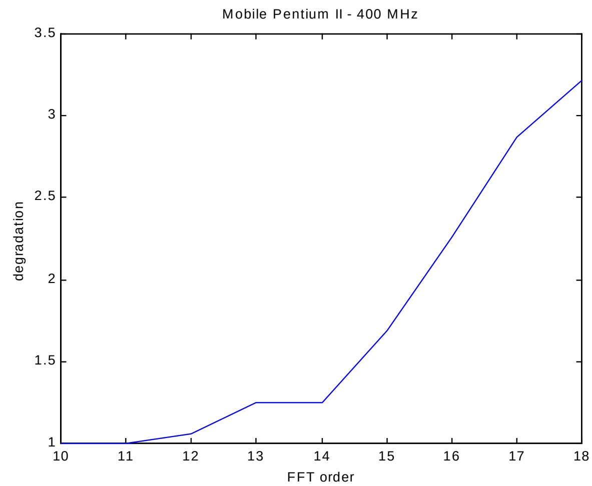

Figure 6. Performance degradation (PC-2)

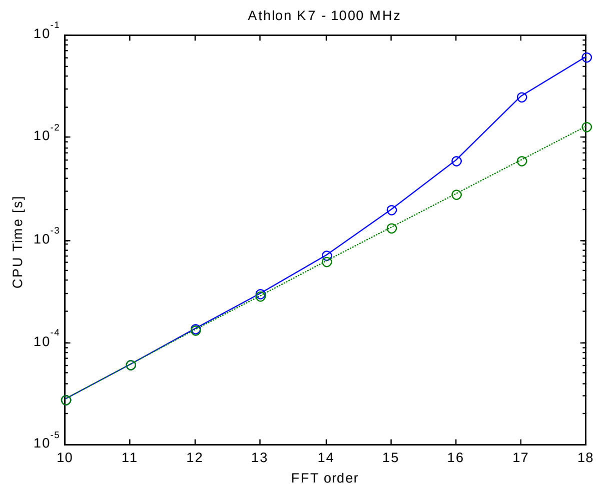

Figure 4. CPU Times for real FFT (PC-3)

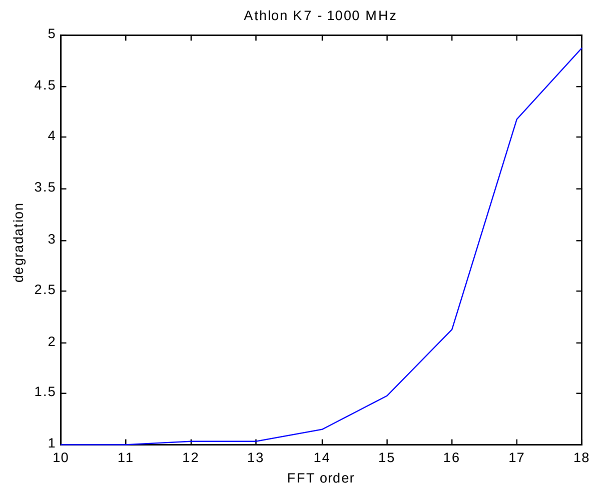

Figure 7. Performance degradation (PC-3)

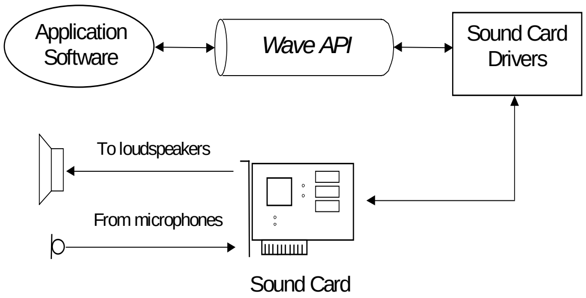

Figure 8. Sound system architecture

Operating system (OS) usually interacts with a multichannel sound card grouping the D/A channels in pairs, therefore an 8 output channels sound card is handled by the OS as 4 stereo devices. The developed software was designed for driving these separate stereo devices, and thus it can drive simultaneously several physically separate sound cards. However, output signals from different sound cards can suffer from lack of synchronism since each sound card follows each own internal clock. After working for a while, time differences between output signals could be appreciated. There are only a few sound cards that allow synchronization between them by means of a wiring system. A multichannel sound card, instead, ensures perfect synchronization between all its stereo devices.

SOFTWARE DESCRIPTION

The software tool has been developed using the Borland Delphi compiler. Delphi uses Object Pascal and has a versatile IDE and powerful debugging tools. This compiler produces very optimized code for the Intel processors family. The software presented along this paper has been named *Ambiovolver*.

Figure 9 shows *Ambiovolver* main window. A friendly GUI (graphic user interface) provides to the user an easy handling and fast learning. From the main window the user can reach different configuration windows, *Load* the bank of filters and *Play* or *Stop* the convolution.

In order to provide the impulse responses (IR) to the software, a simple and flexible method is used. Files containing the IR must be stereo .WAV files. The files have to be named as follow: *'anythingN.wav'* where N represents the number of the output channels. The first two files (N=1 and N=2) must correspond to the stereo-dipole pair. For each stereo WAV file, the left channel contains the impulse response that has to be convolved with the left program signal, and idem for the right channel. Results of two convolutions are added and sent to the output after applying a gain factor selectable by the user.

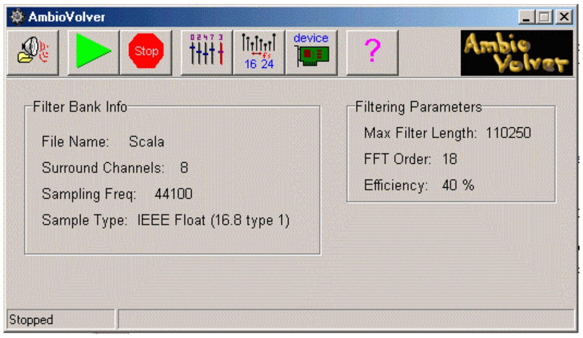

Figure 9. *Ambiovolver* main window

A wide variety of WAV file formats are accepted by the software. Particularly the following integer and floating-point formats are supported:

- 16 bit integer PCM
- 20 bit packed integer PCM
- 24 bit packed integer PCM
- 32 bit integer PCM
- IEEE floating point (16.8) [CoolEdit]
- IEEE floating point (0.24) [type 3]
- IEEE floating point (24.0)

After loading a given bank of filters, two information boxes in the main window show information about them (length, bits, sample format, number of channels, ...), as shown in fig. 9.

Wave Devices

This window gives the possibility of assigning the different surround channels to the wave devices available at the computer. As shown in fig. 10, the first device is the stereo input device that feeds the convolver inputs. The next device is the output for the stereo-dipole, this one is mandatory. As it can be observed, the user has the possibility to assign up to 4 other stereo output devices (8 channels) for the surround loudspeakers.

Open order and start order of the wave devices is selectable, in order to be compatible with certain sound cards. The number of buffers can be increased to avoid reproduction gaps when other applications or the operating system interfere with *Ambiovolver*. However, this possibility increases the output latency.

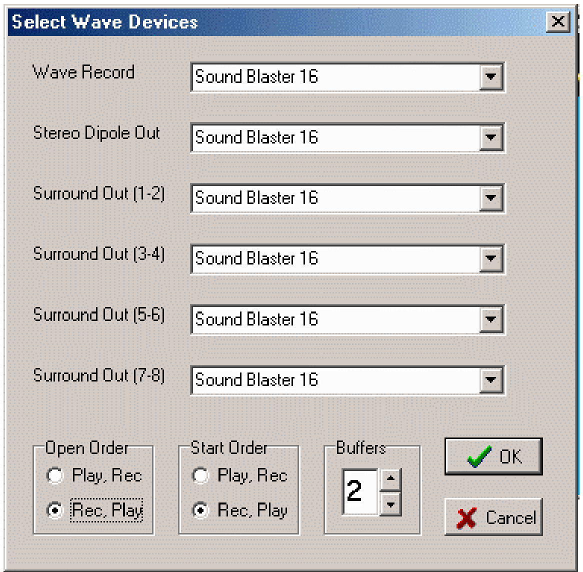

Figure 10. Selection of wave devices

Sampling Frequency and resolution bits

The user can change the default sampling frequency defined by the WAV files. *Ambiovolver* supports soundcards with high sampling frequency (up to 88.1 and 96 KHz) and 24 bits of resolution, fig. 11.

In the process of converting the floating point output samples to 16 bits integer, classic triangular PDF dither can be added for improving sound quality [11][12]. When working at 24 bits output, no dithering is applied.

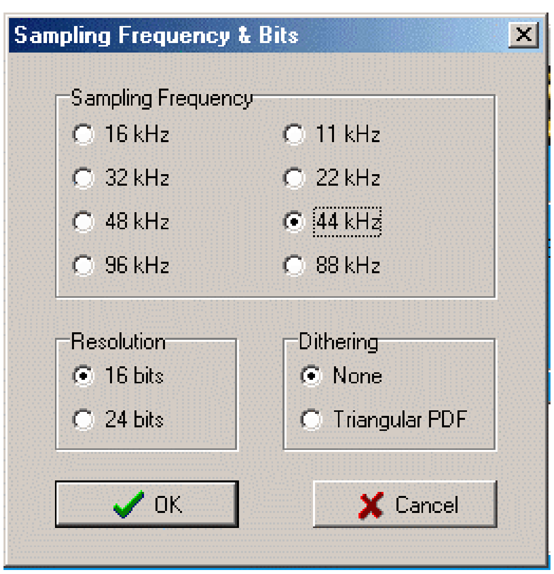

Figure 11. Sampling Frequency and Bits selection

Channel Gain Control

Despite the bank of filters intrinsically define the levels for each loudspeaker, *Ambiovolver* allows the user to modify the gain factor of each output channel before floating-point to integer conversion (prior to send the data to the D/A). This option allows to avoid clipping due to surpassing the maximum dynamic range, fig 12.

It is important to note that these gain factors are applied previously and independently to the output volume controls, which any sound card would provide. Moreover the two channels comprising the stereo-dipole can not be independently modified in order to avoid degradations in cross-talk cancellation.

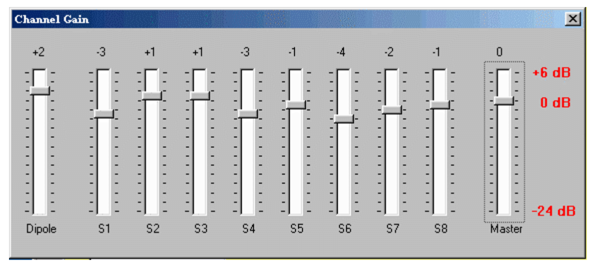

Figure 12. Control of channel gains

CONCLUSIONS

A software tool to carry out multichannel convolutions with long impulses responses has been presented. This software has been addressed to the reproduction of Ambiophonic surround sound. Using state of the art PCs it is possible to achieve set-ups with more than 10 loudspeakers and simulate room impulse responses larger than three seconds.

The main advantage of the software is given by the chance to use affordable multichannel sound cards compatible with the different Win32 based operating systems. This fact provides a high flexibility to select better sound cards newly available in the market. Additionally a multichannel convolver based on this software and a PC platform, is cheaper than other solutions recently appeared, which are based on specialized hardware.

The total latency of the system, that can be around twice the length of the RIRs employed, represents the main drawback of *Ambiovolver*. There also exists well known algorithms for carrying out fast convolutions with low latency which need only a little more computational power [13][14]. Future versions of *Ambiovolver* will include this feature.

Another software (*BruteFIR*) is available for multichannel convolution for the Linux operating system, [15]. Main advantage of this software is the partitioned convolution method [13] that provides low latency. Main drawbacks are: the software does not provide a GUI (graphic user interface) and Linux supports only a very little number of all the multichannel sound cards available in the market.

Ambiovolver is freeware and is available for download at, http://www.dcomg.upv.es/jjlopez/.

ACKNOWLEDGMENTS

This work has been partially supported by CICYT grant no.TIC2000-1683-C03.

The authors want to thank Ralph Glasgal and Angelo Farina for their valuable suggestions and e-mail tutorials about Ambiophonics that helped us to improve the software.

REFERENCES

[1] D. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA, 1994

[2] Gardner W.G., *3-D Audio using loudspeakers*, PhD Thesis, 1997, MIT

[3] Farina A., Ugolotti E., *Subjective comparison between Stereo Dipole and 3D Ambisonics surround systems for automotive applications*, 16th AES Conference, Rovaniemi (Finland), 10-12 April 1999

[4] Farina A., Glasgal R., Armelloni E., *Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music*, Proc. 19th AES Conference, Surround Sound, June 2001, Schol Dalmau (Germany)

[5] Kirkeby O., Nelson P. A., Hamada H., *The "Stereo Dipole"-A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers*, JAES vol. 46, n. 5, 1998 May, pp. 387-395

[6] Oppenheim A.V, Schafer R.W., *Discrete Time Signal Processing*, Prentice-Hall, 1989

[7] M. Frigo, S. G. Johnson *FFTW: An Adaptive Software Architecture for the FFT*. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, 1998, pp. 1381-1384.

[8] *Intel Performance Libraries. Intel Corporation Developers website*. http://developers.intel.com/

[9] Lopez J.J., Gonzalez A., *Software Toolbox for 3D Sound Using Loudspeakers*, Proc. 16th AES Conference, April 1999, Rovaniemi (Finland)

[10] Lopez J.J., Gonzalez A., *Software Toolbox for Multichannel Sound Reproduction*, Proc. of Digital Audio Effects Conference (DAFX), December 1998, Barcelona (Spain)

[11] Gray R.M., Stockham T.G., *Dithered Quantizers*, IEEE Transactions on Information Theory, Vol. 39, No. 3, May 1993, pp. 805-812

[12] Wannamaker R.A., Lipshitz S.P., Vanderkooy J., Wright J.N., *A Theory of Nonsubtractive Dither*, IEEE Transactions on Signal Processing, Vol. 48, No. 2, February 2000, pp. 499-516.

[13] Soo J. S., Pang K. K., *Multidelay block frequency adaptive filter*, IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-38, No. 2, February 1990

[14] Gardner, W. G., *Efficient convolution without input-output delay*, J. Audio Eng. Soc. 43 (3), 1994, pp. 127-136.

[15] Torger A., *BruteFIR: a software for applying long FIR filters to multi-channel digital audio*, http://www.ludd.luth.se/~torger/brutefir.html