Signal Processing Techniques for Audio and Speech Applications

Hector Perez-Meana
National Polytechnic Institute, Mexico

Mariko Nakano-Miyatake
National Polytechnic Institute, Mexico

INTRODUCTION

Since the apparition of the first standalone digital signal processor (DSP) in 1980, the development of very-large-scale integration (VLSI) technology has allowed an impressive improvement on the performance of signal processing devices. This fact has made it possible to implement more efficient systems for storage, transmission, enhancement, protection, and reproduction of speech and audio signals. Some of these successful applications, shown in Table 1, have contributed to improving the performance of communications, storage, and medical systems, as well as security and copyright protection.

BACKGROUND

Since the apparition of electronics technology, several devices have been introduced to improve received, produced, and recorded audio and speech signal quality—such as that of low pass and band pass analog filters, amplifiers, and so forth—that allowed suppression of some kinds of interfering signals. Next, with the development of solid state technology in the late of 1960s and the apparition of the op-amp, several analog signal processes were developed; although there were limitations, these allowed import improvements in audio and speech systems. The limitations of analog technology encouraged the development of digital technology. As a result in 1980 the Nippon Electric Corporation (NEC) and American Telegraph and Telephone (AT&T) released

the first standalone complete digital signal processors, the PD7220 and the DSP1. Since then, digital signal processing technology has experienced impressive growth, allowing performance improvement of already available systems, as well as development of many other successful systems in several other fields, some described in this article.

REVIEW OF MAIN SIGNAL PROCESSING APPLICATIONS IN SPEECH AND AUDIO FIELDS

As mentioned before, the signal processing applications in speech and audio fields have increased, contributing to the solution of many important problems, constituting an important part of many practical systems. To understand the importance of this technology and how it has contributed in the development of audio and speech fields, this section provides a review of some successful signal processing systems.

Echo Cancellation for Long-Distance Transmission

A very successful speech signal processing application is the adaptive echo cancellation used to reduce a common but undesirable phenomenon in most telecommunications systems, called echo. Here when mismatch impedance is present in any telecommunications system, a portion of the transmitted signal is reflected to the transmitter as an echo, which represents an impairment that degrades the system quality (Hansler & Schmidt, 2006; Manolakis, Ingle, & Kogon, 2005). In most telecommunications systems, such as a telephone circuit, the echo is generated when the long-distance portion consisting of two one-directional channels (four wires) is connected with a bidirectional channel (two wires) by means of a hybrid transformer. If the hybrid impedance is perfectly balanced, the two one-directional channels are uncoupled and no signal is returned to the transmitter side (Hansler & Schmidt, 2006; Manolakis et al., 2005). However in general the bridge is not perfectly balanced because the

Table 1. Main audio and speech signal processing applications

• Echo cancellation in telecommunication systems
• Acoustic echo cancellation
• Noise canceling
• Adaptive equalization
• Narrowband speech coding
• Broadband audio and speech coding
• Medical applications
• Watermarking

required impedance to properly balance the hybrid depends on the overall impedance network. In this situation, part of the signal is reflected, producing an echo. To avoid this problem an adaptive filter is used to generate an echo replica, which is then subtracted from the signal to be transmitted. Subsequently, the adaptive filter coefficients are updated to minimize, usually, the mean square value of the residual echo (Madisetti & Williams, 1998; Perez-Meana, 2007). To obtain an appropriate operation, the echo canceller impulse response must be larger than the longer echo path to be estimated. Thus assuming a sampling frequency of 8kHz and an echo delay of about 60ms, an echo canceller with 256 or more taps is required (Manolakis et al., 2005; Perez-Meana, 2007). Besides the echo path estimation, another important problem is how to handle the double talk—that is, the simultaneous presence of the echo and the near speech signal (Hansler & Schmidt, 2006; Manolakis et al., 2005).

Acoustic Echo Cancellation

A critical problem affecting speech communication in a teleconferencing system is the acoustic echo. When a bi-directional line links two rooms, the acoustic coupling between the loudspeaker and microphones in each room causes an acoustic echo perceivable to the users in the other room. The best way to handle it appears to be the adaptive echo cancellation. An acoustic echo canceller generates an echo replica and subtracts it from the signal picked up by the microphones. The residual echo is then used to update the filter coefficients such that the mean square value of approximation error is kept to a minimum (Perez Meana, Nakano-Miyatake, & Nino-de-Rivera, 2002.; Huang & Banesty, 2005).

Adaptive Noise Cancellation

The adaptive noise canceller is a generalization of the echo canceller in which a signal corrupted with additive noise must be restored or enhanced. When a reference signal correlated with the noise signal but uncorrelated with the desired one is available, the noise cancellation can be achieved by using an adaptive filter to minimize the total power of the output of the difference between the corrupted signal and the estimated noise, such that the resulting signal becomes the best estimate, in the mean square sense, of the desired signal. This system works fairly well when the reference and the desired signal are uncorrelated among them. However, appropriate reference signals are not always available. To solve this problem several noise canceling algorithms have been proposed which are resistant to crosstalk situations. A different approach, developed by Dolby Laboratories, is used in the Dolby noise reduction systems in which the dynamic range of the sound is reduced during recording and expanded during the playback (Davis, 2002; Hansler & Schmidt, 2006). Several types of Dolby noise reduction systems have been developed including the A, B, C, and HXpro. Most widely used is the Dolby B, which allows acceptable playback even on devices without noise reduction. The Dolby B noise reduction system uses a pre-emphasis that allows masking the background hiss of a tape with a stronger audio signal, especially at higher frequencies. This effect is called psychoacoustic masking (Davis, 2002; Perez-Meana, 2007).

A related problem to noise cancellation is the active noise cancellation that intends to reduce the noise produced in closed places by several electrical and mechanical pieces of equipment such as home appliances, industrial equipment, air conditioning units, airplanes turbines, motors, and so forth. Active noise is achieved by introducing a canceling

Table 2. Digital speech coding standards

Rate Kb/s	Application	Type of Coder	Year
64	Public Switched Telephone Network	Pulse Code Modulation (PCM)	1972
2.4	U.S. Government Federal Standard	Linear Predictive Coding	1977
32	Public Switched Telephone Network	Adaptive Differential PCM	1984
9.6	Skyphone	Multi-Pulse Linear Predictive Coding (MPLPC)	1990
13	Pan-European Digital Mobile Radio (DMR) Cellular System (GSM)	Regular Pulse Excitation Linear Prediction Coding (RPE-LPC)	1991
4.8	U.S. Government Federal Standard	Codebook Excited Linear Prediction Coding (CELP)	1991
16	Public Switched Telephone Network	Low Delay CELP (LD-CELP)	1992
6.7	Japanese Digital Mobile Radio (DMR)	Vector Sum Excited Linear Prediction Coding (VSELP)	1977

Related Content

Business Technology Strategy for a Major Real Estate and Mortgage Brokerage Company
Charlton Monsantoand Stephen J. Andriole (2010). *Journal of Information Technology Research (pp. 43-53).*
www.irma-international.org/article/business-technology-strategy-major-real/47216

User Opinion and Differentiated Attribute based Ranking in Federated Cloud
C. S. Rajarajeswariand M. Aramudhan (2016). *Journal of Information Technology Research (pp. 78-88).*
www.irma-international.org/article/user-opinion-and-differentiated-attribute-based-ranking-in-federated-cloud/160157

Surveying Mobile Commerce Environments
Jari Veijalainenand Mathias Weske (2005). *Encyclopedia of Information Science and Technology, First Edition (pp. 2702-2711).*
www.irma-international.org/chapter/surveying-mobile-commerce-environments/14679

Engineering Emotionally Intelligent Agents
Penny Baillie-de Byland Mark Toleman (2005). *Encyclopedia of Information Science and Technology, First Edition (pp. 1052-1056).*
www.irma-international.org/chapter/engineering-emotionally-intelligent-agents/14385

See-Through-Sound: Transforming Images into Sonic Representations to Help the Blind
J. Tomás Henriques, Sofia Cavacoand Nuno Correia (2014). *Journal of Information Technology Research (pp. 59-77).*
www.irma-international.org/article/see-through-sound/111252