Trends in Audio and Speech Compression for Storage and Real-Time Communication

Paul Mermelstein
BNR and INRS-Telecommunication (Univ. of Quebec)
Montreal, Canada

ABSTRACT

Advances in speech compression are generally driven by the need to conserve transmission rate or bandwidth, while maintaining an ability to reconstruct the speech at the receiver with good fidelity. For real-time communication a constant transmission rate and low signal delay for processing are additional requirements. For speech storage applications, variable rate compression techniques which match the short-time compression rate to the nature of the current signal are of significant interest. Variable rate systems are rarely used in real-time systems other than packet transmission systems, since the output would have to be buffered for transmission by constant rate transmission pipes and incur a variable or maximum signal delay unacceptable from the point of view of echo considerations on the telephone system.

High-fidelity audio storage systems, such as compact discs require storage rates of roughly 700 kb/s without compression, 1400 kb/s for 2 channel stereo. Reduction to 128 kb/s and even 64 kb/s can be achieved without noticeable degradation by taking advantage of features of the auditory system that permit shaping the quantization noise in time and frequency so that it is best masked by the audio signal. Such compression systems are of interest primarily to increase the program time of recording devices and in some cases for audio program transmission. Since the higher frequencies do not contribute to stereo perception, an additional of 25% reduction is achievable by common processing of the two stereo channels.

The principal focus of speech compression research today is in coding for wireless communication where capacity is limited by the scarce spectrum space available. In particular, a new generation of digital mobile transmission systems is being developed employing 8 kb/s speech coding with a provision for a future doubling of capacity through 4 kb/s speech coding. High speech quality can be maintained by such systems under good propagation conditions, quality superior to that achieved by analog mobile systems. Under fading conditions some degradation will be noted but good intelligibility is still retained.