EXPERIMENTE MIT DEM TEAGER ENERGIE OPERATOR

Robert Vích, Martin Vondra

Institut für Photonik und Elektronik,
Akademie der Wissenschaften der Tschechischen Republik
vich@ufe.cz

Abstract: Der von Kaiser definierte Teager Energie Operator ist eine nichtlineare Transformation des Sprachsignals, die sich in vielen Aufgaben der Sprachsignalverarbeitung als ein wirkungsvolles Instrument erwiesen hat. Als Beispiel dient die Grundfrequenzanalyse, die Spracherkennung, Geräuschreduktion, etc. Besonders der kleine numerische Aufwand des Teager Energie Operators, seine schnelle Adaptivität und seine Filterungseigenschaft ist attraktiv. In dem Beitrag werden einige Eigenschaften der zeitdiskreten Variante des Teager Energie Operators untersucht, Zusammenhänge mit anderen Signalanalyse Verfahren angedeutet und seine Anwendung an bandbegrenzten Signalen präsentiert.

1 Einführung

Verschiedene Laute unterscheiden sich untereinander in der Feinstruktur ihrer Zeitverläufe innerhalb der glottalen Zyklen. Bei der üblichen Frequenzanalyse mit Hilfe der kurzzeitigen Fourier Transformation über mehrere Grundfrequenzperioden geht diese Feinstruktur verloren. Diese Tatsache motivierte die Autoren Teager eine neue nichtlineare Signaltransformation zu definieren, die nur mit kleiner Verzögerung den Verlauf der Signalenergie verfolgt. Diese Transformation wurde später von Kaiser als Teager Energie Operator (TEO) bezeichnet und erweitert [1, 2, 3]. Dieser Operator entspricht der Energie, die das Signal in dem entsprechenden System generiert hat. In diesen Arbeiten und auch in Beiträgen von anderen Autoren wurde diese Signalrepräsentation z.B. für die Analyse der Amplituden- und Phasenmodulation ausgearbeitet. Im Laufe der Zeit erschienen mehrere Artikel über weitere Anwendungen des TEO in der Sprachverarbeitung, z.B. für Grundfrequenzanalyse [4], robuste Spracherkennung [5], [9], Sprachenerkennung [6], Geräuschreduktion [7], Stresserkennung [8], Emotionserkennung [10], Sprache-Pausen Erkennung [11], etc.

In diesem Beitrag wollen wir nicht die Einführung und die klassischen Beispiele der Anwendung des Teager Energie Operators wiederholen, die sind in den Arbeiten [1-3] zu finden. Wir werden den Teager Energie Operator für diskrete Signale ableiten, seine Eigenschaften beschreiben, Zusammenhänge mit anderen Signalanalyse Verfahren andeuten und an einfachen Beispielen auf graphische Weise mit klassischer Signalanalyse vergleichen.

2 Teager Energie Operator

Kaiser definierte den Teager Energie Operator [1-3] für kontinuierliche Signale $x(t)$ mit dem Ausdruck

$$\Psi_c(x(t)) = \left(\frac{d}{dt}x(t)\right)^2 - x(t)\frac{d^2}{dt^2}x(t). \tag{1}$$

Die Form des Energie Operators für diskrete Signale erhalten wir nach der Abtastung des Signals $x(t)$, durch Substitution der ersten Derivation mit der aufsteigenden ersten Differenz

und der zweiten Derivation mit der symmetrischen zweiten Differenz. Das führt nach kleiner Umordnung zu der Form des diskreten nichtlinearen Teager Energie Operators

$$\Psi\left(x[n]\right)=x^2[n]-x[n-1]x[n+1], \tag{2a}$$

wo $\Psi(\bullet)$ den diskreten TEO bedeutet und $x[n]$ sein diskretes Eingangssignal. Für das Ausgangssignal des TEO, die Teager Energie, werden wir die Bezeichnung $\Psi\left(x[n]\right)=t[n]$ einführen. Für diskrete kausale Signale $x[n]$, mit der Voraussetzung $x[n]$=0 für n<0, gilt weiter

$$\Psi\left(x[0]\right)=t[0]=x^2[0]. \tag{2b}$$

Das Ausgangssignal $t[n]$ hat die Dimension der Energie und entspricht der Energiedichte, d.h. der Energie per Abtastwert, kann aber auch negative Werte annehmen, was aus (2a) ersichtlich ist. In manchen Anwendungen muss man damit rechnen.

Der TEO kann auch mit Hilfe der symmetrischen Differenz $d[n]=x[n+1]-x[n-1]$ und der symmetrischen Summe $s[n]=x[n+1]+x[n-1]$ ausgedrückt werden. Es gilt

$$\begin{aligned}\Psi\left(x[n]\right)&=t[n]=x^2[n]-x[n-1]x[n+1]=\\&=x^2[n]-\frac{1}{4}\left(s^2[n]-d^2[n]\right)=x^2[n]-\left(\left(\frac{s[n]}{2}\right)^2-\left(\frac{d[n]}{2}\right)^2\right),\end{aligned} \tag{3a}$$

oder auch

$$t[n]=x^2[n]-\frac{1}{4}\left(s[n]-d[n]\right)\left(s[n]+d[n]\right)=x^2[n]-\frac{s[n]-d[n]}{2}\cdot\frac{s[n]+d[n]}{2}. \tag{3b}$$

Die Varianten (3a) und (3b) können mit Vorteil für physikalische Interpretation verwendet werden.

Die nichtlineare Transformation des Eingangssignals $x[n]$ mit dem TEO in das Ausgangssignal $t[n]$ kann nach Abb. 1 dargestellt und als Signalvorverarbeitung angesehen werden.

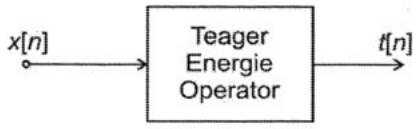

Abbildung 1 –Teager Energie Transformation für diskrete Signale.

Für diskrete im breiten Sinn stationäre Zufallssignale $x[n]$ gilt weiter

$$\begin{aligned}E\left\{t[n]\right\}&=E\left\{\Psi\left(x[n]\right)\right\}=E\left\{x^2[n]-x[n+1]x[n-1]\right\}=\\&=E\left\{x^2[n]\right\}-E\left\{x[n+1]x[n-1]\right\}=r_x[0]-r_x[2],\end{aligned} \tag{4}$$

wo $E\{\bullet\}$ den Erwartungswert und $r_x[n]$ die Autokorrelationsfolge des Signals $x[n]$ bedeuten. Man kann den dimensionslosen Teager Energie Quotienten Q definieren

$$Q = \frac{E\{t[n]\}}{E\{x^2[n]\}} = \frac{r_x[0] - r_x[2]}{r_x[0]} = 1 - \frac{r_x[2]}{r_x[0]} = \frac{\sum_{n=0}^{N-1} t[n]}{\sum_{n=0}^{N-1} x^2[n]}, \tag{5}$$

der für die Klassifizierung von Signalen gebraucht werden kann. Mit Hilfe der Eigenschaften der Autokorrelationsfolge $r_x[n]$ kann man beweisen, dass $Q \geq 0$ gilt. Den Wert $Q = 0$ erhalten wir für Signale, bei denen $r_x[2] = r_x[0]$. Im nächsten Kapitel werden wir als Beispiele einige allgemeine Signale und Sprachsignale analysieren.

Es sei weiter das Signal $x[n] = u[n] + v[n]$ durch additive Mischung von zwei Signalen gegeben. Die Signale $u[n]$ und $v[n]$ sind unabhängig, nicht korreliert und haben die Mittelwerte gleich Null. Dann gilt

$$t[n] = t_u[n] + t_v[n] + \tilde{\Psi}(u[n]v[n]), \tag{6}$$

wo $\tilde{\Psi}(u[n]v[n]) = t_{uv}[n]$ die Kreuz Teager Energie bedeutet. Unter den oben genannten Bedingungen für die Signale $u[n]$ und $v[n]$ gilt $\tilde{\Psi}(u[n]v[n]) = t_{uv}[n] = 0$. Man kann weiter feststellen, dass für viele Rauschsignale $t[n] \cong 0$ ist.

Für die Analyse der Eigenschaften der Teager Energie $t[n]$ transformieren wir die Signale $x[n]$ und $t[n]$ in den Spektralbereich mit Hilfe der diskreten Fourier Transformation. Es sei $X[k] = DFT\{x[n]\}$ und $T[k] = DFT\{t[n]\}$. Um eindeutig die Frequenzverläufe $X[k]$ und $T[k]$ zu vergleichen und zu interpretieren, werden wir das Leistungsspektrum $|X[k]|^2$ mit dem Spektrum $|T[k]|$ darstellen.

Die Signalenergie kann nach dem Theorem von Parseval in der Zeit- oder Frequenzebene berechnet werden und es gilt für das Signal $x[n]$

$$\sum_n x^2[n] = r_x[0] = \frac{1}{K} \sum_k |X[k]|^2. \tag{7}$$

Analog kann man auch das Theorem von Parseval für die Teager Energie $t[n]$ formulieren.

$$\sum_n t^2[n] = r_t[0] = \frac{1}{N} \sum_k |T[k]|^2, \tag{8}$$

wo aber $t[n]$ die Dimension der Energie hat, also die Energiedichte im Zeitbereich bedeutet und $T[k]$ der spektralen Energiedichte entspricht.

In der Auswertung und Interpretation von Signalen benutzt man mit Vorteil die Korrelationsanalyse und Cepstralanalyse. Das Ausgangssignal $t[n]$ des Teager Operators kann auch auf diese Weise transformiert werden. Man kann eine nichtlineare Transformation in dem Spektralbereich, d.h. des Betragsspektrums $|T[k]|$ einführen [12], die nach der Rücktransformation in den Zeitbereich die gesuchte transformierte Folge $\breve{t}[n]$ liefert. Als

Transformationsfunktion kann eine passende Funktion $f(|T[k]|)$ benutzt werden. Für $f(|T[k]|) = \log(|T[k]|)$ erhalten wir nach der inversen diskreten Fourier Transformation das reale Cepstrum und für $f(|T[k]|) = (|T[k]|)^2$ die Autokorrelationsfolge. Man kann auch die allgemeine Potenzfunktion einführen $f(|T[k]|) = (|T[k]|)^{\gamma}$, wo $-2 \le \gamma \le 2$. Die transformierte Folge $\bar{t}[n] = IDFT\{f(|T[k]|)\}$ kann als reales Pseudocepstrum oder auch Pseudo-Korrelationsfolge der Folge $t[n]$ bezeichnet werden.

3 Signal und Teager Energie Eigenschaften im Zeit- und Frequenzbereich

Als Beispiel für die Analyse der Eigenschaften der Teager Energie benutzen wir Segmente von Sprachlauten und auch Segmente von ausgewählten Signalen. Die Signale wurden mit der Abtastfrequenz 8kHz aufgenommen. In Tabelle 1 sind die Energie Quotienten Q für diese Signale angegeben. Die Laute *a, e, i, o, u, m, n, s* entsprechen einer Männerstimme. Das Signal Sinus ist eine Sinusfolge mit der Frequenz 200 Hz, Rand. ist eine Pseudozufallsfolge mit gleichmäßiger Verteilung und PKW bedeutet das Rauschen in einem PKW Skoda Felicia.

Signal	***a***	***e***	***i***	***o***	***u***	***m***	***n***	***s***	**Sinus**	**Rand.**	**PKW**
Q	0,658	0,326	0,082	0,425	0,106	0,036	0,065	1,100	0,049	0,991	0,034

Tabelle 1 – Energie Quotienten für Laute, Sinusfunktion, Zufallsfolge und PKW Rauschen.

Um eine Ansicht über den Zusammenhang der Verläufe des Signals, der Teager Energie und des Leistungsspektrum der Signale zu gewinnen, sind in Abbildungen 2-5 die entsprechenden Verläufe $x[n]$, $t[n]$ und $|T[k]|$ dargestellt. Es wurden wegen Raummangel nur spezifische Laute *a, n,* und *s* und das PKW Geräusch ausgewählt. Wie zu sehen ist, existieren Stellen, wo die Teager Energie $t[n] < 0$. Weiter ist aus den Spektralverläufen die starke signalabhängige Filterungseigenschaft des TEO zu sehen.

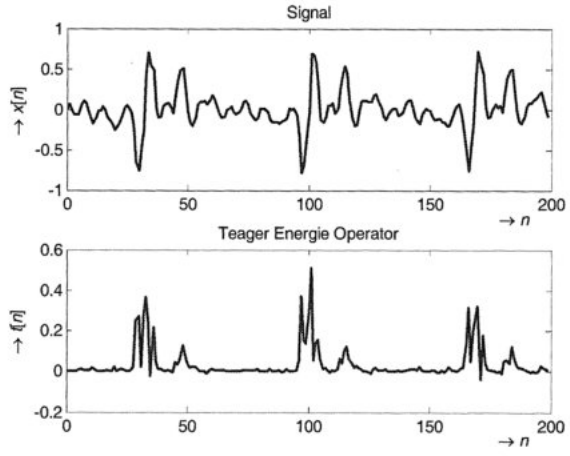

Abbildung 2a – Laut *a* – Signal und Teager Energie.

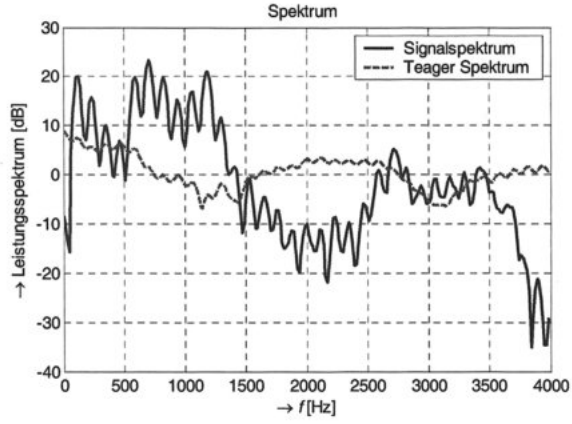

Abbildung 2b – Laut *a* – Spektrum.

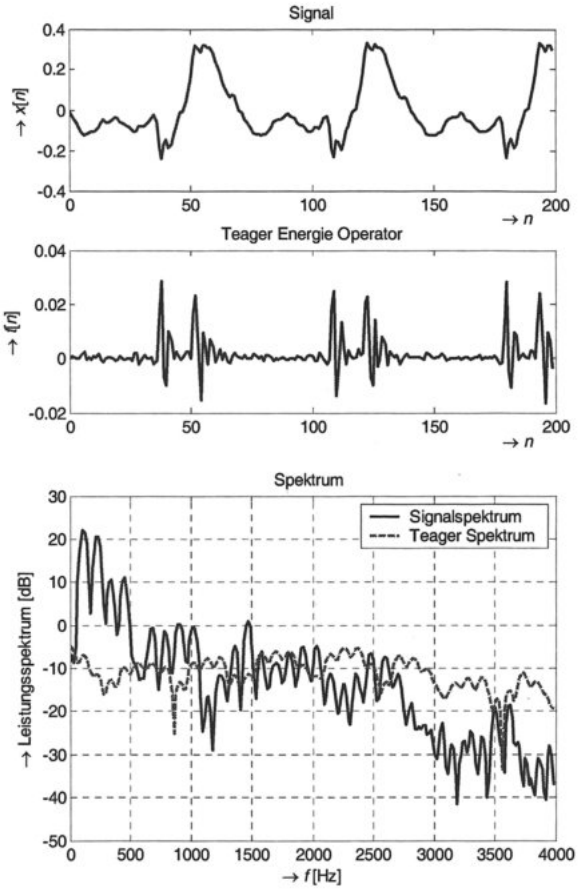

Abbildung 3 – Laut *n*.

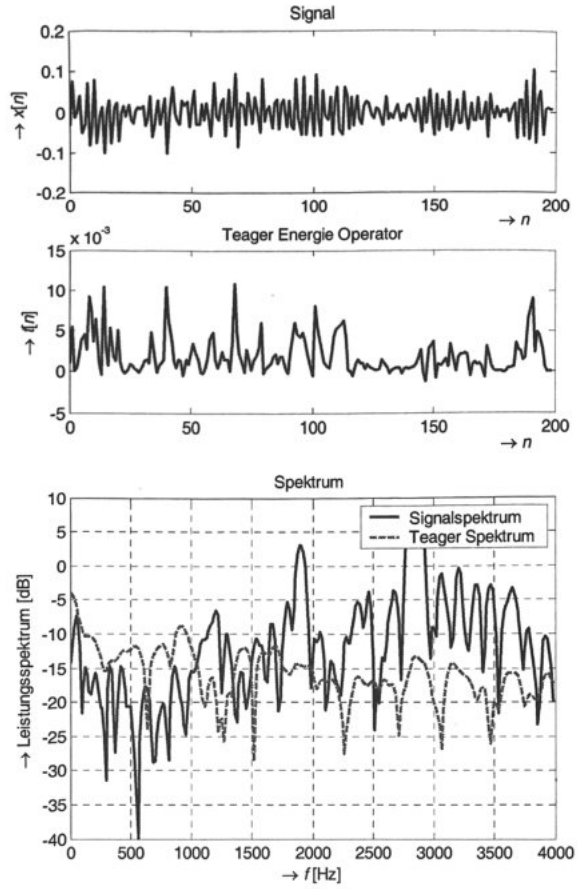

Abbildung 4 – Laut *s*.

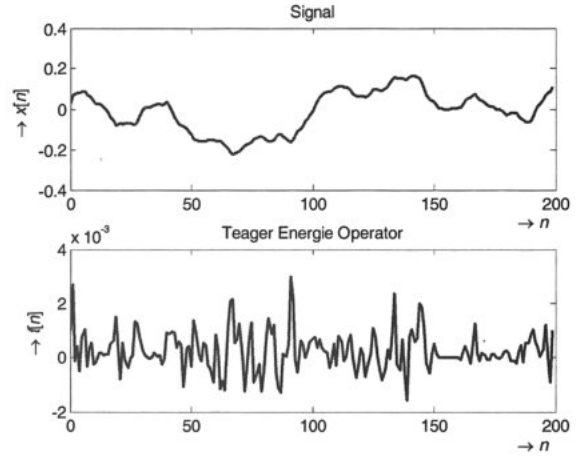

Abbildung 5a – PKW Rauschen, Signal und Teager Energie.

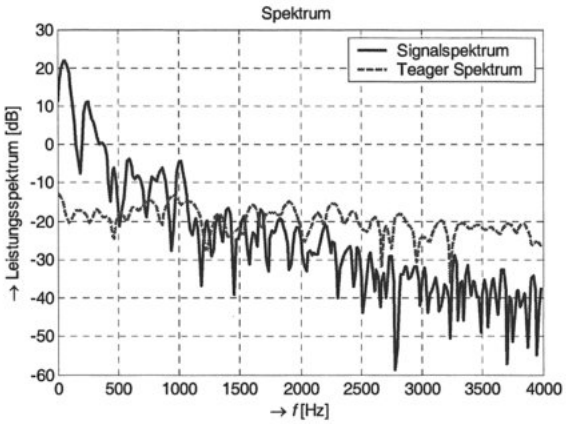

Abbildung 5b – PKW Rauschen – Spektrum.

4 Schlussbetrachtung

In dem Beitrag wurden einige nicht traditionelle Überlegungen und Analysen präsentiert, um einen Einblick in die Eigenschaften des TEO und seine nichtlineare Filterungseigenschaften zu gewinnen. In unserem Beitrag [13] ist die Anwendung der Teager Energie bei der Erkennung von Emotionen ausgewertet. Weitere spezielle hoch entwickelte Verfahren konkreter Anwendungen sind in der zitierten Literatur zu finden. Man kann aber behaupten, dass das volle Erkennen der Möglichkeiten und die praktische Ausnutzung der Eigenschaften des Teager Energie Operators noch zu erforschen ist.

Danksagung

Die vorliegende Arbeit wurde im Rahmen des Projektes „Informationsgesellschaft“ von der Grantagentur der Akademie der Wissenschaften der Tschechischen Republik unter dem Kennzeichen 1ET301710509 unterstützt.

Literatur

[1] Maragos, P., Kaiser, J. F., Quatieri, T. F.: On Amplitude and Frequency Modulation Using Energy Operators. IEEE Transactions on Signal Processing, Vol. 41, No. 40, April 1993, S. 1532-1550.

[2] Maragos, P., Kaiser, J. F., Quatieri, T. F.: Energy Separation in Signal Modulations with Application to Speech Analysis. IEEE Transactions on Signal Processing, Vol. 41, No. 10, October 1993, S. 3024-3050.

[3] Quatieri, T., F.: Discrete-Time Speech Signal Processing. Principles and Practice. Prentice Hall PTR, 2002, Kap. 11.5.

[4] Rouat, J., Liu, Y. C., Morissette, A: Pitch Determination and Voiced/Unvoiced Decision Algorithm for Noisy Speech. Speech Communication, Vol. 21, 1997, S. 192-207.

[5] Jabloun, F., Cetin, A. E., Erzin, E.: Teager Energy Based Feature Parameters for Speech Recognition in Car Noise. IEEE Signal Processing Letters, Vol. 6, No. 10, October 1999, S. 259-261.

[6] Patil, H. A., Basu, T. K.: Cepstral Domain Teager Energy for Identifying Perceptually Similar Languages. Springer, LNCS 4815, 2007, S. 455-462.

[7] Bahoura, M., Rouat, J.: Wavelet Speech Enhancement Based on the Teager Energy Operator. IEEE Signal Processing Letters, Vol. 8, No. 1, January 2001, S. 1-12.

[8] Fernandez, R., Picard, R. W.: Modelling Drivers Speech Under Stress. Speech Communication, Vol. 40, 2003, S. 145-159.

[9] Dimitriadis, D., Maragos, P., Potamianos, A.: Auditory Teager Energy Cepstrum Coefficients for Robust Speech Recognition. Interspeech 2005, Lisbon, September 2005, S. 3013-3016.

[10] Ververidis, D., Kotopoulos, C.: Emotional Speech Recognition: Resources, Features, and Methods. Speech Communication, Vol. 58, 2006, S. 1162-1181.

[11] Chen, S.-H. et al: Robust Voice Activity Detection Using Perceptual Wavelet-Packet Transform and Teager Energy Operator. Pattern Recognition Letters, Vol. 28, 2007, S. 1327-1332.

[12] Vích, R.: Pseudocepstrale Sprachanalyse und Konstruktion eines Vokaltraktmodells mit endlicher Impulsantwort. In: D. Wolf (Ed.): Signaltheorie und Signalverarbeitung, Akustik und Sprachakustik, Informationstechnik. Studientexte zur Sprachkommunikation, 2003, Vol. 29, S. 126-132.

[13] Vondra, M., Vich, R.: Evaluation of Automatic Speech Emotion Recognition Based on Gaussian Mixture Models. In: 19. Konferenz Elektronische Sprachsignalverarbeitung, 8.-10. September 2008, Frankfurt am Main, (in Druck).