Perception-Based Partial Encryption of Compressed Speech

Antonio Servetti, *Student Member, IEEE,* and Juan Carlos De Martin, *Member, IEEE*

***Abstract*—Mobile multimedia applications, the focus of many forthcoming wireless services, increasingly demand low-power techniques implementing content protection and customer privacy. In this paper low complexity perception-based partial encryption schemes for speech are presented. Speech compressed by a widely-used speech coding algorithm, the ITU-T G.729 standard at 8 kb/s, is partitioned in two classes, one, the most perceptually relevant, to be encrypted, the other, to be left unprotected. Two partial-encryption techniques are developed, a low-protection scheme, aimed at preventing most kinds of eavesdropping and a high-protection scheme, based on the encryption of a larger share of perceptually important bits and meant to perform as well as full encryption of the compressed bitstream. The high-protection scheme, based on the encryption of about 45% of the bitstream, achieves content protection comparable to that obtained by full encryption, as verified by both objective measures and formal listening tests. For the low-protection scheme, encryption of as little as 30% of the bitstream virtually eliminates intelligibility as well as most of the remaining perceptual information. Low-power, portable devices could therefore achieve very high levels of speech-content protection at only 30–45% of the computational load of current techniques, freeing resources for other tasks and enabling longer battery life.**

***Index Terms*—Encryption, low-power, multimedia security, speech compression, speech perception, speech transmission.**

I. Introduction

THE increasing relevance of multimedia applications is placing a great demand on content protection and customer privacy. Communications can be intercepted, especially over wireless links. Since encryption can effectively prevent eavesdropping, its use is widely advocated. Unfortunately, encryption and decryption are computationally demanding, a severe problem in mobile, portable devices, where power consumption needs to be reduced as much as possible. The need for encryption in wireless systems has led to intense activity aimed at reducing the complexity of encryption algorithms [1].

One solution to the problem of introducing encryption into power-constrained, real-time multimedia applications is partial encryption. Instead of encrypting multimedia signals in their entirety, only a subset of the bitstream is protected. The subset is chosen to cause, after encryption, the desired degree of degradation after decoding. Proposed partial encryption (sometimes referred to also as selective encryption) techniques have been developed for compressed image and video data. Efficient encryption of MPEG compressed video was proposed in [2]–[4]. More recently, the approach was extended to image compression [5], [6].

Speech and audio signals are an essential component of most multimedia applications. Not only speech services are the basis of the huge wireless telephony industry, but speech is also the most important component of advanced audiovisual services such as videoconferencing and news broadcasting. The benefits of partial encryption of speech signals could thus be very significant. We present partial encryption of speech compressed by the widely used algebraic code-excited linear-prediction technique. The proposed schemes deliver extremely effective content protection and can be straightforwardly extended to a number of international telephony standards based on the same algorithm.

The paper is organized as follows. Partial encryption of multimedia data is presented in Section II. The first part of Section III, after a brief review of speech compression techniques, describes partial encryption of G.729 compressed speech. The rest of Section III is devoted to the analysis of effective ways to evaluate performance. Section IV presents the experimental results obtained in the form of objective distortion measures as well as results from formal listening tests to demonstrate the effectiveness of the proposed partial encryption techniques.

II. Partial Encryption of Multimedia Data

Digital speech, audio, images and video bitstreams are characterized by *nonuniform perceptual importance*: the effects of errors can be much more pronounced for some bits than for others. So far, perhaps the most significant application of the nonuniform sensitivity of multimedia signals to bit errors has been Unequal Error Protection (UEP) schemes for multimedia transmission over wireless channels.

The same principle can also be applied to aid the introduction of content encryption in low-power, wireless multimedia scenarios. Instead of encrypting the multimedia stream (voice, audio, image or video) in its entirety, only a perceptually relevant fraction of the stream is subject to encryption, while the remaining part is transmitted unprotected. Fig. 1 shows the two alternative approaches: full content encryption is shown in Fig. 1(a), while perception-based partial encryption is shown in Fig. 1(b).

Encryption of only a fraction of the bitstream lowers the computational load, thereby freeing resources for other tasks or, in the case of portable devices, extending battery life. With multimedia mobile applications quickly becoming the focus of upcoming wireless services, the possibility of delivering content

Manuscript received August 7, 2001. This work was supported in part by CERCOM, the Center for Wireless Multimedia Communications, Torino, Italy; http://www.cercom.polito.it. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Harry Printz.

A. Servetti is with the Dipartimento di Automatica e Informatica, Politecnico di Torino, 10129 Torino, Italy (e-mail: servetti@polito.it).

J. C. De Martin is with the IEIIT-CNR, Politecnico di Torino, 10129 Torino, Italy (e-mail: demartin@polito.it).

Digital Object Identifier 10.1109/TSA.2002.804300

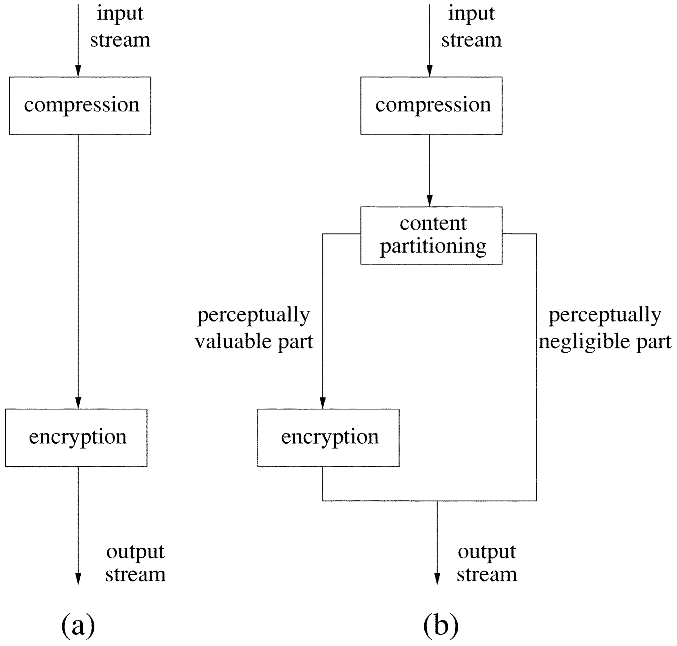


Fig. 1. Comparison between partial and full encryption schema.

protection and privacy with the minimum impact on power consumption is attractive and worth investigating.

III. Partial Encryption of Compressed Speech

The bits of a compressed speech bitstream are not perceptually equally important. Bit errors can have vastly different perceptual impact, depending on specifically which bit is corrupted. Non-uniform bit sensitivity has been exploited to create channel coding schemes that deliver different levels of protection to different classes of bits. Such Unequal Error Protection approach achieves, for a given capacity, significantly better performance than uniform protection of all bits.

Perceptually based partial encryption, too, depends on perceptual classification of the bitstream. The objective, however, is not to preserve perceptual quality above given levels after transmission and decoding, but, on the contrary, to cause the desired degree of signal content degradation. More formally, a pre-requisite of partial encryption is to identify *the smallest subset of the compressed bitstream that, if made unavailable due to encryption, causes the desired amount of degradation at the decoder.* A smaller subset would not offer enough content protection, while a larger one would wastefully increase system complexity, with adverse effects on the battery life of portable devices.

For speech, degradation may mean a decrease of *naturalness*, a decrease of *intelligibility*, or both. While in speech transmission the focus is mostly on the former, for speech protection the attention is essentially on the latter. The assumption is, in fact, that an eavesdropper is mostly, if not only, interested in the semantic *content* of the communication and, possibly, in the identity of the speaker and only marginally interested in the degree of naturalness of the intercepted speech. The protection of other, potentially sensitive, kinds of information such as traffic patterns or the fact that a targeted person is using the phone are not addressed by the proposed technique.

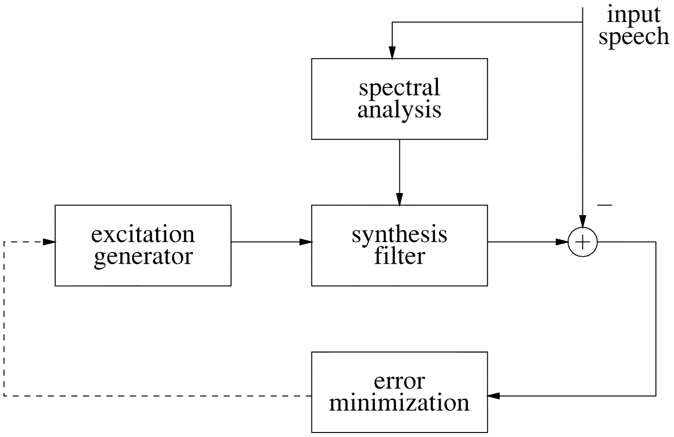


Fig. 2. Block diagram of generic linear prediction based analysis-by-synthesis speech coder.

The analysis of bit error sensitivities for speech transmission is usually carried out by systematically corrupting a given bit and then measuring the corresponding drop in performance over a speech database. Both objective distortion measures and listening tests are normally employed. Such approach leads to an arrangement of the bits in order of perceptual importance.

A. Compression of Narrow-Band Speech

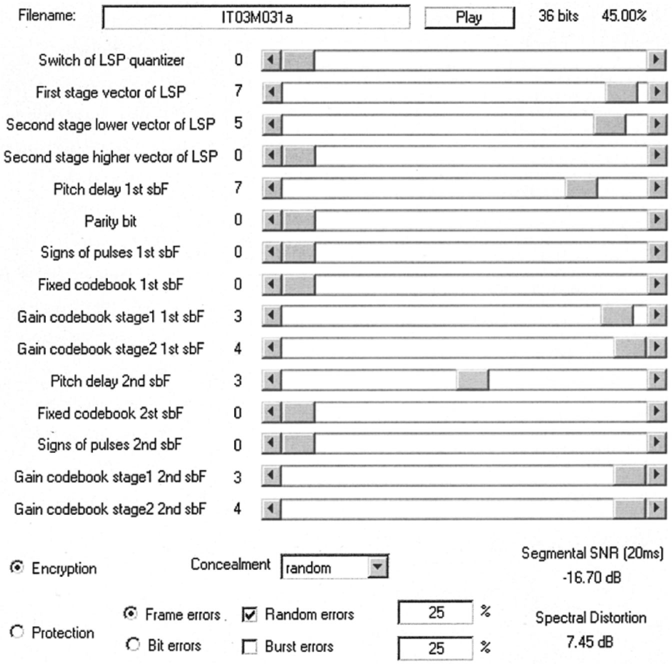
Development of partial encryption techniques for multimedia signals require understanding of the bitstream format as well as of its perceptual relevance. Since in most cases multimedia signals are transmitted compressed, analysis of the compression algorithms is important.

In this work, the focus is on partial encryption techniques for telephone-bandwidth (or narrow-band) speech. Voice remains the fundamental service offered by wireless operators. Even in future media-rich wireless applications, speech will always be the basic functionality characterizing wireless communications. Moreover, in audio-visual communications, such as videoconferencing, it is well known that in case of limited channel capacity, audio quality is to be preserved at the expense of video quality.

Telephone-bandwidth (300–3400 Hz), toll-quality speech can be represented digitally at 64 kb/s with nonuniform PCM, as in the ITU-T standard G.711. During the last decade, however, approximately the same quality has been achieved at bit-rates in the 6–12 kb/s range by linear-prediction based analysis-by-synthesis (LPAS) speech coding techniques (e.g., [7]). Fig. 2 shows the block diagram of a generic LPAS speech coder: excitation signals, usually taken from a codebook, are passed through the synthesis filter and the resulting signal is compared to the input speech. The excitation signal that generates the minimum error is selected and its index transmitted to the decoder. Efficient search techniques are possible when the structure of the codebook is deterministic. Sparse codebooks with unity gain pulses have shown good performance and limited complexity. The speech coding technique based on such codebooks, sometimes referred to as Algebraic Code-Excited Linear Predictive coding (ACELP) [8], [9] is the basis for a number of speech coding telephony standards, including the ETSI GSM 12.2 kb/s Enhanced Full-Rate [10], the ETSI GSM

TABLE I
BIT ALLOCATION FOR ITU-T G.729

Symbol	Description	Bits
L0	Switched MA predictor index of LSP quantizer	1
L1	First stage vector of LSP quantizer	7
L2	Second stage lower vector of LSP quantizer	5
L3	Second stage higher vector of LSP quantizer	5
P1	Pitch delay 1st subframe	8
P0	Parity bit for pitch delay	1
S1	Signs of fixed-codebook pulses 1st subframe	4
C1	Fixed codebook 1st subframe	13
GA1	Gain codebook (stage 1) 1st subframe	3
GB1	Gain codebook (stage 2) 1st subframe	4
P2	Pitch delay 2nd subframe	5
S2	Signs of fixed-codebook pulses 2nd subframe	4
C2	Fixed codebook 2nd subframe	13
GA2	Gain codebook (stage 1) 2nd subframe	3
GB2	Gain codebook (stage 2) 2nd subframe	4



Filename: IT03M031a

36 bits 45.00%

Switch of LSP quantizer 0
First stage vector of LSP 7
Second stage lower vector of LSP 5
Second stage higher vector of LSP 0
Pitch delay 1st sbF 7
Parity bit 0
Signs of pulses 1st sbF 0
Fixed codebook 1st sbF 0
Gain codebook stage1 1st sbF 3
Gain codebook stage2 1st sbF 4
Pitch delay 2nd sbF 3
Fixed codebook 2st sbF 0
Signs of pulses 2nd sbF 0
Gain codebook stage1 2nd sbF 3
Gain codebook stage2 2nd sbF 4

● Encryption
Concealment: random
Segmental SNR (20ms) -16.70 dB

○ Protection
● Frame errors ☑ Random errors 25 %
○ Bit errors □ Burst errors 25 %
Spectral Distortion 7.45 dB

Fig. 3. Matlab tool to apply selective encryption to a G.729 compressed bitstream.

4.75–12.2 kb/s Adaptive Multi-Rate [11], the TIA 7.4 kb/s IS-641 Enhanced Full-Rate [12], and the ITU-T 8 kb/s G.729 [13].

We chose to investigate partial encryption solutions for the ITU-T G.729 CS-ACELP speech coding standard. Extension to the other ACELP-based speech coding standards is straightforward. G.729 is widely used, most conspicuously in Voice over IP applications. It provides toll quality at 8 kb/s with low algorithmic delay and moderate complexity. The standard comprises several Annexes, including lower (6.4 kb/s) and higher (11.8 kb/s) bit-rate extensions, a voice activity detector and a low-complexity version.

To better understand partial encryption applied to G.729 compressed speech, the main features of the standard are briefly reviewed. The frame size is 10 ms (80 samples), divided into two subframes of 5 ms each. The spectral envelope information is computed at every frame, quantized with a predictive, split-VQ scheme and transmitted using 18 bits. Two adaptive-codebook indices are quantized, one per each subframe; the second index is differentially quantized with respect to the first, for a total of 8 and 5 bits, respectively. The fixed-codebook pulse positions and signs are represented with 13 and 4 bits per subframe, respectively. Finally, the gains for both the adaptive and fixed-codebook contributions are vector quantized with 7 bits per subframe, using a two-stage conjugate structure codebook, for an overall total of 80 bits. Table I shows the bit allocation for G.729.

B. Partial Encryption of G.729 Speech

For G.729, several UEP schemes have been proposed (see, e.g., [14] and [15]). The minimum set of class 1 bits reported in [15], i.e., the 36 bits (out of 80) absolutely requiring error protection, was taken as the reference to start to experiment with partial encryption of G.729 compressed speech. A tool developed in MATLAB was used in the selection process: using a graphical interface, shown in Fig. 3, specific subsets of the G.729 bitstream could be selected, on a parameter by parameter basis, encrypted and the result immediately played back for perceptual evaluation. Knowledge of the perceptual significance of each parameter guided the process: spectral envelope and its essential role in intelligibility; gain and its role in discrimination of voiced and unvoiced speech and between speech and silence; pitch period and gender identification.

After systematic informal listening tests, two sets of bits were finally selected: a larger, high-protection set, aimed at offering very strong protection against eavesdropping; and a smaller, medium-protection set, intended to eliminate intelligibility, but perhaps leaving in the bitstream some lesser clues about the speech stream.

The high-protection set (or HIGH scheme) is shown in Fig. 4(a). It covers 36 bits out of 80, 45% of the overall bitrate and coincides with the minimum set of class 1 proposed in [15]. It includes vector quantization indices L1 and L2, line spectral frequencies, the first seven most significant bits (MSB's) of the absolutely quantized pitch period (first subframe) and the first three MSB's of the differentially quantized pitch period (second subframe). The four gain indices are also integrally protected. Three essential features of speech, spectral envelope, pitch contour and gain contour, are, therefore, severely, if not completely, degraded.

The low-protection set (or LOW scheme) is shown in Fig. 4(b). It covers 24 bits out of 80 and it protects the same parameters protected by the HIGH scheme. Protection, however, is weaker for all parameters, except P2 (differentially encoded pitch). The two least significant bits (LSB's) of L1, L2, GB1 and GB2 are now unprotected, together with the LSB of GA1 and GA2. As for the HIGH scheme, the LSF index L3 and pulse signs and locations are completely unprotected.

It is straightforward to adapt the proposed partial encryption schemes to other ACELP speech coders, such as the ETSI GSM EFR and AMR standards, or the TIA IS-641 EFR. The parameters of those coders are, in fact, very similar and the underlying algorithm is essentially the same.

Cryptoanalysis of partially encrypted bitstreams proved to be troublesome. The unprotected bits, in fact, are in general very

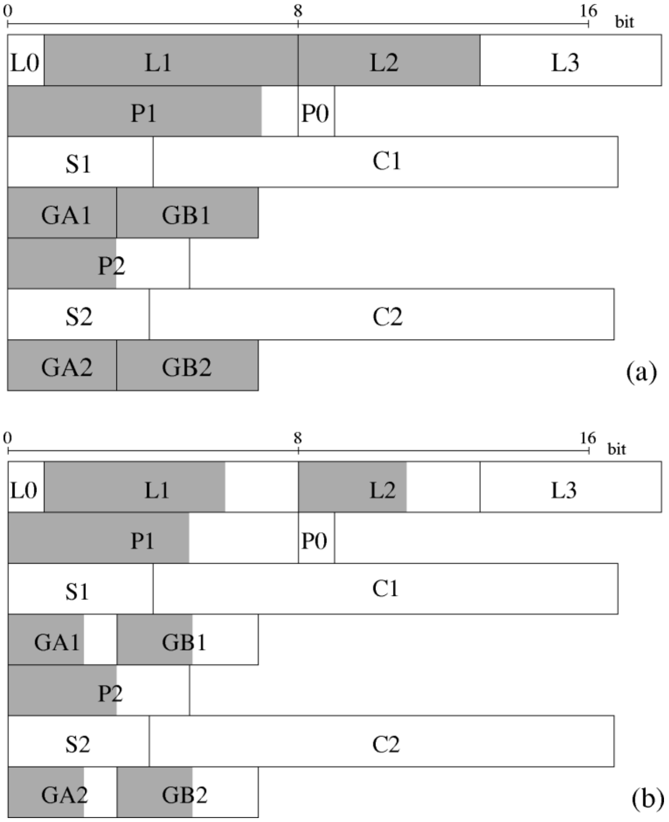


Fig. 4. Partial encryption schemes for G.729: HIGH (a) and LOW (b) bit allocations are shown, MSB to LSB moving from left to right. Bits subject to encryption are shown in gray.

weakly correlated with the encrypted bits. The correlation is low not only between parameters, such as, e.g., the unprotected fixed-codebook pulse locations (C1 and C2 in Fig. 4) and the encrypted line spectral frequencies (L1, L2); it is also negligible between encrypted and unencrypted bits belonging to the same parameter, especially when vector quantization is employed. It is, however, not unconceivable that a determined attacker could try to identify a speaker, particularly for a small number of potential speakers, by means of a Gaussian mixture model using only the unencrypted bits. Although this is unlikely to be an easy problem, the approach could be the subject of further study.

The security of the proposed technique depends on the security of the encryption layer that cyphers the protected bits. For the remainder the this work, we will assume perfect encryption, that is, all protected bits are unavailable for decoding. Even more effective security could be possible if the identity of the encrypted bits in each frame is, instead of being pre-determined, selected by a pseudo-random sequence, resulting in double encryption. The total number of bits encrypted would be the same, resulting in approximately the same power usage, but obtaining partial information based on the unencrypted bits would be even more difficult.

C. Performance Evaluation

Performance of the proposed partial encryption schemes was evaluated:

1) by signal inspection, in both the time and the frequency domains;
2) by means of objective distortion measures;
3) by means of formal listening tests.

The first approach consisted in analyzing specific features of speech signals subjected to partial encryption. Both spectral and temporal features, in fact, carry well-studied perceptual significance: the spectral envelope and in particular formant frequencies and bandwidths, is closely related to phoneme identification. Quasiperiodicity of the time-domain signal, or, equivalently, harmonic structure of the signal spectrum, are connected to voicing and, through the absolute value of the fundamental frequency, to gender identification.

The second evaluation approach was based on objective quality measures. In particular, segmental signal-to-noise ratio (segSNR) and Spectral Distortion (SD) were employed to evaluate the effectiveness of competing partial encryption schemes.

The third approach consisted of formal listening tests concerning the absolute performance of the proposed schemes. Since standard quality-oriented listening tests, such as Mean Opinion Scores (MOS) or A-B comparison tests, would not provide direct information about intelligibility or speaker/gender identification, new subjective tests needed to be designed. The attention was focused on the following tasks:

1) intelligibility;
2) gender identification;
3) plain-text identification;
4) speech/non speech discrimination.

Both high and low-encryption schemes were designed to prevent comprehension. Listeners were therefore asked to listen to a sentence, in their own mother language, but unknown to them and then give an intelligibility score based on a 5-point scale: “5–Full,” “4–Good,” “3–Fair,” “2–Poor,” “1–Nothing.” If the score was greater than one, that is, if the listener deemed to have understood one or more words, he/she was asked to write them down and then to listen to the clear-text sentence, to verify his/her judgment. In case of match between perceived and original words, the score was confirmed, otherwise it was downgraded to “1–Nothing.” Matches were rather loosely evaluated: acoustical similarity between estimates and original words was enough to confirm the first score. Listeners were free to listen to the sentences as many time as desired, thus approaching the condition of an offline analyst rather than a real-time eavesdropper.

The second experiment focused on gender identification. Listeners were asked to listen to encrypted sentences and to guess the speaker gender as “male,” “female,” or “don’t know.” Again, they were free to repeat their listening experience at will.

The third experiment concerned the case of speech communications based on a limited vocabulary, e.g., commands, or prompts of an automated response system. In such cases, the eavesdropper is assumed to know the vocabulary and the objective of the protection scheme is to obtain rates of successful identification comparable to random choice. In our case, listeners heard four clear-text sentences. Successively, they heard one of the four sentences, randomly chosen, encrypted and they were asked to match it with one of the four original sentences, or to vote “don’t know.” Repeated listenings, in any order, were allowed.

Finally, the fourth experiment evaluated the ability to discriminate between speech and nonspeech encrypted signals.

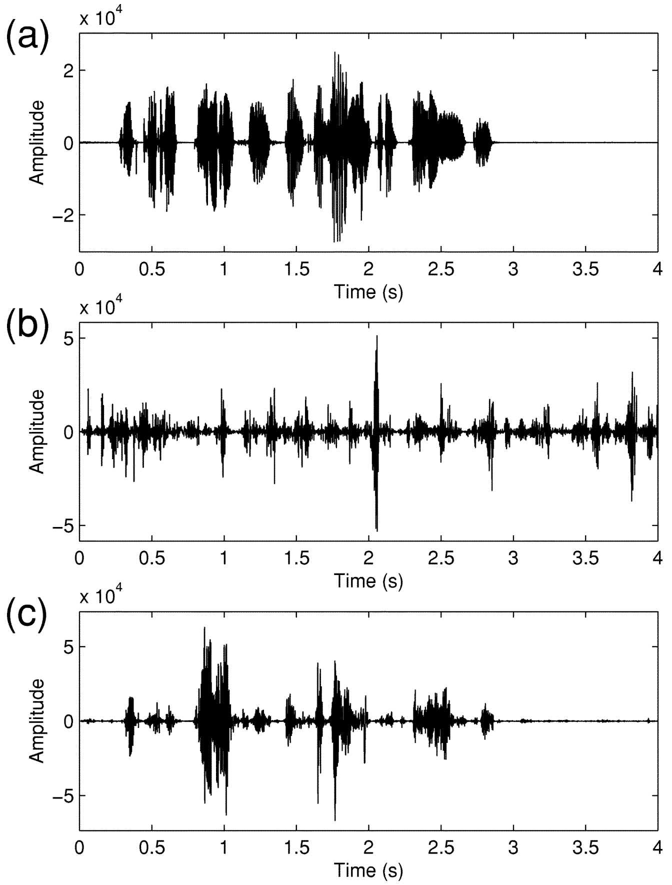


Fig. 5. Partial encryption of a sentence (Italian text "Il carrettiere frusta il cavallo troppo lento"): the original signal is shown in (a). Partially encrypted signals using the (b) HIGH and (c) LOW encryption schemes for G.729 are shown.

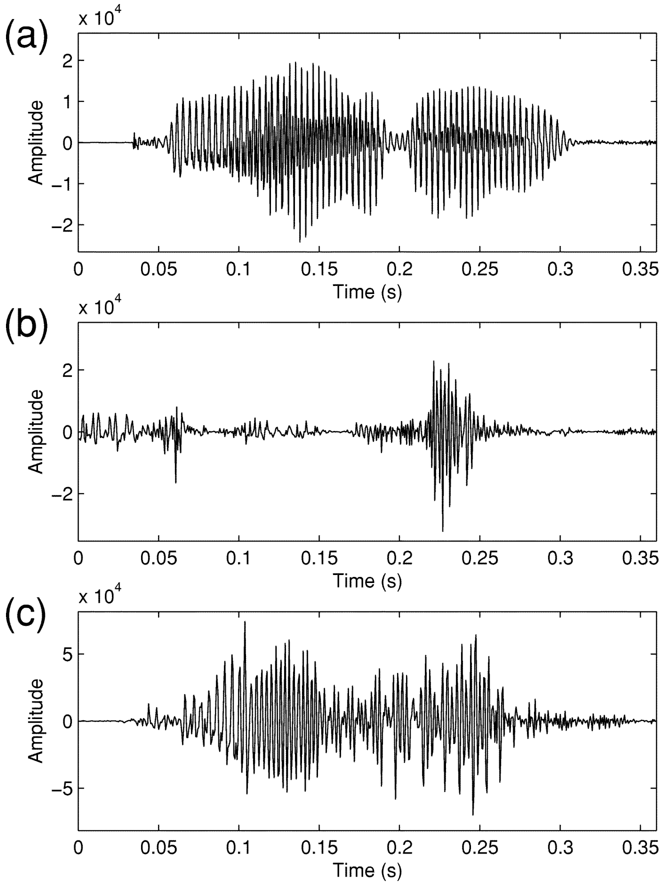


Fig. 6. Partial encryption of speech signal corresponding to the word fragment /tiere/ (final part of Italian word "carrettiere"): the original signal is shown in (a). Partially encrypted signals using the (b) HIGH and (c) LOW encryption schemes for G.729 are shown.

This ability could be of interest to an eavesdropper scanning for speech vs. silence in a database of speech recordings. In this case, listeners were asked to classify the signals as "speech," "nonspeech" or "don't know."

IV. Results and Discussion

Both the HIGH and LOW partial-encryption schemes were applied to flat filtered clean speech taken from the NTT Multilingual Speech Database. The material was encoded using the ITU-T G.729 floating-point reference software. The resulting bitstreams were partitioned according to the schemes shown in Fig. 4, with the encrypted bits assumed to be perfectly protected and thus replaced with a random sequence of binary digits. Finally, standard-compliant G.729 decoding generated the output material.

A. Signal Inspection

Fig. 5 shows an example sentence. Comparison of the original signal, shown in Fig. 5(a), to the signal subject to HIGH partial encryption, shown in Fig. 5(b), indicates a very high degree of content destruction. Analysis of the partially encrypted signal alone does not even permit to discriminate between speech and silence. Informal listening confirms that the signal sounds noise-like, with no hints of perceptual content.

The effects of the LOW partial encryption scheme are shown in Fig. 5(c). In this case, the energy contour of the sentence is fairly well preserved. Inter-word pauses, as well as silences before and after the sentence, are easily identified. Intelligibility is still practically zero, but the perception that the noise-like signal might be speech, albeit highly distorted, is possible.

Fig. 6 shows a fragment of the previous sentence, more specifically the final part, /tiere/, of the Italian word "carrettiere." The signal is characterized by a plosive sound, /t/, followed by strongly voiced phonemes, /iere/. The HIGH encryption scheme erases both the time-domain structure of the plosive and the periodicity of the voiced part, as shown in Fig. 6(b). Significantly more is apparently retained, at least in the time-domain, after LOW partial encryption, as can be seen in Fig. 6(c).

Since most of the analysis performed by the hearing system is in the frequency domain, the spectrum of the penultimate vowel of the same word fragment, /tiere/, has been analyzed. Fig. 7(a) shows the magnitude spectrum of the original vowel. A clear harmonic structure with fundamental frequency of approximately 250 Hz characterizes most of the spectrum. A spectral envelope with at least two formants is distinctly discernible. Both the HIGH and the LOW encryption schemes completely eliminate the harmonic structure, as shown in Fig. 7(b) and (c). The HIGH encryption scheme leaves almost no detectable spectral envelope, confirming the total absence of perceptual content suggested by informal listening.

TABLE II
OBJECTIVE DISTORTION MEASURES OF PARTIALLY ENCRYPTED SPEECH WITH RESPECT TO ORIGINAL SPEECH; COMPUTED OVER 192 4-SECOND SENTENCES (50% MALE, 50% FEMALE). SEGMENTAL SNR VALUES FOR 20-MS SEGMENTS

	FULL encryption	HIGH encryption	LOW encryption
Segmental SNR	-18.74 dB	-18.41 dB	-9.91 dB
Spectral Distortion	8.14 dB	7.60 dB	6.23 dB

TABLE III
FORMAL LISTENING TEST RESULTS FOR EXPERIMENTS 1 THROUGH 4

Experiment	HIGH encryption			LOW encryption		
1. Intelligibility (1 to 5)	1.00			1.11		
	No vote	True	False	No vote	True	False
2. Gender Identification	35.6 %	41.4 %	23.0 %	22.1 %	60.6 %	17.3 %
3. Plain-Text Identification	30.8 %	17.9 %	51.3 %	18.0 %	41.0 %	41.0 %
4. Speech/Non-Speech Discrimination	14.4 %	44.2 %	41.4 %	10.6 %	60.6 %	28.8 %

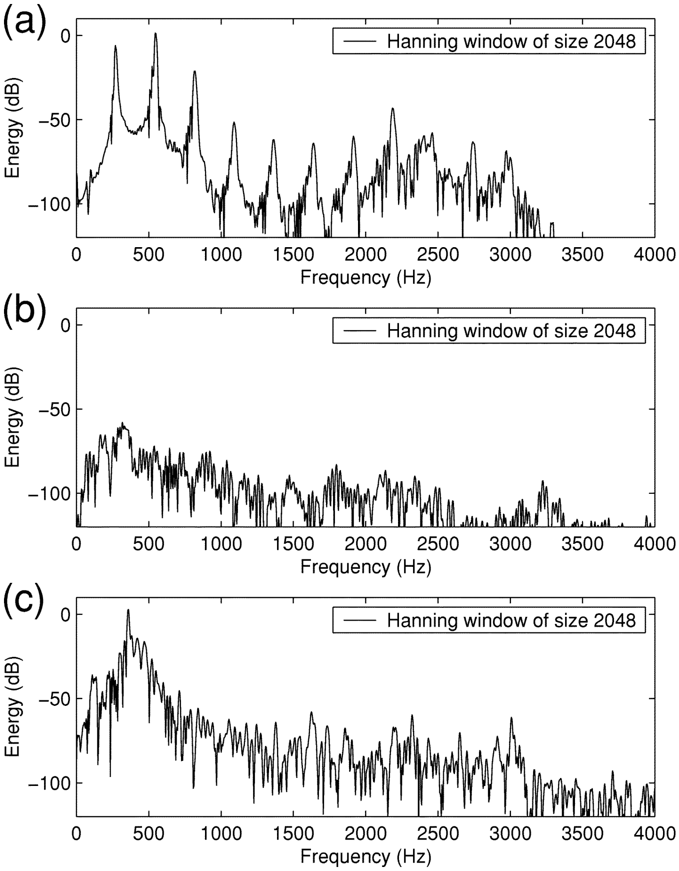


Fig. 7. Partial encryption of speech signal corresponding to a vowel (the penultimate /e/ sound of the Italian word "carrettiere"): magnitude spectrum of the original signal is shown in (a). Magnitude spectra for partially encrypted G.729 signals using the (b) HIGH and (c) LOW encryption schemes are shown.

B. Objective Performance Measures

Segmental signal-to-noise ratio (segSNR) and spectral distortion (SD) were used to objectively assess the performance of the partial encryption schemes under investigation. Speech material consisting of 192 sentences, each four seconds long, spoken by both male and female speakers, were encoded using the ITU-T G.729 reference software and then partially encrypted with the proposed HIGH and LOW schemes. Although absolute segSNR and SD values are not descriptive per se, they are quite useful to compare the performance of different partial encryption schemes and to supplement other performance evaluation data.

Segmental SNR was used to compare partially encrypted signals to corresponding clear-text sentences. The segment size was 20 ms, no overlapping, no threshold. Table II shows the segSNR values for both HIGH and LOW partial encryption schemes. The performance gap between the schemes is almost 9 dB. HIGH encryption is only 0.3 dB below the performance of full encryption.

Table II also reports Spectral Distortion (SD) values for HIGH, LOW and full encryption schemes. As in the case of segSNR, the performance of HIGH encryption is very close to that of full encryption (-0.5 dB) and significantly higher than the performance of LOW scheme ($+1.4$ dB).

In the following Section, we will see how the insight gained by signal inspection and objective performance measures correlate with the responses of listeners in formal listening tests.

C. Formal Listening Tests

The test material was presented to 13 different listeners, all using headphones in a controlled environment. The four experiments described in Section III-C were carried out to assess the subjective performance of HIGH and LOW partial encryption. For the first experiment (intelligibility), sixteen sentences (eight for HIGH and eight for LOW encryption), each four seconds long, were used, as for the second (gender identification) and fourth (speech/nonspeech discrimination) experiments. For the third experiment (plain-text identification), six sets (three for HIGH, three for LOW) of four sentences each were used.

Table III shows the overall results. For HIGH encryption, the intelligibility average score was precisely 1.0, corresponding to "Nothing" in the five point scale described in a previous section: nothing intelligible was ever discerned in the 104 stimuli presented to the listeners. Moreover, no score downgrading ever took place, i.e., no listener ever had the perception of understanding a word. For LOW encryption, the intelligibility average score was 1.11. In a few cases, in fact, the listeners indicated to have understood something and, by a rather loose verification criterion, their perception was close enough to the original unencrypted signal to be classified as successful.

The success rate for gender identification was certainly statistically better than random guesses for LOW partial encryption. For HIGH encryption, too, identification rate is statistically better than random (95% confidence interval). It has to be said, however, that this experiment was run together with the previous experiment, to shorten the test and to minimize listening fatigue. In that experiment, whenever the intelligibility vote was greater than "1," the listener would hear the original signal to verify whether his/her guess was right. In doing so, knowledge of the speaker's gender would be gained and matched to his/her prosody and speaking rate, leading to, according to some listeners, better gender identification skills.

For HIGH encryption, plain-text identification success rate is statistically equivalent to random guesses; a full one-third of the responses were "don't know." The performance is worse for LOW encryption, where the likelihood of identifying the correct sentence is about 50% (compared to 25% for random guesses). "Don't know" votes are down to 20% of the total votes.

Speech/nonspeech discrimination is again statistically equivalent to random guesses for the case of HIGH partial encryption. LOW encryption, instead, leads to 66% success rate, as opposed to 50% for random guesses.

V. Conclusions

Perception-based, partial encryption schemes for telephone bandwidth speech were presented. Speech compressed by a widely-used speech coding algorithm, ITU-T G.729 at 8 kb/s, was partitioned in two classes, one, the most perceptually relevant, to be encrypted, the other, to be left unprotected. Two partial-encryption techniques were developed, a low-protection scheme, aimed at preventing most kinds of eavesdropping and a high-protection scheme, based on the encryption of a larger share of perceptually important bits and meant to perform as well as full encryption of the compressed bitstream.

The high-protection scheme, based on the encryption of about 45% of the bitstream, achieves content protection comparable to that obtained by full encryption of the bitstream, as verified by both objective measures and formal listening tests. For the low-protection scheme, encryption of as little as 30% of the bitstream virtually eliminates intelligibility as well as most of the remaining information.

Low-power, portable devices could therefore achieve very high levels of speech-content protection at only 30–45% of the computational load of current techniques, freeing resources for other tasks and enabling longer battery life.

Acknowledgment

The authors would like to acknowledge Prof. M. Elia of the Politecnico di Torino, as well as the anonymous reviewers, for their insightful comments and suggestions.

References

[1] J. Goodman and A. P. Chandrakasan, "Low power scalable encryption for wireless systems," *Wireless Networks*, no. 4, pp. 55–70, 1998.

[2] G. A. Spanos and T. B. Maples, "Security for real-time MPEG compressed video in distributed multimedia applications," in *Proc. Conf. Computers Communications*, Mar. 1996, pp. 72–78.

[3] C. Shi and B. K. Bhargava, "An efficient MPEG video encryption algorithm," in *Proc. Symp. Reliable Distributed Systems*, Oct. 1998, pp. 381–386.

[4] A. M. Alattar and G. I. Al-Regib, "Evaluation of selective encryption techniques for secure transmission of MPEG-compressed bit-streams," in *Proc. Symp. Circuits Systems*, vol. 4, June 1999, pp. 340–343.

[5] H. C. H. Cheng, "Partial Encryption for Image and Video Communication," M.S. thesis, Univ. Alberta, Edmonton, AB, Canada, 1998.

[6] H. Cheng and X. Li, "Partial encryption of compressed images and videos," *IEEE Trans. Signal Processing*, vol. 48, pp. 2439–2451, Aug. 2000.

[7] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier, 1995, ch. 3.

[8] C. Laflamme, J.-P. Adoul, H. Y. Su, and S. Morissette, "On reducing computational complexity of codebook search in CELP coder through the use of algebraic codes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1990, pp. 177–180.

[9] C. Laflamme, J.-P. Adoul, R. Salami, S. Morissette, and P. Mabilleau, "16 kpbs wideband speech coding technique based on algebraic CELP," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1991, Paper S1.4, pp. 13–16.

[10] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul, "GSM enhanced full rate speech codec," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1997, pp. 771–774.

[11] E. Ekudden, R. Hagen, I. Johansson, and J. Svedberg, "The adaptive multi-rate speech coder," in *Proc. IEEE Workshop Speech Coding for Telecommunications*, 1999, pp. 117–119.

[12] T. Honkanen, J. Vainio, K. Järvinen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul, "Enhanced full rate speech codec for IS-136 digital cellular system," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1997, pp. 731–734.

[13] R. Salami *et al.*, "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 116–130, Mar. 1998.

[14] P. Kroon and Y. Shoham, "Performance of the proposed ITU-T 8 kb/s speech coding standard for a Rayleigh fading channel," in *Proc. IEEE Workshop on Speech Coding for Telecommunications*, Annapolis, MD, Sept. 1995, pp. 11–12.

[15] K. Swaminathan, A. R. Hammons Jr., and M. Austin, "Selective error protection of ITU-T G.729 CODEC for digital cellular channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1996, pp. 577–580.

Antonio Servetti (S'01) was born in Cuneo, Italy, in 1975. He received the Laurea degree in computer engineering from the Politecnico di Torino, Italy, in 1999. He is presently a Ph.D. student with the Dipartimento di Automatica e Informatica of the Politecnico di Torino.

In 2000 he was with CSELT (Centro Studi e Laboratori Telecomunicazioni) in Turin. His primary research interests include speech and audio signal processing, coding and trasmission over IP and wireless networks. He collaborates with the Center for Wireless Multimedia Communications (CERCOM) of the Politecnico di Torino.

Mr. Servetti is a student member of the IEEE Signal Processing Society.

Juan Carlos De Martin (S'91–M'97) graduated in electronic and computer engineering from the Politecnico di Torino, Italy, in 1991. From the same university, he also received the Ph.D. degree in electronic and telecommunications engineering in 1996.

Between 1993 and 1995, he was a Visiting Researcher at the Signal Compression Laboratory of the University of California, Santa Barbara, where he worked on low bit rate speech coding based on interpolative techniques. From 1996 until 1998 he was Member of Technical Staff at the Media Technologies Laboratory, Texas Instruments Inc., Dallas, where he carried out research in the areas of speech coding, speech transmission over noisy channels, and voice over IP networks. In 1998, he joined the National Research Council of Italy (CNR) at the Politecnico di Torino, where he is currently a Principal Researcher and an Adjunct Associate Professor. His research interests include multimedia transmission over packet networks, data, speech, audio and video compression, wireless communications and applications.

Dr. De Martin is a member of the IEEE Signal Processing and IEEE Communications Societies.