2. LSP REPRESENTATION

For a given order p, the LPC analysis of a speech segment results in an all zero, minimum phase, prediction error filter [11] whose response is represented by the following z-transform expression:

$$A_p(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$$

For the purpose of deriving LSP's, two artificial (p+1)-order polynomials are defined as:

$$P_p(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1})$$

$$Q_p(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1})$$

The two polynomials correspond, respectively, to a complete opening and a complete closure of the glottis in an acoustic tube [2]. The polynomials $P_p(z)$ and $Q_p(z)$ have three fundamental properties:

1. All their roots are on the unit circle.
2. The roots are interleaved with each other and satisfy the following relationships:

$$0 = \alpha_0 < \alpha_1 < \alpha_2 < \ldots < \alpha_{p-1} < \alpha_p = \pi$$

where: $\alpha_i$ are roots of $P_p(z)$ for $i$ even

$\alpha_i$ are roots of $Q_p(z)$ for $i$ odd and $\alpha$ is a normalized frequency, with $\pi$ representing the folding frequency.

The roots ($\alpha_0, \ldots, \alpha_p$) are defined as LSP frequencies.

3. The previous properties 1) and 2) ensure the stability of the corresponding synthesis filter $1/A_p(z)$.

Due to the properties 1) and 2), the polynomials $P_p(z)$ and $Q_p(z)$ can be factorized, in the case of $p$ even, in the following way:

$$P_p(z) = (1 - z^{-1}) \prod_{k=2}^{p} (1 - 2 \cos \alpha_k z^{-1} + z^{-2})$$

$$Q_p(z) = (1 + z^{-1}) \prod_{k=1}^{p-1} (1 - 2 \cos \alpha_k z^{-1} + z^{-2})$$

The constant roots at $z=1$ and $z=-1$ can be removed from the polynomials $P_p(z)$ and $Q_p(z)$, respectively. In this way two alternative polynomials, $G_p(z)$ and $Q_p(z)$, are determined:

$$G_p(z) = \frac{P_p(z)}{(1 - z^{-1})}$$

$$G_p(z) = \frac{Q_p(z)}{(1 + z^{-1})}$$
Similar relations can be easily derived for the case of p odd [8,9]. By evaluating the two polynomials $G_1(z)$ and $G_2(z)$ on the unit circle we obtain:

$$G_1(e^{j\omega}) = e^{-j\pi p/2} G_2(e^{j\omega})$$

(9)

where $[G_1(e^{j\omega})]_{p,1}$ are two zero phase series expansions in cosines:

$$G_1(e^{j\omega}) = \sum_{k=0}^{p/2} g_1(k) \cos(\frac{p}{2} - k)\omega$$

(10)

The roots of $[G_1(e^{j\omega})]_{p,1}$ are simply related to the LSP frequencies by a cosine transformation. Furthermore, starting from the frequency mapping $\cos\omega$, $G'$-functions can be evaluated as a sum of Chebyshev polynomials:

$$G'_1(x) = \sum_{k=0}^{p/2} g_1(k) T_p(\cos(\frac{p}{2} - k)\omega)$$

(11)

where:

$$T_0(x) = 1$$

$$T_1(x) = x$$

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$$

(12)

As a result, these functions can be computed, in a given point $x$ of the interval $[-1,1]$, with about $p/2$ multiplications and $p$ additions[8].

![Fig. 1](image)

Fig. 1 Spectrogram over LSP trajectories (b) of the speech waveform (a), corresponding to the Italian word "angebo"("angel"). Spoken by a female speaker (x-axis is time in centiseconds, y-axis is frequency in Hz). $d_{Ang}$ and $d_{pitch}$ distance measures (see Section 5) between adjacent frames are given in (c).

3. ALGORITHMS FOR THE COMPUTATION OF LSP's

Two algorithms are described in the following, which represent a modification of the Kabal and Ramachandran's algorithm, originally presented in [8].

In order to determine reasonable assignments to the algorithm parameters and to evaluate the complexity of the proposed algorithms, a speech database of four speakers (two male and two female) has been used. The data base consisted of 340 words (approximately 250 s), sampled at 15 kHz and representative of a wide range of sounds. Preemphasis of the digitized speech has been accomplished by a first order digital filter whose transfer function is: $H(z)=1-0.95z^{-1}$. A 20 ms Hamming window, spaced every 10 ms, has been used to determine the input signal to the LPC analysis. The LPC coefficients have been obtained by using the Durbin recursive procedure.

Algorithm of Kabal and Ramachandran

In the previous section two polynomials $G_1(x)$ and $G_2(x)$ have been introduced[11]. Their roots (x-coefficients) correspond to the LSP frequencies by the mapping $[\cos x]_{p,1}$.

In the Algorithm of Kabal and Ramachandran (AKR) the roots of the equations $[G_1(x)]=0,1$ are determined by applying a bisection method inside the interval $[-1,1]$. Starting from $x=1$, the search procedure performs successive bisections of an elementary interval of length $h_k$, when at its boundaries different sign values of $G'_2(x)$ have been detected.

Once the first root $x_1$ has been found, the algorithm goes on searching for the first $G'_1(x)$ root, i.e. $x_2$ and so on. The acceptable uncertainty in the root estimation is tuned by a parameter $\varepsilon_k$, which has been fixed to 0.00075, in order:

a) to avoid a root missing, when the algorithm switches from one searching function to another, and

b) to keep the LSP frequency uncertainty between 90 Hz in the low and high range of the spectrum and 10 Hz in the medium range (between 300 and 7200 Hz).

An interval size of $\delta_k=0.02$, as suggested in [8], has been adopted in order both to keep the algorithm complexity relatively constant and to avoid the loss of some roots.

It is worth noting that a great contribution to the computational complexity of AKR is due to the $G'$-function evaluations, even if the searching procedure is extremely simple and suitable to real time applications. For a given frame, AKR ensures a maximum number of $G'$-function evaluations of $[2/p\pi(p-\log_2(\delta_k/\varepsilon_k))$, even if the actual number is somewhat smaller (about 160, on the average, for a 10-th order LPC analysis).

The following algorithms are characterized by a number of function evaluations substantially lower, if compared with the AKR.

Algorithm 1

An algorithm (A1) has been used, which, for a given frame, starts with $x_0=x_0$, and repeats the same search procedure adopted in AKR, with $\delta_k'=\delta_k/2$, until p roots have been found: in this way the average number of $G'$-function evaluations is decreased, especially for low predictor orders. Experiments were conducted to determine reasonable assignments to $\delta_k$, which are given in Tab.1.

<table>
<thead>
<tr>
<th>$p$</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_k$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.08</td>
<td>0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Tab.1 Assignments to $\delta_k$ as a function of the predictor order $p$.

Algorithm 2

A more sophisticated algorithm (A2) has been derived, which exploits the high interframe redundancy of the LSP frequencies (Fig.1(b)), i.e. of the x-coefficients.

Starting from the configuration $[x'_k, x_{k+1}, p]$ (Fig.2(a)) of the previous frame, a set of p intervals, centered on each $x'_k$ and of length $\Delta x_k=1/\Delta t$, are defined (Fig.2(b)). $\Delta l_k$ is updated every frame, as a fraction of the minimum distance between the corresponding $x'_k$ and the coefficients $x_{k+2}, x_{k-2}$, i.e. the adjacent roots extracted from the same $G'_1(x)$ function.

As a second step, starting from $x'_k$, a search procedure is applied to the intervals containing each $x'_k$ by evaluating $G'_1(x)$ on the boundaries of the corresponding central interval and then on the boundaries of the adjacent intervals (Fig.2(c)). When a sign change of $G'_1(x)$ has been detected, a bisection procedure is applied, which exploits a linear interpolation technique to speed up the evaluation of the zero and stops when either the prescribed accuracy $\varepsilon_k$ (i.e., $|G'_1(x=\varepsilon_k)|$) or the assigned tolerance $\varepsilon_k$ of the zero $x_k$ is reached. With a choice of $\varepsilon_k=0.01$, a very low number of imprecise evaluations of the x-coefficients has been found, with a maximum error corresponding to about 50 Hz.
A strong misalignment of the x-configuration from the previous one can constrain the search procedure over the left boundary \( x = -1 \), as indicated in Fig.2(d). The same condition holds when two roots, coming from the same function, are located in the same search interval. In both cases the position of the zeros, which have been missed, cannot be easily identified: hence the procedure is repeated with a new initial x-configuration, which comes from a right translation of \( \delta_k \) from the previous x-configuration; in addition, the searching interval lengths are halved, as shown in Fig.2(e). From our experiments the rate of this critical condition which, by the way, corresponds to significant spectral changes, resulted to be less than 1\%.

In Fig.3 comparisons between the three algorithms are given in terms of the average number of \( G' \)-function evaluations required for each frame.

A further optimization of algorithm \( A_2 \) could be represented by a more sophisticated prediction of the initial x-configuration, which takes into account two or three previous frames. But, preliminary studies have not shown any improvement, in comparison with \( A_2 \), when a simple linear interpolation is applied to the previous two x-sets in order to determine the predicted x'-configuration.

The spectral distance measure used, as reference, throughout this section is the L2-norm \( d_2 \) of the spectral difference between any two given log spectra (rms log spectral measure). This measure is frequently used, as reference, both for its high relevance to subjective human perception and for its computational simplicity (using the cepstral coefficients\[15\]). Recently \( d_2 \) has been used, in a speech coding application, to study the spectral sensitivity as a function of the LSP frequencies\[14\].

In most cases, the Euclidean distance directly measured on LSPs, indicated here as \( d_{eucl} \), provides high performance in speech coding\[15\] and recognition\[16\]. In general, however, its relation with more appropriate spectral measures should be considered, especially when a weighting function is introduced\[16\].

In alternative to the Euclidean distance, a class of LSP-based distance measures \( d_{LSP} \) is described here, which basically approximates \( d_2 \) by evaluating the two power spectra only in some spectral points, related to the LSP frequencies.

Consider two LPC-models \( \alpha(A_p(z)) \) and \( \alpha'(A_p'(z)) \). The rms log spectral measure is defined\[15\] by:

\[
(d_2)^2 = \int_{-\pi}^{\pi} \frac{S'(\omega)^2}{S(\omega)} \frac{d\omega}{2\pi}
\]

depends on the whole LSP distribution, according to the relation:

\[
S(\omega) = \frac{2^{-p} \sigma^2}{[\sin \frac{\omega}{2} \sum_{k=2}^{p} (\cos \omega \cdot x_k)^2 + [\cos \frac{\omega}{2} \sum_{k=1}^{p-1} (\cos \omega \cdot x_k)^2]^2}
\]

for \( p \) even, and:

\[
S(\omega) = \frac{2^{-p} \sigma^2}{[\sin \frac{\omega}{2} \sum_{k=2}^{p} (\cos \omega \cdot x_k)^2 + [\cos \frac{\omega}{2} \sum_{k=1}^{p-1} (\cos \omega \cdot x_k)^2]^2}
\]

for \( p \) odd, where \( \sigma \) is the LPC residual root mean square value and \( \{x_k=\cos k\omega\}_{k=1,p} \). From (14),(15) it appears that, when one or more lines are close to another line characterized by \( \omega_0 \), both denominator terms tend to zero, for \( \omega \to \omega_0 \) and \( S(\omega) \) becomes large. Furthermore, when two LSPs are concentrated, it holds that:

a) for an even predictor order, the odd LSP frequencies, which correspond to the closed glottis model, are located near a formant center frequency, when formants are present, as much as this frequency is low. Conversely, when the formant center frequency is positioned in the higher part of the spectrum, the role of even and odd LSP frequencies are interchanged.

b) for an odd predictor order, the odd LSP frequencies are dominant both in the high and in the low spectrum.

c) in both cases, in the medium spectrum odd and even LSP frequencies influence equally the position of the formant center frequency.

The same properties can be stated by the following relationships:

\[
S(\omega) = \frac{2^{-p} \sigma^2}{[\sin \frac{\omega}{2} \sum_{k=2}^{p} (\cos \omega \cdot x_k)^2 + [\cos \frac{\omega}{2} \sum_{k=1}^{p-1} (\cos \omega \cdot x_k)^2]^2}
\]

for \( p \) even, and:

\[
S(\omega) = \frac{4\sigma^2}{[\sin \frac{\omega}{2} \sum_{k=2}^{p} (\cos \omega \cdot x_k)^2 + [\cos \frac{\omega}{2} \sum_{k=1}^{p-1} (\cos \omega \cdot x_k)^2]^2}
\]

for \( p \) odd.

Both (14),(15) and (16),(17) allow to evaluate the power spectrum by either the LSP frequencies or the \( \alpha \)-coefficients. In addition, (16),(17) suggest the derivation of a spectral distance between two spectral models, which exploits the values \( G_{1,2}(\omega) \).

5. AN LSP-BASED DISTANCE MEASURE

The spectral distance measure used, as reference, throughout this section is the \( L_2 \)-norm \( d_2 \) of the spectral difference between any two given log spectra (rms log spectral measure). This measure is frequently used, as reference, both for its high relevance to subjective human perception and for its computational simplicity (using the cepstral coefficients\[15\]). Recently \( d_2 \) has been used, in a speech coding application, to study the spectral sensitivity as a function of the LSP differences\[14\].

In most cases, the Euclidean distance directly measured on LSPs, indicated here as \( d_{eucl} \), provides high performance in speech coding\[15\] and recognition\[16\]. In general, however, its relation with more appropriate spectral measures should be considered, especially when a weighting function is introduced\[16\].

In alternative to the Euclidean distance, a class of LSP-based distance measures \( d_{LSP} \) is described here, which basically approximates \( d_2 \) by evaluating the two power spectra only in some spectral points, related to the LSP frequencies.

Consider two LPC-models \( \alpha(A_p(z)) \) and \( \alpha'(A_p'(z)) \). The rms log spectral measure is defined\[15\] by:

\[
(d_2)^2 = \int_{-\pi}^{\pi} \frac{S'(\omega)^2}{S(\omega)} \frac{d\omega}{2\pi}
\]
where \( S(a) \) and \( S'(a) \) are the two power spectra. Conversely, the Euclidean distance measure \( d_{esp} \) is defined as:

\[
(d_{esp})^2 = \frac{1}{P} \sum_{k=1}^{P} (e_k - e'_k)^2
\]  

(19)

The distortion measure \( d_2 \) can be approximated by evaluating the two power spectra \( S(a) \) and \( S'(a) \) either, for instance, in the LSP frequencies or, more generally, in a grid of \( N \) spectral points \( \theta_0, \theta_1, \ldots, \theta_{N-1} \). The resulting measure is defined as:

\[
(d_2)^2 = \frac{1}{N} \sum_{k=1}^{N} S'(\theta_k)^2 - \frac{1}{N} \sum_{k=1}^{N} S(\theta_k)^2
\]  

(20)

Here \( d_{esp} \) is referred to the case \( e_k = \{ e_0, e_1, \ldots, e_{N-1} \} \), i.e. the LSP frequencies corresponding to \( A_2(z) \) and \( A'_2(z) \), respectively. In this case the \( d_{esp} \) measure, between successive frames, can be obtained by the LSP extraction procedure \( A_2 \), with only a few additional \( G' \)-function evaluations, since either \( G'(x) \) or \( G'(2x) \) can be considered zero in the spectral frequencies \( \{ \theta_0, \theta_1, \ldots, \theta_{N-1} \} \).

A second choice for \( \theta \), which gives a measure \( d_{esp} \) to locate the spectral frequencies \( \{ \theta_k \}_{k=1}^{N} \), is taken into account the relationships, given in the previous section, between formant center positions and LSP's. Broadly speaking, the grid (of \( N \times 2p \) points) is determined by a set of spectral points inside the intervals which correspond to the minimum LSP frequency differences \( \{ \Delta a_{k-1}, p \Delta e_{k-1}, p \} \) where \( p \) the closer two LSP's are together, the higher the number of spectral points between them. As a result, the distance measure \( d_{esp} \) gives more weight to the spectral peaks than to the spectral valleys and is more related to formant movements than the distance measure \( d_2 \), as shown in Fig. 1(c). As \( N \) increases, both \( d_{esp} \) and \( d_2 \) approach \( d_2 \).

The scatter plot of \( d_2 \) for an example, versus the rms log spectral measure \( d_2 \) is shown in Fig. 4. The cross-correlation coefficients are given in Tab. 2. For comparison purposes, the cross-correlation coefficient between \( d_2 \) and the cepstral distance, obtained with \( p \) cepstral coefficients, has been included[13]. Experiments have been conducted, with a predictor order \( p=16 \), by comparing only successive time frames; similar results have been obtained with other predictor orders. Only measures below 6 dB contributed to the cross-correlation computation. The LPC gain has not been considered.

\[
\begin{array}{l}
0.991 \\
0.994 \\
0.943 \\
0.895 \\
\end{array}
\]

Tab.2. Cross-correlation coefficients between given distances and \( d_2 \).

Further studies should be devoted to quantify the same relationship between a weighted Euclidean distance [16] and \( d_2 \). Nevertheless, an important conclusion can be drawn here: the unweighted Euclidean distance \( d_{esp} \) performs better than \( d_{esp} \). \( d_2 \) in terms of cross-correlation with \( d_2 \). However, this result cannot establish which LSP-based distance is more meaningful for speech processing; for this purpose, work is in progress in order to obtain a definitive comparison in terms of performance of an automatic speech segmentation task, which is being developed.

6. CONCLUSIONS

In this paper the computation of LSP frequencies from LPC coefficients has been considered. By exploiting the high frame redundancy of the LSP's, it is possible to achieve a considerable computational saving, with respect to single frame analysis procedures, even if an application of the resulting algorithm to a real time environment should require a careful analysis of its complexity.

Next, the task of speech segmentation during the LSP extraction has been investigated. Some spectral relationships between the LSP's and the corresponding spectral envelope have been described and used for the definition of an LSP-based spectral distance measure, which lends itself to a meaningful frequency domain interpretation. It turns out that the correlation between the unweighted Euclidean distance measure and the rms log spectral measure is higher. However the LSP-based distance measure seems to be more related to perceptually significant acoustic cues. Further studies are needed to verify if the use of this distance measure, for speech segmentation or for speech recognition, is convenient.

REFERENCES


Fig. 4. Scatter plot of \( d_2 \) versus the rms log spectral measure \( d_2 \).