LSP (Line Spectrum Pair): Essential Technology for High-compression Speech Coding

Takehiro Moriya

Abstract

Line Spectrum Pair (LSP) technology was accepted as an IEEE (Institute of Electrical and Electronics Engineers) Milestone in 2014. LSP, invented by Dr. Fumitada Itakura at NTT in 1975, is an efficient method for representing speech spectra, namely, the shape of the vocal tract. A speech synthesis large-scale integration chip based on LSP was fabricated in 1980. Since the 1990s, LSP has been adopted in many speech coding standards as an essential component, and it is still used worldwide in almost all cellular phones and Internet protocol phones.

Keywords: LSP, speech coding, cellular phone

1. Introduction

On May 22, 2014, Line Spectrum Pair (LSP) technology was officially recognized as an Institute of Electrical and Electronics Engineers (IEEE) Milestone. Dr. J. Roberto de Marca, President of IEEE, presented the plaque (**Photo 1**) to Mr. Hiroo Unoura, President and CEO of NTT (**Photo 2**), at a ceremony held in Tokyo. The citation reads, "Line Spectrum Pair (LSP) for high-compression speech coding, 1975. Line Spectrum Pair, invented at NTT in 1975, is an important technology for speech synthesis and coding. A speech synthesizer chip was designed based on Line Spectrum Pair in 1980. In the 1990s,

Photo 1.   Plaque of IEEE Milestone for Line Spectrum Pair (LSP) for high-compression speech coding.

Photo 2.   From IEEE president to NTT president.

this technology was adopted in almost all international speech coding standards as an essential component and has contributed to the enhancement of digital speech communication over mobile channels and the Internet worldwide."

IEEE Milestones recognize technological innovation and excellence for the benefit of humanity found in unique products, services, seminal papers, and patents, and they have so far been dedicated to more than 140 technologies around the world.

2. Properties of LSP

LSP is an equivalent parameter set of LP (linear prediction) coefficients $a[i]$. Among the various types of linear prediction, AR (auto-regressive) or all-pole systems have mainly been used in speech signal processing. In an AR system, the current sample is predicted by summation (from 1 to p, e.g., 16) of i past sample multiplied by each associated coefficient $a[i]$. A prediction error signal $\hat{x}[n]$ at time n is obtained by the difference between the current sample $x[n]$ and the predicted values of the term $-\sum_{i=1}^{p}$ as

$$\hat{x}[n] = x[n] + \sum_{i=1}^{p} a[i]x[n-i] . \qquad (1)$$

The preferable set of $a[i]$ can be adaptively determined to minimize the average energy of prediction errors in a frame. This relation can be represented by the polynomial of z as

$$A(z) = 1 + \sum_{i=1}^{p} a[i]z^{-i} , \qquad (2)$$

while $1/A(z)$ represents the transform function of the synthesis filter. The frequency response of $1/A(z)$ can be an efficient approximation of the spectral envelope of a speech signal or that of a human vocal tract. This representation, normally called linear prediction coding (LPC) technology, has been widely used in speech signal processing, including for coding, synthesis, and recognition of speech signals. Pioneering investigations of LPC were started independently, but simultaneously, by Dr. F. Itakura at NTT and Dr. M. Schroeder and Dr. B. Atal at AT&T Bell Labs, in 1966 [1].

For the application to speech coding, bit rates for LP coefficients need to be compressed. In 1972, Dr. Itakura developed PARCOR[*1] coefficients to send information equivalent to LP coefficients with low bit rates while keeping the synthesis filter stable. A few years later, he developed LSP [2]–[4], which achieved better quantization and interpolation performance than PARCOR. A set of pth-order LSP parameters is defined as the roots of two polynomials $F_1(z)$ and $F_2(z)$, which consists of the sum and difference of $A(z)$ as

$$F_1(z) = A(z) + z^{-(p+1)} A(z^{-1}) \qquad (3)$$
$$F_2(z) = A(z) - z^{-(p+1)} A(z^{-1}) . \qquad (4)$$

The LSP parameters are aligned on the unit circle of the z-plane, and the angles of LSP, or LSP frequencies (LSFs), are used for quantization and interpolation. An example of 16th-order LSF values $\theta(1),\ldots,\theta(16)$ and the associated spectral envelope along the frequency axis are shown in **Fig. 1**. The synthesis filter is stable if each root of $F_1(z)$ and $F_2(z)$ is alternatively aligned on the frequency axis. It has been proven that LSP is less sensitive to the shape of a spectral envelope; that is, the influence of distortion due to quantization in LSP on the spectral envelope is smaller than it is with other parameter sets, including PARCOR and some variants of it. In addition, LSP has a better interpolation property than others. If we define LSP vector $\Theta_A = \{\theta(1),\ldots, \theta(P)\}$ corresponding to spectral envelope A, the envelope approximated by *envelope*$((\Theta_A + \Theta_B)/2)$ with LSP Θ_A and Θ_B can be a better approximation of the interpolated spectral envelope (*envelope*(Θ_A) + *envelope*(Θ_B))/2 than that with other parameter sets. These properties can further contribute to efficient quantization when they are used in combination with various compression schemes, including prediction and interpolation of LSP itself. These properties of LSP are beneficial for the compression of speech signals.

3. Progress of LSP

After the initial invention, various studies were carried out by Dr. N. Sugamura, Dr. S. Sagayama, Mr. T. Kobayashi, and Dr. Y. Tohkura [5] to investigate the fundamental properties and implementation of LSP. In 1980, a speech synthesis large-scale integration (LSI) chip (**Fig. 2**), was fabricated and used for real-time speech synthesis. Until that time, real-time synthesizers had required large equipment consisting of as many as 400 circuit boards. Note, however, that the complexity of the chip was still 0.1 MOPS (mega operations per second), less than 1/100 of the complexity of chips used for cellular phones in the 1990s.

---

*1 PARCOR (partial auto correlation): Equivalent parameter set of LP coefficients. PARCOR is advantageous in terms of its easy stability checks and better quantization performance than LP coefficients.
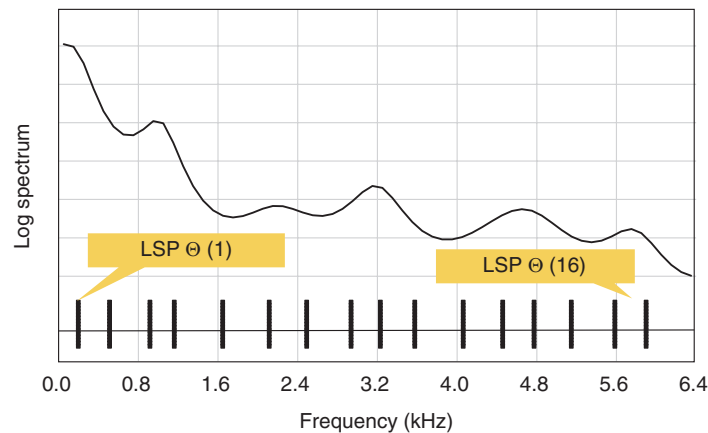
Fig. 1. A set of LSP frequency values and the associated spectral envelope in the frequency domain.
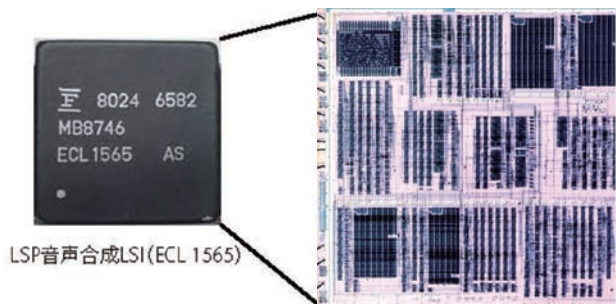

Fig. 2. LSI speech synthesis chip based on LSP in 1980.

Around 1980, low-bit-rate speech coding was achieved with a vocoder scheme that used spectral envelope information (such as LSP) and excitation signals modeled by periodic pulses or noise. These types of coding schemes were able to achieve low-rate (less than 4 kbit/s) coding, but they were not applied to public communication systems because of their insufficient quality in practical environments with background noise. Another approach for low-bit-rate coding was waveform coding with sample-by-sample compression. However, it also could not provide sufficient quality below 16 kbit/s. In the mid 1980s, hybrid vocoder and waveform coding schemes, typically CELP[*2], were extensively studied; these schemes also need an efficient method for representing spectral envelopes such as LSP.

4. Promotion of LSP in worldwide standards

During the 1980s, however, the general consensus was that compression of speech signals would probably not be useful for fixed line telephony, and there was some doubt as to whether digital mobile communications, which requires speech compression, could easily be used in place of an analog system in the first generation. Just before 1990, however, new standardization activities for digital mobile communications were initiated because of the rapid progress being made in LSI chips, batteries, and digital modulation, as well as in speech coding technologies. These competitive standardization activities focusing on commercial products accelerated the various investigations underway on ways to enhance compression, including extending the use of LSP, as shown in **Fig. 3**. These investigations led to the publication of some insightful research papers, including one on LSP quantization by the current president of IEEE, Dr. Roberto de Marco [6].

In the course of these activities, LSP was selected for many standardized schemes to enhance the overall performance of speech coding. The major standardized speech/audio coding schemes that use LSP are listed in **Table 1**. To the best of our knowledge, the federal government of the USA was the first to adopt LSP as a speech coding standard in 1991. The Japanese Public Digital Cellular (PDC) half-rate

---

*2 CELP (code-excited linear prediction): Among large numbers of sets of excitation signals, the encoder selects the most suitable one that minimizes the perceptual distortion between the input and the synthesized signal with LP coefficients. This was initially proposed by AT&T in 1985 and has been widely used as a fundamental structure of low-bit-rate speech coding.

APC-AB: adaptive predictive coding with adaptive bit allocation
ARIB: Association of Radio Industries and Businesses
CS-ACELP: conjugate structure algebraic CELP
EVRC: Enhanced Variable Rate Codec
EVS: Enhanced Voice Service
GSM: Global Standard for Mobile Communications
HVXC: Harmonic Vector Excitation Coding
IETF: Internet Engineering Task Force
MPC-MLQ: Multipulse LPC with Maximum Likelihood Quantization
PSI-CELP: pitch synchronous innovation CELP
QCELP: Qualcomm CELP
RCELP: relaxed CELP
TCX: transform coded excitation
TwinVQ: transform-domain weighted interleave vector quantization
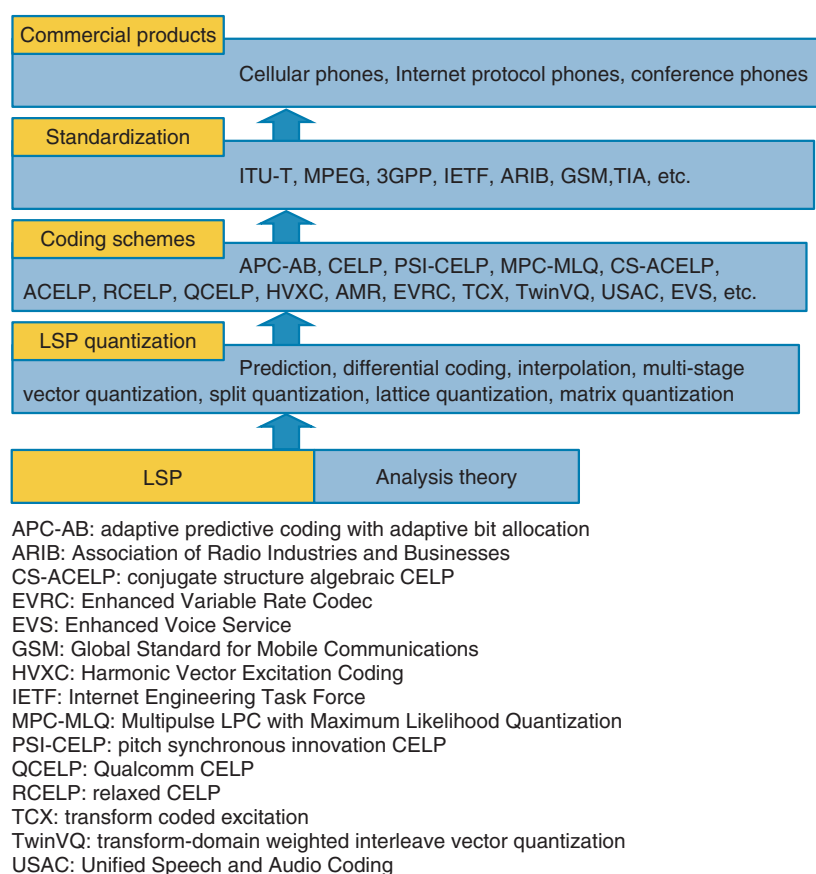USAC: Unified Speech and Audio Coding

Fig. 3. Steps of technologies towards commercial products.

standard in 1993 may have been the first adoption of LSP for public communications systems; the USA and Europe soon followed suit. In 1996, two ITU-T (International Telecommunication Union-Technology Sector) recommendations (G.723.1 and G.729) were published with LSP as one of the key technologies. Both, but especially G.729, have been widely used around the world as default coding schemes in network facilities for Internet protocol (IP) phones. In 1999, speech coding standards for the third generation of cellular phones, which are still widely used around the world, were established by both 3GPP[*3] and 3GPP2[*4] with LSP included.

Furthermore, LSP has proven to be effective in capturing spectral envelopes not only for speech but also for general audio signals [7] and has been used in some audio coding schemes defined in ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) MPEG-4 (Moving Picture Experts Group) in 1999 and MPEG-D USAC (Unified Speech and Audio Coding) in

2010.

5. Future communication

In the VoLTE[*5] service introduced in 2014 by NTT DOCOMO, 3GPP adaptive multi-rate wideband (AMR-WB) is used for speech coding, and it provides wideband speech (16-kHz sampling, the same speech bandwidth as mid-wave amplitude modulation (AM) radio broadcasting). For the next generation of VoLTE, the 3GPP Enhanced Voice Service (EVS) standard is expected to be used, which can

*3 3GPP (3rd Generation Partnership Project): Joint project for third-generation mobile communications by ETSI (European Telecommunications Standards Institute) and Japanese, Korean, and Chinese standardizing bodies. The activities are continuing and are focused on a fourth-generation system.
*4 3GPP2: Joint projects for third-generation mobile communication by the TIA and Japanese, Korean, and Chinese standardizing bodies.
*5 VoLTE: IP-based speech communication system over LTE mobile networks.

Table 1.   Major standards with LSP.

Standardization body	Coding scheme	Bit rate (kbit/s)	Applications	year
Federal govt. of USA	FS1016 CELP	4.8	Govt. communication	1991
Federal govt. of USA	FS1017 MELP	2.4	Govt. communication	1995
Japan RCR (now ARIB)	STD-T27 PSI-CELP	3.4	2nd generation half-rate	1993
USA TIA/EIA	IS-95 RCELP	2,4,8	2nd generation half-rate	1995
Europe GSM	GSM-EFR	12.2	2nd generation enhanced full-rate	1997
ITU-T	G.723.1 MLP-MLQ/ACELP	5.3/6.3	TV (television) phone, IP phone	1996
ITU-T	G.729 CS-ACELP	8	IP phone Cellular phone (PDC)	1996
3GPP	AMR	12.2	3rd generation cellular phone	1999
3GPP2	EVRC	9.6	3rd generation cellular phone	1999
ISO/IEC MPEG-4	14496-3:2009 CELP/HVXC/TwinVQ	2–16	Speech/audio coding	1999
ISO/IEC MPEG-D	23003-3:2012 USAC	8–256	Speech/audio coding	2010
3GPP	AMR-WB	8–23	VoLTE	2001
3GPP	AMR-WB+	6–48	Speech/audio coding	2004
3GPP	EVS	5.9–96	VoLTE	2014

AMR-WB: adaptive multi-rate wideband
EIA: Electronic Industries Alliance
GSM EFR: GSM Enhanced Full Rate
MELP: mixed-excitation linear prediction
RCR: Research and Development Center for Radio Systems
TIA/EIA: Telecommunications Industry Association/Electronic Industries Alliance
VoLTE: voice over Long Term Evolution

handle a 32-kHz sampling rate signal and general audio signals. LSP or a variant of LSP is incorporated in both AMR-WB and EVS. In the near future, it may be possible to achieve all speech/audio coding functions with downloadable software. Even in such a case, we expect that LSP will still be widely used.

In this way, LSP may be a good example of technology that has contributed to the world market. The NTT laboratories will continue to make efforts to enhance communication quality and the quality of services by meeting challenges in research and development.

References

[1]   B. S. Atal, "The History of Linear Prediction," IEEE Signal Processing Magazine, Vol. 23, No. 2, pp. 154–157, March 2006.

[2]   F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals," J. Acoust. Soc. Am., Vol. 57, S35, 1975.

[3]   F. Itakura, "All-pole-type Digital Filter" Japanese patent No. 1494819.

[4]   F. Itakura, "Statistical Methods for Speech Analysis and Synthesis From ML Vocoder to LSP through PARCOR," IEICE Fundamentals Review Vol. 3, No. 3, 2010 (in Japanese).

[5]   F. Itakura, T. Kobayashi, and M. Honda, "A Hardware Implementation of a New Narrow to Medium Band Speech Coding," Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 1982, pp. 1964–1967, Paris, France, May 1982.

[6]   J.R.B. de Marca, "An LSF Quantizer for the North-American Half-rate Speech Coder," IEEE Trans. on Vehicular Tech., Vol. 43, No. 3, pp. 413–419, August 1994.

[7]   N. Iwakami, T. Moriya, and S. Miki, "High-quality Audio-coding at Less Than 64 kbit/s by Using TwinVQ", Proc. of ICASSP 1995, pp. 3095–3098, Detroit, USA, May 1995.

Takehiro Moriya

NTT Fellow, Moriya Research Laboratory, NTT Communication Science Laboratories.

He received his B.S., M.S., and Ph.D. in mathematical engineering and instrumentation physics from the University of Tokyo in 1978, 1980, and 1989, respectively. Since joining NTT laboratories in 1980, he has been engaged in research on medium- to low-bit-rate speech and audio coding. In 1989, he worked at AT&T Bell Laboratories, NJ, USA, as a Visiting Researcher. Since 1990, he has contributed to the standardization of coding schemes for the Japanese PDC system, ITU-T, ISO/IEC MPEG, and 3GPP. He is a member of the Senior Editorial Board of the IEEE Journal of Selected Topics in Signal Processing. He is a Fellow member of IEEE and a member of the Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers, the Audio Engineering Society, and the Acoustical Society of Japan.