Chebychev Polynomials in a Speech Recognition Model

G. CARBALLO
Departamento de Personalidad, Evaluación y Tratamiento Psicológico
Facultad de Psicología, Universidad de Granada, E-18071 Granada, Spain
gloriac@platon.ugr.es

R. ÁLVAREZ-NODARSE
Departamento de Análisis Matemático, Universidad de Sevilla
Apdo 1160, E-41080 Sevilla, Spain
and
Instituto Carlos I de Física Teórica y Computacional
Universidad de Granada, E-18071 Granada, Spain
rankica.es

J. S. DEHESA
Departamento de Física Moderna
and Instituto Carlos I de Física Teórica y Computacional
Universidad de Granada, E-18071 Granada, Spain
dehesa@ugr.es

(Received and accepted June 2000)

Communicated by E. Ortiz

Abstract—Advanced speech information processing systems require further research on speaker-dependent information. Recently, a specific system of discrete orthogonal polynomials \( \phi(l), l = 1, 2, \ldots, L \) has been encountered to play a dominant role in a segmental probability model recently proposed in the speaker-dependent feature extraction from speech waves and applied to text-independent speaker verification. Here, these speech polynomials are shown to be the shifted Chebyshev polynomials on a discrete variable \( l(l-1, L) \), whose structural and spectral properties are discussed and reviewed in light of the recent discoveries in the field of discrete orthogonal polynomials.

Keywords—Discrete Chebyshev polynomials, Speech recognition, Speaker verification, Speech identification, Speech sound, Speech compression

This work has been partially supported by the European Project INTAS-93-219-ext and by the Spanish Dirección General de Enseñanza Superior (DGES) Grant PB-96-0120-C03-01.

This work has been partially supported by the European Project INTAS-93-219-ext, the Junta de Andalucía Grant P00M-207 and by the Spanish Dirección General de Enseñanza Superior (DGES) Grants PB-96-0120-C01-01 (R.A.N.) and PB-96-1205 (J.S.D.)

This paper was done during the stay of the second author in the Universidad de Granada.
1. INTRODUCTION

The classical orthogonal polynomials have been shown to play a relevant role in speech science, particularly, in research on extraction of speaker-dependent features from speech waves. This is the case of the Legendre polynomials which have been used for speech recognition, speech enhancement, speaker adaptation, (see, e.g., [1-4]). The discrete polynomials were first used in speech recognition performance by means of the orthogonal-polynomial-compression technique [5,6], which allows us to independently compress different aspects of the speech spectrum. Indeed, each polynomial corresponds to a different feature of the short-term speech spectrum, for example, the polynomials of first and second degrees correspond to the average slope and quadratic curvature of the spectrum.

Recently, the discrete polynomials have been used in the research of speaker recognition to face both speaker verification and speaker identification problems. Indeed, there has been proposed [7] a segmental probabilistic model which is based on an orthogonal polynomial representation of speech signals. Contrary to the conventional frame-based probabilistic models, the Liu-Wang model concatenates several consecutive frames with similar characteristics into an acoustic segment and represents it by an orthogonal polynomial function. Thus, the speech signal is composed of \( L \) successive \( N \)-dimensional feature vectors, that is, a set of \( N \) trajectories of length \( L \). Each trajectory is represented by a polynomial function. Also, Liu and Wang propose an iterative, self-consistent procedure that performs recognition and segmentation processes for estimating the segment-based speaker model. Moreover, they illustrate its validity not only in the text-dependent speaker recognition, where the speaker is required to issue a predetermined utterance, but also in the text-independent recognition methods which do not rely on a specific text being spoken. These methods, whose aim is to verify the identity of a claimed speaker, have a training phase and a verification phase. For a given speech signal of a specific speaker, the number of segments, the length of each segment, and the appropriate segmental probabilistic model are determined in the training phase. The performance of the model depends on the mixture number (i.e., number of acoustic segments used for modeling the speaker’s voice characteristics) and the degrees of the orthogonal polynomials. This degree affects the accuracy of the model and controls the type of the basic segment for the model and its characteristics. In fact, the degree of the orthogonal polynomial used in the model determines the smallest length of the partitioned segment of a given speech signal. Moreover, the degree and the algebraic properties of the discrete orthogonal polynomials used are crucial for the efficiency (computation time and memory storage) and accuracy of the model. Furthermore, the discrete orthogonal polynomials that have been naturally encountered in some speaker recognition methods (to be called henceforth speech polynomials) correspond to a particular class of the so-called Hahn polynomials [8-11], denoted by \( h_{n}^{\alpha,\beta}(x,N) \). This class is composed by the classical Chebyshev polynomials of a discrete variable \( t_{n}(x,N) = h_{n}^{0,0}(x,N) \), which were introduced by the Russian mathematician Chebyshev in the past century (see [12]). Until now, however, the study of the algebraic and spectral properties of the Chebyshev polynomials is a very interesting mathematical topic which receives much attention in the modern theory of special functions [11,13-16].

The purpose of this paper is to identify the speech polynomials as shifted Chebyshev polynomials. In doing so, we observe that some mathematical tools used in some speaker recognition methods could be considerably reduced (see, e.g., the recurrence relation used in the Liu-Wang paper [7, Appendix A] for the orthogonal polynomials), that can imply a big reduction and simplification in the algorithms inherent to these methods.

The structure of the paper is as follows. In Section 2, some definitions and statements of the general theory of orthogonal polynomials are given. Then, in Section 3 the speech polynomials are identified as shifted Chebyshev polynomials. Finally, in Section 4 the spectrum of zeros of the speech polynomials as a whole is studied by the explicit determination of the moments-around-the-origin of the distribution of zeros of the speech polynomials.
2. BASIC BACKGROUND ON ORTHOGONAL POLYNOMIALS

In this section, we will describe some well-known facts from the general theory of orthogonal polynomials and, specifically, from the Chebyshev polynomials of a discrete variable which will help us in the next section to identify the orthogonal polynomials found in the Lie-Wang model [7].

Let \( \mu(x) \) be a nonconstant and nondecreasing function in \([a, b]\) (if any of \(a, b\) are ±∞, we require that \(\mu(±\infty)\) should be finite). Let us define the scalar product of two real functions \( f \) and \( g \) by the Stieltjes-Lebesgue integral

\[
(f, g) = \int_a^b f(x)g(x) \, d\mu(x),
\]

where we suppose that \( f, g \) belong to \( L_2(\mu) \), i.e., \( \int_a^b f^2(x) \, d\mu(x) < +\infty \). A particular case of special interest corresponds to the ones when \( \mu \) is a step function with jumps at a finite number of points \( \{x_i\}_{i=1}^N, x_i \in [a, b], i = 1, 2, \ldots, N \). In this case, \( (2.1) \) becomes

\[
(f, g) = \sum_{i=1}^N f(x_i)g(x_i)\rho(x_i), \quad \rho(x) > 0, \quad \forall x \in [a, b],
\]

and \( \rho \) is said to be a discrete weight function. Given a sequence of linearly independent functions in \( L_2(\mu) \), it is always possible to obtain an orthogonal sequence. In fact, if we denote by \( \Delta_n \) the following determinant \( \Delta_n = ||\mu_{j+k}||_{n,j=0}^n \), where \( \mu_k = \int_a^b x^k \, d\mu(x) \), \( k = 0, 1, 2, \ldots \), are the moments associated to \( \mu \), then the Gram-Schmidt orthogonalization process leads us to a set of orthogonal polynomials. In the case when \( \mu \) has a finite number \( N \) of point of increase (like in the case of a discrete weight function mentioned above), \( n \) is necessarily finite \( n \leq N \). Furthermore, the following theorem holds (see, e.g., [8, 11]):

**Theorem 1** Given a distribution function \( \mu \) with moments \( \mu_k \), \( k = 0, 1, 2, \ldots \), there exists a uniquely determined up to a constant multiplicative factor sequence of orthogonal polynomials \( \{p_n\} \), each of which have degree exactly equal \( n \), providing that \( \Delta_n > 0 \) for all \( n \geq 0 \). Moreover, if \( \{p_n\}_{n=0}^\infty \) is a monic orthogonal polynomial sequence with respect to a weight function \( \rho(x) \), then the polynomials \( p_n \) satisfy a three-term recurrence relation of the form

\[
p_n(x) = (x - c_n)p_{n-1}(x) - \lambda_np_{n-2}(x), \quad p_{-1}(x) = 0, \quad p_0(x) = 1, \quad n \geq 1,
\]

where \( \{c_n\}_{n=0}^\infty \) and \( \{\lambda_n\}_{n=0}^\infty \) are given by \( c_n = (xp_{n-1}, p_{n-1})/(p_{n-1}, p_{n-1}), \) \( n \geq 1 \) and \( \lambda_n = (xp_{n-1}, p_{n-2})/(p_{n-2}, p_{n-2}), \) \( n \geq 2 \), respectively.

In this paper, we will deal with the classical discrete Chebyshev monic polynomials \( t_n(x, N) \) (a subclass of the Hahn polynomials \( h_n^\alpha,\beta(x, N) \), [11] that with \( \alpha = \beta = 0 \)). These polynomials \( t_n(x, N) \) are polynomials that satisfy an orthogonality relation of the form

\[
\sum_{x=0}^{N-1} t_n(x, N)t_m(x, N) = \delta_{nm}\frac{n!2(N+n)!}{(2n+1)(N-n-1)!(n+1)^2},
\]

where \( \delta_{nm} \) is the Kronecker symbol (\( \delta_{nm} = 1 \) if \( n = m \) and 0, elsewhere) and \( (a)_n = a(a + 1) \quad (a + n - 1) \) denotes the Pochhammer symbol. That is, they are orthogonal with respect to a distribution function \( \mu \), which is a step function with \( N \) jumps at the points \( x = 0, 1, \ldots, N - 1 \), but in this case, as we already pointed out (see, e.g., [8, p 24]), they form a finite family of orthogonal polynomials. Furthermore, they satisfy a three-term recurrence relation (2.3) with coefficients

\[
c_n = \frac{N - 1}{2}, \quad \lambda_n = \frac{(n - 1)^2 [N^2 - (n - 1)^2]}{4[4(n - 1)^2 - 1]}.
\]
3. IDENTIFICATION OF THE SPEECH POLYNOMIALS

In the Lee-Wang model [7, Section 3.1], a sequence of orthogonal polynomials \( \phi_n^L(l) \) is introduced to regenerate a time sequence of \( L \)-length feature vectors. This family of polynomials satisfies an orthogonality relation [7, Appendix A, equation (48)]

\[
\sum_{l=1}^{L} \phi_n^L(l) \phi_k^L(l) = 0, \quad n \neq k, \quad n, k = 0, 1, 2, \ldots, R
\]  

(3.1)

Obviously, the above orthogonality relation corresponds to the discrete scalar product (2.2). Then, Theorem 1 states that the polynomials \( \phi_n^L(l) \) are uniquely determined up to a constant factor. Moreover, since the distribution function \( \mu \) is a step function with \( L \) jumps at points \( x = 1, 2, \ldots, L \), the family \( \phi_n^L(l) \) is a finite family, i.e., in (3.1), \( R \leq L - 1 \). If we now compare the orthogonality relation (3.1) with (2.4), one can easily arrive at the conclusion that the speech polynomials \( \phi_n^L(l) \) are proportional to the Chebyshev polynomials \( \phi_n(x, L) \). In fact, making the change of variable \( x \rightarrow l - 1 \) in (2.4) \( \phi_n(x, L) = \sum_{k=0}^{L-1} \frac{\Gamma(L-k+k+1)}{\Gamma(L-n-l+k+1)\Gamma(L-n-l-k)}(l-1)^{(L-k)}(l-k) \), we arrive at (3.1). Moreover, the square norm of the \( \phi_n^L(l) \), denoted \( \Phi_n^L \), has the explicit form \( \sum_{l=1}^{L} \phi_n^L(l) \phi_k^L(l) = (n^2(L+n))/((2n+1)(L-n-1)(n+1)^2) \).

For simplicity, let us consider the monic polynomials, i.e., the polynomials \( \phi_n^L(l) = l^n \). With this normalization, and using the three-term recurrence relation for the Chebyshev polynomials (2.5), we obtain that the speech polynomials \( \phi_n^L(l) \) satisfy a three-term recurrence relation of the form

\[
\phi_n^L(l) = \left( l - \frac{L+1}{2} \right) \phi_{n-1}^L(l) - \frac{(n-1)^2[L^2-(n-1)^2]}{4(n-1)^2 - 1} \phi_{n-2}^L(l),
\]

which is nothing more than relation (50) in the Liu-Wang model [7]. Most important is to remark that the \( \alpha \) and \( \beta \) values of this model [7, equations (51) and (52)] reduce to \( \alpha = 0, \beta = -(n-1)^2[L^2-(n-1)^2]/(4(n-1)^2 - 1) \). Obviously, from the properties of the Chebyshev polynomials [8,11,13], one can obtain many properties for the \( \phi_n^L(l) \). In fact, we have following

1. Second-order difference equation

\[
(l-1)(L-l+1)\Delta \nabla \phi_n^L(l) + (L+1-2l)\Delta \phi_n^L(l) + n(n+1)\phi_n^L(l) = 0,
\]

\[
\nabla f(x) = f(x+1) - f(x) \quad \text{and} \quad \Delta f(x) = f(x) - f(x-1)
\]

2. Explicit formula

\[
\phi_n^L(l) = \frac{(-1)^n}{(n+1)n} \sum_{k=0}^{n} (-1)^k \frac{n!}{k!(n-k)!} \Gamma(L-l+k+1)\Gamma(n+l-k)\Gamma(L-n-l+k+1)\Gamma(l-k),
\]

\[
\phi_n^L(1) = \frac{n!(L-1)!}{(n+1)(L-n-1)!}, \quad \phi_n^L(L) = \frac{n!(L-1)!}{(n+1)n(L-n-1)!}
\]

3. Symmetry property \( \phi_n^L(L-l+1) = (-1)^n \phi_n^L(l) \)

4. Dette inequality \( |\phi_n^L(l)| \leq (L-n)_n/(n+1)_n \)

4. SPECTRAL MOMENTS OF THE SPEECH POLYNOMIALS

Some important spectral characteristics of the speech polynomials are the moments of their zeros, which are defined by \( \mu_0 = 1, \mu_m^{(n)} = 1/n \sum_{k=1}^{n} x_{k,n}^m, m = 1, 2, \ldots, L-1, n \leq L-1 \), where \( x_{k,n}, k = 1, 2, \ldots, n \) denotes the zeros of the polynomial \( \phi_n^L \). To obtain these quantities,
we can use the method given in [16] Then the first few spectral moments of $\phi_k^L$'s have the simpler expressions

$$
\mu_1^{(n)} = \frac{L + 1}{2}, \quad \mu_2^{(n)} = \frac{(1 + 3L (1 + L)) + (2 + 3L)^2 n + 2n^2 - n^3}{24n - 12},
$$
$$
\mu_3^{(n)} = \frac{(1 + L)( -2L - 4L^2 + 4Ln + 5L^2 n + 2n^2 - n^3)}{16n - 8}
$$

These quantities give different dispersion measures of the distribution of zeros of the speech polynomials The centroid of the distribution is $\mu_1^{(n)}$, and the variance $\sigma^2$ is equal to

$$
\sigma^2 = \mu_2^{(n)} - \left( \mu_1^{(n)} \right)^2 = \frac{(n - 1) (3L^2 + n - n^2 - 1)}{24 n - 12}
$$

Also, it turns out that the skewness $\gamma_1$ vanishes and the excess or kurtosis is positive The former is a straightforward consequence of the symmetric nature of the distribution, while the latter indicates that the distribution of the zeros of the speech polynomials are sharper around the centroid than a Gaussian distribution of the same variance

**REFERENCES**