

# Algorithms and Software for Predictive and Perceptual Modeling of Speech

Venkatraman Atti  
Verance Corporation

*SYNTHESIS LECTURES ON ALGORITHMS AND SOFTWARE ENGINEERING*  
#7



MORGAN & CLAYPOOL PUBLISHERS

## ABSTRACT

From the early pulse code modulation-based coders to some of the recent multi-rate wideband speech coding standards, the area of speech coding made several significant strides with an objective to attain high quality of speech at the lowest possible bit rate. This book presents some of the recent advances in linear prediction (LP)-based speech analysis that employ perceptual models for narrow- and wide-band speech coding.

The LP analysis-synthesis framework has been successful for speech coding because it fits well the source-system paradigm for speech synthesis. Limitations associated with the conventional LP have been studied extensively, and several extensions to LP-based analysis-synthesis have been proposed, e.g., the discrete all-pole modeling, the perceptual LP, the warped LP, the LP with modified filter structures, the IIR-based pure LP, all-pole modeling using the weighted-sum of LSP polynomials, the LP for low frequency emphasis, and the cascade-form LP. These extensions can be classified as algorithms that either attempt to improve the LP spectral envelope fitting performance or embed perceptual models in the LP. The first half of the book reviews some of the recent developments in predictive modeling of speech with the help of Matlab™ Simulation examples.

Advantages of integrating perceptual models in low bit rate speech coding depend on the accuracy of these models to mimic the human performance and, more importantly, on the achievable “coding gains” and “computational overhead” associated with these physiological models. Methods that exploit the masking properties of the human ear in speech coding standards, even today, are largely based on concepts introduced by Schroeder and Atal in 1979. For example, a simple approach employed in speech coding standards is to use a perceptual weighting filter to shape the quantization noise according to the masking properties of the human ear. The second half of the book reviews some of the recent developments in perceptual modeling of speech (e.g., masking threshold, psychoacoustic models, auditory excitation pattern, and loudness) with the help of Matlab™ simulations. Supplementary material including Matlab™ programs and simulation examples presented in this book can also be accessed at [www.morganclaypool.com/page/atti](http://www.morganclaypool.com/page/atti)

## KEYWORDS

Linear Prediction (LP), speech coding standards, perceptual LP, warped LP, cascade-form LP, predictive modeling, LP parameter transformation, psychoacoustic model, masking asymmetry, auditory excitation pattern, and perceptual loudness

# Bibliography

- [1] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Amer.*, vol. 66, no. 6, p. 1647, 1979. DOI: [10.1121/1.383662](https://doi.org/10.1121/1.383662) 1, 3, 55
- [2] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 27, no. 3, pp. 247–253, June 1979. DOI: [10.1109/TASSP.1979.1163237](https://doi.org/10.1109/TASSP.1979.1163237) 1, 4, 10, 53
- [3] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High quality speech at very low bit rates," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 10, Apr. 1985, pp. 937–940. DOI: [10.1109/ICASSP.1985.1168147](https://doi.org/10.1109/ICASSP.1985.1168147) 1, 3
- [4] B. S. Atal and M. R. Schroeder, "Stochastic coding of speech signals at very low bit rates," in *Proc. Int. Conf. Comm.*, May 1984, pp. 1610–1613. DOI: [10.1016/0167-6393\(85\)90043-3](https://doi.org/10.1016/0167-6393(85)90043-3) 1, 3
- [5] ITU, *Coding of Speech at 8kb/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)*, ITU Study Group 15 Std. Draft Recommendation G.729, 1995. 1, 4, 6, 7, 50
- [6] TIA, *Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communications Systems*, TIA Std. TIA IS-893, 2004. 1, 6, 7
- [7] TIA, *Enhanced variable rate codec, Speech service option 3 for wideband spread spectrum digital systems*, TIA Std. TIA IS-127, 1997. 1, 6, 7
- [8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 27, no. 2, pp. 113–120, Apr. 1979. DOI: [10.1109/TASSP.1979.1163209](https://doi.org/10.1109/TASSP.1979.1163209)
- [9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984. DOI: [10.1109/TASSP.1984.1164453](https://doi.org/10.1109/TASSP.1984.1164453)
- [10] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 443–445, Apr. 1985. DOI: [10.1109/TASSP.1985.1164550](https://doi.org/10.1109/TASSP.1985.1164550) 1

- [11] J. D. Markel and A. H. Gray, *Linear prediction of speech*. Springer, 1976. 2, 9, 10, 13, 16, 30
- [12] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, no. 2, pp. 637–655, 1971. DOI: 10.1121/1.1912679 2, 10, 15, 16
- [13] D. Florencio, "Investigating the use of asymmetric windows in CELP vocoders," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, Apr. 1993, pp. 427–430. DOI: 10.1109/ICASSP.1993.319331 2
- [14] F. Itakura, "Line spectrum representation of linear predictive coefficients," *J. Acoust. Soc. Amer.*, vol. 57, no. 1, p. S35, 1975. DOI: 10.1121/1.1995189 3, 10
- [15] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 1, pp. 3–14, Jan. 1992. DOI: 10.1109/89.221363 3, 6, 10, 16, 49
- [16] P. Kroon and W. B. Kleijn, "Linear prediction-based analysis-synthesis coding," in *Speech coding and synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier Science, 1995, ch. 3, pp. 79–119. 4, 16, 19
- [17] R. Salami *et al.*, "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder," *IEEE Trans. Speech Audio Processing*, vol. 6, no. 2, pp. 116–130, Mar. 1998. DOI: 10.1109/89.661471 3, 4, 5, 6, 7, 13
- [18] B. S. Atal and J. Remde, "A new model for LPC excitation for producing natural sounding speech at low bit rates," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 7, Apr. 1982, pp. 614–617. DOI: 10.1109/ICASSP.1982.1171649 3
- [19] S. Singhal and B. S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 9, Mar. 1984, pp. 9–12.
- [20] P. Kroon and B. S. Atal, "Predictive coding of speech using analysis-by-synthesis techniques," in *Twenty-Fourth Asilomar Signals, Systems and Computers Conference*, vol. 2, Nov. 1990, pp. 664–668. DOI: 10.1109/ACSSC.1990.523421 3
- [21] ISO, *Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mb/s, Part 3: Audio*, ISO/IEC JTC1/SC29/WG11 MPEG Std. IS 11 172–3, 1992. 3, 53, 54, 64
- [22] ISO, *Information Technology - Generic Coding of Moving Pictures and Associated Audio, Part 3: Audio*, ISO/IEC JTC1/SC29/WG11 MPEG Std. IS13 818–3, 1994.
- [23] ISO, *Generic Coding of Moving Pictures and Associated Audio: Audio (non backwards compatible coding, MPEG-2 NBC/AAC)*, ISO/IEC JTC1/SC29/WG11 MPEG Std. Committee Draft 13 818–7, 1996. 3

- [24] G. Davidson, "Digital audio coding: Dolby AC-3," in *The Digital Signal Processing Handbook*, V. Madisetti and D. Williams, Eds. CRC Press, 1998, ch. 41, pp. 1–21. [3, 53](#)
- [25] M. Davis, "The AC-3 multichannel coder," in *Proc. 95th Conv. Aud. Eng. Soc.*, no. 3774, 1993. [53](#)
- [26] S. Smyth, W. Smith, M. Smyth, M. Yan, and T. Jung, "DTS coherent acoustics delivering high quality multichannel sound to the consumer," in *Proc. 100th Conv. Aud. Eng. Soc.*, no. 4293, 1996. [3, 53](#)
- [27] Y. Medan, E. Yair, and D. Chazan, "Super resolution pitch determination of speech signals," *IEEE Trans. Signal Processing*, vol. 39, no. 1, pp. 40–48, Jan. 1991. [DOI: 10.1109/78.80763](#) [4](#)
- [28] W. C. Chu, *Speech Coding Algorithms*. New York: Wiley-Interscience, 2003. [4](#)
- [29] ITU, *Dual Rate Speech Coder for Multimedia Communications transmitting at 5.3 and 6.3 kb/s*, ITU Std. Draft Recommendation G.723.1, 1995. [4, 6, 7](#)
- [30] A. Kataoka, T. Moriya, and S. Hayashi, "An 8-kb/s conjugate structure CELP (CS-ACELP) speech coder," *IEEE Trans. Speech Audio Processing*, vol. 4, no. 6, pp. 401–411, Nov. 1993. [DOI: 10.1109/89.544525](#) [4, 6](#)
- [31] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol. 30, no. 4, pp. 600–614, Apr. 1982. [DOI: 10.1109/TCOM.1982.1095501](#) [4](#)
- [32] J. P. Adoul *et al.*, "Fast CELP coding based on algebraic codes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 12, Apr. 1987, pp. 1957–1960. [DOI: 10.1109/ICASSP.1987.1169413](#) [4, 6](#)
- [33] C. Laˆamme *et al.*, "On reducing computational complexity of codebook search in CELP coder through the use of algebraic codes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 1, Apr. 1990, pp. 177–180. [DOI: 10.1109/ICASSP.1990.115567](#) [4](#)
- [34] J.-H. Chen and A. Gersho, "Adaptive post-filtering for quality enhancement of coded speech," *IEEE Trans. Speech Audio Processing*, vol. 3, no. 1, pp. 59–70, Jan. 1995. [DOI: 10.1109/89.365380](#) [5, 6](#)
- [35] V. Atti and A. Spanias, "Speech analysis by estimating perceptually relevant pole locations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 1, Philadelphia, PA, Mar. 2005, pp. 217–220. [DOI: 10.1109/ICASSP.2005.1415089](#)
- [36] V. Atti and A. Spanias, "Rate determination based on perceptual loudness," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 2, May 2005, pp. 848–851. [DOI: 10.1109/ISCAS.2005.1464721](#)

## 102 BIBLIOGRAPHY

- [37] B. Wei and J. Gibson, "A new discrete spectral modeling method and an application to CELP coding," *IEEE Signal Processing Lett.*, vol. 10, no. 4, pp. 101–103, Apr. 2003. DOI: [10.1109/LSP.2003.808550](https://doi.org/10.1109/LSP.2003.808550) 10, 34, 35
- [38] D. Petrinovic, "Discrete weighted mean square all-pole modeling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 1, Apr. 2003, pp. 828–831. DOI: [10.1109/ICASSP.2003.1198909](https://doi.org/10.1109/ICASSP.2003.1198909) 10, 34, 35
- [39] P. Alku and T. Backstrom, "Linear predictive method for improved spectral modeling of lower frequencies of speech with small prediction orders," *IEEE Trans. Speech Audio Processing*, vol. 12, no. 2, pp. 93–99, Mar. 2004. DOI: [10.1109/TSA.2003.822625](https://doi.org/10.1109/TSA.2003.822625) 10, 11, 41
- [40] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. Amer.*, vol. 87, no. 4, pp. 1738–1752, Apr. 1990. DOI: [10.1121/1.399423](https://doi.org/10.1121/1.399423) 10, 14, 19, 20, 21
- [41] H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. Amer.*, vol. 68, no. 4, pp. 1071–1076, Oct. 1980. DOI: [10.1121/1.384992](https://doi.org/10.1121/1.384992) 10, 14, 21
- [42] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Processing*, vol. 39, no. 2, pp. 411–423, Feb. 1991. DOI: [10.1109/78.80824](https://doi.org/10.1109/78.80824) 10, 30, 33, 34, 35
- [43] E. Zwicker and H. Fastl, *Psychoacoustics Facts and Models*. Springer-Verlag, 1990. 14, 18, 53, 58, 84, 85
- [44] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. Academic Press, 2003. 14, 18, 53, 58, 84
- [45] B. Bessette *et al.*, "The adaptive multirate wideband speech codec (AMR-WB)," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 8, pp. 620–636, Nov. 2002. DOI: [10.1109/TSA.2002.804299](https://doi.org/10.1109/TSA.2002.804299) 6, 7, 8, 10, 49, 55
- [46] A. Harma and U. K. Laine, "A comparison of warped and conventional linear prediction," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 5, pp. 579–588, July 2001. DOI: [10.1109/89.928922](https://doi.org/10.1109/89.928922) 10, 21, 29, 30
- [47] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. of the IEEE*, vol. 88, no. 4, pp. 451–513, Apr. 2000. DOI: [10.1109/5.842996](https://doi.org/10.1109/5.842996) 61, 63
- [48] J. Makhoul, "Linear prediction: A tutorial review," *Proc. of the IEEE*, vol. 63, no. 5, pp. 561–580, Apr. 1975. DOI: [10.1109/PROC.1975.9792](https://doi.org/10.1109/PROC.1975.9792) 9, 10, 13
- [49] J. Makhoul, "Spectral linear prediction: properties and applications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 23, no. 3, pp. 283–296, June 1975. DOI: [10.1109/TASSP.1975.1162685](https://doi.org/10.1109/TASSP.1975.1162685) 10, 30

- [50] R. J. McAulay, "Maximum likelihood spectral estimation and its application to narrow-band speech coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 2, pp. 243–251, Apr. 1984. DOI: [10.1109/TASSP.1984.1164318](https://doi.org/10.1109/TASSP.1984.1164318)
- [51] C. Lee, "Robust linear prediction for speech analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 12, Apr. 1987, pp. 289–292. DOI: [10.1109/ICASSP.1987.1169680](https://doi.org/10.1109/ICASSP.1987.1169680)
- [52] J. Picone, "Joint estimation of the LPC parameters and the multipulse excitation," *Speech Commun.*, vol. 5, no. 3/4, pp. 253–260, Dec. 1986. DOI: [10.1016/0167-6393\(86\)90012-9](https://doi.org/10.1016/0167-6393(86)90012-9)
- [53] R. Mizoguchi, "Speech analysis by selective linear prediction in the time domain," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 3, May 1982, pp. 1573–1576. DOI: [10.1109/ICASSP.1982.1171428](https://doi.org/10.1109/ICASSP.1982.1171428)
- [54] F. Itakura and S. Saito, "An analysis-synthesis telephony based on maximum likelihood method," in *Proc. 6th Int. Conf. Acoustics*, 1968, pp. 17–20. 10
- [55] A. Gray and J. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, no. 5, pp. 380–391, Oct. 1976. DOI: [10.1109/TASSP.1976.1162849](https://doi.org/10.1109/TASSP.1976.1162849)  
30
- [56] TIA, *Speech service option standard for wideband spread spectrum digital cellular system*, TIA Std. TIA IS-96, 1996. 6, 7
- [57] A. Das, E. Paksoy, and A. Gersho, "Multimode and variable-rate coding of speech," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier Science, 1995, ch. 7, pp. 257–288.
- [58] A. Gersho and E. Paksoy, "Variable rate speech coding for cellular networks," in *Speech and Audio Coding for Wireless and Network Applications*, B. A. V. Cuperman and A. Gersho, Eds. Kluwer Academic Publishers, 1993, ch. 10, pp. 77–84. 7
- [59] S. Ramprasad, "The multimode transform predictive coding paradigm," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 2, pp. 117–129, Mar. 2003. DOI: [10.1109/TSA.2003.809195](https://doi.org/10.1109/TSA.2003.809195) 6
- [60] R. Vafin and W. B. Kleijn, "Rate-distortion optimized quantization in multistage audio coding," *IEEE Trans. Speech Audio Processing*, vol. 14, no. 1, pp. 311–320, Jan. 2006. DOI: [10.1109/TSA.2005.854104](https://doi.org/10.1109/TSA.2005.854104) 6
- [61] A. Spanias, "Speech coding: A tutorial review," *Proc. of the IEEE*, vol. 82, no. 10, pp. 1541–1582, Oct. 1994. DOI: [10.1109/5.326413](https://doi.org/10.1109/5.326413)
- [62] Y. Bistrantz and S. Pellerin, "Immittance spectral pairs (ISP) for speech encoding," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, Apr. 1993, pp. 9–12. DOI: [10.1109/ICASSP.1993.319215](https://doi.org/10.1109/ICASSP.1993.319215) 10, 49

- [63] A. Harma and U. K. Laine, "Linear predictive coding with modified filter structures," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 8, pp. 769–777, Nov. 2001. DOI: [10.1109/89.966080](https://doi.org/10.1109/89.966080) 10, 35, 36
- [64] A. C. den Brinker *et al.*, "IIR-based pure linear prediction," *IEEE Trans. Speech Audio Processing*, vol. 12, no. 1, pp. 68–75, Jan. 2004. DOI: [10.1109/TSA.2003.815524](https://doi.org/10.1109/TSA.2003.815524) 10, 11, 37
- [65] T. Backstrom and P. Alku, "All-pole modeling technique based on weighted sum of LSP polynomials," *IEEE Signal Processing Lett.*, vol. 10, no. 6, pp. 180–183, 2003. DOI: [10.1109/LSP.2003.811635](https://doi.org/10.1109/LSP.2003.811635) 11, 37, 38
- [66] L. B. Jackson and S. L. Wood, "Linear prediction in cascade form," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 26, no. 6, pp. 518–528, Dec. 1978. DOI: [10.1109/TASSP.1978.1163155](https://doi.org/10.1109/TASSP.1978.1163155) 11, 42, 44, 47
- [67] W. C. Chu, "Window optimization in linear prediction analysis," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 626–635, Nov. 2003. DOI: [10.1109/TSA.2003.818213](https://doi.org/10.1109/TSA.2003.818213) 13
- [68] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Prentice Hall PTR, 1993. 18, 35, 58
- [69] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Processing*, vol. 7, no. 6, pp. 697–708, Nov. 1999. DOI: [10.1109/89.799695](https://doi.org/10.1109/89.799695) 18, 20, 25, 58
- [70] B. C. J. Moore and B. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Amer.*, vol. 74, pp. 750–753, 1983. DOI: [10.1121/1.389861](https://doi.org/10.1121/1.389861) 53, 58, 85, 87
- [71] H. Hermansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 4, pp. 578–589, Oct. 1994. DOI: [10.1109/89.326616](https://doi.org/10.1109/89.326616) 21
- [72] J.-C. Junqua, H. Wakita, and H. Hermansky, "Evaluation and optimization of perceptually-based ASR front-end," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 1, pp. 39–48, Jan. 1993. DOI: [10.1109/89.221366](https://doi.org/10.1109/89.221366) 21
- [73] A. Oppenheim, D. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the fast Fourier transform," *Proc. of the IEEE*, vol. 59, no. 2, pp. 299–301, Feb. 1971. DOI: [10.1109/PROC.1971.8146](https://doi.org/10.1109/PROC.1971.8146) 21
- [74] E. Kruger and H. Strube, "Linear prediction on a warped frequency scale," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 9, pp. 1529–1531, 1988. DOI: [10.1109/29.90384](https://doi.org/10.1109/29.90384) 21
- [75] U. Laine, M. Karjalainen, and T. Altonsaar, "WLP in speech and audio processing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 3, 1994, pp. 349–352. DOI: [10.1109/ICASSP.1994.390018](https://doi.org/10.1109/ICASSP.1994.390018) 21

- [76] A. Harma and U. K. Laine, "Warped low-delay CELP for wideband audio coding," in *Proc. AES 17th Int. Conf. High-Quality Audio Coding*, 1999, pp. 207–215. DOI: [10.1109/ICASSP.1998.681674](https://doi.org/10.1109/ICASSP.1998.681674) 6, 21, 29
- [77] A. Harma, "Implementation of recursive filters having delay free loops," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 3, May 1998, pp. 1261–1264. DOI: [10.1109/ICASSP.1998.681674](https://doi.org/10.1109/ICASSP.1998.681674) 29
- [78] A. Harma, "Implementation of frequency-warped recursive filters," *Signal Processing*, vol. 80, pp. 543–548, Feb. 2000. DOI: [10.1016/S0165-1684\(99\)00151-6](https://doi.org/10.1016/S0165-1684(99)00151-6) 29
- [79] R. Gray, A. Buzo, A. Gray, and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, no. 4, pp. 367–376, Aug. 1980. DOI: [10.1109/TASSP.1980.1163421](https://doi.org/10.1109/TASSP.1980.1163421) 30
- [80] A. El-Jaroudi and J. Makhoul, "Discrete pole-zero modeling and applications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, May 1989, pp. 2162–2165. DOI: [10.1109/ICASSP.1989.266891](https://doi.org/10.1109/ICASSP.1989.266891) 34, 35
- [81] B. Wei and J. Gibson, "Comparison of distance measures in discrete spectral modeling," in *Proc. IEEE Dig. Sig. Proc. Workshop*, 2000. 35
- [82] W. Kautz, "Transient synthesis in the time domain," in *IEEE Trans. Circuit Theory*, vol. 1, 1954, pp. 29–39. DOI: [10.1109/TCT.1954.1083588](https://doi.org/10.1109/TCT.1954.1083588) 36, 37
- [83] R. King and P. Paraskevopoulos, "Digital Laguerre filters," in *J. Circuit Theory Applications*, vol. 5, 1977, pp. 81–91. DOI: [10.1002/cta.4490050108](https://doi.org/10.1002/cta.4490050108) 37
- [84] P. Stoica and A. Nehorai, "The poles of symmetric linear prediction models lie on the unit circle," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1419–1426, Dec. 1986. DOI: [10.1109/TASSP.1986.1164937](https://doi.org/10.1109/TASSP.1986.1164937) 41, 42
- [85] T. Backstrom, P. Alku, T. Paatero, and B. Kleijn, "A time-domain interpretation for the LSP decomposition," *IEEE Trans. Speech Audio Processing*, vol. 12, no. 6, pp. 554–559, Nov. 2004. DOI: [10.1109/TSA.2004.834470](https://doi.org/10.1109/TSA.2004.834470) 42
- [86] P. C. Ching and C. C. Goodyear, "Linear prediction in cascade form," *IEE Proc. pt. E*, vol. 130, 1983. DOI: [10.1109/TASSP.1978.1163155](https://doi.org/10.1109/TASSP.1978.1163155) 42
- [87] A. Nehorai and D. Starer, "Adaptive pole estimation," *IEEE Trans. Signal Processing*, vol. 38, no. 5, pp. 825–838, May 1990. DOI: [10.1109/29.56028](https://doi.org/10.1109/29.56028) 42
- [88] R. Yu and C. C. Ko, "Lossless compression of digital audio using cascaded RLS-LMS prediction," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 532–537, Nov. 2003. DOI: [10.1109/TSA.2003.818111](https://doi.org/10.1109/TSA.2003.818111) 42

- [89] G. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual audio coding using adaptive pre- and post-filters and lossless compression," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 6, pp. 379–390, Sept. 2002. DOI: [10.1109/TSA.2002.803444](https://doi.org/10.1109/TSA.2002.803444)
- [90] P. Prandoni and M. Vetterli, "An FIR cascade structure for adaptive linear prediction," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2566–2571, Sept. 1998. DOI: [10.1109/78.70954842](https://doi.org/10.1109/78.70954842)
- [91] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, no. 4, pp. 299–309, Aug. 1977. DOI: [10.1109/TASSP.1977.1162974](https://doi.org/10.1109/TASSP.1977.1162974)
- [92] K. Rao and P. Yip, *The Discrete Cosine Transform: Algorithm, Advantages, and Applications*. Academic Press, 1990.
- [93] Y. Linde, A. Buzo, and A. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan. 1980. DOI: [10.1109/TCOM.1980.1094577](https://doi.org/10.1109/TCOM.1980.1094577)
- [94] W. Chan and A. Gersho, "Constrained-storage quantization of multiple vector sources by codebook sharing," *IEEE Trans. Commun.*, vol. 39, no. 1, pp. 11–13, Jan. 1991. DOI: [10.1109/26.68269](https://doi.org/10.1109/26.68269)
- [95] T. Moriya *et al.*, "Extension and complexity reduction of TWINVQ audio coder," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, 1996. DOI: [10.1109/ICASSP.1996.543299](https://doi.org/10.1109/ICASSP.1996.543299)
- [96] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
- [97] ITU, *Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss*, ITU Std. Draft Recommendation G.722.1, 1999.
- [98] ITU, *7 KHz Audio Coding within 64 kbits/s*, ITU Std. Draft Recommendation G.722, 1988. 6, 7
- [99] P. Mermelstein, "G.722, a new CCITT coding standard for digital transmission of wideband audio signals," *IEEE Commun. Mag.*, vol. 26, pp. 8–15, Feb. 1988. DOI: [10.1109/35.4177](https://doi.org/10.1109/35.4177)
- [100] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 744–754, Aug. 1986. DOI: [10.1109/TASSP.1986.1164910](https://doi.org/10.1109/TASSP.1986.1164910)
- [101] R. McAulay and T. Quatieri, "Sinusoidal coding," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier Science, 1995, ch. 4, pp. 121–173.
- [102] T. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*. Prentice Hall, 2001.

- [103] T. Painter and A. Spanias, "Perceptual segmentation and component selection for sinusoidal representations of audio," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 2, pp. 155–158, Apr. 2005. DOI: [10.1109/TSA.2004.841050](https://doi.org/10.1109/TSA.2004.841050) 54
- [104] LPC, *Telecommunications: Analog to Digital Conversion of Radio Voice by 2400 Bit/Second Linear Predictive Coding*, National Communication System - Office Technology and Standards Std. Federal Standard 1015, 1985. 7, 47
- [105] J. Campbell and T. E. Tremain, "Voiced/unvoiced classification of speech with applications to the U.S. government LPC-10e algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 11, Apr. 1986, pp. 473–476. DOI: [10.1109/ICASSP.1986.1169060](https://doi.org/10.1109/ICASSP.1986.1169060)
- [106] T. E. Tremain, "The government standard linear predictive coding algorithm: LPC-10," in *Speech Technology*, Apr. 1982, pp. 40–49. 7
- [107] CELP, *Telecommunications: Analog to Digital Conversion of Radio Voice By 4800 Bit/Second Code Excited Linear Prediction (CELP)*, National Communication System - Office Technology and Standards Std. Federal Standard 1016, 1991. 6, 7
- [108] J. Campbell, T. E. Tremain, and V. Welch, "The proposed federal standard 1016 4800 bps voice coder: CELP," in *Speech Technology*, vol. 5, Apr. 1990, pp. 58–64. 6, 7
- [109] GSM, *GSM Digital Cellular Communication Standards: Enhanced Full-Rate Transcoding*, ETSI/GSM Std. GSM 06.60, 1996. 6, 7
- [110] P. Kroon, E. Deprettere, and R. Sluyter, "Regular-pulse excitation—a novel approach to effective and efficient multipulse coding of speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 5, pp. 1054–1063, Oct. 1986. DOI: [10.1109/TASSP.1986.1164946](https://doi.org/10.1109/TASSP.1986.1164946) 6
- [111] GSM, *GSM Digital Cellular Communication Standards: Half Rate Speech; Half Rate Speech Transcoding*, ETSI/GSM Std. GSM 06.20, 1996. 7
- [112] I. Gerson and M. Jasiuk, "Techniques for improving the performance of CELP-type speech coders," *IEEE J. Select. Areas Commun.*, vol. 10, no. 5, pp. 858–865, June 1992. DOI: [10.1109/49.138990](https://doi.org/10.1109/49.138990)
- [113] ITU, *Pulse code modulation (PCM) of voice frequencies*, CCITT Std. Draft Recommendation G.711, 1988. 7
- [114] ITU, *32 kb/s adaptive differential pulse code modulation (ADPCM)*, CCITT Std. Draft Recommendation G.721, 1988. 7
- [115] ITU, *Wideband coding of speech at around 16 kb/s using Adaptive Multi-rate Wideband (AMR-WB)*, ITU Std. Draft Recommendation G.722.2, 2001. 6, 7, 8, 55

- [116] R. Ekudden, R. Hagen, I. Johansson, and J. Svedburg, "The adaptive multi-rate speech coder," in *Proc. IEEE Workshop on Speech Coding*, 1999, pp. 117–119. DOI: [10.1109/SCFT.1999.781503](https://doi.org/10.1109/SCFT.1999.781503)
- [117] ITU, *Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction (LD-CELP)*, ITU Std. Draft Recommendation G.728, 1992. 6, 7
- [118] J.-H. Chen *et al.*, "A low-delay CELP coder for the CCITT 16 kb/s speech coding standard," *IEEE J. Select. Areas Commun.*, vol. 10, no. 5, pp. 830–849, June 1992. DOI: [10.1109/49.1389887](https://doi.org/10.1109/49.1389887)
- [119] ITU, *High Level Description of TTS 4 kb/s Coder*, Texas Instruments Std. ITU-T Q21/SG16 Rapporteur meeting, AC-99-25, 1999. 7
- [120] ITU, *Conexant's ITU-T 4 kbit/s deliverables*, Conexant Systems Std. ITU-T Q21/SG16 Rapporteur meeting, AC-99-20, 1999.
- [121] A. McCree *et al.*, "A 4 kb/s hybrid MELP/CELP speech coding candidate for ITU standardization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 1, May 2002, pp. 629–632. DOI: [10.1109/ICASSP.2002.10058187](https://doi.org/10.1109/ICASSP.2002.10058187)
- [122] TIA, *The 8 kbit/s VSELP Algorithm*, TIA Std. TIA IS-54, 1989. 6, 7, 47
- [123] I. Gerson and M. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8 kbits/s," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 1, Apr. 1990, pp. 461–464. DOI: [10.1109/ICASSP.1990.1157497](https://doi.org/10.1109/ICASSP.1990.1157497)
- [124] Y. Gao *et al.*, "The SMV algorithm selected by TIA and 3GPP2 for CDMA applications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, 2001, pp. 709–712. DOI: [10.1109/ICASSP.2001.9410137](https://doi.org/10.1109/ICASSP.2001.9410137)
- [125] Y. Gao *et al.*, "EX-CELP: A speech coding paradigms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, 2001, pp. 689–692. DOI: [10.1109/ICASSP.2001.941008](https://doi.org/10.1109/ICASSP.2001.941008)
- [126] A. Spanias, "Speech coding standards," in *CRC Mobile Communications Handbook*, J. Gibson, Ed. CRC, 1999.
- [127] D. Knisely, S. Kumar, S. Laha, and S. Navda, "Evolution of wireless data services: IS-95 to CDMA2000," *IEEE Commun. Mag.*, vol. 36, 1998. DOI: [10.1109/35.7221506](https://doi.org/10.1109/35.7221506)
- [128] M. El-Sayed and J. Jaffe, "A view of telecommunications network evolution," *IEEE Commun. Mag.*, vol. 40, no. 12, pp. 74–81, Dec. 2002. DOI: [10.1109/MCOM.2002.11061636](https://doi.org/10.1109/MCOM.2002.11061636)
- [129] L.R. Rabiner and R.W. Schafer, *Digital processing of speech signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

- [130] L. Deng and D. O'Shaughnessy, *Speech Processing*. CRC Press, 2003.
- [131] O. Ghitza, "Auditory models and human performance in tasks related to speech coding and speech recognition," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 1, pp. 115–132, Jan. 1994. DOI: [10.1109/89.260357](https://doi.org/10.1109/89.260357) 53
- [132] B. Paillard *et al.*, "PERCEVAL: Perceptual evaluation of the quality of audio signals," *J. Acoust. Soc. Amer.*, vol. 40, no. 1, 1992. 53, 84, 88
- [133] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224–240, Apr. 1997. 53, 57, 84, 88
- [134] M. Smyth, "White paper: An overview of the coherent acoustics coding system," Digital Theater Systems (DTS), Tech. Rep., 1999.
- [135] D. Sen, D. Irving, and W. Holmes, "Use of an auditory model to improve speech coders," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, Apr. 1993, pp. 411–414. DOI: [10.1109/ICASSP.1993.319327](https://doi.org/10.1109/ICASSP.1993.319327) 53
- [136] J. Thyssen, W. B. Kleijn, and R. Hagen, "Using a perception-based frequency scale in waveform interpolation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, Apr. 1997, pp. 1595–1598. DOI: [10.1109/ICASSP.1997.596258](https://doi.org/10.1109/ICASSP.1997.596258)
- [137] G. Kubin and W. B. Kleijn, "On speech coding in a perceptual domain," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 1, Mar. 1997, pp. 205–208. DOI: [10.1109/ICASSP.1999.758098](https://doi.org/10.1109/ICASSP.1999.758098) 53
- [138] S. Voran, "Observations on auditory excitation and masking patterns," in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995. DOI: [10.1109/ASPAA.1995.482992](https://doi.org/10.1109/ASPAA.1995.482992) 53
- [139] H. Purnhagen, N. Meine, and B. Edler, "Sinusoidal coding using loudness-based component selection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, 2002, pp. 1817–1820. DOI: [10.1109/ICASSP.2002.1006118](https://doi.org/10.1109/ICASSP.2002.1006118) 53, 54
- [140] M. Bosi and R. Goldberg, *Introduction to Digital Audio Coding and Standards*. Springer, 2002. 53
- [141] A. Spanias, T. Painter, and V. Atti, *Audio Signal Processing and Coding*. New York: Wiley-Interscience, 2006. 5, 50, 53, 54
- [142] J. Beerends, A. Hekstra, A. W. Rix, and M. Hollier, "Perceptual evaluation of speech quality (PESQ) the new ITU standard for end-to-end speech quality assessment part II: Psychoacoustic model," *J. Audio Eng. Soc.*, vol. 50, no. 10, pp. 765–778, Oct. 2002. 54

- [143] Thiede *et al.*, “PEAQ – The ITU standard for objective measurement of perceived audio quality,” *J. Audio Eng. Soc.*, vol. 48, no. 1, pp. 3–29, Jan. 2000. 54
- [144] M. Goodwin, *Adaptive Signal Models: Theory, Algorithms, and Audio Applications*. Boston: Kluwer Academic Publishers, 1998. 54
- [145] R. Heusdens, R. Vafin, and W. B. Kleijn, “Sinusoidal modeling using psychoacoustic adaptive matching pursuits,” *IEEE Signal Processing Lett.*, vol. 9, no. 8, pp. 262–265, Aug. 2002. DOI: [10.1109/LSP.2002.802999](https://doi.org/10.1109/LSP.2002.802999) 54
- [146] Y. Hu and P. C. Loizou, “A perceptually motivated approach for speech enhancement,” *IEEE Trans. Speech Audio Processing*, vol. 11, no. 5, pp. 457–465, 2003. DOI: [10.1109/TSA.2003.815936](https://doi.org/10.1109/TSA.2003.815936) 54
- [147] F. Baumgarte, “Improved audio coding using a psychoacoustic model based on a cochlear filter bank,” *IEEE Trans. Speech Audio Processing*, vol. 10, no. 7, pp. 495–503, Oct. 2002. DOI: [10.1109/TSA.2002.804536](https://doi.org/10.1109/TSA.2002.804536) 54, 84
- [148] S. van de Par, A. Kohlrausch, G. Charestan, and R. Heusdens, “A new psychoacoustical masking model for audio coding applications,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, vol. 2, 2002, pp. 1805–1808. DOI: [10.1109/ICASSP2002.1006115](https://doi.org/10.1109/ICASSP2002.1006115) 54, 84
- [149] D. Pickard, Wikipedia: The free encyclopedia, [Online]. Available: <http://en.wikipedia.org/wiki/Image:HumanEar.jpg>, Tech. Rep., 2006. 56
- [150] J. Deller, J. Proakis, and J. Hansen, *Discrete-Time Processing of Speech Signals*. New York: Macmillan, 1993. 58
- [151] S. Stevens and J. Volkman, “The relation of pitch to frequency,” in *American Journal of Psychology*, vol. 53, no. 3, 1940, p. 329. 58
- [152] J. Johnston, “Transform coding of audio signals using perceptual noise criteria,” *IEEE J. Select. Areas Commun.*, vol. 6, no. 2, pp. 314–323, Feb. 1988. DOI: [10.1109/49.608](https://doi.org/10.1109/49.608) 79, 82
- [153] J. Johnston *et al.*, “AT&T perceptual audio coding (PAC),” in *Collected Papers on Digital Audio Bit-Rate Reduction*, N. Gilchrist and C. Grewin, Eds. Audio Engineering Society, 1996, pp. 73–81. 79
- [154] F. Baumgarte, “A nonlinear psychoacoustic model applied to the ISO MPEG layer III codec,” in *Proc. 99th AES Symposium*, 1995. 84
- [155] R. A. Lutfi, “Additivity of simultaneous masking,” *J. Acoust. Soc. Amer.*, vol. 73, no. 1, pp. 262–267, Jan. 1983. DOI: [10.1121/1.388859](https://doi.org/10.1121/1.388859) 84

- [156] E. Ordentlich and Y. Shoham, "Low-delay code excited linear predictive coding of wideband speech at 32 kb/s," in *Proc. IEEE ICASSP-93*, vol. 2, pp. 9–12, 1993. DOI: [10.1109/ICASSP1991.150266](https://doi.org/10.1109/ICASSP1991.150266) 6
- [157] C. Laflamme, R. Salami, and J.-P. Adoul, "9.6 kbit/s ACELP coding of wideband speech," in *Speech and audio coding for wireless and network applications*, Eds., B. Atal, V. Cuperman, and A. Gersho, Kluwer Academic Publishers, 1993.
- [158] J.-P. Adoul and R. Lefebvre, "Wideband speech coding," in *Speech analysis and synthesis*, Eds., W. B. Kleijn and K. K. Paliwal, Kluwer Academic Publishers, 1995.
- [159] N. Jayant, J. Johnston, and Y. Shoham, "Coding of wideband speech," *Speech Commun.*, vol. 11, no. 2–3, pp. 127–138, 1992. DOI: [10.1016/0167-6393\(92\)90007-T](https://doi.org/10.1016/0167-6393(92)90007-T) 6
- [160] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 14496–3 AMD-1: "Coding of Audio-Visual Objects – Part 3: Audio," 2000. ("MPEG-4 version-2"). 6
- [161] N.S. Jayant, V. Lawrence, and D. Prezas, "Coding of Speech and Wideband Audio," *AT&T Tech. J.*, Vol. 69(5), pp. 25–41, Sept.–Oct. 1990. 6
- [162] GSM 06.10, "GSM Full-Rate Transcoding," *Technical Report Version 3.2*, ETSI/GSM, July 1989. 49
- [163] I. Boyd and C. Southcott, "A Speech Codec for the Skyphone Service," *Br. Telecom Technical J.*, vol. 6(2), pp. 51–55, April 1988. DOI: [10.1007/s10550-007-0070-0](https://doi.org/10.1007/s10550-007-0070-0) 49
- [164] Itu, *G.729-based embedded variable bit-rate coder: An 8-32 kbits/s scalable wideband coder bit-stream interoperable with G.729*, ITU Std. Draft Recommendation, G.729.1, 2006. 6, 7, 8