

SUMMARY OF Ph.D. DISSERTATION

School	Student Identification Number	SURNAME, First name EHARA, Hiroyuki
Title Enhanced low-bit-rate speech coding and a scalable coder		
<p>Abstract</p> <p>High-quality low-bit-rate speech coding, with its enhancing technologies and scalable coding, is examined in this study. A high-quality 4-kbit/s speech coding algorithm was developed for realizing speech communication among different networks at different transmission speeds using a single unified speech codec. The algorithm was enhanced to a wideband scalable speech coding algorithm whose bit-stream is decodable at bit rates of 6.8–32 kbit/s.</p> <p>The 4-kbit/s speech coding algorithm was inspired by ITU-T Recommendation G.729, the lowest bit rate high-quality speech coding standard based on an algebraic code-excited linear-prediction. To achieve an equivalent quality to that of ITU-T Recommendation G.726 (32 kbit/s) or G.729 (8 kbit/s), the algorithm has the following features: 1) a fixed codebook (FCB) comprising a constrained algebraic codebook and a random codebook, 2) backward adaptive mode switching for controlling the proportion of the random codebook to the constrained algebraic codebook, 3) dispersed pulse based FCB, and 4) noise post-processing (NPP) at the decoder side. The NPP generates pseudo-stationary-noise and superimposes it on a decoded speech signal. Through extensive subjective listening tests, the effectiveness of NPP has been demonstrated on existing standards such as G.729 and G.723.1.</p> <p>For accommodating VoIP applications, improved algorithms for frame-erasure concealment (FEC) of existing speech coding standards are studied. One is an extrapolation algorithm, in which excitation signal energy of a lost frame is constructed based on past evolution of excitation signal energy. Another is interpolative concealment of parameters that are quantized using moving-average prediction. It was realized by introducing the constraint of minimizing the total distance between parameters decoded on three consecutive frames including pre-and-post frames of an erased frame. Performance improvement was verified in a condition of a 10% frame erasure rate.</p> <p>For bandwidth extension of a speech signal, predictive quantization of wideband line spectral frequency (LSF) was studied. It works in combination with a narrowband LSF quantizer; consequently, it is applicable to technologies intended for enhancing the quality of a speech signal by extending its bandwidth. One feature of predictive quantization is exploitation of the correlation between wideband and narrowband LSFs quantized in the previous frame for estimating wideband LSF in the current frame. Test results show that introduction of the predictive scheme improved the performance in spectral distortion by 0.3 dB. (1.6 dB → 1.3 dB)</p> <p>Finally, audio-signal bandwidth-extension and band-selective modified discrete cosine transform coding algorithms are implemented on top of a 6.8 kbit/s speech coding algorithm, which is based on the 4 kbit/s speech coding algorithm described above; they are formulated as a high-quality speech and audio scalable coding algorithm. The studied NPP, FEC, and LSF quantization are also integrated in the scalable coding algorithm. Subjective listening test results demonstrate that the scalable coder outperforms a state-of-the-art scalable coder G.729.1, which was standardized by ITU-T in 2006.</p>		