Linear prediction on a warped frequency scale

Hans Werner Strube

Drittes Physikalisches Institut, Universität Göttingen, Bürgerstrasse 42-44, D-3400 Göttingen, Federal Republic of Germany

(Received 29 April 1979; accepted for publication 28 April 1980)

Linear prediction is considered with respect to a nonlinear frequency scale obtained by a first-order all-pass transformation. The predictor can be computed from a frequency-warped autocorrelation function obtained from the power spectrum or by a direct linear transformation of the original acf. Three numerical procedures are compared. Alternatively, the predictor can be determined from a covariance matrix or (adaptively) from continuously formed correlations, suitably defined according to the all-pass transformation. Prediction-error minimization and spectral flattening are no longer equivalent criteria. In the synthesis part of a vocoder or APC system, no inverse transformation is required, since the direct form of the analysis and synthesis filters can be modified so as to immediately realize the warped transfer function. Single-word intelligibility is compared for a predictive vocoder on a "Bark" scale and a linear frequency scale. The Bark scale yields results around 90% even at predictor orders of 5 to 7. More possible applications have been given previously by other authors.

PACS numbers: 43.70.Gr, 43.70.Qa, 43.60.Cg

I. INTRODUCTION

The idea of linear prediction with respect to a warped frequency scale is not new. Makhoul and Cosell (1976) have described its usefulness for vocoding since the frequency resolution of the human ear at high frequencies is less sharp than at low ones, also, formant bandwidth increases with frequency. When warping is applied only either in the analysis or in the synthesis part, the spectral envelope can be arbitrarily distorted but the fine structure left unchanged (Makhoul 1976), for instance, in order to unscramble helium speech. Another application has been given by Itahashi and Yokoyama (1978), who traced predictor-derived formants on the subjective mel scale. Recently, Stålhammar (1978) found a connection with his G_2' concept. In all these applications, the frequency-warped predictor is computed from an autocorrelation function obtained by Fourier transforming a warped power spectrum; in the synthesis part of a vocoder, the inverse transformation has to be executed. Here another algorithm will be presented which avoids the resampling of a power spectrum or the cosine-series summation of nonequidistant spectral samples, it differs from the ordinary autocorrelation method of linear prediction only by an additional linear transformation. Alternatively, the "covariance method" or adaptive correlation methods can be modified according to the warped frequency scale. Further, a synthesis filter can be constructed directly with respect to the warped frequency scale, its coefficients are obtained from the predictor coefficients by another linear transformation. This method is not applicable to general warping characteristics but only to first-order all-pass transformations, but this will be sufficient in many cases. The all-pass transformation considered is of the form

$$\tilde{z}^{-1}=(z^{-1}-a)/(1-az^{-1}),\ -1<a<1\,, \tag{1}$$

$$\tilde{\omega}=\omega+2\tan^{-1}[(a\sin\omega)/(1-a\cos\omega)]\,, \tag{2}$$

$$d\tilde{\omega}/d\omega=(1-a^2)/(1+a^2-2a\cos\omega)\,, \tag{3}$$

where $z=e^{i\omega}$, $\tilde{z}=e^{i\tilde{\omega}}$. The inverse transformation is simply given by replacing a with $-a$. As an example, for a sampling frequency of 10 kHz, $\tilde{\omega}$ is a very good approximation to the subjective Bark scale based on the critical bands of the ear (Zwicker and Feldtkeller, 1967) (similar to the mel scale) if $a=0.47$.

II. THE WARPED AUTOCORRELATION METHOD

If s_t is any sequence where t (integer) denotes time (e. g., a sampled time function, an autocorrelation function, numerator or denominator coefficients of a filter), a corresponding frequency-warped sequence $\tilde{s}_t$ is defined by

$$\sum_t \tilde{s}_t\tilde{z}^{-t}(z)=\sum_t s_t z^{-t}\,, \tag{4}$$

where $\tilde{z}^{-t}(z)$ is the tth power of an all-pass transfer function $\tilde{z}^{-1}(z)$ [e. g., as given in Eq. (1)] such that $e^{i\tilde{\omega}}=\tilde{z}(e^{i\omega})$. s_t and $\tilde{s}_t$ are connected by a linear transformation (multiplication with a fixed matrix, see Sec. IV) which, however, is not shift-invariant. The convolution of two sequences is transformed into the convolution of the transformed sequences. All this would hold for any frequency-warping transformation; the special transformation (1), (2) can, according to Oppenheim and Johnson (1972), also be done in a recursive way, corresponding to a cascade of filters whose outputs at the time $t=0$ are the transformed sequence:

$$\begin{aligned}
&g_0^{(k)}=s_{-k}+ag_0^{(k-1)}\,,\\
&g_1^{(k)}=(1-a^2)g_0^{(k-1)}+ag_1^{(k-1)}\,,\\
&g_n^{(k)}=g_{n-1}^{(k-1)}+a(g_n^{(k-1)}-g_{n-1}^{(k)}),\quad n=2,3,\ldots;\\
&\qquad\qquad k=\ldots,-2,-1,0\,.\\
&\tilde{s}_t=g_t^{(0)},\quad t=0,1,2,\ldots
\end{aligned} \tag{5}$$

This transformation holds for a causal sequence s_t; if s_t does not vanish for all $t<0$, the anticausal part has to be transformed in the same way where in (5), s_{-k} is replaced by s_k and $\tilde{s}_t$ by $\tilde{s}_{-t}$. The term s_0 can be split up arbitrarily, $\tilde{s}_0$ is then the sum of $\tilde{s}_0$ (causal) and $\tilde{s}_0$

 0001-4966/80/101071-06$00.80 © 1980 Acoustical Society of America

Downloaded 10 May 2010 to 200.39.9.109. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms

(anticausal). Thus, if s_t is even (e. g., an autocorrelation function), (5) can be applied to the causal sequence $s_0/2$, s_1, $s_2, \ldots$, and in the result, $\tilde{s}_0$ has to be doubled again.

In practice, only a finite-length sequence can be treated exactly but is usually transformed into an infinite one. Therefore the transformation should not, as could be done in principle, be directly applied to the sequence of coefficients of the prediction-error filter, since the result would be a predictor of infinite order. For the same reason, transformation of a windowed signal segment before forming the acf in the autocorrelation method of linear prediction is not recommended, rather, the (finite-length) acf should be transformed. The number of original acf samples R_k to be computed can be kept low by using a lag window. Such a window also prevents the spectral resolution from being increased so much in a certain frequency range that single harmonics appear as spectral poles; further the lag window alleviates undesirable signal-windowing effects in pitch-asynchronous analysis (Tohkura *et al.*, 1978). If R_k is nonzero for $|k| \leq N$ and the desired order of the predictor is p, only $\tilde{R}_0$ to $\tilde{R}_p$ of the warped acf need be computed and the number of operations is proportional to $(N+1)(p+1)$. Whether the transformation is done by multiplication with a prestored matrix (which may include the lag-window effect) or by the recursion (5) depends on whether storage or computing time is more costly. Also the known warping methods employing the power spectrum may still be advantageously used in many cases, see Sec. IV.

From $\tilde{R}_k$, predictor coefficients $\tilde{a}_k$ or PARCOR coefficients can be obtained in the usual way (Markel and Gray, 1976). The prediction-error (inverse) filter is of the form

$$\tilde{A}(z) = \sum_{k=0}^{p} \tilde{a}_k \tilde{z}^{-k}(z), \quad \tilde{a}_0 = 1; \tag{6}$$

it has not only p zeros but also a p-fold pole at $z = a$. It can be realized in the direct form with each unit-delay element replaced by an all-pass $\tilde{z}^{-1}(z)$ according to (1), Fig. 1. The corresponding synthesis filter is $\tilde{A}^{-1}(z)$. This however, cannot be implemented in the usual way as a recursive filter with $-(\tilde{A} - \tilde{a}_0)$ in the feedback loop, since the open-loop gain would then contain a lag-free term, as can be seen from the decomposition

$$\tilde{z}^{-1} = (1 - a^2)z^{-1}/(1 - az^{-1}) - a. \tag{7}$$

Inserting (7) into (6), the filter can be transformed into a polynomial in $z^{-1}/(1 - az^{-1})$, so that all the terms with $k \neq 0$ are delayed:

$$\tilde{A}(z) = \sum_{k=0}^{p} b_k z^{-k}(1 - az^{-1})^{-k}. \tag{8}$$

The new coefficients b_k are obtained from the $\tilde{a}_k$ by a linear transformation, namely, a multiplication with a fixed triangular matrix obtained from the binomial formulas:

$$b_k = \sum_{n=k}^{p} C_{kn}\tilde{a}_n, \quad C_{kn} = \binom{n}{k}(1 - a^2)^k (-a)^{n-k}. \tag{9}$$

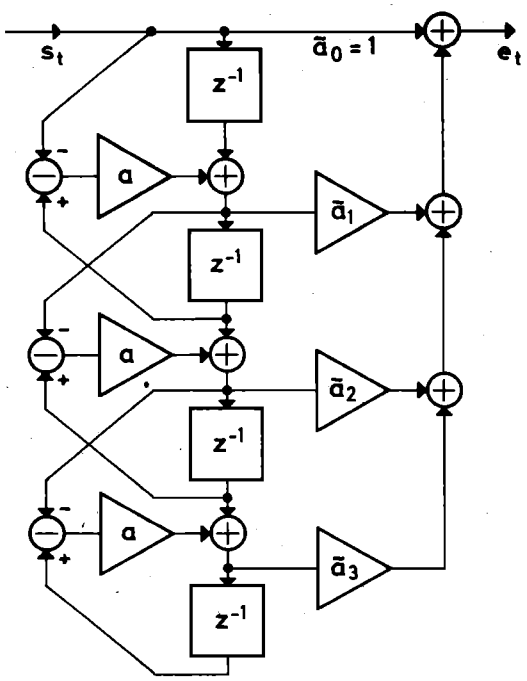

FIG. 1. Direct form of prediction-error filter (order $p = 3$) with unit delays replaced by all-passes $\tilde{z}^{-1}(z)$ to achieve frequency warping.

Again, a recursive algorithm can be used instead of this matrix multiplication:

$$b_p := \tilde{a}_p;$$

$$b_{p-n} := \tilde{a}_{p-n} - ab_{p-n+1},$$

$$\text{if } n > 1\text{: } b_k := (1 - a^2)b_k - ab_{k+1}, \quad k = p - n + 1, \ldots, p - 1,$$

$$b_p := (1 - a^2)b_p, \qquad n = 1, \ldots, p. \tag{10}$$

This recursion can even be done in place. Whether (9) or (10) is preferred will again depend on memory versus

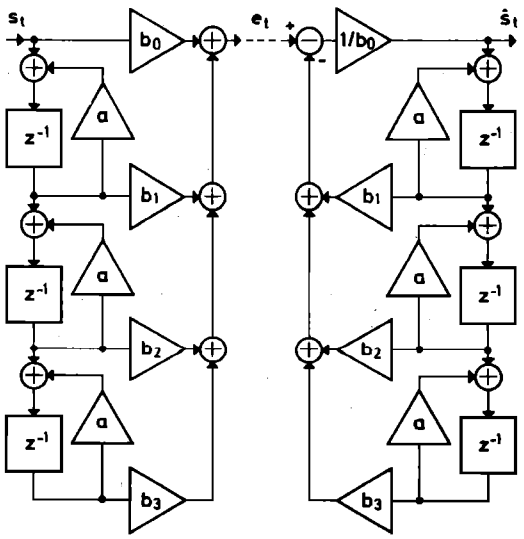

FIG. 2. Modified direct form of analysis (prediction error) and synthesis filter, avoiding algebraic (zero-lag) loops in the synthesis filter. Coefficients b_k are a linear transform of the predictor coefficients, normalization (b_k/b_0) is possible.

wnloaded 10 May 2010 to 200.39.9.109. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms

time saving. The filter structures of $\tilde{A}(z)$ and $\tilde{A}^{-1}(z)$ based on (8) are shown in Fig. 2. Of course, normalized coefficients b_k/b_0 may also be used, omitting the b_0 and b_0^{-1} amplifiers in Fig. 2; then the power of the transmitted error signal (or, in a vocoder system, of the artificial excitation signal) is b_0^{-2} times the prediction-error power.

III. SPECTRAL INTERPRETATION; COVARIANCE AND ADAPTIVE METHODS

Let $P(\omega)$, $E(\omega)$ be the power spectra of the signal s_t and the prediction error, respectively, where the prediction error e_t is obtained by filtering s_t with $\tilde{A}(z)$, so that

$$E(\omega) = |\tilde{A}(e^{i\omega})|^2 P(\omega) . \tag{11}$$

As an example, Fig. 3 shows P, $|\tilde{A}|^{-2}$ and E for a vowel /i/ with prediction orders $p=5$ and 13 and $a=0.47$. Then, since the predictor is computed from $\tilde{R}_k$, the total prediction error power formally defined on the *warped* frequency axis is minimized:

$$\sigma^2 \triangleq \int_{-\pi}^{\pi} E(\omega(\tilde{\omega}))\, d\tilde{\omega} = \int_{-\pi}^{\pi} E(\omega) \frac{1-a^2}{1+a^2-2a\cos\omega} d\omega . \tag{12}$$

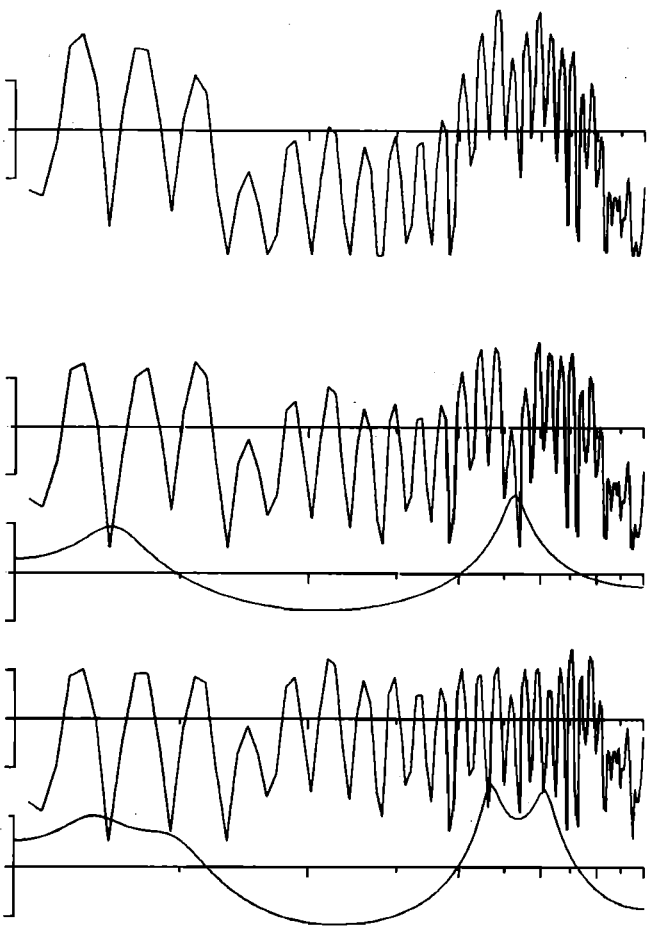

FIG. 3. Logarithmic power spectra, from top to bottom: differentiated signal (vowel /i/); prediction error and synthesis filter, $p=5$; prediction error and synthesis filter, $p=13$. Warping parameter is $a=0.47$. Abscissa: kHz; ordinate: unit = 10 dB.

As usual (Markel and Gray, 1976), $E(\omega(\tilde{\omega}))$ is maximally flat; if E is constant, $\tilde{\sigma}^2$ is equal to the actual error power $\sigma^2 = \Sigma e_t^2 = \int_{-\pi}^{\pi} E(\omega) d\omega$. The terms $\tilde{\sigma}$ and σ are close to each other for sufficiently large prediction order p. However, the predictor as defined here is *not* the solution to the minimization problem of σ^2, but rather, as can be seen from the factor $(1-a^2)/(1+a^2-2a\cos\omega)$ in (12), the power of the error signal filtered by

$$W(z) \triangleq (1-a^2)^{1/2}/(1-az^{-1}) \tag{13}$$

is minimized; $\tilde{\sigma}^2$ is equal to this power.

At first glance, this result might suggest that the effect of frequency-warped prediction is essentially the same as that of pre-emphasis with $W(z)$. However, this is not true, instead it means that spectral flatness and minimum power of the error are no longer equivalent criteria. An error minimizing method would not yield an approximately flat error spectrum but one close to $|W(e^{i\omega})|^2 \equiv d\tilde{\omega}/d\omega$. To prove this, write the signal to be analyzed formally as Ws (meant as W operating on signal s) resulting in the error $\tilde{A}Ws = We$, where $\tilde{A}$ is as above and e has a flat spectrum.

This result has consequences if, instead of the "autocorrelation method," other methods are used for computing predictor or PARCOR coefficients that are based on direct error power minimization. This can be done by the "covariance method" or by networks containing continuously averaging correlators, for instance, updating the predictor coefficients by terms proportional to the cross correlations between delayed signal and error (gradient method), or with Itakura's PARCOR lattice (Markel and Gray, 1976). In principle, such structures have the usual form with unit delays replaced by $\tilde{z}^{-1}(z)$, e. g., in the covariance method the covariances are given by

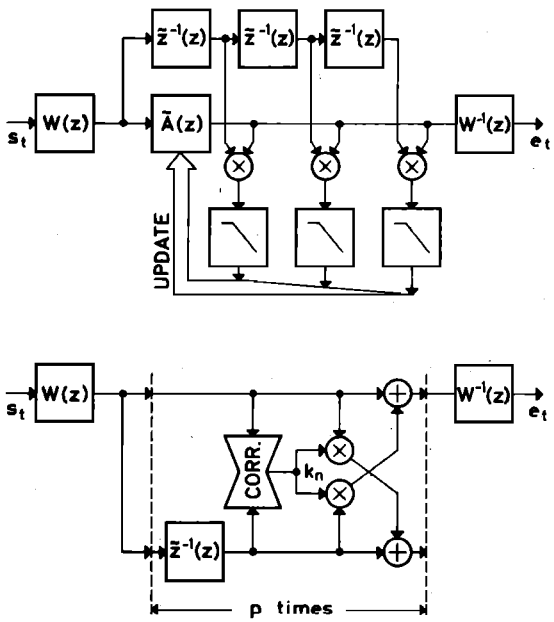

FIG. 4. Two adaptive predictive analyzers containing continuously averaging correlators (top: gradient method ($p=3$), bottom: PARCOR method), modified according to a warped frequency scale $\tilde{\omega}(\omega)$. $|W(e^{i\omega})|^2 = d\tilde{\omega}/d\omega$.

nloaded 10 May 2010 to 200.39.9.109. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/term

$$\phi_{ij}=\sum_t (\tilde{z}^{-i}s)_t(\tilde{z}^{-j}s)_t\,; \qquad (14)$$

however, since the actual error power is minimized, the input signal has to be prefiltered by $W(z)$ from (13) in order to obtain the same $\tilde{A}(z)$ as above. If the (maximally flat) prediction error of the original signal is required, the error signal must be postfiltered by $W^{-1}(z)$ again. Figure 4 shows two such adaptive analyzers. The effects of the finite averaging time constant of the correlators seem to be unimportant, since it is usually much longer than the decay time of the impulse responses of the all-passes $\tilde{z}^{-k}(z)$ and of $W(z)$.

IV. TESTS

A. Comparison of three warping methods

A comparison was carried out concerning the computational and storage requirements of the following three warping methods, all applied to a windowed signal segment of length $L=256$ samples. Typical values for predictor length p and (one-sided) lag-window length N are 10 and 42, respectively. This N value gives a bandwidth of some 230 Hz (for a Hann window at 10 kHz sampling frequency), sufficient to "smear" single harmonics of voiced male speech.

Method 1. From the signal segment, the power spectrum $P_n, n=0,\ldots,L/2$, is computed via the FFT, where P_n belongs to the original frequency $\omega_n=2\pi n/L$, which is warped into the frequency $\tilde{\omega}_n$ according to (2). Then the acf $\tilde{R}_k$ is obtained by matrix multiplication (since $p \ll L$, computation of $\tilde{R}_k$ as FFT of an interpolated and resampled spectrum is not advantageous):

$$\tilde{R}_k=\sum_{n=0}^{L/2} U_{kn}P_n, k=0,\cdots,p; \quad U_{kn}=(d\tilde{\omega}/d\omega)_n\cos(k\tilde{\omega}_n)\,.$$

If the effect of a lag window is to be simulated, the original power spectrum has to be convolved with the Fourier transform of this window, which can simply be achieved by a more complicated matrix U_{kn} without increasing computation time, if the matrix is prestored.

Method 2. From the signal segment, the acf R_t is computed for $t=0,\cdots,N$. Then $\tilde{R}_k$ is obtained by matrix multiplication (the lag window w_t is included in the matrix):

$$\tilde{R}_k=w_0R_0\delta_{k0}+\sum_{t=1}^{N} V_{kt}R_t, \quad k=0,\cdots,p;$$

$k=0$: $V_{0t}=2w_ta^t$ (the factor 2 is due to the even symmetry of R_t and $\tilde{R}_k$),

$$k>0: \quad V_{kt}=w_t\sum_{n=1}^{m_{kt}}\binom{t}{n}\binom{k-1}{n-1}(1-a^2)^n a^{t-n}(-a)^{k-n},$$

$m_{kt}\triangleq \min(k,t)$.

This matrix can be computed by applying the Oppenheim recursion (5) to unit impulses δ_{kt}.

Method 3. From the signal segment, the acf R_t is computed for $t=0,\cdots,N$, weighted by the lag window w_t and transformed by the Oppenheim recursion.

The amount of storage required is largest for method 1 due to the matrix U_{kn}, which has $(L/2+1)(p+1)$ elements, for $L=256$ and $p=10$, this is 1419. The matrix V_{kt} of method 2 has only $N(p+1)$ elements, for $N=42$, this is 462. Method 3 needs no matrix at all, only an array of length N or $N+1$ for the lag window. On the other hand, computation time is least for method 1, method 2 is 17% and method 3, 24% slower for the above values of L, N, and p. For larger p-values, the difference in speed between method 1 and the others decreases; Fig. 5 shows the dependence of computation time on p. The time relations shown will of course, depend on the relative effectiveness of the subroutines for FFT, autocorrelation, matrix multiplication, and Oppenheim recursion. In our case, all were written efficiently in a combination of some Fortran and much machine language. The acf computation is the slowest part in methods 2 and 3; its duration is also roughly proportional to the length of the lag window N, whereas method 1 is independent of N.

It can be concluded that, for small predictor length p or long lag window, warping via the power spectrum is advantageous, provided there is enough storage for the large matrix U_{kn} available. For large p and short lag window, however, the advantage of this method diminishes and the direct transformation of the acf will be preferred because of its small storage requirements, especially in its recursive form.

B. Perceptual comparisons

In order to test the advantage (if any) of the frequency-warped prediction over the usual one and to determine the useful range of the filter order p, intelligibility and quality comparisons must be carried out. Only the Bark-scale case ($a=0.47$) will be considered here. Intelligibility was measured for the following vocoder

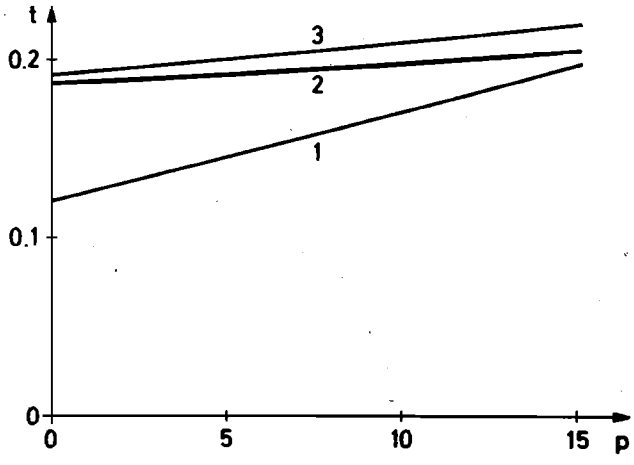

FIG. 5. Dependence of computation time (in seconds for Honeywell H 632) on predictor order p for obtaining a frequency-warped autocorrelation function $\tilde{R}_0, \cdots, \tilde{R}_p$. 1: matrix multiplication of power spectrum; 2: matrix multiplication of acf; 3: recursive Oppenheim transformation of acf. In 2 and 3, the lags of the original (lag windowed) acf ranged from 0 to $N=42$; the time is roughly proportional to N. The frame size is 256 samples.

wnloaded 10 May 2010 to 200.39.9.109. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms

system. The analog speech signal was differentiated once by an RC circuit for pre-emphasis and low-pass filtered at 4.8 kHz (16 poles, Butterworth), then A/D-converted with 12 bits resolution at a sampling frequency of 10 kHz. Analysis frames were 250 samples long and Hamming-windowed, the frame rate was 100 per second. For each frame, the acf was formed and lag windowed ($N=41$), the lag window being realized by every third sample of the (halved) above Hamming-window. For the linear frequency scale ($a=0$), the predictor coefficients and the excitation power were obtained directly from the weighted acf, whereas for the Bark scale, the acf was first transformed by the Oppenheim algorithm. For synthesis, the filter of Fig. 2 (right) was applied, followed by a de-emphasis filter $1/(1-0.92z^{-1})$ and excited by a pulse and noise generator. After D/A-conversion, the signal was low-pass filtered at 5 kHz with an 8-pole Butterworth filter and recorded on tape. Presentation to the subjects was by headphones (Sennheiser HD 424 and 414) in an ordinary room. The speech material consisted of 240 monosyllabic German nouns, separated by pauses of 4–5 s, from groups 1–12 of the "Freiburger" word list (DIN 45621) available on a standardized tape. The filter order p was in the range 4–13. Each of the 20 pairs of (p,a) values occurred once in each group of 20 words in a pseudorandom order which was different for each of the twelve groups. Since each (p,a) would then be represented by a different set of twelve words, the test had to be symmetrized with respect to a by repeating it with $a=0$ and $a=0.47$ interchanged. This was done some days later with a permuted group order. In spite of this, learning effects reduced the average error rate by 60%. Therefore a second test pair was conducted in the opposite order. The number of subjects was eight (four in each test pair), so that each (p,a) condition (represented by 24 different words) was tested by 192 presentations totally. The results are shown in Fig. 6. Our emphasis is on the comparison between the two a-values, not so much on the p dependence, which has irregularities due to the different speech material for each p value (apart from a possibly true nonmonotonicity of the p dependence), but this is the same for both a values. The results show that both sorts of prediction are equally good at prediction orders $p \geq 8$ but Bark-scale prediction is clearly advantageous at $p \leq 7$; even with five coefficients, intelligibility is fair. This can be understood by the fact that the first and an effective second formant are still well represented by such short predictors on the Bark scale (cf. Fig. 3, $p=5$), whereas on the frequency scale, the connection between filter poles and formants becomes irregular.

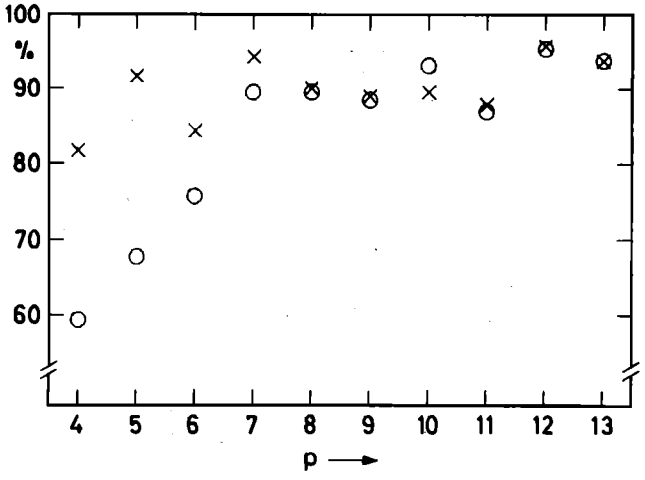

FIG. 6. Intelligibility of monosyllabic German nouns processed by a predictive vocoder ($f_s=10$ kHz, 100 frames/s, autocorrelation method). ○: linear frequency scale ($a=0$), ×: Bark scale ($a=0.47$). Abscissa: predictor order p.

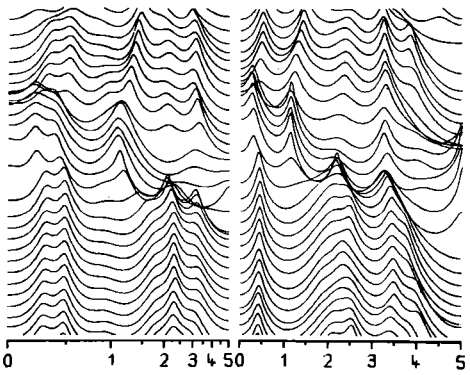

FIG. 7. Predictor spectra for /e:bə/. $p=13$; $a=0$ (right) and 0.47 (left). Time runs from bottom to top, step 10 ms. Note the too high resolution (splitting) of low formants with $a=0.47$.

Another question is speech quality rather than intelligibility. Our quality tests have only been informal, the results indicate the same clear advantage of Bark-scale prediction at low p values; whereas its quality is slightly inferior to frequency-scale prediction at large p values. This may be attributed to the better fit of the frequency-scale prediction to the approximate all-pole structure of speech spectra, whereas on the Bark scale, some poles are wasted by a splitting of the low formants, perhaps due to the harmonic structure (cf. Fig. 3, $p=13$, and Fig. 7). As an example, Fig. 7 shows a segment /e:bə/ on both scales for $p=13$.

So far, no investigations in optimizing the warping parameter a and the pre-emphasis have been carried out. Thus, still better results may be obtained.

V. CONCLUSION

Several methods for modification of the "autocorrelation method" of linear prediction have been considered so that the underlying frequency scale becomes nonlinear in a way as given by a first-order all-pass transformation. The frequency-warped autocorrelation function may either be obtained from the power spectrum, which is fast if a large prestored matrix is employed, or by linear transformation of the original (lag-windowed) acf, especially using the Oppenheim recursion. As an alternative to computing the predictor from the warped acf, also the "covariance method" or continuously averaging adaptive methods (as the PARCOR lattice) can be used for analysis after slight modification. Here a new feature has to be taken into account: for a nonlinear frequency scale, spectral flatness and minimum variance of the prediction error are no longer equivalent criteria.

nloaded 10 May 2010 to 200.39.9.109. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/term

To obtain a synthesis filter from the warped predictor coefficients, no inverse transformation is required, rather, appropriately modified structures of the predictive analysis and synthesis filters in their direct form have been given as an immediate implementation of the frequency-warped transfer function. The filter coefficients are obtained from the predictor coefficients by a linear transformation that can be done either by multiplication with a triangular matrix or recursively "in place."

As a first application test, single-word intelligibility was measured for a Bark-scale and a frequency-scale vocoder, indicating fair performance of the former even in the low predictor-order range $p = 5-7$. Other applications have not been tested, the new methods should work for all known applications of frequency-warped prediction (cf. Sec. I) provided the restriction of the warping function to a first-order all-pass transformation is sufficient for that purpose.

ACKNOWLEDGMENTS

I thank Professor Dr. M. R. Schroeder for his interest in this work. All computations were carried out on a Honeywell H 632, purchased with funds from the Stiftung Volkswagenwerk.

Itahashi, S., and Yokoyama, S. (1978). "A formant extraction method utilizing mel scale and equal loudness contour," Speech Transmission Lab.—Quarterly Progress and Status Report (Stockholm) (4), 17–29.

Makhoul, J. (1976). "Methods for nonlinear spectral distortion of speech signals," Proc. 1976 Int. Conf. Acoust. Speech Sign. Proces. (Philadelphia), 87–90.

Makhoul, J., and Cosell, L. (1976). "LPCW: An LPC vocoder with linear predictive spectral warping," Proc. 1976 Int. Conf. Acoust. Speech Sign. Proces. (Philadelphia), 466–469.

Markel, J. D., and Gray, A. H., Jr. (1976), *Linear Prediction of Speech* (Springer, Berlin).

Oppenheim, A. V., and Johnson, D. H. (1972). "Discrete representation of signals," Proc. IEEE 60, 681–691.

Stålhammar, J. U. J. (1978). "Form factors for power spectra of vowel nuclei II," Speech Transmission Lab.—Quarterly Progress and Status Report (Stockholm) (2–3), 23–34.

Tohkura, Y., Itakura, F., and Hashimoto, S. (1978). "Spectral smoothing technique in Parcor speech analysis-synthesis," IEEE Trans. Acoust. Speech Sign. Process. ASSP–26, 587–596.

Zwicker, E., and Feldtkeller, R. (1967), *Das Ohr als Nachrichtenempfänger* (Hirzel, Stuttgart), 2nd ed., Tab. 27, I.

wnloaded 10 May 2010 to 200.39.9.109. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms