

# NONLINEAR OPERATORS FOR SPEECH ANALYSIS

Jean Rouat

Dépt des sciences appliquées,  
Université du Québec à Chicoutimi,  
CHICOUTIMI, Québec, Canada, G7H 2B1

## 1. Introduction

The research in speech analysis is recognized to be an important aspect in the area of speech processing, with applications in speech coding, speech recognition, etc. Depending on the application, the speech analyzer has to extract the most appropriate parameters. This paper will focus on the problem of speech analysis with possible applications in speech recognition.

The automatic "demodulation" of speech with non-linear operators based on perceptive knowledge is a problem which has not yet been fully addressed, and speech "demodulation" might assist the researcher in the understanding of speech and / or in the design of a simple and efficient speech analysis.

## 2. Modulated tone perception

Since the auditory system does not resolve the higher frequency components, the temporal features of vowel-like sounds are comparable and hence coded similarly as those of amplitude-modulated tones. Research work on automatic demodulation of speech can be motivated because of the hypothesis proposing that the human brain has neural cells specialized in Amplitude Modulation (AM) and Frequency Modulation (FM) detection (Gardner et al. 1979) (Tansley et al. 1983) (Wakefield et al. 1984). More recently (Schreiner and Langner, 1988) have studied the representation of amplitude modulation in the inferior colliculus of the cat and have shown that the inferior colliculus of the cat contains a highly systematic topographic representation of amplitude modulation parameters.

## 3. Basilar membrane nonlinearities

Nonlinearity and perception of intermodulation distortion products ( $f_1-f_2$ ,  $2f_1-f_2$ , etc.) are a live issue in hearing research and it is not easy to understand the origin of these nonlinearities. Recently Robles, Ruggero and Rich have observed distortion products on chinchilla basilar membrane by using a laser-velocimetry technique (Ruggero and Rich, 1991). Their work suggests that the lived basilar membrane is a nonlinear system and thus the perception of distortion products could be due to the basilar membrane response and not only to the neural-postprocessing.

## 4. Nonlinear operators

Recently J.F. Kaiser (1990) proposed a nonlinear operator (called Teager energy operator) able to extract the energy of a signal based on mechanical and physical considerations. It has been shown (Maragos, Quatieri and Kaiser, 1991) that this operator is able to track the amplitude of an A.M. signal or the frequency of an FM signal very quickly .

Another nonlinear operator has been proposed (Rouat, 1991). This operator, called "Dyn", shows ability to enhance the AM-FM modulation in speech, and it is interesting to compare it to the Teager energy operator. Figure 1 illustrates the output of Dyn and Teager energy operators. Depending on the application, one can use the Dyn or the Teager operators. The top section of Fig. 1 (a) shows the original speech ( $/a/$ ) of a male speaker ( three pitch period ). The second section presents the band-pass-filtered speech (Moore and Glasberg, 1983) with a centre frequency of 1400Hz. The third

section shows the output of the Dyn operator on the band-pass-filtered speech. And the fourth section illustrates the output of the Teager energy operator on the same band-pass-filtered speech. Fig. 1 (b) presents the same output, for a /i/ with a centre frequency of 2300Hz ( three pitch period ). Dyn and the Teager energy operators show the modulated energy pulses characteristic of the speech signal. The Teager energy operator has an output which does not need to be post-processed when the speech has been properly recorded and band-pass-filtered.

## 5.The analysis of speech

In this section, we describe how we use the Teager energy and Dyn operators for extraction of amplitude and/or frequency modulation in speech.

### a) filtering

The actual version of the analyzer is comprised of a bank of twenty-four filters centred from 330Hz to 4700Hz. These filters simulate partially the frequency analysis performed by the cochlea. These are rounded exponential filters with the Equivalent Rectangular Bandwidths (ERB) proposed by Patterson ( Patterson, 1976), Moore and Glasberg (Moore and Glasberg, 1983). The output of each filter is a bandpass signal with a narrow-band spectrum centred around  $f_i$  where  $f_i$  is the central frequency (C.F.) of channel  $i$ . According to communication theory (Carlson, 1986) the output signal  $s_i(t)$  from channel  $i$  can be considered has being modulated in amplitude and phase with a carrier frequency of  $f_i$ .

$$s_i(t) = A_i(t) \cos [\omega_i t + \phi_i(t)] \quad (1)$$

$A_i(t)$  is the modulating amplitude and  $\phi_i(t)$  is the modulating phase. It should be noticed that equation (1) is true only for a bandpass signal (bandwidth of  $A_i(t)$  and  $\phi_i(t)$  small in comparison to  $f_i$ ).

### b) AM - FM demodulation

The output of the analog version of the Teager energy operator (Kaiser, 1990) to  $s_i(t)$  is given by

$$(\dot{s}_i(t))^2 - s_i(t) \cdot \ddot{s}_i(t) .$$

It is possible to show (Rouat, 1991) that:

$$\begin{aligned} \text{Teager } (s_i(t)) = & A_i^2(t) [\omega_i + \dot{\phi}_i(t)]^2 + \frac{\dot{A}_i^2(t) - A_i(t) \ddot{A}_i(t)}{2} + \frac{\dot{A}_i^2(t) - A_i(t) \ddot{A}_i(t)}{2} \cos [2\omega_i t + 2\phi_i(t)] \\ & + \frac{A_i^2(t) \ddot{\phi}_i(t)}{2} \sin [2\omega_i t + 2\phi_i(t)] \end{aligned} \quad (2)$$

with  $\dot{A}_i(t)$ , and  $\dot{\phi}_i(t)$  being the time derivative of  $A_i(t)$  and  $\phi_i(t)$ , and  $\ddot{A}_i(t)$ , and  $\ddot{\phi}_i(t)$  being the second time derivatives.

$[A_i(t) (\omega_i + \dot{\phi}_i(t))]^2$  is related to the modulating amplitude times the instantaneous frequency. Thus, depending on the modulation (AM or FM) the Teager energy operator will automatically demodulate the signal. The other terms in equation (2) will be considered as being "noisy" terms ( for the purpose of the paper ) and are , most of the time, much smaller than  $[A_i(t) (\omega_i + \dot{\phi}_i(t))]^2$  when  $\omega_i$  and  $A(t)$  are large enough.

The output of the Dyn operator (Rouat, 1991) to  $s_i(t)$  is given by

$$\begin{aligned} \text{Dyn}(s_i(t)) &= s_i(t) \cdot \dot{s}_i(t) \\ &= \frac{1}{4} \frac{d}{dt} [A_i^2(t)] + \frac{1}{2} A_i^2(t) \sqrt{\left(\frac{\dot{A}_i(t)}{A_i(t)}\right)^2 + \left[\frac{d}{dt}(\omega_i t + \phi_i(t))\right]^2} \cdot \cos [2(\omega_i t + \phi_i(t)) + \zeta(t)] \quad (3) \end{aligned}$$

$$\text{with } \zeta(t) = \arctg \frac{A_i^2(t) [\omega_i + \dot{\phi}_i(t)]}{\dot{A}_i(t)}$$

It can be demonstrated (Rouat, 1991) that it is possible to extract  $\frac{d}{dt} (\log A_i(t))$  (time derivative of the logarithm of the amplitude modulation) and  $\frac{1}{2\pi} \cdot \frac{d}{dt} [\omega_i t + \phi_i(t)]$  (instantaneous frequency modulation) from the output of Dyn ( $s_i(t)$ ) by using standard communication operations (normalisation, envelope detection,...) (Carlson 86), with some restrictions on the speech.

### c) Speech data

The speech data has been provided by the organiser of the workshop and we present the analysis output for a speech segment called "clean.dip". This segment includes a /k/ and a /a/ from the word "can" spoken by a female. The speech has been sampled at 20kHz after low-pass-filtering. The power of the signal is low because the dynamic range of the quantizer in the A/D converter has not been fully exploited. Thus, the quantization effect is important and seems to affect the performance.

### d) Experiment with the Teager Energy operator

After filtering, the output of each channel has been processed by the Teager energy operator and the raw data are plotted as three dimensional images. No smoothing or low-pass-filtering of Teager energy output has been made. As a matter of fact,  $A_i^2(t) [\omega_i + \dot{\phi}_i(t)]^2$  is dominant in expression (2) when  $A_i(t)$  and  $\dot{\phi}_i(t)$  have a bandwidth small in comparison to  $f_i$  ( $\omega_i = 2\pi f_i$ ) and when  $A_i(t)$  and  $\omega_i$  are large. But one should be careful, as depending on the speech, and for low frequency channels, the "noisy" terms of expression (2) can be large and Teager energy operator might not be always reliable.

### e) Experiment with the Dyn operator

The output of each channel has been processed by the Dyn operator. In the first experiment the raw data coming out from Dyn operator are plotted as a three dimensional image. In a second experiment, the output of the Dyn operator has been post-processed (low-pass-filtered) to extract the expression  $\frac{1}{4} \frac{d}{dt} [A_i^2(t)]$  from equation (3) before plotting the image.

## 6. Analysis of the output

Fig. 2, fig. 3 and fig. 4 present the output of the 3D analysis of /k//a/ Teager energy operator, Dyn and low-pass-filtered Dyn. The horizontal scale is the time, the vertical scale is expressed in ERB. The third dimension is the output of Teager energy operator, or Dyn or low-pass-filtered Dyn. The darkness is proportionnal to the output value of the analysis.

The output of Teager energy operator is most of the time positive, as  $A_i^2(t) [\omega_i + \dot{\phi}_i(t)]^2$  is dominant in equation (2). Therefore, the white background in figure 2 corresponds approximately to zero.

The Dyn output can be as well negative or positive, giving a gray background on the images for figures 3 and 4. The white portion on the images correspond to negative values of the output analysis.

Both operators give a rich representation of the /k/ with a typical "v" shape. Teager energy operator enhances the high frequency channels in comparison to Dyn. With the Teager energy operator the plosive part of /k/ dominates the output of the image. Both Teager and Dyn operators allow an easy detection of /k/. On the three figures, one can observe that the medium frequency components are not exactly in phase with the low frequency components, which seems to be typical of the vowel /a/.

## 7. Conclusion

We have proposed a new speech analysis based on nonlinear operators. Results have been presented with the Teager energy and the Dyn operators. These analysis extract automatically information related to an AM or FM signal. With such analysis it seems to be possible to obtain patterns characteristics of phonemes and transition between phonemes, which can not be obtained by using other speech analysis (FFT, LPC) techniques.

From experience on other data provided for the workshop it seems that the Dyn operator is less sensitive to noise and quantization noise than the Teager energy operator, but further experiments have to be made on more speech data. The output of Dyn has to be post-processed in order to obtain the information related to  $A_i(t)$   $\dot{A}_i(t)$  or to the instantaneous frequency. The Teager energy operator does not need such post-processing to get the information related to  $A_i^2(t) [\omega_i + \dot{\phi}_i(t)]^2$ .

In summary, speech analysis experiments have been reported based on two nonlinear operators which are very simple ( 3 points algorithm ) and show ability to enhance AM or FM information in speech . We believe that the understanding of the nonlinearity occurring on the basilar membrane is very important to help in the design of nonlinear analysis enhancing AM and FM components in the speech.

### Acknowledgment:

This work has been supported by the NSERC of Canada under Grant N° OGP0042386, by the "Fonds de Développement Académique Réseau de l'Université du Québec", and by the "fondation" from Université du Québec à Chicoutimi. Many thanks are due to S. Lemieux and Y.C. Liu for their programming work.

## 8. References

- Carlson, A.B. (1986). "Communication systems: An Introduction to Signals and Noise in Electrical Communication". Mc Graw Hill.
- Gardner, R.B. and Wilson J.P. (1979). Evidence for direction-specific channels in the processing of frequency modulation. J. Acoust. soc. Amer. 66, 704-709.
- Kaiser, J.F. (1990). On a simple algorithm to calculate the 'energy' of a signal. Proceedings of IEEE - ICASSP'90, Albuquerque, 381-384.
- Langner, G. and Schreiner, C.E. (1988). "Periodicity coding in the inferior colliculus of the cat.I. Neuronal mechanisms". J. Neurophysiol. Vol. 60, no 6, 1799-1822.
- Maragos, P. Quatieri, T. and Kaiser, J.F. (1991). Speech nonlinearities, modulations and energy operators. Proceedings of the IEEE - ICASSP'91, Toronto, 421-424.
- Moore, B.C.J. and Glasberg, B.R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. 74, 750-753.
- Patterson, R.D. (1976). Auditory filter shapes derived with noise stimuli. Jour. Acoust Soc. Amer. 59 , 3, 640 - 654.
- Plomp, R. (1988). Effect of amplitude compression in hearing aids in the light of the modulation-transfer function. J. Acous. Soc. Amer. 83 (6), 2322-2327.
- Rouat, J. (1991). Dyn: a nonlinear operator for speech analysis. Université du Québec à Chicoutimi, dépt des sciences appliquées, rapport interne.
- Ruggero, M.A. and Rich, N.C. (1991) Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration - Hear. Res. 51, 215-230.
- Schreiner, C.E. and Langner, G. (1988). "Periodicity coding in the inferior colliculus of the cat.II. Topographical organization". J. Neurophysiol., Vol 60, no 6, 1823-1840.
- Tansley, B.W. and Suffield, J.B. (1983). Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. J. Ac. Soc. Am. '74, 765-775.
- Wakefield, G.H. and Viemeister, N.F. (1984). Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels ? J. Ac. Soc. Amer. 75, 1588-1592.