



# Audio Engineering Society Convention Paper

Presented at the 120th Convention  
2006 May 20–23 Paris, France

*This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42<sup>nd</sup> Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

## Reduced Bit Rate Ultra Low Delay Audio Coding

Stefan Wabnik, Gerald Schuller, Jens Hirschfeld, Ulrich Krämer

*Fraunhofer Institute for Digital Media Technology, Ehrenbergstrasse 29, 98693 Ilmenau, Germany*

Correspondence should be addressed to Stefan Wabnik ([wbk@idmt.fraunhofer.de](mailto:wbk@idmt.fraunhofer.de))

### ABSTRACT

An audio coder with a very low delay (6-8 ms) for reduced bit rates is presented. Previous coder versions were based on backward adaptive coding, which has suboptimal noise shaping capabilities for reduced bit rate coding. We propose to use a different noise shaping method instead, resulting in an approach which uses forward adaptive predictive coding. We will show that, in comparison, the forward adaptive method has the following advantages: it is more robust against high quantization errors, has additional noise shaping capabilities, has a better ability to obtain a constant bit rate, and shows improved error resilience.

### 1. INTRODUCTION

The use of digital audio coding in new communication networks as well as in professional audio productions for real-time, bi-directional communication requires very low algorithmic encoding and decoding delay. A typical scenario in which the application of digital audio coding becomes delay-critical is when direct and transmitted (encoded and decoded) signals are used simultaneously. Examples are live productions with wireless microphones and simultaneous (in-ear) monitoring, or distributed productions where artists perform simultaneously in different studios. The tolerable total delay time in these applications is less than ten milliseconds. If, for instance, asymmetric subscriber lines are used for

communication, the bit rate becomes a limiting factor, too!

The algorithmic delay of standard audio coders like MPEG-1 layer 3 (MP3), MPEG-2 AAC, and MPEG-2/4 Low Delay, ranges from 20 milliseconds to several hundred milliseconds [1]. Speech coders, though operating at lower bit rates and with less algorithmic delay, provide only limited audio quality.

Our Ultra Low Delay Coder (ULD) has an encoding/decoding delay of only 5.5 to 8 milliseconds at sampling frequencies from 32 kHz to 48 kHz.

#### 1.1. Goal

Our ULD coder obtains a perceptual quality comparable to standard audio coders, like MP3, for bit

rates of about 80 kbit/s per channel and higher. Our goal is to obtain a version for lower bit rates, for instance 64 kbit/s, with still competitive perceptual quality. Furthermore, a simple scheme to obtain a constant bit rate is desirable, especially for the targeted lower bit rates. Additionally, we also would like to minimize recovery time after a transmission error.

### 1.2. Problem

For redundancy reduction of the psycho-acoustically pre-processed input signal, the ULD coder so far uses closed-loop, sample-wise backward adaptive prediction. This means that the calculation of the prediction coefficients in encoder and decoder relies only on quantized and reconstructed signal samples of the past. To adapt to the signal, a new set of predictor coefficients is calculated for every sample. As an advantage, long predictors can be used since there is no need to transmit the prediction coefficients. However, it also means that the quantized prediction error has to be transmitted to the decoder without loss of accuracy in order to derive prediction coefficients identical to that of the encoder. Otherwise, the predicted values in the encoder and decoder would not be identical which would result in an unstable decoder. To enable random access of the bit stream, as well as to stop propagation of transmission errors, a periodic reset of the predictor in both encoder and decoder is necessary. This, however, leads to bit rate peaks. For a variable bit rate channel this is not a problem. For fixed bit rate channels, this limits the lower bound of a constant bit rate setting.

### 1.3. Approach

The approach we propose in this paper is to use block-wise forward adaptive prediction with a backward adaptive quantization step size, instead of sample-wise backward adaptive prediction. On one hand this has the disadvantage that the predictors have to be shorter in order to limit the necessary side-information for transmitting its coefficients to the decoder. This potentially results in reduced coding efficiency. On the other hand it has the advantage that it still works effectively for higher quantization errors (a result of reduced bit-rates), hence the decoder predictor can be used for quantization noise shaping.

To limit the bit rate we restrict the range of values of the prediction residual before transmission.

This results in a modified noise shaping method and leads to differently sounding artifacts. It also produces a constant bit rate without the use of iteration loops. At the same time, a reset is included automatically for each block of samples as a result of the block-wise forward adaptation. Additionally, we use a coding scheme for both pre-filter coefficients as well as forward prediction coefficients, which applies differential coding with backward adaptive quantization step size control on a line spectral frequency representation of the coefficients. The scheme provides block-wise access of the coefficients, produces a constant side info bit rate, and is robust against transmission errors.

## 2. CODER DESCRIPTION

In what follows, the ULD encoder and decoder structure used for higher constant bit rates as well as the proposed modifications for lower bit rates are described in greater detail.

### 2.1. ULD Coder Structure

The input signal of the encoder is analyzed by a perceptual model in order to obtain information about the perceptually irrelevant parts of the signal. The information is used to control a pre-filter via time-varying filter coefficients. By this, the pre-filter normalizes the input signal with respect to its masking threshold [2]. The filter coefficients are calculated once every block of 128 samples, quantized and transmitted as side information to the decoder.

After the pre-filtered signal is multiplied with a gain factor and the predicted signal is subtracted, the prediction residual is quantized by a uniform quantizer. The predicted signal is derived via closed loop, sample-wise backward adaptive prediction. Therefore, no prediction coefficients need to be transmitted to the decoder. After that, the quantized prediction residual is entropy coded. For constant bit rate coding, a distortion loop is used which repeats the steps of multiplication, prediction, quantization and entropy coding several times for each block of pre-filtered samples. After the iteration, the highest gain factor of a set of pre-defined gain values which still fulfills the constant bit rate constraint has been found [3]. This gain has to be transmitted to the decoder. However, if a gain value smaller than unity is found, the quantization noise is perceivable after decoding, i.e. its spectrum is shaped similar to the

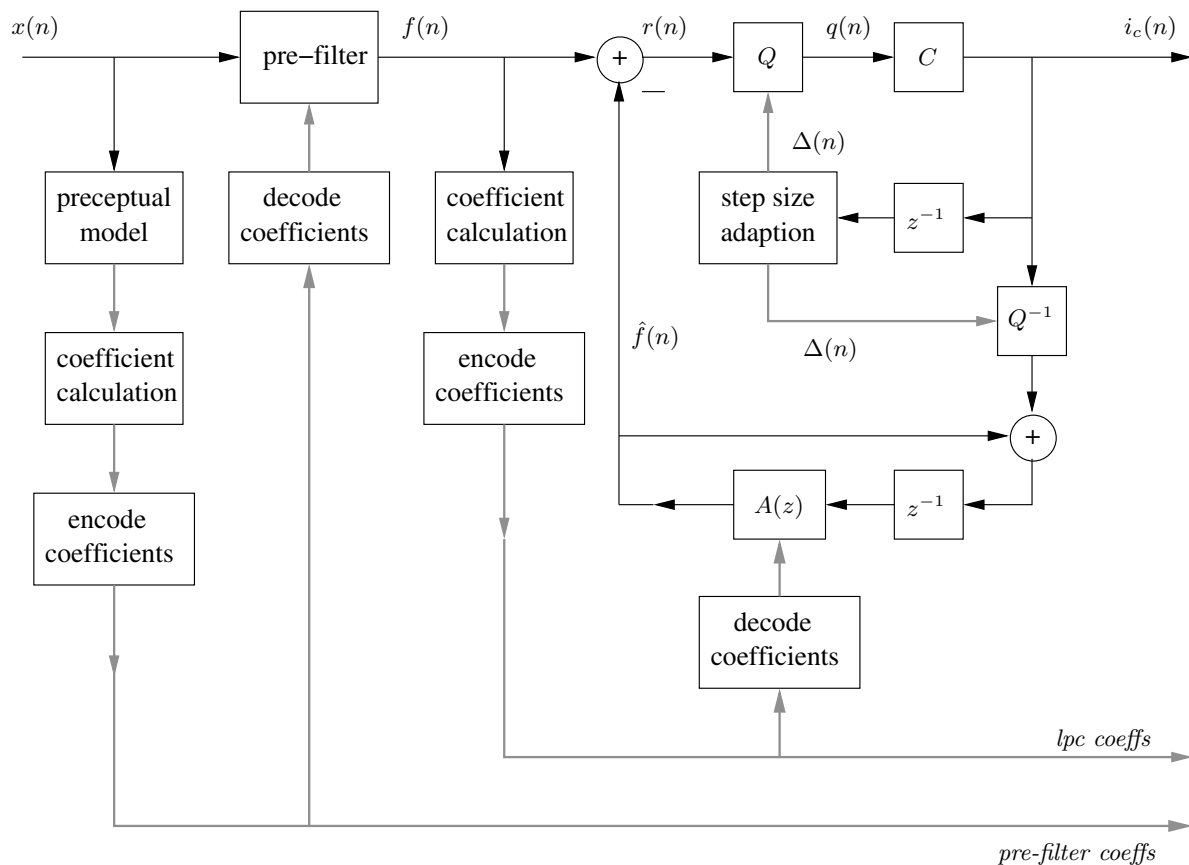


Fig. 1: Proposed structure for the ULD encoder with forward-adaptive prediction and clipping.

masking threshold, but its overall power is higher than suggested by the perceptual model. This can be described as an elevation of the masking threshold. For parts of the input spectrum, the elevated threshold could become even higher than the input signal spectrum itself, producing audible artifacts in parts of the spectrum where otherwise would be no audible signal, due to the use of a predictive coder. The effects an elevated masking threshold causes are the limiting factor when aiming for lower constant bit rates.

The pre-filter coefficients are only transmitted as

first order, intra-frame line spectral frequency (lsf) differences exceeding a certain limit. To prevent transmission error propagation, the system is reset from time to time. Additionally, concealment techniques are applied to minimize perceptual degradation of the decoded signal in case of transmission errors [4]. The transmission scheme produces a variable side information bit rate, which is accounted for in the distortion loop.

Entropy coding of the quantized prediction residual can be done using different methods, e.g. golomb, huffman or arithmetic coding. The entropy coding

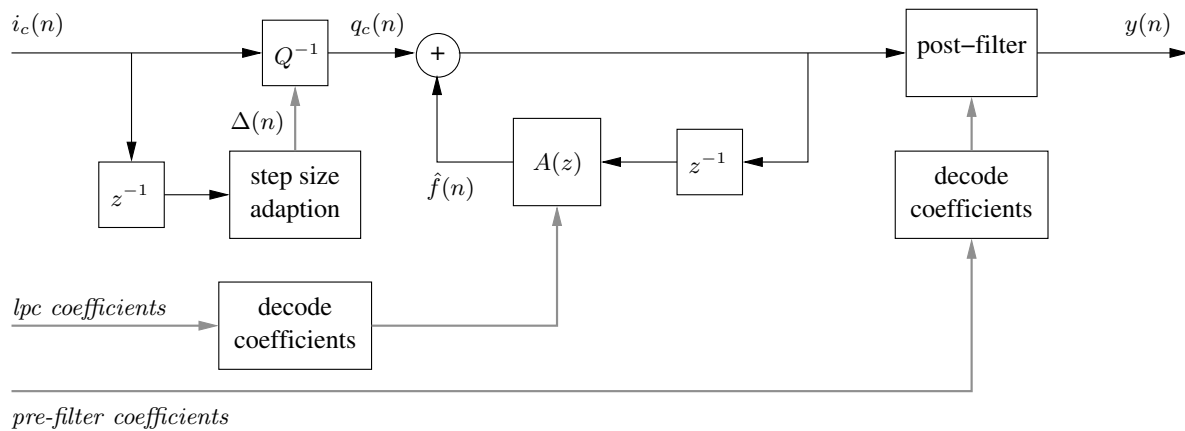


Fig. 2: Proposed structure for the ULD decoder.

has to be reset and inherently produces a variable bit rate.

In the decoder, the quantized prediction residual is retrieved after entropy decoding. The prediction residual and the predicted signal are added, the sum is multiplied with the inverse of the transmitted gain factor. From that sum, the post-filter with a frequency response inverse to that of the pre-filter generates the output signal using the transmitted pre-filter coefficients.

## 2.2. Proposed ULD Coder Structure for lower bit rates

The proposed coder structure (see Fig.1) applies the same pre-/post-filter combination to the input signal  $x(n)$  as described in 2.1. For each block of pre-filtered samples, a linear predictive coding (lpc) analysis is done to obtain the prediction coefficients, instead of using sample-wise backward adaptive prediction. The coefficients found in the lpc analysis are used with a closed-loop predictor to generate a predicted signal  $\hat{f}(n)$  which is subtracted from the pre-filtered signal  $f(n)$ . The prediction residual  $r(n)$  is quantized in  $Q$  with time-varying step size  $\Delta(n)$ , which is calculated backward adaptively [5]. The quantized residual  $q(n)$  can be expressed as  $q(n) = i(n) \cdot \Delta(n)$  with  $i(n)$  as the quantization step index. The quantization index  $i(n)$  is clipped in  $C$ , i.e. for a constant  $c \in \{1, 2, \dots\}$ .

All index values with  $|i(n)| > c$  are set to either  $-c$  or  $c$ . Only the clipped index sequence  $i_c(n)$  is transmitted. The backward adaptive step size control uses past index sequence values to calculate  $\Delta(n) = \beta \Delta(n-1) + \delta(n)$ ,  $\beta \in [0.0; 1.0[$  with  $\delta(n) = \delta_0$  for  $|i_c(n-1) + i_c(n-2)| \leq I$  and  $\delta(n) = \delta_1$  for  $|i_c(n-1) + i_c(n-2)| > I$  with constant  $\delta_0, \delta_1, I$ .

The proposed ULD decoder is shown in Fig. 2. With the transmitted index sequence  $i_c(n)$  and the calculated sequence  $\Delta(n)$ ,  $q_c(n)$  is reconstructed and added to  $\hat{f}(n)$ . From this sum, the decoded sequence  $y(n)$  is generated by the post-filter.

The quantization noise introduced in  $q_c(n)$  through clipping is no longer white. Its spectral shape resembles that of the pre-filtered signal [6], hence after post-filtering the quantization noise spectrum more closely resembles that of the input signal decoded signal. This means that the quantization noise after decoding stays below the signal spectrum. The effect is illustrated in Fig. 3. For backward (a) adaptive prediction and forward (b) adaptive prediction with applied clipping, three curves are shown in the normalized frequency domain: the signal, the quantization noise after decoding and the masking threshold (from top to bottom). For the unmodified ULD coder, as shown in sub-plot a), the quantization noise is shaped like the masking threshold and lies above the signal spectrum for parts of the signal.

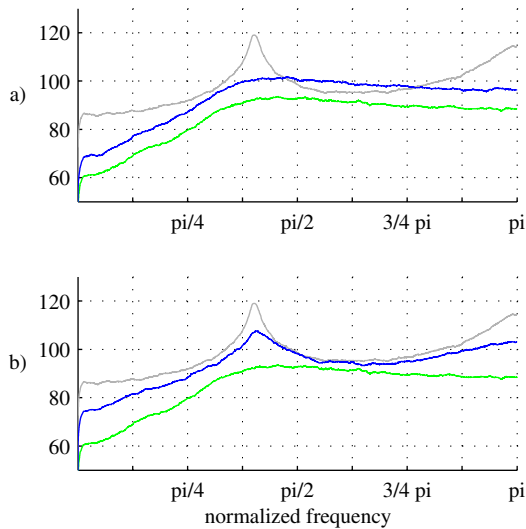


Fig. 3: Signal psd, quantization error psd and masking threshold (top to bottom curves) after decoding for a) backward adaptive prediction, b) forward adaptive prediction and clipping.

The effect of the proposed modifications is shown in sub-plot b): the quantization noise is always smaller than the signal spectrum and its shape is a mixture of signal spectrum and masking threshold. In a listening test, we found that the coding artifacts introduced by this method are less annoying.

For the proposed ULD coder structure, the transmission of both pre-filter and predictor coefficients is done using a constant bit rate coding scheme. The filter coefficients are converted into the line spectral frequency domain. Each line spectral frequency  $l(n)$  is processed as shown in Fig. 4. From the calculated  $l(n)$ , a constant  $l_c$  is subtracted. From the difference  $l_d(n)$  a predicted  $\hat{l}_d(n)$  is subtracted, which is calculated using a closed-loop predictor with fixed coefficients  $A(z)$ . What remains is quantized with an adaptive step size quantizer. The values of the resulting index sequence are clipped such that  $\forall n : l_e(n) \in \{-1, 0, 1\}$ . For the quantization step size adaption  $\Delta_l(n)$  of the lsf residual quantizer, the same algorithm is used as for the prediction residual.

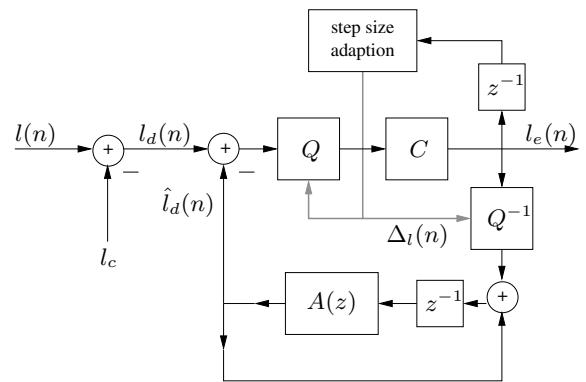


Fig. 4: Encoding scheme for the pre-filter and predictor coefficients used for the proposed ULD encoder.

Although this coefficient coding scheme only transmits differences, the decoded filter coefficients will converge after random access or transmission errors of the bit stream, due to the minimum phase property of the prediction polynomial  $A(z)$  and because  $\beta < 1.0$ .

With the proposed structure, the constant bit rate constraint is met without using a distortion loop. Due to the block-wise forward adaption of the lpc coefficients and the applied coding scheme, no explicit reset of the predictor is necessary.

### 3. LISTENING TEST

Both the old and the proposed coder structures were tested with the ULD coding scheme in a listening test according to the MUSHRA standard [7], where anchors were omitted. The MUSHRA test was implemented on a Laptop computer with external DA-converter and STAX amplifier/headphones in a quiet office environment. The group of eight test listeners consisted of expert and non-expert listeners. Before the subjects started with the listening test, they had the possibility to listen to a test set.

The tests were conducted with 12 mono audio files of the MPEG test set, all with a sampling frequency of 32 kHz: es01 (Suzanne Vega), es02 (male speech, German), es03 (female speech, English), sc01 (trumpet), sc02 (orchestra), sc03 (pop music), si01 (cembalo), si02 (castanets), si03 (pitch pipe), sm01 (bagpipe), sm02 (glockenspiel), sm03 (plucked strings).

For the unmodified ULD structure, backward adaptive prediction with length 64 was used, together with a backward adaptive golomb coder for entropy coding, coded at a constant bit rate of 64 kbit/s. For the proposed modified ULD coding scheme, a forward adaptive predictor with length 12 was used and the number of different quantizer steps was limited to three ( $\forall n : i_c(n) \in \{-1, 0, 1\}$ ). This, together with the coded side information, results in a constant bit rate of 64 kbit/s, too.

The results of the MUSHRA listening test are presented in Fig. 5, showing means 95%-confidence intervals for the twelve test items as well as overall results. As long as the confidence intervals overlap, there is no statistical significant difference between the coding methods.

Item es01 (Suzanne Vega) is a good example for the superiority of the proposed noise shaping method at lower bit rates. The higher parts of the decoded signal spectrum show less audible artifacts compared with the unmodified ULD coding scheme. This results in a significantly higher rating of the proposed scheme.

The signal transients of item sm02 (Glockenspiel) cause a high bit rate demand for the unmodified ULD coding scheme. When used with 64 kbit/s, the unmodified coding scheme produces annoying coding artifacts for complete blocks of samples. With the proposed scheme, however, the perceptual quality significantly improves.

The overall score (all items) of the proposed scheme showed a significantly better rating than the unmodified scheme, too. The overall rating of the proposed scheme is "good audio quality" under the given test conditions.

#### 4. CONCLUSIONS

We presented a modified low delay audio coding scheme which uses block-wise forward adaptive prediction together with clipping instead of backward adaptive sample-wise prediction. Thus, a different noise shaping method is applied for constant bit rate coding. For evaluation at lower bit rates, we conducted a listening test. The listening test showed that the proposed coding scheme outperforms the backward adaptive method in case of lower bit rates. Hence the presented coder is a candidate to reduce

the bit-rate gap between high quality speech coders and low delay audio coders.

#### 5. REFERENCES

- [1] M. Lutzky, G. Schuller, M. Gayer, U. Kraemer, S. Wabnik, "A guideline to audio codec delay", presented at the 116th AES convention, Berlin, May 2004.
- [2] B. Edler, C. Faller and G. Schuller, "Perceptual Audio Coding Using a Time-Varying Linear Pre- and Post-Filter", presented at the 109th AES convention, Los Angeles, September 2000.
- [3] U. Krämer, G. Schuller, S. Wabnik, J. Klier, J. Hirschfeld, "Ultra Low Delay audio coding with constant bit rate", presented at the 117th AES Convention, San Francisco, October 2004.
- [4] S. Wabnik, G. Schuller, J. Hirschfeld, U. Krämer, "Packet Loss Concealment in Predictive Audio Coding", Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, New York, Oct. 16-19, 2005.
- [5] N.S. Jayant and Peter Noll, "Digital Coding of Waveforms", Prentice-Hall Signal Processing Series.
- [6] S. Wabnik, G. Schuller, J. Hirschfeld, U. Krämer, "Different Quantization Noise Shaping Methods for Predictive Audio Coding", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, May 2006.
- [7] ITU-R BS.1534-1, "Method for the subjective assessment of intermediate quality levels of coding system", January, 2003.

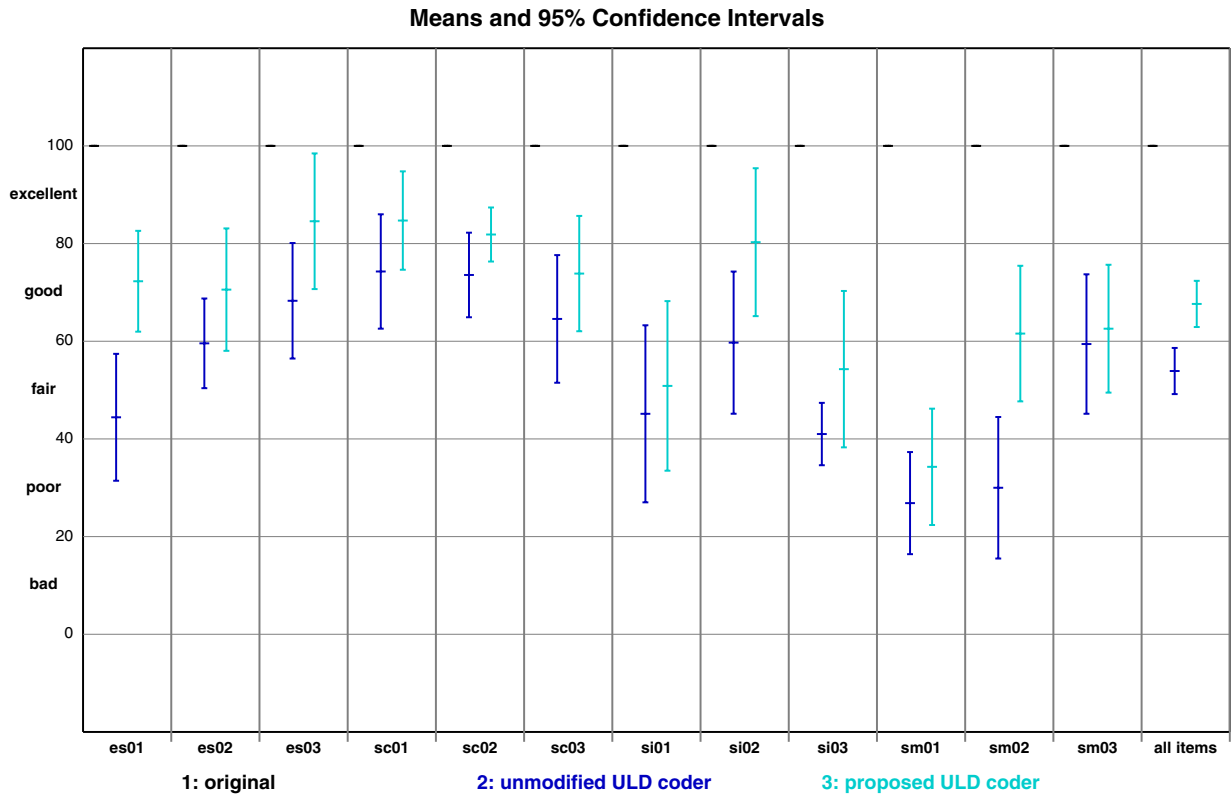


Fig. 5: Listening test results for the unmodified ULD coding scheme and the proposed ULD coding scheme.