Speech compression: a key to low-cost telephony systems for developing countries

Geraldo Lino de Campos
Deptartment of Computer Engineering, Escola Politécnica da Universidade de São Paulo,P. O. Box 8174 - São Paulo 01065-970 - BRAZIL.*
e-mail:RTC@FPSP.FAPESP.BR

Abstract

Speech compression allows one order of magnitude reduction in the bandwidth required for normal telephone conversation without loss of speech quality. This paper presents proposal for utilizing this capability for reducing the cost of telephonic systems for developing countries.

Keyword codes: B.4.1; H.4.3
Keywords: Data Communications Devices; Communications Applications

1. INTRODUCTION

Speech compressing techniques have evolved to a point were a rate of 4.8 Kb/s is sufficient to transmit voice with quality that exceeds normal telephone grade voice. As carriers still use 64 Kb/s for transmission of a voice signal, this effectively wastes more than 90% of the world's long range transmission capacity, a scarce, valuable and expensive resource. This waste is specially significant for developing countries, where capital is the most scarce resource.

This paper proposes the use of voice-only phone lines, using compression techniques and 4.8 kb/s per channel, multiplexing up to 26 subscribers in one line (subscriber loop), using the chips and technologies already developed for ISDN, and multiplexing up to 13 voice channels in each toll connecting trunk, as well in intertoll trunks. This will result in a low-cost system for developing countries.

In this paper all references will be to the international telephony standards; the same reasonings, with slightly changed figures, apply to the American standards.

---

* This work was partially supported by CNPQ grant 521565/93-6 and FAPESP grant 92/1088-0.

This paper is organized as follows. Section 2 presents the fundamentals of speech compression; section 3 briefly describes the CELP coding algorithm; section 4 shows how to apply voice compression for reducing the cost of telephone systems, and section 5 presents some concluding remarks.

2. BASIC MECHANISMS OF SPEECH COMPRESSION

Speech compression is not a new subject; the first analog version dates back to 1939 [1], and many digital systems exist. However, either the compression was small (in the range of 2:1 to 4:1), or the quality poor.

Modern compression systems analyze the speech and reduce it to a set of parameters. These parameters are transmitted and the speech is reconstructed at the receiving end. Almost all systems use knowledge about the process of speech generation in the process of analysis and synthesis.

Human speech is generated by the excitation of the vocal tract either by a train of pulses (voiced sounds), generated by the vibration of the vocal cords, or by white noise generated by a constriction in the vocal tract (unvoiced sounds).

The vocal tract behaves acoustically as a time-varying acoustic tube; its characteristics can be modelled by a sequence of small acoustic tubes, with time varying diameters, generating resonance peaks called *formants*. Modeling the resonances of the vocalic tract is easily and faithfully performed by a technique called *Linear Predictive Analysis* (LPA), which generates an all-pole model form the resonances of the vocal tract. The difficult part is modeling the excitation.

Until recently, the excitation was modeled as a simple train of pulses or pure white noise; the resulting speech was intelligible, but quality was very poor and acceptable only for specialized applications, like secure communications. Moreover, the systems was very sensitive to ambient noise.

In recent years, several algorithms have been proposed for modelling the excitation in a more faithful way. The main innovation is to substitute artificial excitation models, like simulated glottal pulses or white noise, by a sequence of pulses that represent the excitation better. They vary in details of quantization, number of parameters, computational requirements (always high) and offer good to excellent quality. Of these, the CELP algorithm offers excellent quality, robustness, and acceptable computational requirements.

3. CELP CODING

CELP (Code Excited Linear Prediction) is a recent development [2-4], and has been adopted as an US Federal Standard [5]. It uses the usual LPA for determining the resonance of the vocal tract; the resulting parameters are changed to a mathematically equivalent formulation, called *Line Spectrum Pairs,* that exhibits better quantization properties, allowing a representation with fewer bits.

CELP's main innovation is the way to handle the excitation. Very briefly, CELP coding is based on analysis by synthesis search procedures. First, a standard LPA is performed, resulting in a 10^{th} order filter. A set of different excitations are applied to the filter. Each output is compared with the input signal, using a perceptually weighted distance measure. The excitation with the least distance is selected and transmitted. Obviously, it is necessary to reduce the possible excitations to a number as small as possible. This is accomplished by a fixed *stochastic codebook* for representing the unvoiced components of speech, and an *adaptive codebook* for the voiced components.

Although the computational requirements are high, requiring a digital signal processor with 25 MIPs for real-time operation, such processors are readily available and of low cost. In [6], a figure of US$200.00 is quoted as a typical value for a one channel CELP system, if produced in small quantities. Mass market should lower this figure considerably.

The resulting system produces a speech that is almost indistinguishable from natural speech, when recorded in an office environment. As it uses a larger bandwidth, 3.7 instead of 3.2 KHZ, the perceived quality is better than the obtainable with the usual phone system - in this context, called POTS: Plain Old Telephone System. More important, it is very robust to environment noise, a feature that lacked dramatically in older systems.

The required bandwidth can be as small as 4800 b/s without noticeable loss of quality. Since a conventional telephone channel is specified at 64 Kb/s, it can carry up to 13 compressed voice channels.

4. APPLICATION TO LOW-COST TELEPHONY

Use of CELP to the general telecommunications facilities is relatively inexpensive, as compared to expanding a conventional system. Economies result from two applications: in toll connecting trunks and in the subscriber loop.

To put things in perspective, a very simplified description of the telephone system is required. Each telephone has two copper wires coming out of it that go to the telephone company *end office*. This connection is called the *local loop*, and is always an analog connection. If a subscriber call another telephone attached to the same end office, the switching mechanism within the office sets up a direct (analog) connection between the two local loops.

When a call is made to a telephone attached to another end office, a different procedure is required. Each end office has a number of outgoing lines to nearby switching centers, called *toll offices*; these lines are called *toll connecting trunks*. If the caller and callee do not have a toll office in common, higher level connections are provided by *intertoll trunks*. Except for some very old equipment, these trunks are entirely digital. Each analog voice signal is digitized at the end office, producing a bit stream of 64 Kb/s.

4.1. Application in toll connecting

An obvious way of using CELP in the general phone system would be to maintain the present user facilities, and compress the speech signal when the end office determines the need to use any long-range facility (toll and intertoll trunks). This means that the long-range facility may be dimensioned as a fraction of a conventional system. Cost and tariffs should be reduced accordingly. The percentage of

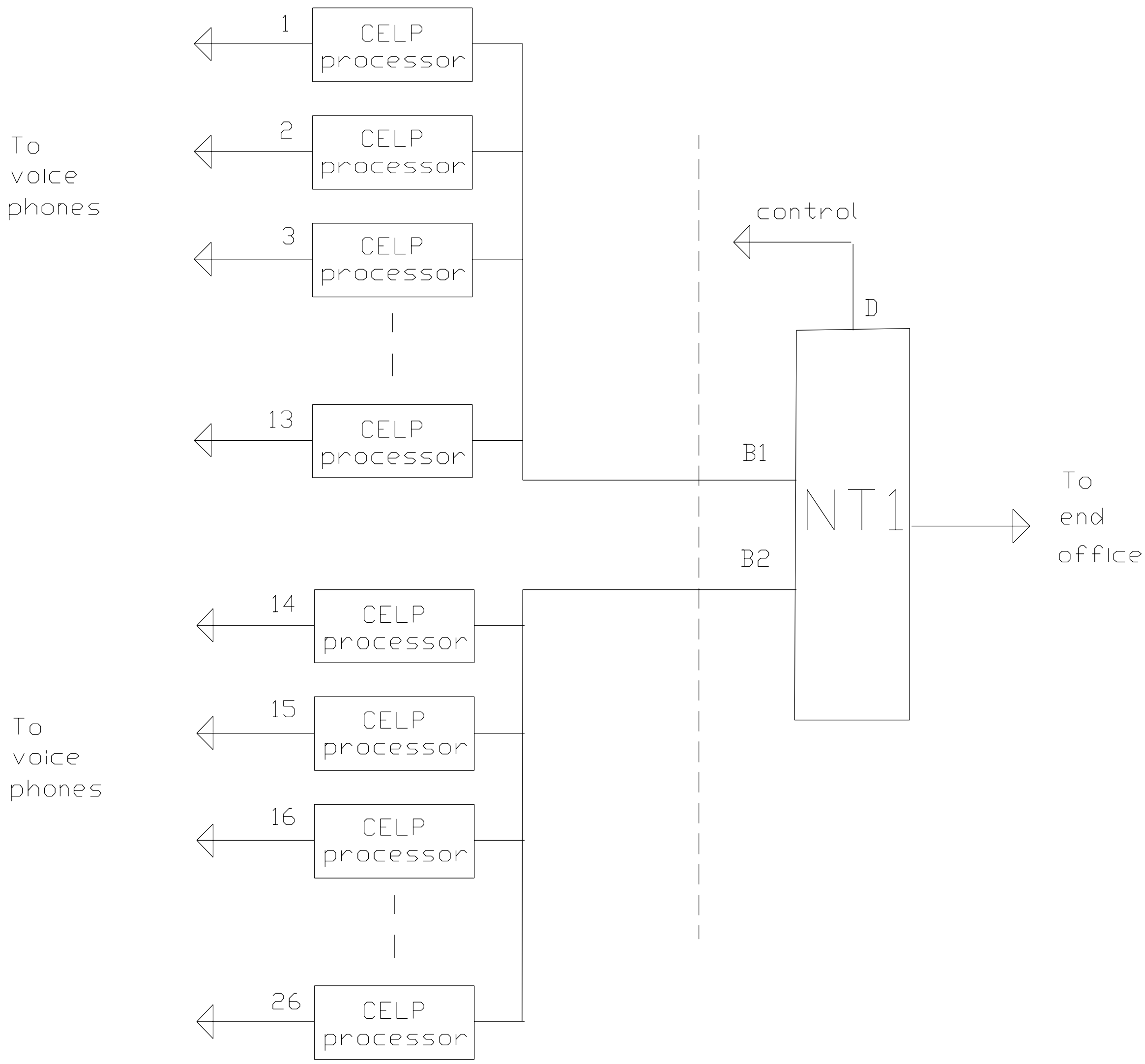

Figure 1 - Proposed NT26
The dashed line represents the standard T point. Components at right are standard ISDN chips.

the channels reserved for full bandwidth transmission is the limiting factor to this reduction.

It has, however, an operational drawback: it requires different circuits for voice only and for voice plus other services, ranging from transmitting music to use FAXes or modems. Nevertheless, this is not a severe problem in developing countries, where the main use of the telephone system is for voice. The user should have, of course, the right to chose between the conventional and more expensive system and the cheaper voice-only system.

The effective reduction may vary according to several factors. As several exchanges already have provisions for subdivision of the 64 Kb/s channel in eight 8 Kb/s channels, this might be a solution dictated by technical constraints. Use of twelve 5 Kb/s channels fits nicely in the ISDN frame: 2 bits out of every three frames; 13 is the maximum value for reserving 4800 b/s per channel; 16 is probably possible, with a negligible loss of quality; this last possibility requires further research, however. In this paper, the number 13 will be used as an example only; it is not implied to be the best solution.

An important consideration is that some kind of data communication is still possible on these voice-only lines: a smart interface can identify a modem operating up to the v.32 level, recover the digital information and transmit in digital form up to the destination, and there reconstruct the modem signals. The processing requirements form doing this are considerable lower than voice compression and decompression, so it adds only to the cost of software.

4.2. Application in the subscriber loop

Another important source of economy can be realized by combining CELP and ISDN [7-8] technologies. ISDN (Integrated Services Digital Network) is a redesign of the telephone system; for the purpose of this paper, it suffices to say that it changes the local loop to a digital format and offers added features.

The conventional telephony system requires a physical connection between each user and the end office, called the subscriber loop. The cost of the subscriber loop is significant, and can also be reduced.

The subscriber loop can be especially expensive in low density areas, where the terminal are spread over a large geographic area, requiring long lines.

Subscriber loop technology for ISDN systems are already developed, and it can handle the equivalent of two conventional lines over one subscriber loop. The technology is ready and there are commercial chips available for immediate implementation.

Consider now a new ISDN box, able to squeeze 13 compressed voice phone channels in each one of an ISDN B (64 Kb/s) channel of a basic rate ISDN installation. Lets call it NT26 - Network Terminator for 26 speech Compressed Channels (fig. 1). Mass produced, a NT26 should be rather inexpensive.

The NT26 should be installed in a short distance from the subscribers. One conventional subscriber loop connects the NT26 to the end office. Up to 26 short-haul loops connect the outputs of the NT26 to up to 26 subscribers in a neighborhood - a solutions that is much less expensive than to have 26 conventional subscriber loops. Moreover, the above proposal allows all of the 26 subscribers to be active at once. If further components are acceptable at the NT26 box, with the inclusion of a small switching central, a larger number of subscribers, in the range of 100, could be accommodated.

5. CONCLUSIONS

The proposal is to have two kinds of telephone lines: conventional full-service lines, and economic voice-only lines. Voice-only lines are implemented trough speech compression, allowing up to 13 lines to be multiplexed in a 64 Kb/s trunk. Further economy can be obtained by concentrating locally up to 26 lines, and using only one subscriber loop, with ISDN technology, to the end office.

The user of the voice-only lines will also be able to use low-speed data transmission, up to 2400 b/s, which is enough for many domestic activities, like using bulletin board systems.

The proposed approach offers many benefits at a relatively small cost. It is clear that one can not expect reduction of 13 times per line installed, but figures in the range of 3 to 5 over conventional lines can be realistically expected.

Cost reduction should be larger in long distance calls, which now are prohibitively high for most people in developing countries.

Bibliography

1. Dudley, H. The Vocoder, Bell Labs Rec 18, 122-126, Dec 1939.
2. Campbell,J.,Welch, V. and Tremain, T. "An Expandable Error-Protected 4800 bps CELP Coder", Proc. of ICASSP, 735-738, 1989
3. Lin, D. "New Approaches to Stochastic Coding of Speech Sources at Very Low Bit Rates", Signal Processing III: Theories and Applications, Proc. of EUSIPCO-86, 445-448, 1986.
4. Cox, R., Kleinj, W. and Kroon, P. "Robust CELP Coders for Noisy Background and Noisy Channels", Proc. of ICASSP, 739-742, 1989.
5. Federal Standard 1016 - Telecommunications: Analog to Digital Conversion of Radio Voice by 4800 bit/second Code Excited Linear Prediction (CELP), National Communications System, Office of Technology & Standards, Washington, DC 20305-2010, Feb. 1991.
6. Details to Assist in Implementation of Federal Standard 1016 CELP. Office of the Manager, National Communications System, Jan. 1992.
7. IEEE Journal on Selected Areas in Communications, special issue, may 1986.
8. Stallings, W. (ed.) Integrated Services Digital Networks. IEEE Computer Society Press, Loa Alamitos, Ca., 1991