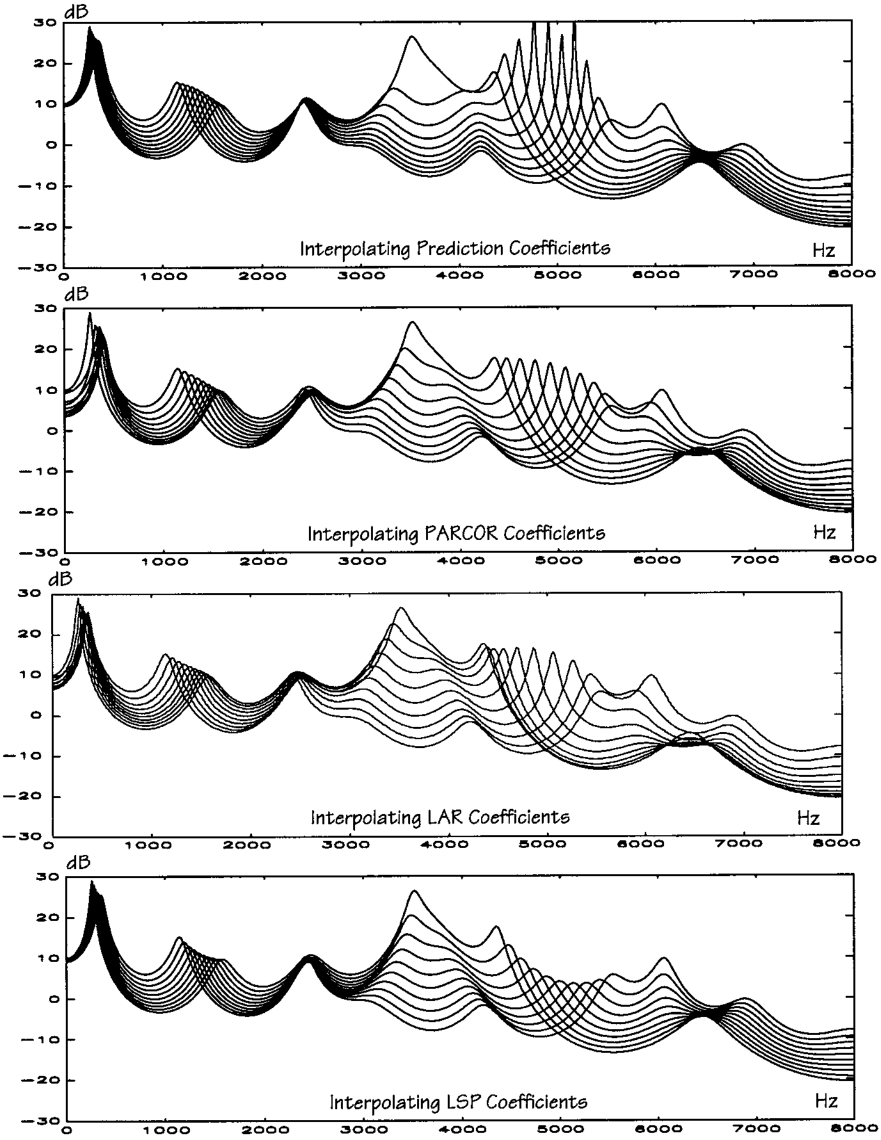

Fig. 8.9. Linear interpolation of filter parameters reveals that even in the case of small acoustic differences, linear concatenation laws may lead to nonoptimal results. Interpolation is performed between /l/ phones, respectively encountered in diphones /al/ and /lo/ and the resulting filter transfer functions are plotted (p=18).

transitions, since nothing keeps the zeros of $A_p(z)$ from moving out of the unit circle. LSPs are generally found to produce smoother transitions than LARs or PARCORs. In practice, all sets of coefficients give perceptually similar results in the context of segments concatenation synthesis. As far as filter stability is concerned, ***interpolating*** two sets of stable LARs or LSPs always results in stable filters, since the corresponding PARCORs maintain a norm lower than one and LSP tracks do not cross. This property, however, is not automatically maintained when linear ***smoothing*** is considered, as described by equations (7.1) and (7.2), though it happens to be the case in practice.

One should not conclude, however, that PARCOR or LSP coefficients lead to optimal interpolations. Transitions are still encountered, in which formant amplitudes are unnaturally affected when their frequencies are modified. In order to avoid such imperfections, the prediction filter poles are sometimes proposed as interpolation parameters (Papamichalis, 1987). [14] This requires the computation and storage of the roots of $A(z)$ instead of its coefficients, an operation that is scarcely considered in vocoders while it is easy to face in TTS synthesis since the database analysis is performed off-line. However, as shown in Fig. 8.10, grouping filter poles by pairs is not always a straightforward operation and interpolating them can be disastrous. Such pairing problems are not encountered with LSP's (see Fig. 8.10, on which LSPs for the left and right segments are almost superimposed).

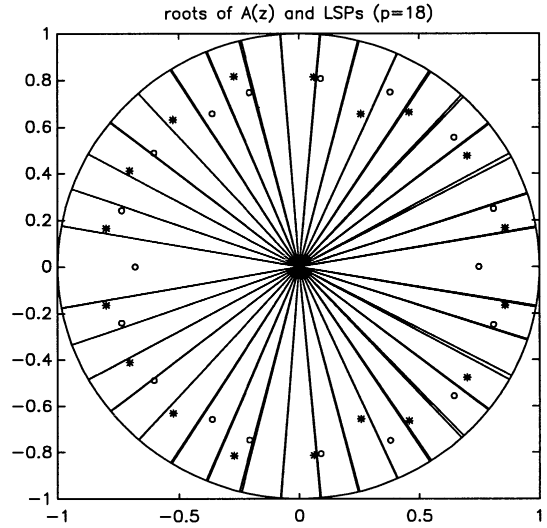

Fig. 8.10. Roots of A(z) for diphones /ɔv/ (o) and /vɛ/ (*), for p=18. LSP sets are presented in the form of radial lines, for they are very close to one another.

[14] They also obviously make it possible to control the filter stability.

Some authors have also proposed to interpret prediction filter poles in terms of the amplitudes, frequencies, and bandwidths of spectral amplitude peaks, and to smooth their trajectories over diphones (see Depalle *et al.*, 1990, for example). However, since the reverse transformation (from peaks to prediction coefficients) is not easy to perform, they implement a frequency domain synthesizer (see Chapter Nine).

8.6. Speech synthesis

Synthesis is performed as in Fig. 8.1, in which parameters are received from the segment concatenation module. Given the bad interpolation properties of the prediction coefficients, the IIR synthesis filter $1/A_p(z)$ is never implemented, either in its direct or transposed form. PARCOR coefficients can be used directly in a lattice synthesis filter derived from the lattice inverse filter (Boite and Kunt, 1987) (Fig. 8.11), which is easily found to require (4*p*-2) elementary operations (additions and multiplications) per sample—70 operations for order 18. On the other hand, an LSP filter has been proposed by Sugamura and Itakura (1986). Its computational load is about 10 percent higher than for the lattice filter.

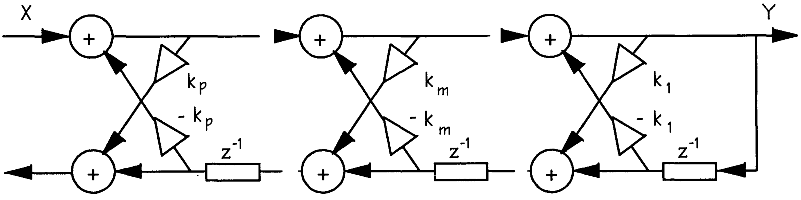

Fig. 8.11. The lattice synthesis filter.

In order to avoid audible transients when coefficients are updated, the filter coefficients are linearly interpolated inside each synthesis frame. It is found experimentally that the update frequency strongly depends on the type of filter used: about once every 5 ms (for speech at normal speed) for PARCOR coefficients, as opposed to once every 0.5 ms for the LSP parameters. Thus, since LSP parameters have no significant advantage over PARCOR coefficients, except their slightly higher compression ratio, and given the additional computational load they imply, they have never been used so far for TTS synthesis.

8.7. Segmental quality

The selection of an analysis method with respect to optimal quality is not simple. The covariance method theoretically is a better spectral estimator than the autocorrelation method (Kay, 1988); it is generally admitted that it provides more naturalness, as far as pitch asynchronous analysis is concerned (Hunt *et al.*, 1989). Differences, however, tend to vanish when $N>>p$.