STOCHASTIC MODELING AND QUANTIZATION OF HARMONIC PHASES IN SPEECH USING WRAPPED GAUSSIAN MIXTURE MODELS.

Yannis Agiomyrgiannakis and Yannis Stylianou

Multimedia Informatics Lab, Department of Computer Science,
University of Crete, Hellas and
Institute of Computer Science, FORTH
{jagiom, yannis}@csd.uoc.gr

ABSTRACT

Harmonic sinusoidal representations of speech have proven to be useful in many speech processing tasks. This work focuses on the phase spectra of the harmonics and provides a methodology to analyze and subsequently to model the statistics of the harmonic phases. To do so, we propose the use of a Wrapped Gaussian Mixture Model (WGMM), a model suitable for random variables that belong to circular spaces, and provide an Expectation-Maximization algorithm for training. The WGMM is then used to construct a phase quantizer. The quantizer is employed in a prototype variable rate narrowband VoIP sinusoidal codec that is equivalent to iLBC in terms of PESQ-MOS, at ~ 13 kbps.

Index Terms— speech analysis, speech coding, source coding, phase coding, transform coding

1. INTRODUCTION

The voiced parts of the speech signal have a quasi-harmonic behavior that is easily observed on the time-frequency plane. The properties and the statistical behavior of the amplitude spectra has been well studied and used in many applications like Speech Recognition, Speaker Identification, etc. On the contrary, phase spectra are usually disregarded, mainly because of the intrinsic difficulties associated with the accurate and robust modeling of phases. However, there are several studies that indicate the importance of phase in speech perception [1].

Understanding and modeling the phase information present in speech signals is a problem related to the "inverse" model techniques that estimate the glottal air flow signal from the speech signal [2] and to techniques which model phase using group delay spectra [3]. These techniques are deterministic in the sense that they rely on a speech production model for justification. This paper presents a stochastic framework to model phases that does not imply am explicit speech production model. The framework allows efficient quantization of the raw phase data of speech harmonics, and has many applications, like Speech Coding for VoIP and Speech Synthesis for small-footprint TTS systems.

Lately, there is a renewed interest in sinusoidal speech coding for VoIP applications, driven by the fact that sinusoidal models are suitable for packet loss concealment [4] and time-scale modifications for adaptive jitter buffer resizing. Sinusoidal (harmonic) speech codecs have shown to have superior quality at low bit rates. These codecs take a source/filter approach and use a *phase model* to reconstruct the phase of the harmonic excitation, like in STC (Sinusoidal Transform Codec) and in MBE (MultiBand Excitation) codecs [5]. In [6], the voiced harmonics are constructed using a Rosenberg glottal pulse model. Another idea is to use all-pass filters to correct the phase response of the minimum phase AR spectral envelope [7]. These model-based approaches work well for low bit-rates (below 4 kbps), but many researchers argue that high-quality sinusoidal speech coding requires the quantization of the phases.

In [8], the *phase residual*, the difference between the phase of the current frame and it's prediction from the previous frame is quantized. Vector quantization of phases was proposed in [9] for the quantization of the harmonic phases of the SEW (Slowly Evolving Waveform) in the context of WI (Waveform Interpolation) coders. However, VQ-based phase quantizers cannot operate at increased bit-rates and, inevitably, the quality of speech is limited. A GMM-based phase quantization algorithm capable of operating at high rates was provided in [4], but the quantizer restricts the GMM to $(0, 2\pi]$. This does not take into account the modulo-2π behavior of the phase data. Phase quantization has also found applications in concantenative sinusoidal Text-To-Speech synthesis. The TTS database in such systems is rather large and small-footprint implementations suitable for low-end terminals requires efficient reconstruction of the harmonic phases. In [10], phase is encoded with 7 bits/harmonic.

In this paper, we decompose the harmonic excitation phases in two terms, a linear phase term and a dispersion phase term. The stochastic behavior of the dispersion phases is well modeled using circular (or directional) statistics [11]; pdfs defined on the surface of the n-Torus manifold which is the extension of the unit circle to multiple dimensions (modulo-2π spaces). We propose to model the dispersion phases using the so-called Wrapped Gaussian Mixture Models (WGMM) which are able to model a wide range of variables that exhibit a modulo-2π behavior. However, only a few recent publications utilize wrapped models to model circular data. In [12], *wrapped Hidden Markov Models* (HMM) are used to track the trajectories of sound sources inside a room. In [13], wrapped (*Normal*, *Cauchy*) mixture models are used to study time series with linear and circular variables.

An Expectation-Maximization (EM) algorithm for wrapped multivariate Gaussians and an extension to HMM is presented in [12] for the case of Gaussian components with diagonal covariance matrices. However, the EM algorithm provided in [12] estimates the parameters by performing the EM steps one dimension at a time. We show that this restriction is not necessary and an EM algorithm for WGMM with diagonal covariance matrices is given in this paper. The WGMM captures the pdf of the dispersion phases, which is used to construct a phase quantizer by employing ideas from GMM-based quantization [14]. The WGMM-based quantizer is then used in a phase quantization scheme that encodes the variable-dimension phase vectors. The proposed method used in a prototype sinusoidal speech codec and evaluated in terms of PESQ-MOS [15] where it is shown that a score of 3.87 can be achieved with an average rate of $\sim$13 kbps for all speech parameters. The extension to wideband signals is not discussed in this paper, but the interested reader could be referred to [16].

The outline of this paper is as following. In Section 2 we present the harmonic model and a method compute the linear phase term. Section 3 presents the WGMM and the corresponding EM algorithm. The construction of a WGMM-based quantizer is presented in Section 4. Finally, Section 5 evaluates the application of this quantizer to narrowband speech coding.

2. HARMONIC MODEL AND PHASE DECOMPOSITION

Harmonic representation is a high quality parametric model used for analysis/synthesis of the speech signal [5]. Signal $x(n)$ is typically analyzed in short intervals called *frames*, where it is assumed to be stationary. Within each frame, the signal is represented as a weighted sum of harmonically related sinusoids:

$$\hat{x}(n) = \sum_{k=1}^{K} A_k \cos(k\omega_0 n + \phi_k), \tag{1}$$

where ω_0 is the fundamental frequency (in radians), K is the number of the harmonics, A_k and ϕ_k are the amplitude and the phase of the k-th harmonic, and n is the time index. The amplitudes and the phases can be obtained using least squares methods [17].

Harmonic amplitudes A_k maybe well represented using a minimum phase RCC (Real Cepstrum Coefficients) spectral envelope with 20 dimensions. The cepstral envelope fits the log-spectra $\log(A_k)$ at the Mel-scale by solving a regularized least squares problem [17].

Let $H_s(\omega)$ be the frequency response of the RCC envelope. The phases ϕ_k are decomposed to a minimum phase term $\angle H_s(k\omega_0)$, a *linear phase term* $k\omega_0\tau$ and a *dispersion term* ψ_k:

$$\phi_k = k\omega_0\tau + \angle H_s(k\omega_0) + \psi_k \tag{2}$$

The phases $\psi_k + k\omega_0\tau$ correspond to the phases of an excitation signal since the subtraction of the minimum phase term corresponds to inverse filtering with the spectral envelope $H(\omega)$. The excitation signal $e(n)$ can be reconstructed according to the formula:

$$e(n) = \sum_{k=1}^{K} \cos(k\omega_o n + k\omega_0\tau + \psi_k). \tag{3}$$

The linear phase term $k\omega_0\tau$ corresponds to the translation of the excitation with respect to a reference point inside the pitch period. As a reference point, we used the peak of the excitation $e(n)$ within a single pitch period. The peak-picking is made to a uniformly sampled version of the excitation $e(n)$, using 128 samples (7 bits). We found that this procedure provided robust reference points within the glottal cycle. The dispersion phases ψ_k, $k = 1, ..., K$ have a distribution that exhibits structure and can be modeled using WGMM. A similar observation was also used in [4] for phase quantization.

3. CIRCULAR STATISTICS AND WRAPPED GMM

Let $\vec{\psi}$ be the vector that holds the phases ψ_k, $k = 1, ..., K$. Vector $\vec{\psi}$ is distributed on the surface of the n-Torus $\mathbb{T}^K = \mathbb{R}^K / 2\pi\mathbb{Z}^K$. The $\mathbb{T}^1$ n-Torus is the unit circle, while $\mathbb{T}^K = \mathbb{T}^1 \times \mathbb{T}^1 \times \ldots \times \mathbb{T}^1$ is the K times product of $\mathbb{T}^1$. The corresponding statistics are called *circular (or directional) statistics* and the random variables $\vec{\psi}$ are called *circular (or directional) random variables* [11].

The statistics of a scalar circular random variable can be captured either by a pdf explicitly defined on the unit circle, like the Von-Mises distribution, or by wrapping the pdf of a linear random variable to the circumference of the unit circle [11]. The pdf of the *linear* Gaussian distribution is:

$$N(\theta; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right). \tag{4}$$

For notational simplicity we will henceforth assume that all circular random variables are confined in $(0, 2\pi]$. The pdf of the scalar *wrapped Gaussian distribution* is given by [11] (pg. 55):

$$N_w(\theta; \mu, \sigma^2) = \sum_{w=-\infty}^{\infty} N(\theta - w2\pi; \mu, \sigma^2), \tag{5}$$

where μ_w and σ_w^2 is the mean and the variance of the wrapped Gaussian. The wrapped-Gaussian pdf can be approximated by the linear Gaussian pdf at small variances $\sigma_w^2 \leq 1$ and by the uniform distribution at large variances $\sigma_w^2 \geq 2\pi$. The wrapped Gaussian pdf is constructed by *infinite wrappings* of the linear Gaussian pdf in the interval $(0, 2\pi]$. In practice, though, a summation over ± 2 tilings provides a sufficient approximation even for large variances σ_w^2.

A pdf for the multivariate wrapped Gaussian distribution can be obtained as a product of scalar pdfs:

$$N_w(\vec{\psi}; \vec{\mu}, \vec{\sigma}^2) = \prod_{k=1}^{K} N_w(\vec{\psi}(k); \vec{\mu}(k), \vec{\sigma}^2(k)). \tag{6}$$

The pdf of the Wrapped Gaussian Mixture Model (WGMM) can then be defined as:

$$p(\vec{\psi}) = \sum_{m=1}^{M} \alpha_m N_w(\vec{\psi}; \vec{\mu}_m, \vec{\sigma}_m^2), \tag{7}$$

where M is the number of Gaussian components and α_m, $\vec{\mu}_m$ and $\vec{\sigma}_m^2$ are the prior probability, the mean and the variances of the m-th wrapped Gaussian component. The pdf assumes that ψ_k are independent random variables. The pdf of a WGMM with full covariance matrices can also be captured by a WGMM with diagonal covariance matrices using more Gaussian components. Furthermore, it leads to algorithms with tractable complexity.

Expectation-Maximization Algorithm

The derivation of the algorithm is omitted due to the lack of space. This section presents the equations that perform a full iteration of the EM algorithm. Let $\vec{\psi}_n$, $n = 1, ..., N$ be N data vectors.

We will define some accessory variables that hold the information related to the expectation step:

$$\delta_{k,m,n,w} = N(\vec{\psi}_n(k) - w2\pi; \vec{\mu}_m(k), \vec{\sigma}_m^2(k)) \tag{8}$$

$$\beta_{k,m,n} = \sum_{w\in\mathbb{Z}} \delta_{k,m,n,w} \tag{9}$$

$$\beta_{k,m,n}^{(\mu)} = \sum_{w\in\mathbb{Z}} \delta_{k,m,n,w} \left(\vec{\theta}_n(k) - w2\pi\right) \tag{10}$$

$$\beta_{k,m,n}^{(\sigma^2)} = \sum_{w\in\mathbb{Z}} \delta_{k,m,n,w} \left(\vec{\theta}_n(k) - \vec{\mu}_m(k) - w2\pi\right)^2 \tag{11}$$

$$\beta_{m,n} = \alpha_m \prod_{k=1}^{K} \beta_{k,m,n} \tag{12}$$

$$\tag{13}$$

We will define some more accessory variables:

$$\omega_m = \sum_{n=1}^{N} \frac{\beta_{m,n}}{\sum\limits_{m'=1}^{M} \beta_{m',n}} \tag{14}$$

The update equations can then be written as:

$$\alpha_m \leftarrow \frac{1}{N}\omega_m \tag{15}$$

$$\vec{\mu}_m(k) \leftarrow \frac{1}{\omega_m} \sum_{n=1}^{N} \frac{\beta_{m,n}}{\beta_{k,m,n}} \beta^{(\mu)}_{k,m,n} \tag{16}$$

$$\sigma^2_m(k) \leftarrow \frac{1}{\omega_m} \sum_{n=1}^{N} \frac{\beta_{m,n}}{\beta_{k,m,n}} \beta^{(\sigma^2)}_{k,m,n} \tag{17}$$

In practice the summation in equations (11),(10),(11) need not be made over the whole $\mathbb{Z}$; only ± 2 tilings are adequate.

4. WGMM-BASED QUANTIZATION

The proposed scheme encodes data vector $\vec{\psi}$ according to the pdf of each of the multivariate wrapped Gaussians. The data $\vec{\psi}$ is quantized with M wrapped multivariate Gaussian *coders* and the "best" quantization is transmitted through the channel along with the corresponding indices. The resulting scheme is a straightforward extension of the GMM-based quantization scheme [14] to WGMM. The design of a wrapped multivariate Gaussian quantizer is the subject of the rest of the section.

Initially, we define a distortion measure suitable for circular random variables. We define the Wrapped-Square-Error (WSE) as:

$$d(\psi, \hat{\psi}) = \min_{w \in \mathbb{Z}} \left\{ \left(\psi - \hat{\psi} - w2\pi\right)^2 \right\}. \tag{18}$$

If both ψ and $\hat{\psi}$ are confined to their principal values in $(0, 2\pi]$, then only ± 1 wrappings are enough in equation (18). The extension of WSE to vectors is straight-forward:

$$d(\vec{\psi}, \hat{\vec{\psi}}) = \sum_{k=1}^{K} d(\vec{\psi}(k), \hat{\vec{\psi}}(k)). \tag{19}$$

Next, we construct a quantizer for scalar circular random variables distributed according to $N_w(\theta; \mu, \sigma^2)$. We construct a circular codebook by wrapping the codepoints of the $N(0, \sigma^2)$ Gaussian codebook around the circumference of the unit circle. This solution works quite well for low variances $\sigma^2 \leq 1$ because the overlapping between the tilled Gaussian components is low but it becomes less accurate in higher variances. Therefore, we constrain the maximum overlapping by restricting the variances to $(0, 2\pi]$ during the training of the WGMM.

Finally, we provide a greedy bit-allocation algorithm. Let R be the rate of the WGMM-based quantizer and $N_m = \lfloor \alpha_m 2^R \rfloor$ be the number of quantization levels is assigned to each of the M components of the WGMM. Within each Gaussian component, the N_m quantization levels are allocated with a greedy algorithm that minimizes the expected component distortion D_m:

$$D_m = \sum_{k=1}^{K} D(N_{m,k}, \sigma^2_m(k)), \tag{20}$$

where $D(N_{m,k}, \sigma^2_m(k))$ is the expected WSE when the k-th variable of the m-th wrapped Gaussian component is encoded with $N_{m,k}$ quantization levels. The minimization is made subject to the rate constrain:

$$\prod_{k=1}^{K} N_{m,k} \leq N_m.$$

When the variances $\sigma^2_m(k) \leq 0.5$, the wrapped univariate Gaussian is well approximated by a linear Gaussian and the well-known distortion-rate formula for linear Gaussians can be used instead [18] (pg. 228):

$$D(N, \sigma^2) = \frac{\sqrt{3}\pi}{2}\sigma^2 N^{-2}. \tag{21}$$

For higher variances, $\sigma^2_m(k) > 0.5$, we use linear interpolation of tabulated distortions, sampled for a wide range of quantization levels and variances. The distortions were computed using 100.000 samples of a wrapped $N(0, \sigma^2)$ and evaluated with the WSE, for quantization levels $l = 1, 2, \ldots, 2^6$ and for densely sampled variances $\sigma^2 = \{0.5, 0.51, 0.52, \ldots, 2\pi\}$.

5. PHASE QUANTIZATION FOR NARROWBAND SPEECH

The presented WGMM-based quantization algorithm was used to quantize the dispersion phases ψ_k of the narrowband speech harmonics below 3700 Hz. Only the phases of voiced frames were quantized, while the phases of unvoiced frames were randomly set. A practical phase quantization scheme has to quantize variable dimension phase vectors. We address this problem by classifying pitch values in 7 classes (continuous intervals) in order to reduce the variability of the dimensions of the phase vectors within each class. Then we use a split-band approach and separate the harmonics to low-frequency harmonics and high-frequency harmonics in order to provide a higher bit-rate to the perceptually important low-frequency harmonics. For each pitch class, a fixed number of low-frequency harmonics are vector quantized using the corresponding low-frequency WGMM. The rest of the harmonics are quantized with a high-frequency WGMM that is constructed on-the-fly for each frame to fit the number of dimensions of the high-frequency harmonics.

Table 1 shows the 7 pitch classes and the corresponding frequency intervals. A WGMM is trained for the low-frequency harmonics of each class. For classes Q1 and Q2, another WGMM is trained for a subset of the higher frequency harmonics (25-th harmonic and above). For every class except Q3 and Q7, the number of harmonics modeled by a WGMM is equal to the minimum size of the phase vectors of that class. For example, class Q1 has phase vectors with sizes ranging between 38 and 52 dimensions, therefore 24+14=38 harmonics are modeled by a WGMM. For class Q3, only the first 24 harmonics are modeled by the low-frequency WGMM. Class Q7 is relatively rare in our training set and we chose to quantize it using a WGMM obtained by truncating the dimensions of the low-frequency WGMM of Q6 to the desired number of harmonics.

The construction of the high-frequency WGMM is made as follows: for pitch classes Q1 and Q2, the WGMM trained for the phases above the 24-th harmonic is *expanded* to the total number of high-frequency harmonics. This is made by replicating the means and the variances of the highest harmonic which is modeled by a WGMM of the specific pitch class. The same strategy is adopted in classes Q3 to Q6 in order to construct the high-frequency WGMM: the means and the variances of the highest harmonics of the low-frequency WGMM are replicated to fit the number of high-frequency harmonics.

The expansion/truncation strategy adopted in this phase quantization scheme is a practical choice dictated by the problem of having

Pitch Class	Pitch Range	Low-Freq. WGMM dims.	High-Freq. WGMM dims.
Q1	70-95 Hz	24	14+
Q2	95-115 Hz	24	8+
Q3	115-142 Hz	24	0+
Q4	142-176 Hz	21	0+
Q5	176-217 Hz	17	0+
Q6	217-250 Hz	14	0+
Q7	250-350 Hz	14-	0

Table 1. Pitch Classes for WGMM-based Vector Quantization of phases. The "+" symbol refers to the variable number of expanded dimensions of the high-frequency WGMM. The "-" symbol refers to the variable number of reduced dimensions of the low-frequency WGMM of class Q6.

variable dimension vector quantization. However, the choices were not made totaly in blind. Some motivation was provided by the following observations: first, the high-frequency phases above 3 kHz have similar statistics, second, the ear is not sensitive to phase distortion in higher frequencies, third, the phases of Q7 have similar statistics with the phases of Q6.

The proposed phase quantization method was evaluated in the context of variable-rate narrowband speech coding. The speech signal was analyzed/synthesized using 20 ms frames with a step of 10 ms (100 frames/sec) using Hanning window. The frames were classified as voiced, transitional or unvoiced. All parameters were quantized. A 20-th order RCC spectral envelope was fitted to the harmonics according to Section 2 and quantized using GMM-based quantization [14]. Unvoiced frames were reconstructed using the RCC envelope, the energy and random phases. Transitional and voiced frames were reconstructed using the RCC envelope, the linear phase term τ, the pitch, the energy and the dispersion phases $\hat{\psi}(k)$. Pitch was quantized with 8 bits, frame energy with 8 bits, the linear phase term τ was quantized with 7 bits, and the voicing condition with 3 bits. The RCC parameters were encoded with 50 bits for transitional frames and 60 bits for unvoiced and voiced frames. Two different cases were examined: **HMCa** with 70 bits for the low-frequency WGMM and 30 bits for the high-frequency WGMM, and **HMCb** with 60 bits for the low-frequency WGMM and 20 bits for the high-frequency WGMM. HMC stands for *Harmonic Model Codec*. Codec HMCa requires an average of 14.2 kbps and codec HMCb an average of 12.9 kbps.

The evaluation was made using PESQ-MOS [15] computed with a test-set of 64 male and 64 female utterances. As a baseline, we also examined the analysis/synthesis system (unquantized parameters and amplitudes sampled from the RCC spectral envelope) (**AS-RCC**) and the iLBC [19] codec (**iLBC**) at the 20 ms mode (15.2 kbps). The results are shown in Figure 1. We can observe that the 12.9 kbps HMCb codec is more-or-less equivalent to iLBC in terms of PESQ-MOS score. This result is also supported by informal subjective listenings. In addition, HMCb has the advantage of having a parametric form suitable for packet loss concealment in VoIP and for small footprint TTS systems.

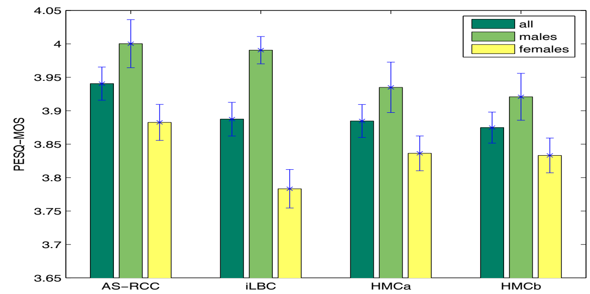

Fig. 1. PESQ-MOS evaluation (mean and 95% confidence interval) of the HMC codec, iLBC and the analysis/synthesis system.

6. CONCLUSION

A novel framework for stochastic modeling and quantization of harmonic phases was presented and evaluated in the context of sinusoidal speech coding. The statistics of phases are captured with a model suitable for variables with modulo-2π behavior. The model is used to construct a high-rate quantizer for harmonic phases. The potential of WGMM in phase modeling is not limited to quantization. Using WGMM, phase information may complement magnitude information in a number of applications like Speaker Recognition, generative models for TTS synthesis, detection of pathological speech, and others.

7. REFERENCES

[1] K. K. Paliwal and Leigh Alsteris, "Usefulness of phase spectrum in human speech perception," in *EUROSPEECH*, 2003, pp. 2117–2120.

[2] C. Gobl, *The voice source in speech communication*, Ph.D. thesis, KTH, 2003.

[3] Baris Bozkurt, *Zeros of the z-transform (ZZT) representation and group delay processing for the analysis of source and filter characteristics of speech signals*, Ph.D. thesis, 2005.

[4] Jonas Lindblom, "A sinusoidal voice over packet coder tailored for the frame-erasure channel," *IEEE Trans. Speech Audio Processing*, 2004.

[5] Thomas F.Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice Hall, 2001.

[6] X.Q Sun, B.M.G. Cheetham, and W.T.K. Wong, "Spectral envelope and phase optimization for sinusoidal speech coding," in *IEEE Workshop on Speech Coding for Telecommunications*, Annapolis, USA, 1995, pp. 75–76.

[7] Sassan Ahmadi and Andreas Spanias, "A new sinusoidal phase modelling algorithm," in *ICASSP*, Munich, Germany, 1997, vol. 3, pp. 1675–1678.

[8] Jorge S. Marques, Luis B. Almeida, and Jose M. Tribolet, "Harmonic coding at 4.8 kb/s," in *ICASSP*, 1990, vol. 1, pp. 17–20.

[9] Oded Gottesman, "Enhanced waveform interpolative coding at low bit-rate," *IEEE Transactions Speech and Audio Processing*, vol. 9, no. 8, pp. 786–798, 2001.

[10] Y. Stylianou, "On the implementation of the harmonics plus noise model for concantenative speech synthesis," in *ICASSP*, Istanbul, Turkey, 2000, vol. 2.

[11] K.V. Mardia, "Statistics of directional data," 1972.

[12] Paris Smaragdis and Petros Boufounos, "Learning source trajectories using wrapped-phase hidden markov models," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 2005.

[13] Hajo Holzmann, Axel Munk, Max Suster, and Walter Zucchini, "Hidden markov models for circular and linear-circular time series," *Environmental and Ecological Statistics (to appear)*, 2006.

[14] D. Subramaniam Anand, *Gaussian Mixture Models in Compression and Communication*, Ph.D. thesis, University of California, 2003.

[15] ITU-T, "Perceptual evaluation of speech quality assessment of narrowband telephone networks and speech codecs," February 2001 2001.

[16] Yannis Agiomyrgiannakis and Yannis Stylianou, "Combined estimation/coding of highband spectral envelopes for speech spectrum expansion," in *ICASSP*, Montreal, Canada, 2004.

[17] Yannis Stylianou, *Harmonic plus Noise models for Speech, combined with Statistical Methods for Speech and Speaker Modification*, Ph.D. thesis, 1998, Phd thesis.

[18] Allen Gersho and Robert M. Gray, *Vector Quantization and Signal Compression*, Kluwer, 1992.

[19] S. V. Andersen, W. B. Kleijn, R. Hagen, J.Linden, M. N. Murthi, and J. Skoglund, "iLBC - a linear predictive coder with robustness to packet losses," in *IEEE Workshop on Speech Coding*, Tsukuba, Ibaraki, Japan, 2002.