Detection of Multi-tone Signals Based on Energy Operators

Edgar F. Velez
Telecom Microelectronics Center, Northern Telecom Limited
P.O. Box 3511, Stn. C, Ottawa, Ontario, Canada, K1Y 4H7
velez@bnr.ca

Abstract

A nonlinear approach based on the Teager-Kaiser energy operator for multi-tone detection in the presence of voice is presented. The approach is shown to provide accurate frequency estimation, robustness against network distortions, good rejection of tone-like speech, with a significantly lower computational load when compared to classical tone detection approaches.

Introduction

The use of single- and multi-component signals is quite common in telecommunications, as a means to transmit additional information or signaling together with voice. Dual Tone Multi-Frequency (DTMF) is used extensively for telephone dialing, and also provides a user interface to request special services from the central office or to input data into automated answering systems. These signals are produced by a combination of two single-frequency tones, each from a set of four tones in a lower (697, 770, 852, 941 Hz) or upper (1209, 1336, 1477, 1633 Hz) frequency bands, representing the digits 0-9, *, #, and A-D reserved tones.

Standards for DTMF reception specify stringent requirements, such as validating the presence of only one frequency component in each of the bands, frequency tolerances, amplitude limits and permissible difference in component levels (twist), minimum duration of tones and cycle times, noise requirements, and rejection of false DTMF signals.

Classical DTMF detection methods use bandpass filter banks and envelope detectors to estimate the level of each of the eight possible frequency components, and select the frequencies with the highest levels as candidates for DTMF signals. Further processing is required to try to discriminate real tones from voice signals or other energy in the voiceband. The properties of the Teager-Kaiser algorithm [1, 2], also known as the energy operator, make it a natural choice for tone processing:

$$\Psi[x(n)] = x(n)^2 - x(n+1)x(n-1) \qquad (1)$$

which for a sinusoid $x(n) = A\cos(\Omega n + \phi)$, gives

$$\psi[x(n)] = A^2 \sin(\Omega)^2 \ . \qquad (2)$$

For detection, this expression provides the required frequency information if the amplitude remains constant. This is indeed the case for each of the components of a DTMF signal, whose amplitude is required to remain stable while the signal is being produced. Furthermore, the amplitudes of both DTMF frequency components are essentially equal at the point of generation. However, there is a large transmission media diversity in the telephone network, which is characterized by transfer functions with nonuniform amplitude response in the voiceband. Thus, the transmission channel produces different level attenuation for each DTMF component that must be tolerated at the receiver. The required frequency and amplitude information can be decoupled by using the Discrete Energy Separation Algorithm (DESA) approach [3, 6].

Detection using the Energy Operator

A DTMF signal has two frequency components, $x(n) = A_1\cos(\Omega_1 n + \phi_1) + A_2\cos(\Omega_2 n + \phi_2)$. The corresponding energy operator has 4 terms, of which 2 are related to the amplitude-frequency product squared of the individual components, and the other 2 are crossterms varying at a rate equal to the sum and difference of the component frequencies. By the use of simple filtering it is possible to eliminate the DC terms, which contain the amplitude and frequency information for both components but cannot be easily decoupled, and the frequency-sum term which is usually heavily attenuated, the frequency-difference term can be isolated:

$$c(n) = 2A_1A_2\sin\left[\frac{(\Omega_1+\Omega_2)}{2}\right]^2 \cos[(\Omega_1-\Omega_2)n+\phi] \ , \qquad (3)$$

where $\phi = \phi_1 - \phi_2$. The frequency of this crossterm, which corresponds to the difference of DTMF tone components, can be detected by a second energy operator. However, this detection method is affected by the reception level of the individual components, since the detected quantity is proportional to the product of the component amplitudes and the squared sine of the frequency difference:

$$\Psi[c(n)] = 4A_1^{\,2}A_2^{\,2}\sin\left[\frac{(\Omega_1+\Omega_2)}{2}\right]^4 \sin(\Omega_1-\Omega_2)^2 \quad (4)$$

An inherent advantage of this crossterm detection approach is that it treats the DTMF tone as a composite signal, without isolating its components. In order for the output of the second energy operator to remain constant during DTMF production, the amplitudes and frequencies of both components have to remain constant simultaneously, thus providing an automatic verification of this requirement. This property enhances the detector discrimination against false detection, as it is highly unlikely that speech can produce a dual tone without fluctuations of amplitude or frequency for any significant length of time.

The level dependence problem could easily be avoided by decoupling the crossterm's amplitude and frequency components using the DESA approach. On the other hand, the frequency sum term in Eq. (4) helps to discriminate between DTMF tones with similar frequency difference between components. Detection based on frequency differences only reduces the tone discrimination capability with respect to that obtained by detection of individual frequency components. The set of frequency differences for all 16 valid DTMF tones is (268, 357, 395, 439, 484, 512, 536, 566, 625, 639, 692, 707, 780, 781, 863, 936 Hz), where the minimum distance between difference frequencies corresponding to two different DTMF tones is only 1 Hz, which would result in the codes '3' and 'C' being easily confused. As a comparison, the minimum difference of individual DTMF frequency components is 73 Hz. Nevertheless, the crossterm method remains as a cost effective alternative for DTMF detection in applications where only a subset of DTMF tones is used or the level of the DTMF components is guaranteed to have a minimal variation from their nominal generation level.

Detection using the Discrete Energy Separation Algorithm (DESA)

DTMF detection can be performed reliably by separating the two frequency components in the low and upper frequency bands by the use of simple filters. Instead of using the energy operator on each component, which would produce level-dependent frequency estimates, the frequency and amplitude estimates for each component are obtained using the DESA approach, as shown in Fig. 1.

Both DESA-1 and DESA-2 [6] produce accurate estimates, although DESA-2 has a slight computational advantage [2]. Although DESA-2 has been employed for DTMF detection in the rest of this paper to minimize computations, it is important to outline the advantages of each of these techniques for frequency identification. Both methods require the computation of the energy operator on the signal and its derivative. The difference consists in the way the derivative is approximated in the discrete-time domain. Although both approximations are accurate and result in almost identical estimation results, the resulting discrete-time analytical expressions for the energy operator outputs are different and have useful properties.

The DESA-1 [3, 5, 6] approach approximates the signal derivative by the average of forward and backward 1-point differences, but instead of computing the difference signal and then apply the operator, which would result identical to DESA-2, it approximates the energy operator of the signal derivative by the average of the energy operators of forward and backward differences. For a sinusoid $x(n) = A\cos(\Omega n + \phi)$, this results in:

$$\frac{\Psi[x_n - x_{n-1}] + \Psi[x_{n+1} - x_n]}{2} = 4A^2\sin\left(\frac{\Omega}{2}\right)^2\sin(\Omega)^2 \quad (5)$$

where the notation $x_n = x(n)$ has been used. This approximation of the energy operator of the signal derivative is combined with the energy operator of the signal in Eq. (2) to produce the frequency and amplitude estimates:

$$\sin\left(\frac{\Omega}{2}\right)^2 = \frac{\left(\Psi\left[\frac{x_n - x_{n-1}}{2}\right] + \Psi\left[\frac{x_{n+1} - x_n}{2}\right]\right)/2}{\Psi[x_n]} \quad (6)$$

$$|A^2| = \frac{\Psi[x_n]}{1 - \left(1 - 2\sin\left(\frac{\Omega}{2}\right)^2\right)^2} \quad (7)$$

where Eqs. (6) and (7) are obtained from the corresponding expressions in [6] by the use of scaling properties of the energy operator [4] to ensure that the computed signals remain bounded in a fixed-point implementation. The DESA-1 estimates are based on only 5 sample points. It can be shown, by using trigonometric simplification, that the above expressions are *exact* for the case of a pure sinusoid.

An important property of DESA-1 frequency estimation using Eq. (6) is that the frequency of any signal up to $\Omega = \pi$ (half the sampling frequency) can be determined. This is due to the fact that the sine function and its inverse have a unique correspondence between 0 and $\frac{\pi}{2}$. Since the sine argument is $\frac{\Omega}{2}$, it follows that the frequency can be uniquely determined for any Ω between 0 and π.

The DESA-2 method requires the computation of the energy operator of the signal derivative, by approximating the derivative by the 2-point symmetric difference $x_{n+1} - x_{n-1}$. The DESA-2 frequency and amplitude estimates are obtained from:

$$\sin(\Omega)^2 = \frac{\Psi\left[\frac{x_{n+1} - x_{n-1}}{2}\right]}{\Psi[x_n]} \tag{8}$$

$$|A|^2 = \frac{\Psi[x_n]^2}{\Psi\left[\frac{x_{n+1} - x_{n-1}}{2}\right]} \tag{9}$$

Again, it is important to note that these expressions are *exact* for a pure sinusoid. The DESA-2 estimates are also based on 5 sample points. From Eq. (8) and in light of the previous discussion for DESA-1, it can be seen that the frequency of the signal can be uniquely determined only up to $\frac{\pi}{2}$ (one quarter of the sampling frequency) for DESA-2. The frequency Ω of a signal beyond $\frac{\pi}{2}$ will be identified as its mirror frequency with respect to $\frac{\pi}{2}$, i.e. as $\pi - \Omega$. This is also true for frequency estimation using the basic energy operator. However, it should not be considered that the energy operator is inherently limited to process signals with frequencies only up to one quarter of the sampling frequency; the use of a different discrete approximation which results in appropriate trigonometric expressions for frequency estimation can avoid this limitation, as is the case in the DESA-1 approach.

The nonlinear DTMF detection approach based on DESA-2 frequency/amplitude estimation is shown in Fig. 1. The bandpass filters separate the lower and upper frequency bands, increase the SNR by eliminating signals out of the DTMF frequency range, and provide sufficient adjacent band rejection to avoid crossterms with the other DTMF component. The energy operator output is almost always positive, but a lower threshold is applied to ensure positivity and avoid division by zero.

Computational Complexity

Traditional DTMF detectors use a bank of bandpass filters for the 8 component frequencies followed by envelope detectors, and identify a tone based on the largest outputs. Each bandpass filter is 2nd order or higher, and the envelope detectors are 1st order LPFs, requiring a minimum of 40 adds and 56 multiplies per sample.

The nonlinear approach reduces the computational load to almost a half by using only 2 detectors for the upper and lower bands. The DESA-2 approach requires only 3 adds, 5multiplies, and 1 divide per sample for frequency estimation. Band-split pre-filters are 4th order IIRs, while both smoothing LPFs are 1st order. Total requirements are 24 adds, 34 multiplies, and 1 divide per sample.

The task of DTMF detection is completed by a simple control scheme which checks the outputs of the frequency and amplitude estimators, and declares the valid detection of a DTMF tone when the amplitude and frequency specification requirements are met and maintained during a minimum length of time, the level twist doesn't extend beyond acceptable limits, and the tone duration is within the limits specified for on/off periods. Since all DTMF frequencies are below 2 KHz, processing can be performed at a 4 kHz sampling rate.

DTMF Detection

The performance of the proposed method for a series of short DTMF pulses is shown in Fig. 2. Each of the 16 DTMF tones tested has a duration of 40 ms followed by a 10 ms silence. The top graph shows two traces for the upper and lower-band instantaneous frequencies, while the lower graph shows two traces for the corresponding amplitudes and a third trace for the original signal offset by -1. The frequency estimates are accurate, remain constant during the tone, and respond immediately to silence or to a new tone. The amplitude estimates are also precise and independent of the DTMF frequencies. The last five pulses represent the same tone (D) with differing levels for each of its components: both at 0 dBm, both at -6 dBm, twist of 12 dB, reverse twist of -6 dB, and both at -25 dBm, respectively. As required, the frequency estimates are insensitive to level variations in the components.

Detection with Noise and Impostor Speech

Additive noise produces a distortion on the energy operator output proportional to the noise variance. This distortion extends to the DESA approach, as it is formed by the ratio of energy operators of the signal and its derivative. To minimize the noise effects on the estimates, the signal is bandpass filtered to include only the DTMF frequency components before DESA processing. An additional stage of soft low-pass filtering is performed on the frequency and amplitude estimates, based on the requirement that the DTMF signals be steady during their production. The design of the filters presents a trade-off between removal of noise, detection speed, impostor signal rejection, and computational complexity.

Speech has fundamentally different characteristics than tones due to complex mechanisms in its production. Both level and frequency of formants vary during a pitch period. In addition, voice within each band usually presents more than one component. The nonlinear frequency estimates of multi-component signals present oscillatory crossterms. These properties are exploited in the system to discriminate valid tones from speech by simply checking the flatness of the estimates. The results are illustrated in Fig. 3, which shows a telephone speech signal known to falsely trigger other detectors, together with DTMF tones with 20 dB SNR. The nonlinear approach easily discriminates between tone-like voice and real tones.

Figure 1 Nonlinear DTMF detection scheme

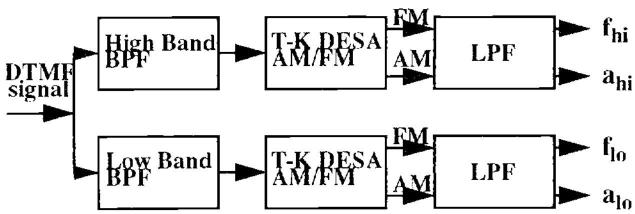


Figure 2 Detection of DTMF signals using DESA-2

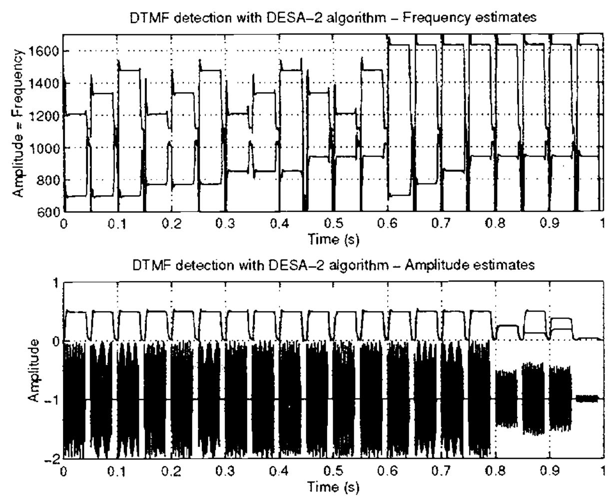
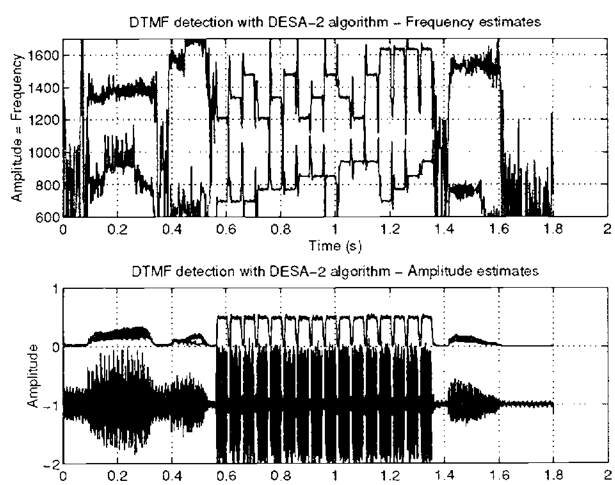


Figure 3 Detection of DTMF signals in noise (SNR=20dB) and impostor speech



Conclusions

A nonlinear method for the detection of DTMF signals in telephone communications based on estimation of instantaneous amplitudes and frequencies using the Teager-Kaiser energy operator is presented. Evaluation on real telephone signals demonstrates the advantages of nonlinear detection techniques over conventional approaches based on linear filtering. Frequency and amplitude estimates based on DESA-2 enable accurate detection of DTMF signals, with a minor amount of processing required to separate the signal components and to handle the variability of noise and level distortions present in the telephone network. Furthermore, the operator properties for sinusoidal signals provide excellent immunity to false detections due to speech or other voiceband energy which attempts to simulate DTMF signals, as well as to the existence of more than one component in each frequency band. The described approach satisfies the DTMF reception standards requirements, with its main advantages being its profound simplicity, a large reduction in computational complexity, faster detection performance, and robust discrimination against false DTMF signals.

Acknowledgments

The author would like to thank Dr. James Kaiser for several insightful discussions about the energy operator.

References

1) H. M. Teager and S. M. Teager, "Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract", in *Speech Production and Speech Modelling*, edited by William J. Hardcastle and Alain Marshal, NATO Advanced Study Institute Series D, vol. 55, Bonas, France, July 17-29, 1989, Kluwer Academic Publishers, Boston, MA, pp. 241-261, 1990.
2) J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal", in *Proc. IEEE ICASSP-90*, Albuquerque, New Mexico, pp. 381-384. April 1990.
3) P. Maragos, T. F. Quatieri, and J. F. Kaiser, "Speech Nonlinearities, Modulations, and Energy Operators", in *Proc. IEEE ICASSP 91*, Toronto, Canada, pp. 421-424, May 1991.
4) J. F. Kaiser, "Some Useful Properties of Teager's Energy Operators", in *Proc. IEEE ICASSP-93*, Minneapolis, Minnesota, pp. 149-152, April 1993.
5) J. T. Foote, D. J. Mashao, H. F. Silverman, "Stop Classification using DESA-1 High-resolution Formant Tracking", in *Proc. IEEE ICASSP-93*, Minneapolis, Minnesota, pp. 720-723, April 1993
6) P. Maragos, J. F Kaiser, and T. F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis", *IEEE Trans. Signal Processing*, vol. 41, no. 10, October 1993.