

SONY

VIDEO COMMUNICATION SYSTEM-TECHNICAL DOCUMENTATION

Superb Sound Technology for the PCS-HG90

IPELA

PCS-HG90 All

Introduction

Sony's PCS-HG90 HD Visual Communication System realizes two-way communication using MPEG-4 AAC Stereo sound ^{*1} with an industry-leading sampling frequency of 96 kHz in the videoconferencing product business as a standard feature. The system also enables music to be delivered in high-quality stereo sound. The external input covers a frequency band of more than twice the 44.1 kHz sampling frequency of compact discs (CDs), which are known for their high-quality sound. The system can also transmit ultrawideband audio such as Super Audio CD (SACD) and DVD Audio.^{*2}

With these features, the PCS-HG90 provides realistic, clear and natural sound appropriate for high-quality HD video.

This document describes the stereo echo canceller technology and multi-rate signal processing technology, which are indispensable for realizing these features.

^{*1}: Three times the sampling frequency of our conventional products such as the PCS-G70/G70P.

^{*2}: The echo canceller works at an internal sampling frequency of 48kHz, still communication accomplishes at 96kHz for the mixed audio (Mic+Aux).

Stereo Echo Canceller Technology

This product is equipped with a wideband (sampling frequency: 48 kHz) 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 speaker at the remote site, where it is picked up by the microphone and sent back to the speaker at the local site. In other words, when you speak into the microphone, you hear your own voice from the speaker 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 speaker and prevents it from being sent back to the remote site is indispensable for realizing full-duplex communication in two-way communication systems using speakers and microphones.

Difference Between Monaural And Stereo

Adaptive filtering technology, which is at the core of the echo canceller, calculates the characteristics between speakers and microphones and performs processing to cancel the sound of the speaker picked up by the microphone. For conventional monaural audio, systems consist of one speaker and one microphone, which means only one adaptive filter is required because there is only one relationship. For stereo sound, systems consists of two (right and left) speakers 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.

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 speakers' 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 resolved this problem.

PCS-HG90's Stereo Echo Canceller

One approach to reducing costs is to reduce the huge amount of calculation at the expense of sound quality. Instead, the PCS-HG90 realizes high-speed calculation by incorporating five sets of high-performance DSP *3 (Digital Signal Processors) with large-capacity memory, achieving natural-sounding audio appropriate for high-end models.

The PCS-HG90 also supports long-latency speakers. The latency amount of your TV speakers can be preset. This is a PCS-HG90's original feature, and it ensures the echo canceller operates properly and echo is eliminated reliably, even when the product is used in conjunction with high-quality televisions.

*3: Five sets of Texas Instruments, Inc.'s TMS320C6727 with external 16-MB memory.

About 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.

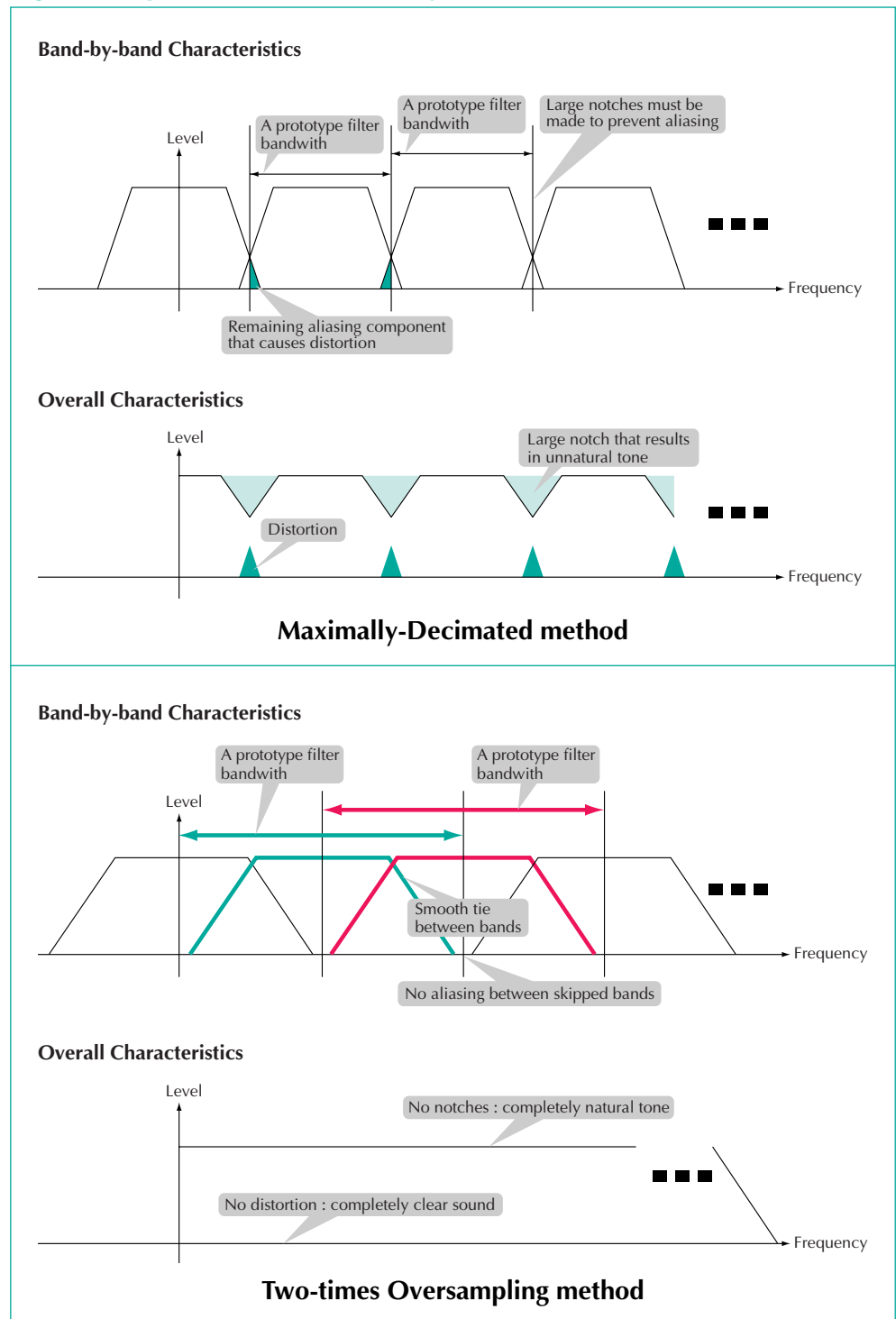
It is known that the quality of the adaptive filter deteriorates if aliasing occurs during downsampling. An anti-aliasing filter is applied before downsampling to prevent this aliasing, but the characteristics of the filter affect the sound quality. When using the "maximally-decimated" method that can 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, 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 Fig.1)

Although this method costs twice as much 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



SONY