

# TEAGER ENERGY BASED FEATURE PARAMETERS FOR SPEECH RECOGNITION IN CAR NOISE

*Firas Jabloun*

Dept of Electrical and Electronics Eng.  
Bilkent University

*A. Enis Çetin*

Dept of Electrical and Electronics Eng.  
Bilkent University, Ankara, Turkey

## ABSTRACT

In this paper, a new set of speech feature parameters based on multirate signal processing and the Teager energy operator is introduced. The speech signal is first divided into nonuniform subbands in mel-scale using a multirate filterbank, then the Teager energies of the subsignals are estimated. Finally, the feature vector is constructed by log-compression and inverse DCT computation. The new feature parameters have robust speech recognition performance in the presence of car engine noise.

## 1. INTRODUCTION

In this paper, a new set of speech feature parameters is proposed. The new parameters are developed using multirate signal processing and the Teager Energy Operator (TEO) which has been successfully used in various speech processing applications [1]-[5]. It is experimentally observed that the TEO can suppress the car engine noise which makes the new feature parameters a good candidate for voice dialing systems in automobiles.

In continuous-time, TEO is defined as

$$\Psi_c[s(t)] = [\dot{s}(t)]^2 - s(t)\ddot{s}(t). \quad (1)$$

where  $s(t)$  is a continuous-time signal and  $\dot{s} = \frac{ds}{dt}$ . In discrete-time TEO can be approximated by

$$\Psi_d[s(n)] = s(n)^2 - s(n+1)s(n-1). \quad (2)$$

In this paper, the discrete-time version is used, and the subscript 'd' is dropped from now on. Let  $s(n)$  be a discrete-time wide-sense stationary random signal. In this case

$$E\{\Psi[s(n)]\} = E\{s^2(n)\} - E\{s(n+1)s(n-1)\} \quad (3)$$

or

$$E\{\Psi[s(n)]\} = R_s(0) - R_s(2). \quad (4)$$

where  $R_s(k)$  is the autocorrelation function of  $s(n)$ .

In general, the car engine noise,  $v(n)$ , is mostly low pass in nature. A typical example is shown in Figure 1. For

this noise signal, the relation between the first three autocorrelation lags are estimated as  $R_v(1) = 0.9997R_v(0)$  and  $R_v(2) = 0.9991R_v(0)$ . Since  $R_v(0) \approx R_v(1) \approx R_v(2)$  we have  $E\{\Psi[v(n)]\} \approx 0$ . Due to this reason, the spectrum of  $\Psi[v(n)]$  shown in Figure 2 is almost negligible compared to the spectrum of the noise  $v(n)$ .

For a typical speech signal,  $s(n)$ , the first three autocorrelation lags are not as close to each other. For example,  $R_s(1) = 0.7415 R_s(0)$  and  $R_s(2) = 0.4584 R_s(0)$  for the first author's /a/,  $R_s(1) = 0.97 R_s(0)$ ,  $R_s(2) = 0.91 R_s(0)$  for the second author's /f/,  $R_s(1) = 0.84 R_s(0)$ , and  $R_s(2) = 0.73 R_s(0)$ , for the second author's /s/.

In practice, the observed signal is the sum of the speech signal and the noise. Let the observed signal be  $x(n) = s(n) + v(n)$  where  $s(n)$  is the noise free speech signal and  $v(n)$  is a zero mean additive noise which is independent from  $s(n)$ . The Teager energy of the noisy speech signal  $x(n)$  is given by

$$\Psi[x(n)] = \Psi[s(n)] + \Psi[v(n)] + 2\tilde{\Psi}[s(n), v(n)] \quad (5)$$

where  $\tilde{\Psi}[s(n), v(n)] = s(n)v(n) - \frac{1}{2}s(n-1)v(n+1) - \frac{1}{2}s(n+1)v(n-1)$ , is the cross- $\Psi$  energy of  $s(n)$  and  $v(n)$ . Since  $s(n)$  and  $v(n)$  are zero mean and independent the expected value of their cross- $\Psi$  energy is zero. Thus,  $E\{\Psi[x(n)]\} = E\{\Psi[s(n)]\} + E\{\Psi[v(n)]\}$ . Furthermore,  $E\{\Psi[v(n)]\}$  is negligible compared to  $E\{\Psi[s(n)]\}$  for the car engine noise, i.e.,

$$E\{\Psi[x(n)]\} \approx E\{\Psi[s(n)]\} \quad (6)$$

Hence, the effect of car engine noise can be eliminated by using TEO in feature extraction. On the other hand, the commonly used energy has no filtering capability because

$$E\{x^2(n)\} = R_s(0) + R_v(0). \quad (7)$$

Due to this reason, we expect a TEO based feature set to produce better recognition rates than the regular energy based features in car engine noise.

In Section II, new TEOCEP feature parameters are formally defined. The speech signal is first divided into

nonuniform subbands in mel-scale using a multirate filter bank. Then the Teager energies are estimated in each subband and the feature vector is constructed by log-compression and inverse DCT computation. In Section III, the new parameters are used in isolated word recognition under car engine noise and it is experimentally observed that the TEOCEP parameters produce better recognition performance than MELCEP's [6] and SUBCEP's [7].

## 2. THE TEOCEP FEATURE PARAMETERS

In our method, multirate subband decomposition [7, 8, 9] is used in a tree structure to divide the speech signal  $x(n)$  according to the mel-scale as shown in Fig. (3), and 21 sub-signals  $x_l(n)$ ,  $l = 1, \dots, L = 21$ , are obtained. The filter bank corresponding to a biorthogonal wavelet transform is used in the analysis [10]. The lowpass  $H_0(z)$  and highpass  $H_1(z)$  filters have the transfer functions:

$$H_{0,1}(z) = \frac{1}{2} \pm \frac{9}{32}(z^{-1} + z^1) \mp \frac{1}{32}(z^{-3} + z^3), \quad (8)$$

respectively. For every sub-signal, the average Teager energy  $e_l$

$$e_l = \frac{1}{N_l} \sum_{n=1}^{N_l} |\Psi[x_l(n)]|; l = 1, \dots, L. \quad (9)$$

is estimated. In (9),  $N_l$  is the number of samples in the  $l^{th}$  band, and  $\Psi[\cdot]$  is the discrete-time Teager energy operator.

Although it is possible that the instantaneous Teager energy have negative values in very rare circumstances, the average value  $e_l$  is a positive quantity for most natural signals [4, 11] as  $R_x(0)$  is usually larger than  $R_x(2)$ . Nonetheless, the magnitude of the Teager energy is used to ensure the non-negativity of  $e_l$ .

At the last step, log compression and inverse DCT computation is applied to obtain the TEO-based cepstrum coefficients

$$TC(k) = \sum_{l=1}^L \log(e_l) \cos\left[\frac{k(l-0.5)\pi}{L}\right]; k = 1, \dots, N. \quad (10)$$

We call the new feature set TEOCEP parameters. The first 12  $TC(k)$  coefficients are used in the feature vector. Twelve more coefficients obtained from the first-order differentials are also appended. A final feature vector with dimension 24 is obtained and is used for training and recognition.

The SUBCEP parameters used in [7] differ from the TEOCEP parameters in the definition of the energy measure used in Equation (9). In [7],  $\ell_1$  energy

$$\varepsilon_l = \frac{1}{N_l} \sum_{n=1}^{N_l} |x_l(n)|; l = 1, \dots, L \quad (11)$$

is used instead of  $e_l$ .

It is shown that the SUBCEP parameters perform slightly better than the well-known MELCEP parameters [7, 9, 8]. For this reason, the performance of the TEOCEP parameters are evaluated with respect to that of SUBCEP parameters.

## 3. SIMULATION RESULTS

A continuous density Hidden Markov Model based speech recognition system with 5 states and 3 Gaussian mixture densities is used in simulation studies. The recognition performances of the TEOCEP feature parameters are evaluated using the *TI-20* speech database of *TI-46 Speaker Dependent Isolated Word Corpus* which is corrupted by various types of additive noise. The *TI-20* vocabulary consists of ten English digits and ten control words. The data is collected from 8 male and 8 female speakers. There are 26 utterances of each word from each speaker of which are 10 are designated as training tokens and 16 designated as testing tokens.

Speaker dependent isolated word speech recognition simulations are presented in Tables 1, 2 and Table 3. In the first two tables car noise is added on the speech signal and in Table 3 the speech signal is corrupted by additive white noise. The first car noise is recorded inside a Volvo 340 on a rainy asphalt road by the *Institute for Perception-TNO, Holland*. The spectrum of this noise signal is shown in Figure 1. The second set of results in Table 2 is obtained for the noise recorded inside a Mazda 626 on an asphalt road traveling at 90 km/hour (55 miles/hour).

The same filterbank is used to generate the SUBCEP and TEOCEP parameters. The frame size is chosen as 48 msec with an overlap of 32 mseconds. In the car noise case, the superiority of the TEOCEP parameters over the SUBCEP parameters is obvious especially at low SNR values. However, in white noise, just a slight improvement is achieved at low SNR values. This is expected because for white noise  $v(n)$ , the autocorrelation function  $R_v(k) = 0$  for  $k \neq 0$ , and the TEO does not perform any filtering.

In Table 4, speaker independent experiment results with the Volvo car noise are shown. The utterances of five men and five women were used for training. The utterances of the rest of the speakers are used to test the performance of the system. Again the TEOCEP parameters outperform the SUBCEP parameters especially at low SNR's.

#### 4. REFERENCES

- [1] H. M. Teager, "Some observations on oral air flow during phonation," *IEEE Trans. on Speech and Audio Proc.*, Oct. 1980.
- [2] H. M. Teager and S. M. Teager, "Evidence for nonlinear speech production mechanisms in the vocal tract," *NATO Advanced Study Institute on Speech Production and Speech Modelling, Bonas, France*, July 1989.
- [3] A. C. Bovik, P. Maragos, and T. Quatieri, "AM-FM energy detection and separation in noise using multi-band energy operators," *IEEE Trans. on Signal Proc.*, vol. 41, pp. 3245–3265, December 1993.
- [4] P. Maragos, J. F. Kaiser, and T. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. on Signal Proc.*, vol. 41, pp. 3025–3051, Oct. 1993.
- [5] P. Maragos, T. Quatieri, and J. F. Kaiser, "On amplitude and frequency demodulation using energy operators," *IEEE Trans. on Signal Proc.*, vol. 41, pp. 1532–1550, April 1993.
- [6] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. on Acoust., Speech, and Signal Proc.*, vol. 28, pp. 357–366, August 1980.
- [7] E. Erzin, A. Çetin, and Y. Yardımcı, "Subband analysis for robust speech recognition in the presence of car noise," *Proc. of the Int. Conf. on Acoust., Speech and Signal Proc. 1995 (ICASSP '95)*, May 1995.
- [8] R. Sarikaya, B. L. Pellom, and J. H. Hansen, "Wavelet Packet Transform Features with Application to Speaker Identification," *NORSIG'98*, pp. 81–84, 1998.
- [9] R. Sarikaya and J. N. Gowdy, "Subband Based Classification of Speech Under Stress," *Proc. of the Int. Conf. on Acoust., Speech and Signal Proc. 1998 (ICASSP '98)*, vol. 1, pp. 596–572, 1998.
- [10] C. W. Kim, R. Ansari, and A. E. Çetin, "A class of linear-phase regular biorthogonal wavelets," *Proc. of the Int. Conf. on Acoust., Speech and Signal Proc. 1992 (ICASSP '92)*, vol. IV, pp. 673–677, 1992.
- [11] A. C. Bovik and P. Maragos, "Conditions for positivity of an energy operator," *IEEE Trans. on Signal Proc.*, Feb 1994.

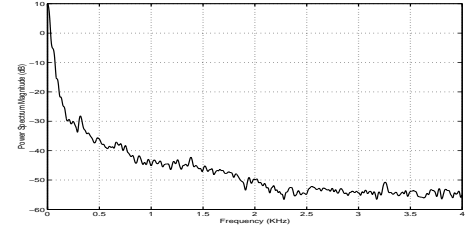


Figure 1: Power Spectrum Density of the car noise signal recorded inside a Volvo 340 on a rainy asphalt road by the *Institute for Perception-TNO, The Netherlands*

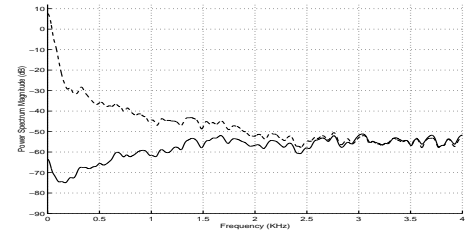


Figure 2: Spectrum of the car noise  $v(n)$  (dashed line) and the spectrum of the Teager energy  $\Psi[v(n)]$  (continuous line)



Figure 3: The sub-band frequency decomposition of the speech signal

SNR (dB)	TEOCEP	SUBCEP
30	99.66	99.15
10	99.26	99.05
7	99.37	97.98
5	99.05	97.02
3	98.84	96.41
0	98.17	95.14
-3	97.83	93.12
-5	96.86	90.62

Table 1: Table 1: The average recognition rates of speaker dependent isolated word recognition system with SUBCEP and TEOCEP features for various SNR levels under additive Volvo 340 noise.

SNR (dB)	TEOCEP	SUBCEP
30	99.54	99.43
10	99.41	99.10
7	99.12	98.30
5	99.11	97.17
3	98.93	96.83
0	98.05	95.20
-3	97.81	93.05
-5	96.57	90.54

Table 2: Table 2: The average recognition rates of speaker dependent isolated word recognition system with SUBCEP and TEOCEP features for various SNR levels under additive Mazda 626 noise.

SNR (dB)	TEOCEP	SUBCEP
20	97.79	98.37
10	87.07	87.70
7	86.12	85.17
5	82.97	81.70
3	79.83	79.50

Table 3: Table 3: The average recognition rates of speaker dependent isolated word recognition system with SUBCEP and TEOCEP features for various SNR levels under additive white noise.

SNR (dB)	TEOCEP	SUBCEP
30	91.22	91.25
10	91.13	90.96
7	90.74	89.94
3	89.10	88.40
0	87.13	86.63
-3	85.26	80.17

Table 4: Table 4: The average recognition rates of speaker independent isolated word recognition system with SUBCEP and TEOCEP features for various SNR levels under additive Volvo 340 noise.