MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 7, No. 2, Feburary 2012

IEEE COMMUNICATIONS SOCIETY

CONTENTS

MESSAGE FROM MMTC CHAIR

Dear MMTC fellow members,

At the beginning of year 2012, as always I wish all of you a healthy, prosperous, and happy new year! The past year 2011 has witnessed many inspiring new activities in MMTC, such as the first TC annual workshop MMCOM'11, IG keynote talks, as well as many journal Special Issues and so on, I wish 2012 become another successful year for MMTC community with strong growth and many achievements in both research efforts and professional activities.

First of all, I would like to thank the following leaders who setup a record in history for TC's first annual workshop, International Workshop on Multimedia Communications (MMCom), which was held in Dec. 2011 at Houston, Texas with a big success:

- Prof. Dr. Thomas Magedanz, Fraunhofer FOKUS/ T. U. Berlin, Germany
- Prof. Jiangtao (Gene) Wen, Tsinghua University, China
- Dr. Xiaoli Chu, King's College, UK
- Prof. Yung-Hisang Lu, Purdue University, USA

I am glad to report that the 2nd version of MMCom is currently in preparation by the following members with the main theme of QoE Multimedia Communication:

- Dr. Periklis Chatzimisios, Alexander TEI of Thessaloniki, Greece
- Prof. Tasos Dagiuklas, TEI of Mesolonghi, Greece
- Prof. Chang Wen Chen, University at Buffalo, State University of New York, USA
- Prof. Luigi Atzori, University of Cagliari, Italy

Lets support this event by submitting papers and volunteering to join the TPC team. The MMCom'12 will be held in conjunction with GLOBECOM'12 at Anaheim, California (pending approval).

In Dec. 2011, we had the election for Steering Committee Voting Member of IEEE ICME representing MMTC (term 2012-2013), 3 members were nominated to compete for this position. I am in the position to report to you the final outcome of the votes, Dr. Jin Li (Microsoft) has been elected to represent MMTC in IEEE ICME Steering Committee. Lets congratulate Dr. Li and thanks him for taking the leadership to promote our TC's involvement in future ICME co nferences.

In addition, the TC's Award Board recently reviewed the papers recommended in R-Letter during the past 2 years and selects the following 2 papers for the Year 2011 Best Paper Award winners:

- W.-H. Kuo, W. Liao, and T. Liu, "Adaptive Resource Allocation for Layer-Encoded IPTV Multicasting in IEEE 802.16 WiMAX Wireless Networks," *IEEE Transactions on Multimedia*, vol. 13, no. 1, pages 116—124, Feb. 2011.
- Z. Liu, Y. Shen, K. W. Ross, S. Panwar, and Y. Wang, "LayerP2P: Using Layered Video Chunks in P2P Live Streaming," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pages 1340—1352, November 2009.

Lets congratulate the winning authors for their high-quality work. I would like to encourage again our members to recommend your own (or others) good work to our R-Letter editors, once selected, this paper will get the chance to compete the Best Paper Award for future years. It is always respectful efforts to promote good work inside our community to our members.

In a couple of months, we will start the process of electing MMTC's new leadership team (term 2012-2014). I hope all TC members can consider this election from now on and nominate the best team to lead this TC for the coming future.

It is important to realize that E-Letter has been in a good shape during the past 3 years, I would like to thank the current E-Letter Co-Directors, Dr. Chonggang Wang and Kai Yang, and the editors in the Editorial Board, for their hard work and

consistent efforts to make this letter very successful and maintain strong impact in the community. We hope to see a further growth for this E-Letter, which requires much more resources and contribution from our members, as a strategic move to strengthen the leadership team, I propose to have a dedicated TC Vice Chair (other than the regional Vice-Chairs) in future TC leadership team who can concentrate to promote our Letters and strengthen the link between our IG team and the Letters. I hope our fellow members can give this proposal a thought and we can discuss and make final call during our next TC meeting at ICC'13 (Ottawa, Canada).

At last but not the least, I would like to encourage our members to submit papers to IEEE ICME 2012 workshops at Melbourne, Australia (http://www.icme2012.org/). The paper submission deadline is **March 5, 2012**.

Thank you very much!

Haohong Wang
Chair of Multimedia Communication TC of IEEE ComSoc

SPECIAL ISSUE ON AUDIO AND ACOUSTIC SIGNAL PROCESSING

Audio and Acoustic Signal Processing

Zhiqiang Wu, Wright State University, USA
Pei-Jung Chung, University of Edinburgh, UK
Zhiqiang.Wu@wright.edu, pchung@ed.ac.uk

While many new applications emerge every day, audio is still the essential application in communication networks. Although audio and acoustic signal processing has gone a long way, there are exciting new advances we'd like to share with our readers. Researchers around the world have been working on providing us better audio experience and better acoustic communication systems: new speech and audio codecs are being designed, new standards on high quality digital audio are being developed, new techniques are being adopted from other areas into audio and acoustic signal processing arena, to name a few. The objective of the special issue of the E-Letter on audio and acoustic signal processing is to present some of the recent progresses in this area.

In the first paper, Synthesis Filter/Decoder Structures in Speech Codecs, the author discusses a new type of decoder/synthesizer structures for common linear predictive speech models. Using the Shannon backward channel result from rate distortion theory, it is shown that the new decoder/synthesizer structures require zeros, in contrast to the usual all pole decoders/synthesizers used in popular CELP codecs. The author calculates the transfer functions for these new structures and discusses the effect on the reconstructed signal. The parameters in the perceptual weighting filter and the average distortion both change the shaping of the excitation.

In the second paper, A high Fidelity Channel Simulator for Underwater Acoustic Communication, the authors proposes a novel high-fidelity channel simulator for high frequency acoustic communication. In the high frequency Acoustic communication, various physical processes, including surface waves, subsurface bubbles and ocean volume fluctuations, can significantly affect the communication channel. Currently the research community is still lacking a realistic channel representation. This paper develops a high-fidelity channel simulator through parabolic equation modeling of acoustic propagation and scattering. After calibration by experimental data, it is shown that the channel simulator is capable of generating realistic time varying impulse responses and the output agrees well with measurements.

In the third paper, The IETF Opus Speech and Audio Codec – Yet Another Codec?, the author discusses a new codec called "Opus" developed by the Internet Engineering Task Force (IETF), which is ready to be standardized. The author first discusses available codecs and argues the necessity and need for the development of Opus. Requirements, codec design, and unique features of Opus are discussed next. Opus is available as open source and probably will be available royalty free, making it very attractive. It is expected that Opus is likely to dominate the sector of interactive audio and speech coding, if RTCweb is successful and opus remains free of royalty payments.

In the fourth paper, Introduction to AVS Audio Lossless Coding/Decoding Standard, the authors provide us an overview of the new high quality lossless audio coding standard of the Audio Video Coding Standard Workgroup of China. Authors first present the history of this ongoing effort, then discuss the framework and the technical details of the lossless audio coder and decoder. Some experimental results have shown that the compress rate of this standard is comparable and better than several popular coders.

In the fifth paper, Blind Cyclostationary Carrier Frequency and Symbol Rate Estimation for Underwater Acoustic Communication, the authors present their effort to apply cyclostationary analysis on underwater acoustic signals for blind parameter estimation. While cyclostationary analysis has been accepted as an important and useful tool in RF signal processing, it is not clear how it will perform in underwater acoustic signal processing. Due to the extremely complex and dynamic environment of underwater acoustic communication, authors develop a short term dynamic resolution algorithm for cyclostationary analysis. They also show that the proposed method is capable of blindly estimate the carrier frequency (and Doppler shift) and symbol rate of underwater

acoustic signals accurately. The effectiveness of the proposed method is validated with real data collected at sea.

Dr. Zhiqiang Wu received his BS from Beijing University of Posts and Telecommunications in 1993, MS from Peking University in 1996, and PhD from Colorado State University in 2002, all in electrical engineering. He has worked at West Virginia University Institute of Technology as assistant professor from 2003 to 2005. He joined Wright State University in 2005 and currently serves as associate professor. He co-authored one of the first books on multi-carrier transmission for wireless communication. He has published more than 70 papers in journals and conferences. His work on software defined radio implementation of cognitive radio received the Best Demo Award at IEEE Globecom 2010.

Dr. Pei-Jung Chung received Dr.-Ing. Degree in 2002 from Ruhr-University Bochum, Germany with distinction. From 1998 to 2002 she was with Signal Theory Group, Ruhr-University Bochum, Germany. From 2002 to 2004 she held a post-doctoral position at Carnegie Mellon University and University of Michigan, Ann Arbor, USA, respectively. From 2004 to 2006 she was Assistant Professor with National Chiao Tung University, Hsin Chu, Taiwan. In 2006 she joined the Institute for Digital Communications, School of Engineering, the University of Edinburgh, UK as Lecturer. Her research interests include array processing, statistical signal processing, and wireless MIMO communications and distributed processing in wireless sensor networks.

Synthesis Filter/Decoder Structures in Speech Codecs

Jerry D. Gibson, Electrical & Computer Engineering, UC Santa Barbara, CA, USA
gibson@ece.ucsb.edu

Abstract
Using the Shannon backward channel result from rate distortion theory, we derive new decoder/synthesizer structures for common linear predictive speech models. We calculate the transfer functions for these new structures for common source models and discuss the effect on the reconstructed signal. The perceptual weighting filter and the average distortion both play a prominent role. We show that common CELP encoders and decoders lack these components, and note that at least at the decoder, postfiltering as currently used in CELP may partially compensate for the missing structure.

1. Introduction

[1]Codebook excited linear predictive coding (CELP) is the underlying principle used in all narrowband and some wideband standardized speech codecs today [1]. Most recent efforts toward the development of fullband codecs (20 Hz to 20 kHz) utilize a combination of the CELP approach and the transform/filter bank approaches with well-designed switching between the coding methods, as is evident in the recently standardized USAC codec [2] and as is expected in the EVS codec for LTE Mobile Systems [3]. While linear prediction had long had success for speech waveform coding and was later used for low bit rate voice codecs by modeling the vocal tract [4], codebook excited analysis-by-synthesis coding using linear prediction was first motivated by rate distortion theory principles [5-10].

In this paper, we return to rate distortion theory fundamentals to examine the structure of the decoder and the synthesizer used in the encoder for optimal speech coding subject to the squared error fidelity criterion. It is shown that the usual CELP decoder should have additional excitation filtering, currently not present in CELP codecs, that is dependent on the perceptual weighting filter and on the average distortion.

The paper is organized as follows. Section II contains a quick review of the linear prediction model and CELP codecs. Section III then uses the classical Shannon lower bound and Shannon backward channel concept to derive the form of the optimal code generator or decoder for sources satisfying the linear prediction model with perceptual weighting of the distortion. Section IV presents examples of the new structures for three simple representative source models subject to weighted and unweighted squared error fidelity criteria. Section V presents comparisons to current CELP codec structures, and Section VII discusses possible implementation approaches. Section VII contains some conclusions.

2. Linear predictive Voice Codecs

Linear predictive coding (LPC) is the dominant paradigm for narrowband speech coding in the last 40 years. In LPC, the decoder or synthesizer has the form shown in Fig. 1, wherein

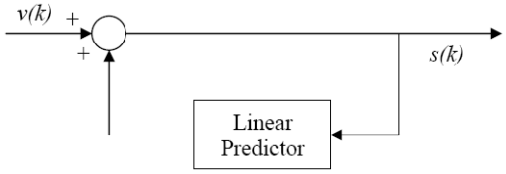


Figure 1. The Linear Prediction Model

The output speech is reconstructed according to

$$s(k) = -\sum_{i=1}^{m} a_k s(k-i) + v(k)$$

This decoder structure has been carried over to the code-excited linear predictive (CELP) analysis-by-synthesis codecs with encoders of the form shown in Fig. 2 and decoders as shown in Fig. 3. In these figures, the Synthesis filter is the linear predictor given in Fig. 1.

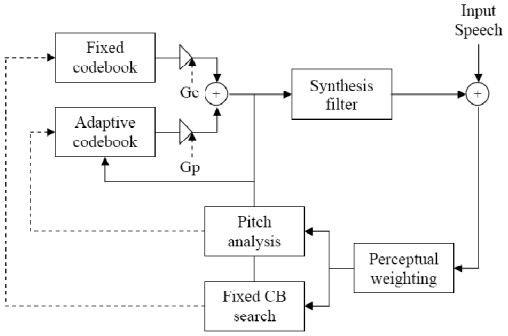


Figure 2. CELP Encoder

[1] This research has been funded by NSF Grant Nos. CCF-0728646 and CCF-0917230

Having a synthesis filter that mimics the linear prediction model appears to be intuitive and well-motivated. However, it is well known that other modifications such as postfiltering following the decoder may improve the reconstructed speech in some instances. In this paper we explore alternative decoder structures implied by rate distortion optimal lossy source coding theory.

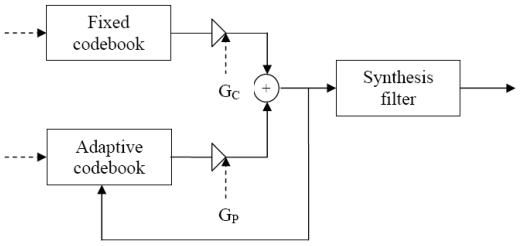


Figure 3. CELP Decoder

3. Rate Distortion Analysis for Difference Distortion Measures

To introduce the basic underlying principle from rate distortion theory, we begin by considering the problem of encoding a memoryless Gaussian source *X* subject to the mean squared error (MSE) fidelity criterion. A classic lower bound on the rate distortion function for difference distortion measures, first introduced in Shannon's original paper on rate distortion theory in 1959, is called the Shannon lower bound [11].

When this bound is satisfied with equality for an average distortion D_1, the optimally encoded output, $\hat{X}_1$, satisfies the Shannon backward channel condition expressed by [5]

$$X = \hat{X} + Z \tag{1}$$

where if *X* is zero mean, Gaussian with variance σ^2, then $\hat{X}$ is zero mean, Gaussian with variance $\mathrm{var}(\hat{X}) = \sigma^2 - D$, and Z a zero mean Gaussian random variable that is statistically independent of $\hat{X}$ with variance D. Memoryless sources do not give us much insight into the coding of actual speech signals so we turn our attention now to sources that satisfy the linear prediction model. However, after suitable transformations, the Shannon backward channel condition can still be imposed to provide the essential results.

An *m*th-order, time-discrete AR source can be expressed as

$$X_t = -\sum_{k=1}^{m} a_k X_{t-k} + Z_t \tag{2}$$

where $a_1, \ldots, a_m$ are the AR coefficients, and $\{Z_t\}$ is a sequence of independent and identically distributed (iid) random variables, and X_r and Z_s are statistically independent if *s*>*r*. The Shannon backward channel formulation has been used by Berger to analyze optimal tree encoding of Gaussian autoregressive (AR) sources subject to the MSE distortion measure. In his analysis, not repeated here, Berger shows that the power spectral density (psd) of the optimal reconstructed value is of the form [5]

$$\Phi_Y(z) = \Phi_X(z) - D \tag{3}$$

for average distortion *D*. For an AR source, the source psd is

$$\Phi_X(z) = \frac{\sigma^2}{|A(z)|^2} \tag{4}$$

which upon substitution into the above yields

$$\Phi_Y(z) = \frac{\sigma^2 - |A(z)|^2 D}{|A(z)|^2} = \frac{|B(z)|^2}{|A(z)|^2} \tag{5}$$

This perhaps surprising result shows that the reconstructed output is no longer purely AR, but it is now an autoregressive moving average (ARMA) sequence, where the MA part is dependent on the average distortion and on the linear prediction coefficients through the numerator polynomial *B(z)*.

The analysis can be extended to the optimal encoding of this AR source subject to a weighted squared error distortion measure to obtain

$$\Phi_Y(z) = \Phi_X(z) - \frac{D}{|W(z)|^2} \tag{6}$$

where $W(z)$ is the frequency weighting of the reconstruction error (the distortion). Substituting as before for the psd of the input, we obtain the expression [12]

$$\Phi_Y(z) = \frac{\sigma^2 - D\dfrac{|A(z)|^2}{|W(z)|^2}}{|A(z)|^2} = \frac{|B(z)|^2}{|A(z)|^2} \tag{7}$$

The results in Eqs. (5) and (7) are subject to the small distortion condition, namely,

$$\frac{D}{|W(z)|^2} \le \frac{\sigma^2}{|A(z)|^2} \tag{8}$$

Which also guarantees that the *B(z)* polynomial in the numerator of the needed form exists. The numerator polynomial in the expression for the psd of the reconstructed source now depends upon the perceptual weighting function as well as the average distortion and linear prediction coefficients.

These results imply that the synthesizer in the encoder and the decoder in CELP codecs should not simply be the linear prediction model with appropriate excitation. In the following sections, we investigate the impact of the change in the numerator on the reconstructed spectrum.

4. The Excitation Shaping Filter

We denote the numerator polynomial *B(z)* as the excitation shaping filter and we illustrate the form of this filter with and without perceptual weighting for three different source spectra, a 3rd order Butterworth shaping, a 3rd order AR shaping based on the coefficients in [6,13], and a 10th order AR model. The power spectral densities of these sources are shown in Fig. 4.

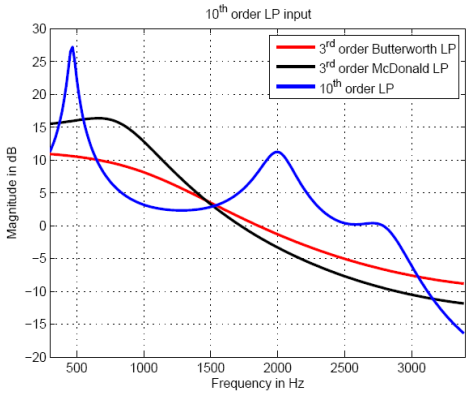


Figure 4. Power Spectral Densities of Example Sources

For the unweighted case, the excitation shaping filter has the form shown in Fig. 5. Expanding both sides of the numerators in Eq. (5), we obtain n+1 nonlinear equations in n+1 unknowns. Using a nonlinear optimization technique, these equations can be solved for the coefficients of *B(z)* [12].

For the case with weighting, we use the weighting function

$$W(z) = A(z/\gamma_1)/A(z/\gamma_2) \qquad (9)$$

where γ_1 and γ_2 are 0.94 and 0.6, respectively, since these parameters are used for some modes of the AMR-NB codec. With weighting, the structure of *B(z)* becomes more complicated and has the form shown in Fig. 6. In this case, upon expanding the numerator of the expression in Eq. (7), where *R(z)* is the numerator of *B(z)* and *Q(z)* is the denominator of *B(z)*, we obtain 2n+1 nonlinear equations in 2n+1 unknowns. The solutions of this nonlinear optimization are nonunique and depend on the initial conditions and the optimization scheme [12].

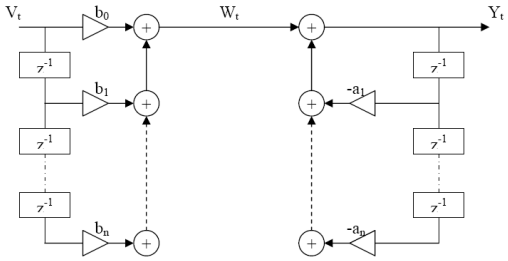


Figure 5. Excitation Shaping Filter *B(z)* for No Perceptual Weighting

The details of calculating the *B(z)* expressions are given elsewhere, but we exhibit the magnitude response of the resulting filters for the unweighted and weighted cases and different average distortions for each of the three sources in Figs. 7, 8, and 9.

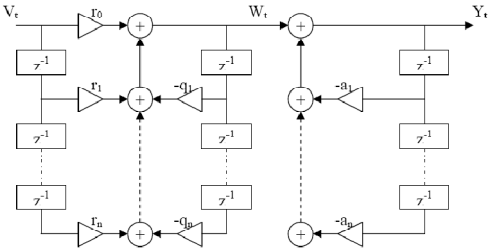


Figure 6. Excitation Shaping Filter *B(z)* with Perceptual Weighting

From Figs. 7 and 8 for the 3rd order sources, we observe the following:

- *B(z)* has a low-pass filtering effect, the intensity of which increases with an increase in distortion.
- For the same distortion, the frequency response of *B(z)* with no weighting has a more severe low-pass filtering effect relative to the frequency response of *B(z)* with weighting.
- As distortion *D* is increased, we reach a point where the small distortion condition is not valid for MSE distortion but remains valid for weighted MSE distortion. This is true in the case of *D*=0.15 for the Butterworth coefficients and *D*=0.1 for the McDonald coefficients.

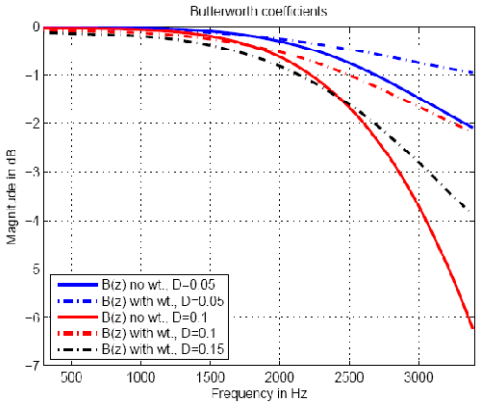


Figure 7. *B(z)* for 3rd Order Butterworth Source

The low-pass filtering effect of *B(z)* has been attributed in [6] to the rate distortion theory trading off the high frequency signal component (where the quantization effects are centered) against a reduction in noise. When weighting is used, the weighted distortion at high frequencies is reduced due to a redistribution of the noise across the spectrum. This may explain why the low-pass effect is less severe in case of weighted MSE, relative to when no weighting is used.

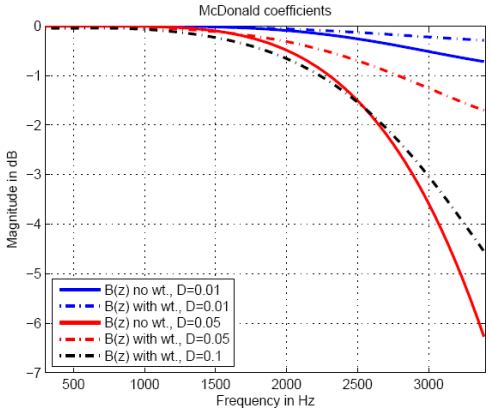


Figure 8. *B(z)* for 3rd Order MacDonald Source

For the AR(10) source in Fig. 9, the observations are similar to those that we had observed earlier for AR(3), but the weighting filter parameters play a more prominent role.

From Fig. 9, we note that:

- The low pass filtering effect in the frequency response of *B(z)* increases in severity as the distortion increases for both MSE and weighted MSE.
- In comparing the frequency response of *B(z)* without and with weighting, we see that weighting reduces the severity of the low pass filtering, relative to the case when no weighting is used. This is true for the cases where both the unweighted and weighted MSE satisfy the small distortion condition (D=0.005).
- For D=0.01 and D=0.05, the case without weighting does not satisfy the small distortion condition, and hence *B(z)* cannot be determined. However, for these specified values of *D*, when weighting is used, the small distortion condition is satisfied, allowing us to determine *B(z)*.
- We observe some shaping in the low frequencies visible for *B(z)* with weighting when D=0.05. To investigate this effect,we adjusted the weighting filter parameters, and the *B(z)* for a value of γ_2 =0.2 is shown in Fig. 10.

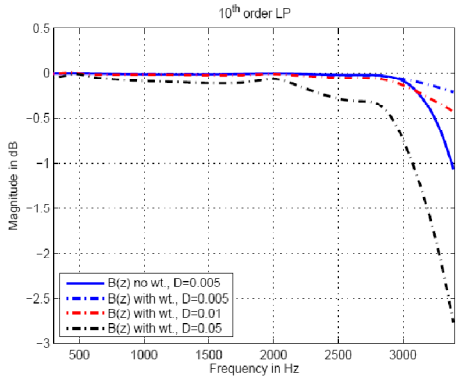


Figure 9. *B(z)* for 10th Order AR Source

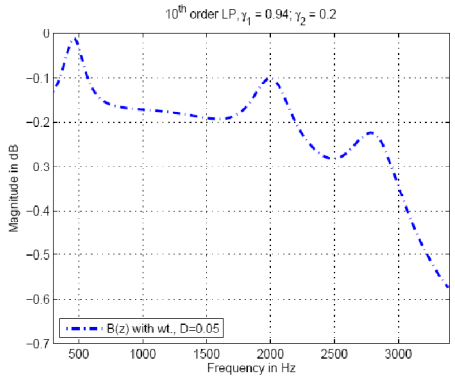


Figure 10. *B(z)* for 10th Order AR Source with Different Weighting

We see that the shaping is more prominent and the formants on the source spectrum start to appear. This value for the parameter γ_2 may appear due to adaptation of the perceptual weighting filter, but some codecs put upper and lower bounds on the value of this parameter. For

example, G.729 bounds γ_2 to be between 0.4 and 0.7 [14]. Following the explanation in [6] for the high-frequency effect, the shaping of the *B(z)* observed in Fig. 10 that emphasizes formant frequencies and de-emphasizes formant valleys and high frequency components can be said to be trading-off signal fidelity at formant valleys and high frequencies against a reduction in perceived noise.

5. Decoder Structures in CELP Codecs

For comparison purposes, we now examine the reconstructed output of the common standardized CELP codecs. The weighted reconstruction error in the analysis by synthesis calculation has the form

$$E(z) = W(z)[S(z) - \hat{S}(z)] \tag{10}$$

which upon rewriting yields,

$$S(z) = \hat{S}(z) + \frac{1}{W(z)} E(z) \tag{11}$$

By comparing to the Shannon backward channel condition, we see that this has the same form, and if the sequences are Gaussian and if we argue that $\hat{S}(z)$ is selected to satisfy the orthogonality condition from optimal estimation, the error will be independent of the reconstructed output. However, the expression for the transfer function of the decoder is

$$H(z) = \frac{1}{A(z)} \tag{12}$$

Therefore, the common CELP decoders (and encoders) lack the excitation shaping filter implied by rate distortion optimal encoding.

CELP decoders often have a postfilter with a numerator polynomial that may inadvertently compensate for this oversight, although the dependence on the average distortion is not explicit and of the same form. The rate distortion motivated structures will have the excitation shaping at the encoder too, within the analysis by synthesis loop, which is always lacking in current CELP codecs.

6. Implementations

The process to determine the rate distortion theory motivated excitation filter is quite involved, and therefore one has to determine how this approach might be incorporated into a practical voice codec. One approach would be to use the structure in Fig. 6, with the weighting parameter updated as in CELP, but with the remaining coefficients adapted according to algorithms as in Gibson [15]. Several other approximate structures are possible.

7. Conclusions

The Shannon backward channel result from rate distortion theory is shown to require zeros in the decoder/synthesizer structures for common all pole, linear predictive speech models. This is in contrast to the usual decoders/synthesizers used in popular CELP codecs. We calculate the transfer functions for these new structures for common source models and discuss the effect on the reconstructed signal. The parameters in the perceptual weighting filter and the average distortion both change the shaping of the excitation. Although the common CELP encoders and decoders lack these components, it is noted that, at least at the decoder, postfiltering as currently used in CELP may partially compensate for the missing structure.

References

[1] J. D. Gibson, "Speech Coding Methods, Standards, and Applications," *IEEE Circuits and Systems Magazine*, vol. 5, no. 4, pp. 30 – 49, 2005.

[2] M. Neuendorf, et al, "A Novel Scheme for Low Bitrate Unified Speech and Audio Coding – MPEG RM0, Convention Paper 7713, AES 126th Convention, Munich, Germany, May 7-10, 2009.

[3] K. Jarvinen, I. Bouazizi, L. Laaksonen, P. Ojala, A. Ramo, "Media coding for the next generation mobile system LTE," *Computer Communications*, vol. 33, pp. 1916-1927, 2010.

[4] J. D. Gibson, T. Berger, T. Lookabaugh, D. Lindbergh, R. Baker, *Digital Compression for Multimedia: Principles & Standards*, Morgan Kaufmann Publishers, Inc., 1998.

[5] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*, Prentice Hall, Englewood Cliffs, NJ, 1971.

[6] J. B. Anderson, J.B. Bodie, "Tree encoding of speech," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 379-387, July 1975.

[7] S. G. Wilson, S. Husain, "Adaptive Tree Encoding of Speech at 8000 Bits/s With a Frequency-Weighted Error Criterion," *IEEE Trans. Comm.*, vol. COM-27, pp. 165-170, Jan. 1979.

[8 M. R. Schroeder, B. S. Atal, "Rate Distortion Theory and Predictive Coding," ," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1981, pp. 201-204.

[9] L. C. Stewart, R. M. Gray, Y. Linde, "The design of trellis waveform coders," *IEEE Trans. Commun.*, vol. COM-30, pp. 702-710, Apr. 1982.

[10] M. R. Schroeder, 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*, 1985, pp. 25.1.1-25.1.4.

[11] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Conv. Rec*., vol. 7, 1959, pp. 142-163.
[12] N. A. Shetty, "Tandeming in Multi-Hop Voice Communication," Ph. D. Dissertation, ECE Dept, UCSB, Dec.2007.
[13] R. A. McDonald, "Signal-to-noise and idle channel performance of differential pulse code modulation systems-Particular applications to voice signals," *Bell Syst. Tech. J.*, vol. 45, pp. 1123-1151, Sept. 1966.
[14] R. Salami, et al, "Design and Description of CS-ACELP: A Toll Quality 8 kb/s Speech Coder," *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 116-130, March 1998.
[15] J. D. Gibson, "Adaptive Prediction in Speech Differential Encoding Systems," Proceedings of the IEEE, vol. 68, pp. 488-525, April 1980.

Jerry D. Gibson is Professor and Chair of Electrical and Computer Engineering at the University of California, Santa Barbara. He has been an Associate Editor of the *IEEE Transactions on Communications* and the *IEEE Transactions on Information Theory*. He was President of the IEEE Information Theory Society in 1996, and he has served on the Board of Governors of the IT Society and the Communications Society. He was a member of the Speech Technical Committee of the IEEE Signal Processing Society from 1992-1994, and he is currently a member of the Multimedia Communications Technical Committee of the Communications Society. He was an IEEE Communications Society Distinguished Lecturer for 2007-2008, a member of the IEEE Awards Committee (2008-2010), and a member of the IEEE Medal of Honor Committee (2009-2010). He is an IEEE Fellow, and he has received The Fredrick Emmons Terman Award (1990), the 1993 IEEE Signal Processing Society Senior Paper Award, the 2009 IEEE Technical Committee on Wireless Communications Recognition Award, and the 2010 Best Paper Award from the *IEEE Transactions on Multimedia*.

A High-Fidelity Channel Simulator for Underwater Acoustic Communication

Aijun Song and Mohsen Badiey, College of Earth, Ocean, and Environment, University of Delaware, Newark, DE 19716, USA
{ajsong, badiey}@udel.edu

Abstract
High frequency acoustic communication (8-50 kHz) has attracted much attention recently. At these high frequencies, various physical processes, including surface waves, subsurface bubbles, and ocean volume fluctuations, can significantly affect the communication channel. While there is on-going work, the research community is still lacking adequate models that can provide realistic representations of the dynamic channel in the ocean. A high-fidelity communication channel simulator is developed here through the use of parabolic equation (PE) modeling of acoustic propagation and scattering. The channel simulator consists of a PE model and a linear surface model. The linear surface model generates an evolving surface based on theoretical or experimental directional surface spectrum and feed the surface displacement and its derivatives to the acoustic model. The time-varying acoustic field is calculated using successive PE runs when the surface evolves. It also accounts for propagation through the water column and through the sediment based on other environmental measurements such as sound speed profile, bathymetry, and bottom properties. The channel simulator is calibrated by experimental data obtained in the Pacific Ocean in 2008. The surface model simulates a time-evolving surface from the directional surface spectrum obtained by a Waverider buoy in the experiment. Based on the surface input and other environmental measurements, the channel simulator can generate realistic time-varying impulse responses. The output agrees well with the acoustic measurements in terms of arrival time structure, intensity profile, and fluctuation characteristics.

1. Introduction

Underwater acoustic communication technology is critical for many scientific, industrial, and naval applications such as ocean exploration and observation, navigation and telemetry for autonomous underwater vehicles, etc. This is especially true for the coastal regions, where high speed acoustic communication is of high interest to multiple communities, including oil and gas industries and oceanographic research communities. However, achieving high data rate acoustic communication in the ocean is still a challenging task [1]. One of the main features of underwater acoustic channels is the limited available bandwidth. At the medium communication range of 1-10 km, the bandwidth can only be a few tens of kilohertz, compared with a few hundreds of megahertz bandwidth or more in radio wireless communication. The major obstacle to bandwidth-efficient communication is the large delay spread, which often leads to significant inter-symbol interference (ISI). Further, various physical processes, including surface waves, subsurface bubbles, and ocean volume fluctuations, can significantly affect the channel, making acoustic communication even more challenging.

Although there is a large body of research literature dealing with acoustic wave propagation, there are few reported efforts to model acoustic communication channels, for example [2]. The research community is still lacking adequate numerical models that can provide realistic representations of both deterministic and stochastic channel properties in the dynamic ocean. Advancements of underwater acoustic communication technology mainly rely on at-sea experiments. These experiments often introduce high cost, although they can provide ultimate algorithm validation. However, the acoustic channel is also highly dependent on the oceanographic condition and the location. It is nearly impossible to test communication algorithms for all ocean conditions and in every part of the ocean.

Some efforts have focused on using experimental data to establish acoustic channel libraries for algorithm development and evaluation [3]. Acoustic channels generated from this method are still limited to the physical measurements available (range, receiving element spacing, number of source and receiving elements, ocean condition etc.). Numerical models are free of these physical limits. For example, they can provide a large number of receiving elements with arbitrary element spacing.

The other issue is performance comparison among different communication algorithms. A large number of high data rate transceivers have been developed since the introduction of coherent communication in the 1990s [4], including multichannel decision feedback equalizers (DFEs) [4], time reversal receivers [5-6], orthogonal frequency-division multiplexing methods [7], etc. These algorithms were tested in different ocean locations, environmental conditions, and source-receiver settings. A channel simulator can provide a common platform for algorithm performance comparison.

Furthermore, a channel simulator can be used to investigate channel limits. It has been shown that long-term oceanographic variability can generate significant performance variation for acoustic communication systems [5-6]. For example, during a 2008 experiment in the Pacific Ocean, i.e. KAM08 [6], source depth and receiver location were found to have significant impact on communication performance (up to 6-8 dB) amid oceanographic events such as tidal internal waves. This suggests the channel capacity is affected by ocean condition and source-receiver geometry.

For medium communication range (1-10 km) and high acoustic frequency (8-50 kHz), the dynamic sea surface is often responsible for the rapid channel fluctuations in shallow water. Due to their importance to acoustic communication, surface effects on acoustic transmissions have been considered in several efforts. For example, different ray-based methods were used to simulate Doppler effects resulting from a moving sea surface and from moving sources and receivers [2].

For communication use, both intensity of the acoustic signal and its coherence over the scale of several seconds are important. In this work, parabolic equation methods are used to model both aspects as a result of sea surface dynamics. Particularly, we use the Miami-Monterey Parabolic Equation (MMPE) model [8], which employs a split-step Fourier marching algorithm and provides fast implementation among various parabolic equation codes. The MMPE model is combined with linear surface model to simulate time-varying acoustic field.

During the KAM08 experiment, extensive acoustic communication signals were tested for different source-receiver geometries and frequency bands. Concurrent environmental measurements were obtained including surface wave spectra, wind speed, sound speed profiles and bottom properties. Using the experimental data, it is shown that the channel simulator can generate realistic impulse responses. Measured wave spectra from the experiment are used to generate a linear time-evolving surface. The modeled impulse responses are compared with field acoustic data. Using experimental data for surface wave generation and also in validation of the model distinguishes our efforts from other modeling work in the literature [2].

2. Methodology

The simulator consists of the MMPE model and a linear surface model. The surface model generates an evolving surface based on directional surface spectra. The surface displacement and its derivations are then fed to the acoustic model. The acoustic field is calculated using successive MMPE runs as the surface evolves. At each single run, the acoustic model accounts for surface scattering effects based on the surface input at that time instant. It also accounts for propagation through the water column and sediment based on other environmental measurements such as sound speed profiles, bathymetry, and bottom properties. The water column and sediment properties are set as static during the successive MMPE runs since they change at a much slower rate than the sea surface. Thus, a time-varying acoustic field is generated. Broadband calculations at multiple frequency bins from MMPE then give time-varying impulse responses.

Based on a gravity wave model [9], a linear surface is generated from the frequency and direction spectrum given by the surface Waverider buoy. The surface wave power spectrum is converted to the surface in physical space, where phase terms are randomly generated. Then an evolving surface is generated using fourth order Runge-Kutta time integration techniques.

3. Results

This section presents our experimental data from KAM08. Based on experimental input, the model is utilized to generate time-varying impulse responses. Comparison between the experimental data and the model output is also presented.

The KAM08 experiment was conducted from June 16 to July 2, 2008 west of Kauai, Hawaii [6]. The water depth was about 100 m at the site. The experimental setting is illustrated in Fig.1 for JD181 (June 29) 12:30:00Z.

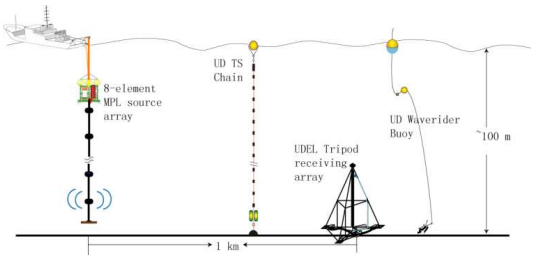
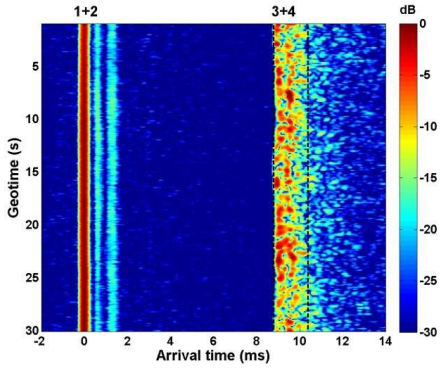


Figure 1. KAM08 Experimental Setting.

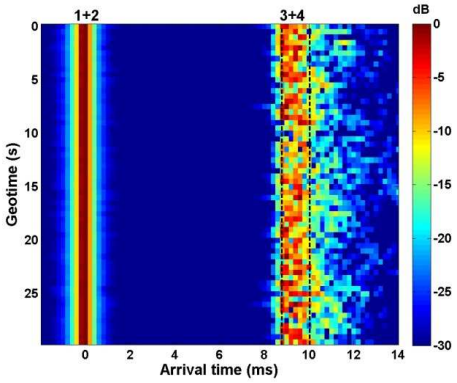
As shown, an 8-element source array was deployed off the stern of the research vessel Melville. A 30 second long maximum length sequence from the bottom source is used here in modeling and data analysis. The center frequency of the sequence was 15 kHz and the chip rate was 5 kHz. A 5-element receiving array was mounted on a rigid tripod structure at the seafloor, 1 km away from the source along the 100 isobath. Along with the acoustic measurements, detailed environmental data including surface wave spectrum and water column temperature profiles were collected. The surface wave spectrum was measured by a directional wave-rider buoy deployed close to the receiving array. The sea surface was relatively calm, with a significant wave height of about 0.7 m during the considered period. A thermistor string was deployed to measure the water column temperature profiles.

Figure 2 provides an example model result for the KAM08 setting. The measured impulse responses from 30 second maximum length sequences are shown in Fig. 2(a). The impulse responses in Fig. 2(a) were obtained through correlating the received signal with the transmitted maximal length sequence every 0.1022 second (or 511 chips). At the receiving array, the first four major paths were direct, bottom, surface, and surface-bottom paths, confirmed by ray code simulations. Since the receiver was positioned just 2 m above the seafloor, the acoustic arrivals came in pairs. As shown in Fig. 2(a), the first two arrivals, i.e., direct and bottom paths marked as "1+2" overlapped with each other and formed a single strong peak. The arrivals around 10 ms (marked as "3+4" and also indicated by the black dashed lines) corresponded to the surface and surface-bottom paths, which were highly fluctuating.

The model calculated the acoustic field based on the environmental measurements during KAM08 including the bathymetry, bottom property, and sound speed profile.



(a) Data



(b) Model output

Figure 2. Impulse response over a 30 second period for (a) Tripod data and (b) PE model in the KAM08 setting. Numbers atop figures indicate direct (1), bottom (2), surface (3), and surface-bottom (4) paths. Dotted black lines indicate the surface paths.

A time-evolving rough surface was generated from the directional surface spectrum given by the Waverider buoy in the experiment. Successive MMPE simulations every 0.125 second based on the evolving surface generated 30 seconds of impulse responses in Fig. 2(b). At

each MMPE run, 512 frequency points evenly distributed in the 5 kHz signal band were calculated for a 2-D domain with a 260 m depth and a 1 km range.

The model output in Fig. 2(b) largely reproduced the arrival-time structure, compared with experimental data shown in Fig. 2(a). Some weak returns existed after the direct and bottom paths in the experimental data that were not present in the model results. The difference was attributed to the measured sound speed profile, which might not reflect the range-dependency of the water column. The model also reproduced the time-varying property of the surface paths. Similar to the experimental data, the model output showed strong, but fluctuating, specular returns around arrival time 10 ms. The model also generated weak dispersive signals following the specular returns as a result of non-specular scattering. To make further comparison, average intensity profiles for both experimental data and model results were calculated. The model generated proper intensity levels for each arrival. The two surface paths had initial peaks and their intensity decreased similarly in the data and model results.

In summary, the channel simulator can generate realistic time-varying impulse responses, using environmental parameters collected during at-sea experiments. The output agreed well with the acoustic measurements in terms of arrival time structure, intensity profile, and fluctuation characteristics.

Acknowledgement

The authors wish to thank Dr. Kevin B. Smith from Naval Postgraduate School for his help on use of the MMPE model. The research work has been supported in part by Office of Naval Research (ONR) Code OA322 (Grants Nos. N00014-10-1-0396 and N00014-10-1-0345) and by University of Delaware Research Foundation.

References

[1] D. B. Kilfoyle and A. B. Baggeroer. The state of the art in underwater acoustic telemetry. *IEEE J. Oceanic Eng.*, 25(1):4–27, Jan. 2000.

[2] M. Siderius and M. B. Porter. Modeling broadband ocean acoustic transmissions with time-varying sea surfaces. *J. Acoust. Soc. Am.*, 124(1):137–50, Jul. 2008.

[3] P. van Walree, T. Jenserud, and M. Smedsrud. A discrete-time channel simulator driven by measured scattering functions. *IEEE J. Sel. Areas Commun.*, 26(9):1628–1637, Dec. 2008.

[4] M. Stojanovic, J. A. Catipovic, and J. G. Proakis. Phase-coherent digital communications for underwater acoustic channels. *IEEE J. Oceanic Eng.*, 19(1):100–111, Jan. 1994.

[5] A. Song, M. Badiey, H. C. Song, W. S. Hodgkiss, M. B. Porter, and the KauaiEx Group. Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx). *J. Acoust. Soc. Am.*, 123(2):856–865, Feb. 2008.

[6] A. Song, M. Badiey, H. C. Song, and W. S. Hodgkiss. Impact of source depth on coherent underwater communications. *J. Acoust. Soc. Am.*, 128(2):555–558, Aug. 2010.

[7] B. Li, J. Huang, S. Zhou, K. Ball, M. Stojanovic, L. Freitag, and P. Willett. MIMO-OFDM for high rate underwater acoustic communications. *IEEE J. Ocean. Eng.*, 34(4):634–644, Oct. 2009.

[8] K. B. Smith. Convergence, stability, and variability of shallow water acoustic predictions using a split-step Fourier parabolic equation model. *J. Comp. Acoust.*, 9(9):243–285, 2001.

[9] D. Dommermuth and D. Yue. A high-order spectral method for the study of nonlinear gravity waves. *J.Fluid Mech.*, 184:267–288, 1987.

Dr. Aijun Song (S'02–M'05) received his Ph.D. degree in Electrical Engineering at the University of Delaware, Newark, DE in 2005. From 2005 to 2008, he was a Postdoctoral Research Associate at the College of Earth, Ocean, and Environment (CEOE), University of Delaware. During this period, he was also an ONR postdoctoral fellow supported by the special research award from the Ocean Acoustics program. Since 2008, he has been an Assistant Professor of the Physical Ocean Science and Engineering program, University of Delaware. His general interests include underwater acoustic signal propagation, digital communication theory, and advanced signal processing in mobile radio frequency and underwater acoustic environments.

Dr. Mohsen Badiey ('94) received his Ph.D. degree in applied marine physics and ocean engineering from the University of Miami, Rosenstiel School of Marine and Atmospheric Science, Miami, Florida, in 1988. He was a Post-Doctoral Fellow at the Port and Harbour Institute, Ministry of Transport, Tokyo, Japan, from 1988 through 1990. In 1990, he became a faculty member at the CEOE, University of Delaware, where he presently is a full Professor. From 1992 to 1995, he directed the ocean acoustics program at the ONR. Dr. Badiey's research interests are physics of sound and vibration, underwater acoustics in shallow water regions, acoustical oceanography, underwater acoustic communications, seabed acoustics, and geophysics. He is a Fellow of the Acoustical Society of America.

The IETF Opus Speech and Audio Codec - Yet Another Codec?

Christian Hoene, 31. October 2011, Tübingen
christian.hoene@symonics.com

1. Introduction

Codecs, which compress and decompress multimedia data streams, are important as they make efficient use of transmission and storage resources. Many codecs have been developed: The ITU-T Media Coding Summary Database counts 54 different speech and audio codecs [1] that have been standardized at ITU-T, MPEG, 3GPP and other Standard Developing Organizations (SDOs).

For Internet Telephony, only a few of them are actually used. For narrow band speech transmissions, µ- and A-law logarithmic coding has been used for decades. These codecs compress speech sampled at 8 kHz down to a rate of 64 kbit/s. They are standardized in ITU-T G.711. Google Talk and Skype are using proprietary codecs such as ISAC and Silk, which support wideband speech using a sampling rate of 16 kHz and superwideband at 32 kHz. Also, the ITU-T study group 16 has developed and standardized many advanced codes [2] that support wide- and superwideband speech or fullband audio such as ITU-T G.711.1, G.718, G.719, and G.729.1.

In the view of these facts, it can be assumed that enough codecs have been developed and hence there is no need for yet another one. However, the experienced VoIP user might confirm that Internet Telephony does not always work as good as it should do:

1.Muffled sounding narrow band transmissions are used even if plenty of transmission bandwidth could be used.

2. For good sounding speech and audio codecs royalties must be paid. However, most softphones can be downloaded for free and calling via VoIP does not introduce additional costs beside those for the Internet access.

3. Frequently, the call quality of VoIP is below acceptable quality levels, especially if bandwidth is limited, if packet loss rates are high and if on wireless links temporal interruptions occur.

4. The most demanding telephone application scenario, playing music together over the Internet, is not supported by any standardized codec. Only the non-efficient PCM coding and the Bluetooth Subband codecs can be used for this purpose as they have an algorithmic delay of only a few milliseconds while supporting transparent audio quality.

Of course, not all the issues discussed above can be solved by the codec alone. However, a clean-slate approach in designing and standardization of a codec might lead a great step forward. In 2008, we tried to convince traditional SDOs to start the work on a royal-free high quality codec [3] – the open-source audio codec CELT by Jean-Marc Valin [4] seemed to be a good starting point – but our request failed. Thus, we asked at the Internet Engineering Task Force (IETF) under which conditions they would support the standardization of a codec. As compared to other SDOs, the standardization process at the IETF has advantages:

First, they have a long tradition of preferring royal-free or even patent-free solutions against those, which are available at Reasonable and Non Discriminatory Licensing (RAND) or other terms. As a consequence, the most important Internet protocols can be implemented without any financial IPR burden.

Second, the intended codec is to be used on the Internet and the IETF is the standardization body for making Internet protocols. Thus, it would make sense to put the codec developers and Internet experts on the same table to foster a mutual knowledge exchange.

Third, the IETF did hardly have any experiences on standardizing codecs. Thus, new innovative ways of the beaten path are more likely to be adopted so that a disruptive technologies can be developed.

In January 2009, Skype also went to the IETF to start the standardization of its newly developed Silk codec [5]. This step convinced the IETF that there is a considerable interest from the industry to start the work on codecs. As a consequence, in November 2009, the IETF Codec Working Group was founded, and now, two years later, a new codec called “Opus” is ready to be standardized.

Opus is having unique features compared to traditional codecs. It supports mono and stereo speech or audio transmission, frame lengths between 2.5 ms and 60 ms, Forward Error Correction (FEC), Discontinuous Transmission (DTX), different algorithmic complexities, different sampling rates, variable bit rate encoding, and many other features. Opus is available as open source and – probably – will be available royalty free.

2. Requirements

Opus is intended for the Internet. Thus, Opus should help to deal with the unique features of Internet transmission paths. In the following, the three most important difference are explained.

2.1 Operational Range

Traditionally, speech and audio codecs have been developed for circuit switched links. For example, G.711 is used for ISDN reserved links and AMR and AMR-WB are applied on GSM and 3G networks in the circuit switched transmission domain. Circuit switched links have a constant bit rate and allow for only for moderate changes of the bit rate. 3GPP phone links may change their bit rate between 4.75 and 12.2 kbit/s depending on the quality of the wireless link and the amount of redundancy needed [6].

The Internet – on the other side – providing only best effort services and does not guaranty any transmission quality at all. In fact, the transmission quality may change rapidly over time. Thus, Internet transmission control protocols must be able to cope with any given transmission rate – whether high or low – and change their sending rate to the current network condition on the fly. For example, the transport protocol TCP works well for transmission paths having a few throughput of a few bits per second up to the gigabit high speed links [7]. Consequently, an audio codec intended for the Internet should also support a large operational range with bit rates from a few bit/s up to the full available bandwidth.

The only constrains should be by the usability of the service and whether the perceptual transmission quality still can be improved.

2.2 Packet vs. Bit Rate

The quality of speech and audio codecs is traditionally measured by looking at the speech/audio quality vs. bit rate. The better the quality at a given bit rate, the better the codec. For circuit switched networks, this metric works fine. However, in the Internet, this metric is misleading as data is sent in packets that have a considerable overhead. The packet overhead is caused by packet headers including those at the physical, link, network, transport and application layer. IP header compression algorithm may shrink the average size of IP, UDP and RTP headers but they do not reduce the impact of lower layers on the packet overhead [8].

In fact, if Internet links are congested, TCP does not control the bit rate of an application flow but limits the frequency of sending TCP segments. Typically, the TCP segment size is set to the Maximum Transmission Unit (MTU) that can be send without fragmentation on a given path [9].

For codec design, this means that a codec should support different packet rates. More precisely, in a given application scenario and on a given transmission link, both bit and packet rate should be controllable. Traditional codecs support a fixed frame size of typically 20ms only. For ensemble performance over the Internet an algorithmic delay of about 20ms is too high to meet the target of 25ms end-to-end delays. On congested low rate links, a frame size of 20ms is too large and a VoIP application should put multiple frames into one RTP packet [10]. If this concatenation is done by the application this would potentially reduce the compression efficiency. A codec might be able to compress longer speech segments better than shorter segments. Thus, if a codec is designed for the Internet, besides quality and bit rate also its frame rate is of importance.

2.3 Variable transmission delay

It is a well-known fact that transmission latency of IP packets is variable. Also, if the packet rate is variable then the algorithmic delay may change too. Thus, before decompression, a dejittering buffer delays received speech frames so that they can be played out in a timely manner. Dynamic playout buffers change the buffering delays to adopt to changed delay conditions. They change the point of play during periods of silences and by removing or adding pitch periods during continuous speech [11]. Thus, to support the dejittering buffers operation, the decoder shall provide information about the currently transmitted audio content via an API.

2.4 Other requirements

Packet loss concealment is also considered as an important feature to support transmission over the Internet. However, current effort in the IETF tries to reduce the frequency of packet losses. Traditionally packet loss events were used as an indication of congestion. Nowadays, with the introduction of the Explicit Congestion Notification (ECN) flag in the IP headers [12], packet losses can be avoided and routers can notify clients about upcoming congestions. Thus, one can assume that the importance of packet loss concealment will reduce if more Internet nodes are going to support ECN.

In the IETF documents [13] and [14], a couple of other features are mentioned, which are important for an Internet speech and audio codec. Those include support of emergency calls, DTMF signals, VAD and FEC. Other features, such as hierarchical compression schemes as in G.729.1 or support of tandem coding, are of minor importance.

3. Codec Design

The Opus codec has been mainly developed by Valin, Kos, Terriberry [15]. In addition, various other developers, users, and testers have

contributed to the current design and implementation.

Opus is based on the CELT und Silk codecs and combines two compression algorithms based on the Linear Prediction (LP) of Silk and the Discrete Cosine Transform (MDCT) of CELT. It supports three different coding modes: LP, MDCT and a hybrid mode.

The LP-only mode is intended for speech signals at lower rates, the hybrid mode is also for speech but used the LP for lower frequency components of the speech signal and MDCT for higher frequencies. The MDCT-only mode is good for music and supports frame sizes down to 2.5ms.

The operational range of Opus is large. Ranging from 6kbit/s up to 510 kbit/s for a fullband transparent stereo coding. Opus can generate frames with sizes from 2.5 ms up to 60 ms while having an algorithmic delay of 5 to 65.2 ms.

If lower rates are needed, a Push-To-Talk like mode can be supported via special transport protocols [16]. If higher rates are available, more audio channels can be transmitted.

Depending on the transmission mode, Opus has internal sampling rates of 8, 12, 16,24 or 48 kHz. In addition, most of the features needed for efficient or high quality Internet transmission, such as DTX, VAD, support for dynamic time adjustment and FEC are included.

4. Quality

As compared to other standardized codecs, Opus is not going to be tested in similar formal manners. Instead of testing Opus in numerous formal listening-only tests, Opus was released already to the public for testing and usage at early stages. Thus, many users have been checking CELT, Silk and now Opus for bugs and weaknesses. However, to have a verifiable statement on the quality performance, various laboratories have conducted formal tests. Google, HydrogenAudio, Nokia and the Universität Tübingen have made listening tests on the latest versions of Opus [17]. The results show that Opus outperforms other codecs. For example, for wide speech at 20kbps Opus is significantly better AMR-WB. For music, Opus shows a similar performances to AAC-LD and MP3 and outperforms G.719. However, for binaural speech Opus was not able to match the coding performance of AMR-WB+.

5. Summary

The codec Opus promises to have unique features as compared to other standardized codecs. Thus, no other codec can be adopted more easily on the Internet transmission.

Its coding performance is comparable with other codecs such as AMR-WB, MP3, and AAC-LD. Modern codecs such as AAC-eUD, USAC, AMR-WB+ and others may outperform Opus at their specific operational ranges.

The recent RTCweb [18] initiative by W3C and IETF aims to add Opus and the video codec VP8 to any browser so that browsers are able to support voice and video calls. Thus, if RTCweb will be successful and if Opus remains free of royality payments, Opus is likely to dominate the sector of interactive audio and speech coding.

Nevertheless, further research must be done on developing effective rate control algorithms for Opus, to enhance packet loss concealment algorithms, and to improve the encoder.

References

[1] C. Lamblin, “ITU-T Media Coding Summary Database”, 30 July 2010, http://www.itu.int/en/ITU-T/studygroups/com16/multimedia/Pages/mcsd.aspx

[2] R. V. Cox, S. F. De Campos Neto, C. Lamblin, M. H. Sherif, "ITU-T coders for wideband, superwideband, and fullband speech communication [Series Editorial]," IEEE Communications Magazine, vol.47, no.10, pp.106-109, October 2009

[3] M. Bartl and C. Hoene, “An open-source softphone for musicians playing over IP”, ITU-T Workshop 'From Speech to Audio: bandwidth extension, binaural perception', September 2008

[4] J. M. Valin, T. B. Terriberry, C. Montgomery, G. Maxwell, "A High-Quality Speech and Audio Codec With Less Than 10-ms Delay," IEEE Transactions on Audio, Speech, and Language Processing, , vol.18, no.1, pp.58-67, Jan. 2010

[5] Skype Journal (2009-01-07) Skype for Windows 4.0 Beta 3 Hotfix Introduces New Audio Codec, Retrieved October 2011

[6] J. Makinen, J. Vainio, "Source Signal based Rate Adaptation for GSM AMR Speech Codec," International Conference on Information Technology: Coding and Computing (ITCC'04), p. 308, Volume 2, 2004

[7] T. Kelly, “Scalable TCP: improving performance in highspeed wide area networks”, ACM SIGCOMM Computer Communication Review, Volume 33 Issue 2, April 2003, NY, USA

[8] C. Mahlo, C. Hoene, A. Rostami, and A. Wolisz, “Adaptive coding and packet rates for TCP-friendly VoIP flows”, In Proc. of 3rd International Symposium on Telecommunications (IST2005), Shiraz, Iran, September 2005

[9] J. Postel, “The TCP Maximum Segment Size and Related Topics“, IETF Request for Comments: 879, November 1983

[10] C. Hoene, H. Karl, and A. Wolisz, “A perceptual quality model intended adaptive VoIP applications”, International Journal of Communication Systems, Wiley, August 2005.

[11] Y. J. Liang, N. Farber, B. Girod, "Adaptive playout scheduling and loss concealment for voice

communication over IP networks," IEEE Transactions on Multimedia, vol.5, no.4, pp. 532- 543, Dec. 2003
[12] S. Floyd, "Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) Field", IETF Request for Comments: 4774, November 2006
[13] C. Hoene, "Requirements of an Audio Communication System (ACS)", IETF draft-hoene-avt-acs-requirements-00, work in progress, August 2009
[14] J. Valin, K. Vos, "Requirements for an Internet Audio Codec", IETF RFC 6366, August 2011
[15] JM. Valin, K. Vos, T. Terriberry, "Definition of the Opus Audio Codec", IETF draft-ietf-codec-opus-08, work in progress, August 2011
[16] C. Hoene and P. Schreiner, "A DTN mode for reliable internet telephony", in ACM Network and Operating System Support for Digital Audio and Video workshop (NOSSDAV), Vancouver, British Columbia, Canada, June 2011
[17] C. Hoene, Ed., J- M. Valin, K. Vos, J. Skoglund, "Summary of Opus listening test results", IETF draft-ietf-codec-results-00, work in progress, October 2011
[18] RTCweb, http://rtc-web.alvestrand.com/, Retrieved October 2011

Dr. Christian Hoene studied computer engineering at the Technical University of Berlin. He wrote his dissertation at the research group of Adam Wolisz on the topic "Internet Telephony over Wireless Links". Since 2005, he is at Universität Tübingen, where he is leading the research group „Interactive Communication Systems". He has written about 40 publications and his research merits are confirmed by about 650 citations and a Hirsch factor of 12 according to Google scholar. Since 2011, he is founder and CEO of the startup Symonics.

Introduction to AVS Audio Lossless Coding/Decoding Standard

Wenxin He, Yi Gao, Tianshu Qu, The Key Laboratory of Machine Perception (Ministry of Education), Peking University, China, 100871
qutianshu@cis.pku.edu.cn

1. Introduction

Recently, massive storage technology, especially Blu-ray Disk, has become more sophisticated. The demand for high quality digital audio can now be met. Audio lossless coding compresses digital audio data without any loss in quality due to a perfect reconstruction of the original signal. Several lossless audio coding technologies have been developed, including MPEG4-ALS, Monkey, WavPack, FLAC, and TAK.

Even though the lossless audio coding technologies mentioned above can satisfy certain requirements, they also have some problems, such as copyright protection and encryption support. Therefore, in December 2009, AVS (Audio Video coding Standard Workgroup of China) issued a call for audio lossless coding proposals [1]. After algorithm assessment, performance evaluation, and technology integration, in July 2010, the proposals from the Institute for Infocomm Research (Singapore), CASKY eTech Co. LTD (China), Peking University (China) were chosen to be combined as the first working draft of the audio lossless coding standard [2]. Since then, further improvements and extensions have been integrated. The final technical specification was issued in July 2011 [3].

The following chapters provide more details of the audio lossless coding standard. Section 2 introduces an overview of the codec structure. Section 3 introduces the algorithm to decorrelate multichannel signals. Section 4 introduces the integer lifting wavelet transform. Section 5 introduces the algorithm entropy coding. At last, the conclusion is given in section 6.

2. Framework of AVS Audio Lossless Coding and Decoding System

2.1 AVS audio Lossless Coder

The AVS audio lossless coder is shown in Fig.1 [3]. A multichannel sound is decorrelated firstly. Secondly, the decorrelated signal is filtered into a high frequency band and a low frequency band by a lifting wavelet filter. Then, both subband signals are processed separately by linear prediction coding (LPC), and a residual signal is produced. Then, the residual signal is sent to the preprocessor, where it is normalized. Lastly, the normalized signal is processed by entropy coding and a bitstream is generated.

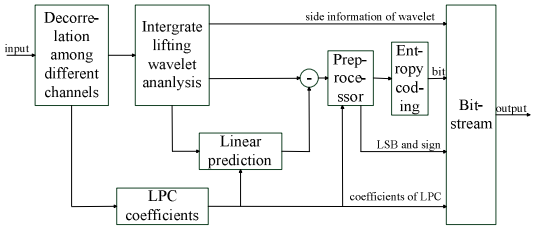


Fig. 1 Framework of AVS audio lossless coding[3]

2.2 AVS audio Lossless Decoder

The AVS audio lossless decoder is shown in Fig.2 [3]. Firstly, a bitstream is processed by the entropy decoder. Then, the residual signal is reconstructed using the output signal of the entropy decoder, LSB, and sign bits. Secondly, using the residual signal and the coefficients of the LPC, subband signals are constructed. Thirdly, the subband signals are processed by the lifting wavelet construct filter. Lastly, the decorrelated multi-channel signals are reconstructed perfectly.

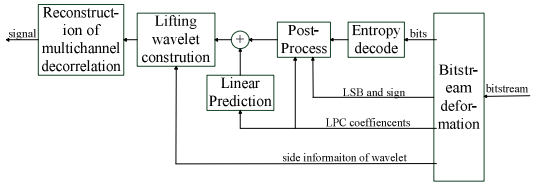


Fig. 2 Framework of AVS audio lossless ecoding[3]

3. Decorrelation among Different Sound Channels

3.1 Coder for Decorrelation

A sum and subtraction coding method is adopted to decorrelate multichannel signals sounds [3,4]. The decorrelate algorithm is as follows,

$$mid = (L + R)/2$$
$$side = L - R \quad (1)$$

where L is left channel signal, and R is right channel signal.

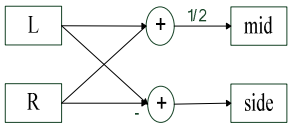


Fig. 3 Framework of decorrelation between L and R

3.2 Decoder of decorrelation

In the process of lossless reconstruction for sum and subtraction coding, the rounding error of the sum channel is compensated by odd and even information from the subtraction channel [3,4]. The algorithm is described as follows,

$$L = mid + (side + side \,\&\, 1)/2$$
$$R = mid - (side - side \,\&\, 1)/2 \quad (2)$$

4. Integer lifting Wavelet Transform

4.1 Coder of Lifting Wavelet Transform

The coder of the lifting wavelet transform has two steps. Firstly, an input signal is processed by the lifting wavelet transform to generate a detail signal (high frequency) and a scale signal (low frequency). Secondly, the two subband signals are processed by the LPC algorithm, respectively [3,5]. In the first step, the integer lifting wavelet is adopted to avoid the rounding error.

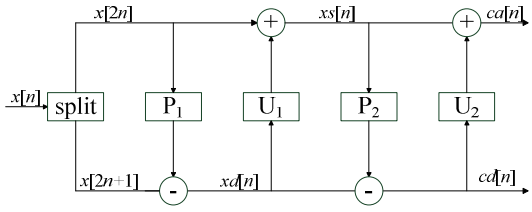


Fig.4 Analysis filter of lifting wavelet

In Fig.4, P1 and P2 are prediction filters, and U1, U2 are update filters.

For lossless coding, the parameters of the prediction filters and the update filters are formatted as Q14 fix-point. The integer lifting wavelet algorithm is detailed as follows,

The output of the first prediction filter is

$$xd[n] = x[2n+1] - \left\lfloor \frac{2731 \cdot (x[2n] + x[2n+2])}{2^{14}} \right\rfloor \quad (3)$$

The output of the first update filter is

$$xs[n] = x[2n] + \left\lfloor \frac{-9216 \cdot (xd[n-1] + xd[n])}{2^{14}} \right\rfloor \quad (4)$$

The output of the second prediction filter is

$$cd[n] = xd[n] - \left\lfloor \frac{21845 \cdot (xs[n] + xs[n+1])}{2^{14}} \right\rfloor \quad (5)$$

The output of the second update filter is

$$ca[n] = xs[n] + \left\lfloor \begin{matrix} 126 \cdot (cd[n-3] + cd[n+2]) \\ -938 \cdot (cd[n-2] + cd[n+1]) \\ +3372 \cdot (cd[n-1] + cd[n]) \end{matrix} \Big/ 2^{14} \right\rfloor \quad (6)$$

After the integer lifting wavelet filter, the output signal, ca and cd, are sent to the linear prediction coder, and processed as follows,

$$d[n] = \begin{cases} x[n] & n = 0 \\ x[n] - \left\lfloor \dfrac{2^{19} + \sum_{k=1}^{n} c[n][k] \cdot x[n-k]}{2^{20}} \right\rfloor & 1 \le n < lpc\_order \\ x[n] - \left\lfloor \dfrac{2^{19} + \sum_{k=1}^{lpc\_order} c[lpc\_order][k] \cdot x[n-k]}{2^{20}} \right\rfloor & lpc\_order \le n < N \end{cases} \quad (7)$$

where x[n] is ca[n] or cd[n], and d[n] is the residual signal, respectively.

4.2 Decoder of Lifting Wavelet Transform

In the decoder, the signals, ca and cd, are reconstructed by the residual signals as follows [3,6],

$$x[n] = \begin{cases} d[n] & n = 0 \\ d[n] + \left\lfloor \dfrac{2^{19} + \sum_{k=1}^{n} c[n][k] \cdot x[n-k]}{2^{20}} \right\rfloor & 1 \le n < lpc\_order \\ d[n] + \left\lfloor \dfrac{2^{19} + \sum_{k=1}^{lpc\_order} c[lpc\_order][k] \cdot x[n-k]}{2^{20}} \right\rfloor & lpc\_order \le n < N \end{cases} \quad (8)$$

Then, the signals, ca and cd, are sent to wavelet decoding filter. Signal x is reconstructed as follows,

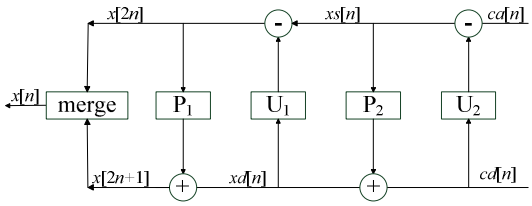


Fig.5 Construction filter of lifting wavelet

The output of the second update filter is

$$xs[n]=cd[n]-\left\lfloor \begin{matrix} 126\cdot(cd[n-3]+cd[n+2]) \\ -938\cdot(cd[n-2]+cd[n+1]) \\ +3372\cdot(cd[n-1]+cd[n]) \end{matrix} \Big/ 2^{14} \right\rfloor \quad (9)$$

The output of the second prediction filter is

$$xd[n]=cd[n]+\left\lfloor \frac{21845\cdot(xs[n]+xs[n+1])}{2^{14}} \right\rfloor \quad (10)$$

The output of the first update filter is

$$x[2n]=xs[n]-\left\lfloor \frac{-9216\cdot(xd[n-1]+xd[n])}{2^{14}} \right\rfloor \quad (11)$$

The output of the first prediction filter is

$$x[2n+1]=xd[n]+\left\lfloor \frac{2731\cdot(x[2n]+x[2n+2])}{2^{14}} \right\rfloor \quad (12)$$

5. Entropy coding

5.1 Coder of Entropy Coding System

Because of wavelet filter and LPC, there often are a few large-valued samples in the first part of residual signal. Therefore, before entropy coding, the power of residual signal should be normalized to attenuate the first few samples. Then, the normalized signal is sent to the entropy coder.

The Entropy coder firstly segregates the input sequence, calculates the mean of each sub-sequence, and then quantifies the mean value. The MSB of mean index and residual signal is coded according to the probability table, which is generated by the reversed quantification and probability model. The output bitstream and LSB of the residual signal are combined to form the output bitsteam [3,6].

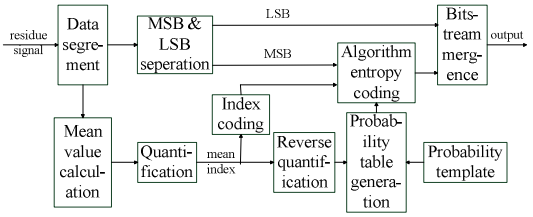


Fig.6 Framework of entropy coding

5.2 Decoder of Entropy Coding System

The models of the entropy decoder are the same as those of the entropy coder. The bitstream is dispatched into the LSB and algorithm bits. The algorithm bits are decoded to generate the MSB of the residual signal according to the probability table, which is generated by the reverse quantification and probability model. The MSB and LSB of residual signal are combined to generate the output signal. [3,6]

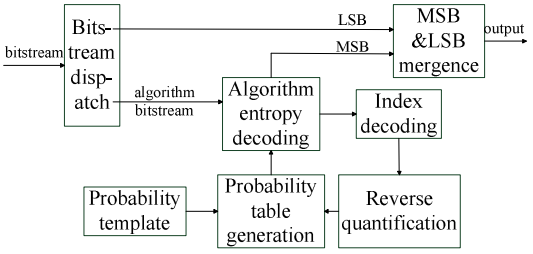


Fig.7 Framework of entropy decoding

6. AVS-ALS Evaluation

The AVS-ALS was compared with several popular lossless coding methods, including Monkey's Audio (extra high/normal), TAK (normal), ALS RM21 (Rice/BGMC 1024 sample), FLAC (normal), and WavPack (default) [4]. The results show that the average compress rates of AVS-ALS are comparable to or better than others.

Table 1 Average compress rate of coder (%)[4]

Coder	32kHz/16bit	44.1kHz/16bit	96kHz/24bit	192kHz/24bit
AVS-ALS	51.44	47.98	46.37	35.22
Monkey EH	50.77	47.02	47.19	35.57
Monkey normal	52.23	48.32	49.08	36.31
TAK	52.29	47.95	47.97	35.01
MPEG4 ALS(RICE)	52.49	48.6	47.89	35.91
MPEG4 ALS(BGMC)	51.95	48.09	47.57	35.54
FLAC	53.31	49.38	51.68	40.33
WavPack	54.39	50.64	51.38	46.03

7. Conclusion

The AVS Audio Lossless Coding Standards is a new standard produced by the Chinese AVS work group for losslessly compressing digital audio data. AVS uses Lifting Wavelet Analysis, Linear Prediction and Entropy Coding to compress digital audio. Experimental results show that the compress

rate of AVS-ALS is comparable and even better than several popular lossless coders.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (61175043; 31170985; 90920302; 60811140086), the "973" National Basic Research Program of China (2010DFA31520; 2011CB70 7805), the Chinese Ministry of Education (20090001110050), the research program from Microsoft China.

References

[1] AVS Requirement Group, "Technology requirement of audio lossless coding," AVS_N1660, Beijing, Dec., 2009

[2] AVS Audio Group, "Resolution of AVS audio group at Sanxia," AVS_N1720, Sanxia, June, 2010

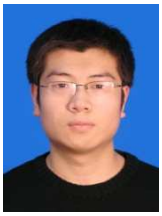
[3] AVS Audio Group, "AVS audio lossless coding final committee draft ," AVS_N1809, Wulumuqi, June, 2011

[4] Xinhui Yang, Haiyan Shu, Tianshu Qu, Tao Zhang, Weipei Dou, "From lossy to lossless: a framework of audio codec," Audio Engineering, vol. 34(12), pp. 60-64, 2010

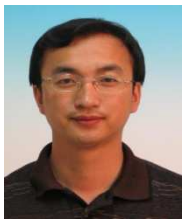
[5] Tianshu Qu, Wenxin He, Yi Gao, Bo Zhang, Xihong Wu, "Lifting wavelet transform based on audio lossless coding method," Audio Engineering, vol. 34(12), pp. 65-68, 2010

[6] Haibin Huang, Haiyan Shu, Rongshan Yu, Diyou Chen, Rahardja Susanto, "Introduction to AVS lossless audio coding standard: entropy coding," Audio Engineering, vol. 34(12), pp. 69-71, 2010

Wenxin He received the B.S. degree in Electrical Engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2009. He is currently pursuing the Ph.D. degree at the Speech and Hearing Research Center, Peking University, Beijing, China. His principal interest is in audio coding and speech enhancement.



Yi Gao obtained the B.S. degree from Beijing University of Aeronautics and Astronautics of China, Beijing, China, in 2008. He is currently pursing the Ph D. degree at the Speech and Hearing Research Center, Peking University, Beijing, China. His research interests include audio coding and speech separation.



Tianshu Qu got his B.Eng. degree from Jilin University of Technology, China, in 1993. In 2002, he got Ph.D. degree from Jilin University, China. Currently, he is an associate professor in the Key Laboratory of Machine Perception (Ministry of Education), Peking University, China. His principal interests are in acoustic signal processing, auditory model and virtual sound.

Blind Cyclostationary Carrier Frequency and Symbol Rate Estimation for Underwater Acoustic Communications

Zhiqiang Wu, Department of Electrical Engineering, Wright State University, USA
T. C. Yang, Inst. of Applied Marine Physics and Undersea Tech., Nat. Sun Yet-Sen Univ., Taiwan

1. Introduction

Cyclostationary analysis has been accepted as an important tool to perform signal detection, signal parameter estimation, and modulation detection of radio frequency (RF) signals [1]-[8]. Cyclostationary analysis is based on the fact that communication signals are not accurately described as stationary, but rather more appropriately modeled as cyclostationary. While stationary signals have statistics that remain constant in time, the statistics of cyclostationary signals vary periodically. These periodicities occur for signals of interest in well defined manners due to underlying periodicities such as sampling, modulating, multiplexing, and coding. This resulting periodic nature of signals can be exploited to detect the existence of the signal, estimate important parameters of the signal, and determine the modulation scheme of the unknown signal.

However, due to the extremely complex and dynamic environment of underwater acoustic communications, it is not clear if the cyclostationary analysis is still applicable in these tasks. The underwater acoustic channels are extremely complex and dynamic [9-11]. The underwater acoustic communication signals experience severe Doppler shifts, multipath effects, phase noise and variations over time. As a direct result, the cyclostationary analysis faces significant challenges in analyzing underwater acoustic communication signals.

In this paper, we apply cyclostationary analysis to simulated data obtained by convoling the measured time-varying channel impulse with the transitted symbols; we also include the symbol phase variations as found in the data [12]. We then estimate the signal parameters such as the carrier frequency including the Doppler shift, and symbol rate. As shown previously, the Dopper shift for a moving source can vary rapidly with time [12]. As a result, the conventional analysis using the second order spectral correlation function (SCF), which treats the channel as time invariant, does not always work. To estimate the fast varying Doppler shift, one must apply the cyclostationary analysis to short segements of the signal. To observe the spectral lines in cyclic frequency, very high cyclic frequency resolution is required which demands a high sampling rate. To overcome the computational burden in obtaining high frequency resolution, a dynamic varying resolution algorithm is adpated where high resolution is obtained only around the cyclic frequencies of interest. We note that most of the signal parameter estimation methods reported in the literature do not consider extensive multipaths (frequency selective fading) as found in the underwater channel. While the multipaths are expected to modify the features of the SCF, we find, based on our analysis, that the effect of multipaths can be minimized using the spectral coherence function (SOF) as suggested before [1-2]. As a result, the proposed method using the high resolution SOF is shown to be able to blindly estimate the carrier frequency (and Doppler shift) and symbol rate of the signal.

The rest of the paper is organized as follows: In section II, we present the system model and the cyclostationary analysis. Section III describes the short time window cyclic analysis using the dynamic resolution algorithm. Section IV shows numerical results of the estimation of carrier frequency and symbol rate via simulated data.

2. Cyclostationary Analysis and SCF/SOF

It is well known that most of communication signals are cyclostationary, and cyclostationary features such as spectral correlation function (SCF) and spectral coherence function (SOF) can be used in signal detection, parameter estimation and modulation detection.

SCF can be measured by the normalized correlation between two spectral components of the signal $x(t)$ at $f+\frac{\alpha}{2}$ and $f-\frac{\alpha}{2}$ over Δt interval where f and α are the spectral and cyclic frequencies. With the finite time Fourier transform of the signal, denoted by $X_T(t,f)$, the measurement of SCF for a signal $x(t)$ can be expressed as:

$$S_X^{\alpha}(f) = \lim_{T\to\infty}\lim_{\Delta t\to\infty}\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2}\frac{1}{T}X_T(t,f+\frac{\alpha}{2})X_T^*(t,f-\frac{\alpha}{2})dt$$

where T represents the integration range of the finite time Fourier transform

$$X_T(t,f) = \int_{t-T/2}^{t+T/2} x(u)e^{j2\pi fu}du$$

and Δt denotes the integration range of the correaltion.

A normalized version of the SCF, the Spectral Coherence Function (SOF) is given as:

$$C_X^\alpha(f) = \frac{S_X^\alpha(f)}{[S_X^0\left(f+\frac{\alpha}{2}\right)^* S_X^0\left(f-\frac{\alpha}{2}\right)]^{1/2}}$$

The SOF has a proper coherence value with a magnitude in the range of [0, 1]. An additional benefit of using the SOF is its insensitivity to channel effects. Taking channel effects into consideration, the SCF of a received signal is as:

$$S_Y^\alpha(f) = H\left(f+\frac{\alpha}{2}\right)H^*\left(f-\frac{\alpha}{2}\right)S_X^\alpha(f),$$
$$y(t) = x(t) \otimes h(t)$$

where $h(t)$ is the unknown channel response, and $H(f)$ is the Fourier Transform of $h(t)$. Hence, the resulting SCF of the received signal can be significantly distorted by the channel. However, the channel effect is minimzed through the normalization process in forming the SOF.

3. Two Stage Parameter Estimation

The cyclic frequency features in SCF and SOF are discrete spectral lines, hence requiring very high resolution to observe them. To relieve the computational burden, we propose a twostage parameter estimation method and a dynamic varying resolution SCF/SOF algorithm. As shown in Fig. 1, a spectral analysis based coarse parameter estimation is first performed to obtain a coarse estimation of carrier frequency $\hat{f}_c$ and symbol rate $\hat{f}_b$ (where $f_b = 1/T_b$, with T_b denoting the symbol duration). Next, a cyclostationary analysis based estimation algorithm is used to obtain high resolution estimates $\hat{\hat{f}}_c$ and $\hat{\hat{f}}_b$.

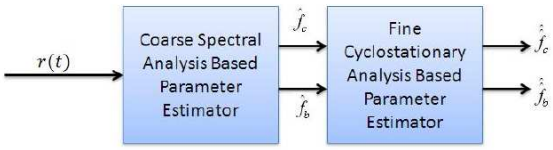


Fig. 1. Block Diagram of Parameter Estimation Algorithm

3.1. Coarse Estimation via Spectral Analysis

In the coarse estimation stage, we first obtain the spectrogram of the signal. Fig. 2 shows the spectrogram of a real underwater communication signal with carrier frequency $f_c = 17000Hz$ and symbol rate $f_b = 4000Hz$. Let $B(f,t)$ denote the spectrum at frequency f and time t. It is clear that the signal of interest is contained in the red rectangular area centered at f_c, with bandwidth approximately the vertical width of the red area. There exist some interferences and distortions out of the red rectangular area. To coarsely estimate the f_c and the f_b, we

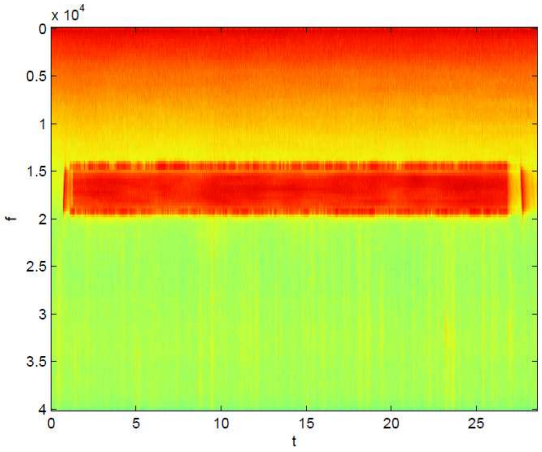


Fig. 2. Spectrogram of Received Signal

1) Find the maximal and minimal levels of the spectrum for different times t:

$$P(t) = max[B(f,t)]$$
$$L(t) = \min[B(f,t)]$$

2) Find the threshold Th_1 to highlight out the red area in this figure. For example, we can set

$$Th_1 = \bar{L} + 0.02 * (\bar{P} - \bar{L})$$

where $\bar{L}$ and $\bar{P}$ denote the time average value of $L(t)$ and $P(t)$, respectively.

3) Update the spectrum by setting the spectrum values to be zero if the magnitude of the spectrum is less than the threshold.

4) Estimate $\hat{f}_c$ and $\hat{f}_b$ as following:

a) Find the highest frequency f_{max} from the non-zero spectrum:

$$f_{max}(t) = max_f\{f: |B(f,t)| > 0$$

b) Compute the estimated highest frequency bound $\bar{f}_{max}$ as the time average of $f_{max}(t)$.

c) Find the lowest frequency bound as:

$$\bar{f}_{min} = max_f\{f: f \in [0, \bar{f}_{max}], \mathrm{B(f,t)} \le 0.5Th_1\}$$

d) Obtain the coarse estimates $\hat{f}_c$ and $\hat{f}_b$ as:

$$\hat{f}_c = (\bar{f}_{max} + \bar{f}_{min})/2;$$
$$\hat{f}_b = \bar{f}_{max} - \bar{f}_{min};$$

3.2. Fine Estimation via Dynamic Varying Resolution Cyclostationary Analysis

With the coarsely estimated carrier frequency and symbol rate $\hat{f}_c$ and $\hat{f}_b$, we propose to use a dynamic varying resolution SCF/SOF algorithm to estimate these parameters accurately. The cyclic frequency features we are interested in occur around $\alpha = 0$

and $\alpha = \pm 2f_c$. Hence, we don't need to calculate the whole SCF/SOF 3-d image with the same resolution. Instead, we only need to provide ultra-high resolution around cyclic frequency at $\alpha = 0$ and $\alpha = \pm 2f_c$ and use much more coarse resolution elsewhere. The algorithm is as follows:

1) Set the cyclic frequency resolution to be $d_\alpha = 1/T_w$(which is the highest cyclic frequency resolution available given the window length of T_w) in $[2\hat{f}_c - \Delta\alpha \le \alpha \le 2\hat{f}_c + \Delta\alpha]$ and $[\hat{f}_b - \Delta\alpha \le \alpha \le \hat{f}_b + \Delta\alpha]$, and $d_\alpha = 0.01$ elsewhere.

2) Calculate the SCF and SOF with the varying resolution along the cyclic frequency axis.

3) Find the peaks in the cyclic frequency of SOF. The location of these peaks are high resolution estimates of carrier frequency and symbol rate, $\hat{\hat{f}}_c$ and $\hat{\hat{f}}_b$.

4. Numerical Results

Now we use direct-sequence code-division-multiple-access data [12] to validate the effectiveness of the proposed algorithm. The transmitted signal has a carrier frequency of $f_c = 17000Hz$ and symbol rate of $f_b = 4000Hz$. The transmitted signal has a near flat spectrum but the received signal shows a spiky spectrum (Fig. 3) due to multipath interferences.

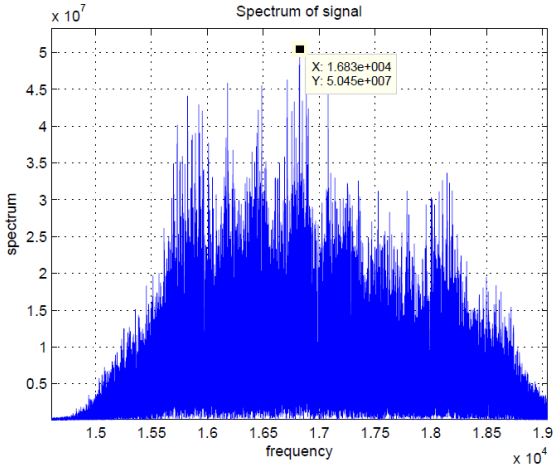


Fig. 3. PSD of the received signal

Fig. 4 illustrates the high resolution SOF around $\alpha = 2f_c$, using data samples of about 0.128 second. As shown in this figure, the SOF exhibits a spectral line at $33{,}970Hz$. Hence, the carrier frequency can be estimated as $33970/2 = 16985Hz$ and Doppler shift can be estimated as $-15Hz$ which is the correct value. Although shown only for one data segment, the method can be applied to consecutive segments to estimate the time varying Doppler shift.

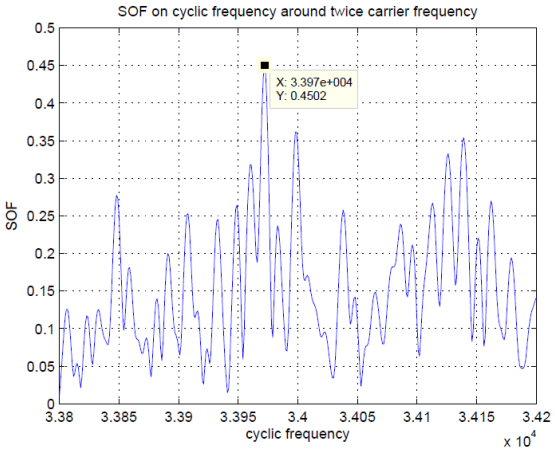


Fig. 4. SOF of received signal around twice the carrier frequency

Fig. 5 shows the high resolution SOF around $\alpha = f_b$. As expected, the SOF also exhibits a spectral line at 4,000 Hz which is the correct symbol rate of the transmitted signal.

5. Conclusion

It is shown in this paper that second order cyclostationary analysis (especially the SOF), using dynamic high resolution sampling, is able to blindly estimate the time varying carrier frequency (including the Doppler shift) and symbol rate of underwater communication signals. The method does not require any a priori knowledge of the transmitted signal.

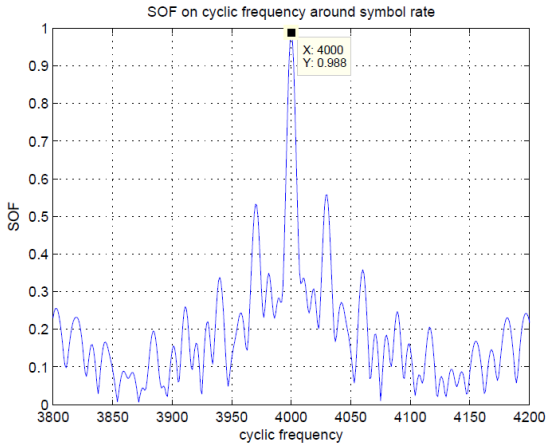


Fig. 5. SOF of the received signal around the symbol rate

Specifically, a two-stage estimation algorithm is developed: First, a coarse estimation based on spectral analysis is performed; second, a fine estimation based on cyclostationary analysis is performed to obtain high resolution estimation of the carrier frequency and symbol rate. Taking

advantage of the coarse estimation in the first stage, a dynamic varying resolution spectrum correlation function and spectrum coherence function estimation algorithm is developed to significantly decrease the computational complexity by only calculating the SOF around the locations of expected features in the cyclic frequency domain. Numerical results from the data analysis validate the effectiveness of the proposed method.

References

[1] W. A. Gardner, *Cyclostationarity in Communications and Signal Processing*. IEEE Press, Piscataway, NJ, USA, 1993.

[2] W. A. Gardner, W. A. Brown, and C.-K. Chen, "Spectral correlation of modulated signals part II: digital modulation," *IEEE Trans. Commun.*, 1987.

[3] W. A. Gardner, "Signal interception: A unifying theoretical framework for feature detection," *IEEE Trans. Commun.*, 1988.

[4] A. Fehske, J. Gaeddert, and J. Reed, "A new approach to signal classification using spectral correlation and neural networks," *IEEE DySPAN*, pp. 144 –150, Nov. 2005.

[5] M. Tsatsanis and G. Giannakis, "Blind estimation of direct sequence spread spectrum signals in multipath," *IEEE Trans. Signal Process.*, vol. 45, pp. 1241–1252, May 1997.

[6] S. Buzzi and V. Massaro, "Parameter estimation and multiuser detection for bandlimited long-code CDMA systems," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2307–2317, June 2008.

[7] W. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Process. Mag.*, vol. 8, pp. 14–36, Apr 1991.

[8] E. Like, V. D. Chakravarthy, P. Ratazzi, and Z. Wu, "Signal classification in fading channels using cyclic spectral analysis", *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, 2009.

[9] D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," *IEEE J. Oceanic Eng.*, 25, 4-27 (2000).

[10] T. C. Yang, "Correlation-based decision feedback equalizer for underwater acoustic communications," *IEEE J. Ocean. Eng.*, vol. 30, pp. 865–880, Oct. 2005.

[11] T. C. Yang, "Properties of underwater acoustic communication channels in shallow water," *J. Acoust. Soc. Am.*, in press (Jan. 2012).

[12] T. C. Yang, and W.-B. Yang, "Low probability of detection underwater acoustic communications using direct-sequence spread spectrum," *J. Acoust. Soc. Am.*, 124, 3632-3647 (2008).

Dr. Zhiqiang Wu received his BS from Beijing University of Posts and Telecommunications in 1993, MS from Peking University in 1996, and PhD from Colorado State University in 2002, all in electrical engineering. He has worked at West Virginia University Institute of Technology as assistant professor from 2003 to 2005. He joined Wright State University in 2005 and currently serves as associate professor. He co-authored one of the first books on multi-carrier transmission for wireless communication. He has published more than 70 papers in journals and conferences. His work on software defined radio implementation of cognitive radio received the Best Demo Award at IEEE Globecom 2010.

Dr. T. C. Yang is a National Science Counsel Chair Professor at the Inst. of Applied Marine Physics and Undersea Technology, Nat. Sun Yet-Sen Univ. Kaohsiung, Taiwan. He is a fellow of the Acoustical Society of America.

E-LELLER EDITORIAL BOARD

MMTC OFFICERS