

1

APPARATUS AND METHOD FOR IMPROVED VOICE ACTIVITY DETECTION

TECHNICAL FIELD

This invention relates to the transmission of digitally encoded voice, and in particular, to the transmission of digitally encoded voice so as to maintain speech quality.

BACKGROUND OF THE INVENTION

Because of the popularity of the Internet, a growing need for remote access, and the increase in data traffic volume that has exceeded the voice traffic volume through the voice and data communication networks, the transmission of voice as data rather than circuit switched voice is becoming more important. The problem that exists when voice is transmitted as data such as voice-over-packet technology or voice-over-the-Internet is to guarantee the quality of service. To reduce the bandwidth required to carry voice, voice-over-packet systems employ a voice activity detection to suppress the packetization of voice signals between individual speech utterances such as the silent periods in a voice conversation. Such techniques adapt to varying levels of noise and converge on appropriate thresholds for a given voice conversation. Use of voice activity detection reduces the required bandwidth of an aggregation of channels 50% to 60% for conversations that are essentially half-duplex, only one person speaks at a time in a half-duplex conversation.

When silence suppression is being used, a noise generator at the receiving end compliments the suppression of silence at the transmitting end by generating a local noise signal during the silent periods rather than muting the channel or playing nothing. Muting the channel gives the listener the unpleasant impression of a dead line. The match between the generated noise and the true background noise determines the quality of the noise generator.

Within the prior art, it is well known that voice activity detection to determine silence and the removal of those silent periods can cause speech utterances to sound choppy and unconnected when cutting in or out of the speech. Two terms are utilized to express this problem. First, front-end clipping refers to clipping the beginning of an utterance. Second, holdover time refers to the time the activity detector continues to packetize speech after the voice signal level falls below the speech threshold. The holdover time is normally set to the period between words as has been determined for a particular conversation so as to avoid front-end clipping at the beginning of each word. However, excessive holdover times reduce network efficiency and too little causes speech to sound choppy.

SUMMARY OF THE INVENTION

This invention is directed to solving these and other problems and disadvantages of the prior art. In an embodiment of the invention, the problems of front-end clipping and excessively long holdover times is resolved by the introduction of a history queue at the transmitting end of the digital conversation.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 illustrates an embodiment of the invention;
FIG. 2 illustrates an embodiment of the invention;
FIG. 3 illustrates an embodiment of the invention;

2

FIG. 4 illustrate, in flow chart form, the steps performed in implementing an embodiment of the invention; and

FIGS. 5-6 illustrate, in flow chart form, the steps performed in implementing another embodiment of the invention.

GENERAL DESCRIPTION

Problems of front-end clipping and long holdover times are resolved by the introduction of a history at the transmitting end. The history queue is equal in length to the normal front-end clipping time. That is to say that there are sufficient samples in the history queue to equal the normal time that would be devoted to front-end clipping. When the speech threshold is reached indicating silence, the transmitter no longer transmits packets to the receiving end of the conversation. However, the speech samples being generated indicating silence or voice are continuously stored in the history queue. However, it should be realized that only the last period of time of the speech is stored in the history queue during this period of operation. When the speech threshold is reached indicating the transition from silence to voice, the transmitter begins once again to remove samples from the history queue and transmit packets to the receiving end of the voice conversation. Since the history queue includes the normal front-end clipping time of samples prior to the detection of voice, the transition from silence to speech appears to the listener to be excellent since this transition includes the normal front-end clipped speech. Advantageously, not only is the front-end clipping problem resolved, but the holdover time that is allowed for the determination of silence can be reduced. Advantageously, this method and apparatus greatly increases the efficiency of the transmission of voice through a packetized system.

DETAILED DESCRIPTION

FIG. 1 illustrates a system for implementing an embodiment of the invention. Synchronous physical interface **101** is exchanging digital samples with IP switched network **107** via voice encoder **106**. Voice samples being received from IP switched network **107** are received by voice coder **106** and processed by elements **102-104** before being transferred to interface **101** in a manner well known by those skilled in the art. This processing allows insert/remove circuit **102** to maintain a steady synchronous stream of voice samples to interface **101** in accordance with the requirements of interface **101**.

Interface **101** is also transmitting a steady synchronous stream of voice samples to history queue **108** and low energy detector **109**. However, voice coder **106** is packetizing voice samples for transmission to the receiving end of the voice conversation via IP switched network **107**. The number of samples stored in history queue **108** is equal to the holdover time between utterances that has been determined for the user of the system that is speaking into a microphone not shown that eventually communicates voice samples to interface **101**. The length of the queue of history queue **108** would adapt to the speaking characteristics of different users, resulting in the number of samples being processed by history queue **108** varying for individual users and during the conversation for the same user. Low energy detector **109** determines the thresholds that specify the presence of silence or voice activity in the speech samples being received from interface **101**. History queue **108** is continuously accepting samples from interface **101** and attempting to transmit these samples to control circuit **111**. Control