Advances in Speech and Audio Processing and Coding

Andreas Spanias
SenSIP Center, School of ECEE
Arizona State University
Tempe, AZ 85287. USA
spanias@asu.edu

***Abstract*—This plenary session will cover speech processing research advances with the emphasis on speech and audio coding methods. In the session, we will discuss the fundamental principles, techniques, and algorithms used in current coding applications including a summary of codecs for telecommunication standards. The session will start with a discussion on: the basic speech representation methods, the performance measures used to evaluate coded speech, and the role of the standards. Brief algorithm descriptions include: ADPCM, sub-band coding, adaptive transform coding, sinusoidal transform coding (STC), linear predictive coding (LPC), and analysis-by-synthesis LPC (sparse excitation, code excited LPC, and ACELP). The presentation will feature audio, and computer demonstrations of recent speech coding standards including voice-over IP algorithms. The plenary session will also cover wideband audio standards such as MPEG audio and other layers (e.g., MP3, AAC). Recent algorithms will also be described including the following: Variable-Rate Multimode Wideband (VMR-WB), Speex, G722.1, OGG Vorbis 2012, iLBC, SELT, SILK, Opus 2013, Qualcomm wideband 5G codecs. At the end of the session, we will cover briefly recent applications that use voice features for detecting speech pathologies, and also discuss how long-term speech parameters can be used as predictors of other diseases such as tremors, Alzheimer's etc.**

Keywords—Speech and Audio Coding, Standardized Codecs

BRIEF SURVEY

In the following, we provide a brief survey with citations relevant to the plenary session. Basic speech analysis-synthesis principles are given in [1,2]. The basics of linear prediction are covered in [3,4,5,6]. Various first generation algorithms including the LPC-10 standard [6] are described in [7] and the references therein. The analysis-by-synthesis CELP algorithm was first introduced in [8]. The Federal standard 1016 CELP at 4.8 Kbits/s is described in [9] and MATLAB simulations with this algorithm are presented in [10]. A more recent 2.4kbits/s FS1016 is presented in [16]. Basics of vector quantization are described in [11]. Line spectral frequencies and their application in speech coding are introduced in [12,13]. Algebraic codebooks for CELP (ACELP) which use sparse excitation sequences were introduced in [14], and pitch interpolation methods for LPC are described in [15]. First generation cell phone standards are described in [7] and references therein. Second generation and third generation speech coding standards are summarized in [17-23]. Pitch estimation is described in [4,7,24,64] and sinusoidal representations of speech are described in [25,26,27]. Educational software and simulations for speech and audio coders are presented in [28]. Perceptual determination of LPC poles is given in [29]. Bandwidth extension for low bit rate speech codecs are described in [30,31] and references therein. Recent standards including open source algorithms are described in [32-36]. Audio coding for high fidelity applications is covered in [37-39] and references therein. Principles of psychoacoustics and critical bands are described in [40-43]. Perceptual entropy theory is covered in [44]. Various audio coding standards are described in [45-56]. New applications of speech processing which provide diagnostics that can be used for detecting pathologies are presented in [57,58]. Audio analysis, synthesis and content search methods including tools for speech and audio processing are provided in [59-65].

BIOGRAPHY

Andreas Spanias is Professor in the School of Electrical, Computer, and Energy Engineering at Arizona State University (ASU). He is also the director of the Sensor Signal and Information Processing (SenSIP) center and the founder of the SenSIP industry consortium (an NSF I/UCRC site). His research interests are in the areas of adaptive signal processing, speech processing, and sensor systems. He and his student team developed the computer simulation software Java-DSP and its award winning iPhone/iPad and Android versions. He is author of two textbooks: Audio Processing and Coding by Wiley and DSP; An Interactive Approach (2nd Ed.). He served as Associate Editor of the IEEE Transactions on Signal Processing and as General Co-chair of IEEE ICASSP-99. He also served as the IEEE Signal Processing Vice-President for Conferences. Andreas Spanias is a co-recipient of the 2002 IEEE Donald G. Fink paper prize award and was elected Fellow of the IEEE in 2003. He served as Distinguished lecturer for the IEEE SPS in 2004. He is editor of the Morgan and Claypool algorithms and software series.

ACKNOWLEDGMENT

We acknowledge several sponsors including NSF, the SenSIP Center and NCSS I/UCRC site, Intel, Freescale, Raytheon, and Sprint for research awards and contracts that supported studies in signal and speech/audio processing.

REFERENCES

[1] J. Flanagan, Speech Analysis, Synthesis and Perception, Springer-Verlag, 1972.

[2] G. Fant, Acoustic Theory of Speech Production, Mounton and Co., Gravenhage, The Netherlands, 1960.

[3] Jayant, N. S., and Noll, P. Digital coding of waveforms, Prentice-Hall, Englewood Cliffs, NJ,1984.

[4] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, No. 4, pp. 561-580, April 1975.

[5] P.P. Vaidyanathan, Theory of Linear Prediction, Morgan & Claypool, Ed. J. Mura, 2008.

[6] T.E. Tremain, "The Government Standard Linear Predictive Coding Algorithm: LPC-10," *Speech Technology*, pp. 40-49, Apr. 1982.

[7] A. Spanias, "Speech Coding: A Tutorial Review," *Proc. IEEE*, Vol. 82, No. 10, pp. 1441-1582, October 1994.

[8] M.R. Schroeder and B. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates," *Proc. ICASSP-85*, pp. 937-940, Tampa, Apr. 1985.

[9] J. Campbell, T.E. Tremain, V. Welch, "The Proposed Federal Standard 1016 4800 bps Voice Coder: CELP," *Speech Tech*., pp.58-64, Apr. 1990.

[10]. Ramamurthy and A. Spanias, MATLAB® Software for the Code Excited Linear Prediction Algorithm: The Federal Standard-1016, Morgan and Claypool Publishers, vol. 2, no. 1, ISBN 1608453847, Jan 2010.

[11] A. Gersho and R.M. Gray, Vector Quantization and Signal Compression, Kluwer Academic Publishers, 1992.

[12] P. Kabal and R. Ramachandran, "The Computation of Line Spectral Frequencies Using Chebyshev Polynomials," *IEEE Trans. ASSP-34,* pp. 1419-1426, Dec. 1986.

[13] G. Kang and L. Fransen, "Application of Line-Spectrum Pairs to Low-Bit-Rate Speech Encoders," Proc. IEEE ICASSP, , pp. 244-247, 1985.

[14] J. P. Adoul et al., "Fast CELP coding based on algebraic codes," Proc. IEEE ICASSP-87, vol. 12, pp. 1957-1960, Apr. 1987.

[15] W. Kleijn et al., "Generalized Analysis-by-Synthesis Coding and its Application to Pitch Prediction," *Proc. ICASSP-92,* p. 337, Mar. 1992.

[16] L.M. Supplee, R.P. Cohn, J.S. Collura, A.V. McCree , "MELP: The New Federal Standard 1016 at 2400bps, *Proc. ICASSP-97*, Munich.

[17] A. Spanias, Digital Signal Processing; An Interactive Approach – 2nd Edition, ISBN 978-1-4675-9892-7, Lulu Press, NC, May 2014..

[18] A. Spanias, Vocoders, Encyclopedia of Telecommunications, Ed. J. Proakis, Wiley Interscience, John Wiley & Sons, Inc., 2003.

[19] W. Chu, Speech Coding Algorithms, Wiley Interscience, 2003.

[20] TIA/EIA/IS-96, "QCELP, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems," TIA 1992.

[21] R. Ekudden, R. Hagen, I. Johansson, J. Svedburg, "The Adaptive Multi-Rate Speech Coder," IEEE Workshop Speech Coding, pp. 117-119, 1999.

[22] Y. Gao, E. Schlomot, A. Benyassine, J. Thyssen, H. Su, and C. Murgia, "The SMV algorithm selected for TIA and 3GPP2 for CDMA applications," Proc. IEEE ICASSP-01, Vol. 2, Salt Lake City, May 2001

[23] TIA/EIA/IS-127, "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems," TIA 1997.

[24] S. Ahmadi+ and A.S. Spanias, "Cepstrum-Based Pitch Detection Using a New Statisctical V/UV Classification Algorithm**,"** *IEEE Trans. on Speech and Audio*, Vol. 7, No. 3, pp. 333-338, May 1999.

[25] H. Krishnamoorthi, A. Spanias, "Sinusoidal component selection based on partial loudness," *IEEE ICASSP 2013*, pp.575-579, Vancouver, 2013.

[26] R. McAulay and T. Quatieri, "Speech analysis synthesis based on a sinusoidal representation," *IEEE Trans. ASSP*-34,pp.744-754, Aug. 1986.

[27] M. Brandstein, P. Monta, J. Hardwick, J. Lim, "A Real-Time Implementation of the Improved MBE Coder," *Proc. ICASSP-90*, pp. 5-8, New Mexico, April 1990.

[28] A. Spanias and V. Atti+, "Interactive On-line Undergraduate Laboratories Using J-DSP," *IEEE Trans. on Education Special Issue on Web-based Instruction*, pp. 735-749, Nov. 2005.

[29] V. Atti, Algorithms and Software for Predictive and Perceptual Modeling of Speech, Morgan & Claypool Publishers, Editor A. Spanias, 2011.

[30] V. Berisha and A. Spanias, "Wideband Speech Recovery Using Psychoacoustic Criteria," *EURASIP J. on Audio, Speech, and Music Pro*, Vol. 2007, ID 16816, 18 pages, doi:10.1155/2007/16816, August 2007.

[31] V. Berisha, and A. Spanias, Split-Band Speech Compression Based On Loudness Estimation, ASU, Tempe, AZ, US 8,392,198 B1, Issued 2012.

[32] S. Andersen, A. Duric, H. Astrom, R. Hagen, W. Kleijn, J. Linden, Internet Low Bit Rate Codec (iLBC), *IETF RFC3951*, 2004.

[33] J. Valin, G. Maxwell, T.Terriberry1, K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec," *135th AES Conv.,* NY, Oct. 2013.

[34] Vorbis I specification, Xiph.Org Foundation Feb. 3, 2012.

[35] V. Atti et al. "Improved error resilience for volte and voip with 3gpp evs channel aware coding," *Proc. ICASSP-2015*, Brisbane, April 2015.

[36] V. Malenovsky, T. Vaillancourt; W. Zhe,K. Choo, V. Atti. "two-stage speech/music classifier with decision smoothing and sharpening in the evs codec," *Proc. ICASSP-2015*, Brisbane, April 2015.

[37] P. Noll, "Wideband Speech and Audio Coding," *IEEE Comm. Mag.,* pp.34-44, Nov. 1993.

[38] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," *Proc. of the IEEE,* vol. 88, pp. 451-513, Apr. 2000.

[39] A. Spanias, T. Painter, V. Atti, Audio Signal Processing and Coding, Wiley, 2007.

[40] H. Fletcher, "Auditory Patterns," Rev. Mod. Phys., pp. 47-65, Jan. 1940.

[41] B.C.J. Moore, Introduction to the Psychology of Hearing, University Park Press, 1977.

[42] B. Scharf, "Critical Bands," Foundations of Modern Auditory Theory, New York: Academic Press, 1970.

[43]E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models, Springer-Verlag, 1990.

[44] J. Johnston, "Estimation of Perceptual Entropy Using Noise Masking Criteria," in Proc. ICASSP-88, pp. 2524-2527, May 1988.

[45] M. Bosi, et al., "ISO/IEC MPEG-2 Advanced Audio Coding," J. Audio Eng. Soc., pp. 789-813, Oct. 1997.

[46] M. Bosi and R. Goldberg, Introduction to Digital Audio Coding and Standards, Kluwer Academic Publishers, 2002.

[47] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio," *J. AES,* p. 780, Oct. 1994.

[48] ISO/IEC JTC1/SC29/WG11 MPEG, IS11172-3 "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mb/s, Part 3: Audio," 1992. ("MPEG-1").

[49] ISO/IEC JTC1/SC29/WG11 MPEG, IS13818-3 "Information Technology - Generic Coding of Moving Pictures and Associated Audio, Part 3: Audio," 1994. ("MPEG-2 BC-LSF").

[50] ISO/IEC 13818-7, "Information Technology – Generic Coding of Moving Pictures and Associated Audio, Part 7: Advanced Audio Coding," 1997.

[51] K. Konstantinides, "An Introduction to Super Audio CD and DVD-Audio," IEEE Signal Proc. Mag., vol. 20, no. 4, pp. 71–82, 2003.

[52] J. J. Thiagarajan, A. Spanias, Analysis of the MPEG-1 Layer III (MP3) Algorithm Using MATLAB, Morgan and Claypool Publishers, vol. 3, no. 3, ISBN-10: 1608458016, ISBN-13: 978-1608458011, Nov. 2011

[53] Karlheinz Brandenburg, "MP3 and aAAC explained," AES 17 International Conference on High Quality Audio Coding.

[54] Lossless and Lossy Windows Media Algorithms, Micrpsoft Research, http://en.wikipedia.org/wiki/Windows_Media_Audio "ISO/IEC 13818-7, Fourth edition, Part 7 - Advanced Audio Coding (AAC)"

[55] M. Bosi, K. Brandenburg, Sch. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Yoshiaki Oikawa. ISO/IEC MPEG-2 Advanced Audio Coding. In Proc 101st AES-Convention, 1996.

[56] Apple Lossless Audio Codec (ALAC), http://en. wikipedia.org/wiki/ Apple_Lossless.

[57] V. Berisha, S. Sandoval, R. Utianski,, J. Liss, A. Spanias,, Selecting disorder-specific features for speech pathology fingerprinting,"Proc. *IEEE ICASSP 2013*, 7562 - 7566, Vancouver, May 2013.

[58] An interview with Visar Berisha and Julie Liss, Parsing Ronald Reagan's Words for Early Signs of Alzheimer's, www.nytimes.com/.../parsing-ronald-reagans-words-for-early-signs-of- alzheimers.html.

[59] I. Rodomagoulakis, G. Potamianos, P. Maragos, "Advances in Large Vocabulary Continuous Speech Recognition in Greek: Modeling and nonlinear features," SP Conference (EUSIPCO), Sep. 2013.

[60] G Wichern, J. Xue, H Thornburg, B Mechtley, A Spanias, "Segmentation, indexing, and retrieval for environmental and natural sounds*," IEEE Trans. SLP,* 18 (3), pp. 688-707, 2010.

[61] T Giannakopoulos, A Pikrakis, S Theodoridis, "A multi-class audio classification method with respect to violent content in movies using bayesian networks," *IEEE 9th Workshop on MMSP 2007*, 90-93, 2007.

[62] T Giannakopoulos, A Pikrakis, "Introduction to Audio Analysis: A MATLAB® Approach," Academic Press, 2014.

[63] A. Fink, A. Spanias, P. Cook, "Derivation of a new banded waveguide model topology for sound synthesis," J. Acoust. Soc. Am., Volume 133, Issue 2, pp. EL76-EL81 (2013).

[64] L. Rabiner and R. Schafer, Theory and Applications of Digital Speech Processing, Pearson, 2011.

[65] Julius O. Smith III, Spectral Audio Signal Processing, W3K Publishing, 2011.