A review of line spectral pairs

Ian Vince McLoughlin

School of Computer Engineering, Nanyang Technological University, N4 #02a-32 Nanyang Avenue, Singapore 639798

Abstract

Linear prediction-based coders commonly utilise line spectral pairs (LSP) to represent linear prediction coefficients for reasons of filter stability and representational efficiency. Line spectral pairs have other useful properties such as an ordering related to the spectral properties of the underlying data, which leads to advantages when used for analysing speech and other signals. This paper reviews the LSP representation, conversion and quantization processes, computational issues associated with the implementation of LSP-based methods, and their use in speech analysis and processing. In addition, this paper presents LSP manipulation methods that can be used to alter frequencies within the represented signal in a consistent and relevant way, and considers the use of LSPs for analysis of non-speech information.

Key words: Line spectral pair, line spectral frequency, LSP, LSF, speech modification, speech analysis, speech processing
PACS: Speech processing, 43.72.p, 43.72.+q

1 Introduction

Line spectral pairs (LSP) are a direct mathematical transformation of a set of linear prediction coefficients such as are generated within many speech compression systems, including CELP (codebook-excited linear prediction) analysis-by-synthesis coders [1]. LSP usage is popular due to the good quantization characteristics [2] and consequent efficiency of the representation. Individual lines are commonly referred to as Line Spectral Frequencies (LSF).

LSFs collectively describe the two resonance conditions arising from an interconnected tube model of the human vocal tract. This includes mouth shape and nasal cavity, and forms the basis of the underlying physiological relevance of the linear prediction representation [3]. The two resonance conditions are those that describe the vocal tract being either fully open or fully closed at the glottis respectively. The model in question being constructed from a set of

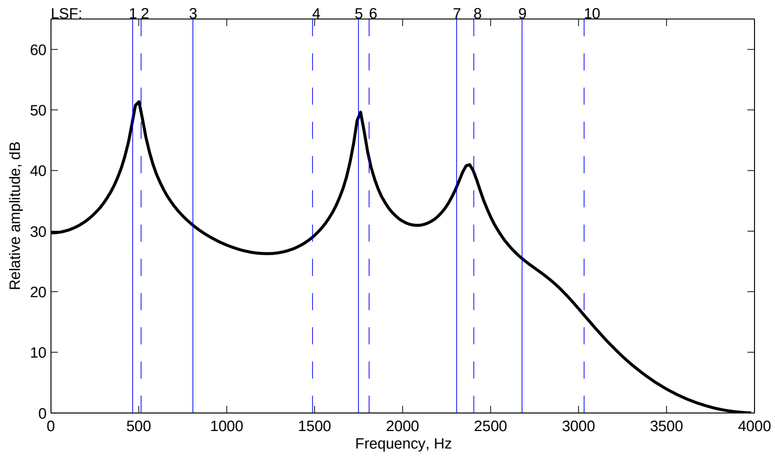

Fig. 1. An example LPC spectrum overlaid with the corresponding vertical LSP frequencies. Odd lines are drawn solid and even lines are drawn dashed.

equal-length but different diameter tubes, with the source end either closed or open. The two conditions give rise to two sets of resonant frequencies, with the number of resonances in each set being determined by the number of joined tubes (which in turn is a function of the order of the analysis system). The resonances of each condition are the odd and even line spectra respectively, and are interleaved into a monotonically increasing set of line spectral frequencies.

In reality, the human glottis opens and closes rapidly during voiced speech: it is neither fully closed nor fully open over an analysis frame. Hence actual resonances occur at frequencies located somewhere between the two extremes of odd and even LSP condition. Nevertheless, this relationship between vocal resonance and LSP position endows them with a significant interpretation which we will build upon as we review the representation.

Fig. 1 illustrates LSPs overlaid on a power spectrum plot. The ten vertical lines were drawn at the LSFs, and show the odd (solid) and even (dashed) frequencies being interleaved. Both the lines and the spectrum were derived from the same set of linear prediction parameters which were in turn obtained from 10^{th} order linear predictive analysis of a 20ms frame of voiced speech.

Apart from the natural interleaving of the line frequencies, it is notable that peaks in the underlying spectrum of Fig. 1 tend to be bracketed by a narrow pair of lines. By contrast, local minima in the spectrum tend to not have LSPs overlaid nearby. This relationship between line location and spectral resonance is one reason for the popularity of LSPs for the analysis, classification and transmission of speech.

This paper will investigate the line spectral pair representation and some of its many uses. Section 2 will derive the LSP equations, section 3 will discuss the traditional usage of and motivation for LSPs, while sections 4, 5 and 6 will consider the interpretation of the representation including static and dynamic analysis. Section 7 introduces speech modification through LSP adjustment. Section 8 looks at non-speech applications of LSPs before section 9 reviews implementation issues related to the analysis, quantization and modification of LSPs and section 10 concludes the paper.

2 The LSP representation

Line spectral frequencies are derived from the Linear predictive coding (LPC) filter representing vocal tract resonances in analysed speech, for p^{th} order analysis:

$$A_p(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_p z^{-p} \tag{1}$$

We will define two $(p+1)^{th}$ order polynomials related to $A_p(z)$, which we shall name $P(z)$ and $Q(z)$. These are referred to as antisymmetric (or inverse symmetric) and symmetric in turn based on observation of their coefficients. The polynomials represent an interconnected tube model of the human vocal tract. They correspond in turn to complete closure at the source end of the interconnected tubes and a complete opening, defined by the $(p+1)^{th}$ extra term. In the original model, the source end is the glottis, and as discussed in section 1, this is neither fully open nor fully closed during the period of analysis, and thus the actual resonance conditions encoded in $A_p(z)$ are a linear combination of the two boundaries. In fact this is simply stated:

$$A_p(z) = \frac{P(z) + Q(z)}{2} \tag{2}$$

The two polynomials are created from the LPC polynomial with an extra feedback term being positive to model energy reflection at a completely closed glottis, and negative to model energy reflection at a completely open glottis:

$$P(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1}) \tag{3}$$

$$Q(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1}) \tag{4}$$

The roots of these two polynomials are the set of line spectral frequencies (LSF), ω_k. These relate back to the symmetric and antisymmetric polynomials as will be shown later in (6).

It can be shown that the complex roots of the polynomials will lie on the unit-circle in the z-plane if the original LPC filter was stable [2], and alternate in order around the unit circle. It is also important to note that *any* equivalent size set of roots that alternate in this way around and on the unit circle will represent a stable LPC filter, and we shall consider the implications of this for LSP-based adjustment and processing in sections 3 and 7.

There are many different ways of calculating or estimating roots, such as trial and error, Newton-Raphson approximation, evaluating around the unit circle looking for sign changes and so on. There has been a significant amount of research effort expended on finding computationally efficient solutions for real-time speech processing algorithms using LSPs, discussed further in section 9. Here we simply assume that an appropriate method, such as the MATLAB *roots()* function, has been employed.

If we denote the set of complex roots as θ_k, then the line spectral frequencies are determined from (3) and (4):

$$\omega_k = tan^{-1}\left(\frac{Re\{\theta_k\}}{Im\{\theta_k\}}\right) \tag{5}$$

ω_k are then the p line spectral frequencies expressed in radians.

2.0.1 Generation of LPC coefficients from LSPs

Conversion from LSPs back to LPCs is a simple process [2], since we can easily use the ordered LSPs ω_k to recreate the polynomials that they are roots of, namely $P(z)$ and $Q(z)$ [4]:

$$\begin{aligned} P(z) &= (1-z^{-1}) \prod_{k=2,4,\cdots,p} (1-2z^{-1}cos\omega_k + z^{-2}) \\ Q(z) &= (1+z^{-1}) \prod_{k=1,3,\cdots,p-1} (1-2z^{-1}cos\omega_k + z^{-2}) \end{aligned} \tag{6}$$

And then these can be substituted into (2) expressing $A_p(z)$ as the linear combination of $P(z)$ and $Q(z)$. Since (6) involves the cosine of the LSPs, and bearing in mind that some of the more efficient LSP calculation algorithms will yield the roots in the cosine domain [5], it is common to also perform the reverse conversion from LPC to LSP in the cosine domain.

If q_k denotes the array of correctly ordered, lowest-first, LSP frequencies ω_k, in the cosine domain:

$$q_k = cos\omega_k \tag{7}$$

then we can use these to solve the following set of recursive equations as an intermediate step towards the calculation of LPCs:

$$\begin{aligned}
&for \quad k = 1 \ldots p \\
&\quad f_k = -2f_{(k-1)}q_{(2k-1)} + 2f_{(k-1)} \\
&\quad for \quad m = (k-1) \ldots 1 \\
&\quad\quad f_m = f_m - 2f_{(m-1)}q_{(2k-1)} + f_{(m-2)}
\end{aligned} \tag{8}$$

Bearing in mind initial conditions of $f_0 = 1$ and $f_{-1} = 0$, and calculate the coefficients of g similarly as follows:

$$\begin{aligned}
&for \quad k = 1 \ldots p \\
&\quad g_k = -2g_{(k-1)}q_{(2k)} + 2g_{(k-1)} \\
&\quad for \quad m = (k-1) \ldots 1 \\
&\quad\quad g_m = g_m - 2g_{(m-1)}q_{(2k)} + g_{(m-2)}
\end{aligned} \tag{9}$$

Once the values of f and g have been determined, they form a second set of equations:

$$f'_k = f_k + f_{(k-1)} \qquad for \quad k = 1 \ldots p/2 \tag{10}$$

$$g'_k = g_k + g_{(k-1)} \qquad for \quad k = 1 \ldots p/2 \tag{11}$$

which are then averaged to form LPC coefficients from the following:

$$a_k = \frac{1}{2}f'_k + \frac{1}{2}g'_k \qquad for \quad k = 1 \ldots p/2 \tag{12}$$

$$a_k = \frac{1}{2}g'_{(k-p/2)} - \frac{1}{2}g'_{(k-p/2)} \qquad for \quad k = p/2 + 1 \ldots p \tag{13}$$

2.0.2 LSP power spectrum

Given the transfer function of the all-pole LPC filter:

$$H(z) = \frac{1}{A_p(z)} \tag{14}$$

the power spectrum can be calculated from:

$$|H(e^{j\omega})|^2 = \frac{1}{|A_p(e^{j\omega})|^2} \tag{15}$$

Obviously this can be directly evaluated if LPC coefficients are available, but the intention here is to develop a formulation based on LSP coefficients. So using the relationship between $A_p(z)$ and the $P(z)$ and $Q(z)$ polynomials in (2) we can see that [6]:

$$|H(e^{j\omega})|^2 = 4|P(e^{j\omega}) + Q(e^{j\omega})|^{-2} \tag{16}$$

Then substituting in the standard expression for $P(z)$ and $Q(z)$, and knowing that these two polynomials, given in (6), are odd and even respectively, we can calculate a power spectrum [4] directly from the LSP values:

$$|H(e^{j\omega})|^2 = 2^{-p} \Big/ \left\{ \begin{array}{cc} sin^2(\frac{\omega}{2}) \prod_{k=2,4,\cdots,p} & (cos\omega - cos\omega_k)^2 \\ + \, cos^2(\frac{\omega}{2}) \prod_{k=1,3,\cdots,p-1} & (cos\omega - cos\omega_k)^2 \end{array} \right\} \tag{17}$$

If the above equation were evaluated for the LSPs, ω_k, of Fig.1 in 100 steps ranging from angular frequency of $\omega = 0 \ldots \pi$, the result would be identical to the power spectral plot in that figure.

Next, considering the dependencies of the power spectrum equation, it can be seen that (17) reaches a maximum when both terms of the denominator $(cos\omega - cos\omega_k)$ are minimum. Since one term is indexed with odd k and one with even k, this occurs when two adjacent lines have similar value. This minimum also occurs near angular frequencies of 0 or π since the first and second denominator terms respectively will be zero at these extremes. Thus an odd numbered line approaching 0 or an even numbered line approaching π would result in resonance peaks at the corresponding extreme. These properties corroborate the evidence that spectral peaks are bracketed by closely-spaced lines, and also that line placement near either frequency extreme will introduce a low or high frequency spectral bias respectively. It can be therefore be considered that there are implicit hidden lines at angular frequencies of 0 and π.

MATLAB functions, for converting bidirectionally between LPC and LSPs, and for plotting the LSP-with-spectrum as in Fig.1 are provided in Appendix A.

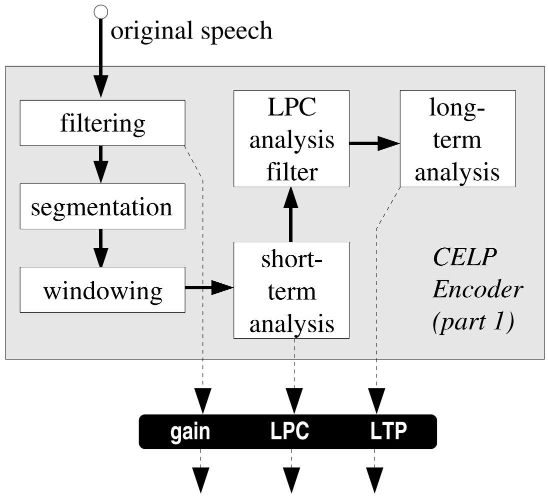

Fig. 2. Block diagram of CELP coder analysis process

3 Traditional use of LSPs

The method of LSP analysis was initially developed for application within speech codecs [3], where they are typically derived either through reflection coefficients or from LPC coefficients. In this section we explore this traditional usage, based around the quantization properties of LSPs, before we explore other more diverse proporties of, and applications for LSPs in subsequent sections.

3.1 LSPs within speech codecs

Many CELP speech compression systems convey pseudostationary human vocal tract spectral information between coder and decoder through the transmission of a set of LSPs. In such systems, other transmitted information may include gain and pitch [7]. Figs. 2 and 3 show a block diagram of a generic CELP speech codec, firstly the analysis section which determines four transmission parameter sets per frame from the input speech, namely pitch information, codebook index, gain and LSPs. Fig. 3 shows the analysis-by-synthesis process of the codebook search loop used to determine the optimal codebook index. In Fig. 4, the decoder is shown using the transmitted data to recreate decoded speech.

It is common to describe the CELP coder parameter sets as having physiological equivalents: if the codebook index conveys a lung excitation signal, a pitch parameter may convey glottal information, the gain value will convey overall vocal amplitude, and LSPs convey vocal tract spectral information including mouth shape, tongue position and the contribution of nasal cavity.

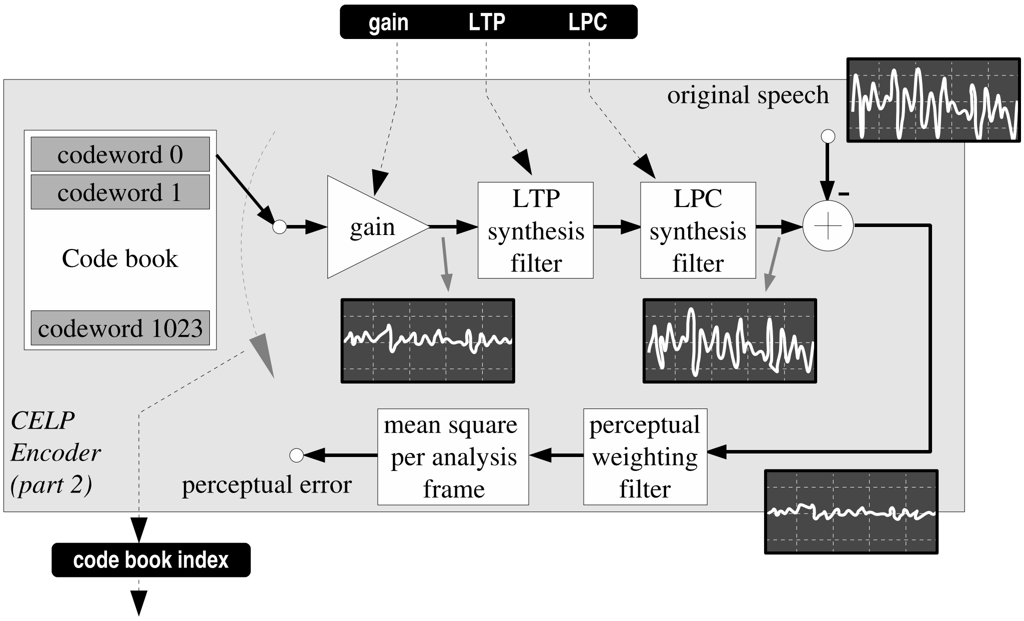

Fig. 3. Block diagram of CELP codebook search loop

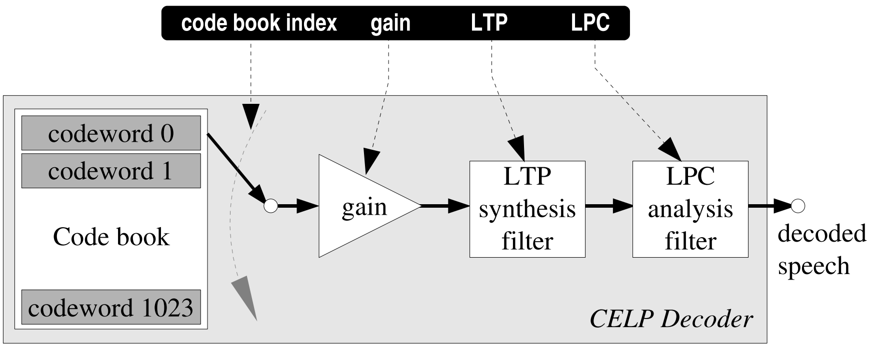

Fig. 4. Block diagram of CELP decoder

LSPs are used within the CELP coder for two main reasons. Firstly, they always result in a stable LPC filter, even after undergoing severe quantization [2], and secondly that they may be quantized with fewer bits than representations such as reflection coefficients or log-area ratios, while maintaining a given speech quality [8].

The stability feature of LSPs even allows the parameters to be interpolated across frames [7], and it is this robustness along with their relative efficiency on a bits-quality scale that has made their use popular. However one disadvantage of LSPs is undoubtedly the numerical complexity involved in their calculation, explored in section 9.

Table 1
SEGSNR resulting from uniform quantization of LPC and LSP parameters

Bits/parameter	LPC (dB)	LSP (dB)
4		-6.26
5	-535	-2.14
6	-303	1.24
7	-6.04	8.28
8	-10.8	15.9
10	19.7	20.5
12	22.2	22.2
14	22.4	
16	22.4	22.4

3.2 Quantization of LSPs

Speech from the large TIMIT database [24] was analysed and used to compare the relative susceptibility of raw LPCs and LSPs to uniform quantization at different resolutions.

Uniform quantization represents each line by the same number of bits, ranging from 4 to 16 bits per line in this case, for 10^{th} order analysis of 8kHz, 16bit sampled 240ms analysis frames. The segmental signal-to-noise ratio (SEGSNR - the log of the mean-squared difference between the original speech before LPC/LSP and the recreated speech resulting from synthesis) was performed for each test case and the degree of SEGSNR difference in dB determined.

The results, presented in table 1, show that LPCs are far more susceptible to quantization than LSPs: down around 5 bits per parameter, the recreated speech resulting from LPC quantization exhibits sharp spikes of noise and oscillation - totally unlike the original speech. Hence the difference between original and recreated speech evidenced by the large negative SEGSNR value. By contrast the LSP representation, with an SEGSNR of -2.14 indicates a likelihood of understandable speech. This substantiates the assertion of section 1 that LSPs are favoured for their superior quantization characteristics. Finally, note that both approaches in the table achieve a SEGSNR level of 22.4 when more bits are used for quantization, establishing the limit of achievable SEGSNR for the analysis process used (i.e. segmentation, Hamming windowing, autocorrelation, 10^{th} order analysis and so on).

LSPs in speech codecs may be available as an angular frequency, or the cosine of the angular frequency depending on the method used to solve the polynomial roots (as discussed in section 2). Each line value can be quantized independently (scalar quantization) on either a uniform or a non-uniform scale [3] which can also be dynamically adapted. Alternatively, lines can be grouped and vector quantized (VQ) with either static or adaptive codebooks [9]. VQ groups sets of lines together such as the typical (2,3,3,2) or (3,3,4) arrangements for 10^{th} order systems.

Both scalar and vector quantization can be applied either to the raw LSP values themselves, or to differential values, where the difference is either that between a line's current position and its position in the previous frame or between its current position and its mean position [10]. We can refer to these as the short-term and long-term average differences respectively [11].

An adaptation of long-term average differential quantization, is to recalculate a reference LSF distribution each frame, where the $(p-2)$ centre lines are evenly distributed within the range defined by the position of ω_1 and ω_p in that frame. This is known as Interpolated LSF (LSFI) [12]. A different form of interpolation is that applied by the TETRA standard CELP coder [7], which quantizes LSPs interpolated between subframes (of which there are four per standard-sized frame). This approach is desired to provide a degree of immunity to the effects of subframes lost due to transmission burst errors.

An effective quantization scheme will generally maximise either the signal-to-quantization noise ratio for typical signals, or will minimise a more perceptually-relevant measure. Such measures could be spectral distortion (SD) or variants. Some published examples are the LSP distance (LD) [10], LSP spectral weighted distance measure (LPCW) [13], local spectral approximation weights (LSAW) [14] and inverse harmonic mean weights (IHMW) [15].

In each case, it is necessary to appreciate the dynamics of the signal to be quantized, and optionally to assign different levels of importance to critical spectral regions, either directly, or by emphasising LSPs who's frequency locus is within such a critical region. For example, lines 9 and 10 in a 10^{th} order analysis may relate to format F3, if present. This formant can be considered less important to speech intelligibility than formants F1 and F2. Therefore lines 9 and 10 may be quantized with fewer bits than lines 2 and 3.

By plotting the LSP line frequency locus for a number of TIMIT speech recordings, as shown in Fig. 5, we can see that line localisation in frequency is fairly limited. It is naturally those lines located predominantly in less important frequency regions that would be chosen for reduced quantizer bit allocations. The plot was obtained through 10^{th} order LPC analysis on 40ms frames with 50% overlap for different male and female speakers. These LPC coefficients were then transformed into LSP values, with their cumulative distribution computed in analysis bins of size 40Hz and plotted in the vertical axis for

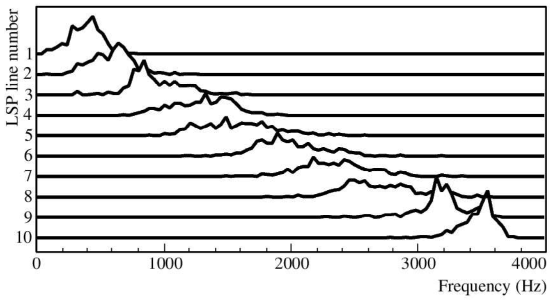

Fig. 5. Histogram of relative frequency placement of each of 10 LSPs in speech

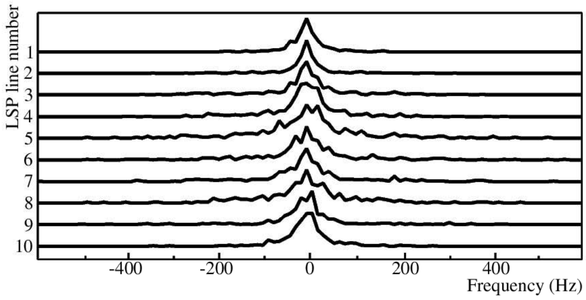

Fig. 6. Histogram of the relative frequency difference between LSP line locations for consecutive analysis frames.

each of the line traces.

Similarly, a differential analysis has been plotted in Fig. 6, where the relative frequency of frame-to-frame difference is computed for each line. Again, it can be seen that the frame-to-frame variation for individual lines is relatively localised. Both Figs. 5 and 6 show that lines 1 and 10 exhibit less variation in their absolute value, but also in their frame-to-frame variation than do the centre lines.

To examine further, table 2 gives the mean frequency for each line, and the median frequency for a flat spectral frame. It also lists the standard deviation between the average frequency and actual frequency of each line, σ_1, and the frame-to-frame standard deviation of each line, σ_2. The test database was the same as for Figs. 5 and 6. In the table, the standard deviation of lines 1 and 10 is significantly smaller than that of other lines, such as line 6, to corroborate the visual evidence in the figures. In a uniform scalar quantization scheme, this would mean that lines 1 and 10 require fewer quantization bits than line 6 for equivalent signal-to-quantization noise ratio.

The spread of values for the mean analysis is similar to that for the differential analysis, in fact only 2.5% less. If the quantization range was defined as, for example, a two standard deviation span around the mean, then both schemes would therefore have similar quantization requirements for given SNR [9].

Table 2
Mean frequency, standard deviation to mean (σ_1), standard deviation between frames (σ_2) and median ($\bar{\omega}$) for 10^{th} order LSPs from a test database

no.	mean (Hz)	σ_1 (Hz)	σ_2 (Hz)	$\bar{\omega}$ (Hz)
1	385	117	91	364
2	600	184	175	727
3	896	241	241	1091
4	1269	272	279	1455
5	1618	299	292	1818
6	1962	306	312	2182
7	2370	284	314	2545
8	2732	268	285	2909
9	3120	240	253	3272
10	3492	156	183	3636
	Total:	*2368*	*2425*	

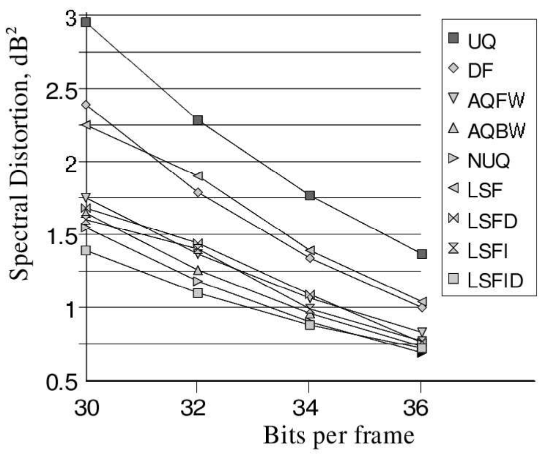

Fig. 7. A plot of average spectral distortion due to nine different methods of LSP quantization for allocations of between 30 and 36 quantization bits per frame.

However, experimental evidence reveals that dynamically choosing one or other of the two quantizer bases (frame-to-frame differential, or distance-from mean) for representing quantized LSP data can improve performance further [16]. In practice this implies that the quantized parameters are calculated using both methods for each analysis frame, and that the method which results in the closest match is then chosen. A single bit from the quantization bit pool must be reserved to flag which method is in use for a particular frame. The best method can be determined through total frame quantization error, or from a more perceptually relevant measure such as spectral distortion [10].

In a standard CELP coder employing scalar quantization, around 3 or 4 bits are used to quantize each LSP [9][17], or around 30 to 40 bits in total for a 10^{th} order representation. A number of techniques have been published in order to allow representation with fewer bits at equivalent levels of quality. A survey of published results can assess the spectral distortion associated with a number of different coding methods at between 3.0 and 3.6 bits per LSP. This is shown in Fig. 7 which plots the spectral distortion due to quantizing a 10^{th} order LSP representation using the following methods:

- Uniform quantization (UQ) [18] also referred to as LSF scalar quantization (LSF) [12].
- Differential quantization (DF or DQ) [18] also known as differential line spectral frequencies (LSFD) [12][19].
- Adaptive forward sequential quantization - adaptive quantization, forward direction (AQFW) [18].
- Adaptive backward sequential quantization - adaptive quantization, backward direction (AQBW) [18].
- Non-uniform quantization (NUQ) [18].
- Dynamically interpolated quantization - interpolated line spectral frequencies (LSFI) [12].
- Dynamically interpolated differential quantization - interpolated, differential line spectral frequencies (LSFID) [12].

From Fig.7 it can be seen that all methods improve at a similar rate as bit allocation increases, and that differential methods are better than non-differential methods. Furthermore non-uniform quantization improves over uniform quantization. For near-transparent quantization at a spectral distortion of 1dB, it seems that more than 33 quantization bits are required even for the best of the schemes shown, and for several extended combinations [4].

In order to extract greater compression levels with higher quality, researchers turned to vctor quantization (VQ) in which a candidate vector, comprising the set of parameters to be quantized, is compared in turn to each entry in a codebook of equal sized vectors. The index of the codebook entry nearest to the candidate vector (in a Euclidean or perhaps more perceptual sense), is then transmitted to represent that set of parameters. This was initially applied directly to LPC values [20], then to hybrid and cascaded methods before being applied to LSPs by Pailiwal and Atal [13] who achieved near-transparent 1dB spectral distortion quantization at only 24 bits per frame. This compares to at least 33 bits per frame in the scalar quantization methods above. The 24 bits per frame result relied upon a split-VQ technique to quantize sets of lines.

Split VQ divides the set of LSPs into two or more sub-vectors which are compared to different codebooks. A considerable amount of research has been performed to determine optimal vector and codebook splits, with a typical split for a 10^{th} order system being into three subvectors of 3 lower, 3 middle and 4 upper LSPs respectively. In fact this has been shown to be optimal for a 3-subvector system, whilst a 4-6 split is optimal for a 2 subvector system, such as that in the Paliwal and Atal system [13]. Note that such a split is based on separating the vector into frequency regions, something that could also be accomplished at a higher computation cost through a frequency-specific distance measure.

VQ is well known to improve over scalar quantization methods [9] and is thus the method of choice for CELP-based speech compression systems.

4 Understanding and interpreting LSPs

Moving away from the traditional quantization-led rationale for using LSPs, it is instructive to consider their properties, with the aim of developing analysis and measurement techniques based upon LSP data. As explained previously, LSPs exist in the frequency domain, derived from an interconnected tube model of the human vocal tract. Thus it is unsurprising that they relate to speech features, however, in some cases this may not be immediately obvious from the equations describing them. This section will review the more obvious relationships before developing and discussing some higher-order features, for use in both single-frame, and consecutive-frame analysis.

Firstly, spectral peaks are usually bracketed quite closely by LSP line pairs. Furthermore, the degree of line closeness tends to relate to the sharpness of the underlying spectral peak, and its amplitude. Note in Fig. 1 that the three tallest spectral peaks correspond to the three narrowest pairs of LSPs (between lines {1:2}, {5:6} and {7:8} respectively), and that ordering this set by line separation matches the bandwidth ordering of the corresponding spectral peak. In other words, wider bandwidth spectral peaks tend to be described by wider spaced line pairs, and narrower spectral peaks by more narrowly spaced line pairs.

Secondly, since the LSPs in a CELP coder loosely represent the shape of the human vocal tract, it is reasonable to expect that potential exists for speech analysis [21] and processing [22] based upon this. For example, detection or classification of speech can be performed, especially when co-classified with other available parameters (for example, gain and pitch), or more traditional analysis features [22].

LSPs can be interpreted instantaneously, or continuously. The former is based on their intra-frame distribution, and the latter based on their inter-frame dynamics. Such analysis is discussed in sections 5 and 6 respectively.

Finally, there is the possibility of altering the location of single or multiple lines to effect deliberate changes in the underlying spectrum. As LSPs tend to cluster in pairs around resonant frequencies, and if resonant frequencies can usually be associated to formants (at least in voiced speech), then such actions would naturally target speech formants. In other words, this would be a method of formant manipulation, discussed further in section 7.

Use of speech codec transmission parameters such as LSPs for speech analysis and processing has implicit advantages. If we consider an example of a typical CELP coder:

- The transmitted gain, LSP, pitch delay/multiplier and codebook index have lower combined bit-rate than the raw speech, thus require fewer operations to measure or process.
- Coded parameters display some degree of orthogonality, having been engineered to represent speech as efficiently, and therefore as uniquely as possible. Although not necessarily true for all speakers and all coded material, the ideal of orthogonality between parameters assists in the measurement/adjustment of a basic set of partitioned speech features.
- At the decoder, parameters are available prior to the reconstructed speech. Parameter analysis conducted in parallel with decoding will therefore yield information in advance of reconstructed speech without imposing additional latency.

5 Instantaneous LSP analysis

LSP-based speech analysis is the process of extracting usable and interpretable information from the LSP data representing a frame of speech. This can be through static analysis of a single frame, or through examination of the time-evolution of analysis measures.

In this section, we consider the static extraction of three features for individual analysis frames of speech, whereas section 6 will extend to the evolution of the measures in time. The following equations present three instantaneous LSP measures, in which subscript i refers to one of the p LSFs ω_i which represent speech in analysis frame n.

The overall shift in LSP location from one frame to the next, $Shift$, indicates predominant spectral movements between frames [23]:

$$Shift[n] = \left\{\sum_{i=1}^{p} \omega_i[n]\right\} - \left\{\sum_{i=1}^{p} \omega_i[n+1]\right\} \tag{18}$$

The mean frequency of all LSPs representing an analysis frame is also a useful measure of frequency bias, $Bias$:

$$Bias[n] = \frac{1}{p}\sum_{i=1}^{p} \omega_i[n] \tag{19}$$

It is also possible to specify a reference LSP distribution, $\bar{\omega}_i$, and calculate a deviation Dev between this and the current LSP set, raised to an arbitrary power β, as follows:

$$Dev[n] = \sum_{i=1}^{p} (\omega_i[n] - \bar{\omega}_i)^{\beta} \tag{20}$$

where $\bar{\omega}_i$ could be a set of nominally positioned LSPs located at evenly spaced locations:

$$\bar{\omega}_i = i\pi/(p+1) \qquad for \quad i = 1 \ldots p \tag{21}$$

This has the p lines distributed evenly across the spectrum. If this LSP distribution were transformed into LPC coefficients and their power spectrum calculated, it would be flat. Dev thus determines how close each frame is to this distribution, such that with $\beta = 2$, it becomes the Euclidian distance between the actual and comparison distributions. Odd values such as $\beta = 1$ or $\beta = 3$ attributes a sign to the deviation from $\bar{\omega}_i$. In this case, a positive measure denotes a high frequency spectral bias or tilt, and a negative measure a low frequency spectral bias.

Each of these measures provides useful information regarding the underlying speech signal, and are illustrated when applied to a speech recording from the TIMIT database [24], in Fig.8 (the deviation plot is given for $\beta = 1$).

$Shift$ indicates predominant LSP frequency distribution shifts between consecutive frames. Considering an example of two adjacent frames containing unvoiced and voiced speech, the LSP distributions in the two frames will be low frequency, and high frequency biased respectively. This will yield a large measure value, indicating a sharp transition between speech types. The gross measure as shown in Fig. 8, peaks at obvious speech waveform changes, and can thus be used for speech segmentation [23].

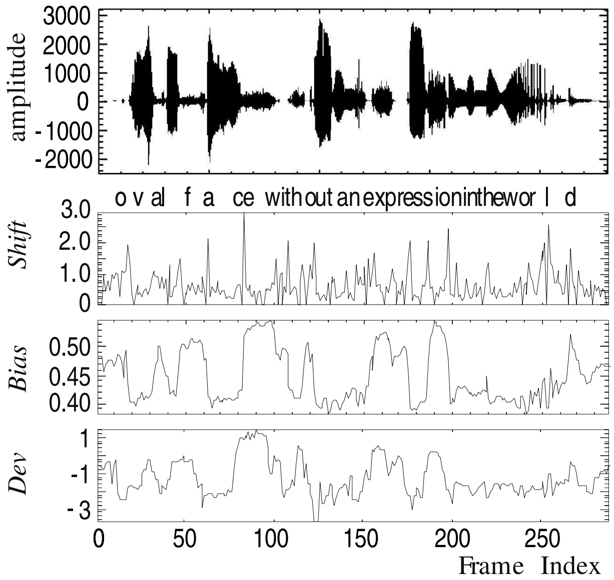

Fig. 8. Gross, average and deviation LSP analysis features of a speech utterance, with waveform as plotted, extracted from the TIMIT database.

$Bias$ is an indication of the frequency trend within the current frame - that is whether the spectrum is predominantly high frequency or low frequency. This is logically similar to Dev in (20) which determines how close the LSP distribution of the current frame is to a predetermined comparison distribution. In Fig. 8 this registers high values for fricatives, indicating the predominance of their high frequency components [25].

Where the speech is contaminated by noise of a particularly well-defined shape, if the comparison distribution, $\bar{\omega}$, of (20) is set to represent a spectrum of this shape, then the Dev measure will be relatively noise insensitive when averaged. In other words analysing the noise itself will produce a zero mean output.

It is also possible to use LSP data to estimate the position of spectral peaks within an analysis frame. As discussed in section 1, peaks are located approximately halfway between pairs of closely spaced lines, with the peak power related to the closeness of the lines. Typically, ordering the first few most closely-spaced lines in terms of narrowness, will correspond to the ordering of the corresponding peaks by power. It must be noted, however, that in some cases, especially where three lines are close together, the correspondence is less predicable. For unvoiced frames, other speech resonances are similarly reflected in the LSP distribution, although the visual correspondence when plotted is far less dramatic than in the case of strongly voiced speech.

6 Time-evolution LSP analysis

Each of the static measures can be tracked to build up a statistical model of the signal under analysis. Similarly, spectral peaks detected by a measure of line-closeness, can be tracked and post-processed to extract likely formant tracks. Analysis of LSP feature changes over time may be combined with consideration of traditional speech features to determine dynamic speech characteristics.

Applications of speech feature tracking can include the classification of speech analysis frames into voiced, unvoiced or non-speech periods. There is evidence to suggest that simple instantaneous speech features derived from LSP analysis can be complementary to frame power measures, but are similar under low noise conditions, to zero-crossing rate [25].

One simple example of speech classification uses a binary LSP decision measure of upward or downward deviation. This LSP vote measure is equivalent to counting the number of lines in a frame above their median value, and provides a computationally simple alternative to (20). The measure has been used by the author to classify speech into phonetic groupings [25]. In this application the LSP vote measure categorises speech analysis frames into six basic classes of voiceless fricatives and africatives, fricatives, plosives, nasals and glides and silence. Whilst the technique shows promise, the long-time averaging required for the classification of features would generally limit its usefulness in the classification of individual utterances, or words, unless combined with alternative features such as frame power.

However, by far the greatest indication of the potential of continuous LSP analysis is in the speech recognition field. As early as 1990, Paliwal had investigated the use of line spectral frequencies as classification features for hidden Markov model (HMM)-based speech recognition systems [26], and found them to be significantly better in terms of recognition score than the common cepstral coefficient representation. Many other authors have used LSPs for speech recognition, often finding advantages in LSP use over and above other parameters. In particular Palomar and Fukuda [27] found the LSP to slightly outperform other LPC-derived features, such as the LPC-cepstrum, and the postfilter-cepstrum. Choi et.al [28] used a Mel-scale pseudo-cepstrum, a summation of the line spectral frequencies on a cosine Mel-scale. This significantly improved recognition performance over either raw LSPs or quantized LSPs. This transformation of raw LSFs was further advocated by Peláez-Moreno et.al [29] who stated in 2006 that "LSP have proved to be unsuitable for ASR, and they must be transformed into cepstrum-type parameters". Even transformed in some way, LSP-based measures continue to be chosen among feature-vectors for automatic speech recognition

In general, speech recognition uses techniques such as template matching, statistical modelling or multi-feature vector quantization. It is possible to use LSP-derived data with each of these techniques, although the statistical matching of the HMM [26] appears to be a more common application for LSP-derived features. One interesting alternative is a vector-quantization based approach, acting upon mean-corrected LSPs, similar to the *Dev* measure of (20), with a dynamically updated comparison distribution, $\bar{\omega}_i$ derived from the average position of each line over a set of previous analysis frames [30].

Paliwal [21] provides one of the better descriptions of an LSP-based feature vector, and compares this to the cepstral coefficient representation, for use with a generic maximum-likelihood classifier (designed to represent the class of statistical, in particular HMM-based, systems). Through many experiments, he concludes that, although instantaneous frame analysis based on the two alternatives leads to very similar results, time-evolution analysis using LSPs is better. However, it should be noted that Paliwal's work predated the publication of Mel-scale pseudo-cepstrum techniques [28], and related advances in the post-processing of cepstral coefficients, both of which will tend to improve the respective results [29].

Finally, continuous LSP analysis has been used to collect statistics on the line frequency distribution over time in a CELP coder and relate this to the language being conveyed. It has been shown that LSP statistics differ markedly across several languages [31], leading to the potential for LSP-based features to be used in linguistic detection algorithms.

7 LSP modification

When used in speech compression systems, LSPs will be quantized prior to transmission or storage. It is thus common to find that decoded LPC coefficients differ from the originals in some way. The use of an intermediate LSP representation for quantization ensures that LPC filter instabilities do not result from this quantization error.

A comparison of the original and final spectra shows differences in the immediate frequency regions of the lines that were changed most. From these observations it has been found possible to alter the values of particular LSPs to change the underlying spectral information which they represent. Methods used to achieve these alterations are shown to have great potential for enhancement of speech in the presence of noise.

To illustrate some of the modifications possible, Fig. 9 plots the original spec-

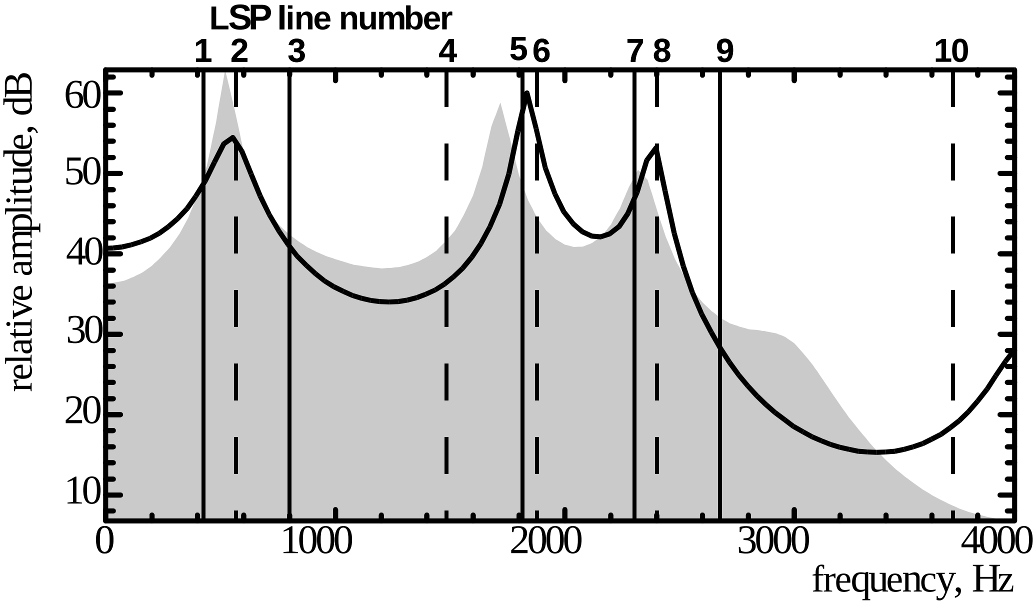

Fig. 9. The altered set of LSPs, and the resulting LPC spectrum are plotted over the original spectrum, which is shown as a grey area.

trum of Fig. 1 in grey, and a power spectrum derived from an altered set of LSPs as drawn. The LSP adjustments made to cause these spectral changes were namely:

> The separation of line pair {1:2} has been increased, resulting in a wider, lower amplitude spectral peak between them.
> The separation of line pair {5:6} has been decreased, and the pair has also been translated slightly upward in frequency, causing a sharper peak between them, now at a higher frequency.
> Line 10 has been moved closer to the Nyquist frequency of 4kHz, inducing a spectral peak at that frequency.

All of the line alterations shown were performed manually.

It was probably Paliwal again [21] who first reported that the effects, on the underlying spectrum, of modifying a line are predominantly confined to the immediate frequency region of that line. However amplitude changes in one region will always cause compensatory power redistribution in other regions. Despite this, as long as line alterations are minimal, the effects on other spectral regions can be limited. This is saying that, for small movements, and small movements only, localised spectral adjustments can be made through careful LSP manipulation.

The example of Fig. 9 shows a spectrum of voiced speech. The three spectral peaks represent formants, and as such we can see that the operations we performed have affected those formants. In fact, LSP operations have demonstrably altered formant bandwidths and positions.

The LSP operations to derive the changes shown in Fig. 9 can be formalised as follows [22]. If ω_k are the LSP frequencies and ω'_k the altered frequencies, then narrowing line pair {k:k+1} by degree α would be achieved by:

$$\omega'_k = \omega_k + \alpha(\omega_{k+1} - \omega_k) \tag{22}$$

$$\omega'_{k+1} = \omega_{k+1} - \alpha(\omega_{k+1} - \omega_k) \tag{23}$$

and increasing the frequency of line k by degree γ may be achieved with:

$$\omega'_k = \omega_k + \omega_k(\gamma - 1)(\pi - \omega_k)/\pi \tag{24}$$

When altering the line positions it is important to avoid forming unintentional resonances by narrowing the gaps between lines that were previously separated. This problem may be obviated either by moving the entire set of LSPs, or providing some checks to the adjustment process. In the former case, movement of lines 1 and 10 closer to angular frequencies of 0 and π may also induce an unintentional resonance. Eqn. (24), designed for upward shifting, progressively limits the degree of formant shift as a frequency of π is neared. A similar method may be applied to downward shifting.

Adjusting lines in this way alters the frequency relationship between any underlying formants, and therefore will tend to degrade the quality of encoded speech.

A perceptual basis for line shifting has also been introduced [32] to minimise quality degradation through adjustments that may be uneven in the perceptual sense. In this scheme, frequencies are altered by constant Bark. If B_k is the bark corresponding to frequency ω_k, then that line shifted by degree δ is:

$$\omega'_k = 600 sinh\{(B_k + \delta)/6\} \tag{25}$$

Furthermore, a hard limit must still be applied to prevent LSP values approaching or even exceeding an angular frequency of π. Fig. 10 illustrates the upward shift factor applied to LSPs by (24) and (25) with $\gamma = 1.5$ and $\delta = 1$, with a progressive linear cutoff applied to bark shifts on lines located above 3kHz. These values were chosen through subjective listening tests to provide an obvious demonstration of shifting without degrading audio quality excessively. This degree of shifting was also applied to the middle formant resonance in the plot of Fig. 9.

The methods of LSP adjustment described here have been successfully applied in the intelligibility enhancement of speech [22].

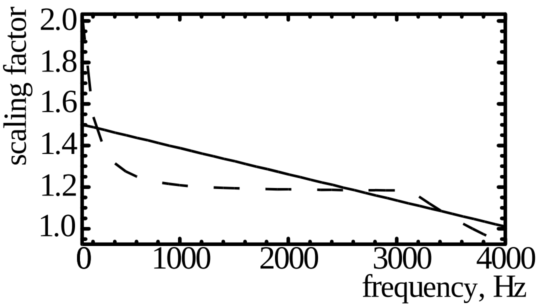

Fig. 10. Degree of LSP upward shift plotted against frequency for a linear shift of $\gamma = 1.5$ derived from (24) as a solid line and a bark-domain shift of $\delta = 1$ from (25) as a dashed line.

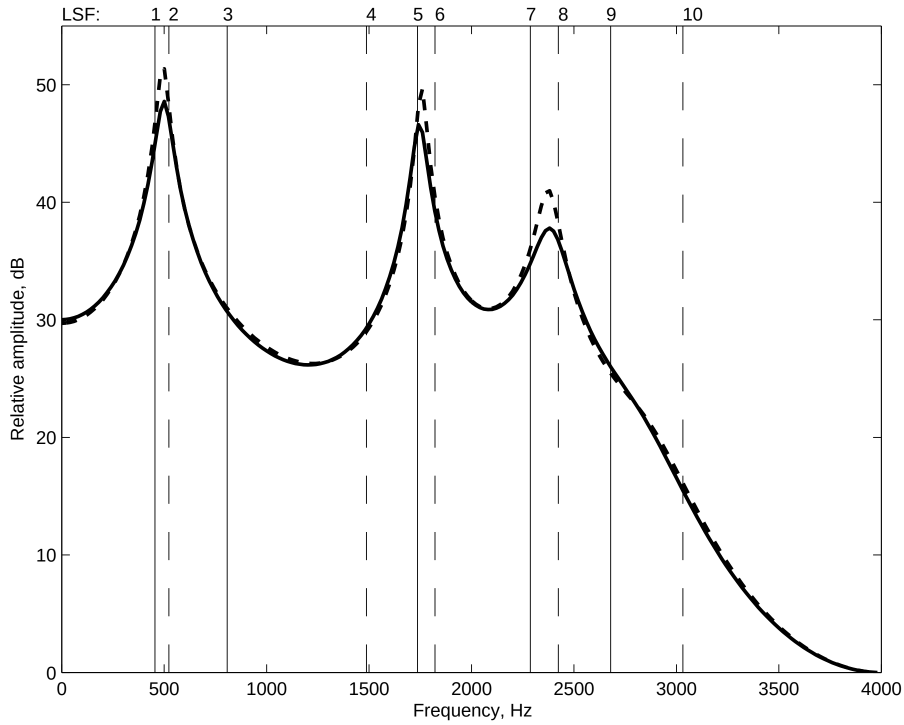

Fig. 11. The result of applying an LSP-based bandwidth altering function to the spectrum of Fig. 1 The LPC power spectrum is drawn solid and the original drawn dashed. The new LSP positions are also plotted.

In particular, the use of LSP shifting for altering the balance of spectral power between formant peaks and valleys has shown promise. The plot of Fig.11 illustrates the effect of increasing the separation between the three most closely-spaced line pairs by 20%. This is shown in the resultant power spectrum (solid) compared to the original (dashed). The bandwidth-altering adaptive filter of [33] performs the same task, again, at a much higher computational cost than the LSP process [22].

Further applications of LSP adjustment may include the quality enhancement of speech/audio and voice masking. In one intelligibility enhancement test

[32], spoken vowels /a/ and /o/ were recorded to computer, and then framed by identical plosive delimiters. Copies of the composite words (i.e. plosive-vowel-plosive) were processed by LSP-based formant narrowing, and LSP-based formant shifting corresponding to $\gamma = 1.5$ in (24) and $\alpha = -0.7$ in (22) and (23) respectively. In all cases overall speech energy was maintained.

Each of the three conditions - non-processed, shifted and narrowed - were then occluded by simulated vehicle interior noise (predominantly low frequency) at each of three levels of SNR, and presented in random order to a listener in an anechoic chamber. A two-alternative forced choice test was conducted to determine the ability of listeners to correctly identify the vowels, and the test repeated for 18 listeners. Over more than 1000 test words, it was found that formant shifting and bandwidth adjustment improved intelligibility under the tested conditions by 15% and 21% respectively, to a 95% confidence level [32].

Further tests identified the ability of the formant shifting technique to dynamically shift speech formants away from single-tone interfering noise, and even to be adapted successfully to account for additive noise based on three common police sirens: *wailer, yelper* and *two-tone* [23].

These techniques have been applied in a digital radio communications system [34] based on the TETRA [7] standard. In this application, a radio handset samples the local background noise to determine spectral characteristics. It then adjusts LSP data being received, and about to be decoded into speech, to change formant bandwidths and positions. This technique shifts formant locations up to 6% to avoid interfering noise, and 'sharpens' formants found to be below the threshold of audibility as determined by a psychoacoustic model. The system has been evaluated using the Diagnostic Rhyme Test (DRT) of ANSI S3.2-1989 [35], and found to improve intelligibility under conditions of low SNR.

Apart from modification of existing LSPs, they have also been used for synthesis of artificial speech, for reasons of representational efficiency - i.e. word prototypes can be stored efficiently as LSPs, and potentially modified to adjust speech characteristics during replay. As early as 1990, Nippon Telegraph and Telephone had developed a working speech synthesis system employing LSPs [36]. For many years this system provided speech response services for automated call-handling systems at Japanese banks. The NTT system utilised compressed LSP sets to recreate rule-based speech from consonant-vowel-consonant syllable templates of Japanese speech.

8 Non-speech applications

While by far the most significant use of LSPs has been in speech compression, with a notable extension into the field of speech and speaker recognition, these techniques have found limited application elsewhere.

Firstly, several groups have investigated musical instrument recognition and coding. Krishna and Sreenivas [37] evaluated three methods of recognising a set of individual instruments with the line spectral frequency method is shown to be the superior method among those compared. The authors state several advantages of LSPs as being the localised spectral sensitivities, the fact that they characterise both resonance locations and bandwidths (which are characteristics of the timbre of the instruments), and the emphasis on the important aspect of spectral peak location.

To investigate further, a recording was made of a violin open A string (tuned to 440Hz), and analysed using LPC. A sample rate of 16kHz was used with a 16-bit resolution. Various orders of analysis were attempted as shown in Fig. 12, ranging from 12 to 48 coefficients, revealing greater and greater levels of spectral detail. The plot in the case covers a 4096 sample selection, beginning mid-recording, and containing an uninterrupted upward bowing action of regular characteristic. For each of the analyses, corresponding LSP locations have been found and overlaid on the power spectrum using the *lpcsp.m* MATLAB script given in Appendix A. Note in each case the good correspondence of the LSP positions to spectral peaks - in particular the location of narrow line pairs around many major peaks. As more detail is revealed through increased order analysis, the harmonic structure of the played note becomes clearer, although a triple (rather than pair) at the fundamental of the 48^{th} order plot probably indicates that the analysis order is greater than the note complexity warrants.

A plot showing the played note is shown in Fig. 13, where the upper graph displays the 36^{th} order LSP values as they evolve over time (and where the lines do not cross since they are monotonically ordered), for each 512 sample analysis window. The lower plot shows a spectrogram of the signal using a 256 sample window size with 50% overlap. It is interesting to note the LSP placement with respect to the obvious 440Hz fundamental and harmonics shown in the spectrogram. Furthermore at the end of the played note, some resonance continues to sound a 440Hz fundamental but with almost no harmonics (since the abrasion of the rosin-coated horsehair bow on the aluminium-clad string has ceased). During this resonance period the upper LSPs gradually lose their tendency to narrow into pairs. However the fundamental continues to be marked by the narrow spacing of the lower two lines.

The obvious visual correspondence between the spectrogram and the LSP

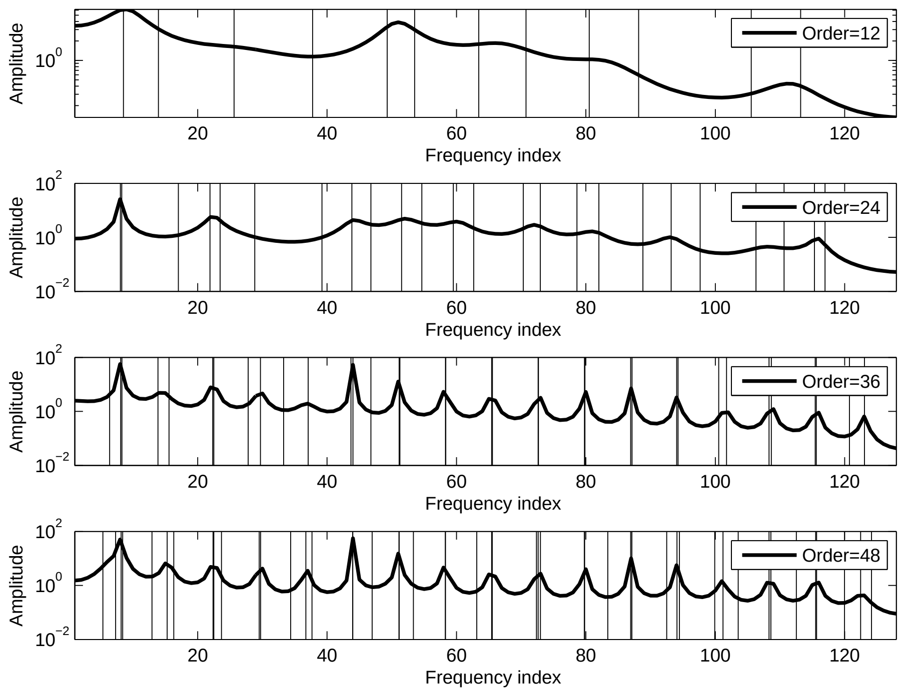

Fig. 12. LSP locations and LPC power spectrum plotted for 4096 samples of violin open A string sampled at 16kHz and analysed with 12, 24, 36 and 48 coefficients respectively against 128 frequency index points spanning the range from 0Hz to 8kHz

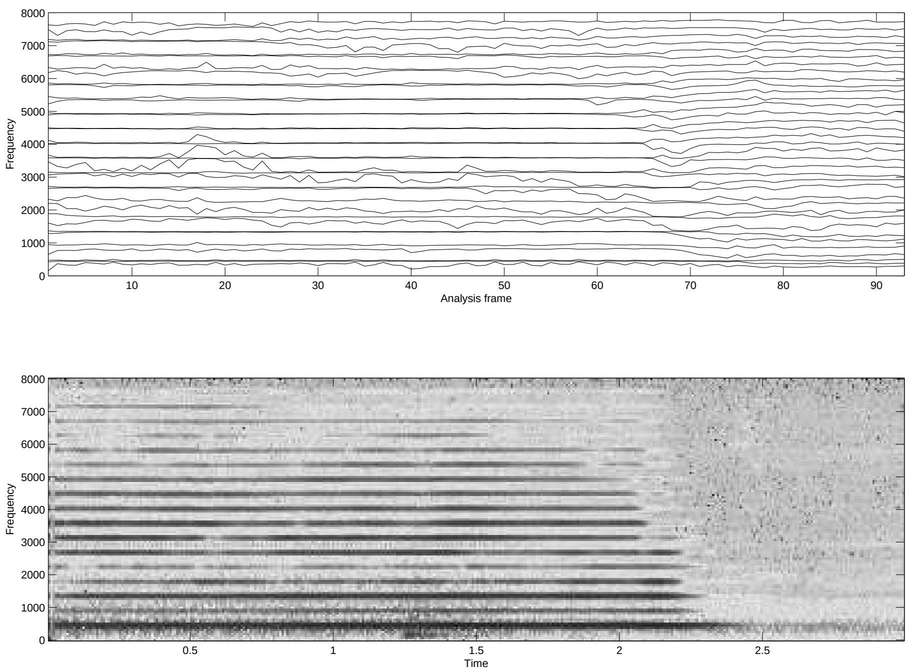

Fig. 13. LSP tracks for a violin note played on the open A string sampled at 16kHz (top) compared to a spectrogram of the same recording (bottom)

evolution of Fig.13 supports the secondary conclusion of [37] in that LSP values are usable for instrument recognition. It is also likely that both fundamental and harmonic frequency identification is similarly achievable.

In a second example of LSP use outside the speech coding field, a recording of birdsong from the BBC [38] was analysed in a similar fashion. In this instance, 18^{th} order LPC analysis was found to be optimal for LSP track plotting. The LSP evolution over time was determined for the approximately 4.5 seconds of a Blackbird's warble. The results, plotted in Fig. 14 indicate a very clear visual identification of the effect of the birdsong syllables shown in the spectrogram and on the plotted LSP trajectories. In this case, the analysis used 256 sample 50% overlapped windows for the spectrogram, and 256 non-overlapped windows for the LSP tracking. Harmonic analysis has been shown useful in the classification of birdsong [39], and the potential for LSP analysis to be used in this area is clear: the spectrogram may show features clearly to the human eye, but is non-trivial to analyse automatically, and due to both the large volume of data involved, and the Fourier Transform necessary in its calculation, is less likely to be usable in a real-time scenario. LSP data, by contrast, exists in a reduced form as a set of 10 to 30 integers.

In general, the research trend over recent years has been for LSP data to form part of the feature set applied to neural network classifiers. This applies across the board from speech recognition, speaker recognition, through musical instrument identification, animal noise identification and even to financial market analysis. In many cases, the LSP representation provides a near-minimal data set for subsequent classification.

9 Implementation Issues

Since LSPs represents spectral shape information at a lower data rate than raw input samples, it is reasonable that judicious use of processing and analysis methods in the LSP domain could lead to a complexity reduction against alternative techniques operating on the raw input data itself. A typical speech coder sampling 12-bit data at 8kHz would have a 96kbits/s input data rate but perhaps only a 1.5kbit/s LSP data rate (at 3 bits quantizing each of 10 lines). For this reason, analysis or processing performed directly on LSPs will generally require fewer operation per second than when performed on raw speech frames.

For example, the method of LSP-based spectral sharpening discussed in section 7 requires only $p + 5$ multiply-accumulates, and $p - 1$ comparison operations when performed in a p^{th} order system [22]. By contrast, the original LPC-based processing method [33] requires $2Np$ multiplications on an N sam-

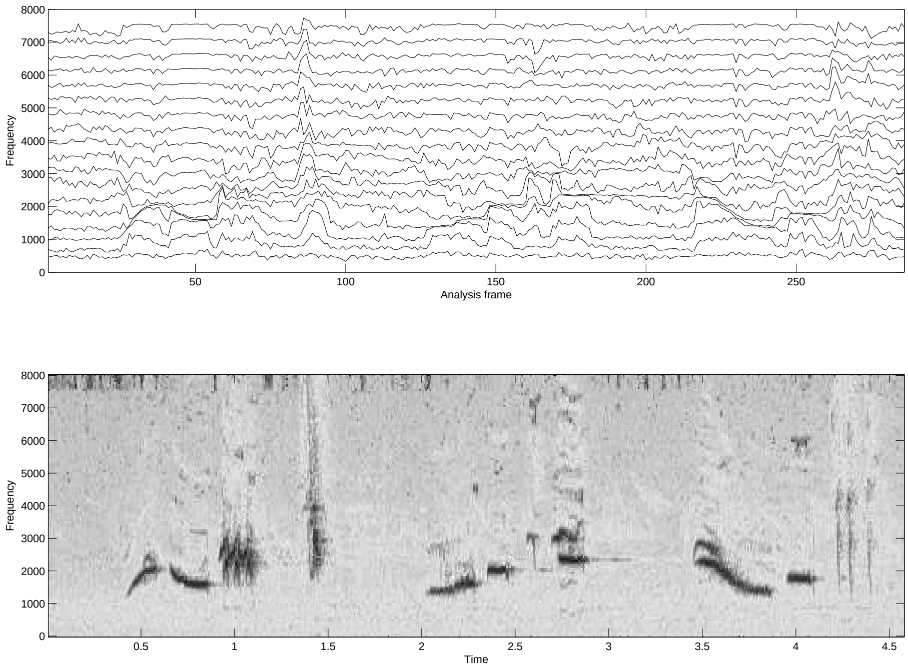

Fig. 14. LSP tracks for a few seconds of a Blackbirds' song sampled at 16kHz (top) compared to a spectrogram of the same recording (bottom)

ple analysis window. For 10^{th}-order analysis operating on 240 sample frames, the LSP adjustment may be between 40 and 400 times more efficient than the adaptive filter.

However the complexity of LSP-generation from raw speech is such that complexity advantages only hold true when LSP data is already available, such as in many speech coders. The generation of LSP parameters can be accomplished using several methods, ranging in complexity [1][7][5][40][8][41][19][42]. The heart of the problem lies in finding the roots of the P and Q polynomials defined in (3) and (4), which can be through standard root solving methods, or more esoteric methods, and is often performed in the cosine domain [7]. When cosine or other transformed parameters are used, conversion may be required before applying the analysis and processing described in this paper.

Saoudi et. al. [5] analysed the complexity of each step in the calculation of LSPs, and considered several alternative methods of reflection coefficient calculation, polynomial solving, and eigenvalue computation. His results are given as a function of analysis order and window size.

Grassi et. al. [41] also compared the computational complexity of standard LSP derivation methods: specifically the methods of calculation via Chan's reflection coefficient recursion [40], via Kabal et. al.'s Chebyshev polynomial expansion [8] and also Saoudi et. al.'s analysis. The detailed results indicate the number of multiplications, additions, divisions and square roots involved in each step for a 10^{th} order system operating on 240-sample analysis frames, and show that, following the calculation of autocorrelation coefficients, the total number of floating-point operations required to derive a set of 10 LSPs is of the order 1200 to 8000, with the refined Chebyshev polynomial method being most efficient.

10 Conclusion

Line Spectral Pairs occupy an important role in the transmission of vocal tract information from speech coder to speech decoder with their widespread use being a result of their excellent quantization properties. They have been a popular research topic for many years, especially related to implementation complexity and quantization efficiency. In recent years, many publications have addressed the analysis of LSP information, particularly with regard to speech recognition. However the modification of LSP values has received little research attention, despite the very promising low-complexity nature of such modifications.

This paper described the common use of LSPs in speech coders, and extended the discussion to the analysis and processing of line spectral pairs, and to use for the analysis of non-speech data. The advantages of using LSPs for both analysis of speech and for the alteration of speech have been discussed, and issues such as computational efficiency, representational space efficiency and stability were included.

It is clear that LSP use is widespread and the representation remains an active research area for speech and audio analysis, as well as a promising investigative tool for non-speech analysis and processing.

A MATLAB scripts

Three scripts are given. The first two convert bidirectionally between LPC and LSP for an even order system, and the final one draws the spectral overlay plot as shown in Fig. 1. In MATLAB, LPC coefficients of order p can be found directly from an audio vector using the autocorrelation method with the function *a=lpc(audio, p);*.

```
%This function takes an array of LSP values as input
%and transforms these to LPC values.
%Note: number of LSPs must be EVEN
%
```

```
%Ian McLoughlin, 10/10/96
%
function a=lsp_lpc(theta)
p=length(theta);
q=cos(theta(1:p));
f1(10)=1;  f1(9)=0;
for i=1:p/2
        f1(10+i)=-2*q(2*i-1)*f1(10+i-1) + 2*f1(10+i-2);
        for k=i-1:-1:1
                f1(10+k)=f1(10+k) - 2*q(2*i-1)*f1(10+k-1) + f1(10+k-2);
        end
end
f2(10)=1;  f2(9)=0;
for i=1:p/2
        f2(10+i)=-2*q(2*i)*f2(10+i-1) + 2*f2(10+i-2);
        for k=i-1:-1:1
                f2(10+k)=f2(10+k) - 2*q(2*i)*f2(10+k-1) + f2(10+k-2);
        end
end
f1b(1)=f1(11)+1;
f2b(1)=f2(11)-1;
for i=2:p/2
        f1b(i) = f1(10+i) + f1(10+i-1);
        f2b(i) = f2(10+i) - f2(10+i-1);
end
for i=1:p/2
        a2(i)       = 0.5*( f1b(i) + f2b(i) );
        a2(i + p/2) = 0.5*( f1b((p/2)-i+1) - f2b((p/2)-i+1) );
end
a=[1,a2];

%This function takes an array of LPC values as
%input and transforms these to LSP values.
%The order must be EVEN (ie a=[1 ,even number]);
%
%Ian McLoughlin, 10/10/96
%Here is an example value for a from a speech frame:
%a=[1;-1.618682;2.317888;-2.955535;2.88624;-2.533107;\
     2.22996;-1.327088;0.988561;-0.612581;0.23539];
%
function lsp=lpc_lsp(a)
p=length(a);
%derive the coefficients for P'(z) and Q'(z)
A(1)=1; B(1)=1;
for k=2:p
```

```
        A(k)=(a(k) - a(p+2-k)) + A(k-1);
        B(k)=(a(k) + a(p+2-k)) - B(k-1);
end
r1=roots(A);
r2=roots(B);
for k =1:p-1
        if (real(r1(k)) < 0)
                theta1(k)=pi-abs(atan(imag(r1(k))/real(r1(k))));
        else
                theta1(k)=abs(atan(imag(r1(k))/real(r1(k))));
        end
        if (real(r2(k)) < 0)
                theta2(k)=pi-abs(atan(imag(r2(k))/real(r2(k))));
        else
                theta2(k)=abs(atan(imag(r2(k))/real(r2(k))));
        end
end
%Test vectors on 3/11/95 show 'correctness'
p=p-1;
for k=1:p/2
        theta(k)=theta1(k*2);
        theta(k+(p/2))=theta2(k*2);
end
lsp=sort(theta);      %Sort into ascending order
```

```
% This program plots the LPC pwr spectrum on
% the current subplot of the current figure,
% and then overlays the LSPs on to this.
%
% Ian McLoughlin,  11th November 1996
%
function lpcsp(a, lsp)
[HH, FF]=freqz(1, a, 100); %cheating - simpler than eqn.(17)?
semilogy(abs(HH),'m-');
hold on
V=axis;
axis([1,length(FF),V(3),V(4)]);
hold on;
lsc=100/pi;
for lp=1:length(lsp)
        line([1+lsp(lp)*lsc,1+lsp(lp)*lsc], [V(3),V(4)]);
end
hold off;
```

References

[1] A. Goalic and S. Saoudi, "An intrinsically reliable and fast algorithm to compute the line spectrum pairs (LSP) in low bit-rate CELP coding," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, 1995.

[2] S. Saito and K. Nakata, *Fundamentals of speech signal processing.* Academic Press, 1985.

[3] N. Sugamura and F. Itakura, "Speech analysis and synthesis methods developed at ECL in NTT - from LPC to LSP," *Speech Communications*, pp. 213–229, 1986.

[4] N. Sugamura and N. Favardin, "Quantizer design in LSP speech analysis-synthesis," *IEEE Journal Selec. Areas Comms.*, vol. 6, no. 2, pp. 432–440, Feb. 1988.

[5] S. Saoudi, J. Boucher, and A. Guyader, "A new efficient algorithm to compute the LSP parameters for speech coding," *Signal Processing*, no. 2, pp. 201–212, 1995.

[6] S. Furui, *Digital Speech Processing, Synthesis and Recognition.* Marcel Dekker Inc., 1985.

[7] ETSI, "Trans-European trunked radio system (TETRA) standard," *A European Telecommunications Standards Institute (ETSI) standard.*

[8] P. Kabal and R. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 34, no. 6, pp. 1419–1425, 1986.

[9] J. S. Collura and T. E. Tremain, "Vector quantizer design for the coding of LSF parameters," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 29–32, 1993.

[10] I. V. McLoughlin, "LSP parameter interpretation for speech classification," *Proc. 2nd IEEE Int.l Conf. on Information, Communications and Signal Processing*, Dec. 1999.

[11] I. V. McLoughlin and F. Hui, "Adaptive bit allocation for LSP parameter quantization," *Proc. IEEE Asia-Pacific Conf. Circuits and Systems*, no. 231, Dec. 2000.

[12] Q. Zhao and J. Suzuki, "Efficient quantization of LSF by utilising dynamic interpolation," *IEEE Int. Symp. on Circuits and Systems*, pp. 2629–2632, June 1997.

[13] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits per frame," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 661–664, 1991.

[14] D. Chang, S. Ann, and C. W. Lee, "A classified vector quantization of LSF parameters," *Signal Processing*, no. 3, pp. 267–273, June 1997.

[15] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 641–644, 1991.

[16] H. Zarrinkoub and P. Mermelstein, "Switched prediction and quantization of LSP frequencies," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 757–760, 1996.

[17] C. S. Xydeas and K. K. M. So, "A long history quantization approach to scalar and vector quantization of LSP coefficients," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 1–4, 1993.

[18] N. Sugamura and N. Favardin, "Quantizer design in LSP speech analysis and synthesis," *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, pp. 398–401, 1988.

[19] F. K. Soong and B. W. Juang, "Line spectrum pair (LSP) and speech data compression," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, no. 10, pp. 1–4, 1984.

[20] B. H. Juang, D. Y. Gray, and R. M. Gray, "Distortion performance of vector quantization," *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, pp. 294–303, Apr. 1982.

[21] K. K. Paliwal, "On the use of line spectral frequency parameters for speech recognition," *Digital Signal Processing*, vol. 2, pp. 80–87, 1992.

[22] I. V. McLoughlin and R. J. Chance, "LSP analysis and processing for speech coders," *IEE Electronics Letters*, vol. 33, no. 99, pp. 743–744, 1997.

[23] I. V. McLoughlin, *CELP and speech enhancement: PhD Thesis*. Birmingham, UK: The University of Birmingham, 1997.

[24] T. I. and M. I. T., "TIMIT database," *a CD-ROM database of phonetically classified recordings of sentences spoken by a number of different male and female speakers*, disc 1-1.1 1990.

[25] I. V. McLoughlin, "LSP parameter interpretation for speech classification," *Proc. of the 6th IEEE Int. Conf. on Electronics, Circuits and Systems*, no. 113, Sept. 1999.

[26] K. K. Paliwal, "A study of LSF representation for speaker-dependent and speaker-independent HMM-based speech recognition systems," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, vol. 2, pp. 801–804, 1990.

[27] L. R. Palomar, T. Fukuda, and E. P. Dadios, "A comparative analysis of the topological structures of different LPC feature-based speech models," *Int. Joint Conf. on Neural Networks*, vol. 5, pp. 2954–2958, 1999.

[28] S. H. Choi, H. K. Kin, H. S. Lee, and R. M. Gray, "Speech recognition method using quantised LSP parameters in CELP-type coders," *IEE Electronics Letters*, vol. 34, no. 2, pp. 156–157, Jan. 1998.

[29] C. Peláez-Moreno, A. Gallardo-Antolín, D. F. Gómez-Cajas, and F. Diáz-de-María, "A comparison of front-ends for bitstream-based ASR over IP," *Signal Process.*, vol. 86, no. 7, pp. 1502–1508, 2006.

[30] S. M. Ashgar and L. Cong, "Matrix quantization with vector quantization error compensation for robust speech recognition," USA Patent 957 902, Oct. 27, 1997.

[31] J. Parry, I. Burnett, and J. Chicharo, "Linguistic mapping in LSF space for low-bit rate coding," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, vol. 2, pp. 653–656, Mar. 1999.

[32] I. V. McLoughlin and R. J. Chance, "LSP-based speech modification for intelligibility enhancement," *13th International Conference on DSP*, July 1997.

[33] A. Schaub and P. Straub, "Spectral sharpening for speech enhancement/noise reduction," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 993–996, 1991.

[34] I. V. McLoughlin and R. J. Chance, "Method and apparatus for speech enhancement in a speech communications system," *PCT international patent*, no. PCT/GB98/01936, July 1998.

[35] ANSI, "ANSI S3.2," *Method for Measuring the Intelligibility of Speech over Communication Systems*, 1989.

[36] R. Nakatsu, "Anser - an application of speech technology to the Japanese banking industry," *IEEE Computer*, pp. 43–47, Aug. 1990.

[37] A. G. Krishna and T. V. Sreenivas, "Musical instrument recognition: from isolated notes to solo phrases," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, vol. 4, pp. 265–268, 2004.

[38] (2007, May) The British Broadcasting Corporation - BBC Radio 4: Brett Westwood's guide to garden birdsong. [Online]. Available: http://www.bbc.co.uk/radio4/science/birdsong.shtml

[39] A. Harma and P. Somervuo, "Classification of the harmonic structure in bird vocalization," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, vol. 5, pp. 701–704, 2004.

[40] C. Chan, "Computation of LSP parameters from reflection coefficients," *Electronics Letters*, no. 19, pp. 1773–1774, 1991.

[41] S. Grassi, A. Dufaux, M. Ansorge, and F. Pellandini, "Efficient algorithm to compute LSP parameters from 10th-order LPC coefficients," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 1707–1710, 1997.

[42] V. Semenov, A. Kalyuzhny, and A. Kovtonyuk, "Efficient calculation of line spectral frequencies based on new method for solution of transcendental equations," *IEEE Int. Conf. Acoustics Speech and Signal Proc.*, pp. 457–460, 2003.