Dottorato in Ingegneria dell'Elettronica Biomedica, dell'Elettromagnetismo
e delle Telecomunicazioni (XXI Ciclo)

Università di Roma3
Rome, Italy

Audio Digital Signal Processing: techniques and applications

Carlo Belardinelli

Tutor Professor Gaetano Giunta

A dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Philosophy
in
Engineering
in the
Applied Electronics Division
of the
UNIVERSITY OF ROMA3, ROME; ITALY

Author's Address

Carlo Belardinelli

Viale Marco Polo 84, 00154, Roma, Italy
Galvanistraat 126, 2517RE, The Hague, The Netherlands

EMAIL: cbelardinelli@epo.org

Table of contents

I

Abstract in Italiano

In questa tesi viene descritta la mia attività di ricerca svolta durante il corso di dottorato in Elaborazione Elettronica del Segnale Audio. Durante questi tre anni ho sviluppato il quadro di riferimento teorico in questo ambito realizzando poi diverse applicazioni pratiche che vengono riportate in dettaglio nei capitoli seguenti dopo un'introduzione generale (Capitolo II). In particolare sono stati sviluppati i concetti seguenti:

- Nel terzo capitolo viene presentato un sistema di marchiatura digitale audio che permette la valutazione della qualità di una trasmissione in sistemi di comunicazione di terza generazione senza influire sul payload dei contenuti inviati né tantomeno peggiorare la qualità dei contenuti musicali trasmessi, sviluppato durante il primo anno di corso presso il dipartimento di Elettronica Applicata dell'Università di RomaTre. I risultati raggiunti confermano l'efficacia del sistema e ne permettono l'applicazione in sistemi di comunicazione radio mobile di terza generazione (UMTS) e internet WIFI (IEEE 802.11).
  In particolare un watermark fragile viene inserito tramite tecniche spread-spectrum nel dominio DCT all'interno della regione delle frequenze medio-alte dei frame del segnale musicale. Il principio base dell'idea consiste nell'ipotesi che, poiché marchio e segnale audio marchiato sono entrambi sottoposti alle stesse operazioni di codifica/decodifica e viaggiano sullo stesso canale di trasmissione, il livello di degradazione del watermark dopo la ricezione è supposto essere proporzionale al livello di deteriorazione subita dal segnale audio stesso. In ricezione è infatti a disposizione la versione originale del marchio (avente un payload trascurabile): una volta estratti i watermark ricevuti dai frame del segnale audio, questi vengono mediati di modo da ottenere un'unica versione del marchio ricostruito che puo' essere utilizzato come riferimento per il calcolo del QoS (Quality of Service) della trasmissione.
  Infatti, mentre l'uso del PEAQ (Perceptual Evaluation of Audio Quality, standard ITU BS-1387) è stato utilizzato per provare l'effettiva completa trasparenza del marchio all'udito umano, i risultati delle sperimentazioni testimoniano la sensibilità del sistema di marchiatura fragile proposto nell'ottenere una stima *blind* della bontà della

comunicazione: l'errore quadratico medio del marchio (MSE), difatti, non solo si rivela essere proporzionale al numero di errori introdotti dal canale, ma é strettamente proporzionale all'errore quadratico medio del segnale audio rivevuto in cui é stato inserito.

- Nel periodo da visiting student presso il Signal Processing Lab della University of Cambridge (UK), ho implementato e verificato le prestazioni di un Phase Vocoder Probabilistico applicato alla ricostruzione dei campioni andati perduti all'interno di un file musicale. Nel quarto capitolo viene illustrato come questa applicazione abbia prestazioni incoraggianti per quel che riguarda la stima di informazioni musicali a contenuto informativo residuo nullo tramite strumenti di calcolo probabilistico bayesiano.
  Dopo aver utilizzato il Phase Vocoder come modello generativo del segnale audio, viene definita la versione stocastica dello spazio di stato. Il Phase Vocoder Probabilistico permette l'applicazione di un filtraggio di Kalman e di uno smoothing secondo le equazioni di Rauch-Tung-Striebel. L'algoritmo di Expectation Maximization permette la derivazione di una procedura ricorsiva che permette infine di stimare iterativamente tramite *maximum likelihood* i parametri del suddetto spazio di stato che meglio approssimano i campioni mancanti.

- Infine presso il Fraunhofer Institut di Magdeburg in Germania ho sviluppato delle ricerche sul tema "Interaction of Sound in Virtual Reality". In questo contesto:
  - Ho sviluppato e testato la funzione audio della piattaforma grafica per realtà virtuale, elaborata dal Virtual Development and Training Centre in seno all'Istituto stesso, che permette in maniera agile di ottenere suoni 3D all'interno di scenari virtuali (capitolo V).
    La core function è implementata utilizzando l'API OpenAL in maniera da rendere la realizzazione di contesti sonori virtuali, che si integrano con la rispettiva rappresentazione visiva, intuitiva ma al tempo stesso completa per la creazione e diffusione tridimensionale del segnale acustico. Per quanto riguarda il punto di vista del fruitore degli scenari virtuali, la spazializzazione del suono è implementata in maniera efficace di modo tale che l'elaborazione sonora avviene in real-time, anche con hardware non dedicato. L'Audio Core Function permette usi innovativi del suono in rappresentazioni 3D, testati tramite diversi scenari virtuali qui di seguito riportati.
  - Ho ideato e implementato il workflow per un plugin che permetta di visualizzare nello spazio virtuale (in particolare all'interno della Digital Factory) il campo acustico, simulato in maniera fedele da efficienti strumenti di CAE appositi, generato da macchinari industriali (capitolo VI).
    Il livello di pressione sonora (SPL) viene calcolato in maniera fedele tramite accurate simulazioni ed elaborati strumenti di calcolo ricorrendo al Boundary e Finite Element Methods. I valori di SPL ottenuti al termine dell'analisi acustica possono venire importati facilmente dalla piattaforma di Realtà Virtuale, poiché vengono salvati in formato altamente editabile. In questo modo é possibilie verificare in 3 dimensioni la conformità di progetti di nuovi siti industriali alle norme sulle soglie di rumore in ambienti lavorativi e apportare eventuali

correzioni (applicando anche tecniche di Active Noise Control, ANC), prima che la fabbrica sia concretamente realizzata.

L'attività di ricerca da me svolta durante il dottorato è stata oggetto di presentazioni in conferenze internazionali (quali la IEEE European Signal Processing Conference 2006 a Firenze o l'IEEE Workshop on Application of Signal Processing on Audio and Acoustics 2007 a New Paltz, New York) e ha mostrato l'efficacia delle tecniche di DSP per quanto riguarda il segnale audio.

Nella società odierna in cui si sovraccaricano i sensi -e quindi anche e soprattutto l'udito- di contenuti informativi, una migliore gestione del contenuto dell'informazione sonora può dunque essere una chiave per un progresso che aiuti effettivamente a vivere meglio.

II

Introduction

This thesis about Digital Signal Processing (DSP) applied to the audio signal is the outcome of three years of researches I carried out and a keen interest for music and audio in general.

Scientific investigations and passion for music have mixed together in my PhD topic. This implies the acquisition of the analog (as everything created by Mother Nature) audio content, its digitalization and its subsequent modification and processing by means of a computers in order to achieve the most various aims. During the first decades of computer development (i.e. until the end of the Seventies) this research field, if compared to the field of Digital Signal Processing in general (for example applied to the images), has not seen a wide evolution and diffusion. Although analog processing allowed yet covering fundamental needs (think about telephones invented in the nineteenth century), audio digital signal processing had not wide applications (also because of the particular intrinsic complexity of the audio signal which made inadequate the computational resources available at that time).

In the early Eighties the Compact Disc invention and diffusion, introducing to the large markets the digital music concept, and most notably the introduction of audio processing dedicated chips (with a simple and limited instruction set but tailored to the operation to be performed on the sound signal, as in the prominent case of the Motorola 56001), changed this situation. Audio DSP stopped being an exclusive field for specialized laboratories and centers of excellence and became instead a rather hot topic for researchers and developers across the world. This scenario has lead to a technological boom during the Nineties, providing the consumers with innovations as the MP3(©) format or the 5.1 home audio, while new affordable commercial sound cards have started allowing everyone owning a consumer computer to put to the test the newly discovered software audio manipulation findings.

Audio DSP is thus a relatively young but dynamic research area in which applications and new developments are continuously defined and worked out. These incremental research dynamics are reflected in this thesis together with two other life aspects which technology innovations (along with human progress) made easier, which are relocating and co-working. In

fact, the aforementioned interest for this scientific field has brought me to carry out the researches I am about to disclose in three different areas of conceivable application of the DSP techniques to audio, within three different laboratories, in three different countries in Europe. Therefore, in the frame of a clear common point, i.e. the need of modifying some audio content, in the form of a sound wave, my thesis is focused on four particular applications spanning the three aforementioned fields in order to get some particularly useful aims.

During the first period of my PhD studies I researched in my hometown, Rome, working at the Digital Signal Processing, Multimedia, and Optical Communications Lab at the University of Roma3 under the direct supervision of Professor Gaetano Giunta. Here, I focused on the field of digital audio watermarking, and in particular the processing of the audio signal in such a way, that the quality of a transmission of this very signal could be assessed without the need of investigating the physical limitations and influences of the transmission channel.

I was then visiting student at the University of Cambridge (UK). Here I had the opportunity to join the Signal Processing and Communications Lab directed by Professor William Fitzgerald. I got there an amazing insight about the statistical Bayesian approach for audio DSP. In particular I worked to the implementation of a system for the restoration of corrupted samples within a musical signal envelope.

Finally, I became Marie Curie Fellow and worked at the Fraunhofer IFF, in Magdeburg, Germany, directed by Professor Michael Schenk. Supported also by the Otto-von-Guericke University (supervisor Professor Ulrich Gabbert), I investigated the various applications of the audio signal processing within the Virtual Reality. Firstly I implemented the necessary functions to modify the sound signal and its reproduction so that, becoming three-dimensional it could sound the most real possible. Later on, I focused my research work on the numerical calculation and visual representation of sound fields within virtual factories, developing an application devoted to the noise reduction and control, another rising topic in Audio DSP.

Acknowledgements

All the above mentioned scientific and human experiences were made possible by the helpfulness of several people.

First of all I wish to thank my supervisor and tutor, Professor Gaetano Giunta for his constant scientific support through those three years. Not only he has been constantly available and full of precious teachings (very often even remotely!) but he shared his international working links fostering my international research career. I am very grateful to him.

To the doctoral Teaching Board at the Department of Applied Electronics of the University of Roma3, I am obliged for the availability shown as regards my needs as a post graduate student abroad, allowing me to pursue my scientific aims in different places with their valuable practical and academic support: without their open minds I would not be here writing this thesis.

Thanks to the grant delivered by the International Liaisons Office of Roma3, I was able to spend a research period in Cambridge: to Professor William Fitzgerald, for introducing me in his Lab with full enthusiasm and Dr. Taylan Cemgil for driving me with involvement and patience through a scientific field as interesting as completely new to me, my thankful thoughts.

Finally, I would like to show my appreciation and commitment to all People I have worked with in Germany. In particular to Doctor Eberhard Blümel, for giving me the possibility to join the Fraunhofer Community: the trust he showed allowing me to follow my research pursue while

still providing me with useful insights and practical hints related to the world of the industry were of the greatest importance. His determination and enthusiasm concerning new ideas arisen during our talks, together with his solid scientific background will always be a reference point to me. A special thank goes to Professor Ulrich Schmucker who, as a manager and supervisor along my fellowship at the Fraunhofer, introduced me into the world and technology of Computer Assisted Engineering, stimulating my researches with pressing but wise and always coherent questioning about my developments and, even more, about their weak points.

I would like to show all my gratefulness to Professor Ulrich Gabbert, Professor Tamara Nestorovic, and all People at the Machine Simulation Lab of the Otto-von-Guericke University. The support in every theoretical and computer related matter concerning mechanics and machines I was continuously given has been just precious.

Last but not least, the European Commission is also greatly acknowledged for the funding within the Marie Curie Actions (Frame Program VI).

III

Audio Digital Watermarking

The subject of this chapter consists of the theoretical fundamentals, the ideas and the experimental implementation with subsequent results coming from the first application of Audio DSP techniques depicted in this thesis. The field referred to is that of the Digital Watermarking.

1. Motivations

Last decades have witnessed the wide spreading of the use and distribution of multimedia digital contents, mostly due to the diffusion of Internet connections with increasing bandwidth. In a world where the possibility to access information resides in the reliability of access to the World Wide Web, a need of finding an effective solution to the emerging need of defending the intellectual property (i.e. the so called copyright) from the hackers' attacks has arisen. All digital contents are concerned by this issue, from digital documents, to pictures and videos, as well as audio files (one of the first planetary case in this matter was indeed in the music field, with Napster and its service allowing users to download no matter what song in digital format for free and regardless to possible copyrights) [1]. From this background, the development of appropriate DSP techniques, like Digital Watermarking, has allowed the achievement (although not yet in a definitively acceptable way and with variable results among the different fields) of the purpose to control the distribution and protection of the various multimedia contents. Watermarking of digital data in such a controversial context has become an active and variegated research field, providing interesting results and solutions [2]-[7] for safe data storage and transmission robust against malicious attacks.
The idea reported here is nevertheless not oriented to the improvement of data security but to an innovative use of audio watermarking, that is providing an assessment of the quality of an audio signal after an MP3 (MPEG-1 Layer III, ©) coding, transmission, and respective decoding.

Although, as already reported, copyright protection has been the first application of Digital Watermarking, different uses of this technique have been recently proposed; digital fingerprinting and content dampers, encipherment of data (as a new stenography expression), or

applications for data retrieval [8]-[12] are just few of the new watermarking utilizations. For what concerns the techniques used to protect the copyright avoiding unauthorised data duplications, the inserted watermark need to be detectable and not extractable without deterioration of the data themselves (i.e. it needs to be *robust*). In an opposite way, a *fragile* kind of watermark has been recently proposed [10] by Professors Campisi, Carli, Giunta e Neri of Roma3 University of Rome to assess the *Quality of Service* (*QoS*) of a video communication in a *blind* way (which means being the original version of the watermarked file unavailable for extraction at the receiver side).

For what concerns the audio field, the main focus of the research on watermarking so far has been posed on either the direct marking of the audio stream or of its compressed version [13]. While designing a watermarking scheme, the various characteristics of the file to be marked have to be kept into consideration as well as the particular application which the watermarking needs to realize. Many of the requirements for audio watermarking may be analogous to those of the picture watermarking, as for example *transparency* (i.e. the presence of the mark is supposed being not noticeable) and the robustness of the mark to manipulations and processing like compression, filtering, and A/D or D/A conversion. This means that in theory the mark should always be inserted in the data in a way which guarantees it is always not audible (a particularly strict requirement for music contents, as nearly perfect audio quality is generally considered as a fundamental point). In addition the mark embedding process should be chosen so to not blunt the mark robustness for what concerns eventual malicious attack expressly aimed to the mark removal.

It is for those reasons that the audio watermarking technique proposed in the following for assessing the audio quality received after a coding-transmission process, is supposed to be applicable without the watermarking itself affecting neither the overall characteristics of the audio signal in which it is embedded nor the transmission effectiveness. In particular a fragile watermark has been inserted in an audio data stream audio of MP3-like type (MPEG-1 layer III, [15]-[18]) using a *spread spectrum* approach [57]. On the receiving side, the watermark is extracted and compared to the original version (available on the decoding side as well, differently from the original version of the sent audio signal). The rationale supporting this approach is that the alterations suffered by the watermark as consequence of the coding and transmission process are likely to be the same as those suffered by the audio signal (or at least proportional to them) since both have been transmitted on the same communication channel. The QoS estimation is based on the calculation of the *Mean Square Error* (*MSE*) between the extracted watermark after the transmission and decoding and the original one before embedding, whereas the non audibility of the mark is tested and assessed using the most recent perceptual techniques, like the PEAQ (*Perceptual Evaluation of Audio Quality*) that is the ITU (*International Telecommunication Union*) standard for measuring the objective quality of an audio file [14]. In this case “perceptual quality” is obviously referred to the possibility of noticing the mark during the use of the music content and to the guarantee that the overall audio quality of the MP3 file is not corrupted after the watermark embedding. The proposed watermarking technique has been designed for application in multimedia communication systems. In fact, such a quality index can be used for a number of scopes in the newest telecommunications services ranging from the automatic and continuous feedback of the effective channel quality between user and radio mobile station to the real-time constant feedback to the server about the effective quality of service offered to the user.

2. Idea: fragile watermarking for QoS assessment

In this chapter the role of the watermark within the watermarking technique will be described before switching to the actual procedure of embedding and extraction of the mark itself.

In particular, it is now shown in which way the non conventional use of digital watermarking (similar to the one proposed in [10] for video signals) can trace the alterations suffered by data in MP3 format through a communication channel. Traditional techniques, as aforementioned, make use of a *robust* watermarking which implies that the imbedded mark is supposed to be spottable, noticeable and audible (for example, a watermark with copyright protection should not in theory be extractable or erasable while at the same time preventing the use or modification of the file where it is embedded, unless it is correctly detected and removed -which means the availability of a decryption key or the right extraction algorithm or procedure is needed-). In the case in point, on the other hand, a *fragile* watermarking is employed (i.e. the mark should not be noted while normally using the marked file). The rationale behind the concept of the watermarking procedure for an effective "blind" QoS assessment as said is the following: it is supposed that the modifications suffered by the watermark are likely to be similar to the ones suffered by the data, having both "travelled" on the same communication channel and therefore having undergone the same physical processes (e.g. the both have suffered from the same interferences and noises). On the receiver side, the extracted watermark is compared to the original one (since the mark has low payload it can be stored on both communication sides without relevant memory usage), in order to assess the overall degradation of the audio quality. In this way a QoS index for the communication link can be provided. Knowing the *end-to-end* signal degradation level, it becomes possible to the service provider to adopt optimal billing schemes related to the QoS profile offered to the various users involved within a communication. The QoS assessment obtained in such a way can be used at the same time for several issues in the multimedia communications using the intuitive principle that the more reliable the communication is the higher the user should be billed. On the other hand, the less effective it becomes (because of a weak or faulty link or external agents) the more resources should provided by the server to improve it [19], so to aim at an effective network administration.

.

3. Implementation

Watermark embedding

The basic scheme of the audio watermarking procedure in question is depicted in Figure 1. It is possible noticing there how firstly the audio stream is divided in M frames. A 2 seconds long time slot has been chosen for each frame so that for audio tracks sampled at the standard frequency of 44.1 KHz, 88200 samples per frame are obtained. The watermark $w[k]$ employed consists of a K bit long binary sequence.

A pseudo-random noise vectors set PN (for each frames a distinct one is employed and all are known at the receiving side) is multiplied by the reference mark (unique for each transmission process and known at the receiving side as well), in the following way:

$$w_i^{(s)}[k] = w[k] \cdot p_i[k], \quad i = 1, 2, ..., N \tag{1}$$

Here $p_i[k]$ is the *i*-th *PN* vector while $w_i^{(s)}[k]$ is the *spread* version of the mark to be imbedded within the *i*-th block or frame.
In an analogous way to what is generally known for other spread spectrum techniques, using multiple different spreading vectors PN ensures that the watermark embedding is performed differently from one frame of the host file to the other. It is worth noticing that in this way the watermark can be considered as non perceivable modification of the audio signal, being at the same time robust for what concerns the permanent bit errors caused by the physical structure of the network or by the management of the network itself (e.g. following to different paths in the transmission channel, multiuser interferences or overload factors and so on).

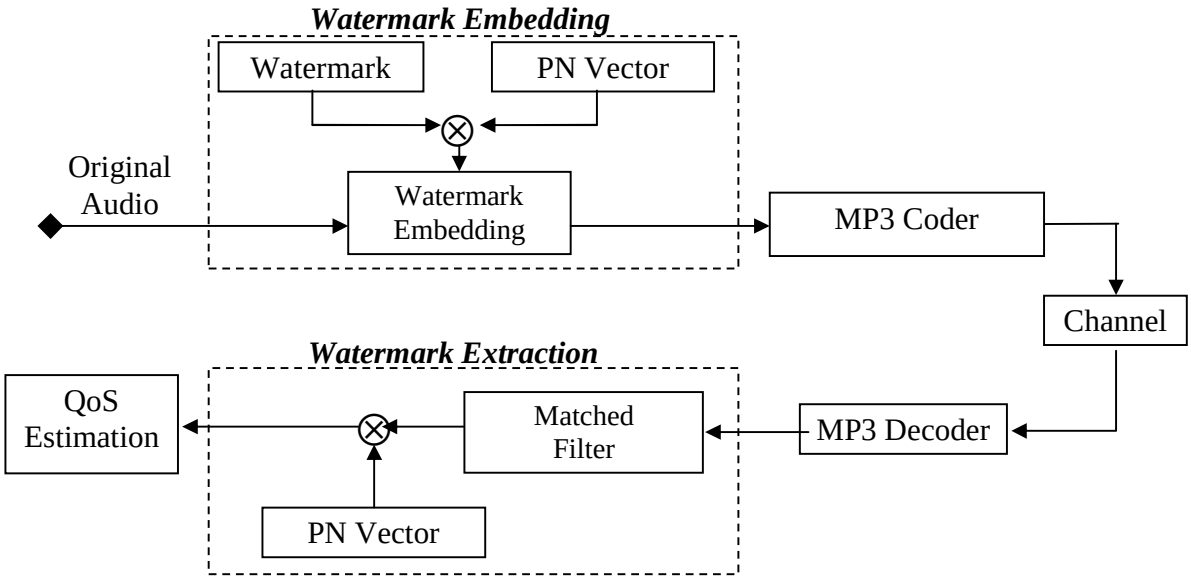


Fig. 1: *Block scheme of tracing watermarking for channel quality assessment in digital audio communications.*

After the watermark itself has been generated in the Discrete Cosine Transform (DCT) domain, the mark embedding procedure is performed in the way which follows.

Let us define the DCT of the *i-th* audio block $g_i[k]$ as

$$G_i[k] = \mathrm{DCT}\{g_i[k]\}$$

The watermark, randomized by the PN vectors is then multiplied by a coefficient α (representing the watermark *robustness*) and then added to the DCT of every and each audio frame within he middle-high frequency region *F*, according to the relation:

$$G_i^{(w)}[k] = \begin{cases} G_i[k] + \alpha w_i^{(s)}[k], & k \in F \\ G_i[k], & k \notin F \end{cases} \tag{2}$$

Here $G_i^{(w)}[k]$ indicates the *i*-th marked block. The embedding in the DCT domain makes relatively easy the operation of selecting the frequency band *F* most adapt to host the watermark payload. In fact modifying the highest frequencies of the audio signal guarantees that the modifications of the envelope due to the presence of the embedded watermark have the slightest effect to the way the watermarked audio content will be actually perceived by humans.

It is to be noted how, irrespective of the frequency bandwidth used to carry the watermark, the increase of the parameter α, will cause the watermark itself to become more audible causing a noticeable degradation of the audio signal in which it is embedded. Decreasing its value on the other hand, will make the watermark easily removable by the coding process and/or channel errors. For this reason the value of the coefficient *alfa* has been set in the following experiments (see for more details paragraph 2.5) in such a way to be a trade-off between those two antithetical constrains. After the inverse DCT transformation (IDCT), the watermarked audio signal is compressed by an MP3 encoder (MPEG layer III at various compression rates) and finally sent on the transmission channel (see Figure 1).

Watermark Extraction

At the receiving side the decoding is performed together with the watermark detection (see once more Figure 1). For this purpose, after the decoded audio stream is filtered by a matched filter which extracts the watermark from the region of interest (as aforementioned, the middle-high frequency range) within the DCT transform of the received MP3 audio signal frames. The extracted watermark is then dispread (thanks to the fact the relative vector PN is known) so as to be comparable to the reference one. The matched filter is calibrated to the actual embedding procedure, so that it can detect the region where the spread watermark is located. In particular, the dispreading operation for the generic *i*-th block is performed following the relation:

$$\hat{w}_i[k] = \hat{w}_i^{(s)}[k] \cdot p_i[k] \tag{3}$$

Here $\hat{w}_i^{(s)}[k]$ represents an estimate of the spread version regarding the watermark inserted in the *i*-th frame. It is thus possible to get an overall estimate of the received watermark $\hat{w}[k]$ on a certain number M of transmitted frames starting from the following relation:

$$\hat{w}[k] = \frac{1}{M} \sum_{i=1}^{M} \hat{w}_i[k] \tag{4}$$

Here $\hat{w}_i[k]$ is the estimate of the mark following its extraction from each *i*-th audio block. The watermark calculated by means of the equation (4) is then compared to the reference one (i.e. the original embedded on the coding side) so that a "degradation index" suffered by the mark itself because of the transmission process can be obtained. It is worth noticing that by means of the described procedure the watermark is affected just by channel errors, and therefore the watermark corruption can be used for obtaining an assessment of the received audio signal quality after an MP3 coding/decoding process and transmission.

4. Employed QoS index

Quality of the communication

In this paragraph the way to effective obtain a coherent quality assessment of an audio signal after that it has been MP3 coded/decoded, transmitted and received (and this without affecting in any way the communication quality) it is reported in details. At first an objective metric to rate the quality of service an objective such as the Mean Square Error (MSE) is used, whereas to give a subjective assessment the PEAQ is employed.

A number of different metrics could be used to evaluate the quality of a multimedia communication (see for examples [20]). For giving an intuitive and immediate proof of concept as regards to the innovative watermarking system proposed herein, the MSE between the extracted watermark and the original one has been referred to. By considering equation (4), the following relation holds

$$MSE_i = \frac{1}{K^2}\sum_{k=1}^{K}(w[k]-\hat{w}[k])^2. \tag{5}$$

Here MSE_i represents the *MSE of the i*-th frame. The various values obtained for each frame taken into consideration are averaged so that the single generic index used can be written as:

$$MSE = \frac{1}{M}\sum_{i=1}^{M} MSE_i. \tag{6}$$

It is important noticing that the value given by solving (6) and obtained using the extracted mark from the *M* transmitted blocks is the one used to provide an assessment of the overall quality for what regards the audio signal received after an MP3 coding/decoding process.

Quality of the watermark embedment

A concept which has been already mentioned as being fundamental when it comes to the digital audio coding field is that the mark embedding it is supposed to not create any perceivable corruptions (i.e. distortions while listening) in comparison to the original signal. The Human

Auditory System (HAS) is in fact much more sensitive to artefacts and sudden changes in the expected signal envelope compared to the sight. The *perceptual quality* is exactly referred to the non audibility of the mark once embedded in the host audio signal. An assessment about this crucial concern can be given by means of repetitive and systematic subjective tests, requiring nonetheless anechoic rooms and multiple subjects of every sex and age (as known the human hear frequency capabilities sink with the increase of the age and differs between male and female). Finally, different trials which could simulate the various listening conditions would have to be performed and the use of different musical genres should be as well taken into consideration. This would be obviously a demanding and challenging procedure and apparatus to be realized in a scientific way. For this reason the PEAQ (Perceptual Evaluation of Audio Quality) has been chosen, having proved extensively its reliability for that task. This is the ITU standard for the objective assessment of audio quality (Standard for Objective Measurement of Audio Quality - ITU BS-1387) [14]. It consists of a set of recommendations on how to optimally implement by software the human acoustic experience. In this way it becomes possible to compare the subjective perception of two similar musical files, so as to objectively measure their difference (or similarity) degree. It is usually employed to measure the fidelity of audio coders or audio broadcastings in real time. Within the standard 2 versions are described: the Basic one uses an FFT based (Fast Fourier Transform) human hear model and it is tailored for real-time calculation. The Advanced one is based on a hear filter bank model. The Model Output Variables (MOVs, 9 in the Basic version, 5 in the Advanced one) are evaluated while comparing the reference audio signal to the one under test. Afterwards, the MOVs become the input of a neural net built and trained in such a way to simulate the psychical process regulating the HAS acoustic experience. In the Basic version this neural net has 11 input nodes, one hidden level with 3 nodes and a single output, while the advanced version has 5 input nodes, one hidden level with a single output. The MOVs are processed by the neural net so as to obtain a single measurement, the so-called Objective Difference Grade (ODG). The latter allows to quantify the overall differences between the two signals as what regards how much they are really differently perceived by the HAS. The ODG is finally related to a merit scale provided by the standard itself (and represented in Table 1) so that it becomes possible to assess the alleged not perceptibility of the mark embedded within a host audio signal.

5. Results

With the aim of giving a concrete proof of the hypothesis formulated in the previous chapters, an extensive number of experimental tests have been carried out. The effective capacity of the mark in tracing the degradations of the audio signal -without affecting the audio quality by means of the embedding procedure- following a coding and transmission process had to be demonstrated.

During the experimentations, the mark has been embedded into the audio signal using the procedure described in paragraph 3. Considering the Nyquist-Shannon sampling theorem and the perceivable frequency range of the HAS, only audio files sampled at the frequency of 44.1 KHz e 48 KHz have been considered. A low payload file (for example a binary sequence as a bitmap logo) has been employed and embedded in every frame (as already said the frames are 2 seconds long each one) modulated by the parameter α (the watermark *robustness*). As a trade-off

between too large α values (which would distort the audio signal while making the watermarking robust, i.e. useless for the scope proposed) and too short ones (making the watermark easily removable) a value $\alpha = 0.04$ has been chosen.

The tests have considered the whole range of musical genres and compression rates, with preference given to the most common and used sampling rates which are 64, 96, 128, 160, and 192 kbps.

Besides of that, systematic simulations have been carried to objectively demonstrate that the simulation parameter just described are really effective in embedding the watermark using the proposed insertion procedure in a transparent way; this means that the embedded watermark is actually non audible. In addition to the fact that all the experimental subjects having to choose between the original music file and its watermarked version have not been able to recognize which was the non marked signal, reliable objective results proving this point have been given through the PEAQ. Those values are represented in Table 2; it easy to note that, adopting as explained the value α=0.04, differences in the envelopes of original signal and watermarked one are not perceivable by the HAS. The values –obtained as average of a number of repeated trials- are in fact to be related to the merit scale provided by the standard itself and shown in Table 1.

Objective Difference Grade	Description of Impairments
0	Imperceptible
-1	Perceptible but not annoying
-2	Slightly annoying
-3	Annoying
-4	Very annoying

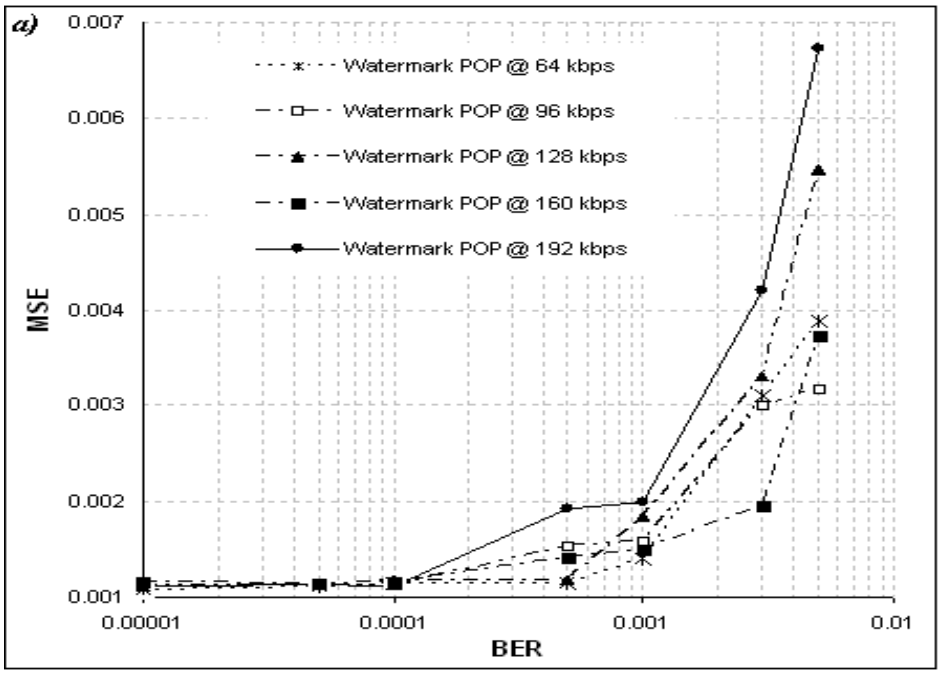
***Tab. 1**: Scale of merit of the Objective Difference Grade in the PEAQ standard*

Genre	ODG
Pop	-0,166
New Wave	-0,167
Soul	-0,197
Rock	-0,165
Roots Reggae	-0,173
Classic	-0,156
Hip-hop	-0,142
Cross-over	-0,169
Techno	-0,157

***Tab. 2**: Objective Difference Grade for different music genres. The differences between the original signals and the watermarked ones tend to be completely imperceptible.*

The following results on the other hand show that the presented digital fragile watermarking technique can be used to provide a reliable measure of the signal quality after an MP3 coding and transmission process, i.e. the starting hypothesis. Following the experimental procedure, the bit stream is sent on a noisy transmission channel simulated by means of a Poisson random transmission error generator. This introduces a Bit Error Rate (BER) having value in the range from 10^{-5} to $5 \cdot 10^{-3}$. In Fig. 2 and 3 the extracted watermark MSE with respect to the original one is considered, and it is related to the BER introduced within audio file of various musical genres. The results from all those multiple genres (e.g. classical music, pop, reggae, swing) at multiple compression rates (from 64 to 192 Kbps) are reported. It is worth noticing that the MSE of the extracted watermark increases as the channel BER increases and as the bit rate decreases. This is coherent to the initial hypothesis that the MSE in the watermark reflects the degradation of the host audio signal, due to the increase of the errors introduced by the channel during transmission and to lower sampling rates. Moreover, as depicted by Fig. 4 and 5, the corruption of the embedded watermark quality has the same behaviour (in terms of values trend) with respect to the audio quality degradation. As easily noticed by the graphs, the extracted watermark MSE is found to be strictly linked to the amount of error introduced within the audio bit stream: the increase in the BER value is followed by a proportional rising in the MSE of the watermark, which is in turn proportional to the rising in the MSE of the audio content.

The numerical results provided in the following show that is possible to obtain a blind quality assessment for the transmission and coding processes. By means of the depicted watermarking system, without having any previous clue as regards the original audio signal quality (i.e. before watermark embedding and MP3 coding), it is thus possible on the receiver side to calculate a quality estimate of the transmission service without corrupting the audio fidelity of the transmitted musical data. Notably, thanks to the agile extraction and MSE calculation procedure, this information can be available real time with common computational power; the average overall calculation time being proportional and comparable to the duration of the marked sound wave. This means that details about the effective transmission efficiency can be immediately transmitted as a feedback to the sender of the watermark contents, so that various optional services in the field of the multimedia mobile communications (as for example third generation mobile phone communications -UMTS- and even the future fourth generation, or internet WI-FI IEEE 802.11 technologies) can be added to the data transmission provided. Those new uses of the watermarking can be found in the employment of those feedback data by the service provider for a more rational management of the resources provided, so to focus on weak or faulty links more than on users having large QoS values. In another scenario this technique can be used by the service provider as an immediate or delayed monitor for billing purposes so to charge adequately the multiple users referring to the effective measured Quality of Service offered.



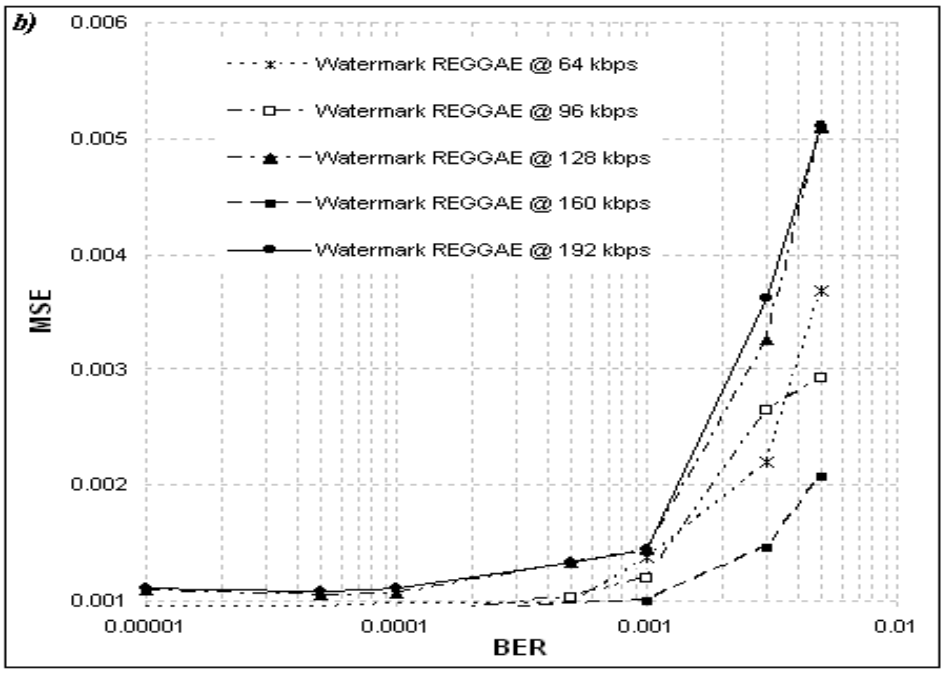
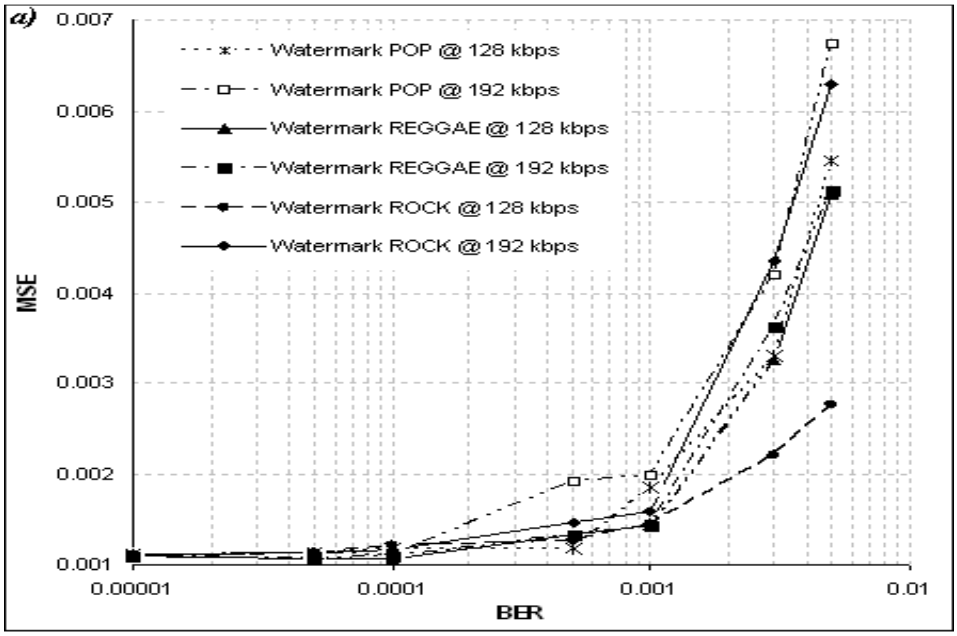


Fig. 2: *MSE of the watermark extracted from MP3 audio signals at different compression ratios: (a) POP music and (b) REGGAE music versus the BER of the channel for different compression ratios of the MP3 coder.*



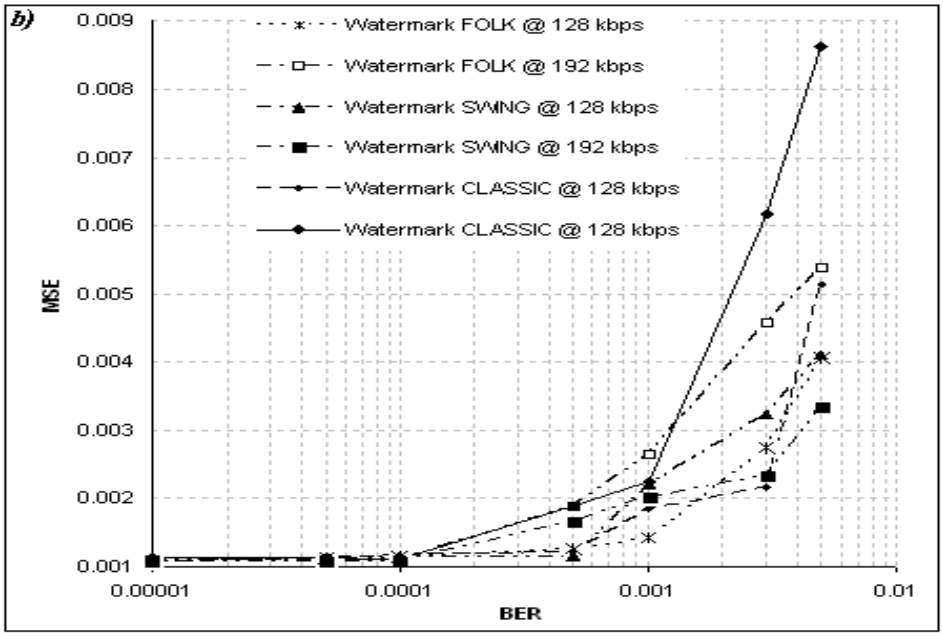


Fig. 3: *MSE of the watermark extracted from MP3 audio signals of different genres: (a) POP, REGGAE and ROCK music; (b) FOLK, SWING and CLASSIC music versus the BER of the channel for different compression ratios (128 and 192 kbps) of the MP3 coder.*

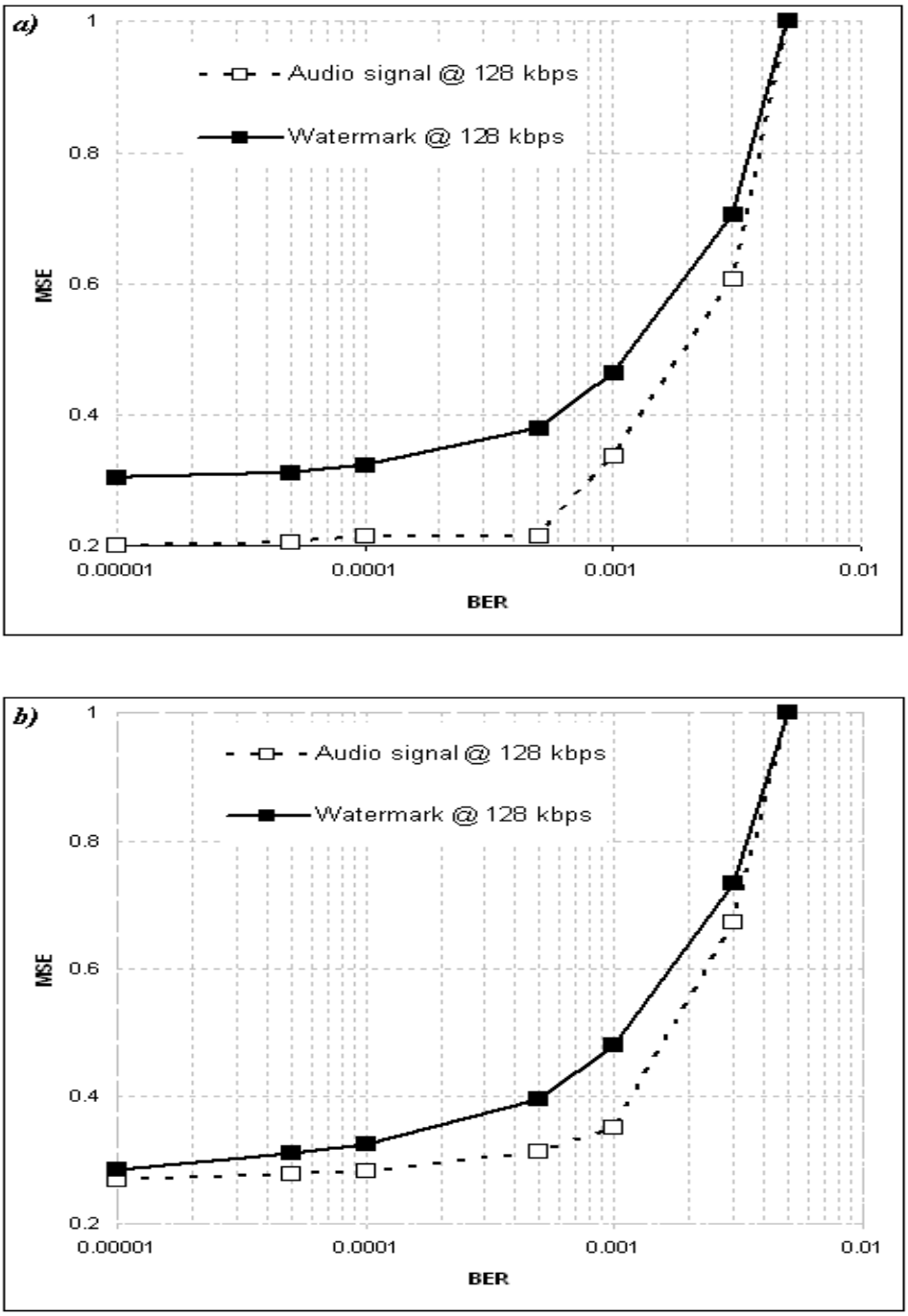


Fig. 4: *MSE (normalized to 1) of both the watermarked MP3 signal and the extracted watermark: (a) RAGGAE music; (b) CLASSIC music, evaluated at a compression ratio of 128 kbps.*

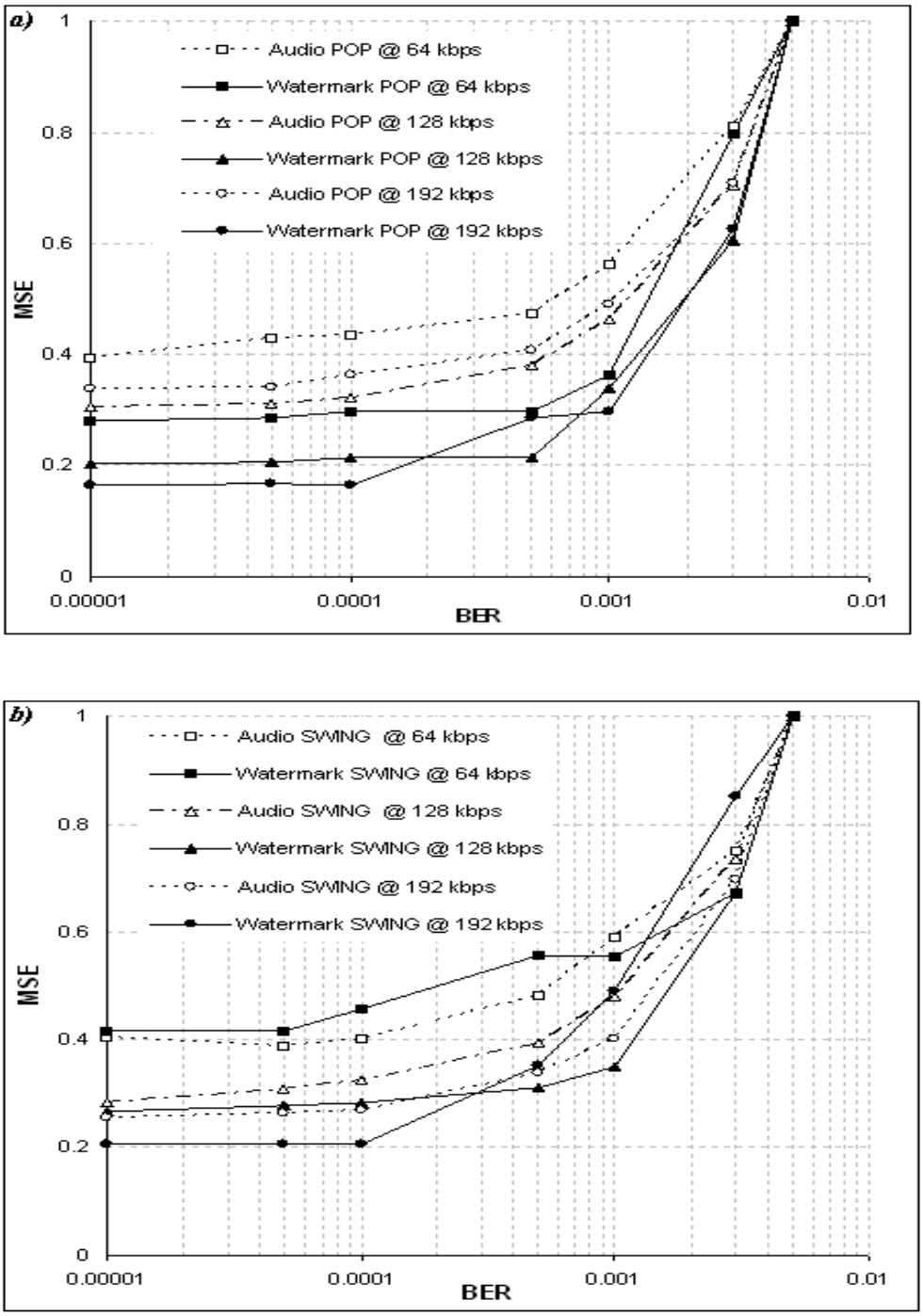


Fig. 5. *MSE (normalized to 1) of both the watermarked MP3 signal and the extracted watermark: (a) POP music; (b) swing music, evaluated at a compression ratio of 64, 128 and 192 kbps.*

6. Conclusions

In this chapter an application of audio digital signal processing techniques to the research field of digital watermarking has been described. The proposed audio watermarking technique has been proved to allow getting a blind assessment referred to the audio quality of musical signals received after being compressed in MP3 format, and transmitted on a normal (i.e. not ideal), noisy channel. A fragile watermark is embedded into an audio uncompressed data stream by means of a spread spectrum approach. The alterations which affect the audio signal during the coding and transmission are supposed to be proportional to the ones the watermark suffered since they are both transmitted on the same communication channel. The watermark degradation can be therefore used to assess the decrease of audio quality to which the whole musical file has undergone. At the receiving side, the watermark is extracted and compared to the original counterpart. The QoS service assessment is performed using both objective and subjective measures: the Mean Square Error between the original reference watermark and the transmitted extracted one has been considered for getting a degree of the modifications caused by the coding and transmission. The analysis about the perceived audio quality has been carried out following the recent ITU standard on the perceptual audio experience assessment that is the PEAQ (ITU-R BS.1387).

The result of the performed extensive simulations confirms the approach effectiveness in providing a coding/transmission Quality of Service estimate within audio communications with no affection of the audio quality caused by the watermarking technique itself.

Further developments of the idea are forecasted, comprising the joint consideration of multiple uses for the watermark. As an example the watermark could consists of info about the host audio contents (e.g. for archive purposing, fingerprinting or even decryption keys in case the communication needs to be secured).

Moreover the question is open if the present system could be used also for spoken communication. In this case, consideration in the particular envelope and limited bandwidth of the speech signal must be taken. Within this scenario, the two communicating side could be mutually sending each other information about the communication link real time.

IV

Audio Restoration

This chapter introduces another application of digital signal processing techniques within the field of digital music. In this case the problem are still corrupted samples (even a complete sequence of them) which could be again generated by a noisy transmission or most likely by a conversion analogical to digital of old recordings (magnetic tapes or vinyl records). The latter as known, differently from their digital successor versions (which for this and other obvious advantages took quickly their place in the last decades), were and still are prone to degradations of the audio quality. The loss of entire sample chunks is in those analogical formats is definitely not so unusual (as experienced by everyone who used them). Starting from this problem, techniques for audio restoration, using interpolation of known samples within the envelope of a digital audio record, come into use to retrieve or at least to assess the value of the missing samples.

The research topic which follows has been developed at the Signal Processing Lab of University of Cambridge (the supervision and support of Dr. Taylan Cemgil and of Professor William Fitzgerald is credited). In this Lab the Audio Restoration problem has been researched during the last years obtaining solid results [21]. In a field like the audio and music one on the other hand, where high fidelity is a serious matter, improvements are always requested in order to obtain the best reconstructed quality.

For this reason the audio restoration is a very active research field where definitive and optimal results are still to be achieved. In this chapter, following the general Bayesian approach to DSP problems adopted within the Signal Processing Lab, the implementation of a probabilistic phase Vocoder will be depicted.

1. Motivations

Degradations of audio sources are generally defined as each and every undesired modification of the audio signal resulting from (or as a consequence of) a registration or coding process. The restoration of the audio contents consists therefore of multiple techniques for the reconstruction of the original audio source starting from the form it has been received in -or output by- the transduction coding instruments (for example a microphone or the input of an A/D converter). The modern digital Audio Restoration methods allow achieving a better degree of freedom compared to the analogical precursor. To work properly those digital devices need nonetheless a wide knowledge of the audio processing and experience in the field in order to avoid undesired drawbacks (i.e. a further degradation of the treated audio file). To the aim of restoring the original audio quality, the first works in the digital domain used the simple convolution technique for reinforcing the solo voice within a musical track (refer for examples to [22] and [23]).
The various kinds of audio source degradations can be broadly classified within the following groups: 1) local degradations (e.g. the so called clicks or the low frequency noise transients) and, 2) global degradations affecting all the samples along an audio signal (for example broad band noise, perceived normally as a background hiss, or defects caused by the pitch variations and distortions, i.e. a wide class of nonlinear defects).
Despite, the commonly listened music is nowadays almost exclusively digital (after the Compact Disk diffusion we are now in the MP3 era), even these new formats can present some missing samples issues: letting alone the concerns deriving from corrupted analogical track leading to problematic digitalisations, this problem can affect even digital audio contents (although techniques like Error Correcting Codes, Cyclic Redundancy Check -CRC-, or Channel Coding are now able to avoid part of those issues). Just imagine CD with scratches or audio files wrongly downloaded, i.e. missing some chunks of musical contents. Even in such cases a time slot within the envelope of the musical signal is completely unknown (i.e. the known samples are anyway to be considered as independent from the missing ones). If the original content is unavailable for a second coding (differently from what happens with the noise superimposing on the musical information which is additive to it) the information payload brought by the missing samples is completely void. On the other hand it can be easily retrieved the pointer to the sample where the "dark" window starts and ends (see [24] and [25]). As regards this, the audio restoration consists of interpolation techniques for assessing missing samples by means of the info provided by the immediately previous and next known series of samples.

Using a probabilistic-Bayesian approach, the just described problem can be formulated very generically in the following form:

$$p(x_{-k}|x_k) \propto \int d\mathrm{H} p(x_{-k}|\mathrm{H}) p(x_k|\mathrm{H}) p(\mathrm{H}) \tag{7}$$

which implies that the learning of the information related to the unknown samples x_{-k}, being the uncorrupted samples x_k available, depends on the unknown parameters set *H* of the model Θ and on the other unobserved state variables *S* describing the sound generation mechanism

$p(x_{0:K-1}, \mathrm{H})$. Here it is assumed that $x_{0:K-1} = x_k \cup x_{-k}$ starting from the assumption that dealing with audio signal it is licit/coherent assuming that the same mechanisms regulates the generation of both missing and available samples.
Following the idea of T. Cemgil and S. Godsill [26] this probabilistic formulation is used together with the concept of the probabilistic Phase Vocoder a generative model for the audio signal so that the problem of the missing samples reconstruction can be treated in a fully probabilistic way.

2. Theory background

In this paragraph, the basic concepts regarding the developed idea are reported in a rather agile but clear way. They consist of the famous Bayes' law and the theoretical notions needed for the implementations of a phase vocoder.

Bayesian Approach
Given two events *A* and *B* (with no null probability) it is known that the probability of the event *B* conditional to the event *A* is $P(B|A) = \frac{P(B \cap A)}{P(A)}$, while the probability of *A* conditional to *B* is $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Since obviously $P(A \cap B) = P(B \cap A)$, by means of elementary replacing, the notable Bayes' law is obtained:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{8}$$

This equation, being at the same time immediate, is fundamental for all those assessment procedures because it tells how to proceed from the probabilistic description of an event B, which has been directly observed, to an unobserved event A, the probability of which is not known and needed. *P(A)* is the so-called *prior probability*, reflecting what is known of A previous of the observation of the even B. $P(A|B)$ is called *posterior probability*, since it represents the probability of A after the event B has been observed. $P(B|A)$ is known as *likelihood* while *P(B)* is the total or *marginal probability* of *B*.
The total probability can be found partitioning the sampling space $\Omega = \{A_1, A_2, ..., A_n\}$, which means that a group of mutual excluding events (i.e. where there is no overlapping among the partition members) are found and their union represents the whole sampling space. The total probability is therefore given by the following equation

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) \tag{9}$$

Conditional distributions together with the Bayes' law are fundamental to the assessment process in general terms and in particular to the analysis of complex random systems, as the one here depicted, where the technical problem assumes the shape of the question "what is it really possible to assess about the parameters *b* of the given system, having the observation *a* available"?

Phase Vocoder

The phase vocoder -contraction of the words voice and coder- was described for the first time in 1966 [27] as a number of coding techniques suitable for the speech signals in order to get vocal signals having low bit rate. It was anyway just one decade later that its outstanding efficiency in processing the musical signal was noticed. In particular it shows unexpected good properties for temporal scale modifications and pitch transpositions. The phase vocoder, in fact, has the characteristic of allowing arbitrary control of the single harmonics of which the sound signal is composed.

The phase vocoder can be after all classified as a mere technique for the analysis and synthesis of musical signals. The basic hypothesis is that the input signal can be represented by a model, the parameters of which are time variant. The analysis is devoted to the determination of the parameters of the signal to be processed, while the synthesis simply consists of the model output itself. Moreover, the parameters which are obtained by means of the synthesis can be modified in order to allow countless elaborations starting from the original signal. In the phase vocoder the signal is modelled as a sin-wave sum and the parameters to be determined are the time varying amplitudes and the frequencies of each sin-wave. This model is very well suitable to be used with a large amount of musical signals since the aforementioned sin-waves do not necessarily need to be harmonically correlated.

Two complementary interpretations of the phase vocoder can be considered. Those are the one which uses a filter bank and the one obtained by means of the Fourier Transform. The latter is here disclosed, since it is apparently more suitable and meaningful for the application on which next paragraphs are focused. Following this interpretation, the Phase Vocoder analysis phase turns into a series of overlapping Fourier Transforms taken on a time slot having finite extension. In the filter bank interpretation the focus is given to the temporal succession of the amplitude and phase values within the single filter bands. In the Fourier Transform interpretation, on the other hand, the emphasis resides in the phase and amplitude values of each sin wave calculated at a single point in time.

The synthesis is then performed by means of the inverse conversion (to the form having real and imaginary part) followed by the overlapping and summing of the following inverse Fourier Transform. The amount of filter bands used is simply the number of bins within the Fourier Transform. The big advantage in this interpretation compared to the filter bank one is that the filter bank calculation can be performed using the Fast Fourier Transform (the FFT being a much more computational efficient technique).

The phase vocoder basic aim is separating temporal and spectral information in the best way possible. The operative approach consists of dividing the signal in a number of spectral bands and characterizing the time-varying signal within each band. This does not give the expected results if the signal oscillates too quickly within one single band, which means if the amplitude and frequency are not relatively constant on the FFT interval. If this condition can be avoided,

the phase vocoder can be effectively used for several applications ranging from the tone analysis in musical instruments, the determination of the single components time -varying amplitudes and frequencies, and in general the possibility of modifying and process audio content. In particular, the temporal envelope of a sound can be decelerated without modifying its pitch. This is obtained spacing more the inverse FFT with respect to the FFT obtained by the analysis stage so that the spectral variations take place in the synthesized sound more slowly than in the original. The phase is scaled of the same factor, so that each band undergoes to the same frequency variation. The inverse operation too is obviously expected to be possible, that is the pitch tuning without changing the time duration of the sound. For obtaining this result, a time scaling of the same amount of the frequency one is performed and then the sound is played while sampling at the rate corresponding to the pitch modification.

3. Implementation

Probabilistic Phase Vocoder as a generative model

In [26] the limited performances of the normal phase vocoder are highlighted. In particular a serious border is placed by the fact that in the original formulation the sound signal is supposed to have characteristics which match well the sin-waves model. An explication of the model underlying the signal to be reproduced is therefore proposed. This is done in order to make the original algorithm more applicable. In detail, the Discrete Fourier Transform (DFT) is employed as a generative process for the acoustic signal so to realize a probabilistic phase vocoder

Given a sequence $x \equiv (x_{0,} x_{1,} ..., x_{k,} ...., x_{K-1})^T$, its Fourier Transform $s \equiv (s_{0,} s_{1,} ..., s_{k,} ..., s_{W-1})^T$ is given by

$$s = Fx \tag{10}$$

Here $F = \{F_\nu^k\}$ is the DFT matrix with elements $F_\nu^k = e^{-2\pi j \nu k / K}$. If the transformation matrix F is squared (which means that the time and frequency indexes are equal, i.e. $K=W$), the mapping between time and frequency domains are revertible. In this way, the inverse DFT for the signal reconstruction is obtained with the following equation

$$x = F^H s \tag{11}$$

Here H represents the transposed Hermitian and the elements of $F^H = \{F_k^{*\nu}\}$ are $F_k^{*\nu} = e^{2\pi j \nu k / K}$. Those elements can therefore be recursively generalized writing

$$F_k^{*\nu} = e^{j\omega\nu} F_{k-1}^{*\nu} \tag{12}$$

Here $\omega = \frac{2\pi}{W}$ and $F_0^{*\nu} = 1$, while the k-th sample in x is

$$x_k = \sum_{\nu} s_k^{\nu} \equiv \sum_{\nu} F_k^{*\nu} s^{\nu} \equiv \sum_{\nu} B(\omega\nu) s_k^{\nu} \tag{13}$$

Where $B(\theta)$ represents the Givens rotation matrix, that is

$$B(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{14}$$

Considering *x* as real (this hypothesis can be considered as coherent as long as the signal are audio and single channel) we can write the deterministic transition and observation equations, which allows calculating the inverse Fourier Transform, as

$$\begin{aligned} s_k &= As_{k-1} \\ x_k &= x_{\Re,k} = Cs_k \end{aligned} \tag{15}$$

where the observation matrix *C*, is defined as $C \equiv \begin{pmatrix} 1 & 2 & 0 & 2 & 0 & \dots & 2 & 0 & 1 \end{pmatrix}$

whereas *A*, the transition matrix is defined as $A \equiv blkdiag\left\{ B(0), B(\omega), \dots, B(\nu\omega), \dots, B(\frac{W}{2}\omega) \right\}$.

Finally, the stochastic version of this state space is defined, obtaining in this way the equations regulating the Probabilistic Phase Vocoder (PPVOC)

$$\begin{aligned} s_k \left| s_{k-1} \right. &\sim \mathcal{N}\left(s_k; As_{k-1}, Q\right) \\ x_k \left| s_k \right. &\sim \mathcal{N}\left(x_k; Cs_k, R\right) \\ s_0 &\sim \mathcal{N}\left(s_0; 0, P\right) \end{aligned} \tag{16}$$

Here $\mathcal{N}(\mathrm{x}, \mu, \Sigma)$ is a Gaussian distribution on the overall data with index *x*, mean μ and covariance matrix Σ.

Assessment by means of the Expectation Maximization

The assessment of the missing samples (i.e. the inference, which in this particular case implies the recursive reconstruction of the musical signal by means of probabilistic calculations [28]) requires the calculation of the posterior distribution to be performed, that is

$$p(S,\Theta|x_k) = \frac{1}{Z_k} p(x_k|S,\Theta)\, p(S|\Theta)\, p(\Theta) \equiv \frac{1}{Z_k}\phi(S,\Theta) \tag{17}$$

(where $S = s_{0:K-1}^{0:W-1}, \Theta = (A,Q)$ and $Z_x = p(x_k)$ is a normalization constant) as well as the predictive distribution

$$p(x_{-k}, x_k) = \int dSd\Theta p(x_{-k}|S,\Theta)\, p(S,\Theta|x_k)\,. \tag{18}$$

The problem resides in the fact the posterior distribution is hard to be computed in a precise way because of the mutual dependence between Θ and *S*.
For this reason the *Mean Field* approximation is employed in combination with the Expectation Maximisation algorithm (*EM* algorithm).

4. Missing samples computation: Kalman filter and Smoothing

The Phase Vocoder in the model typology disclosed in the previous paragraphs can be related to the factorial model of the Kalman Filter [29]. Starting from this assumption such a hybrid model has been implemented in Matlab (©, see Annex A for the source code).
In detail, the loss of consecutive samples are simulated, creating gaps within the original audio signal envelope so that this pre-system output is a music signal having some samples carrying no information (i.e. the no knowledge of the original signal is considered).

First the calculation of the mean values and covariance for each time bin at each iteration *n* is performed. From that, at the every iteration end, updated generation and observation matrices for the processed digital audio signal are calculated together with the relative noise matrices (generation noise and measurement noise). This is done by means of the expected value maximization of the *log joint likelihood* given by the equation which follows, which depends on a number of parameters, namely the mean of the *distribution* μ, the covariance *matrix* Σ, the *observation matrix C*, the *covariance matrix of the measure/observation error Q*, and the *forecasting error covariance matrix R*:

$$\log L \equiv -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(s_0-\mu)'\Sigma^{-1}(s_0-\mu) - \frac{n}{2}\log|Q| - \frac{1}{2}\sum_{t=1}^{n}(s_k - As_{k-1})'Q^{-1}(s_k - As_{k-1}) + \tag{19}$$
$$-\frac{n}{2}\log|R| - \frac{1}{2}\sum_{t=1}^{n}(x_k - Cs_k)'R^{-1}(x_k - Cs_k)$$

The *conditional expected value* can be at its turn defined as

$$x_k^s = E(x_k | s_1,...,s_k) \tag{20}$$

and the *covariance functions* respectively as

$$P_k^s = \mathrm{cov}(x_k | s_1,...,s_k) \tag{21}$$

and

$$P_{k,k-1}^s = \mathrm{cov}(x_k, x_{k-1} | s_1,...,s_k) \tag{22}$$

At first, the forecasting and updating algorithm is used (relative to the Kalman filter cited above), which means that a direct (*forward*) process is performed: the missing samples creating the information gap are estimated by means of the samples preceding them (and the information therein retrievable) [30]

$$x_k^{k-1} = A x_{k-1}^{k-1} \tag{23}$$

$$P_k^{k-1} = A_k P_{k-1}^{k-1} A_k^T + Q_t \tag{24}$$

$$DD_k = P_k^{k-1} C_k^T (C_k P_k^{k-1} C_k^T + R_k)^{-1} \tag{25}$$

$$x_k^k = x_k^{k-1} + DD_k (s_k - C_k x_k^{k-1}) \tag{26}$$

$$P_k^k = P_k^{k-1} - DD_k C_k P_k^{k-1} \tag{27}$$

here the following initial estimations are supposed as being $x_0^0 = \mu$ e $P_0^0 = \Sigma$.

After the performing of this step, the signal undergoes to a *smoothing* process (by using the *Rauch-Tung-Striebel* [31] equations), i.e. a backward process: the missing samples are once again estimated starting from the temporally last one coming in time, based on the values obtained from the statistical analysis of the samples which precedes in this backward time scale them, i.e. the one following along the usual direct time scale (in this case the sample index is going obviously to be *k = K, K-1, ..., 1*)g

$$J_{k-1} = P_{k-1}^{k-1} A_k^T (P_k^{k-1})^{-1} \tag{28}$$

$$x_{k-1}^K = x_{k-1}^{k-1} + J_{k-1} (x_k^K - A_k x_{k-1}^{k-1}) \tag{29}$$

$$P_{k-1}^K = P_{k-1}^{k-1} - J_{k-1} (P_k^K - P_k^{k-1}) J_{k-1}^T \tag{30}$$

and

$$P_{k-1,k-2}^{K} = P_{k-1}^{k-1} J_{k-2}^{T} + J_{k-1}(P_{k,k-1}^{K} - A_k P_{k-1}^{k-1}) J_{k-2}^{T} \tag{31}$$

where $P_{K,k-1}^{K} = (I - K_K M_K) A_k P_{k-1}^{k-1}$.
If we calculate the following fictive quantities for each iteration

$$U = \sum_{k=1}^{K} \left(P_{k-1}^{K} + x_{k-1}^{K} \left(x_{k-1}^{K} \right)^T \right) \tag{32}$$

$$V = \sum_{k=1}^{K} \left(P_{k,k-1}^{K} + x_{k}^{K} \left(x_{k-1}^{K} \right)^T \right) \tag{33}$$

$$Z = \sum_{k=1}^{K} \left(P_{k,k-1}^{K} + x_{k}^{K} \left(x_{k-1}^{K} \right)^T \right) \tag{34}$$

it is obtained that for each iteration the maximum value (i.e. the best one) in (19) is given by

$$A(k+1) = VU^{-1} \tag{35}$$

$$Q(k+1) = K^{-1} \left(Z - VU^{-1} V^T \right) \tag{36}$$

and

$$R(k+1) = K^{-1} \sum_{k=1}^{K} \left[\left(s_k - C_k x_k^K \right) \left(s_k - C_k x_k^K \right)^T + M_k P_k^K M_k^T \right] \tag{37}$$

By means of the signal mean values and respective covariance calculation for each iteration *n* and each time bin, maximizing the expected likelihood value in (19), the optimal μ, Σ, C, Q, and R, are obtained. Those values allow the best reconstruction of the missing sample within the signal.

5. Results and conclusions

The implementation I have realized (again, refer to Annex A for the source code) gives the opportunity, although in a pretty immediate and rather simplistic way, to verify and appreciate the correctness of the theoretical hypothesis and the efficiency of the algorithm. As a measure of the implementation performances, the intuitive objective estimate given by the Mean Square Error has been employed once again. The comparison has been considered between all the samples of the original defects free signal and its corrupted version, in which gaps had been created in order to simulate lost samples. Of the overall samples, an amount between 10 and 20%

has been removed. The gaps created have each one a maximal extension up to 4% of the musical track total duration.
Those data-sets and respective experimental parameters, although quite limited, have confirmed that the proposed method can be considered as reliable for the restoring of corrupted tracks. The MSE values for the restored signal compared to the original one (without lost samples) has always resulted relatively small when compared to the amount of missing samples. In fact MSE values as small as being in the range between 10^{-2} and 10^{-4} have been obtained.

In the following graph (Figure 6, [26]), as an example of the algorithm efficiency, a musical signal before and after undergoing the depicted audio restoration processing for the missing samples retrieval is shown (original samples are in the black dotted line, whereas the reconstructed ones are represented by the red dotted one).

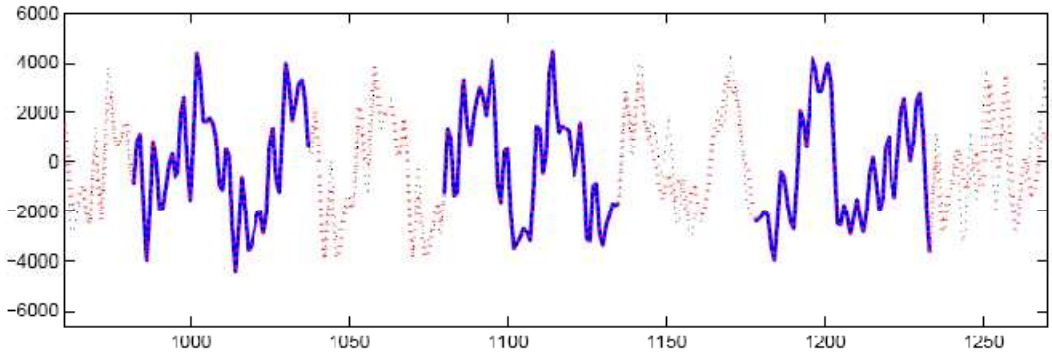


Fig. 6: *example of an audio signal restored with the disclosed technique. On the x axe there are the time bins while on the y one there are the amplitudes of the signal.*

V

Audio in VR: 3D Sound in Virtual Scenarios

This and the following chapter will depict the experimental developments achieved within the field of Acoustics at the Fraunhofer Institut for Factory Logistics and Automation (Fraunhofer Institut fuer Fabrikbetrieb und –automatisierung, IFF) during the period spanning from August 2006 to August 2008.
The Fraunhofer IFF is part of the "Fraunhofer Gesellschaft", the German national institution for applied research (it is estabilished with 56 instituts in 40 different cities in Germany), developing novel technical products and solutions for private and public enterprises. It generates almost the 70% of its research budget by means of contracts with companies or public/private research projects.
The Fraunhofer Institut Fabrikbetrieb und –automatisierung is located in Magdeburg and since 2006 counts among its facilities the newly built and futuristic Virtual Development and Training Centre (VDTC) where innovative Virtual Reality techniques -in general industrial planning and interactive training or interactive learning oriented- can be developed, combined and tested.
In particular the following paragraphs will describe the researches combining audio signal processing and virtual rendering of fictive scenarios. In this chapter, starting from the general issues rising from the three-dimensional (3D) rendition of general digital audio, the implementation and experimentation of the core function dedicated to the virtual audio rendering within the VDTC Virtual Reality engine are disclosed.
In the domains of the emerging VR field, audio DSP is winning an ever growing importance contribuiting to the aim of making the virtual closer and closer to its real counterpart.

1. Introduction

Representing reality in fictive way has always been a basic aim of the human genre. The ambition of artificially recreating in a convincing way real situations or even rendering in convincing manner unreal and fantasy environment, has brought age by age to the rising and development of the figurative arts. Thanks to the emerging and improving of computers and following their ever growing computational capabilities, this representation can become every day closer to what inspired it in the real world.

One of the earliest Virtual Reality (VR) systems, as an example, was the Sensorama realized by Morton Heilig [32] in the early sixties. This was nevertheless simply a mechanic device targeting to stimulate the five senses. The real ancestor of the modern multimedia system recreating virtual scenarios can be considered the Aspen Movie Map developed at MIT in 1977 [33].

Nowadays, several VR systems are commercially available on the market ranging from the so called "goggles n´ gloves" (interactive glasses and gloves) to driving simulators. Starting from the construction of the Electronic Visualization Laboratory by the University of Illinois at Chicago [34] in 1992, last decades have witnessed the diffusions of an alternative way of virtual rendering, which is the *Cave Automatic Virtual Environment* (the so called *CAVE* [35]). With this new technique the virtual environment can be rendered within a closed room.

Since the very first VR systems, the main sense which was intended to be stimulated in 3 dimensions was obviously the sight. Visual scenarios, having different degree of definition and multidimensionality, had been thus created to give the user the feeling of a real space in which he or she is supposed to be able to interact. Not considering that, it is important to point out that the spatial 3-dimensionality and graphic definition are not the overall quality measure for a VR representation. This is instead defined by means of the *immersivity degree*, that is a highly subjective index referring to the VR system capability in giving the user the impression that the non-existing is actually there, i.e. real.

Investigations focused on using audio clues for the improvement of this immersivity and therefore investigations on 3D sound in general, started relatively late, for the reasons stated in the introduction chapter which also delayed the whole development of audio DSP. The audio Digital Signal Processing indeed (together with fundamentals of physics and computer science) is the field which is related the most to the realization of a realistic audio experience reproduction considering the spatial structures interaction, called room acoustics.

One of the first system fully dedicated to 3D audio was e famous Convolvotron [36] developed in the eighties by the NASA Labs while the advent of new and more computationally powerful chips (as the aforementioned Motorola 56001) gave consequently the possibility to collocate the audio processing directly on the computing engine sound card. This advancement allowed the rising of several researches investigating the physical reproduction (e.g. by mean of headphones or loudspeakers) of 3D Audio (see for example the well known Ambisonic [37] or Wave Field synthesis [38] techniques).

As regards the software interfaces on the other hand, an intuitive implementation, easy to use and regarded as an accepted standard, is still missing. While more and more Universities and research institutions are developing their own VR systems, most of the real and concrete

innovations and pratical inventive contributions in this field, in particular for what concerns the virtual audio, are actually brought by the videogames industry.

2. Motivations

Correctly spatialized audio increases the overall immersivity of a scenario. In fact, hearing a sound coming from the point (or the area) where the sight, or the subconscious, has located (or at least suppose it to be located) the virtual sound source, gives inevitably the feeling the rendering is more realistic. Furthermore, 3D audio can result in being an effective source of information to the user (as for example is normally happening in the airplanes cockpits) or in enhancing speech intelligibility (for example by means of the frequency and time interaural imparities reproduction as it happens in the famous "cocktail party effect" [39]).

In the following paragraphs the *audio core function* implementation targeted to the VR platform developed at the Fraunhofer IFF within the VDTC will be described. It allows allocating in an easy, intuitive but yet performing way, sound sources inside virtual scenarios and manipulating properly the sound diffusion so that the rendering is effectively 3D sounded. In the very next chapter a general overview on the just mentioned Virtual Reality render and in particular on the respective audio extension is given. The audio function and its characteristics will be then extensively described while two practical applications are disclosed. Finally conclusions are drawn together with a review of the future planned developments.

3. The IVS_VDT platform and its audio-extension

The IVS_VDT (acronym meaning Interactive Visualization System Virtual Development Tool), realized within the VDTC, is implemented by means of the OpenSG API (Application Program Interface). OpenSG is a *scene graph* (a general data structure to be used for vector based visual applications) employed for real time graphic programs. It is OpenGL based (the standard API for several operating systems for what concerns 3D graphics) but differs from it since it supports multithreading and clustering (allowing the visualization of a virtual representation on different displays or by means of several beamers, which is often the case for the aforementioned CAVE) [40]. The IVS_VDT employs GTK+ as widget toolkit (a building units set for creating a Graphical User Interface), it is C++ implemented, and works both in Windows and Linux operating systems.

The IVS_VDT audio extension has been implemented as an internal *core function* of the platform itself, using the complementary API to OpenGL, which is OpenAL 1.1. OpenAL operates as a software interface to the audio hardware (like as an example Microsoft DirectSound 3D in Windows XP). It was developed by the Creative Labs, featuring initially Loki Entertainment. Its paradigm describes the necessity defining three basic objects. Those are *User* (unique and always defined), *Buffers*, storing the audio data in a total amount limited by

the RAM memory resources available, and finally sound *Sources*, in an overall amount exclusively limited by the amount of samples which can be played simultaneously by the reproduction device, that is the sound card. Audio samples are periodically loaded into the buffers and are then played taking into consideration the properties and characteristics defined for each source they are referred to and to the User (i.e. the listener) position.

The audio core function has been integrated into the general structure of the IVS_VDT platform which is depicted in Figure 7. By means of the dedicated Authoring Tool, the operation of linking audio data within a buffer to a visual object (referred to as *Parent*) can be performed simply specifying the 3D object name. All the scenario details and newly defined characteristics are finally saved in a text files with the special extension *.tws* (TextPad Workspace file) readable by the IVS_VDT.

For what concerns the actual audio signal processing and consequent spatialization capabilities, the audio core function inherits the basic working scheme from OpenAL [42] (see Figure 8). Firstly the needed buffers are created and the audio samples are loaded. The most common audio data extensions are supported (wave, mp3, Ogg vorbis, IMA ADPCM). Furthermore, the possibility to query the frequency in Hz of the loaded samples, the extension in bytes of the data within the buffer, the amount of bit per sample (8 or 16), and the number of channels characterizing the audio data is given.

After the linking to the buffer, the *author* of the scenario, i.e. the person in charge for authoring its virtual audio effects in this case) can take care of the sound propagation. By means of the function *AudioCreate*, the parameters regulating the sound 3D diffusion from the visual parent source are pinpointed.

Figure 9 shows the *AudioCreate* function windows in the IVS_VDT platform. Besides the parameters for the spatial position (S_x, S_y, S_z) and the velocity (v_{Sx}, v_{Sy}, v_{Sz}) of the source, which are directly borrowed from the parent, various characteristics can be set in a simple way. Among others, the precise area of the 3D parent object from which the sound is generated (a *relative position,* one can think to the mouth as the origin for the voice within a human Parent Object) can be specified so that the sound propagation has also a precise orientation. In addition an internal and external 3D diffusion cone can be created (defining their respective degree of aperture), so that the volume where the sound is to be perceivable can be set. The *max gain* and *min gain* values describe the fluctuations of the sound intensity between the inner cone (inside of which the sound has maximal level) and the transition zone (where it tends to fade), to reach the outer cone (which consists in the threshold area for the perception of the sound). The sound field amplitude modulation is carried out by means of the *Volume* coefficient tuning while the pitch can be shifted inserting a frequency multiplication factor (e.g. a value of 2.0 will determine a doubling of the sound frequency, i.e. an octave increment). Finally the effects of the movement of sound sources and listener on the hearing experience have been taken into account.

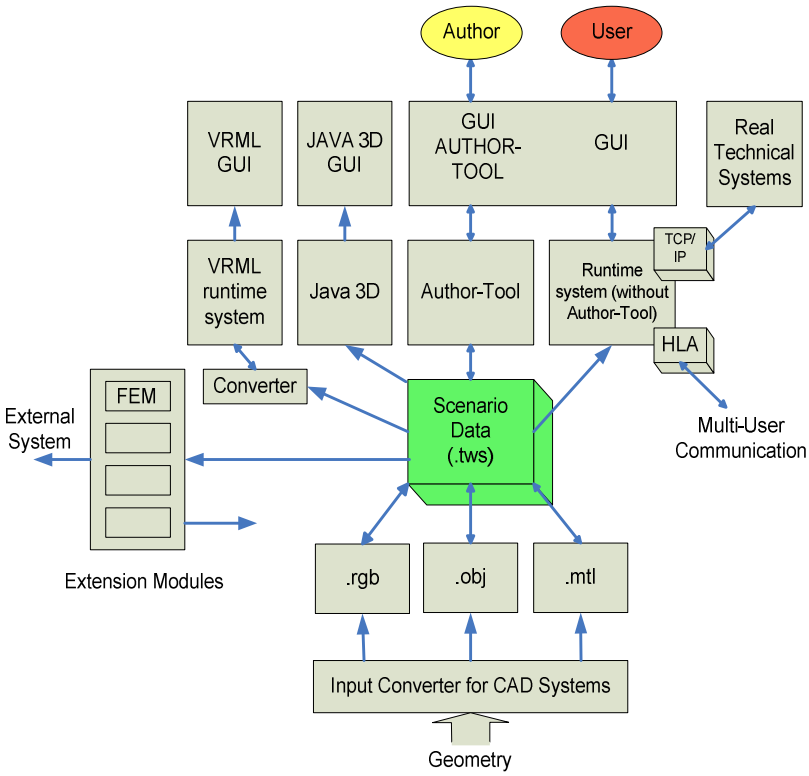


Fig. 7: *IVS_VDT VR platform block scheme structure*

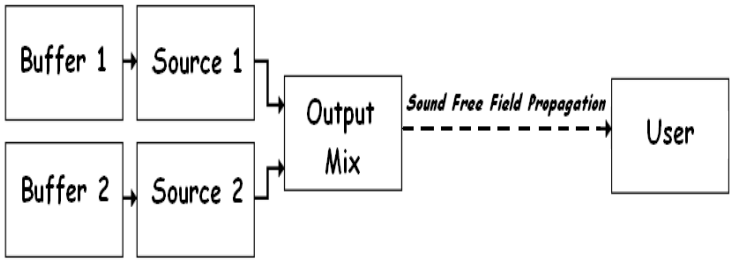


Fig. 8: *OpenAL basic functional scheme*

In fact, the implementation of the *Doppler Effect* allows taking into consideration the real life actual perception of a sound frequency which depends on the mutual respective velocity between sources and hearing subject. Once the user hearing attributes are defined by means of spatial position (L_x, L_y, L_z), velocity (v_x, v_y, v_z) and orientation (e.g. the so called *Azimuth* and *elevation* angles in the vectors representation), the actual frequency shift because of the Doppler Effect is calculated implementing the following relationship [51]:

$$f_D = \frac{f_S\left(v_{Sound} - D_F \cdot V_{LS}\right)}{\left(v_{Sound} - D_F \cdot V_{SS}\right)}, \tag{38}$$

here f_s is the sample frequency for the source in question, v_{sound} is the speed of sound (set to 343.3 m/s in air), D_F is the Doppler factor (usually set to 1, it can be tuned so to boost or bound the effect), while V_{LS} and V_{SS} represent the projections of user velocity and source velocity respectively on the vector connecting them.

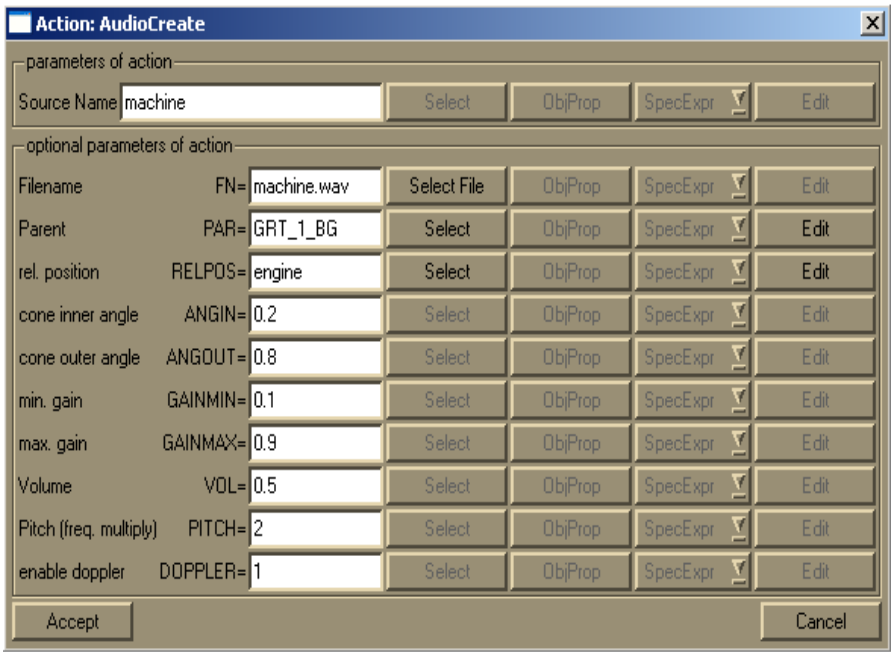


Fig. 9: *AudioCreate function window(© Fraunhofer IFF)*

Yet another basic factor giving the user a natural hearing feeling and depending on the mutual position with regard to the sound source has been considered in the implementation. That is the attenuation of the sound while covering distances in air. Two different models has been implemented which are the linear sound intensity decreasing while the user-source distance increases and the quadratic one (less impressive on the audio effects side but more realistic,

since, as known, the sound amplitude follows the inverse square law with respect to the distance covered by the sound wave).

All the parameters described in the previous lines are updated continously at each visual frame of the rendering, which means the refresh actually depends on the speed of the processor, and the load on the processing device brought by the VR visual counterpart.

This sound emission procedure can be activated, paused, or interrupted by means of the *AudioPlay* action separately for each sound source. Furthermore it is possible to define the amount of times the audio data within the buffers are reproduced every time an AudioPlay action is queried (a -1value would, on the other hand, cause a looped reproduction of the sound).

The so edited audio can be reproduced in a proper 3D manner either using headphones or by means of acoustic loudspeaker (stereo, 5.1 or special configuration such as loudspeakers arrays) since the spatialization is totally performed by the sound card.

4. Realizations, Examples and Results

As already stated, sound within virtual scenarios can achieve several results and realize a number of applications. As a concrete example two VR scenarios are presented in the following in which the sound has the general aim of increasing the overall rendering immersivity.

As a first model, depicted in Figure 10, a scenario consisting of the virtual reproduction of an enormous ship loader machine (located within the real world in the Rotterdam commercial harbor) is described. Since the virtual representation target resides in the tutoring of employees on the machinery control, the realistic and trustworthy rendering of the sounds originated from the real counterpart utilization as a consequence of each working operation, is of great importance. In fact it increases the workers confidence when, once terminated the training, they will use the real machine. Moreover, the spatialized sounds allow the trainee to start learning how to recognize which different operations the machine is performing, without having to visually monitor the moving objects for following them at sight. Finally a number of interaction possibilities being offered by the scenario, the audio responses to the actions performed by the user enhance the training entertaining and dynamism.

The second example is shown in Figure 11 and reproduces a touristic sightseeing tour across Martin Luther's birth town, *Eisleben* (the historical centre of which was designated World Heritage Site by UNESCO in 1997). During the tour, together with the normal realistic sound effects generated by the ordinary sound sources (as for example the bells tolling which comes from the Cathedral tower), the user is given the opportunity to hear different characteristic music coming from each notable historical building or house while simply passing virtually in their proximity. Each musical diffusion interacts in this way with the user, not only catching his/her attention but also providing useful information about the scenario salient elements, in particular the buildings he/she is about to visit (see [54] for further details).

For both the introduced scenarios, as it is the case for each virtual rendering, a fundamental role is played by the supporting representation techniques and related hardware. The actual

implementations have been tested with a particular and challenging experimental setting, which is the newly build ElbeDom CAVE in the IFF VDTC. Here, an amount of 6 high coherence Jenoptic laser projectors render the video representation on a 360 degrees cylindrical, 3 meters tall screen. On the audio side, for creating a wide "*sweet spot"* a sound system consisting of 11 loudspeakers is available. The audio input is handled by a Creative Soundblaster X-Fi Platinum Fatal1ty sound card, and featuring a total amount of 2 GB dedicated RAM memory space for the sound processing. This configuration allows the correct spatialization even on such a widespread space of each sound source respective to the user/listener position, orientation and speed (by means of the use of an ultra red tracking system achieving 2 mm precision) in a trustworthy and totally realistic way all over a *sweet spot* of about 4 meters of diameter.

It is worth nevertheless stressing the fact that, thanks to the agile implementation, a correct sound 3D processing can be obtained without particular problems (although with obvious drawbacks due to the limited physical and calculation resources) using standard state of the art commercial computer (a computer with Windows XP OS having a 2.2 GHz Pentium processor and 1 Gb RAM extension, with a simple consumer sound card can be considered) and audio rendering systems (the aforementioned loudspeaker of headphones). Even with this configuration the end-to-end latency (the basic measure for what concerns the realism of the sound implementation and 3D playback) is not perceivable to human beings (which means that it is shorter than 50 ms [42]), as shown in [55].

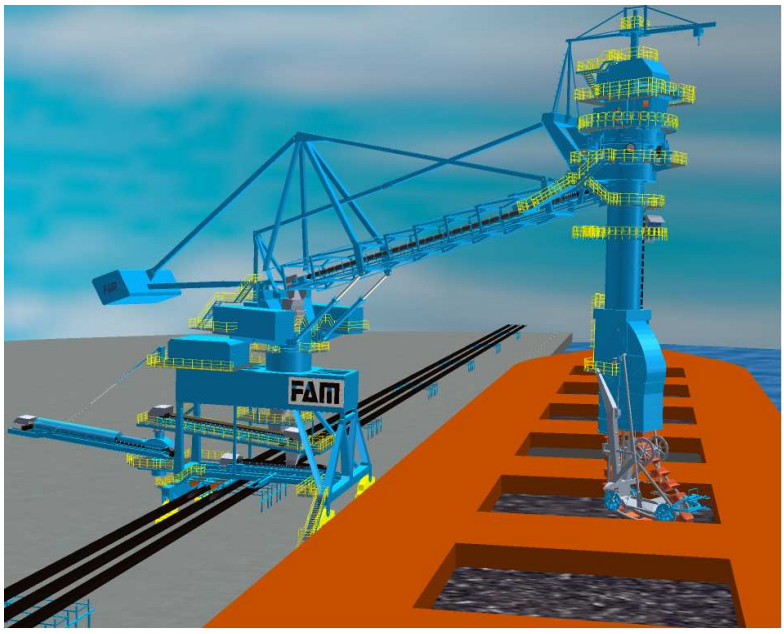


Fig. 10: *Representation of the Ship loader virtual model (© Fraunhofer IFF)*

Fig. 11: *Virtual model of Eisleben city as shown inside the ElbeDom CAVE at Fraunhofer IFF (photo courtesy: Peter Förster/Fraunhofer IFF)*

5. Conclusions

In the previous paragraphs the *core-function* realizing the audio extension for the IVS_VDT VR system has been presented. Its benefits do not reside indeed just in the effective immersivity degree boos within the virtual rendering only. In fact by means of 3D sound the audio extension allows providing the user with important information, in particular about -but not limited to- the surrounding space, while experiencing the virtual sound. The software structure of general platform and audio function has been introduced. The highlight has been given to the explanation of the characteristics making possible to an author implementing a virtual scenario with 3D sounds by spatially processing audio signals. In this way the sound effects have a precise source location and a propagation path within the virtual space so that the user can not distinguish the visual source and audio one whilst being immersed in 3D sounded scenarios; this also thanks to the lean implementation providing for a particularly low processing latency. Besides that, the implemented scenarios described have shown a number of concrete innovative and practical uses of audio spatialization in VR applications.

Still, several further improvements are yet needed and are therefore forecasted to be implemented. For the audio core function to be really completed, the interaction between audio signals propagating and the physical objects (as for example the so called room acoustics) has to be taken into account, with the implementation of the real acoustic phenomena as reverberations,

echoes, and sound occlusions. In this very field the *sofware development kit* and respective *effect extension* provided together with OpenAL 1.1 can be considered as a solid starting point.

VI

Audio in VR: The Digital Factory, the new outpost of virtual design

This chapter discloses a more concrete and actual problem which has been considered together with its possible solutions during my last part of research activities as a PhD student. A workflow for the calculation and the representation within virtual environments of the noise field generated by machinery supposed to be located in new, digitally designed factories will be described.
It is noted that this project could have not been proficiently carried out without the valuable technical and most of all scientifical support offered, besides the internal one within the Fraunhofer IFF, by the Otto-von-Guericke University of Magdeburg. Here the investigation and study of the electronic of working industrial plant and machinery is extensively performed. In particular, precious help and support came within the machine building field, namely from Professors Ulrich Gabbert and Tamara Nestorovic.

The last developed application which is going to be introduced here is completely novel and futuristic, in a research field, the "Noise Control", winning more and more importance and attention (decades ago one could not even think about "acoustic pollution"). This research area considers DSP techniques together with the new possibilities offered by VR as a solid possible solution to the problems generated by noise [53].

Within modern virtual environments every aspect of a production process can be simulated. It is for this reason that the so called Digital Factory is widely regarded as a resource able to bring dramatic improvements to industrial sites since it allows considering their limitations,

issues and potential conflicts as soon as the building project starts. For what concern the audio, the main problem resides in the factory machinery noise which can result dangerous for the human hearing system. Forecasting the overall amount of noise sound field produced becomes therefore a crucial factor which has to be considered to abide by the legal criteria while designing and building a factory.

1. Motivations

The continuous growth of productivity is a main target for most if not all companies involved in the high competitive global market. For this reason the automation and lean planning of the production line (implying often tight cooperation and synchronism among different companies) is considered as fundamental for achieving higher efficiency degree. As evident this is actually a wide target requiring the involving of several scientific branches. In fact for the success of a company, the consideration of each and every operational detail has to be guaranteed so that the production efficiency and the supply chain could be optimized given the eventual conditions (economic or environmental, for example). One of the first steps in this direction implies the virtual redesign of entire industrial sites. Having to think again every single aspect within a factory, the industry world is really keen on taking advantage of the resources offered by the newest virtual reality techniques. As regards the mere industrial applications, VR technologies have been successfully employed during the past years to the design of single products [43]. Thanks to the recent developments nevertheless, those techniques are proving being versatile and more and more useful for the general production planning [44]. The high degree of immersivity which the current virtual representations can achieve offers a solid ground for the realization and completion of the so called *Digital Factory* [45]. Within a 3D environment not only the layout of a company but every aspect of its production flow can be represented, having the opportunity to test at the same time the production flow performances or check the single critical details and processes. For this representation to have a satisfactory degree of pertinence (so to be practically reliable and trustworthy), all the possible multiple limitations regarding the industrial site under consideration have to be taken into account, without narrowing in this way the analysis to the mere design issues.

The acoustical problem related to industrial noise has strongly arisen during last century [46]. Starting from the *Noise Pollution and Abatement acts* promulgated by the Congress of the United States in 1972, the governments of every industrialized country have issued their own regulations for what concerns the sound levels to whom the human being are supposed to be exposed within working environments (see for example [58]). The consideration of the problems and linked issues regarding the acoustic field in a factory, where various industrial machines are simultaneously at work, is therefore not only an effort for the sake of instantaneous safety (as an example exaggerate noise levels might even mask fundamental acoustic alarms, like danger warnings). It is indeed and most of all a legal requirement every company needs to comply to. An application giving reliable results to face and solve this problem is currently needed for filling in a precise and systematic way this gap (as described in [47]). In fact, while software for acoustic simulation of industrial environments is commercially available (see for example [48]),

the control and analysis supported by VR techniques of noise generated by complex machinery is so far a field in which complete and reliable products are still missing.

Starting from this background, next paragraphs will describe the realization of an application for noise control within the digital factory. In particular, the workflow the implementation of a plug-in for the previously reviewed IVS_VDT virtual reality system is presented.

2. The noise problem and the noise control plug-in rationale

In a particular research field such as the Digital Factory implementation (refer to Figure 12, showing a concrete visual model), 3D sound researches and application can play a fundamental role. Not only, as seen in the previous chapter, virtual sound contributes greatly to the immersivity degree enhancement, but it also provides various kinds of information (about the surrounding space, the objects, and the performed actions, to give some examples). At the same time it can allow verifying if a certain factory digital design will end in a real environment where the several noise regulations in force in each country are obeyed to.

In particular those laws generally define a threshold concerning the Sound Pressure Level (SPL) which is not to be overcome. The SPL is an objective measure for the sound intensity in a determined spatial point, defined in the following way [51]:

$$SPL = 10\log_{10}\left(\frac{p_{rms}^2}{p_0^2}\right) = 20\log_{10}\left(\frac{p_{rms}}{p_0}\right) \quad dB \tag{38}$$

here p_0 is the reference sound pressure ($20\mu Pa$ in air) while p_{rms} is the mean square root of the measured sound pressure.

The SPL should be normalized to the HAS different behavior to the various frequency bands within the normal audible range of 20-20000 Hz which is explicated by the equal-loudness levels which determines the *Phon* unit of measurement (i.e. 1 dB SPL for a sound having 1 KHz of frequency, see Figure 13).

Usually, an extended exposition to a noise disturbance higher than 85 dB SPL (normalized by means of the A-weighting; the filter response for a filter implementing it is shown in Figure 14) is considered eventually dangerous for the human auditory system [46], i.e. causing the risk of permanent hearing diseases (e.g. from tinnitus to hearing impairments) and even total hearing loss.

The module for the assessment of the noise generated sound field has been thought to be a plug-in of the IVS_VDT. This has the scope of calculating in a reliable and effective way the sound pressure level which will be perceivable inside the factory and transferring those results to the VR platform so that they can be displayed together with the virtual representation of the factory itself within the 3D space.

Fig. 12: *Example of a Digital Factory model within the IVS_VDT platform (© Fraunhofer IFF)*

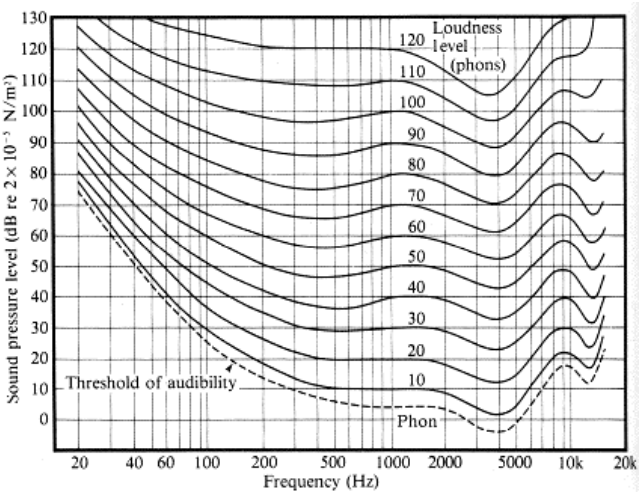


Fig 13*: Curves of equal loudness determined experimentally by Robinson & Dadson in 1956, following the original work of Fletcher & Munson*

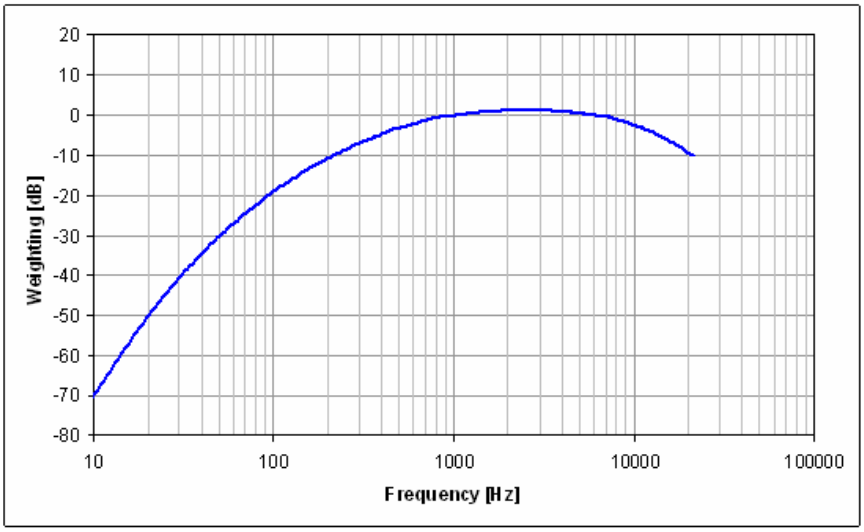


***Fig. 13**: Equal loudness contour following the A-weighting as defined in the IEC61672:2003 International standard.*

3. Implementation: Techniques and workflow

The noise control plug-in is characterized by the innovative workflow depicted in Figure 20.

The raw 3D model of a machine (in the case in example a real robot for multiple industrial utilizations as shown in Figure 15(a), designed by means of ordinary CAD software like *ProEngineer* (©, as for example the one in Figure 15) resulting in either a *Sat, Iges, Vrml or Parasolid* file, that is, having an extension respectively *.sat, .iges, .wrl,* or *.x_t* –as known a standard has not yet been found in the field-) is imported in *Ansys* (©), a software for Computer Assisted Engineering (CAE) which allows the analysis of the machine structural behavior with the calculation of the structure own oscillation normal modes. Those represent the maximum displacements that the machine can show because of vibration causing noise.

Ansys is then employed also for the operation of model meshing. This consists of the fragmentation of the 3D continuous structure in a mesh made by a discrete number of simple and small 2D polygons and 3D elements having polygonal faces (see for example Fig. 16). The operation is performed so as to have the possibility to use the *Finite Element Method* (FEM) and the *Boundary Element Method* (BEM) [49] for the produced noise field calculation. BEM and FEM have been preferred to the other methods as the geometrical ones (like the *Ray or Beam Tracing*, which are effective at high frequencies where the sound wave is highly directional and can be considered as a coherent beam) and the statistical methods (based on the probability theory to calculate the energy exchanges between the system components, i.e. sound wave and physical entity it interacts with).

a) b)

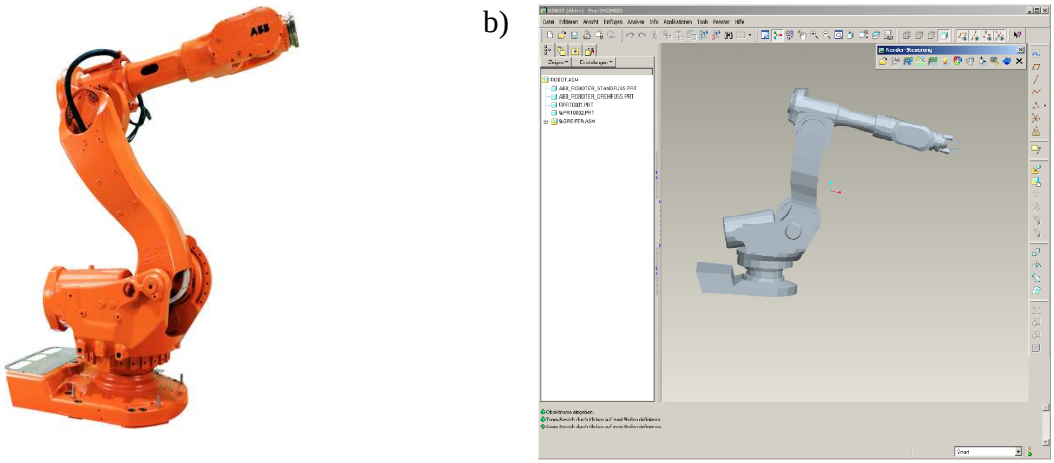


Fig.15*: A picture of a real (a) and CAD (b) model of the industrial ABB Robot, model IRB 6600, used as example for the testing (robot model courtesy of ABB Robots Company, ©)*

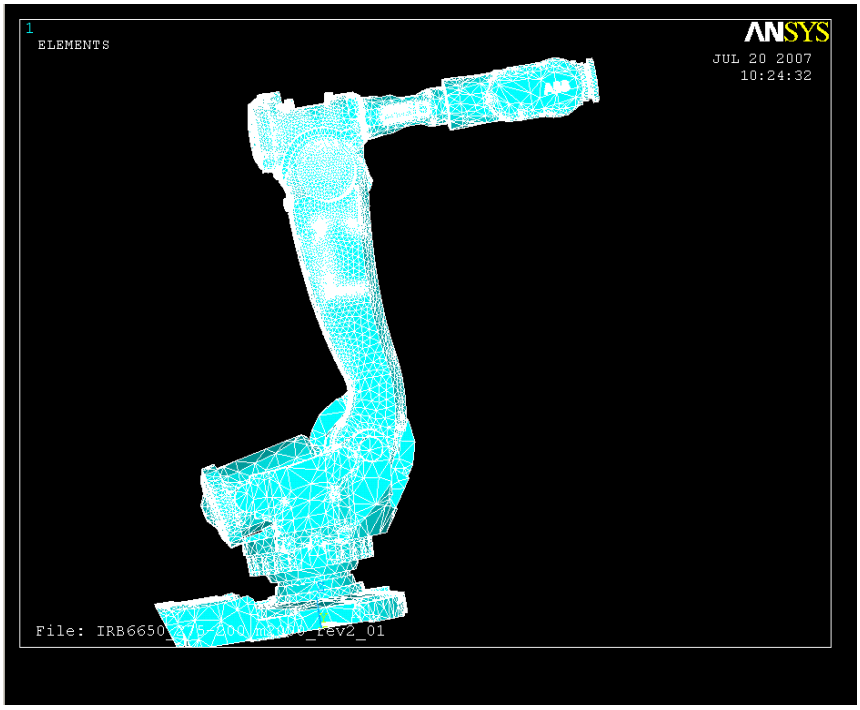


Fig.16: *The ABB Robot meshed structure, consisting of 2D and 3D elements, as obtained by means of Ansys CAE software*

The reason behind this choice is that BEM and FEM are the most effective and precise methods for calculating the acoustic behavior of a structure in the frequency range where the

HAS is most sensible and the industrial noise is stronger and therefore more dangerous, that is between 50 and 4000 Hz. The drawback consists of a huge amount of computational resources needed which makes real time calculation for normal structure rather complicated, unless the frequency band in which the investigation is carried gets extremely narrow.

Those numerical computing methods are used in acoustics for solving, by means of the appropriate boundary conditions, the huge systems differential equations, each of the latter defining the physical vibrating behavior of a single mesh element: combining the results obtained from each equation provides the overall solution.

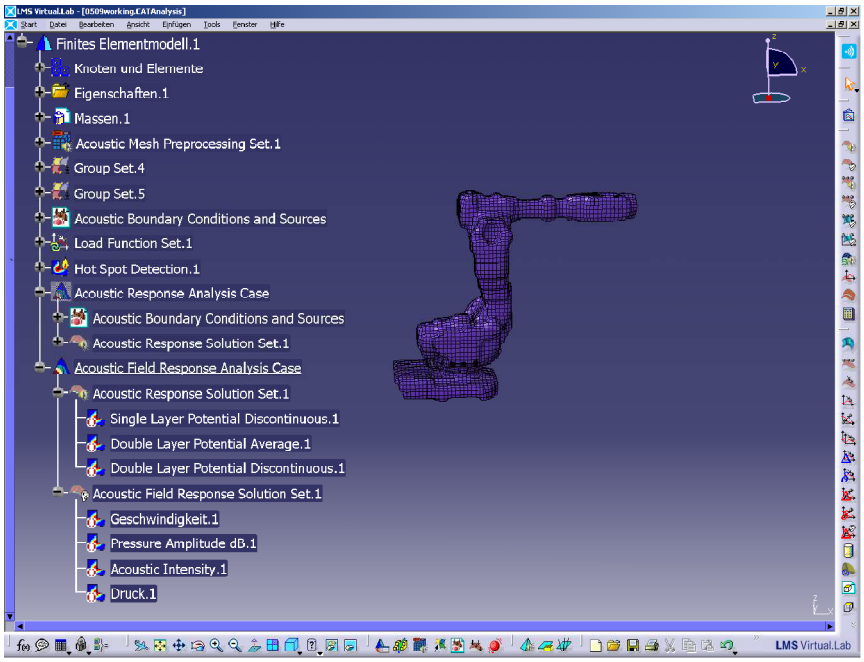


Fig 17: *The coarsened model of the ABB Robot; only the acoustic influent details of the structure are kept and further considered*

For those methods to be applied mesh and modal coordinates obtained in Ansys are imported to another CAE environment, that is *LMS Virtual.Lab* (©). By means of its *Noise and Vibrations* package the 3D model mesh can be simplified so that only the structural characteristics and properties which are of interest for the acoustical analysis are preserved. This means that small holes or ribs on the surface of the structure, usually having negligible effect on the sound production and transmission processes, while making the calculations to be performed much more complicated -and consequently the needed computational resources much more intensive-, can be expressly removed. In this way a coarsened mesh tailored for the acoustic analysis is obtained.

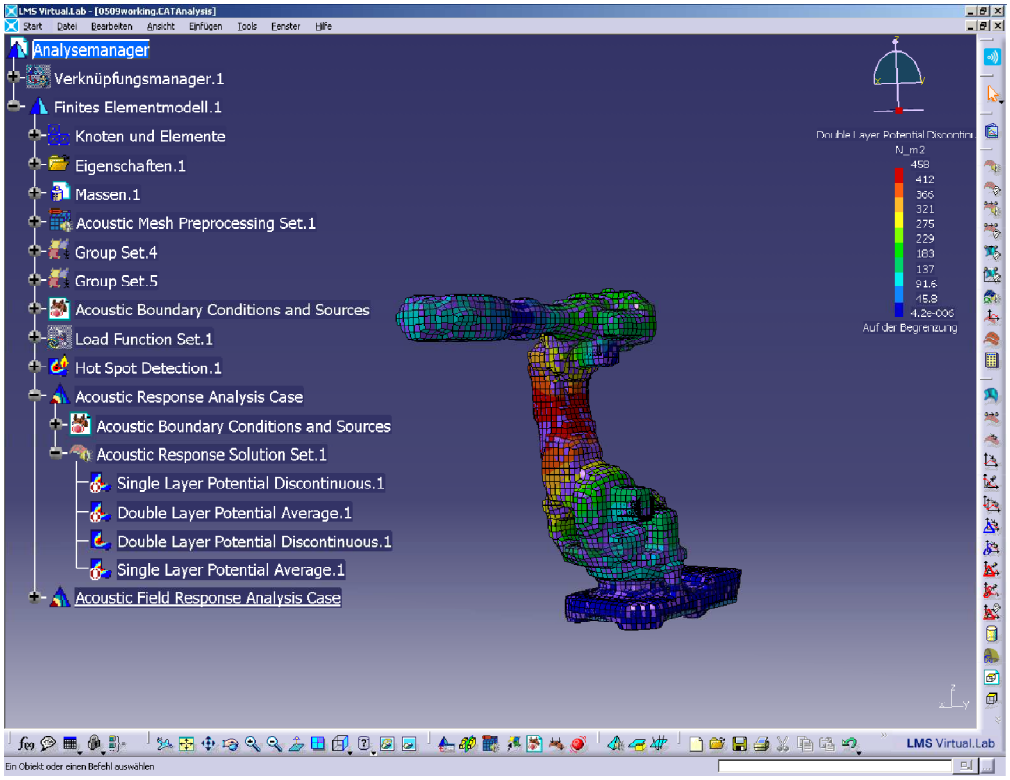


Fig. 18: *The acoustic response of the structure obtained using the FEM method.*

The LMS Virtual.Lab package *Acoustics* using as calculation tool *Sysnoise* (©) -i.e. the reference software for the noise related problem numerical solution- has been then used. The definition of the proper values following the set of boundary conditions required by the problem is needed. Those boundary conditions are [56]:

$$p = \bar{p} \tag{39}$$

which is the *Dirichlet* condition, imposing the normal pressure value on the acoustic structure surface.

$$v_n = \bar{v}_n \Rightarrow \frac{\partial p}{\partial n} = -\rho_0 \dot{v}^T n = -\rho_0 \dot{v}_n = -\rho_0 \dot{\bar{v}}_n \tag{40}$$

which is the *Neumann* condition, imposing the normal velocity value on the acoustic structure surface:

$$\frac{p}{v_n} = Z_n = \bar{Z}_n \tag{41}$$

which is the *Robin* condition, imposing the normal acoustic impedance value on the vibrating structure surface.
The acoustical sources generating the noise inside the machine (originated by a variety of elements ranging from combustion forces, to crank bearing forces, valve train, gear whine and rattle) have been accurately defined and modeled thanks to the cooperation with the team of machine simulation from the department of Machine Building within the University of Magdeburg. The effective values for the parameters have been identified through measurements and evaluate through extensive multi-body simulations and NVH-like (Noise, Vibration and Harshness) analysis. Furthermore the materials and their physical properties of the structure have been defined.

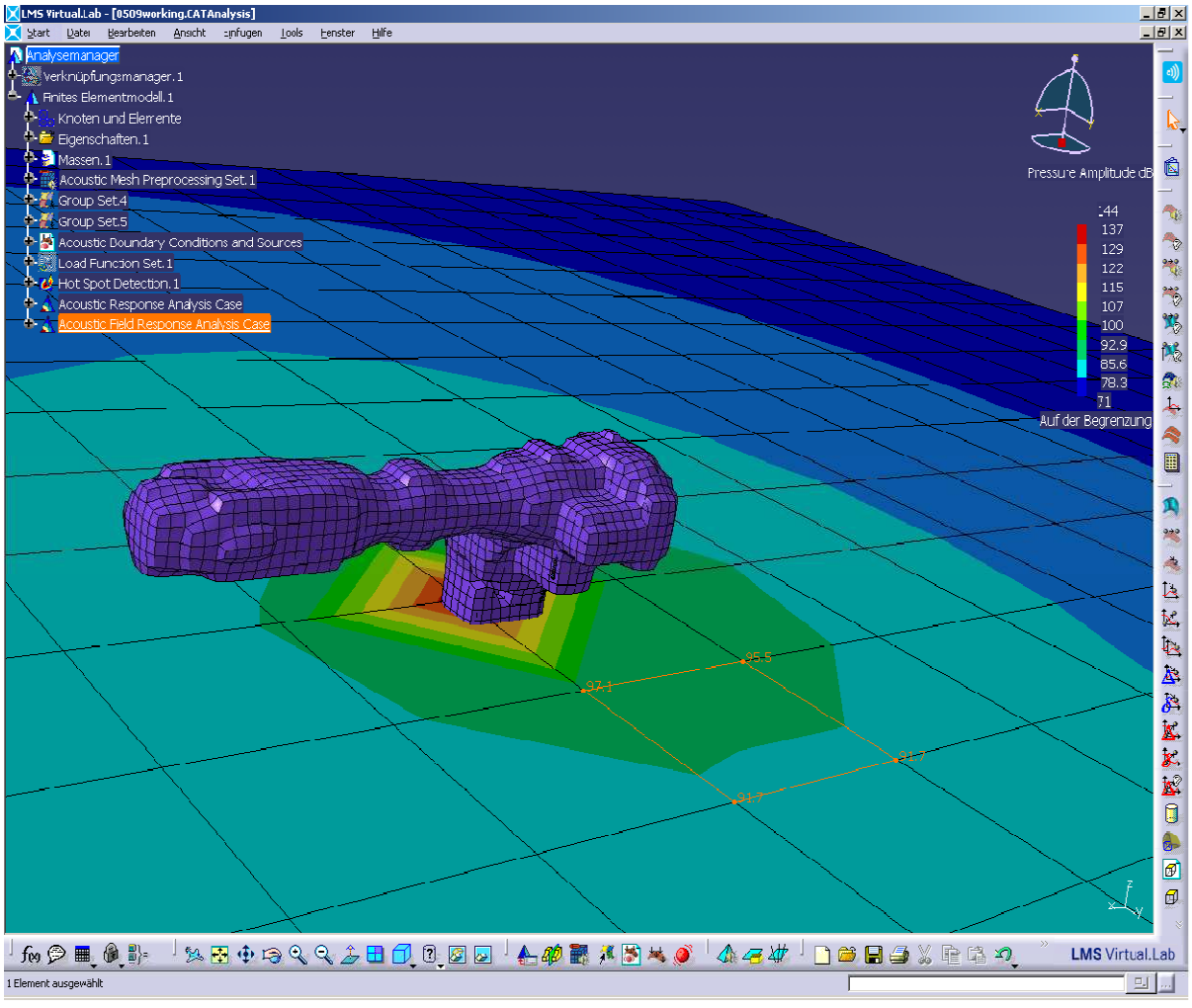


Fig. 19: *Diffused sound pressure field generated by the acoustic structure with an example of numerical values at the nodes*

4. Results

Those model settings, together with definition of the speed of sound in air, and the already calculated modal coordinates allows applying the above introduced FEM. With the help of the modal coordinates retrieved by means of Ansys (and defining how structural modes contribute to the acoustic response) the so called internal problem can be numerically solved. That is, after the necessary calculations, the Finite Elements Method applied to the machine provides the sound generation diffusion and consequently the acoustic response generated in the internal space of the machine by the sound sources of the structure and, most important, on the surfaces of the machines (refer to Fig. 18 for an example).

The BEM is on the other hand more efficient and reliable for the diffused sound field calculation within the surroundings of the machine starting from the results obtained on its surface using the FEM. The originally defined sound sources, considered as point sources, plane sources and cylindrical sources inside the structure are considered for the following investigations about the external sound field taking into account their effect on the surface. This means that every superficial element of the mesh is now considered as an acoustic source.

Together with the boundary conditions previously referred to for the FEM, the Sommerfeld condition needs in addition to be imposed, that is

$$\lim_{r\to\infty}\left[r\left(\frac{\partial p}{\partial r}\right)+\frac{1}{c}\frac{\partial u}{\partial t}\right]=0 \tag{42}$$

This simply implies the necessary assumption that the sound field fades to zero at infinite distance from the sound source.

In this scenario, the famous Acoustic Wave Equation where the transmission mean has been supposed as homogeneous,

$$\nabla^2 p-\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}=0\,, \tag{43}$$

(in the equation, $c = \omega/k$ is the speed of sound, where k is the wave number and ω the pulsation angular frequency) can be solved in a precise manner for each mesh element as well as the following equation defined by the *Helmholtz-Kirchhoff integral theorem*

$$p(x)=\frac{1}{4\pi}\int_{Surface}\int (G(|\,x-y\,|,\,f)\hat{n}\cdot v(y))-p(y)\hat{n}\cdot\nabla' G(|\,x-y\,|,\,f)dy \tag{44}$$

Here, *x* is a point in the surrounding space where the diffused sound is supposed to get measured whereas *y* is a point on the mesh element surface on which the normal $\hat{n}$ is taken, while

$$G(|x-y|, f) = \frac{e^{i\frac{2\pi f}{c}|x-y|}}{|x-y|} \tag{45}$$

is the Green's function describing how sound radiates from a point.

To perform the output calculation (that is the sound field created by the sound pressure levels in the proximity of the machine) a *"net"* made by *virtual microphones* is defined around the examined machinery. Those microphones are implemented to simulate compliancy to the respective ISO standard [50] for the SPL measurements and are basically the points where the sound pressure levels are calculated (by means of the pressure values in these points, retrieved thanks to the acoustic sound wave equation).

In Figure 19 an example of the obtained final configuration is given. The different levels of sound pressure amplitude generated by a working industrial machine are shown. At each node point a "virtual microphone" provides the calculated exact value for the pressure.

The numerical values obtained by means of this last operation in Virtual.Lab Acoustics are finally saved as Universal Files (UFF, Universal data File Format, having the *.unv* extension), which are industrially recognized text files allowing easy storage and transfer of technical calculation results (having several identification numbers related to their fixed different contents) being ASCII data (and therefore editable). In particular, referring also to Table 3, the formal properties of the analysis together with its units of measurement are first defined (Universal File datasets 151 and 164 is therefore employed) and the positions relative to the virtual microphones within the space are reported (Universal File dataset 2411). The information about the structure generating the noise is then recorded (Universal File dataset 2412) and finally the results of the whole analysis are stored; to be more precise the nodes position within the 3D space (linking the info in the UNV 2411) together with their SPL values (in the UNV 2414) are saved in the resulting *.unv* files.

The resulting files containing the salient point of the previously performed operations are then ready to be imported in the UNV_IVS platform by means of a dedicated plug-in. Here they are displayed within the 3D environment of the Digital Factory. The maintaining of the correct spatial location and orientation, thanks to the matching between the model used for the acoustical analysis and the one rendered in the virtual space by means of the platform, needs simply the definition of the so-called *pivot point* for the machine.

An example of the UNV file resulting from the calculation of the noise field generated from an industrial machine is reported in Annex B.

Considering the obvious presence of multiple industrial machines, it can be supposed that the several noise fields are uncorrelated (the most probable condition by far, and at the same time the most dangerous for the HAS -since it generates the highest SPL-, which has always to be considered while dealing with security measure issues). Therefore, since the mean square pressures from independent sources are merely addictive [51], the overall SPL calculation (for

each frequency sub band) in a certain point of the Digital Factory environment (corresponding to where the virtual microphones are placed) due to all the noise sources will have for each frequency sub band the following expression:

$$SPL_{tot} = 10 \log \left(\sum_i 10^{0.1 SPL_i} \right), \tag{46}$$

In the equation 46, the index *i* is referred to each single sound field calculated from each industrial machine.

Those final overall values undergo in the end to an A-weight filtering and are then ready to be displayed for each node within the 3D space in an immediate way. For this purpose it has been decided to employ a chromatic scale, so that each node assumes a color corresponding to the calculated SPL thereto.

As an example, green spots are used for values well below the legal threshold, yellow is used for values below but approaching the legal threshold, whereas red represents the areas of the digital factory where a certain machinery configuration is revealed to be not legally compliant.

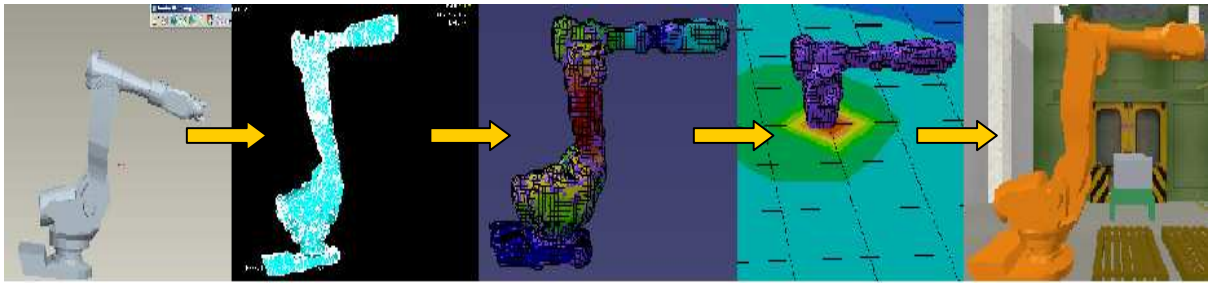


***Fig. 20**: Noise control plug-in general workflow. The different steps from left to right: 3D graphic project (CAD), meshing, structural analysis, sound field calculation, representation in the virtual environment (robot model courtesy of ABB Robots Company, ©).*

5. Conclusions and future developments

By means of the described plug-in, it turns much easier to assess in a trustworthy way an industrial factory design and project for what concerns the respect and compliance to the specific noise law for working environments. The SPLs which are going to be created within the production site by the future machine configurations can be recreated and the possibility to spot in advance sound critical areas where the use of passive sound reduction techniques (as for example silencing panels) is given. In this way corrections to the design can be applied at a stage where modifications are dramatically cost effective and relatively easy to perform compared to the situation in which the actual factory already existed.

UNV File Type	Results Format (reduced and with comments –in red)	Description of the Contents
151	-1 151 D:\LMS\20080514.CATAnalysis NONE LMS Virtual.Lab Rev 7B 14-May-08 10:53:52 -1	Header and model name with declaration of the used software
164	-1 164 1 SI - mks (Newton) 2 *//International System (Meter, Kilo, Second and Newton) is employed* 1.00000000000000000e+000 *//Conversion factors for the units of measure* 1.00000000000000000e+000 1.00000000000000000e+000 0.00000000000000000e+000 *//Temperature Offset* -1	Units of Measure used
2420	-1 2420 1 *//Cartesian coordinate system* -1	Definition of the coordinate system
2411	-1 2411 707757 1 1 1 -7.096568630293243e-001 -4.999999999999949e-003 2.225429284265770e-001 : : -1	Position of the virtual microphones for the analysis
2412	-1 2412 711236 94 1 1 5 4 *//Node number, element type, displacement and color* 721715 721717 721718 721716 *//Nodes connected by the element* : : -1	Definition of the structural elements (2D surfaces and 3D solid polygons)
2414	-1 2414 1 Exported by LMS: Modal data: complex nodal pressure 1 NONE Data written by LMS Virtual.Lab NONE Wed May 14 11:19:59 2008 NONE 1 2 1 117 5 1 *//Results are scalar single precision complex pressures* 0 0 1 0 0 1 0 0 //Data characteristics are normal modes 0 0 0 0 0 0 0 0 *//Integer and real analysis type specific data* 0.00000e+000 1.00000e+002 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 6.28319e+002 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 1 *//Node number* -9.46084e-002 6.84137e-002 *//Complex nodal pressure* 2 *//Node number* -1.11550e-001-4.25880e-002 *//Complex nodal pressure* 3 *//Node number* : -1	Actual results of the acoustical analysis with pressure values at the different nodes.

Table 3: *Schematic representation of the final results of the acoustic analysis of the ABB industrial Robot saved in the UNV extension format (for the extended version please refer to Annex B)*

In addition, this implementation has been thought to be further developed in order to allow the simulation, the test, and the comparison among various Active Noise Reduction (ANR) techniques. For example in the one defined in [52], the noise attenuation is achieved by means of piezoelectric transducer patches working both as sensor and actuator (their stiffness being tuned following the vibrations measured by a Linear Quadratic Regulator -LQR- with feedback, implemented on a computer connected to the machine). The whole piezoelectric structure can be totally modeled by means of the FEM [56] and therefore can be easily included in the sound analysis previously described. Other DSP for ANR paradigms can be simulated and tested.

The different concepts following the basic concept of the directional "*antinoise*" sound wave creation (a microphone measures the noise field and then the info is processed so to create a perturbation having same amplitude but opposite phase, see for example [53]) can be implemented thanks to the sound spatialization capabilities of the IVS_VDT platform already disclosed at the beginning of this chapter.

In this way a powerful, effective and versatile tool can be provided to the industrial world for solving in a practical manner the noise concerns by means of digital signal processing techniques.

VII

Conclusions

This thesis has summed up the research activities performed in different international laboratories during my doctoral studies. In particular, within the frame of the Digital Signal Processing I focused my research on the Audio Signal. Specifically, a number of applications have been disclosed in which the audio wave, represented in its digital form as a finite sequence of numbers, is expressly and properly modified. My overall aim was to show how digital signal techniques can be applied effectively in pretty diversified fields providing for bringing a specific contribution to the solution of diverse technical problems (as stated in Chapter II).

The researches range among three major topics within the digital signal processing, namely Watermarking, Audio Restoration and Virtual Reality.

In Chapter III, the implementation of an Audio Digital Watermarking system has been described. This application has been realized at the Digital Signal Processing, Multimedia, and Optical Communications Lab at the University of Roma3.

The idea is to use a *fragile* digital mark (i.e. not perceivable but prone to be modified) to be hosted by the middle-high frequencies of a musical file. The watermark is then used on the receiver side in order to *blindly* quantify the Quality of Service offered by the telecommunication provider. In fact, once the file is compressed in a MP3-like format and then sent on a communication channel, several causes (as for example noise and disturbances while transferring the data) can contribute to its degradation. Having stored the original version of the used watermark (which can be the same for every communication and has low pay-load), the receiver extracts the sent watermark from the downloaded musical digital content, and compare the two versions. The results presented in this thesis (as well as in [59]) show how, even using a simple metric like the Mean Square Error between the marks, a reliable assessment of the received audio quality can be obtained. The performed experiments widely prove the direct proportionality between the Mean Square Error of the received mark (as compared to the inserted one) and the received audio data (as compared to the original unmarked one). Furthermore, the watermarking technique has been tested following the ITU-R recommendations for the *Objective Measurements of Perceived Audio Quality*. In this case the watermark inserted into the audio wave has proved to be completely not audible by the Human Auditory System (a concern that

has lately decreed the failure of state of the art Digital Right Management techniques for musical data).

In Chapter IV, the use of a probabilistic Phase Vocoder as a generative model for the audio signal, in order to statistically assess the missing samples along a corrupted musical track, has been depicted. My research activity at this purpose has been carried out at the Signal Processing at Communication Lab of the University of Cambridge (UK). There I brought together the bases of the Bayesian Theory, vastly used in that Laboratory, with the Vocoder background. Once defined the model, an expectation maximization procedure is performed, estimating missing samples by means of a forward process (which can be equated to a factorial Kalmann filtering) and a backward process (i.e. a smoothing by means of the Rauch-Tung-Striebel equations). The implementation has provided satisfactory results, since the efficiency of the approach has been clearly proved. Refinements to the generative model and estimation procedure have to be developed in order to make it more versatile; in this way the technique can become totally reliable for performing high quality audio restoration for any kind of corrupted complex musical wave.

Chapter V introduces two main applications of Audio DSP within the futuristic research field of the Virtual Reality technologies. These applications have been developed at the Virtual Development and Training Centre of the Fraunhofer Insitut for Factory Logistics and Automation (IFF), with the support of the Otto-von-Guericke University, both in Magdeburg, Germany.

Directly in chapter V the audio core function complementing a visual VR engine (the IVS_VDT) has been described. This implementation allows agile and realistic implementation of 3D sound effects within virtual environments. Notably, the author of the scenario is given the possibility to define several parameters regulating not only the position of the sound sources but also the fictive diffusion of the various sound waves. The uploaded audio files are in fact processed so that the created audio perturbations own a spatial definition (by means of an amplitude stereo modulation). In addition, particular pitch and general frequencies tunings realize different acoustic phenomena (e.g. the Doppler Effect), rendering the sound experience for the user closer and closer to reality (see [55]). The immersivity degree enhancement of the overall virtual representation being just the major of the possible benefits brought by virtual sound, two scenario realized using the described audio function have been introduced. These two examples broaden the improvements which correct auralization can bring to the VR rendering (see [54]) while yet remaining open air scenarios; for a correct audio experience within closed environment, room acoustics has to be taken into account in the next implementations.

Chapter VI, on the other hand, concerns an innovative idea joining the VR 3D realistic visualization and the aforementioned properties of properly spatialized virtual audio, with the calculation capabilities of different Computer Assisted Software and acoustical physics. In order to consider the noise pollution and subsequent compliance to noise regulations while designing a new industrial site, the noise field created by the machinery has to be assessed. For this task a workflow among different dedicated software has been described. Starting from the raw 3D model of each machine, the vibration modes are calculated and the structure is meshed. Modes and mesh (together with the proper boundary condition) are of use for the application of the Finite Element Method to calculate the acoustic behavior inside of the structure. The Boundary

Element Method is then applied to solve the acoustic wave equation and retrieve the Sound Pressure Levels determined by each working machine within the factory space. The final results are saved as easily editable UNV text files, making possible a successful and smooth transferring the calculated overall produced noise levels to the virtual model of the factory. Within the 3D space it becomes direct and immediate to verify the SPL levels present in the various areas of the industrial site spotting problematic areas (see [60]). It is in this way possible to design cost effective modifications of faulty projects for what regards future possibly non-compliant noisy factories. Moreover, thanks to the versatility of VR techniques, this idea can constitute the basis for coherent Active Noise Reduction and Cancellation methods testing.

All those effective applications, besides the mere technical and scientific results achieved, have shown the potential of the newest technique of Audio Digital Signal Processing to the solution of various real life problems. This is what research should be carried out for, improving life quality by means of the understanding of nature. With Audio DSP it is for sure possible.

VIII

Annex A

Audio restoration Matlab source code for the PPVOC implementation

```
function EM( )
hold off
clear
C=[1 0];
n=0.1;

A=[ cos(n) -sin(n);  sin(n) cos(n)];
K=size(C,2);
N=size(C,1);
Q=(0.02^2)*eye(K);
R=(0.02^2)*eye(N);

T=100;
s(1,:)=[1 1];
x(1)=1;

for k=1:1:T
    s(k+1,:)=A*s(k,:)'+(0.001*randn(K,1));
    x(k+1,:)=C*s(k+1,:)'+(0.001*randn(N,1));;
end

A=randn(K);
xi=x;
xi(16:19,:)=NaN;
xi(35:39,:)=NaN;
xi(64:68,:)=NaN;
xi(91:93,:)=NaN;


logL=0;
for i=1:1:10000
```

```
sup(1,:)=ones(1,K);

sn(1,:)=ones(1,K);
Pup(1:2,1:K)=cov(sup);
for k=1:1:T
    sup(k+1,:)=A*sup(k,:)' ;
    Pup(2*k+1:2*k+2,:)=A*Pup(2*k-1:2*k,:)*A'+Q;
    Pkf(2*k+1:2*k+2,:)=Pup(2*k+1:2*k+2,:);
    skf(k+1,:)= sup(k+1,:);
    if ~isnan(xi(k+1,:))
       sn=sup(k+1,:);
       P(2*k+1:2*k+2,:)=Pup(2*k+1:2*k+2,:);
       error=xi(k+1,:)'-C*sn';
       S=C*Pkf(2*k+1:2*k+2,:)*C'+R;
       KG=P(2*k+1:2*k+2,:)*C'*inv(S);
       sup(k+1,:)=sn'+KG*error;
       Pup(2*k+1:2*k+2,:)=(eye(K)-KG*C)*P(2*k+1:2*k+2,:);

    end


end
res(i,:)=sup(:,1);

%%smoothing process%%

ssm(T+1,:)=sup(T+1,:);
Psm(2*(T+1)-1:2*(T+1),:)=Pup(2*(T+1)-1:2*(T+1),:);
spl=zeros(T,1);
if i>2
    spl=ssm(:,1);
end
for k=T:-1:2
    J(2*k-1:2*k,:)=Pup(2*k-1:2*k,:)*A'*inv(Pkf(2*k+1:2*k+2,:));
    ssm(k,:)=sup(k,:)+(J(2*k-1:2*k,:)*(ssm(k+1,:)'-A*sup(k,:)'))';
    Psm(2*k-1:2*k,:)=Pup(2*k-1:2*k,:)+J(2*k-1:2*k,:)*( Psm(2*k+1:2*k+2,:)-
Pkf(2*k+1:2*k+2,:))*J(2*k-1:2*k,:)';
end
ssm(1,:)=1;
Pbkw(2*(T+1)-1:2*(T+1),:)=(eye(K)-KG*C)*A*Pup(2*(T+1)-3:2*(T+1)-2,:);

for k=T:-1:2
    Pbkw(2*k-1:2*k,:)=Pup(2*k-1:2*k,:)*J(2*k-3:2*k-2,:)'+J(2*k-
1:2*k,:)*(Pbkw(2*k+1:2*k+2,:)-A*Pup(2*k-1:2*k,:))*J(2*k-3:2*k-2,:)';
end
Pbkw(1:2,:)=Pup(1:2,:);
DA=0;
NA=0;
CA=0;

for k=2:1:T+1
DA=DA+(Psm(2*k-3:2*k-2,:)+ssm(k-1,:)*ssm(k-1,:)');
NA=NA+(Pbkw(2*k-1:2*k,:)+ssm(k,:)*ssm(k-1,:)');
CA=CA+(Psm(2*k-1:2*k,:)+ssm(k,:)*ssm(k,:)');
end
```

```
A=NA*inv(DA);
Q=(1/(T+1))*(CA-A*NA');
h=0;
SOM=0;
for k=2:1:T+1
    if ~isnan(xi(k))
    SOM=SOM+((xi(k)-ssm(k,1))*(xi(k)-ssm(k,1))'+(C*Psm(2*k-1:2*k,:)*C'));
    h=h+1;
    end
end
R=(1/(h))*SOM;


logLa=0;
for k=1:1:T+1
   if ~isnan(xi(k))
    logLa=logLa-0.5*(log(norm(C*Pkf(2*k-1:2*k,:)*C'+R)))-0.5*((xi(k)-
skf(k,1))'*inv(C*Pkf(2*k-1:2*k,:)*C'+R)*(xi(k)-skf(k,1)));
   end
end


if i>2
    if logLa<=logL
    break
else
    logL=logLa
    end
    end
end
i
A
Q
R
figure
plot(s(:,1),'r')
hold
plot(xi,'g')
plot (spl,'b')
mse(s(:,1),res(1,:));

%mse(s(:,1),spl);
mse(s(:,1),ssm(:,1));
[JH HJ]= qr(A)
```

IX

Annex B

Universal File Format obtained from the analysis of an industrial machine irradiated sound field for an ABB Robot, model IRB 6600 (©).

```
   -1
   151
D:\LMS\20080514.CATAnalysis
NONE
LMS Virtual.Lab Rev 7B


LMS Virtual.Lab Rev 7B
 14-May-08  10:53:52
    -1
    -1
   164
         1 SI - mks (Newton)  2
 1.00000000000000000e+000 1.00000000000000000e+000 1.00000000000000000e+000
 0.00000000000000000e+000
    -1
    -1
  2420
         1

    -1
    -1
  2411
    707757         1         1         1
  -7.096568630293243e-001  -4.999999999999949e-003   2.225429284265770e-001
    707758         1         1         1
  -7.096568630293241e-001  -5.499999999999979e-002   2.225429284265770e-001
    707759         1         1         1
  -7.097356642223135e-001  -4.999999999999943e-003   2.003188653424680e-001
  .
  .
  .
  .
    721984         1         1         1
  -7.266866179495249e-001  -3.554132750990159e-001   1.985446375912874e-001
    721985         1         1         1
  -7.104933091760628e-001  -3.546271272956723e-001   2.218083352339221e-001
    721986         1         1         1
```

```
  -7.101200227780007e-001  -3.357189713653798e-001   1.796647481281080e-001
    -1


  -1
  2412
    711236        94         1         1         5         4
    721715    721717    721718    721716
    711237        94         1         1         5         4
    721717    721719    721720    721718
    .
    .
    .
    .
    721977    721983    721984    721979
    725486        94         1         1         5         4
    721979    721984    721985    721980
    725487        94         1         1         5         4
    721983    721986    707762    721984
    -1

    -1
-1
  2414
         1
Exported by LMS: Modal data: complex nodal pressure
         1
NONE
Data written by LMS Virtual.Lab
NONE
Wed May 14 11:19:59 2008

NONE
         1         2         1       117         5         1
         0         0         1         0         0         1         0         0
         0         0         0         0         0         0         0         0
 0.00000e+000 1.00000e+002 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
 0.00000e+000 6.28319e+002 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
         1
-9.46084e-002 6.84137e-002
         2
-1.11550e-001-4.25880e-002
         3
-3.17267e-002-1.17615e-001
    207138
 7.25249e-002-1.00497e-001
         5
 1.24818e-001-1.47378e-002
    207140
 1.01641e-001 7.61839e-002
         7
 3.06123e-002 1.24195e-001
    207142
-4.54259e-002 1.20041e-001
    207143
-9.85231e-002 8.22731e-002
    207144
-1.23016e-001 3.53706e-002
    207145
-1.27297e-001-4.04558e-003
    207146
-1.23083e-001-2.93627e-002
    207147
-1.19090e-001-3.99494e-002
        14
-1.19124e-001-3.66689e-002.
.
.
       874
-2.40997e+004-9.79111e-002
```

```
       875
-2.40880e+004-9.85247e-002
  -1


  -1
  2414
         1
Exported by LMS: Complex nodal translational and rotational velocity
         1
NONE
Data written by LMS Virtual.Lab
NONE
Wed May 14 11:19:59 2008
NONE
         1         2         3        11         5         6
         0         0         1         0         0         1         0         0
         0         0         0         0         0         0         0         0
 0.00000e+000 1.00000e+002 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
-0.00000e+000 6.28319e+002 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
         1
-1.80867e-002 2.74775e-003-1.76956e-002 2.36775e-003 2.19236e-003-1.20370e-004
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
         2
-9.09769e-003-1.77763e-002-7.66272e-003-1.57294e-002 8.72461e-004 2.22262e-003
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
         3
 1.39456e-002-1.66585e-002 1.11102e-002-1.26907e-002-1.85445e-003 1.80328e-003
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
    207137
 2.30213e-002 4.79548e-003 1.54863e-002 3.63399e-003-2.65224e-003-8.54145e-004
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
         5
 9.62793e-003 2.33636e-002 5.15582e-003 1.36221e-002-8.24717e-004-2.86173e-003
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
    207139
-1.20350e-002 2.40498e-002-6.12227e-003 1.12476e-002 1.72305e-003-2.63793e-003
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
         7
-2.61764e-002 1.06788e-002-9.99872e-003 3.59115e-003 3.17315e-003-8.72677e-004
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
    207141
-2.88019e-002-5.27553e-003-7.83354e-003-1.89332e-003 3.22975e-003 1.01383e-003
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
.
.
.       440
 9.16233e-004 4.79712e-004 8.40173e-004 4.70452e-004 7.45036e-005 1.21614e-005
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
    207574
 1.02971e-003 1.42340e-004 1.06557e-003 1.50277e-004 7.60581e-005 1.05624e-005
 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000 0.00000e+000
    -1
```

X

References

[1] C. I. Podilchuk, E. J. Delp, "Digital Watermarking: Algorithms and Applications", *IEEE Sign. Proc. Mag*., pp. 33-46, July 2001.

[2] F. Hartung and M. Kutter, "Multimedia Watermarking Techniques", *Proc. of the IEEE*, vol. 87, no. 7, pp. 1079-1107, 1999.

[3] S. Cheng, H. Yu, Zixiang Xiong, "Enhanced spread spectrum watermarking of MPEG-2 AAC", *IEEE Int. Conf. on Acoustics, Speech, and Signal Proc., ICASSP'02*, vol. 4, pp. 3728-3731, May 2002.

[4] P. Bassia, I. Pitas, N. Nikolaidis, "Robust audio watermarking in the time domain", *IEEE Trans. on Multimedia*, vol. 3, no. 2, pp. 232-241, June 2001.

[5] M.D., Swanson, B. Zhu, A.H. Tewfik, "Current state of the art, challenges and future directions for audio watermarking", *IEEE Int. Conf. on Multimedia Computing and Systems*, vol. 1, pp. 19-24 June 1999.

[6] Y. Yaslan, B. Gunsel, "An integrated decoding framework for audio watermark extraction", *Proc of the 17th Int. Conf. on Pattern Recognition*, ICPR 2004, vol. 2, pp. 879-882, Aug. 2004.

[7] Ching-Te Wang; Tung-Shou Chen; Wen-Hung Chao, "A new audio watermarking based on modified discrete cosine transform of MPEG/audio layer III", *IEEE Int. Conf. on Networking, Sensing and Control*, vol. 2, pp. 984-989, 2004.

[8] J. Haitsma, T. Kalker, "Speed-change resistant audio fingerprinting using auto-correlation", *Proc. Int. Conf. on Acoustics, Speech, and Sign. Proc.*, vol. 4, pp. 728-731, Apr. 2003.

[9] A.N. Lemma, J. Aprea, W. OOmen, L. van de Kerkhof, "A Temporal Domain Audio Watermarking Technique", *IEEE Trans. on Sign. Proc*., vol. 51, no.4, pp. 1088-1097, Apr. 2003.

[10] P. Campisi, M. Carli, G. Giunta and A. Neri, "Blind Quality Assessment System for Multimedia Communications Using Tracing Watermarking", *IEEE Trans. on Sign. Proc.*, vol. 51, no. 4, pp. 996-1002, Apr. 2003.

[11] D. Kirovski, and H. Malvar, "Robust Spread-Spectrum Audio Watermarking," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, May 2001, pp. 1345–1348.

[12] M. Arnold, "Audio Watermarking: Features, Applications and Algorithms," *Proc. of the IEEE Int. Conf. on Multimedia and Expo (ICME 2000)*, pp. 1013–1016, New York, July 2000.

[13] G. C. Rodriguez, M. N. Miyatake, H. M. Perez Meana, "Analysis of Audio Watermarking Schemes", *Proc. of 2nd Int. Conf. on Electrical and Electronics Engineering*, pp. 17-20, Sept. 2005.

[14] ITU-R Recommendation BS.1387, *Method for Objective Measurements of Perceived Audio Quality*, Dec. 1998.

[15] S. Hacker, *MP3: The Definitive Guide*, O'Reilly, May 2000.

[16] D. Pan, "Tutorial on MPEG/Audio Compression" *IEEE Multimedia J.*, Summer 1995 issue, May 1995.

[17] N. J. Thorwirth, P. Hmatic, R. Weis, J. Zhao, "Security Methods for MP3 Music Delivery", *in Proc. of 34th Asilomar Conf. on Signals, Systems and Computers*, vol. 2, pp. 1831-1835, Nov. 2000.

[18] ISO11172-3 ISO/IEC International Standard Information Technology, *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits – Part 3: Audio*, ISO11172-3.1996.

[19] G. Giunta, "Quality of service assessment in new generation wireless video communications", *Digital Image Sequence Processing, Compression, and Analysis*, chap.5, pp. 135-150, July 2004, CRC press.

[20] S. Winkler, "Visual fidelity and perceived quality: Toward comprehensive metrics," *Proc. SPIE*, vol. 4299, 2001.

[21] S. J. Godsill and P.J.W. Rayner, "Digital Audio Restoration - a statistical model based approach", *Cambridge University Press*, September 21, 1998

[22]. N. J. Miller, "Recovery of singing voice from noise by synthesis". *Thesis Tech. Rep.* , ID UTEC-CSC-74-013, Univ. Utah, May 1973.

[23]. T. G. Stockham, T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing." *Proc. IEEE*, 63(4): 678–692, April 1975.

[24]. W. Etter, “Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1124{1135, 1996.

[25]. P. J. Wolfe and S. Godsill, “Interpolation of missing values using a gabor regression model," in *Proceedings of the ICASSP*, 2005

[26]. T. Cemgil and S. J. Godsill, “Probabilistic phase vocoder and its application to interpolation of missing values in audio signal”, *EUSIPCO 2005, September 2005.*

[27]. J. L. Flanagan and R. M. Golden, “Phase vocoder," *Bell System Technical Journal*, pp. 1493-1509, November 1966.

[28]. R. Shumway and D. Stoffer, “An approach to time series smoothing and forecasting using the em algorithm", *J. Time Series Analysis*, vol. 3, no. 4, pp. 253-264, 1982

[29]. R. E. Kalman. “A new approach to linear filtering and prediction problems”, *Trans. ASME J. of Basic Eng.*, 8:35–45, 1960

[30]. A.H. Jazwinski, “Stochastic Processes and Filtering Theory”, *Academic Press*, New York, NY, 1970

[31]. H. E. Rauch, F. Tung, C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems”, *J. Amer. Inst. Aeronautics and Astronautics 3* (8) (1965) 1445--1450

[32]. H. Rheingold, *Virtual Reality*, Simon & Schuster, New York, N.Y., 1991.

[33]. E. Sutherland. “The Ultimate Display”, *Proceedings of IFIP65*, Vol. 2, pp. 506-508, 1966

[34]. W. Bender, *“Computer animation via Optical Video Disc,”* Thesis Arch, 1980, M.S.V.S., Massachusetts Institute of Technology.

[35]. C. Cruz-Neira, D. Sandin, T. DeFanti, R. Kenyon, J. Hart, “The CAVE: Audio Visual Experience Automatic Virtual Environment”, *Communications of the ACM 35*, vol. 6, pp. 65-72, June 1992

[36]. R. Begault, “3D sound for virtual reality and multimedia.” *Durand National Aeronautics and Space Administration*, NASA/TM, 2000.

[37]. M. *Gerzon,* “Multidirectional sound reproduction systems” *US Patent 3,997,725*. December 14 1976

[38]. A.J. Berkhout, D. De Vries; P. Vogel, “Acoustic Control by Wave Field Synthesis”, *Journal of the Acoustic Society of America*, vol. 93, pp. 2764-2778, Mai 1993

[39]. R. Gilkey, T.R. Anderson, “Binaural and Spatial Hearing in Real and Virtul Environments*”*, *Psychology Press*, January 1997

[40]. G. Voß, J. Behr, D. Reiners, M. Roth**, “**A multi-thread safe foundation for scene graphs and its extension to clusters”, *Proceedings of the Fourth Eurographics Workshop on Parallel Graphics and Visualisation*, pp. 33–37, 2002.

[41]. “OpenAL 1.1 Specification and Reference*”*, *Creative Labs*, June 2005

[42]. M. O’Donnel, “Producing Audio for Halo*”*, *MS Press*, 2002

[43]. E. Blümel, S. Straßburger, R. Sturek, and I. Kimura, "Pragmatic Approach to Apply Virtual Realty Technology in Accelerating a Product Life Cycle", *Proceedings of the International Conference INNOVATIONS 2004*, June 11-12, 2004, Slany, Czech Republic, pp. 199-207.

[44]. M. Schenk, S. Straßburger, and H. Kissner. "Combining Virtual Reality and Assembly Simulation for Production Planning and Worker Qualification", *Proceedings of the International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV 2005)*, eds. M. Zaeh and G. Reinhart, München, Germany, September 22-23, 2005, pp. 411-414.

[45]. W. Kuhn, "Digital Factory - Simulation Enhancing the Product and Production Engineering", *Proceedings of the 2006 Winter Simulation Conference*, Session: Manufacturing applications: manufacturing systems design, Monterey, CA, December 3-6, 2006 pp. 1899 – 1906.

[46]. K. D. Kryter, "The Effects of Noise on Man", Orlando, FL, Academic Press, 1985.

[47]. G. Rosenhouse, "Active Noise Control – Fundamentals for Acoustic Design", WIT Press, January 2001, pp. 224-225.

[48]. J. H. Rindel, and C. L. Christensen, "ODEON, A Design tool for noise control in indoor environments", *Proceedings of the Symposium "Noise at work"*, Lille, France 3-5 July, 2007.

[49]. O. Zaleski, W.C. von karstedt, and O. von Erstorff, "Zur modellierung mit Boundary Elementen und Finiten Elementen bei Schallabstrahlungsberechnungen", *Proceedings of the 1st Deustchsprachige Anwenderkonferenz Sysnoise*, Bühlerhöhe, Germany, February 1999.

[50]. ISO 3744, *Acoustics -- Determination of sound power levels of noise sources using sound pressure*, International Organization for Standardization, 1994.

[51]. F. Jacobse, T. Poulsen, J. H. Rindel, A. C. Gade, and M. Ohlirch, *Fundamentals of acoustics and noise control*, Ørsted, DTU, August 2006, pp. 18-20.

[52]. J. Lefévre, U. Gabbert, "Finite Element Modelling of Vibro.Acoustic Systems for Active Noise Reduction", *Technische Mechanik*, December 2005, Vol 25, Issue 3-4, pp. 241-247.

[53]. S. M. Kuo, D. R. Morgan, "Active Noise Control: A Tutorial Review", *Proceedings of the IEEE*, June 1999, Vol. 87, Issue 6, pp 943-973.

[54]. A: Hoepner, C. Belardinelli and E. Bluemel, "Methods and Technologies For Virtual Representation Of Urban Environments To Foster City Development", *Proceeding of the IAPS2008*, Rome, Italy, 28 July – 1 August 2008.

[55]. C. Belardinelli, "The Flexible Audio Extension for the IVS VDT Virtual Reality System and Its Different Uses", *DemoSession, Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA 2007*, New Paltz, New York, US, 21-24 October 2007.

[56]. S. Ringwelski, "Numerische Untersuchung und experimentelle Erprobung von Methoden zur adaptiven Schallreduktion mittels piezoelektrischer Patchaktoren", *Master degree Thesis*, University Otto-von-Guericke, Magdeburg, Germany, October 2006.

[57]. K. K. Parhi and T. Nishitani, "Digital Signal Processing for Multimedia Systems", *Signal Processing Series*, Marcel Dekker Inc, New York, US, 1999, pp. 204-206.

[58]. "Directive 2005/88/EC of the European Parlment and of the Council of 14 December 2005", *Official Journal of the European Union*, Strasbourg, France, 27 December 2005.

[59]. F. Benedetto, G. Giunta, A. Neri, and C. Belardinelli "Digital Audio Watermarking for QoS Assessment of MP3 Music Signals", *Proceeding of the IEEE European Signal Processing Conference, EUSIPCO 2006*, Florence, Italy, September 2006.

[60]. C. Belardinelli, "Interaction of Sound in Virtual Reality: applications to the Digital Factory and other uses", *IV IFF-Kolloquium, Forschung vernetzen – Innovationen beschleunigen*, pp. 73-77, Magdeburg, Germany, September 2007.