SONY

Superb Sound Technology for Sony HD Visual Communication System

IPELA ™

PCS-HG90	All
PCS-XG55S	All
PCS-XG80S	All

Introduction

The Sony HG90/XG55S/XG80S HD Visual Communication System realizes two-way communication using MPEG-4 AAC stereo sound with a high sampling frequency of 96 kHz for the PCS-HG90 and 48 kHz for the PCS-XG55S/XG80S as a standard feature. This covers a frequency band exceeding the 44.1 kHz sampling frequency of compact discs (CDs), which are known for their high-quality sound. The system is equipped with external input terminal(s), enabling music to be delivered in high-quality stereo sound.[1] It is also equipped with a wideband stereo echo canceller.[2]

With these features, the PCS-HG90/XG55S/XG80S provides realistic, clear and natural sound appropriate for high-quality HD video. The PCS-XG55S/XG80S incorporates these features into an extremely compact package.

This document describes the stereo echo canceller technology, multi-rate signal processing technology and new technologies adopted by the PCS-XG55S/XG80S, which are indispensable for realizing these features.

[1]: The PCS-HG90 can also transmit ultrawideband audio such as Super Audio CD (SACD) and DVD Audio.

[2]: The echo canceller works at an internal sampling frequency of 48kHz (for the PCS-HG90) or 32kHz (for the PCS-XG55S/XG80S).

Stereo Echo Canceller Technology

This product is equipped with a wideband stereo echo canceller.

What Is The Acoustic Echo Canceller?

In videoconferencing, for example, in the case of two-way hands-free communication, the voice of a participant is sent to the remote site. The audio is output from the loudspeaker at the remote site, where it is picked up by the microphone and sent back to the loudspeaker at the local site. In other words, when you speak into the microphone, you hear your own voice from the loudspeaker as an echo. Known as acoustic echo, this phenomenon is considered annoying and can disrupt communication. In the worst cases, howling is created by feedback, making communication impossible. Therefore, acoustic echo canceller technology that eliminates sound from the loudspeaker and prevents it from being sent back to the remote site is indispensable for realizing full-duplex communication in two-way communication systems using loudspeakers and microphones.

Difference Between Monaural And Stereo

Adaptive filtering technology, which is at the core of the echo canceller, calculates the characteristics between loudspeakers and microphones and performs processing to cancel the sound of the loudspeaker picked up by the microphone.

For conventional monaural audio, systems consist of one loudspeaker and one microphone, which means only one adaptive filter is required because there is only one relationship. For stereo sound, systems consist of two (right and left) loudspeakers and two (right and left) microphones, meaning four adaptive filters are required because there are four relationships. Compared to the monaural echo canceller, the stereo echo canceller needs to perform four times the amount of calculation on a conceptual basis. Sony's unique technology overcomes this difficulty.

PCS-HG90/XG55S/XG80S's Stereo Echo Canceller

The PCS-HG90 realizes high-speed calculation by incorporating five sets of high-performance DSP[*3] (Digital Signal Processors) with large-capacity memory. Meanwhile, the PCS-XG55S/XG80S incorporates two sets of high-performance DSP[*3] with large-capacity memory. Building upon the high-end model PCS-HG90's technology with a new algorithm developed specifically for the PCS-XG series, along with optimized processing of DSP, the PCS-XG55S/XG80S achieves a compact size and natural-sounding audio that compares favorably with the PCS-HG90, without affecting the performance of the stereo echo canceller, which requires an enormous amount of processing.

[*3]: Texas Instruments, Inc.'s TMS320C6727 with external 16-MB memory.

Long-latency Speakers

Flat-panel TVs such as LCDs and plasma displays perform a variety of internal video signal processing to reproduce a high-quality picture. However, this video signal processing can result in a delayed picture output (150 ms or more in some cases), causing the audio and video to become out of sync. To prevent this, their TVs insert the same amount of latency (lip sync) to the loudspeakers' output for video and make them synchronize.

However, conventional echo cancellers cannot eliminate echo when long latency speakers are used. They attempt to recognize the latency as an elongated echo. But the echo time including the latency is too long to cover with the adaptive filtering, hence the echo cannot be fully eliminated.

The PCS-HG90/XG55S/XG80S's echo canceller, however, resolved this problem. The latency amount can be preset according to your TV speakers. This ensures the echo canceller operates properly and echo is eliminated reliably, even when the product is used in conjunction with high-quality televisions.
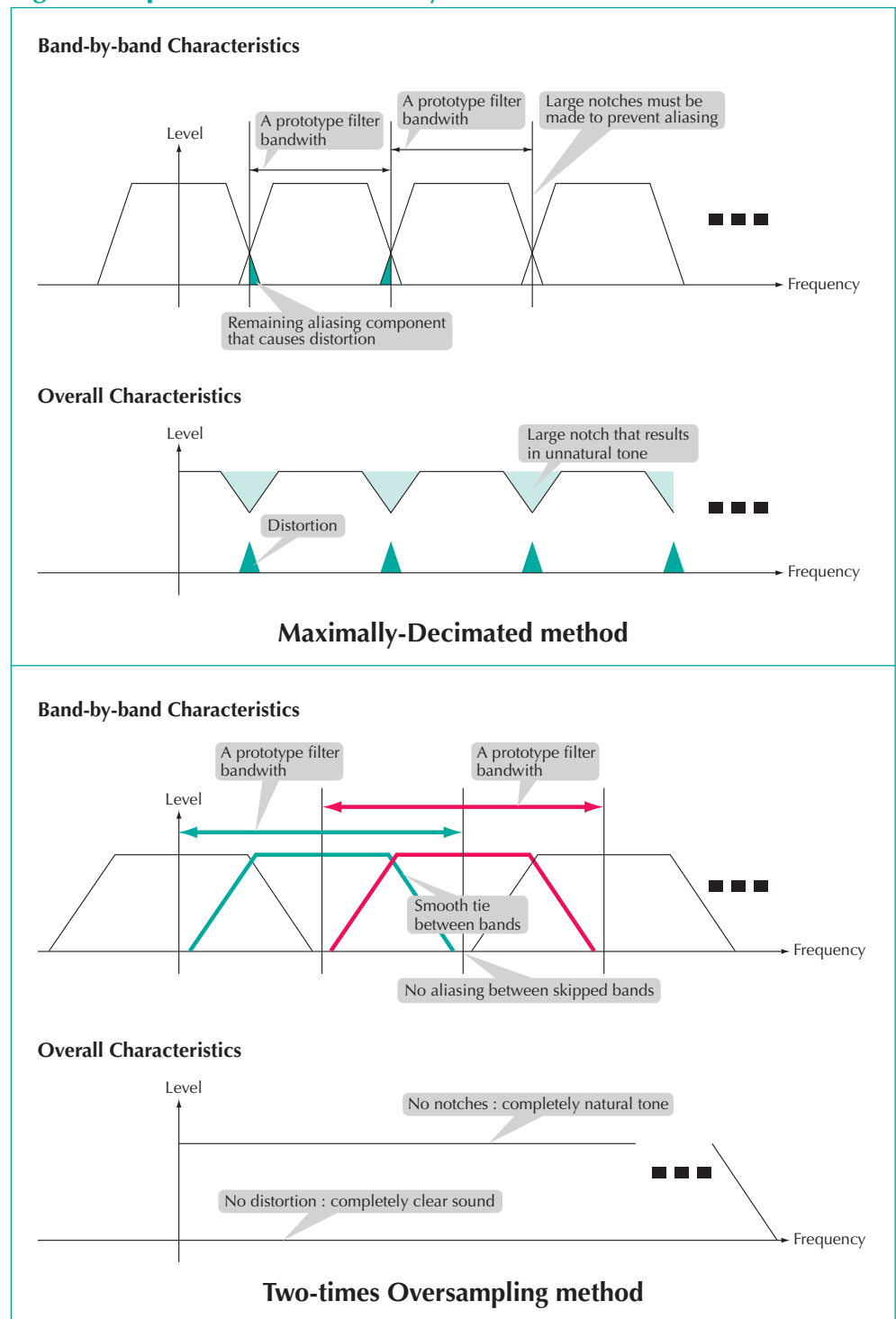
Multi-rate Signal Processing Technology

One way to reduce the calculation amount is to use multi-rate signal processing technology, which allows downsampling to lower sampling frequencies by splitting the frequency band for processing. For example, if the sampling frequency is halved by downsampling, the processing amount of the adaptive filter is halved.

It is known that the quality of the adaptive filter deteriorates if aliasing occurs during downsampling. An anti-aliasing filter is applied before downsampling to prevent this aliasing, but the characteristics of the filter affect the sound quality. When using the "maximally-decimated" method that can most reduce the calculation amount, one of the filter characteristics is that the stop band starts from the border with the adjacent split bands, thereby sacrificing part of the pass band to obtain sufficient attenuation within the stop band. As a result, when band splitting/synthesis is performed, notches are formed around the borders between adjacent bands, which adversely affects the sound quality.

To Realize Natural Sound

The PCS-HG90/XG55S/XG80S, which is designed with sound quality as the top priority, also addresses this problem. To solve the problem of the maximally-decimated method, a two-times oversampling method is adopted. (See Figure 1.) Although the calculation amount with this method becomes twice as large for adaptive filtering calculation, the stop band can be doubled, which makes the boundary between adjacent bands flatter. In addition, aliasing is prevented. As a result, more natural, clearer sound with less distortion compared to the maximally-decimated method is realized.

Fig.1: Comparison with Maximally-Decimated Method

Band-by-band Characteristics

Level

A prototype filter bandwith

A prototype filter bandwith

Large notches must be made to prevent aliasing

Frequency

Remaining aliasing component that causes distortion

Overall Characteristics

Level

Large notch that results in unnatural tone

Distortion

Frequency

Maximally-Decimated method

Band-by-band Characteristics

A prototype filter bandwith

A prototype filter bandwith

Level

Smooth tie between bands

Frequency

No aliasing between skipped bands

Overall Characteristics

Level

No notches : completely natural tone

No distortion : completely clear sound

Frequency

Two-times Oversampling method

New Technology For High-performance Stereo Echo Canceller (PCS-XG series)

Improved Response to Sound Field Changes

"Sound field changes" refers to the changes in characteristics that occur between microphones and loudspeakers when rearranging microphones and/or changing the amplifier volume. The PCS-XG55S/XG80S's echo canceller uses an innovative residual echo suppression technology and a new reinforcement learning algorithm for the acoustic environment, which improve the performance for returned echo and realize quicker echo reduction. (See Figure 2.)

Enhanced Double-talk Suppression

"Double-talk" refers to the situation in which both parties speak at the same time. The PCS-XG55S/XG80S's innovative residual echo suppression technology significantly reduces the residual echo, resulting in a more natural, comfortable conversation. (See Figure 3.)
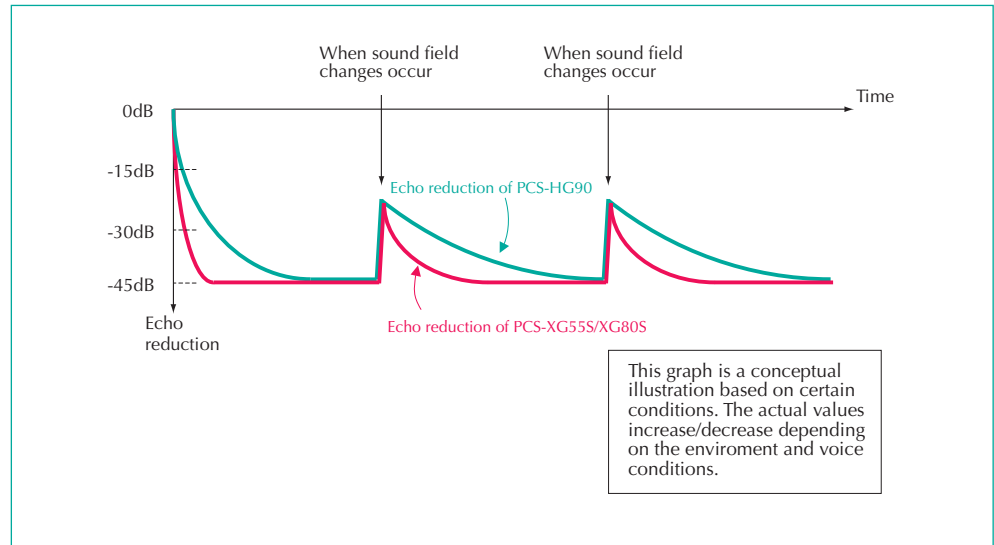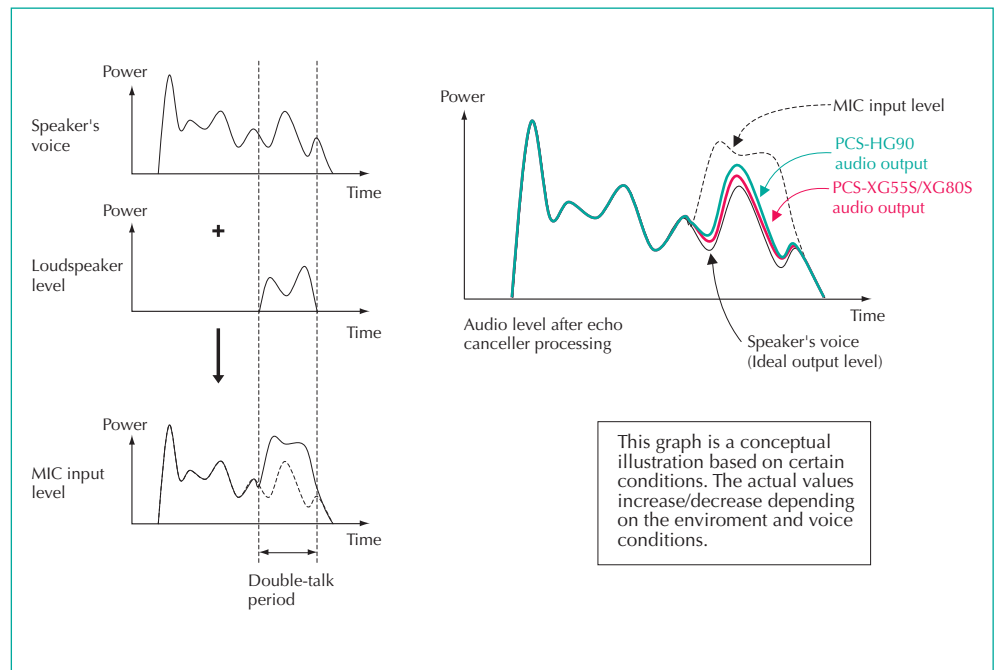
Fig.2: Response To Sound Field Changes

When sound field changes occur

When sound field changes occur

Time

0dB

-15dB

Echo reduction of PCS-HG90

-30dB

-45dB

Echo reduction

Echo reduction of PCS-XG55S/XG80S

This graph is a conceptual illustration based on certain conditions. The actual values increase/decrease depending on the enviroment and voice conditions.

Fig.3: Echo Suppression During Double-talk Period

Power

Speaker's voice

Time

Power

+

Loudspeaker level

Time

Power

MIC input level

Time

Double-talk period

Power

MIC input level

PCS-HG90 audio output

PCS-XG55S/XG80S audio output

Time

Audio level after echo canceller processing

Speaker's voice (Ideal output level)

This graph is a conceptual illustration based on certain conditions. The actual values increase/decrease depending on the enviroment and voice conditions.

IPELA is a trademark of Sony Corporation.

SONY