IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 5, May 2016

Speaker Recognition Using G.723 Coder

Mrs. Minaj S. Shikalgar[1], Mr. N.B.Sambre[2]

Department of Electronics & Telecommunication KIT'S College of Engineering, Kolhapur, India[1,2]

Abstract: This paper presents the speaker recognition system using G.723 coder. Speaker recognition is the computing task of validating a user's claimed identity using characteristics extracted from their voices. In this paper, G.723.1 coder is used to recognize the speaker. G.723.1 coder is a CELP based ITU-T floating point codec at 5.3 Kb/s.

Index Terms: G.723.1 coder, CELP.

I. INTRODUCTION

The task of speaker identification is to determine the identity of a speaker by machine. To recognize voice, the voices must be familiar in case of human beings as well as machines. The process of "getting to know" speaker is referred to as training and consists of collecting data from utterances of people to be identified. The second component of speaker identification is testing; namely the task of comparing an unidentified utterance to the training data and making the identification. The speaker of a test utterance is referred to as the target speaker. The problem of speaker identification can be solved by using G.723.1 at 5.3 kb/s.

I. G.723.1 CODER

Input: Input to this coder is a speech file in wave format and sampled at 8Khz sampling rate.

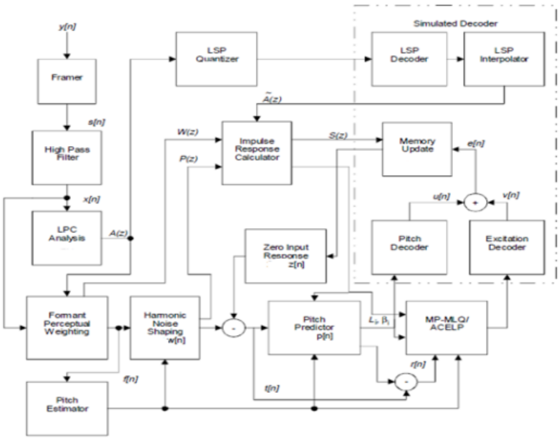
A. Review Stage

This coder operate with a digital signal obtained by first performing filtering of the analogue input, then sampling at 8000 Hz and then convert to 16-bit linear PCM for the input to the encoder. The output of the decoder is converted back to analogue. The coder is based on the principles of linear prediction analysis-by-synthesis coding and attempts to minimize a perceptually weighted error signal. The encoder operates on blocks (frames) of 240 samples each. That is equal to 30 msec at an 8 kHz sampling rate. First each block is high pass filtered to remove the DC component and then divided into four sub frames of 60 samples each.

For every two sub frames (120 samples), the open loop pitch period, is computed using the weighted speech signal. This pitch estimation is performed on blocks of 120 samples. The pitch period is searched in the range from 18 to 142 samples. From this point the speech is processed on a 60 samples per subframe basis.

Using the estimated pitch period computed previously, a harmonic noise shaping filter is constructed. The combination of the LPC synthesis filter, the formant perceptual weighting filter, and the harmonic noise shaping filter is used to create an impulse response. The impulse response is then used for further computations.

Using the pitch period estimation, LOL, and the impulse response, a closed loop pitch predictor is computed. A fifth order pitch predictor is used. The pitch period is computed as a small differential value around the open loop pitch estimate. The contribution of the pitch predictor is then subtracted from the initial target vector. Both the pitch period and the differential value are transmitted to the decoder. Finally the non-periodic component of the excitation is approximated. For the high bit rate, Multi-pulse Maximum Likelihood Quantization (MP-MLQ) excitation is used, and for the low bit rate, an algebraic-code-excitation (ACELP) is used. The G.723.1 codec is the floating point1 CELP-based ITU-T multi-media standard codec at 5.3 kb/s The coder diagram is shown in following figure.



Block diagram of the speech coder – For each block the corresponding reference number is indicated

Figure 1: Coder Diagram

Under three different conditions the coder is tested against a baseline condition in which no cosing is performed for training and testing.

- Condition A: Fully matched
- Condition B : Partially mismatched
- Condition C : Fully mismatched

II. SPEAKER RECOGNITION

In speaker recognition the experimentation is done with 10 different subjects. Among which 8 are adults. and 2 are kids. Again in that 8 adults subjects 4 are male and 4 are female. Similarly, in 4 kids 2 are boys and 2 are girls. The individual recording was recorded.

DOI 10.17148/IJIREEICE.2016.4522

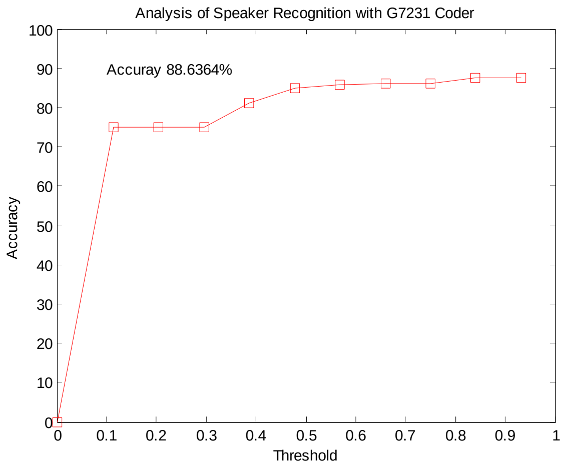


Figure 2: Analysis of G.723.1

III. CONCLUSION

The performance of the system totally depends on the bit rate of that coder. For G.723.1 coder the accuracy which is obtained is about 88.63% while sensitivity and specificity is 0.8611 and 1 respectively.

REFERENCES

[1]. J.M. Huerta and R.M. Stern, \Speech recognition from GSM coder parameters," Proc. 5th Int. Conf. on Spoken Language Processing, Vol 4, pp 1463-1466, 1998.

[2]. ITU-T Recommendation G.729, \Coding of speech at 8 kb/s using conjugate-structure algebraic-code-excited linear pre- diction," June 1995.

[3]. M.A. Zissman, \Predicting, Diagnosing, and Improving Automatic Language Identification Performance," Proc. Eurospeech97, Vol 1, pp 51-54, 1997.

[4]. ITU-T Recommendation G.723.1, \Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3kb/s," March 1996.

[5]. M.A. Zissman, "Comparison of four approaches to automatic language identification of telephone speech." IEEE Trans. Speech and Audio Proc., SAP-4(1), Jan. 1996, pp. 31-44

[6]. Y. Yan and E. Bernard, "An approach to automatic language identification based on language-dependent phone recognition." Proc. ICASSP '95, vol. 5, May 1995, pp. 3511-3514. .

[7]. Gaussian Mixture Models, Douglas Reynolds, MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA

[8]. D.A. Reynolds, \Comparison of Background Normalization Methods for Text-Independent Speaker Verification," Proc. Eurospeech97, Vol 1, pp 963-967, 1997.

[9]. Universal Background Models, Douglas Reynolds, MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA.

[10]. Alejandro Abejón González **"**Phonotactic speaker and language recognition".

DOI 10.17148/IJIREEICE.2016.4522