Accepted Manuscript

A Subspace Based Progressive Coding Method for Speech Compression

Serkan Keser , Ömer Nezih Gerek , Erol Seke ,
Mehmet Bilginer Gülmezoğlu

PII: S0167-6393(17)30055-9
DOI: 10.1016/j.specom.2017.09.002
Reference: SPECOM 2490

To appear in: *Speech Communication*

Received date: 15 February 2017
Revised date: 13 August 2017
Accepted date: 4 September 2017

Please cite this article as: Serkan Keser , Ömer Nezih Gerek , Erol Seke , Mehmet Bilginer Gülmezoğlu , A Subspace Based Progressive Coding Method for Speech Compression, *Speech Communication* (2017), doi: 10.1016/j.specom.2017.09.002

A Subspace Based Progressive Coding Method for Speech Compression

Serkan Keser, Ömer Nezih Gerek, Erol Seke, Mehmet Bilginer Gülmezoğlu

The corresponding author: Serkan Keser
Department of Electronic and Automation, Ahi Evran University, Kırşehir 40300, Turkey
Email: skeser@ahievran.edu.tr

Co-authors : (Professor) Ömer Nezih Gerek
Department of Electrical & Electronics Engineering, Anadolu University, Eskişehir 26555, Turkey
Email: ongerek@anadolu.edu.tr

Co-authors : (Assistant Professor) Erol Seke
Department of Electrical & Electronics Engineering, Eskişehir Osmangazi University, Eskişehir 26480, Turkey
Email: eseke@ogu.edu.tr

Co-author: (Professor) Mehmet Bilginer Gülmezoğlu
Department of Electrical & Electronics Engineering, Eskişehir Osmangazi University, Eskişehir 26480, Turkey
Email bgulmez@ogu.edu.tr

***Abstract*-- In this study, two novel methods, which are based on Karhunen Loeve Transform (KLT) and Independent Component Analysis (ICA), are proposed for coding of speech signals. Instead of immediately dealing with eigenvalue magnitudes, the KLT- and ICA-based methods use eigenvectors of covariance matrices (or independent components for ICA) by geometrically grouping these vectors into fewer numbers of vectors. In this way, a data representation compaction is achieved. Further compression is achieved through discarding autocovariance eigenvectors corresponding to the small eigenvalues and applying vector quantization on the remaining eigenvectors. Additionally, this study proposes an iterative error refinement process, which uses the rest of the available bandwidth in order to transmit an efficient representation of the description error for better SNR. The overall process constitutes a new approach to efficient speech coding, with ICA being used in subspace speech coding for the first time. Constant bit rate (CBR) and variable bit rate (VBR) coding algorithms are employed with the proposed methods. TIMIT speech database is used in the experimental studies. Speech signals are synthesized at 2.4 kbps, 8 kbps, 12.2 kbps, 16 kbps, 16.4kbps and 19.85 kbps rates by using various frame lengths. The qualities of synthesized speech signals are compared to those of available speech codecs, i.e., LPC (2.4 kbps), G.728 (LD-CELP, 16 kbps), G.729A (CS-CELP, 8 kbps), EVS (16.4 kbps), AMR-NB (12.2 kbps) and AMR-WB (19.85 kbps).**

***Keywords* — Independent Component Analysis (ICA), Karhunen Loeve Transform (KLT), Speech codecs, Subspace methods.**

1. Introduction

The goal of speech coding is to represent digital speech waveform with as few bits as possible while maintaining the intelligibility and quality that is required for the particular application (Gibson, 2005). In addition, most applications of speech coding require low coding delays, which is an undesirable property since long coding delays interfere with speech interaction (Chen et al., 1992). Major speech coders can be classified into two categories as waveform and parametric coders. The former includes speech coders such as PCM and ADPCM, and latter class (also known as vocoders) includes very low bit-rate synthetic speech coders (Kondoz, 2007). LPC-based coder is a parametric coder which is mostly used in audio signal processing and speech processing to represent the spectral envelope of speech waveform in a compressed form. This coder uses the information of a finite extent linear predictive model (Deng and O'Shaughnessy, 2003). Linear prediction based speech coding techniques (CELP, MELP, VSELP etc.) have been widely researched in the literature (Vasuki and Vanathi, 2006; Supplee et al., 1997; Gerson and Jasiuk, 1990; Chen et al., 1992). These types of speech coders are capable of synthesizing good quality speech at a reasonably low bit rate. An LPC variant, CELP, has evolved to become the dominant paradigm for real time speech compression (Devalapalli et al., 2003), which is capable of achieving high quality speech coding at rates from 16 kbps to 32 kbps. A further variant, namely the low delay-CELP (LD-CELP) algorithm was adopted by the International Telephone and Telegraph Consultative Committee (CCITT) for speech coding at 16 kbps with toll quality and became a standard as G.728 (Chen et al., 1992). Similarly, G.729A (CS-ACELP Annex A) is a high quality low bandwidth codec at 8 kbit/s with low complexity. ITU-T (International Telecommunication Union) has standardized G.729 as the standard speech coding algorithm for VoIP, DSVD (Digital Simultaneous Voice over Data) and multimedia applications (Rashed et al., 2013). The mixed excitation linear prediction (MELP) coder was chosen by the Digital Voice Processing Consortium to replace the existing 2400 bps Federal Standard FS-1015 (LPC-10). The MELP coder is based on the traditional LPC model, with additional features to improve its performance (Supplee et al., 1997). The vector sum excited linear prediction (VSELP) speech coder utilizes a codebook with a structure that allows for a very efficient search procedure (Gerson and Jasiuk, 1990). The coder uses two VSELP excitation codebooks, a gain quantizer which is robust to channel errors, and a novel adaptive pre/postfilter arrangement.

Being the fundamental application medium of speech coders, GSM networks started with Full Rate (FR) speech codec and evolved into Enhanced Full Rate (EFR). The Adaptive Multi-Rate (AMR) codec was added to 3GPP (The 3rd Generation Partnership Project) Release 98 for GSM to enable codec rate adaptation to radio conditions (Holma and