REFERENCES

- [1] P. Agathoklis, E. I. Jury, and M. Mansour, "The margin of stability of 2-D linear discrete systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 869-873, Dec. 1982.
- [2] E. Walach and E. Zeheb, "N-dimensional stability margins computation and a variable transformation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 887-893, Dec. 1982.
- [3] M. N. S. Swamy, L. M. Roytman, and J. F. Delansky, "Finite word length effect and stability of multidimensional digital filters," *Proc. IEEE*, vol. 69, pp. 1370-1372, Oct. 1981.
- [4] E. Zeheb and E. Walach, "Zero sets of multiparameter functions and stability of multidimensional systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 192-206, Apr. 1981.
- [5] E. Zeheb and D. Hertz, "Another proof and a generalization of a theorem on N-dimensional stability," *Proc. IEEE*, vol. 72, pp. 745-746, June 1984.
- [6] D. Hertz and E. Zeheb, "A generalization of Rudin's multivariable stability theorem," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 725-728, June 1985.
- [7] W. Rudin, *Function Theory in Polydiscs*. New York: Benjamin, 1969, p. 87.

Experimentation with Synthesized Speech Generated from Line-Spectrum Pairs

GEORGE S. KANG AND LAWRENCE J. FRANSEN

Abstract—Line-spectrum pairs (LSP's) are frequency-domain parameters similar to formant frequencies. Thus, they have frequency-selective spectral-error characteristics which allow LSP quantization in accordance with auditory perception. In addition, ease of estimating the spectral-error sensitivity of each line spectrum makes possible encoding each line spectrum efficiently. This correspondence, for the first time, demonstrates that a 31 bit representation of LSP's provides similar intelligibility as a 41 bit representation of reflection coefficients in a current 2400 bit/s LPC. Even with a 12 bit quantization of LSP's, the loss of speech intelligibility is minor, only 2.4 points below that of a 41 bit quantization of reflection coefficients as measured by the diagnostic rhyme test (DRT) which tests initial-consonant discrimination.

I. INTRODUCTION

Prediction and reflection coefficients are the most frequently used parameters for the all-pole filter representation of the speech-spectral envelope. Recently, however, line-spectrum pairs (LSP's) have been gaining interest as alternative parameters because of their intrinsic properties which permit more efficient encoding than the more often used reflection coefficients [1], [2]. It has been said that the LSP's are more efficient parameters than reflection coefficients. So far, there have been no formalized perceptual test scores available from any LSP-based voice synthesizer. This correspondence should fill this gap. We used the diagnostic rhyme test (DRT) to evaluate speech intelligibility because: 1) impartial and trained listeners are used to evaluate speech, 2) test results are repeatable, 3) DRT has proven to be effective in revealing weaknesses and strengths of voice processors under development, and 4) there exist numerous DRT data from other voice processors for comparison.

Manuscript received June 19, 1986; revised October 4, 1986. This work was supported by the Office of Naval Research and managed by Dr. J. Davis of the Naval Research Laboratory.

The authors are with the Naval Research Laboratory, Washington, DC 20375.

IEEE Log Number 8612865.

Fig. 1. Decomposition of the roots of a tenth-order LPC analysis filter with a speech sampling rate of 8 kHz. Each root of the LPC analysis filter (indicated by X) is decomposed into two roots along the unit circle of the complex z plane (indicated by ■ and ●). The related roots are encircled with a faint-line enclosure.

A significant conclusion of this correspondence is that, for the same level of speech intelligibility, encoding of LSP's requires 20 percent fewer bits than reflection coefficients in a pitch-excited, 2400 bit/s speech encoder. Furthermore, speech degradation is gradual as fewer bits are used to represent the LSP's.

II. LINE-SPECTRUM PAIRS

A direct way of defining the LSP is through the decomposition of the impulse response of the LPC analysis filter into even and odd functions. In terms of prediction coefficients, the transfer function of the n th-order LPC analysis filter which transforms speech samples to prediction residual samples is

$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_n z^{-n} \quad (1)$$

$$= \prod_{i=1}^{n/2} (1 - z_i z^{-1}) \quad (2)$$

where a_n is the n th prediction coefficient and z_i is the i th root. A change in each prediction coefficient affects the frequency response of the analysis filter over the entire passband, whereas a change in each root affects the spectrum near that frequency. The roots of the LPC analysis filter have never been used as filter parameters because of difficulties with extracting roots.

Root finding becomes simplified if the LPC analysis filter $A(z)$ is decomposed to a sum of two filters in which each filter has roots along the unit circle of the complex z plane. This can be accomplished by taking a sum and difference between $A(z)$ and its conjugate function [i.e., the transfer function of the filter whose impulse response is a mirror image of $A(z)$]:

$$P(z) = A(z) - z^{-(n+1)} A(z^{-1}) \quad (3)$$

and

$$Q(z) = A(z) + z^{-(n+1)} A(z^{-1}). \quad (4)$$

The LPC analysis filter, reconstructed by the sum of these two filters, is

$$A(z) = \frac{1}{2} [P(z) + Q(z)]. \quad (5)$$

Equation (5) is an equivalent representation of the LPC analysis filter $A(z)$ in which $P(z)$ and $Q(z)$ are component filters. Fig. 1 illustrates how each root of $A(z)$ is decomposed into two roots, each having a unit modulus.

Since roots of both $P(z)$ and $Q(z)$ are along the unit circle of the z plane, roots may be found by searching null frequencies of their amplitude spectra. We use a single complex fast Fourier transform to compute both amplitude spectra at once. A transform size of 512 provides a frequency resolution of 15.625 Hz for the speech bandwidth of 4 kHz. The estimated line spectrum, however, may be refined through a simple parabolic approximation based on the three consecutive spectral points near the null frequency.

Fig. 2. Formant and LSP trajectories of original speech where $F1$, $F2$, $F3$, and $F4$ are the first through fourth formant frequencies and LSP1, LSP2, LSP3, LSP4, and LSP5 are the first through fifth LSP's. (a) Spectrogram, (b) LSP trajectories.

TABLE I
LIST OF FREQUENCIES WITH A 6 PERCENT FREQUENCY RESOLUTION. WE USED A SINGLE-ZERO PREEMPHASIS FILTER, $1 - 0.9375z^{-1}$, PRIOR TO LPC ANALYSIS. IF A PREEMPHASIS FACTOR SMALLER THAN 0.9375 IS USED, THE LOWEST FREQUENCY MUST BE REDUCED

Index	Freq. (Hz)	Index	Freq. (Hz)	Index	Freq. (Hz)	Index	Freq. (Hz)	Index	Freq. (Hz)
1	300	3	400	15	800	27	1600	39	3200
2	350	4	420	16	850	28	1700	40	3390
		5	450	17	900	29	1800	41	3590
		6	480	18	950	30	1900	42	3810
		7	500	19	1010	31	2020		
		8	530	20	1070	32	2140		
		9	570	21	1130	33	2260		
		10	600	22	1200	34	2400		
		11	640	23	1270	35	2540		
		12	670	24	1350	36	2690		
		13	710	25	1430	37	2850		
		14	760	26	1510	38	3020		

Fig. 2 shows formant and LSP trajectories from actual speech samples. The two line spectra stemming from the two real roots [$z = 1$ from $P(z)$ and $z = -1$ from $Q(z)$] are artifacts, and they are not included in this plot. Closely spaced line spectra are located near speech resonant frequencies, whereas sparsely spaced line spectra are located near valleys of the speech spectral envelope. The closely spaced line spectra are spectrally more critical to $A(z)$ [3].

III. EFFECTS OF LOGARITHMICALLY QUANTIZED LSP'S ON DRT

Frequency-related speech parameters have always been quantized logarithmically to encode them more efficiently without compromising the perceptual quality of the synthesized speech. For example, the *pitch frequency* of the narrow-band vocoder is often quantized from 12 to 20 steps per octave (i.e., a 6–3.5 percent frequency resolution). Likewise, *formant frequencies* of a formant vocoder may be encoded more efficiently if a logarithmic scale is used. The just-noticeable differences (JND's) of formant frequencies are on the order of 3–5 percent, although this is greatly dependent on the proximity of the formants to one another [4].

We are interested in the effect of logarithmically quantized LSP's on the DRT score. We tested a 6 percent frequency resolution (i.e., 12 steps per octave). For this case, frequencies from 400 to 3200 Hz (3 octaves) are quantized to $3(12) + 1 = 37$ values including both end frequencies (Table I). Two additional frequencies below 400 Hz and three additional frequencies above 3200 Hz cover the entire frequency range of interest. Thus, there are 42 frequencies. The two additional frequencies below 400 Hz are 350 and 300 Hz. The lower end is quantized rather coarsely because these frequencies are used for producing murmur or nasal sounds which are not too critical to speech intelligibility.

The number of combinations for choosing 10 frequencies out of 42 possible frequencies is $1\,471\,442\,974 \approx 2^{31}$, and the resultant DRT score is 88.0 (Fig. 3). Even if line spectra are quantized in-

Fig. 3. DRT scores from three male speakers when LSP's are quantized logarithmically with a 6 percent frequency resolution. The bits referred to are the number of bits per frame used to encode filter parameters. Frame rate is 44.44 Hz. The excitation signal is identical to that used in the 2400-bit/s LPC. LSP's are more efficient parameters because they save as many as eight bits (20 percent) when compared to a typical 2400 bit/s LPC using reflection coefficients.

dependently, a similar level of speech intelligibility can be attained without requiring many additional bits. For example, if the first through tenth line spectra are independently quantized to 4, 4, 4, 4, 4, 3, 3, 3, 2, and 2 bits, respectively (i.e., 33 bits), the resultant DRT score is 88.4 (Fig. 3). This score is identical to that of a 2400 bit/s LPC that uses reflection coefficients quantized to 41 bits per frame (Fig. 3).

Since fewer bits with LSP are required over the conventional

LPC (operating at 2400 bits/s) for the same intelligibility, a useful application of LSP would be for improving the speech quality at 2400 bits/s by making the best use of the saved bits. We will not discuss this aspect because 2400 bit/s improvement probably would involve the excitation signal which is beyond the scope of this correspondence.

IV. EFFECTS OF ELIMINATING HIGHER INDEXED LSP'S ON DRT

Encoding a fewer number of filter parameters is beneficial for both speech transmission and speech synthesis by rule. Therefore, we investigated the effect of using a reduced number of LSP's on speech intelligibility. Most formant vocoders do not transmit any information on the fourth and fifth formant frequencies because: 1) they are sometimes absent in speech, 2) they are difficult to track, and 3) speech intelligibility is adequate without using them (although the presence of higher formant frequencies is decidedly better). Higher indexed LSP's describe the speech-spectral envelope in the high-frequency region, similar to higher formant frequencies. We would like to know the DRT score sensitivity when higher indexed line spectra are not transmitted.

Unlike formant frequencies, however, missing indexed line spectra can be approximated at the receiver to create a full-band speech sound. This is possible because there are always a fixed number of line spectra in any speech at any time, and they are naturally ordered. The eliminated line spectra may be reintroduced by assuming values determined by their average statistics or even simply at equal spacing between the highest line spectrum transmitted and the upper-cutoff frequency.

Synthesized speech is highly natural even though two or three upper line spectra are not transmitted, but the DRT shows some degradation. When the ninth and tenth line spectra are eliminated (but artificially reintroduced during synthesis), there is a reduction of 3.3 points in the score. When the upper four line spectra are not transmitted, there is an additional 1.5 point drop in the score.

A superior approach for reducing the number of filter parameters is to eliminate some of the difference frequencies of LSP pairs. In this approach, each LSP is represented by its *center frequency* and *difference frequency*. Again, the eliminated LSP difference frequency may be reintroduced during speech synthesis based on their respective mean values. Fig. 4 shows DRT scores as higher LSP difference frequencies are eliminated. Note that the two highest difference frequencies may be omitted from encoding without degrading speech intelligibility. Thus, a tenth-order LPC system can have only eight filter parameters.

V. EFFECTS OF VECTOR QUANTIZATION OF LSP'S ON DRT

According to (1), spectral-error sensitivities of all LPC coefficients are coupled. This is also true for all LSP members. Therefore, it is not efficient to quantize individual LSP's based on some fixed resolution because insensitive LSP's (widely separated LSP's) may be subjected to an unnecessarily fine quantization. It is more efficient to quantize LSP's vectorially [5]. Since there is a *frequency tolerance* around each line-spectral value within which there is no perceivable sound change, many LSP sets can be represented by one LSP set. The magnitude of frequency tolerance is dependent on two major factors: hearing sensitivity to frequency differences and spectral-error sensitivity of the LSP.

Hearing Sensitivity to Frequency Differences

Because the ear cannot resolve differences at high frequencies as accurately as at low frequencies, we may quantize higher frequency LSP's more coarsely than lower ones without introducing audible speech degradation. It is well known that the amount of frequency variation that produces a JND is approximately linear from 0.1 to 1 kHz, and increases logarithmically from 1 to 10 kHz [6].

Fig. 4. DRT scores obtained from the elimination of higher indexed LSP difference frequencies. The excitation signal is identical to that used in the 2400 bit/s LPC. The scores are based on one male speaker contained in the previous DRT scores.

Spectral-Error Sensitivity of the LSP

When each line spectrum is perturbed, there is a corresponding spectral error in $A(z)$. The spectral-error sensitivity is a factor relating error in each line spectrum (in Hz) and the average spectral error of $A(z)$ (in dB). To derive such an expression is not only untractable, but also a cross coupling of all line-spectrum errors into the overall spectral error makes the use of such an expression impractical. Therefore, we derived numerically a relationship which relates the average spectral error of $A(z)$ to all the line-spectrum errors (hence including the effect of cross couplings) from various speech samples. No approximation is involved in computing the average spectral error of $A(z)$ from given line-spectrum errors. We, however, imposed one condition that each line spectrum have an error proportional to the frequency separation to its closest neighbor. Fig. 5 is a resultant scatter plot. In our judgment, a 2 dB average spectral error is as much as we can tolerate. Thus, the allowable frequency tolerance of each line spectrum as obtained from Fig. 5 is approximately 20 percent of the frequency separation to its closest neighbor.

Combining the effect of the hearing sensitivity to frequency difference and the allowable LSP errors, we have allowable frequency tolerance of approximately 20, 30, and 40 percent of the frequency separation to closest neighbor for line spectra located below 1 kHz, between 1 and 2 kHz, and above 2 kHz, respectively. To verify this, we listened to many synthesized speech samples while perturbing each line spectrum by this amount. Indeed, we began to notice some speech quality degradation when the perturbation exceeded the above-mentioned tolerances.

We collected 4096 LSP templates (i.e., a 12 bit representation of filter parameters) from speech samples from 54 male speakers and 12 females uttering 5 sentences each. Initially, we stored the first LSP as a reference template. Subsequently, we compared each new LSP set to all of the stored reference LSP templates. If any member of the new LSP set falls outside the respective allowable frequency tolerance, then the new LSP becomes another reference template. During template collection, the number of LSP sets falling into each template were counted. At the end, the templates representing the fewest sets were eliminated to reduce the total number of templates to 4096. Fig. 6 shows DRT scores of synthesized speech in which ten line spectra are jointly encoded to 12 bits.

VI. CONCLUSIONS

We have shown that a 31 bit representation of LSP's is equivalent to a 41 bit representation of reflection coefficients when speech

Fig. 5. Scatter plot of average spectral error caused by an error in each line spectrum. To obtain one dot in this figure, we derived ten line spectra from speech samples once per frame (22.5 ms). We introduced an error in each line spectrum with the magnitude of error proportional to the frequency separation to its closest neighbor (i.e., error is larger for sparsely spaced line spectra and smaller for closely spaced line spectra). For this scatter plot, the percentage LSP error is varied from frame to frame from 0 to 25 percent.

Fig. 6. Three male DRT scores for 800 and 2400 bit/s vocoders are presented. The 800 bit/s vocoder has 12 bits of information for the line spectra per frame. For comparison, a similar score for the 2400 bit/s LPC is shown. Both vocoders use identical sets of unquantized reflection coefficients, voicing decision, and speech rms values as input data. Previously reported LPC-based 800 bit/s vocoders have DRT scores in the lower 80's.

intelligibility is compared by DRT which evaluates speech in terms of initial-consonant discrimination. Most significantly, the speech degradation is gradual when fewer bits are used to represent LSP's.

A useful application of the LSP's is with voice encoders operating at very low bit rates. Another application, heretofore not widely recognized, is with speech synthesis by rule where LSP's could assume the role of formant frequencies which have been used extensively.

REFERENCES

- [1] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, suppl. 1, p. S35, 1975.

- [2] F. K. Soong and B. Juang, "Line-spectrum pair (LSP) and speech data compression," in *IEEE ICASSP'84*, CH 1945-5, 1984, pp. 1.10.1-1.10.4.
- [3] G. S. Kang and L. J. Fransen, "Application of line-spectrum pairs to low-bit-rate speech encoder," in *IEEE ICASSP'85*, 1985, pp. 7.3.1-7.3.4.
- [4] J. L. Flanagan, *Speech Analysis, Synthesis and Perception*. New York: Springer-Verlag, 1972.
- [5] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proc. IEEE*, vol. 73, pp. 1551-1588, Nov. 1985.
- [6] P. Ladeford, *Elements of Acoustic Phonetics*. Chicago and London: Univ. Chicago Press, 1974.

A Computationally Efficient Approach to the Removal of Impulse Noise from Digitized Speech

AMLAN KUNDU AND SANJIT K. MITRA

Abstract—A new two-step algorithm for removing impulse noise from speech data is outlined. In the first step, a threshold detection scheme is used to determine whether or not the speech sample in question is corrupted by an impulse. The threshold is derived using a statistical theory. In the second step, when the sample has been determined to be noise corrupted, the noisy sample is replaced by a sample generated using a simple least square interpolation scheme. Otherwise, no filtering is done if the sample is determined to be noise free. The performance of the proposed algorithm is shown to be comparable to the more well-known DDC scheme in reducing the noise.

I. INTRODUCTION

In digital speech communication, transmission errors generally introduce impulsive distortions in the received speech waveform. Impulse noise does not corrupt each data point. But at the point where it corrupts the signal, the signal value is suppressed by the high value of the impulse noise. Straightforward smoothing of such a noise-corrupted waveform results in a squelching of the distortion component as well as an undesirable smearing of the speech. In the case of DPCM or PCM codes, for error rates on the order of 2.5 in 100 or more, Jayant [1] proposed a linear smoothing scheme based on running averages and a nonlinear smoothing scheme based on running median to reduce the smearing effect.

Ideally, we would like to filter only those points in the data which are corrupted by impulse noise and leave the uncorrupted data points as they are. This necessitates a technique for thresholding [2]–[4]. If the thresholding procedure yields a positive decision, the data point is declared to be noise corrupted and is subsequently filtered. Two schemes for statistical hypothesis testing using thresholding procedure are considered in [2] and [3]. Our proposed hypothesis testing procedure is different from those considered in [2] and [3]. Our filtering scheme is a linear interpolation scheme whose performance is comparable to that of the median filter [4].

Manuscript received November 26, 1984; revised October 1, 1986. This work was supported by the National Science Foundation under Grant ECS 85-08017.

A. Kundu is with the Department of Electrical Engineering, State University of New York, Amherst Campus, Buffalo, NY 14260.

S. K. Mitra is with the Signal Processing Laboratory, Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.

IEEE Log Number 8612879.