

# A PERFORMANCE STUDY OF SPEECH CODERS FOR TELEPHONE CONFERENCING IN DIGITAL MOBILE COMMUNICATION NETWORKS

M. S. Lee\*, G. C. Lee\*, K. C. Kim\*, H. S. Lee\*, D. S. Lyu\*\*, D. J. Shin\*\* and Hun Lee\*\*

\*Department of Information and Communication Engineerin,  
Korea Advanced Institute of Science and Technology,  
207-43, Cheongryangri-Dong, Dongdaemun-Gu,  
Seoul 130-012, Korea

\*\*Mobile Switching Technology Section,  
Electronics and Telecommunications Research Institute,  
P.O.Box 8, Daeduk Science Town, Taejeon 305-606, Korea

**ABSTRACT** This paper describes two methods to assess the output speech quality of vocoders for telephone conferencing in digital mobile communication networks. The proposed methods are the sentence discrimination method and the modified degraded mean opinion score (MDMOS) test. We apply these two methods to Qualcomm code excited linear prediction (QCELP), vector sum excited linear prediction (VSELP) and regular pulse excited-long term prediction (RPE-LTP) vocoders to evaluate which vocoding algorithm can process mixed voice signal from two speakers better for telephone conferencing. From the experiments we obtain that the VSELP vocoding algorithm reveals superior output speech quality to the other two.

## 1. INTRODUCTION

In recent years, the conventional analog mobile communication system is rapidly transitioning toward the digital one to increase channel capacity. The analog mobile communication system can transmit mixed voice signal from two speakers for three-way on telephone line. However in the narrowband digital mobile communication system, we can normally take limited operating modes for telephone conferencing such as the simplex broadcast because of the inability of coding algorithms to represent mixed voice signals from multiple speakers [1].

The three-way conferencing service has already been provided in the conventional telephone networks. In the near future, this service must be provided in digital mobile communication too. Therefore, we investigate speech coders for digital mobile communication that can model mixed voice signals from two speakers successfully. The selected coding algorithms for this purpose are the QCELP, the VSELP, and the RPE-LTP vocoders.

In order to compare the output speech quality of each vocoder for mixed voice signals from multiple speakers, we can not use the conventional speech quality measures since they are for a single speaker. Thus we propose two speech quality measures for vocoding mixed speech signals from multiple speakers. They are the sentence discrimination method and the modified degraded mean opinion score (MDMOS) method. In the sentence discrimination method, listeners hear the mixed voice spoken by two speakers, and then write down the sentences that they have heard. In the conventional degraded mean opinion score (DMOS) test, the listeners hear two sentences: the first sentence is the original one not

processed by a vocoder, and the second sentence is the one processed by the vocoder. The proposed MD MOS measure differs from the conventional DMOS measure in the sentence test procedure, which will be described in Section 3.

In Section 2, we provide a brief description of the conventional speech coders used in digital mobile communication. The quality assessment procedures for the vocoding algorithms to encode mixed voice signals from two speakers are described in Section 3. In Section 4, we present and discuss the assessment results. Finally, conclusions are made in Section 5.

## 2. SPEECH CODERS FOR DIGITAL MOBILE COMMUNICATION

Digital mobile communication provides many advantages over the conventional analog one. For example, it increases channel capacity and improves noise immunity, and also offers the ability to use encryption. Thus much effort has been made toward the development of a high quality voice coder operating at a relatively low bit rate.

There are three methods to code speech signal: waveform coding, voice coding (vocoding), and hybrid coding [2, 3]. Waveform coding methods can produce high quality speech but require relatively high transmission bit rates. Voice coding is usually based on the speech production model. It can significantly reduce the bit rate, but its speech quality is synthetic. The hybrid coding methods are based on the time-varying excitation model by using both the waveform coding techniques in computing excitation signals and vocoding techniques in estimating vocal tract model parameters. The hybrid coders can produce high quality speech with relatively low transmission bit rate. These include VSELP [4], CELP [7], RPE-LTP [5, 6], residual excited linear prediction (REL P), and improved multi-band excitation (IMBE) vocoders.

In this paper, we select the QCELP, the RPE-LTP and the VSELP coders to compare the output speech quality for telephone conferencing. They are adopted as the standards for digital cellular systems in many countries. The 8 kbps VSELP coder is used as the standard for time division multiple access (TDMA) digital cellular in the United States, the 13 kbps RPE-LTP coder is selected as the European digital cellular standard, and the QCELP is selected as the code division multiple access (CDMA) digital cellular standard.

## 3. QUALITY ASSESSMENTS

One of the most difficult problems in speech coding is the assessment of the relative performance of different coding systems. In order to evaluate the output performance of speech coding methods fairly, a variety of quality assessment techniques have been formulated. The assessment techniques are generally fall into two classes: subjective measures and objective measures. A group of listeners compare the original speech signals with the vocoded ones, then subjectively rank the quality of speech on the basis of a predetermined scale. These subjective measures are adequate in performance comparison between vocoders [3]. On the other hand, the objective speech quality measures are based on the mathematical comparison of the original speech signals with the coded ones. These measures are appropriate to performance test of waveform coders [3].

In telephone conferencing, it is natural to transmit the mixed voice signals from multiple speakers. In this case, the speech quality measures stated above are not suitable to evaluate the output speech quality of the vocoders in the telephone conferencing since those measures are mainly for encoding voice signal from a single speaker.

In this paper, we propose two methods to evaluate the performance of vocoders for voice signal from two speakers. The first one is the sentence discrimination method. In this method, listeners should write down two different sentences from the mixed voices spoken by two speakers. The purpose of this method is to test the voice discrimination capability of a specific vocoding algorithm. Then, the test results are obtained by computing the scores per sentence with the score given in Table 1. For this test, each sentence is composed of 3 phrases.

Table 1: Scores assigned to each test in the sentence discrimination method.

Score/Sentence	Case
1	correct sentence
0.5	2 phrases : correct, 1 phrase : error
0	otherwise

The second one is the MDMOS test. The MDMOS test uses pairs of speech signals which are differently processed by vocoders unlike the DMOS test. The first one of each pair is obtained by vocoding each sentence and then mixing them, and the second of each pair is obtained by mixing two sentences first and then vocoding the mixed utterance. The quality of the second one is compared with the first one and scored using Table 2. Using the two kinds of assessment measures proposed in this paper, we can determine which vocoding algorithm is more suitable to model the mixed voice signals from two speakers for three-way telephone conferencing.

Table 2: Score table used in the MDMOS test.

Score	Description
5	degradation is inaudible
4	degradation is audible but not annoying
3	degradation is slightly annoying
2	degradation is annoying
1	degradation is very annoying

## 4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we present results of the MOS test for a single speaker, and those of the sentence discrimination method and the MDMOS test for two speakers.

### 4.1 Results of the MOS Test for a Single Speaker

For the MOS test, we use a database which consists of two sentences with the duration of 6 and 10 seconds, respectively. The vocoded speech samples are obtained using the three vocoding algorithms, and ten listeners join in the assessment test. The results of the test are listed in Table 3.

In this test, we should consider the transmission rate of each vocoder when we evaluate the output speech quality. The average transmission rate of QCELP is 5 kbps for the first

Table 3: Results of the MOS Test for Three Vocoders.

Vocoder	QCELP	VSELP	RPE-LTP
Score	3.65	3.64	3.37

sentence and 6.7 kbps for the second sentence. In VSELP and RPE-LTP, the transmission rates are fixed to 8 kbps and 13 kbps, respectively.

#### 4.2 Assessment Results of Vocoders for Mixed Voice Signals from Two Speakers

Ten simple and meaningful sentences composed of 3 phrases are used as a database in the sentence discrimination method and the MDMOS test. For the fair comparison, the transmission rate of the QCELP vocoder is fixed to 8 kbps.

##### 4.2.1 Test Results of the Sentence Discrimination Method

The test results of the sentence discrimination method are shown in Table 4, where each score is represented as a percent of cumulated scores obtained by using Table 1.

Table 4: Results of the Sentence Discrimination Method.

Vocoder	QCELP			VSELP		
	F7F8	M4M6	M5M9	F5F6	M2M3	M1F4
Score(%)	87.5	45	85	90	75	87.5
Total(%)	72.5			84.17		

We omit the results for RPE-LTP since its scores are very low. In the table, F and M represents female and male speakers, respectively. For instance, M5F9 represents the mixed voice of the 5th sentence spoken by a male speaker and the 9th sentence spoken by a female speaker. As shown in Table 4, the QCELP and the VSELP vocoders discriminate the sentences well. In the QCELP test, the low score of M4M6 is caused by the fact that the two male speakers have very similar tones.

##### 4.2.2 Results of the MDMOS Test

The speech signals for the MDMOS test are obtained in two ways. The first one is to vocode two utterances separately and then mix them as shown in Fig.1(a), while the other is to mix two utterances first and then vocode the mixed output signal as shown in Fig.1(b).

The results of this test are given in Table 5. We obtained very similar performance results of the vocoding algorithms to those of the MOS test for a single speaker.

## 5. CONCLUSIONS

In this paper, we propose two quality assessment methods: the sentence discrimination method and the modified DMOS (MDMOS) test for evaluating vocoder performance in the telephone conferencing environment. Using these two methods, we obtain the performance of three vocoding algorithms, QCELP, VSELP and RPE-LTP for telephone conferencing. From the two performance tests, we obtain the consistent results that VSELP reveals the best performance and then QCELP follows. As a result, we can use the VSELP and the QCELP vocoding algorithms for three-way telephone conferencing.

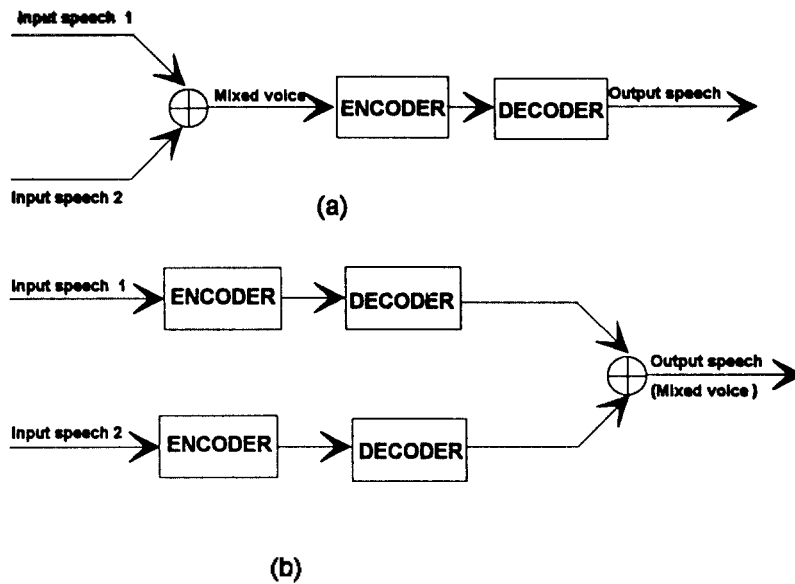


Figure 1: Test Speech Signal Generation for the MDMOS Test : (a) Test speech from mixing vocoded output speech signals, (b) Test speech from vocoding mixed input speech signal.

Table 5: Results of the MDMOS Test for Mixed Voice Signals.

Vocoder	QCELP	VSELP	RPE-LTP
Score	4.07	4.28	4.05

## References

- [1] J. D. Tardelli, P. D. Gatewood, E. W. Kreamer and P. A. La Follette, "The Benefits of Multi-Speaker Conferencing and the Design of Conference Bridge Control Algorithm," *Proc. ICASSP 93*, pp. 435-438, 1993.
- [2] C. Wheddon and R. Linggard, *Speech and Language Processing*, Chapman and Hall, 1990.
- [3] Panos E. Papamichalis, *Practical Approaches to Speech Coding*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1987.
- [4] *Motorola Inc., Digital Signal Processor Operations Austin, Texas.* "Principles of Vector-Sum Excited Linear Predictive(VSELP) Speech Coder and Its Implementation on the DSP56156".
- [5] Peter Kroon, Ed F. Deprettere and Rob J. Sluyter, "Regular-Pulse Excitation-A Novel Approach to Effective and Efficient Multipulse Coding of Speech," *IEEE trans. on Acoust. Speech Signal Processing*, Vol. 34, no.5, pp.1054-1063, Oct., 1986.
- [6] Ed F. Deprettere and Peter Kroon, "Regular Excitation Reduction for Effective and Efficient LP-Coding of Speech," *Proc. ICASSP 85*, pp. 25.8.1-25.8.4, 1985.
- [7] Manfred R. Schroeder and Bishnu S. Atal, "Code Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates," *Proc. ICASSP 85*, pp.25.1.1-25.1.4, 1985.