Multimode coding: A novel approach to narrow- and medium-band coding

Tomohiko Taniguchi, Shigeyuki Unagami and Robert M. Gray

Citation: The Journal of the Acoustical Society of America **84**, S12 (1988); doi: 10.1121/1.2025766
View online: https://doi.org/10.1121/1.2025766
View Table of Contents: https://asa.scitation.org/toc/jas/84/S1
Published by the Acoustical Society of America

Session G. Speech Communication I: Analysis and Synthesis Part A (Poster Session)

Astrid Schmidt-Nielsen, Cochairman
Naval Research Laboratory
Code 5532
Washington, DC 20375-5000

Fumitada Itakura, Cochairman
School of Engineering
Nagoya University
Furo-cho, Chikusa-ku
Nagoya, 464 Japan

Contributed Papers

Posters must be set up before 8:00 a.m. (before Opening Plenary Session). All posters will be displayed from 8:30 to 10:00 a.m. To allow contributors the opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 to 9:15 a.m. and contributors of even-numbered papers will be at their posters from 9:15 to 10:00 a.m.

G1. Efficient subband coding of speech with optimized uniform DFT filter banks. A. Satt and D. Malah (Department of Electrical Engineering, Technion, Haifa 32000, Israel)

Uniform DFT filter banks (FB) offer reduced implementation complexity of subband coders (SBC) as compared with QMF-based SBC. However, because known design techniques aim at minimizing the overall response error of the FB, using either deterministic [V. K. Jain and R. E. Crochiere, Proc. ICASSP **83**, 228–231 (1983)] or statistical [A. Dembo and D. Malah, IEEE Trans. Acoust. Speech Signal Process. **ASSP-36**, 328–341 (1988)] error measures, the performance of DFT-based SBC was found to be subjectively inferior. A new approach is presented for DFT-FB design, which is based on a frequency-doman distortion function in which different weights are applied to the different error components (including interband aliasing and leakage of both signal and quantization noise) thus allowing an optimized design for SBC. The minimization of the proposed distortion function results in two sets of linear equations that are solved iteratively to obtain the optimal analysis and synthesis prototype filters. A 16-band DFT-based SBC operating at 16 kb/s with the designed filters was found in simulations to have subjective and objective performance similar to a QMF-based SBC at the same rate, but with less than half of the computations.

G2. Multimode coding: A novel approach to narrow- and medium-band coding. Tomohiko Taniguchi, Shigeyuki Unagami (Speech Signal Processing Section, Fujitsu Laboratories Ltd., 1015 Kamikodanaka, Nakahara-ku, Kawasaki, 211 Japan), and Robert M. Gray (Information Systems Laboratory, Stanford University, Stanford, CA 94305)

Most research on narrow-band coding is concentrated on how to transmit excitation parameters efficiently. However, the important thing in obtaining a good reproduced speech quality is how to control the balance of the transmission bit rate between the excitation and the LPC parameters. Multimode coding, which is proposed here, has two coding modes: One is a mode that transmits the LPC parameters in every frame, as conventional coders (A mode). The other is a mode that avoids the transmission of the LPC parameters by using the same coefficients as the previous frame and increases the bits allocated to the residual quantization instead of to the LPC transmission (B mode). In each frame, the mode selection takes place based on an evaluation of the reproduced speech quality, and the assignment of transmission information is dynamically controlled by switching between the two modes. This coding algorithm is applied to a 7.2 kb/s CELP coder, and approximately 3 dB of improvement is achieved in SNR compared with a conventional CELP coder. The B mode was used 78%–82% of the time.

G3. Realization of a multirate speech codec utilizing spectrum peak emphasis. Tetsu Taguchi (Radio Application Division, NEC Corporation, Fuchu, 183 Japan)

One of the most promising means of realizing multirate speech coding is multipulse excited linear predictive coding (MPELPC) adopting an auditory weighting filter (AWF). This AWF effectively shapes quantization noise to be masked by the speech signal and ameliorates speech quality at approximately 8 kb/s or more. However, below this rate, the AWF is not effective because the SNR is so low that the signals scarcely mask the noise. When the SNR is rather low, only the high-power frequency components of speech signals (whose levels are higher than those of the noise) should be encoded. In order to emphasize these components effectively upon encoding, a spectrum peak emphasis (SPE), which is performed by a prefilter and utilized in MPELPC, has been proposed. Below 8 kb/s, it has been shown by computer simulation that compared to the AWF, the SPE improves SNR at least by 2.2 dB. Subsequently, a multirate code from 16 to 4.8 kb/s, basically utilizing the AWF but with the SPE below 8 kb/s, has been realized on a DSP, NECμPD77230. This coder is applicable to mobile satellite communication.

G4. Speech coding by Model Reference Adaptive Control. Kiyoshi Hashimoto and Makoto Yasuhara (The University of Electro-Communications, Chofu, 182 Japan)

Model Reference Adaptive Control (MRAC) coding [M. Yasuhara, Trans. IEICE **E71**(1), 34–42 (1987)] was applied to speech signals. The primary feature that characterizes MRAC coding is that it is not a predictive coding, but control coding. It determines the control signal so the output of a plant (an ARMA speech production model) follows the speech waveform. The control signal is quantized and transmitted to the receiver. Since at the receiver only a noisy state vector of the plant is observable, a Kalman filter is embedded in the plant to estimate the optimal state vector. Experiments were conducted to evaluate the transmission performances in the S/D ratio for the MRAC coding in comparison with those for the forward ADPCM. With a rate of 2.25 bits/sample, the AR order of the plant between 5 and 12, and the MA order of 1, the average performance for the MRAC coding is around 20 dB, which is 6 dB higher than for the FADPCM.

G5. On improving voice periodicity prediction in codebook-excited LPC coders. Daniel Lin (International Mobile Machines, 2130 Arch Street, Philadelphia, PA 19103)

Codebook-excited predictive coders are able to synthesize high-quality speech at bit rates of 8 kbits/s and above. However, at lower bit rates (e.g. 4.8 kbits/s), speech enhancement techniques, such as adaptive post-filtering, are needed to improve subjective performance [cf. P. Kroon and B. S. Atal, Proc. ICASSP (1987)]. The main source of quality degradation at 4800 b/s is due to the reduced transmission bandwidth of the excitation parameters, which results in an inaccurate representation of the LPC innovation signal. In this paper, methods are proposed for improving the voiced speech excitation and prediction in the codebook-excited