

ICASSP-93

Speech Processing

Volume I of V

**1993
IEEE International
Conference
on
Acoustics,
Speech, and
Signal Processing**



April 27-30, 1993
Minneapolis Convention Center
Minneapolis, Minnesota, USA

93CH3252-4

ICASSP-93

Speech Processing

Volume II of V

1993

IEEE International Conference

on

**Acoustics, Speech, and
Signal Processing**

江苏工业学院图书馆
藏书章



April 27-30, 1993

Minneapolis Convention Center
Minneapolis, Minnesota, USA

93CH3252-4

Sponsored by
THE INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS
SIGNAL PROCESSING SOCIETY

ICASSP-93 Proceedings in 5 Volumes:

Volume I	Plenary, Special, Audio, Underwater Acoustics, VLSI, Neural Networks
Volume II	Speech Processing
Volume III	Digital Signal Processing
Volume IV	Statistical Signal and Array Processing
Volume V	Image and Multidimensional Signal Processing

Additional copies may be ordered from:

IEEE Order Dept.
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08854

IEEE Catalog No. 92CH3252-4

ISBN	0-7803-0946-4	Softbound
ISBN	0-7803-0947-3	Casebound
ISBN	0-7803-0948-0	Microfiche
ISBN	0-7803-0949-9	CD Rom

Library of Congress Catalog Card No. 83-645139 Serial

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 29 Congress St., Salem, MA 01971. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. For other copying, reprint or republication permission, write to Director, Publishing Services, IEEE, 345 E. 47 St., New York, NY 10017. All rights reserved. Copyright © 1993 by The Institute of Electrical and Electronics Engineers, Inc.

VOLUME II

Speech Processing

Table of Contents

Spectral Quantization for Speech Coding

Chair:

N. Farrardin, *University of Maryland, (USA)*

- A Long History Quantization Approach to Scalar and Vector Quantization of LSP Coefficients** II-1
C. Xydeas and K. So, *University of Manchester, (United Kingdom)*

- Efficient Coding of LPC Parameters Using Adaptive Prefiltering and MSVQ with Partially Adaptive Codebook** II-5
Y. Tanaka and T. Taniguchi, *Fujitsu Labs., Ltd., (Japan)*

- Immittance Spectral Pairs (ISP) for Speech Encoding** II-9
Y. Bistriz and S. Peller, *Tel-Aviv University, (Israel)*

- Robust Vector Quantization in Spectral Coding** II-13
R. Hagen and P. Hedelin, *Chalmers University of Technology, (Sweden)*

- Vector Quantization of Speech, LSP Parameters Using Trellis Codes and/norm Constraints** II-17
J. Pan and T. Fischer, *Washington State University, (USA)*

- A Combined Quantization-Interpolation Scheme for Very Low Bit Rate Coding of Speech LSP Parameters** II-21
J. Lopez-Soler, *University of Grenada;* and N. Farvardin, *University of Maryland, (USA)*

- Interframe Differential Vector Coding of Line Spectrum Frequencies** II-25
E. Erzin and A. Çetin, *Bilkent University, (Turkey)*

- Vector Quantizer Design for the Coding of LSF Parameters** II-29
J. Collura and T.E. Tremain, *DOD, (USA)*

Language Modeling

Chair:

S. Roukos, *IBM T.J. Watson Research Center, (USA)*

- The Estimation of Powerful Language Models from Small and Large Corpora** II-33
P. Placeway, R. Schwartz, P. Fung and L. Nguyen, *BBN Systems and Technologies, (USA)*

- Statistical Language Modeling Combining N-Gram and Context-Free Grammars** II-37
M. Meeter, *Rensselaer Polytechnic Institute;* and J. Rohlicek, *BBN Systems & Technologies, (USA)*

- Automatic Word Classification Using Simulated Annealing** II-41
M. Jardino and G. Adda, *LIMSI-CNRS, (France)*

- Trigger-Based Language Models: A Maximum Entropy Approach** II-45
R. Lau, R. Rosenfeld and S. Roukos, *IBM T.J. Watson Research Center, (USA)*

- Flexible Use of Semantic Constraints in Speech Recognition** II-49
W. Ward and S. Young, *Carnegie Mellon University, (USA)*

- Probabilistic Parse Scoring with Prosodic Information** II-51
N. Veilleux and M. Ostendorf, *Boston University, (USA)*

- Learning Speech Semantics with Keyword Classification Trees** II-55
R. Kuhn and R. DeMori, *McGill University, (Canada)*

- Better Alignment Procedures for Speech Recognition Evaluation** II-59
W. Fisher and J. Fiscus, *NIST, (USA)*

- Semantics and Constraint Parsing of Word Graphs** II-63
M. Harper, L. Jamieson, C. Zoltowski and R. Helzerman, *Purdue University, (USA)*

- An Heuristic-Based Context-Free Parsing Algorithm** II-67
A. Fred, J. Leitão, *CAPS-INIC, (Portugal)*

Speech Recognition in Noise

Chair:

S. Young, *University of Cambridge, United Kingdom*

- Influence of Background Noise and Microphone on the Performance of the IBM Tangora Speech Recognition System** II-71
S. Das, R. Bakis, A. Nadas, D. Nahamoo and M. Picheny, *IBM Research, (USA)*

- Noise-Robust HMMS Based on Minimum Error Classification** II-75
K. Ohkura, M. Sugiyama and D. Rainton, *ATR Interpreting Telephony Research Labs., (Japan)*

- Subband or Cepstral Domain Filtering for Recognition of Lombard and Channel-Distorted Speech** II-79
B. Hanson and T. Applebaum, *Panasonic, (USA)*

- Recognition of Speech in Additive and Convolutional Noise Based on a RASTA Spectral Processing** II-83
H. Hermansky, *US West & ICSI;* N. Morgan, *ICSI;* and H. Hirsch, *University of Aachen, (USA)*

- Noise Masking in a Transform Domain** II-87
B. Mellor and A. Varga, *DRA, (England)*

- Multi-Microphone Correlation-Based Processing for Robust Speech Recognition** II-91
T. Sullivan and R. Stern, *Carnegie Mellon University, (USA)*

- Adaptive Source Compensator and Enhancement for Speech Recognition in Noisy, Stressful Environments** II-95
J. Hansen, *Duke University, (USA)*

ROOT Homomorphic Deconvolution Schemes for Speech Processing in Car-Noise Environments P. Alexandre, J. Boudy and P. Lockwood, <i>Matra Communication, (France)</i>	II-99	Speech Classification Embedded in Adaptive Codebook Search for CELP Coding C. Kuo, F. Jean and H. Wang, <i>National Tsing Hua University, (Taiwan-ROC)</i>	II-147
Noisy Speech Recognition Based on HMMs, Weiner Filters and Re-evaluation of Most Likely Candidates S. Vaseghi and B. Milner, <i>University of East Anglia, (United Kingdom)</i>	II-103	Vector Quantized MBE with Simplified V/UV Division at 3.0 Kbps M. Nishiguchi, J. Matsumoto, R. Wakatsuki and S. Ono, <i>Sony, (Japan)</i>	II-151
Direct Modification on LPC Coefficients with Application to Speech Enhancement and Improving the Performance of Speech Recognition in Noise C. Guan, Y. Chen and B. Wu, <i>Southeast University, Nanjing, (China)</i>	II-107	Variable Rate Speech Coding with Phonetic Segmentation E. Paksoy, K. Srinivasan and A. Gersho, <i>University of California, Santa Barbara, (USA)</i>	II-155
Spoken Language Systems		Implementation and Evaluation of a 2400 bps Mixed Excitation LPC Vocoder A. McCree and T. Barnwell, <i>Georgia Institute of Technology, (USA)</i>	II-159
Chair: G. Doddington, <i>DARPA, (USA)</i>		Evaluation of Low Rate Speech Coders for HF T. Tremain, D. Kemp, J. Collura and M. Kohler, <i>DOD, (USA)</i>	II-163
The BBN/HARC Spoken Language Understanding System M. Bates, R. Bobrow, P. Fung, R. Ingria, F. Kubala, J. Makhoul, L. Nguyen, R. Schwartz and D. Stallard, <i>BBN Systems & Tech., (USA)</i>	II-111	High Quality Speech Coding at 2.4 to 4.0 Kbps Based on Time-Frequency Interpolation Y. Shoham, <i>AT&T Bell Labs, (USA)</i>	II-167
Noisy Spontaneous Speech Understanding Using Noise Immunity Keyword Spotting with Adaptive Speech Response Cancellation Y. Takebayashi, Y. Nagata and H. Kanazawa, <i>Toshiba R&D Center, (Japan)</i>	II-115	Dynamic Bit Allocation in CELP Excitation Coding T. Eriksson and J. Sjöberg, <i>Chalmers University of Technology, (Sweden)</i>	II-171
Words Graphs: An Efficient Interface Between Continuous Speech Recognition and Language Understanding M. Oerder and H. Ney, <i>Philips Research Laboratories, (Germany)</i>	II-119	A Mixed Prototype Waveform/CELP Coder for sub 3 Kb/s I. Burnett and R. Holbeche, <i>University of Bath, (United Kingdom)</i>	II-175
Partial Parsing as a Robust Parsing Strategy P. Baggia and C. Rullent, <i>CESLT, (Italy)</i>	II-123	Voiced Speech Coding at Very Low Bit Rates Based on Forward-Backward Waveform Prediction (FBWP) G. Yang, H. Leich and R. Boite, <i>Lab. T.C.T.S. Faculté Polytechnique de Mons, (Belgium)</i>	II-179
New Techniques for Speech Understanding P. Bianchini, P. Ferragina, M. Notturmo and L. Tarricone, <i>IBM SEMEA, (Italy)</i>	II-127	Speech Synthesis	
Gisting Conversational Speech in Real Time L. Denenberg, H. Gish, M. Meteer, T. Miller, J. Rohlicek, W. Sadkin and M. Siu, <i>BBN, Inc., (USA)</i>	II-131	Chair: Y. Sagisaka, <i>ATR Interpreting Telephony Research Labs, (Japan)</i>	
Test and Evaluation of a Spoken Dialogue System E. Gerbino, P. Baggia, A. Ciaramella and C. Rullent, <i>CESLT, (Italy)</i>	II-135	A Text-to-Speech System for Spanish with a Frequency Domain Based Prosodic Modification E. Banga, <i>Universidad de Vigo</i> ; E. Lopez-Gonzalo, <i>Univ. Politec. Madrid</i> ; and C. Garcia-Mateo, <i>Universidad de Vigo, (Spain)</i>	II-185
ATREUS: A Comparative Study of Continuous Speech Recognition Systems at ATR A. Nagai, K. Yamaguchi, S. Sagayama and A. Kurematsu, <i>ATR Interpreting Telephony Research Labs., (Japan)</i>	II-139	Multilingual PSOLA Text-to-Speech System D. Bigorgne, O. Boeffard, B. Cherbonnel, F. Emarard, D. Larreur, J. Le Saint Mion, I. Metayer, C. Sorin and S. White, <i>CNET, (France)</i>	II-187
Speech Coding at Rates Below 4 KB/S		Tree-Based Unit Selection for English Speech Synthesis W. Wang, W. Campbell, N. Iwahashi and Y. Sagasaka, <i>ATR, (Taiwan ROC)</i>	II-191
Chair: J. Gibson, <i>Texas A&M University, (USA)</i>		Waveform-Based Speech Synthesis Approach with a Formant Frequency Modification H. Mizuno, M. Abe and T. Hirokawa, <i>NTT Human Interface, (Japan)</i>	II-195
A 2.4 kb/s CELP Speech Codec with Class-Dependent Structure P. Lupini, <i>Simon Fraser University</i> ; H. Hassanein, <i>Communications Research Centre</i> ; and V. Cuperman, <i>Simon Fraser University, (Canada)</i>	II-143	An Efficient Way to Learn English Grapheme-to-phoneme Rules Automatically K. Torkkola, <i>IDIA, (Switzerland)</i>	II-199
		Inference of Letter-Phoneme Correspondences with Pre-Defined Consonant and Vowel Patterns R. Luk and R. Dampier, <i>University of Southampton, (United Kingdom)</i>	II-203

Application of the Analysis of Glottal Excitation of Stressed Speech to Speaking Style Modification II-207
K. Cummings and M. Clements, *Georgia Institute of Technology, (USA)*

Analysis and Modeling of Word Accent and Sentence Intonation in Swedish II-211
H. Fujisaki, *Science University of Tokyo*; and M. Ljungqvist, *Infovox AB, (Japan)*

Speech Recognition Modeling II

Chair:

B. Hanson, *Speech Technology Laboratory, (USA)*

Use of Temporal Correlation Between Successive Frames in a Hidden Markov Model Based Speech Recognizer II-215
K. Paliwal, *TIFR, (India)*

Phoneme HMMs Constrained by Frame Correlations II-219
S. Takahashi, T. Matsuoka, Y. Minami and K. Shikano, *NTT, (Japan)*

Automatically Generated Word Pronunciations from Phoneme Classifier Output II-223
P. Schmid, R. Cole and M. Fanty, *CSLU, CSE, (USA)*

A New Framework for Recognition of Mandarin Syllables with Tones Using Sub-Syllabic Units II-227
C. Lin, L. Lee and P.-Y. Ting, *National Taiwan University, (Taiwan ROC)*

A Comparison of Auditory Models for Speaker Independent Phoneme Recognition II-231
T. Anderson, *Armstrong Lab WPAFB, (USA)*

Speaker-Independent Spelling Recognition over the Telephone II-235
D. Jouvet, A. Laine, J. Monne and C. Gagnoulet, *CNET, (France)*

Improvements in Connected Digit Recognition Using Linear Discriminant Analysis and Mixture Densities II-239
R. Häcb-Umbach, D. Geller and H. Ney, *Philips Research Lab., (Germany)*

Inter-Word Coarticulation Modeling and MMIE Training for Improved Connected Digit Recognition II-243
R. Cardin, Y. Normandin and E. Millien, *CRIM, (Canada)*

Limited Parameter Hidden Markov Models for Connected Digit Speaker Verification over Telephone Channels II-247
J. deVeth, G. Gallopyn and H. Bourlard, *LHS, Belgium, (Belgium)*

An Algorithm for the Dynamic Inference of Hidden Markov Models (DIHMM) II-251
P. Lockwood and M. Blanchet, *Matra Communication, (France)*

Speech Recognition by Neural Networks

Chair:

A. Waibel, *Carnegie Mellon University, (USA)*

Multi-Speaker/Speaker-Independent Architectures for the Multi-State Time Delay Neural Network II-255
H. Hild and A. Waibel, *Carnegie Mellon University, (USA)*

Performance Through Consistency: Connectionist Large Vocabulary Continuous Speech Recognition II-259
J. Tebelskis, *Carnegie Mellon University, (USA)*

A Multilayer Perception Postprocessor to Hidden Markov Modeling for Speech Recognition II-263
J. Guo and H. Lui, *ISS, NUS, (Singapore)*

Exploiting Prediction Error in a Predictive-Based Connectionist Speech Recognition System II-267
B. Petek and A. Ferligoj, *University of Ljubljana, (Slovenia)*

An Optimal Learning Method for Minimum Spotting Errors II-271
T. Komori and S. Katagiri, *ATR, (Japan)*

Feature Extraction Based on Minimum Classification Error/Generalized Probabilistic Descent Method II-275
A. Biem and S. Katagiri, *ATR, (Japan)*

Speech Discrimination in Adverse Conditions Using Acoustic Knowledge and Selectively Trained Neural Networks II-279
Y. Anglade and D. Fohr, *CNRS/INRIA*; and J. Junqua, *Speech Tech. Laboratory, (France)*

Speech Recognition Using Dynamical Model of Speech Production II-283
K. Iso, *NEC Corp., (Japan)*

Recurrent Input Transformations for Hidden Markov Models II-287
V. Valtchev, S. Kapadia and S. Young, *Cambridge University, (United Kingdom)*

Prototype-Based MCE/GPD Training for Word Spotting and Connected Word Recognition II-291
E. McDermott and S. Katagiri, *ATR A&V Labs., (Japan)*

Search and Modeling Techniques

Chair:

P. Gopalakrishnan, *IBM T.J. Watson Research Labs, (USA)*

Matrix Parser and its Application to HMM-Based Speech Recognition II-295
H. Singer and S. Sagayama, *ATR, (Japan)*

A New Word Pre-Selection Method Based on an Extended Redundant Hash Addressing for Continuous Speech Recognition II-299
A. Ito and S. Makino, *Tohoku University, (Japan)*

Using Search to Improve Hidden Markov Models II-303
M. Galler and R. DeMori, *McGill University, (Canada)*

An Improved Search Algorithm Using Incremental Knowledge for Continuous Speech Recognition II-307
F. Alleva, X. Huang and M. Hwang, *Carnegie Mellon University, (USA)*

Predicting Unseen Triphones with Senones II-311
M. Hwang, X. Huang and F. Alleva, *Carnegie Mellon University, (USA)*

Analysis and Improvement of the Partial Distance Search Algorithm II-315
L. Fissore, *CSELT*; P. Laface, P. Massafra and F. Ravera, *Politecnico di Torino, (Italy)*

Large Vocabulary Dictation Using SRI's DECIPHER Speech Recognition System: Progressing-Search Techniques II-319
H. Murveit, J. Butzberger, V. Digalakis and M. Weintraub, *SRI International, (USA)*

Exploiting Variable-Width Features in Large Vocabulary Speech Recognition
M. Jones and P. Woodland, *University of Cambridge, (United Kingdom)*

II-323

A Comparison of Trajectory and Mixture Modeling in Segment-Based Word Recognition
A. Kannan and M. Ostendorf, *Boston University, (USA)*

II-327

Modeling Duration in a Hidden Markov Model with the Exponential Family
C. Mitchell and L. Jamieson, *Purdue University, (USA)*

II-331

Enhancement of Noisy Speech

Chair:

B. Juang, *AT&T Bell Labs, (USA)*

An Algorithm for Statistical Signal Mapping and its Assessments
Y. Cheng, *INRS, (Canada)*

★

Noise Robustness in the Auditory Representation of Speech Signals
K. Wang, S. Shamma and W. Byrne, *University of Maryland, (USA)*

II-335

Dynamic Time Warping Comb Filter for the Enhancement of Speech Degraded by White Gaussian Noise
J. Graf and N. Hubing, *University of Missouri, Rolla, (USA)*

II-339

Acoustic Signal Separation of Statistically Independent Sources Using Multiple Microphones
M. Engebretson, *Central Institute for the Deaf, (USA)*

II-343

Microphone Array Speech Enhancement in Overdetermined Signal Scenarios
R. Slyh and R. Moses, *Ohio State University, (USA)*

II-347

A New Two-Sensor Algorithm for Active Noise Cancellation Algorithm
K. Zangi, *Massachusetts Institute of Technology, (USA)*

II-351

A Signal Subspace Approach for Speech Enhancement
Y. Ephraim and H. Van Trees, *George Mason University, (USA)*

II-355

Speech Enhancement Using Psychoacoustic Criteria
D. Tsoukalas, M. Paraskevas and J. Mourjopoulos, *University of Patras, (Greece)*

II-359

Frequency Domain Noise Suppression Approaches in Mobile Telephone Systems
J. Yang, *NovAtel Communications Ltd., (USA)*

II-363

Speech Enhancement Using the Dual Excitation Speech Model
J. Hardwick, C. Yoo and J. Lim, *Massachusetts Institute of Technology, (USA)*

II-367

Speaker and Language Identification

Chair:

F. Soong, *AT&T Bell Labs, (USA)*

A Comparison of Composite Feature under Degraded Speech in Speaker Recognition
J. Openshaw, Z. Sun and J. Mason, *University of Wales, (United Kingdom)*

II-371

Voice Identification Using Nearest-Neighbor Distance Measure
A. Higgins, L. Bahler and J. Porter, *ITT Aerospace, (USA)*

II-375

Robustness Studies in Free-Text Speaker Identification and Verification
Y. Kao, *University of Maryland & Texas Instruments*; J. Baras, *University of Maryland*; and R. Rajasekaran, *Texas Instruments, (USA)*

II-379

Identification of Speakers Engaged in Dialog
G. Yu and H. Gish, *BBN, (USA)*

II-383

Speaker Identification Experiments Using HMMs
J. Webb and E. Rissanen, *AT&T, (USA)*

II-387

Concatenated Phoneme Models for Text-Variable Speaker Recognition
T. Matsui and S. Furui, *NTT, (Japan)*

II-391

Speech Segmentation and Clustering Based on Speaker Features
M. Sugiyama and J. Murakami, *ATR*; and H. Watanabe, *Hokkaido University, (Japan)*

II-395

Automatic Language Identification Using Gaussian Mixture and Hidden Markov Models
M. Zissman, *Lincoln Lab., MIT, (USA)*

II-399

Text-Independent Speaker Identification Using Fuzzy Mathematical Algorithm
Z. Yuan, C. Yu and Y. Fang, *Nanjing University, (P.R. China)*

II-403

A Novel Approach to the Speaker Identification over Telephone Networks
H. Wang, M. Chen and T. Yang, *National Tsing Hua University, (Taiwan, ROC)*

II-407

Topics in Speech Coding

Chair:

B. Kleijn, *AT&T Bell Labs, (USA)*

Use of an Auditory Model to Improve Speech Coders
D. Sen, D. Irving and W. Holmes, *University of NSW, (Australia)*

II-411

Low-Delay Wideband Speech Coding Using a New Frequency Domain Approach
V. Sanchez, *University of Granada*; and J. Adoul, *University of Sherbrooke, (Spain)*

II-415

A Probabilistic Framework for Optimum Speech Extrapolation in Digital Mobile Radio
C. Gerlach, *Aachen University of Technology, (Germany)*

II-419

How Goos is your Index Assignment?
P. Knagenhjelm, *Chalmers University of Technology, (Sweden)*

II-423

Investigating the Use of Asymmetric Windows in CELP Vocoders
D. Florencio, *Georgia Institute of Technology, (USA)*

II-427

Using Neural Networks for Vector Quantization in Low Rate Speech Coders
J. Thyssen and S. Hansen, *Technical University of Denmark, (Denmark)*

II-431

The Benefits of Multi-Speaker Conferencing and the Design of Conference Bridge Control Algorithms
J. Tardelli, P. Gatewood, E. Kreamer and P. LaFollette, *ARCON Corp., (USA)*

II-435

The Application of Subband Coding to Improve Quality and Robustness of the Sinusoidal Transform Coder
R. McAulay and T. Quatieri, *MIT-Lincoln Lab., (USA)*

II-439

Analysis of the Smoothed Residual Driven Algorithm for Speech Coders II-443
S. Nam and J. Gibson, *Texas A&M University, (USA)*

Reducing the Computational Complexity of the LPC Vector Quantizer Using the Fast Tree Search Algorithm
V. Ramasubramanian and K. P. Bhalu, *CVC-TIFR, India*

Word Spotting

Chair:

R. Rohlicek, *BBN Systems and Technology, (USA)*

A Segmental Speech Model with Applications to Word Spotting II-447
H. Gish and K. Ng, *BBN, (USA)*

A Two-Pass Classifier for Utterance Rejection in Keyword Spotting II-451
U. Sukkar and J. Wilpon, *AT&T, (USA)*

Rejection Techniques for Digit Recognition in Telecommunications Applications II-455
L. Villarrubia and A. Acero, *Telefonica I+D, Spain*

Phonetic and Language Modeling for Word Spotting II-459
J. Rohlicek, P. Jeanrenaud, K. Ng, H. Gish, M. Musicus and M. Siu, *BBN, (USA)*

Keyword-Spotting Using SRI's DECIPHER Large-Vocabulary Speech-Recognition System II-463
M. Weintraub, *SRI International, (USA)*

Task Independent Word Spotting Using Decision Tree Based Allophone Clustering II-467
R. Rose and E. Hofstatter, *MIT Lincoln Lab, (USA)*

Application of Large Vocabulary Continuous Speech Recognition to Topic and Speaker Identification Using Telephone Speech II-471
L. Gillick, J. Baker, J. Baker, J. Bridle, M. Hunt, Y. Ito, S. Lowe, J. Orloff, B. Peskin, R. Roth and F. Scattone, *Dragon Systems, (USA)*

Improving the MS-TDNN for Word Spotting II-475
T. Zeppenfeld, R. Houghton and A. Waibel, *Carnegie Mellon University, (USA)*

Speech Recognition Modeling III

Chair:

W. Ward, *Carnegie Mellon University, (USA)*

Phonetic Recognition in a Segment-Based HMM II-479
J. Marcus, *BNR, (Canada)*

A Linguistic Feature Representation of the Speech Waveform II-483
E. Eide, *BBN & Massachusetts Institute of Technology*; J. Rohlicek and H. Gish, *BBN*; and S. Mitter, *Massachusetts Institute of Technology, (USA)*

Allophone Modeling for Vocabulary-Independent HMM Recognition II-487
W. Holmes, L. Wood and D. Pearce, *GEC, (United Kingdom)*

MMI Training for Continuous Phoneme Recognition on the TIMIT Database II-491
S. Kapadia, V. Valtchev and S. Young, *Cambridge University, (United Kingdom)*

Sinuated State Hidden Markov Models II-495
D. Kimber and M. Bush, *Xerox PARC, (USA)*

A Segmental HMM for Speech Pattern Modelling II-499
M. Russell, *DRA Malvern, (United Kingdom)*

Golden Mandarin (II)—An Improved Single-Chip Real-Time Mandarin Dictation Machine for Chinese Language with Very Large Vocabulary II-503
L.S. Lee, I.-J. Hung, M.-Y. Lee, L.-F. Chien, Y. Lee, R. Lyu, H.-M. Wang, Y.-C. Wu, T.-S. Lin, H.-Y. Gu, C.-P. Nee, C.-Y. Liao, Y.-J. Yang, Y.-C. Chang, R.-C. Yang, *National Taiwan University*; and C.Y. Tseng and K.J. Chen, *Institute of History & Philology Academia Sinica, (Taiwan ROC)*

Cross-Lingual Experiments with Phone Recognition II-507
L. Lañel and J. Gauvain, *LIMS-CNRS, (France)*

Beam Search and Partial Traceback in the Frame Synchronous Two Level Algorithm II-511
O. Bezie and P. Lockwood, *Matra Communication, (France)*

Spanish Phone Recognition Using Semicontinuous Hidden Markov Models II-515
I. Torres, *Univ. del Pais Vasco*; and F. Casacuberta, *Univ. Politècnica de Valencia, (Spain)*

Speech Analysis & Synthesis

Chair:

D. Kahn, *Bell Comm. Research, (USA)*
C. Sorin, *CNET, France*

Glottal Pulse Alignment in Voiced Speech for Pitch Determination II-519
J. Harris and D. Nelson, *Department of Defense, (USA)*

Pseudo-Multi-Tap Pitch Prediction Filters II-523
Y. Qian, *Université de Québec*; and P. Kabal, *McGill University, (Canada)*

Determining the Pitch Period of Speech Using No Multiplications II-527
D. Sharp and R.L. While, *Imperial College, (England)*

Pitch from Zeros of Bank-Filtered Signals II-530
C. Peng and S. Ando, *University of Tokyo, (Japan)*

Estimation of the Parameters of a Long Term Model for Accurate Representation of Voiced Speech II-534
T. Stettiner and D. Malah, *Technicon*; and D. Chazan, *IBM, (Israel)*

Optimal Estimation of Vocal Tract Area Functions from Speech Signal Constrained by X-Ray Microbeam Data II-538
Q. Guo and P. Milenkovic, *University of Wisconsin, Madison, (USA)*

Glottal-Source Estimation: Methods of Applying the LF-Model to Inverse-Filtering II-542
E. Riegelsberger and A. Krishnamurthy, *Ohio State University, (USA)*

Acoustic Model of the Vocal Tract with Boundary Layer Corrections II-546
A. Mohsen, T. El-Mistikawy and M. Hesham, *Cairo University, (Egypt)*

HNS: Speech Modification Based on a Harmonic Plus Noise Model II-550
J. Laroche, Y. Stylianou and E. Moulines, *Télécom Paris, (France)*

An Overlap-Add Technique Based on Waveform Similarity (WSOLA) for High Quality Time-Scale Modification of Speech II-554
W. Verhelst and M. Roelands, *Vrije University of Brussels, (Belgium)*

Adaptive Modeling

Chair:

X. Huang, *Carnegie Mellon University, (USA)*

Speaker Adaptation Based on MAP Estimation of HMM Parameters II-558

C. Lee, *AT&T*; and J. Gauvain, *LIMSI/CNRS, (USA)*

A New Speaker Adaptation Technique Using Very Short Calibration Speech II-562

Y. Zhao, *Panasonic, (USA)*

Speaker Adaptation Using Improved Speaker Markov Models II-566

G. Rigoll, *NTT, (Japan)*

New Word Addition and Adaptation in a Stochastic Explicit-Segment Speech Recognition System ★

A. Asadi and H. Leung, *NYNEX, (USA)*

Rapid Speaker Adaptation Using Speaker-Mixture Allophone Models Applied to Speaker-Independent Speech Recognition II-570

T. Kosaka, J. Takami and S. Sagayama, *ATR, (Japan)*

An All-phoneme Ergodic HMM for Unsupervised Speaker Adaptation II-574

Y. Miyazawa, *ATR, (Japan)*

Utterance Normalization Using Vowel Features in a Spoken Word Recognition System for Multiple Speakers II-578

S. Ohno and K. Hirose, *University of Tokyo*; and H. Fujisaki, *Sci. U. Tokyo, (Japan)*

A Neural Network Controlled Adaptive Search Strategy for HMM-based Speech Recognition II-582

K. Yamaguchi, *ATR, (Japan)*

On the Dynamic Adaptation of Stochastic Language Models II-586

R. Kneser and V. Steinbiss, *Philips Research Laboratories, (Germany)*

Learning New Words from Spontaneous Speech II-590

S. Young and W. Ward, *Carnegie Mellon University, (USA)*

Medium Rate Speech Coding

Chair:

J-H. Chen, *AT&T Bell Labs, (USA)*

An 8 kbit/s Speech Coder Based on Conjugate Structure CELP II-592

A. Kataoka, T. Moriya and S. Hayashi, *NTT, (Japan)*

A 5.85 kb/s CELP Algorithm for Cellular Applications II-596

W. Kleijn and P. Kroon, *AT&T*; and L. Cellario and D. Sereno, *CSELT, (USA)*

Tree Coding Combined with Harmonic Scaling of Speech at 6.4 kbps II-600

I. Lee and J. Gibson, *Texas A&M, (USA)*

Techniques for Low Bit Rate Speech Coding Using Long Analysis Frames II-604

T. Minde, T. Wigren, J. Ahlberg and H. Hermansson, *Ericsson Radio System AB, (Sweden)*

Use of Low-Delay Code Excited Linear Prediction Technology in Circuit Multiplexed Networks II-608

S. Dimoultsas, *COMSAT Labs*; F. Corcoran, *COMSAT Labs*; M. Baraniecki, *INTELSAT*; and J. Phipps, *COMSAT Labs, (USA)*

8 Kbit/s Coding of Speech with 6 ms Frame-Length II-612

R. Lefebvre, R. Salami, C. Laflamme and J. Adoul, *University of Sherbrooke, (Canada)*

An 8 kb/s LD-CELP with Improved Excitation and Perceptual Modelling II-616

R. Soheili, A. Kondoz and B. Evans, *University of Surrey, (United Kingdom)*

Wideband Speech Coding in 7.2 KB/S II-620

C. McElroy, B. Murray and A. Fagan, *University of College Dublin, (Ireland)*

Transform Coding at 4.8 kbits/sec Using Interleaving of Transformed Frames and Dual Gain-Shape Vector Quantisation II-624

P. McCourt and H. Kaouri, *Queen's University of Belfast, (Northern Ireland)*

Phonetically-Driven CELP Coding Using Self-Organization Maps II-628

L. Hernandez-Gomez, *ETSI Telecommunication-UPM*; and E. Lopez-Gonzalo, *(Spain)*

Speech Recognition Modeling I

Chair:

C. Lee, *AT&T Bell Labs, (USA)*

Contact Dependent Vector Quantization for Continuous Speech Recognition II-632

L. Bahl, P. DeSouza, P. Gopalakrishnan and M. Picheny, *IBM Research, (USA)*

Unified Stochastic Engine (USE) for Speech Recognition II-636

X. Huang, M. Berlin, F. Alleva and M. Hwang, *Carnegie Mellon University, (USA)*

Large Vocabulary Continuous Speech Recognition of Wall Street Journal Data II-640

R. Roth, J. Baker and J. Baker, *Dragon Systems*; and L. Gillick, M. Hunt, Y. Ito, S. Lowe, J. Orloff, B. Peskin and F. Scattone, *(USA)*

A Supervised Approach to the Construction of Context Sensitive Acoustic Prototypes II-644

J. Bellegarda, P. DeSouza, D. Nahamo, M. Picheny and L. Bahl, *IBM, (USA)*

Continuous Mixture Densities & Linear Discriminant Analysis for Improved Context-Dependent Acoustic Models II-648

X. Aubert, R. Haeb-Umbach and H. Ney, *Philips Research, (Germany)*

Minimum Error Rate Training Based on N-Best String Models II-652

W. Chou, C. Lee and B. Juang, *AT&T Bell Labs, (USA)*

A New Fast Match for Very Large Vocabulary Speech Recognition II-656

P. Labute and P. Kenny, *INRS*; R. Hollan and M. Lennig, *BNR*; and D. O'Shaughnessy, *INRS, (Canada)*

The Lincoln Large-Vocabulary Stack-Decoder HMM CSR II-660

D. Paul and B.F. Necioglu, *Lincoln Lab, MIT, (USA)*

Feature Representation for Recognition

Chair:

J. Campbell, *Department of Defense, (USA)*

Speech Recognition Using the Modulation Model II-664
K. Assaleh, R. Mammone, M. Rahim and J. Flanagan, *Rutgers, The State University of New Jersey, (USA)*

A Dynamic Cepstrum Incorporating Time-Frequency Masking and Its Application to Continuous Speech Recognition II-668
K. Aikawa, H. Singer, H. Kawahara and Y. Tohkura, *ATR, (Japan)*

Two-Dimensional Cepstral Distance Measure for Speech Recognition II-672
H. Pai and H. Wang, *National Tsing Hua University, (Taiwan, ROC)*

A New Speech Recognition Method Based on VQ-Distortion Measure and HMM II-676
S. Nakagawa and H. Suzuki, *Toyohashi University of Technology, (Japan)*

A Comparative Study of Signal Representations and Classification Techniques for Speech Recognition II-680
H. Leung and B. Chigier, *NYNEX; and J. Glass, Massachusetts Institute of Technology, (USA)*

In Search for the Relevant Parameters for Speaker Independent Speech Recognition II-684
J. Smolders and D. Van Compernelle, *K.U. Leuven, (Belgium)*

A Non-Metrical Space Search Algorithm for Fast Gaussian Vector Quantization II-688
E. Schukat-Talamazzani, M. Bielecki, H. Niemann, T. Kuhn and S. Rieck, *University of Erlangen, (Germany)*

Vector Quantization for the Efficient Computation of Continuous Density Likelihoods II-692
E. Bocchieri, *AT&T Bell Labs, (USA)*

Speaker-Independent Speech Recognition Using Nonlinear Predictor Codebooks II-696
T. Kawabata, *NTT, (Japan)*

Auditory Model Representation for Speaker Recognition II-700
J. Colombi, *AFIT/ENG; T. Anderson, AL/CFBA; S. Rogers and D. Ruck, AFIT/ENG; and G. Warhola, AFIG/MA, (USA)*

Speech Analysis

Chair:

J. Tubach, *ENST, (France)*

H. Javkin, *Speech Technology Lab, (USA)*

Central Auditory Model for Spectral Processing II-704
T. Huang, *Institute of Automation; and Y. Gao and J. Haton, CRIN, (France)*

A New Approach to the Estimation of Vocal Tract Area Function ★
M. Asi and M. Mrayati, *HIASST, (Syria)*

A Wave Digital Filter Model of the Entire Auditory Periphery II-708
C. Giguere and P. Woodland, *University of Cambridge, (England)*

A Perceptual Metric for Masking II-712
A. Alwan, *University of California, Los Angeles, (USA)*

Finding Speech Formants and Modulations via Energy Separation: With Application to a Vocoder II-716
H. Hanson, P. Maragos and A. Potamianos, *Harvard University, (USA)*

Stop Classification Using DESA-1 High-Resolution Formant Tracking II-720
J. Foote, D. Mashao and H. Silverman, *LEMS, Brown University, (USA)*

Analysis and Automatic Recognition of False Starts in Spontaneous Speech II-724
D. O'Shaughnessy, *INRS-Telecommunications, (Canada)*

Speech Cryptanalysis Using DSP Techniques
S. Sridha, B. Goldberg and E. Dawson, *QUT, (Australia)*

Optimal Multi-Pitch Estimation Using the EM Algorithm for Co-Channel Speech Separation II-728
D. Chazan, *IBM; and Y. Stettiner and D. Malah, Technicon, (Israel)*

Endpoint Detection of Isolated Utterances Based on a Modified Teager Energy Measurement II-732
G. Ying, L. Jamieson and C. Mitchell, *Purdue University, (USA)*

★ This paper may appear in Late Papers Section at the rear of Volume V.

VOLUME I

Plenary, Special, Audio, Underwater Acoustics, VLSI, Neural Networks

Table of Contents

Plenary

A Model for Moving University Research to Industrial Application	I-1
J.L. Flanagan, <i>Rutgers University, (USA)</i>	
Measurement of Neurocognitive Networks of the Human Brain	I-2
A. Gevins, <i>EEG Systems Laboratory, (USA)</i>	
Digital Video Compression: Research and Standardization	★
D.L. LeGall, <i>C-Cube Microsystems, (USA)</i>	

Signal Processing Education

Chair:

J. McClellan, *Georgia Institute of Technology (USA)*
S. Wood, *University of Santa Clara, CA, (USA)*

All-Thru DSSP Provision, Essential for the Modern EE	I-4
G. Cain, A. Yardim, D. Morling, <i>University of Westminster, (United Kingdom)</i>	
An Evaluation of the New Fortran 90 Standard Relative to DSP Algorithm Implementations	I-8
D. Etter, <i>University of Colorado, (USA)</i>	
Investigating Signal Processing Theory with Mathematica	I-12
B. Evans and J. McClellan, <i>Georgia Institute of Technology</i> ; and J. Trussell, <i>North Carolina State University, (USA)</i>	
A Multimedia Image Database for Image Processing Instruction	I-16
R. Bamberger, <i>Washington State University</i> ; S. Eddins, <i>University of Illinois, Chicago</i> ; and S. Reeves, <i>Auburn University, (USA)</i>	
Teaching Filter Design Using Matlab	I-20
C. Burrus, <i>Rice University, (USA)</i>	
A Computer-Oriented Random Processes Course for a Mix of Graduates and Undergraduates	I-24
L. Kirlin and B. Hedstrom, <i>University of Victoria, (Canada)</i>	
Real-Time Digital Signal Processing in Education	I-28
R. Chassaing, <i>Roger Williams University</i> ; W. Anakwa, <i>Bradley University</i> ; and A. Richardson, <i>California State University, Chico, (USA)</i>	
Practical DSP for Scientists	I-32
R. Stewart, <i>Strathclyde University, (United Kingdom)</i>	
Teaching Signals and Systems Using the Virtual Laboratory Environment in ECE at CMU	I-36
V. Stonick, <i>Carnegie Mellon University, (USA)</i>	
SP Education Laboratory	★
G. Wakefield, B. Feng, K. Raghavan and M. Blommer, <i>University of Michigan, (USA)</i>	

Role of Signal Processing in Advanced Manufacturing

Chair:

J. Cozzens, *National Science Foundation, (USA)*

Fast Signal Processing Hardware for Industrial Inspection	I-40
F. Waltz, <i>3M, (USA)</i>	
Defect Characterization and Classification for Web Materials	I-44
D. Brzakovic, H. Beck and P. Shanmugam, <i>University of Tennessee, (USA)</i>	
Beyond 2000—Imaging for Automated Production Test of PCB Assemblies	I-48
J. Adams, <i>Four Pi Systems, (USA)</i>	
Acoustical Design Criteria for Active Noise Control Systems	★
J. Abbott, <i>SRI International, (USA)</i>	
Active Sound and Vibration Control Using Adaptive Digital Signal Processing	I-51
L. Eriksson, <i>Digisonix, (USA)</i>	
Signal Processing Research in Automatic Tool Wear Monitoring	I-55
L. Heck, <i>SRI International, (USA)</i>	
Application of Non-Stationary Analysis to Machinery Monitoring	I-59
M. Dowling, <i>Liberty Technologies, Inc., (USA)</i>	
Using Signal Processing to Improve Communications on the Manufacturing Shop Floor	I-63
J. Treichler, <i>Applied Signal Technology, Inc., (USA)</i>	
Signal Processing and Control in Semiconductor Fabrication	★
T. Kailath, <i>Stanford University, (USA)</i>	

Biomedical Signal Processing

Chair:

R. Leahy, *University of Southern California, (USA)*

Segmentation of High-Resolution ECGs Using Hidden Markov Models	I-67
D. Coast, <i>Allegheny-Singer Research Institute, (USA)</i>	
Some New Results on the Inverse Problem for Electrocardiography	I-71
C. Nikias, S. Gao and S. Heidari, <i>University of Southern California, (USA)</i>	

A LONG HISTORY QUANTIZATION APPROACH TO SCALAR AND VECTOR QUANTIZATION OF LSP COEFFICIENTS

C. S. Xydeas and K. K. M. So

Department of Electrical Engineering, University of Manchester, Dover Street, Manchester M13 9PL, United Kingdom.

ABSTRACT

It is essential that in low bit rate speech coding systems, codec parameters are quantized with a minimal number of bits without a corresponding reduction in the quality of the decoded signal. In an effort to increase coding efficiency, a new and general methodology for the quantization of codec parameters, called Long History Quantization (LHQ), has recently been proposed. LHQ, when applied in conjunction with scalar quantization for the coding of LSP coefficients in a CELP system, offers "transparent" quantization at an average bit rate of 25 bits per frame. This paper presents certain improvements to the above LHQ-Scalar quantization scheme, which further reduce the average LSP bit rate to about 22 bits per frame. In addition, a new LHQ-Vector quantization scheme is proposed which allows the transparent quantization of LPC coefficients with only 19 bits per analysis frame.

1. INTRODUCTION

The line spectrum pair (LSP) coefficients have been extensively used as an equivalent representation of the LPC coefficients in medium and low bit rate speech coding systems. As a consequence, several LSP quantization schemes have been proposed in the past which exploit the interframe [1,2,3] and intraframe [4,5] correlation of LSP parameters. More recently, a general approach for the adaptive quantization of codec parameters, called Long History Quantization (LHQ), has been proposed. LHQ exploits the constraints imposed on the signal by i) the speech production mechanism characteristics of individuals and ii) language and phonetic considerations [6].

This paper presents first certain improvements in the LHQ-Scalar quantization (LHSQ) of the LSP coefficients scheme reported in [6] and then combines LHQ with vector quantization in a new and efficient LHVQ coding of the LSP coefficients. The general LHQ approach is briefly reviewed in Section 2. The LHSQ and LHVQ schemes are described in Section 3 and their compression characteristics are discussed in Section 4.

2. LONG HISTORY QUANTIZATION (LHQ)

Conventional low bit rate coders operate on successive speech frames and derive a set of parameters for each frame. These parameters are quantized and used to drive a speech synthesis model and thus produce an approximation of the

corresponding speech frame. Optimized scalar [5] and vector [7] quantization is usually applied to model parameters. In this case, quantizers are designed and optimized according to long term statistics and are therefore fixed rather than adaptive. Fixed quantization ignores the fact that for each "signal event", in a speech frame, it is likely that another previous frame contains a very similar or exactly the same signal event. LHQ exploits this long-term redundancy present in the model parameters to obtain further compression.

In particular, consider that E_k represents the k th speech frame to be encoded and $\{P_k\}_1, 1=1, 2, \dots, M$ are M sets of model parameters "describing" E_k . In addition, let E'_{k-j} and $\{P'_{k-j}\}_1, j=1, 2, \dots, N$ be the N previously decoded speech frames and the corresponding quantized parameter sets respectively. A parameter set $\{P_k\}_1$ can be LHQ-quantized to one of the N $\{P'_{k-j}\}_1$ parameter sets representing part of the signal's history. In this way, the information transmitted regarding $\{P_k\}_1$ is the index j of the entry $\{P'_{k-j}\}_1$ of what is effectively an adaptive codebook of size N . The LHQ approach can be applied to a single parameter, individual sets of model parameters or combinations of them. It can lead to fixed or variable bit rate coding schemes and can considerably reduce the bit rate of the underlying codec. LHQ can be used together with conventional fixed scalar or vector quantization and the choice between LHQ and fixed quantization is transmitted in the form of a binary flag, $F_{k,j}$, on a frame by frame basis. Thus, $\{P_k\}_1$ is first quantized, using a fixed quantizer, to $\{P''_k\}_1$ which is then compared to the quantized entries $\{P'_{k-j}\}_1$ of the LHQ codebook. The system then decides whether to transmit $\{P''_k\}_1 = \{P''_k\}_1$ with $F_{k,j} = 0$ or the index j of the $\{P'_{k-j}\}_1 = \{P'_{k-j}\}_1$ assignment with $F_{k,j} = 1$. In this case, the LHQ codebook search is based on quantized parameter values and involves only the indices of the fixed quantizer output levels. This leads to an effective codebook search strategy which is discussed in the next section as a specific application of LHQ to LSP parameters.

The performance of an LHQ scheme is determined by the accuracy of the fixed scalar or vector quantizer employed by the system, the size of the LHQ adaptive codebook and the threshold values used in the LHQ codebook search. The LHQ codebook is updated, in our experiments, using the simple Least Recently Used (LRU) caching technique [8] which maximizes the coding relevance of the information stored in the codebook.

LSP coeff. p	Bits B _p	LSP coeff. p	Bits B _p
1	3	6	4
2	4	7	4
3	4	8	4
4	4	9	4
5	4	10	3

Table 1

Bit allocation for 38-bit fixed scalar quantization

LSP sub-vector p	Dimension	Bits B _p
1	3	10
2	3	10
3	4	10

Table 2

Bit allocation for 30-bit fixed split-VQ

3. APPLICATION OF LHQ TO LSP COEFFICIENTS

LHQ has been applied, in conjunction with scalar (LHSQ) or vector quantization (LHVQ), for coding 10 LSP coefficients. These parameters are estimated every 20ms and they are initially quantized using a total of 38 bits(scalar) or 30 bits(vector) per frame respectively, see Tables 1 and 2.

In general, the LHQ search algorithm "compares" and "attempts to match" the output index of the fixed quantizer to those stored in the LHQ codebook. The LHSQ search algorithm is described first and a discussion on the LHVQ search procedure then follows.

Consider that $SQ(O_p)_k$ represents the index (quantization output level number, $SQ(O_p)_k = 1, \dots, 2^{B_p}$) of the pth scalar quantized LSP coefficient, $C''_{k,p}$, in the kth LPC analysis frame. In the same way, $SQ(O_p)_{k,j}$ represents the index of the pth element of the jth vector, $C'_{k,j,p}$, stored in the LHQ codebook.

The LHSQ search algorithm compares sequentially $\{C''_{k,p}\}$, $p=1, 2, \dots, 10$ to the N codebook vectors $\{C'_{k,j,p}\}$, $j=1, 2, \dots, N$. This means that the process starts by comparing $SQ(O_1)_k$ of $C''_{k,1}$ to the index $SQ(O_1)_{k,1}$ of $C'_{k,1,1}$ element of the first codebook vector. If $|SQ(O_1)_k - SQ(O_1)_{k,1}| \leq T_1$, where T_1 is a fixed integer threshold, then $|SQ(O_2)_k - SQ(O_2)_{k,1}| \leq T_2$ is examined. Thus the algorithm examines the mth, $m=2, 3, \dots, 10$, inequality,

$$|SQ(O_m)_k - SQ(O_m)_{k,1}| \leq T_m \quad (1)$$

only if the (m-1)th inequality is satisfied. When (1) is not satisfied, the search on the current codebook vector is aborted and the next vector is tested. Furthermore, a codebook entry is identified as a possible candidate to represent $\{C''_{k,p}\}$ if (1) is satisfied for all $m=1, 2, \dots, 10$. When none of the LHQ codebook vectors can be identified to be useful, $\{C''_{k,p}\}$ is assigned to represent $\{C'_{k,p}\}$ and F_k is set to zero. On the other hand, one or several codebook entries may satisfy the search criterion, and a spectral distortion (SD) is formed for each candidate entry. In general,

$$SD = \sqrt{\frac{1}{N_s} \sum_{k=1}^{N_s} [\log_{10} S_k(\omega_1) - \log_{10} \hat{S}_k(\omega_1)]^2} \quad (\text{dB}) \quad (2)$$

where $S_k(\omega_1)$ and $\hat{S}_k(\omega_1)$ are the spectral values at frequency ω_1 of the original and "quantized" LPC spectra respectively.

N_s denotes the number of spectral values used from 0 to π . When several candidate vectors are available at the end of the search process, the vector with minimum SD is selected. This gives a single candidate codebook vector, and its SD value is compared to a threshold value TSD which effectively controls the "transparent" quantization of the input LSP vectors. F_k is set to 1 only when

$$SD \leq \text{TSD} \quad (3)$$

A similar search procedure is performed in the LHVQ coding of the LSP coefficients. Let $VQ(O_p)_k$ represent the index of the pth vector-quantized LSP sub-vector, $C''_{k,p}$, in the kth LPC analysis frame, $p=1, 2, 3$. In the same way, $VQ(O_p)_{k,j}$ represents the index of the pth sub-vector of the jth vector, $C'_{k,j,p}$, stored in the LHQ codebook. Unfortunately, (1) cannot be used as the search criterion in this case because the absolute difference between the indices of two LSP sub-vectors does not usually provide any information on their spectral discrepancy. In order to utilize the $\{VQ(O_p)_{k,j}\}$ indices in the LHVQ search process, a new search method has been devised. During the fixed vector quantization of $\{C_{k,p}\}$; $p=1, 2, 3$, N_s best candidate indices, $\{VQ(O_{p,q})_k$; $q=1, 2, \dots, N_s\}$, for each p sub-vector are selected. The search commences with the possible matching of any of the N_s candidate indices, $\{VQ(O_{1,q})_k\}$, produced from the quantization of the first sub-vector, $C_{1,p}$, to the $\{VQ(O_{1,q})_{k,j}\}$ elements of the LHVQ codebook. The same search process is then repeated for the second and third sets of N_s candidate indices with the corresponding $\{VQ(O_{2,q})_{k,j}\}$ and $\{VQ(O_{3,q})_{k,j}\}$ subsets of the LHVQ codebook. The entries of the LHVQ codebook for which

$$\{VQ(O_p)_{k,j}\} \in \{VQ(O_{p,q})_k$$
; $q=1, 2, \dots, N_s\} \quad (4)$

is satisfied for all 3 sub-vectors are considered further, in terms of their SD and the one with minimum SD is subsequently tested for satisfying (3). It is only then F_k is set to 1 indicating the use of LHVQ in the quantization of the input frame.

4. RESULTS AND DISCUSSION

The proposed LHSQ and LHVQ schemes have been evaluated using computer simulation and their compression potential examined. When coding 10 LSP coefficients obtained from a large database, the performance of a conventional scalar LSP quantizer and a three-way split vector quantizer is shown in Figure 1, in terms of Average Spectral Distortion (ASD) measure and the number of bits allocated every 20ms, where

$$ASD = \frac{1}{N_f} \sum_{k=1}^{N_f} SD \quad (\text{dB}) \quad (5)$$

and the ASD is measured over $N_f=12800$ frames. Transparent quantization is defined to have been obtained when i) the $ASD \leq 1\text{dB}$, ii) less than 2% of decoded frames satisfy $2\text{dB} \leq SD \leq 4\text{dB}$ and iii) none of the decoded frames has $SD > 4\text{dB}$. This is achieved with 32 or more bits per frame in the case of scalar quantization and 24 or more bits if vector quantization is to be used. With bit allocations of 38 bits and 30 bits for fixed scalar and vector quantization respectively, ASD values of 0.631 dB and 0.647 dB are obtained.

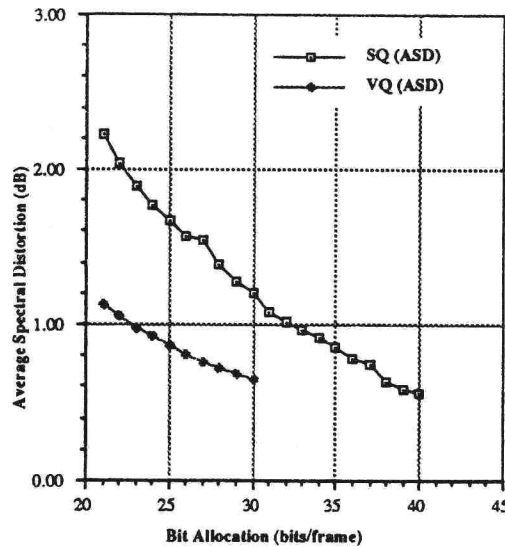


Figure 1

Figures 2, 3, 4 and 5 refer to the LHSQ coding of the LSP coefficients. Figure 2 shows the ASD performance of the system for different values of the TSD threshold used in the codebook search and for two codebook sizes, i.e. 512 and 1024. T_p is set to 2 and the LRU caching technique is used to update the codebook. As expected, higher TSD values relax the distortion constraint, imposed by the quantization algorithm, on selecting previous codebook entries, and allow the ASD of the system to increase. A TSD value of 1.6dB, however, ensures that ASD is less than 1dB. Using this TSD value and also $TSD=\infty$, which represents the system reported in [6], Figures 3, 4 and 5 illustrate the LHSQ performance, for different codebook sizes, in terms of ASD, Average Bit Rate (AvBR) and Percentage of Frames (%) with $\alpha=1$, respectively. Transparent LSP quantization is achieved with $TSD=1.6$ dB at an average bit rate of 22.4 bits per frame. Notice that the TSD assumes different "optimum" values when the fixed quantization bit allocation given in Table 1 is modified.

The performance of the LHVQ scheme described in the previous section is summarized in Table 3. Again, TSD is set to 1.6dB, while $N_s = 60$. It has been found that LHVQ achieves transparent LSP coding with an average bit rate of 19 bits per frame.

Notice that both the LHSQ and LHVQ algorithms described in this paper represent a particular variable bit rate implementation of the LHQ approach, which can easily be integrated within the framework of an LPC based constant bit rate codec. Furthermore, the codebook search in both algorithms is mainly based on the comparison of codeword indices and may therefore be implemented efficiently.

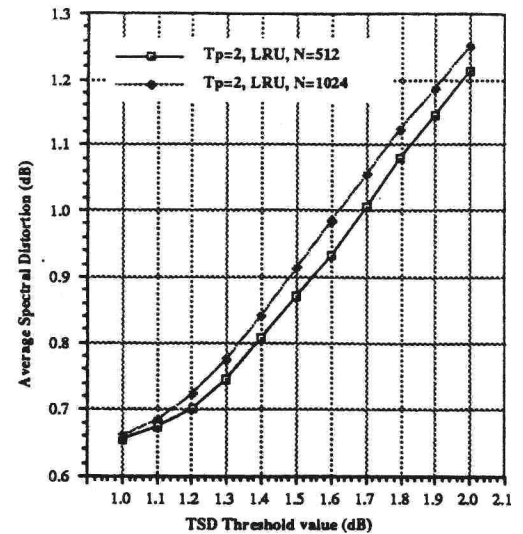


Figure 2

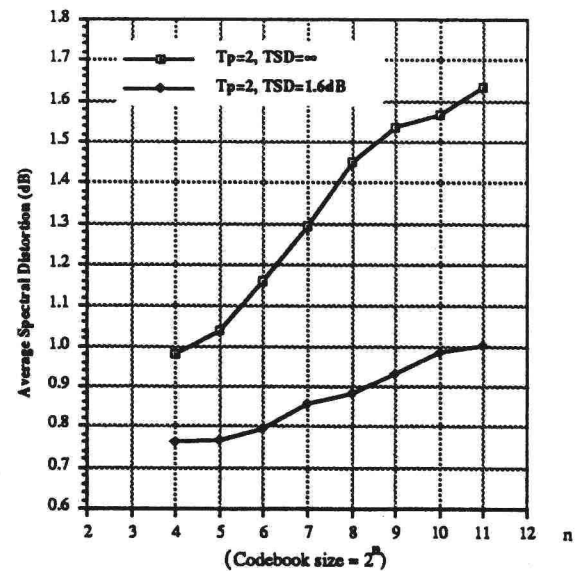


Figure 3

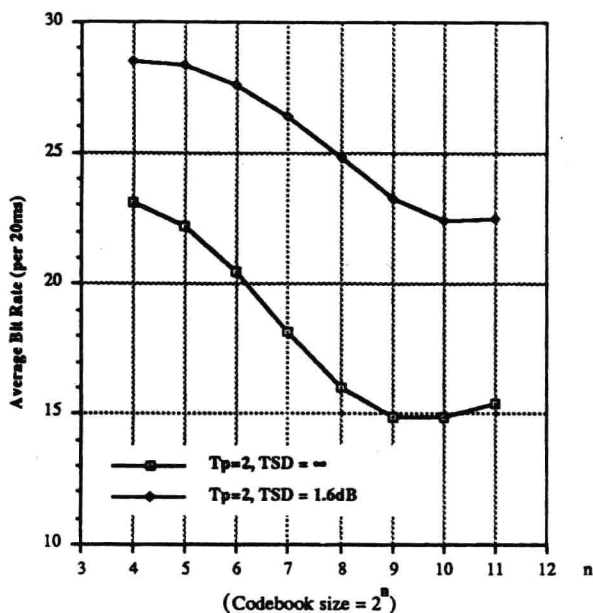


Figure 4

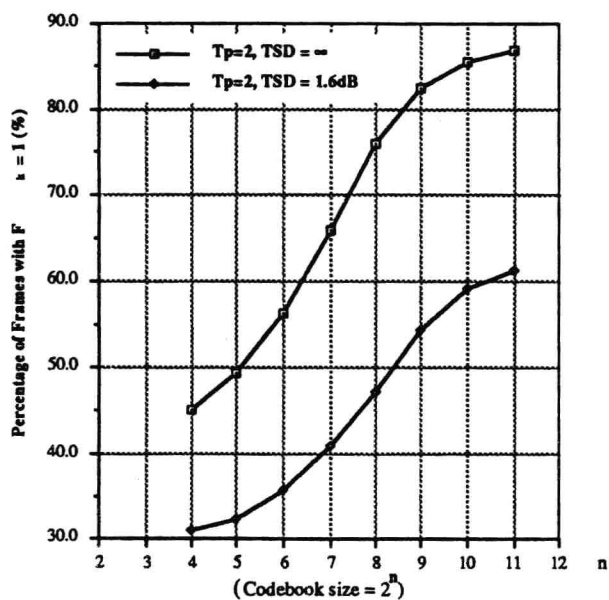


Figure 5

Codebook size, N	ASD (dB)	% of F _k =1	Average Bit Rate
16	0.791	34.49	22.03
32	0.804	36.25	21.94
64	0.829	39.37	21.55
128	0.866	44.19	20.84
256	0.916	50.12	19.97
512	0.968	56.37	19.16
1024	1.019	61.81	18.64

Table 3

REFERENCES

- [1] C. Kuo, F. Jean and H. Wang, "Low Bit Rate Quantization of LSP Parameters using Two-dimensional Differential Coding", IEEE Proc. ICASSP, pp. 197-200, 1992.
- [2] J. Grass and P. Kabal, "Methods of improving Vector-Scalar Quantization of LPC Coefficients", IEEE Proc. ICASSP, pp. 657-660, 1991.
- [3] R. Laroia, N. Phamdo and N. Farvardin, "Robust and Efficient Quantization of Speech LSP Parameters using Structured Vector Quantizers", IEEE Proc. ICASSP, pp. 641-644, 1991.
- [4] N. Phamdo and N. Farvardin, "Coding of Speech LSP Parameters using TSVQ with Interblock Noiseless Coding", IEEE Proc. ICASSP, pp. 193-196, 1990.
- [5] N. Sugamura and N. Farvardin, "Quantizer Design in LSP Speech Analysis and Synthesis", IEEE Proc. ICASSP, pp. 398-401, 1988.
- [6] C. Xydeas and K. So, "A Long History Quantization Approach to Low Bit Rate Speech Coding", EURASIP Proc. EUSIPCO, vol. 1, pp. 479-482, 1992.
- [7] K. Paliwal and B. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame", IEEE Proc. ICASSP, pp. 661-664, 1991.
- [8] K. Kwang and F. Briggs, "Computer Architecture and Parallel Processing", McGraw Hill Computer Science Series, pp. 98-118, 1987.

EFFICIENT CODING OF LPC PARAMETERS USING ADAPTIVE PREFILTERING AND MSVQ WITH PARTIALLY ADAPTIVE CODEBOOK

Y. Tanaka and T. Taniguchi

Fujitsu Laboratories Ltd.

1015 Kamikodanaka, Nakahara-ku, Kawasaki, 211 Japan

ABSTRACT

In this paper, we propose an efficient vector quantization scheme and a new LPC analysis scheme, both of which exploit interframe correlation in the successive spectrum envelope of speech signals. The first quantization scheme proposed is a multi-stage vector quantization of LSP parameters with a partially adaptive codebook (MSVQ-AC). The second new algorithm is an LPC analysis scheme, with closed-loop adaptive prefiltering (LPC-PF), which realizes temporary higher order analysis than the standard LPC with a few additional transmission bits. A combined system of the LPC-PF and 2-split, 2-stage VQ with the adaptive codebook can quantize 10-th order LSP parameters at around 23 bits/frame, realizing sufficient quality and reasonable complexity.

1. INTRODUCTION

Linear predictive coding is widely used as a short time spectral envelope estimation in various speech processing applications. For low bit rate speech coding, it is important to quantize LPC parameters using as few bits as possible without sacrificing the speech quality and within a reasonable complexity. Various quantization schemes have been proposed for such objectives. Scalar quantization of individual coefficients results in acceptable levels of spectral distortion at 32 to 36 bits/frame. Vector quantization is a more efficient scheme which utilizes intraframe correlation among the LPC parameters [1]. For higher efficiency, it is necessary to exploit the interframe correlation of successive parameter sets. Shoham studied VQ with vector predictive quantization (VPQ) [2]. In their VQ, the current spectral envelope is predicted using past quantized spectra and its prediction residual spectrum is quantized. Matrix quantization (MQ) is a direct scheme to quantize a set of parameter vectors [3]. MQ is a very expensive scheme, however, in terms of computation and memory and in that it introduces a large encoding delay. Farvardin et al. studied two-dimensional DCT coding of LSP vectors [4]. Encoding is possible at 21 bits/frame by using appropriate bit allocation based on the distribution of each DCT coefficient. However, this scheme also introduces a large encoding delay. Grass et al. studied vector-scalar quantization with an adaptive codebook, which does not require extra transmission information or encoding delays [5]. We have extended this idea for a multi-stage VQ framework as described in Section 2.

On the other hand, an LPC analysis itself is usually performed in a memoryless fashion, and no one has yet considered using the interframe correlation of spectral envelope in the analysis stage.

In this paper, we propose a new efficient vector quantization scheme and a new LPC analysis scheme, both of which

exploit interframe correlation of the spectrum. The new quantization scheme is a multi-stage vector quantization with a partially adaptive codebook (MSVQ-AC). The first stage codebook has a small adaptive part as well as a large fixed part, which stores the quantized parameters of the past frames.

To exploit the interframe correlation in the analysis stage, a new LPC analysis scheme called LPC-PF, which employs closed-loop adaptive prefiltering prior to LPC analysis, is also proposed.

This paper is organized as follows. In Section 2, as a practical implementation of MSVQ-AC, we present a 2-split, 2-stage VQ with an adaptive codebook for 10-th order LSP parameters. In Section 3, we present a detailed algorithm and do a performance evaluation of LPC-PF.

2. MULTI-STAGE VQ WITH PARTIALLY ADAPTIVE CODEBOOK (MSVQ-AC)

2.1 Configuration

A multi-stage quantization structure can make use of a partially adaptive codebook [5] to exploit interframe correlation. A multi-stage VQ configuration, with this partially adaptive codebook, is shown in Fig. 1. The codebook of first stage VQ has fixed and adaptive parts. The adaptive part stores the quantized vectors of the past $(n-J)$ -th frames, where J is the adaptive codebook size. It is sufficient to add a small size adaptive part to a fixed codebook to efficiently exploit interframe correlation, because a coefficient vector with the highest correlation to that of the current frame can be usually found within the parameter sets of the most recent frames.

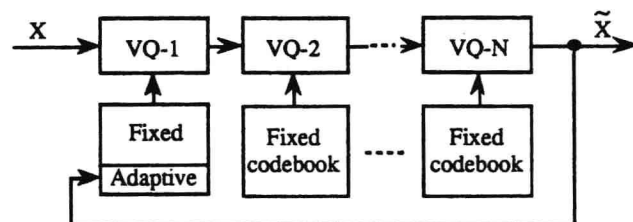


Figure 1. MSVQ with partially adaptive codebook

Before designing the VQ, we assumed that more than 20 bits/frame will be necessary to quantize 10-th order LSP parameters with an acceptable distortion even if a highly efficient VQ is used. Therefore, to reduce the VQ's computational complexity and memory requirements to within a reasonable range, we considered it preferable to decompose the entire quantization process into more than three parts using methods, such as MSVQ or split VQ.

In our study, we have developed a 2-split, 2-stage VQ (Fig. 2). At the first stage, the 10-dimensional LSP vector is split into 2 subvectors, one of length 4 (lower band) and the other of length 6 (higher band). Each subvector is separately quantized. Split VQ has been reported as an efficient scheme for LSP quantization due to its localized spectral sensitivity to quantization error [6]. At the second stage, the 10-dimensional error vector between the input vector and the reconstructed vector from the first stage VQ is quantized. The distance measure used in the search of this VQ is the same as in reference [6].

In the first stage split VQ, both the lower band (CB1-L) and higher band (CB1-H) codebooks have small independent adaptive parts. This configuration allows separate searches of the adaptive codebooks in the lower and higher bands.

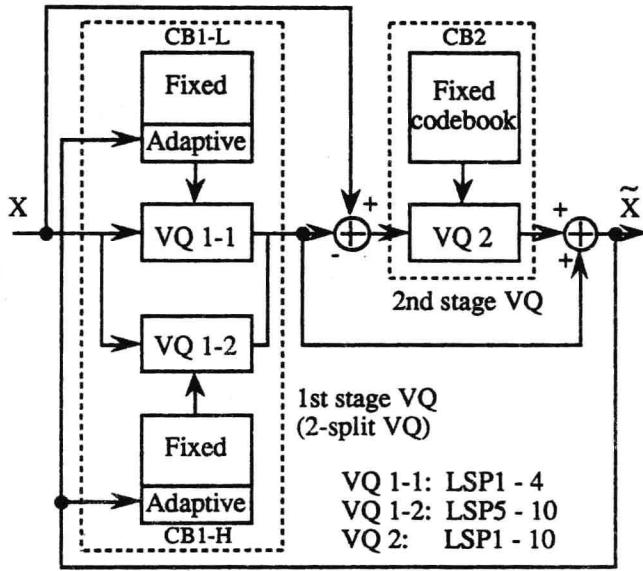


Figure 2. 2-split, 2-stage LSP-VQ with partially adaptive codebook

2.2 Codebook design

Fixed part of each codebook is designed by using the LBG algorithm on the training data. A weighted Euclidian distance is used for the design as well as for the codebook search. For training the second stage codebook, we found that the quantization error vectors of the first stage, whose spectral distortion is greater than a certain threshold value, are useful for the training data to reduce the percentage of outliers. We used a simulation to determine the appropriate threshold of 1.5 dB.

2.3 Codevector reconstruction in split VQ

In a sequential search of CB1-L and CB1-H, invalid unstable combinations of codevectors are present which should be excluded in the CB1-H search. To prevent the loss of valid codebook size from this, CB1-H codevectors are reconstructed. In this method, unstable codewords are modified to satisfy the LSP parameter stability conditions by nonlinearly transforming each coefficient according to the selected CB1-L codevector.

2.4 Evaluation

Thirty minutes of Japanese speech was used as the database in this study. The first 25 minutes were used for codebook training and the last 5 minutes were used for evaluation. 10-th

order LPC analysis, based on the autocorrelation method with high frequency compensation, was performed every 20 ms using a 20 ms Hamming window.

The performance is summarized in Table 1. Average spectral distortion (SD) and outlier percentage (calculated for frames having SD in the range of 2-4 dB and for SD above 4 dB) are used for evaluation. The VQ with second stage codebook training using first stage outliers reduces the outliers percentage with slight SD degradation. Average spectral distortion equivalent to 30 bits/frame SQ was achieved at 24 bits/frame (8 bits for each codebook) for the proposed VQ.

The appropriate adaptive codebook size was determined experimentally. Improvement in performance saturated when the size was larger than 4. So, we set the size to 4 in the following simulation. Using the adaptive codebook significantly reduces outliers as well as average spectral distortion. The proposed 24 bits/frame VQ seems to be sufficient for speech coding applications.

Table 1. Performance of various LSP quantizers

Quantization scheme	Bits/20 ms	SD (dB)	Outlier (%)	
			2dB<SD<4dB	4dB<SD
SQ	34	0.97	1.87	0.0
	32	1.12	4.26	0.0
	30	1.23	5.63	0.01
	28	1.38	9.36	0.04
3-MSVQ	24	1.21	5.81	0.0
2-split VQ	24	1.25	5.83	0.0
2-split, 2-stage VQ	24	1.23	3.81	0.0
2-split, 2-stage VQ with O.T.	24	1.25	3.53	0.0
2-split, 2-stage VQ with O.T. & A.C. (M=4)	23	1.26	3.19	0.0
	24	1.19	1.90	0.0
	26	1.10	0.81	0.0

O.T.: 2nd stage codebook is trained by outliers from the 1st stage VQ
A.C.: Adaptive codebook

2.5 Complexity

The complexity of the proposed 24 bits/frame VQ (with three 8-bit codebooks) is 8 times smaller than that of the conventional 2-split VQ, since it only requires three 8-bit codebook searches rather than two 12-bit codebook searches. Moreover, the memory required for codebook storage is also reduced from 40 K words to 5 K words.

3. LPC ANALYSIS WITH ADAPTIVE PREFILTERING (LPC-PF)

3.1 LPC-PF configuration

This section presents an efficient LPC analysis scheme, LPC-PF, which is an alternate scheme for exploiting interframe correlation. The LPC-PF configuration is shown in Fig. 3. In this scheme, LPC analysis and quantization are included in an adaptive feedback loop. The input speech signal is first fed into a low order prefilter, which is a kind of an inverse filter whose coefficients are determined from past quantized LPC parameters. This prefilter preceding LPC analysis performs rough spectral flattening of the input signal. The overall spectrum envelope, $H(z)$, estimated by this system is represented by the cascade characteristics of the prefilter $P(z)$ and the following LPC analysis filter $A(z)$ as follows.