Preface

With the development of the VLSI technology, the performance of signal processing devices (DSPs) has greatly improved making possible the implementation of very efficient signal processing algorithms that have had a great impact and contributed in a very important way in the development of large number of industrial fields. One of the fields that has experience an impressive development in the last years, with the use of many signal processing tools, is the telecommunication field. Several important developments have contributed to this fact, such as efficient speech coding algorithm (Bosi & Goldberg, 2002), equalizers (Haykin, 1991), echo cancellers (Amano, Perez-Meana, De Luca, & Duchen, 1995), and so forth. During the last several years very efficient speech coding algorithms have been developed that have allowed reduction of the bit/s required in a digital telephone system from 32Kbits/s, provided by the standard adaptive differential pulse code modulation (ADPCM), to 4.8Kbits/s or even 2.4Kbits/s, provided by some of the most efficient speech coders. This reduction was achieved while keeping a reasonably good speech quality (Kondoz, 1994). Another important development with a great impact on the development of modern communication systems is the echo cancellation (Messershmitt, 1984) which reduces the distortion introduced by the conversion from bidirectional to one-directional channel required in long distance communication systems. The echo cancellation technology has also been used to improve the development of efficient full duplex data communication devices. Another important device is the equalizers that are used to reduce the intersymbol interference, allowing the development of efficient data communications and telephone systems (Proakis, 1985).

In the music field, the advantages of the digital technology have allowed the development of efficient algorithms for generating audio effects such as the introduction of reverberation in music generated in a studio to do it more naturally. Also the signal processing technology allows the development of new musical instruments or the synthesis of musical sounds produced by already available musical instruments, as well as the generation of audio effects required in the movie industry.

The digital audio technology is also found in many consumer electronics equipments to modify the audio signal characteristics such as modifications of the spectral characteristics of audio signal, recoding and reproduction of digital audio and video, edition of digital material, and so forth. Another important application of the digital technology in the audio field is the restoration of old analog recordings, achieving an adequate balance between the storage space, transmission requirements, and sound quality. To this end, several signal processing algorithms have been developed during the last years using analysis and synthesis techniques of audio signals (Childers, 2000). These techniques are very useful for generation of new and already known musical sounds, as well as for restoration of already recorded audio signals, especially for restoration of old recordings, concert recordings, or recordings obtained in any other situation when it is not possible to record the audio signal again (Madisetti & Williams, 1998).

One of the most successful applications of the digital signal processing technology in the audio field is the development of efficient audio compression algorithms that allow very important reductions in the storage requirements while keeping a good audio signal quality (Bosi & Goldberg, 2002; Kondoz, 1994). Thus the researches carried out in this field have allowed the reducing of the 10Mbits required by the WAV format to the 1.41Mbits/s required by the compact disc standard and recently to 64Kbits/s required by the standard MP3PRO. These advances in the digital technology have allowed the transmission of digital audio by Internet, the development of audio devices that are able to store several hundreds of songs with reasonable low memory requirements while keeping a good audio signal quality (Perez-Meana & Nakano-Miyatake, 2005). The digital TV and the radio broadcasting by Internet are other systems that have taken advantage of the audio signal compression technology.

During the last years, acoustic noise problem has become more important as the use of large industrial equipment such as engines, blowers, fans, transformers, air conditioners and motors, and so forth increases. Because of its importance, several methods have been proposed to solve this problem, such as enclosures, barriers, silencers, and other passive techniques that attenuate the undesirable noise (Tapia-Sánchez, Bustamante, Pérez-Meana, & Nakano-Miyatake, 2005; Kuo & Morgan, 1996). There are mainly two types of passive techniques: the first type uses the concept of impedance change caused by a combination of baffles and tubes to silence the undesirable sound. This type, called reactive silencer, is commonly used as mufflers in internal combustion engines. The second type, called resistive silencers, uses energy loss caused by sound propagation in a duct lined with sound-absorbing material. These silencers are usually used in ducts for fan noise. Both types of passive silencers have been successfully used during many years in several applications; however, the attenuation of passive silencers is low when the acoustic wavelength is large compared with the silencer's dimension (Kuo & Morgan, 1996). Recently, with the developing of signal processing technology, during the last several years have been developed efficient active noise cancellation algorithms using single- and multi-channel structures, which use a secondary noise source that destructively interferes with the unwanted noise. In addition, because these systems are adaptive, they are able to track the amplitude, phase, and sound velocity of the undesirable noise, which are in most cases non-stationary. Using the active noise canceling technology, headphones with noise canceling capability, systems to reduce the noise aircraft and cabins, air condition ducts, and so forth have been developed. This technology, which must be still improved, is expected to become an important tool to reduce the acoustic noise problem (Tapia et al., 2005).

Another important field in which the digital signal processing technology has been successfully applied is the development of hearing aids systems, speech enhancement of persons with oral communication problems such as the alaryngeal speakers. In the first case, the signal processing device performs selective signal amplification on some specific frequency bands, in a similar form as an audio equalizer, to improve the patient hearing capacity. While improving the alaryngeal speech several algorithms have been proposed. Some of them intend to reduce the noise produced by the electronic larynx, which is a widely used for alaryngeal persons, while the second group intends to restore the alaryngeal speech providing a more natural voice, at least when a telecommunication system, such as a telephone, is used (Aguilar, Nakano-Miyatake, & Perez-Meana, 2005). Most of these methods are based on patterns recognition techniques.

Several speech and audio signal processing applications described previously, such as the echo and noise canceling; the reduction of intersymbol interference, and the active noise canceling, strongly depend on adaptive digital filters using either time domain or frequency domain realization forms that have been a subject of active research during the last 25 years (Haykin, 1991). However, although several efficient algorithms have been proposed during this time, some problems still remain to be solved, such as the development of efficient IIR adaptive filters, as well as non-linear adaptive filters, which have been less studied in comparison with their linear counter parts.

The development of digital signal processing technology, the widespread use of data communication networks, such as the Internet, and the fact that the digital material can be copied without any distortion, has created the necessity to develop mechanisms that permit the control of the illegal copy and distribution of digital audio, images, and video, as well as the authentication of a given digital material. A suitable way to do that is by using the digital watermarking technology (Bender, Gruhl, Marimoto, & Lu, 1996; Cox, Miller, & Bloom, 2001).

Digital watermarking is a technique used to embed a collection of bits into a given signal, in such way that it will be kept imperceptible to users and the resulting watermarked signal remains with nearly the same quality as the original one. Watermarks can be embedded into audio, image, video, and other formats of digital data in either the temporal or spectral domains. Here the temporal watermarking algorithms embed watermarks into audio signals in their temporal domain, while the spectral watermarking algorithms embed watermarks in certain transform domain. Depending on their particular application, the watermarking algorithms can be classified as robust and fragile watermarks, where the robust watermarking algorithms are used for copyright protection, distribution monitoring, copy control, and so forth, while the fragile watermark, which will be changed if the host audio is modified, is used to verify the authenticity of a given audio signal, speech signal, and so forth. The watermarking technology is expected to become a very important tool for the protection and authenticity verification of digital audio, speech, images, and video (Bender et al., 1996; Cox et al., 2001).

Another important application of the audio and speech signal processing technology is the speech recognition, which has been a very active research field during the last 30 years; as a result, several efficient algorithms have been proposed in the literature (Lee, Soong, & Paliwal, 1996; Rabiner & Biing-Hwang, 1993). As happens in most pattern recognition algorithms, the pattern under analysis, in this case the speech signal, must be characterized to extract the most significant as well as invariant features, which are then fed into the recognition stage. To this end several methods have been proposed, such as the linear predictions coefficients (LPC) of the speech signal and LPC-based cepstral coefficients, and recently the used phonemes to characterize the speech signal, instead of features extracted from its waveform, has attracted the attention of some researchers. A related application that also has been widely studied consists of identifying not the spoken voice, but who spoke it. This application, called speaker recognition, has been a subject of active research because of its potential applications for access control to restricted places or information. Using a similar approach it is possible also to identify natural or artificial sounds (Hattori, Ishihara, Komatani, Ogata, & Okuno, 2004). The sound recognition has a wide range of applications such as failure diagnosis, security, and so forth.

This book provides a review of several signal processing methods that have been successfully used in speech and audio fields. It is intended for scientists and engineers working in enhancing, restoration, and protection of audio and speech signals. The book is also expected to be a valuable reference for graduate students in the fields of electrical engineering and computer science.

The book is organized into XIV chapters, divided in four sections. Next a brief description of each section and the chapters included is provided.

Chapter I provides an overview of some the most successful applications of signal processing algorithms in the speech and audio field. This introductory chapter provides an introduction to speech and audio signal analysis and synthesis, audio and speech coding, noise and echo canceling, and recently proposed signal processing methods to solve several problems in the medical field. A brief introduction of watermarking technology as well as speech and speaker recognition is also provided. Most topics described in this chapter are analyzed with more depth in the remaining chapters of this book.

Section I analyzes some successful applications of the audio and speech signal processing technology, specifically in applications regarding the audio effects, audio synthesis, and restoration. This section consists of three chapters, which are described in the following paragraphs.

Chapter II presents the application of digital filters for introducing several effects in the audio signals, taking into account the fact that the audio editing functions that change the sonic character of a recording, from loudness to tonal quality, enter the realm of *digital signal processing* (DSP), removing parts of the sound, such as noise, and adding to the sound elements that were not present in the original recording, such as reverb, improving the music in a studio, which sometimes does not sound as natural as for example music performed in a concert hall. These and several other signal processing techniques that contribute to improve the quality of audio signals are analyzed in this chapter.

Chapter III provides a review of audio signal processing techniques related to sound generation via additive synthesis, in particular using the sinusoidal modeling. Here, firstly the processing stage required to obtaining a sinusoidal representation of audio signals is described. Next, suitable synthesis techniques that allow reconstructing an audio signal, based on a given parametric representation, are presented. Finally, some audio applications where sinusoidal modeling is successfully employed are briefly discussed.

Chapter IV provides a review of digital audio restoration techniques whose main goal is to use digital signal processing techniques to improve the sound quality, mainly, of old recordings, or the recordings that are difficult to do again, such as a concert. Here a conservative goal consists on eliminating only the audible spurious artifacts that either are introduced by analog recording and playback mechanisms or result from aging and wear of recorded media, while retaining as faithfully as possible the original recorded sound. Less restricted approaches are also analyzed, which would allow more intrusive sound modifications, such as elimination of the audience noises and correction of performance mistakes in order to obtain a restored sound with better quality than the original recording.

Section II provides an analysis of recently developed speech and audio watermarking methods. The advance in the digital technology allows an error free copy of any digital material, allowing the unauthorized copying, distribution, and commercialization of copyrighted digital audio, images, and videos. This section, consisting of two chapters, provides an analysis of the watermarking techniques that appear to be an attractive alternative to solving this problem.

Chapters V and **VI** provide a comprehensive overview of classic watermark embedding, recovery, and detection algorithms for audio and speech signals, providing also a review of the main factors that must be considered to design efficient audio watermarking systems together with some typical approaches employed by existing watermarking algorithms. The watermarking techniques, which can be divided into robust and fragile, presented in these chapters, are presently deployed in a wide range of applications including copyright protection, copy control, broadcast monitoring, authentication, and air traffic control. Furthermore, these chapters describe the signal processing, geometric, and protocol attacks together with some of the existing benchmarking tools for evaluating the robustness performance of watermarking techniques as well as the distortion introduced in the watermarked signals.

Section III. The adaptive filtering has been successfully used in the solution of an important amount of practical problems such as echo and noise canceling, active noise canceling, speech enhancement, adaptive pulse modulation coding, spectrum estimation, channel equalization, and so forth. Section III provides a review of some successful adaptive filter algorithms, together with two of the must successful applications of this technology such as the echo and active noise cancellers. Section III consists of four chapters, which are described in the following paragraphs.

Chapter VII provides an overview of adaptive digital filtering techniques, which are a fundamental part of echo and active noise canceling systems provided in Chapters VIII and IX, as well as of other important telecommunications systems, such as equalizers, widely used in data communications, coders, speech and audio signal enhancement, and so forth. This chapter presents the general framework of adaptive filtering together with two of the most widely used adaptive filter algorithms—the LMS (least-mean-square) and the RLS (recursive least-square) algorithms—together with some modification of them. It also provides a review of some widely used filter structures, such as the transversal FIR filter, the transform-domain implementations, multirate structures and IIR filters realization forms, and so forth. Some important audio applications are also described.

Chapter VIII presents a review of the echo cancellation problem in telecommunication and teleconference systems, which are two of the most successful applications of the adaptive filter technology. In the first case, an echo signal is produced when mismatch impedance is present in the telecommunications system, due to the two-wires-to-four-wires transformation required because the amplifiers are one-directional devices, and as a consequence a portion of the transmitted signal is reflected to the transmitter as an echo that degrades the system quality. A similar problem affects the teleconference systems because of the acoustical coupling between the speakers and microphones, in each room, used in such systems. To avoid the echo problem in both cases, an adaptive filter is used to generate an echo replica, which is then subtracted from the signal to be transmitted. This chapter analyzes the factors to consider in the development of efficient echo canceller systems, such as the duration of the echo canceller impulse response, the convergence rate of adaptive algorithm, and computational complexity, because these systems must operate in real time, and how to handle the simultaneous presence of both the echo signal and the near end speaker voice.

Chapter IX provides a review of the active noise cancellation problem together with some of its most promising solutions. In this problem, which is closely related with the echo canceling, adaptive filters are used to reduce the noise produced in automotive equipment, home appliances, industrial equipment, airplanes cabin, and so forth. Here active noise canceling is achieved by introducing an antinoise wave through an appropriate array of secondary sources, which are interconnected through electronic adaptive systems with a particular cancellation configuration. To properly cancel the acoustic noise signal, the adaptive filter generates an antinoise, which is acoustically subtracted from the incoming noise wave. The resulting wave is then captured by an error microphone and used to update the adaptive filter coefficients such that the total error power is minimized. This chapter analyzes the filter structures and adaptive algorithms, together with other several factors to be considered in the development of active noise canceling systems; this chapter also presents some recently proposed ANC structures that intend to solve some of the already existent problems, as well as a review of some still remaining problems that must be solved in this field.

Chapter X presents a recurrent neural network structure for audio and speech processing. Although the performance of this artificial neural network, called differentially fed artificial neural network, was evaluated using a prediction configuration, it can be easily used to solve other non-linear signal processing problems.

Section IV. The speech recognition has been a topic of active research during the last 30 years. During this time a large number of efficient algorithms have been proposed, using hidden Markov models, neural networks, and Gaussian mixtures models, among other several paradigms to perform the recognition tasks. To perform an accurate recognition task, besides the paradigm used in the recognition stage, the feature extraction has also great importance. A related problem that has also received great attention is the speaker recognition, where the task is to determine the speaker identity, or verify if the speaker is who she/he claims to be. This section provides a review of some of the most widely used feature extraction algorithms. This section consists of four chapters that re described in the following paragraphs.

Chapters XI and **XII** present the state-of-the-art automatic voice recognition (ASR), which is related to multiple disciplines, such as processing and analysis of speech signals and mathematical statistics, as well as applied artificial intelligence and linguistics among some of the most important. The most widely used paradigm for speech characterization in the developing of ASR has been the phoneme as the essential information unit. However, recently the necessity to create more robust and versatile systems for speech recognition has suggested the necessity of looking for different approaches that may improve the performance of phoneme based ASR. A suitable approach appears to be the use of more complex units such as syllables, where the inherent problems related with the use of phonemes are overcome to a greater cost of the number of units, but with the advantage of being able to approach using the form in which really the people carry out the learning and language production process. These two chapters also analyze the voice signal characteristics in both the time frequency and domain, the measurement and extraction of the parametric information that characterizes the speech signal, together with an analysis of the use of artificial neuronal networks, vector quantification, hidden Markov models, and hybrid models to perform the recognition process.

Chapter XIII presents the development of an efficient speaker recognition system (SRS), which has been a topic of active research during the last decade. SRSs have found a large number of potential applications in many fields that require accurate user identification or user identity verification, such as shopping by telephone, bank transactions, access control to restricted places and information, voice mail and law enforcement, and so forth. According to the task that the SRS is required to perform, it can be divided into speaker identification system (SIS) or speaker verification systems (SVS), where the SIS has the task to determine the most likely speaker among a given speakers set, while the SVS has the task of deciding if the speaker is who she/he claims to be. Usually a SIS has M inputs and N outputs, where M depends on the feature vector size and N on the size of the speaker set, while the SVS usually has M inputs, as the SRS, and two possible outputs (accept or reject) or in some situations three possible outputs (accept, reject, or indefinite). Together with an overview of SRS, this chapter analyzes the speaker features extraction methods, closely related to those used in speech recognition presented in Chapters XI and XII, as well as the paradigms used to perform the recognition process, such as vector quantizers (VQ), artificial neural networks (ANN), Gaussian mixture models (GMM), fuzzy logic, and so forth.

Chapter XIV presents the use of speech recognition technologies in the development of a language therapy for children with hearing disabilities; it describes the challenges that must be addressed to construct an adequate speech recognizer for this application and provides the design features and other elements required to support effective interactions. This chapter provides to developers and educators the tools required to work in the developing of learning methods for individuals with cognitive, physical, and sensory disabilities.

Advances in Audio and Speech Signal Processing: Technologies and Applications, which includes contributions of scientists and researchers of several countries around the world and analyzes several important topics in the audio and speech signal processing, is expected to be a valuable reference for graduate students and scientists working in this exciting field, especially those involved in the fields of audio restoration and synthesis, watermarking, interference cancellation, and audio enhancement, as well as in speech and speaker recognition.

References

Aguilar, G., Nakano-Miyatake, M., & Perez-Meana, H. (2005). Alaryngeal speech enhancement using pattern recognition techniques. *IEICE Trans. Inf. & Syst., E88-D*(7), 1618-1622.

Amano, F., Perez-Meana, H., De Luca, A., & Duchen, G. (1995). A multirate acoustic echo canceler structure. *IEEE Trans. on Communications, 43*(7), 2173-2176.

Bender, W., Gruhl, D., Marimoto, N., & Lu. (1996).Techniques for data hiding. *IBM Systems Journal, 35*, 313-336.

Bosi, M., & Goldberg, R. (2002). *Introduction to digital audio coding and standards.* Boston: Kluwer Academic Publishers.

Childers, D. (2000). *Speech processing and synthesis toolboxes.* New York: John Wiley & Sons.

Cox, I., Miller, M., & Bloom, J. (2001). *Digital watermark: Principle and practice.* New York: Morgan Kaufmann.

Hattori, Y., Ishihara, K., Komatani, K., Ogata, T., & Okuno, H. (2004). Repeat recognition for environmental sounds. In *Proceedings of IEEE International Workshop on Robot and Human Interaction* (pp. 83-88).

Haykin, S. (1991). *Adaptive filter theory.* Englewood Cliffs, NJ: Prentice Hall.

Kondoz, A. M. (1994). *Digital speech.* Chinchester, England: Wiley & Sons.

Kuo, S., & Morgan, D. (1996). *Active noise control system: Algorithms and DSP implementations.* New York: John Wiley & Sons.

Lee, C., Soong, F., & Paliwal, K. (1996). *Automatic speech and speaker recognition.* Boston: Kluwer Academic Publishers.

Madisetti, V., & Williams, D. (1998). *The digital signal processing handbook.* Boca Raton, FL: CRC Press.

Messershmitt, D. (1984). Echo cancellation in speech and data transmission. *IEEE Journal of Selected Areas in Communications, 2*(3), 283-297.

Perez-Meana, H., & Nakano-Miyatake, M. (2005). Speech and audio signal applications. In *Encyclopedia of information science and technology* (pp. 2592-2596). Idea Group.

Proakis, J. (1985). *Digital communications.* New York: McGraw Hill.

Rabiner, L., & Biing-Hwang, J. (1993). *Fundamentals of speech recognition.* Englewood Cliff, NJ: Prentice Hall.

Tapia-Sánchez, D., Bustamante, R., Pérez-Meana, H., & Nakano-Miyatake, M. (2005). Single channel active noise canceller algorithm using discrete cosine transform. *Journal of Signal Processing, 9*(2), 141-151.