Proc. of the 1993 IEEE Int. Conference on Acoustics, Speech, and Signal Processing, ICASSP-93 vol. 2 pp. 9-12 Minneapolis, Minnesota, April 1993.

IMMITTANCE SPECTRAL PAIRS (ISP) FOR SPEECH ENCODING

Yuval Bistritz and Shlomo Peller

Department of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, ISRAEL

ABSTRACT

Immittance Spectral Pairs (ISP) form a new set of parameters proposed for presenting the LPC filter. ISP is close to LSP and owns similar 'niceties', yet it is different. For a filter of order n it consists of a 'gain' and $n-1$ 'frequency' parameters instead of n 'frequency' parameters for LSP. In regarding LPC as a pseudo model for the vocal tract ISP may represent the immittance at the glottis, without posing, like LSP, artificial boundary conditions. In quantization experiments, ISP has been found to compare favorably with LSP. A study of interframe differentiation coding for ISP and LSP is included and it demonstrates the respectives performances of the two sets.

1. Immittance Spectral Pairs

The LPC filter $1/A_p(z)$ is realizable by a cascade of p lattice sections that implement the following Levinson recursions for $n = 1, \ldots, p$, with $A_n^\sharp(z) := z^{-n}A_n(z^{-1})$

$$A_n(z) = A_{n-1}(z) - k_n z^{-1} A_{n-1}^\sharp(z) \qquad (1a)$$

$$A_n^\sharp(z) = -k_n A_{n-1}(z) + z^{-1} A_{n-1}^\sharp(z) \qquad (1b)$$

This scheme can also be understood as a pseudo model for the vocal tract that consist of p lossless sections of different immittances counted from the lips to the glottis with A_n, $A_n^\sharp$ and k_n representing, respectively, wave moving forward, backward in section n and the amount of A_n that is reflected at the boundary (e.g. [1]).

It is possible, however, to replace in the Levinson and other related algorithms the forward and backward variables, the so called "scattering" variables by their sums and differences

$$F_n(z) = A_n(z) + A_n^\sharp(z)\,; \quad G_n(z) = A_n(z) - A_n^\sharp(z) \quad (2)$$

called "immittance" variables (as they may represent the wave field variables sound pressure and volume velocity in the current context, whose ratios are *im*pedance or ad*mittance*) and obtain alternative (and more efficient) immittance variable versions to these algorithms [2, 3].

If the model for the vocal tract is reformulated into immittance variables the input immittance of the model (the immittance at the glottis) becomes [4],

$$\mathcal{I}_p(z) = \frac{A_p(z) - A_p^\sharp(z)}{A_p(z) + A_p^\sharp(z)}\,. \qquad (3)$$

This immittance function is supported by the following theorem [5].

Theorem 1 (ISP). *$A_p(z)$ is stable if and only if $\mathcal{I}_p(z)$ (3) can be written, when $p = 2m$ as*

$$\mathcal{I}_{2m}(z) = \frac{K(1-z^{-2})\prod_{i=1}^{m-1}(1 - 2z^{-1}x_{2i} + z^{-2})}{\prod_{i=1}^{m}(1 - 2z^{-1}x_{2i-1} + z^{-2})}$$

and when $p = 2m+1$ as (4)

$$\mathcal{I}_{2m+1}(z) = \frac{K(1-z^{-1})\prod_{i=1}^{m}(1 - 2z^{-1}x_{2i} + z^{-2})}{(1+z^{-1})\prod_{i=1}^{m}(1 - 2z^{-1}x_{2i-1} + z^{-2})}\,.$$

with p parameters that are real and satisfy

$$-1 < x_{p-1} < \ldots < x_2 < x_1 < 1 \quad , \quad K > 0 \qquad (5)$$

The Immittance Spectral Pairs (ISP) parameters for the LPC polynomial $A_p(z)$ are essentially the p parameters that this theorem defines. The actual parameters for coding purposes may be

$$\omega_i := cos^{-1}(x_i)\,,\ i = 1, \ldots, p-1 \quad ;\ k_p := \frac{K-1}{K+1} \qquad (6)$$

in which case the necessary and sufficient conditions for stability become

$$0 < \omega_1 < \omega_2 < \ldots < \omega_{p-1} < \pi, \text{ and } -1 < k_p < 1 \quad (7)$$

The parameters ω_i represent (angular) frequencies (they may be replaced by normalized frequencies $f_i = \omega_i/2\pi$ used in the reported experiment below or by actual frequencies). They originate from the conjugate pairs of unit circle zeros $exp(\pm j\omega_i)$ of $F_p(z)$ and $G_p(z)$.

This Work was partly supported by the UNITED STATES -ISRAEL BSF under Grant No. 88-425.

0-7803-0946-4/93 $3.00 © 1993 IEEE

2. ISP versus LSP

Since Line Spectra Pairs (LSP) has been proposed for speech analysis [6, 7, 8, 9] it has been studied by many researchers and has become the most common set of parameters in quantization and encoding algorithms. Some of the more recent developments being reported in [10, 11, 12, 13].

The LSP represents the LPC filter by the unit circle zeros of the two polynomials

$$P(z) = A_p(z) - z^{-1}A_p^\sharp(z); Q(z) = A_p(z) + z^{-1}A_p^\sharp(z) \quad (8)$$

Usually, the two LSP polynomial are viewd as representing two 'snapshots' of the filter at two *separate* and artificial situations neither of which may represent a state along any point on the model. Indeed, the two LSP polynomials in (8) correspond to extending the recursions (1) with a p+1-th reflection coefficient set once to $k_{p+1} = +1$ and once to $k_{p+1} = -1$.

An alternative interpretation may now emerge by viewing the ratio of the two LSP polynomials as an immittance function (3) of a polynomial of degree $p+1$, say $A_{p+1}(z)$, obtained by augmenting $A_p(z)$ with a zero at the origin [4]. First we note that $A_{p+1}(z) = A_p(z)$ (the predictor polynomial is by assumption a polynomial in z^{-1} with monic free term) and $A_{p+1}^\sharp(z) = z^{-1}A_p^\sharp(z)$. Next, it is also clear that $A_{p+1}(z)$ is stable if and only if $A_p(z)$ is stable. Consequently, the LSP theorem [8, 9] follows at once from Theorem 1 applied to this $A_{p+1}(z)$.

Theorem 2 (LSP). *$A_p(z)$ is stable if and only if*

$$L_p(z) := \frac{A_p(z) - z^{-1}A_p^\sharp(z)}{A_p(z) + z^{-1}A_p^\sharp(z)} \quad (9)$$

may be written when $p = 2m-1$ as

$$L_{2m-1}(z) = \frac{(1-z^{-2})\prod_{i=1}^{m-1}(1-2z^{-1}x_{2i}+z^{-2})}{\prod_{i=1}^{m}(1-2z^{-1}x_{2i-1}+z^{-2})}$$

and when $p = 2m$ as (10)

$$L_{2m}(z) = \frac{(1-z^{-1})\prod_{i=1}^{m}(1-2z^{-1}x_{2i}+z^{-2})}{(1+z^{-1})\prod_{i=1}^{m}(1-2z^{-1}x_{2i-1}+z^{-2})}.$$

with p parameters that are real and satisfy

$$-1 < x_p < \dots < x_2 < x_1 < 1 \quad (11)$$

Again $\omega_i := cos^{-1}(x_i)$ bear a frequency meaning and the coded parameters may be the normalized frequencies $f_i = \omega_i/2\pi$. Note that increasing the degree of $A_p(z)$ with a prescribed zero (at $z=0$) adds no information so that the new polynomial is, too, determined by just p parameters. However the modification constrains the 'gain' in (4) to a fixed value ($K = 1$) and instead, a p-th 'frequency' parameter completes the characterization of $A_p(z)$.

The way we arrived to Theorem 2 offers a new modeling interpretation to LSP. Namely, $L_p(z)$ of (9) may be viewed as the immittance at the 'glottis' in the vocal tract model when the termination at the 'lips' is constrained by $k_{p+1} = 0$ [4].

Our experimental study of the properties of ISP parameters for encoding LPC has been motivated by the advantage that ISP has over LSP in being free from any augmentation to the LPC polynomial from a mathematical standpoint, or artificial boundary conditions in the 'vocal tract model'. We were also encouraged by the fact that any quantization gain that these advantages may yield would be without sacrificing any of the properties, like control over stability, ordering and a frequency meaning, that made LSP attractive for coding.

3. Experimental Results

In order to learn on the relative merits of LSP and ISP we ran statistical and quantization experiments under identical conditions with each of the two sets of parameters. The data base was derived from the TIMIT CD-ROM Speech Corpus and included 24 speakers from 8 major dialect regions of the USA (16 male and 8 female). Two different sentences were included for each speaker, summing up to 48 sentences with duration of about 4 minutes. The results reported here are based on uniform scalar quantization with inter-frame differentiation under the further conditions summarized in Table 1.

We chose inter-frame differentiation because LSP parameters are more highly correlated inter-frame wise than intra-frame wise [13]. Fig. 1 and Table 2 describe the distributions of ISP and LSP parameters. The last ISP parameters that is of a different type is depicted separately.

Table 1. Experimental conditions.

data base	4 minutes , 12000 frames.
data	16 bit Linear PCM, 16 kHz sampling, decimated to 8 kHz, no pre-emphasis.
analysis	10*th* order LPC, Burg's algorithm.
framing	20 ms Hamming window, no overlap
update rate	50 frames/sec
distortion	MSE of Log-spectra
quantization	inter-frame differentiation with error feedback

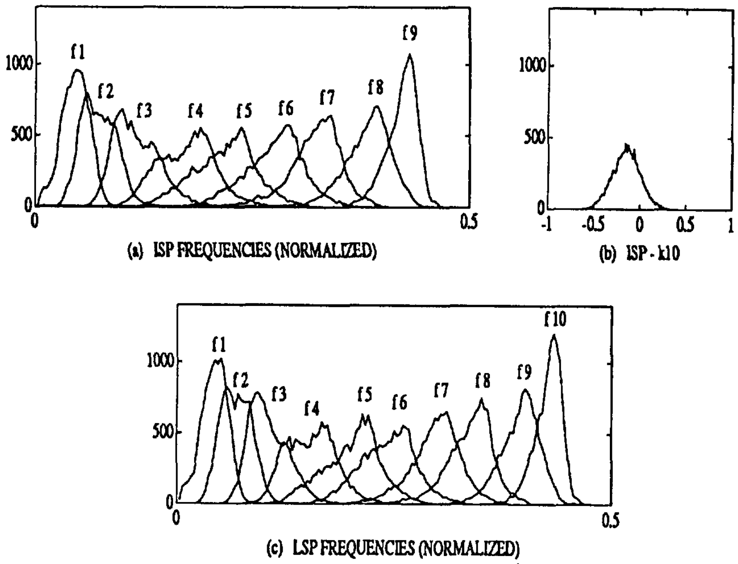

Figure 1: Parameters Distribution

Table 2. Parameters Distribution.

parameter #		1	2	3	4	5	6	7	8	9	10
ISP $\times 10^3$	mean	48	75	118	176	220	277	324	384	423	-141
	std. dev.	15	20	26	32	35	33	29	27	18	153
LSP $\times 10^3$	mean	45	70	107	156	204	244	299	339	396	428
	std. dev.	15	17	22	28	32	31	28	27	23	16

Exhaustive iterations were run to find minimal distortion for bit rates in the interval between 20 bits/ frame and 40 bits/frame. The distortion performance is affected by both bit location and by quantization overload [14]. We carried out an elaborate optimization scheme to take care of the interaction per each parameter between the quantizer's overload points and the number of bits allocated to that parameter. At 20 bits/frame optimal bit allocation was obtained by running iterations over all the possibilities of transferring one bit at a time from each parameter to the others, for a varying quantization intervals, till no further decrease in distortion was possible. From this initial optimal allocation, the optimization at each subsequent bit rate instance involved a bi-dimensional search for the location of the new bit and for new quantization intervals that yield the minimal distortion measure. Fig. 2 compares the distortion per bit-rate of ISP and LSP. ISP precedes consistently LSP by 1 bit roughly at all distortion levels. To be more specific, we examine in more detail the 1dB distorsion level, an accepted perceptual threshold for coding LPC spectral information transparently. ISP is seen to cross the 1dB threshold at 34 bits/ frame, and LSP at 35 bits/frame. The distortion statistics for the two sets of parameters at their respective 1dB threshold are brought in Fig 3 and Table 3. Fig 3 shows the distortion distribution histograms for ISP and LSP. Table 3 reveals that the statistical numbers are slightly in favor of ISP, including lower outliers that is an indication for better speech quality [15]. These experiments, as well as others that we held, indicate that ISP is able to match accepted LSP qualities at lower bit rates.

It should be emphasized that with neither LSP nor ISP we have not tried to determine the best coder or try to hit the lowest bit rate possible. However, the results reported here along with the affinity of the two models may suggest that anywhere that LSP is successful ISP has the potential of performing somewhat better.

Table 3. Spectral Distortion Statistics.

	LSP @ 35 bits	ISP @ 34 bits
mean	0.972	0.989
std. dev.	0.706	0.675
> 3 dB	2.60 %	2.59 %
> 5 dB	0.44 %	0.42 %

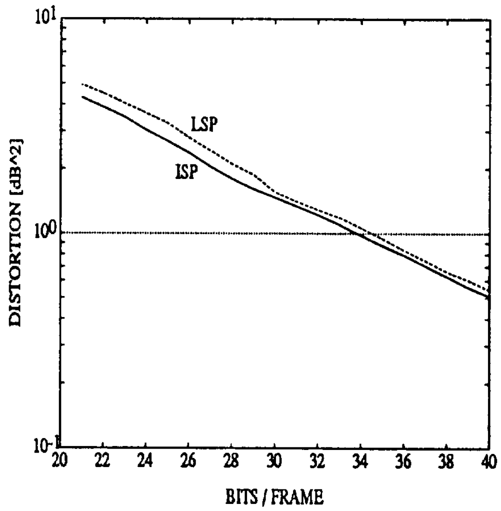

Figure 2: Rate-Distortion Performance

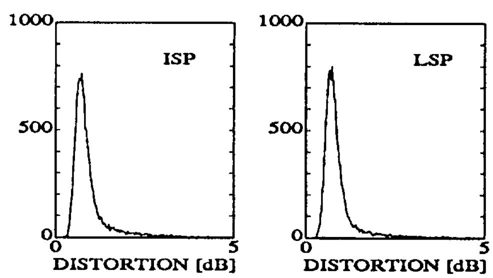

Figure 3: Distortion Distributions

4. Conclusions

The ISP parameters are a new set of parameters proposed for speech encoding. ISP shares with LSP desirable features like stability conditions, ordering and frequency meaning. Our initial experimental study indicates that ISP compares *favorably* with LSP in all the quantization tests that we performed. This may suggest ISP as a substitute for LSP in many LSP coding schemes with the prospect of offering better performance without an increase in computational complexity. In fact, even in computation it has an advantage margin as it needs one less root computation than LSP.

References

[1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice- Hall, Englewood Cliffs, New Jersey, 1978.

[2] Y. Bistritz, H. Lev-Ari and T. Kailath, "Immittance - domain Levinson algorithms", *IEEE Trans. Inform. Theory.*, Vol. IT-35, May 1989. See also *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.* , pp. 253-256, Tokyo, Japan, 1986.

[3] Y. Bistritz, H. Lev-Ari and T. Kailath, "Complexity reduced lattice filters for digital speech processing" *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 21-24, Dallas Texas, 1987.

[4] Y. Bistritz, "Immittance domain modelling for linear predictive coding", in preparation.

[5] Y. Bistritz, "A discrete stability equation theorem and method of stable model reduction", *Systems & Control Letters*, Vol. 1, pp. 373-381, 1982.

[6] F. Itakura, "Line spectrum representation of linear predictive coefficients", *J. Acoust. Soc. AM.* , 535(a), s35(A), 1975.

[7] N. Sugamura and F. Itakura, "Speech data compression by LSP speech analysis and synthesis technique", *IECE Trans.*, Vol. J 64-A, pp. 599-605, 1981.

[8] S. Sagayama, "Stability conditions of LSP speech synthesis digital filters", *Proc. of ASJ*, pp. 153-154, 1982.

[9] F. K. Soong and B-H Juang, "Line Spectrum Pairs (LSP) and speech data compression", *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 1.10.1:4, 1984.

[10] F. K. Soong and B-H Juang, "Optimal quantization of LSP parameters" *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 394-397, 1988.

[11] N. Farvardin and R. Laroia, "Efficient encoding of speech LSP parameters using the discrete cosine transformation", *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 168-171, 1989.

[12] F. K. Soong and B-H Juang, "Optimal quantization of LSP parameters using delayed decisions", *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 185-188, 1990.

[13] C-C Kuo, F-R Jean and H-C Wang, "Low bit-rate quantization of LSP parameters using two-dimensional differential coding", *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. I97-I100, 1992.

[14] N. S. Jayant and P. Noll, *Digital Coding of Waveforms, Principles and applications to speech and video*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984.

[15] B. S. Atal, R. V. Cox and P. Kroon, "Spectral quantization and interpolation for CELP coders" *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, pp. 69-72, 1989.