1. INTRODUCTION

In our society, mobile communications have become a need for people and a major objective of research. The perception of coded speech under real-world (noisy) transmission conditions is an important aspect of this area, with several implications into the intelligibility and quality of existing and/or new services (e.g., mobile satellite communications, cellular mobile telephony, etc.) and into the design of efficient and robust speech coding systems.

In this work, two speech digitizers, that is a Residual Excited Linear Predictive (REL) coder at 2.4 kbit/s [1], and a Linear Predictive Coder (LPC) at 2.4 kbit/s [2] have been simulated, and incorporating an error protection scheme, provides a moderately good quality, while the 2.4 kbit/s LPC vocoder yields a quality that is felt to be commercially acceptable.

2. CODING TECHNIQUES

Both speech compression algorithms and channel models were simulated in real time on an array processor FPS-120 B connected to a VAX 11/785. Speech systems run in half duplex and use a specific audio processing front end with 16-bit A/D/D/A converters. The input speech is band-limited to 200–3400 Hz and sampled at 8 kHz. An automatic gain control circuit permits a suitable reduction of the input dynamic range. The 2.4 kbit/s LPC is based on a 16th order autocorrelation analysis performed every 22.5 ms. An AMFR pitch extractor with median smoothing technique is used in these experiments to obtain voiced/unvoiced information, the energy ratio between high and low frequency regions.

The 2.9/6.6 kbit/s RELP coder used in this study is a perceptual model error protection scheme based upon the available detailed information concerning the speech. The residual signal is quantized and transmitted, and the regeneration of the full band excitation signal is performed at the receiver using the spectral folding method [8]. The 9.6 kbit/s RELP incorporates a 16-state HMM recognizer for bit interleaving and bit protection with error correcting codes. The former mechanism is aimed at splitting a long error burst into several bursts (ideally, into isolated errors), thus allowing, through a sort of "divide-and-conquer" strategy, easier protection of the most important parameters in the data frame. The latter mechanism protects the remaining code and through the use of four (15,5) BCH codes, and the r.m.s. value of each frame using a (12,4) code. The first code can correct 1 or 2 errors, while the second can correct 1 or 2 errors. Residual samples are left to the channel mercy. Overall, the frame format of the 2.4 kbit/s RELP is 121 bits, of which 45 bits correspond to voice information, 48 bits of error protection and 2 bits for synchronization. Filtering the LPC vocoder and the 1.2 kbit/s RELP do not exploit error protection.

3. PROCEDURE

A set of four DRT lists was selected for the experiments, which were performed according to the procedures of English isolated words, read by native American speakers (2 lists read by males, and 2 read by females). These lists performed in a rural environment using an Altic 659A dynamic microphone without a phone screen.

Four different circuit conditions have been examined, combining the three coding bit-rates with two typical channels, as stated in the introduction. Output signals of the processed stimuli have been recorded on analog tapes and then used for the subjective test. Eight listeners took part in the DRT experiments, that were conducted at the Dynacore Inc. (Austin, Texas) in-house speech evaluation facility.

3.1 Structure of the DRT

The DRT of Voiers [3] is based on discrimination between two rhyming monosyllabic words that differ, mainly, in voicing, and can be difficult or impossible to distinguish. In these circumstances the task is simply to indicate which word has been presented. Word pairs are chosen so that initial consonants differ for only one distinctive feature. The sign of the gain means "does not apply". Table 1 shows an example of stimulus words used in the DRT test.

Scores for the voiceless (absent) state of the feature, the sign means "(absent)", and the circle means "does not apply". The scores obtained for only one distinctive feature are given in Table 1, in which the sign + means positive (present) state of the feature, while the sign - means negative (absent) state. Therefore, the feature compactness relates to the voicelessness of the consonants. The nasality feature, which distinguishes /n/ from /m/ and /n/ from /n/, is given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants.

4. RESULTS AND INTERPRETATIONS

The gross scores of the six critical phonemic features considered in the DRT are plotted in Fig. 1. The scores differences over subcategories are high, especially for the voiceless feature. The scores according to the voiceless state or specific Phoneme cues are shown in Table 3.

Noteworthy are the consistent depressions on the voicing, graveness and sustentation components for the conditions 5 and 6. In fact, these phenomena separate the RELP coders from the LPC vocoder.

The voicing feature distinguishes the voiced counterparts from their voiceless counterparts /p/ from /f/, /b/ from /v/, /d/ from /t/, /m/ from /m/, etc. For the conditions 5 and 6, there is a strong bias towards the friction feature. The graveness scores show a bias favoring the absent state, which is similar for both voiced and voiceless consonant pairs. This bias is larger for male speakers than for female speakers.

The gross scores of the six critical phonemic features of voice signals are given by the voicelessness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants. The voiceless, graveness and compactness features are given by the loudness of the consonants.
differences between the voiced and unvoiced states of the compactness feature.

The sibilation feature, which distinguishes /s/ from /z/, /θ/ from /ð/, etc., shows a bias towards the absent state, indicating that strident consonants can be reproduced with more "s" cues. This effect is due to the incompleteness of the excitation signal, as discussed for the sibilance feature.

The maximum degradation in going from the channel No.1 to the channel No.2 is about 5 points for the sibilance feature. In particular, comparing the performance of RELP coders, we note that the error protection implemented by RELP at 9.6 kbit/s seems to be more useful to preserve this feature along with sibilance for (for the channel No.1) and graveness (for the channel No.2). In fact, large amounts of sibilance feature information are carried in the duration and spectral characteristics of adjacent vowels, as well as in the acoustical manifestations of the consonants. Therefore, the error protection of spectral parameters from k(1) to k(4), particularly adequate for vowels, gives benefits also to certain consonants. Of course, loss of information in the upper frequency formants may cause significant degradations. The robustness of nasality for all the conditions, and of sibilance for RELP configurations, is clearly evident. Also compactness, which depends on, among other things, the higher second-formant frequencies, appears somewhat robust for all the conditions.

Overall, the DRT scores show the remarkable robustness of the 9.6 kbit/s RELP system, even in case of multipath fading degradation.

The performance of the LPC system is mainly impaired on the voicing, graveness and sibilance features, which are generally quite fragile in all vocoders and sensitive to various forms of speech degradation.

5. CONCLUSIONS

We have simulated in real time two speech coding systems at low bit-rates, suitable for mobile satellite communications. We have evaluated their robustness against channel impairments, using the DRT facility, and got useful information to tradeoffs between important issues such as power, bandwidth, voice quality, and delay. It turns out that a 9.6 kbit/s RELP coder is capable of preserving very good intelligibility, provided that the most important parameters of the side-information are protected with a combination of bit interleaving and error-correcting codes. Short codes must be used. In fact, in addition to being simpler to decode, short codes are more adequate than long ones when the error probability of the channel is large. In particular, a (15,5) BCH code and a (12,4) code have proven to be suitable for our purpose.

Comparing the DRT scores, it results that two subjective categories are gained by the 9.6 kbit/s RELP over the 2.4 kbit/s LPC system. Indeed, the ability to yield fair quality at 2.4 kbit/s using conventional vocoders remains to be seen. Should this happen, however, it could allow an additional reduction of 4 in power and bandwidth.

Recent speech coding algorithms [5-13] provide high quality speech somewhere between 4 and 8 kbit/s, under ideal transmission conditions. Therefore, future problems to be addressed are those associated with their subjective performance in presence of environmental noise, channel errors and multipath fading.

ACKNOWLEDGMENTS

This study has been supported in part by ESA under ESTEC Contract No. 6098/84/NL. The authors would like to thank W. Krietsch, Agency's representative, for his cooperation. Thanks also go to E. Biglieri and G. Albertengo, from Politecnico of Torino, for the simulation of channel models and error-correcting codes.

REFERENCES


