A NON LINEAR ANALYSIS FOR CLEAN AND NOISY SPEECH

Jean Rouat, Yong Chun Liu and Sylvain Lemieux
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. It is known that speaker-independent recognition of continuous speech is a very complicated task which has not yet been fully mastered. The better the quality of the analysis, the easier it becomes to recognize what has been spoken.

The automatic "demodulation" of speech with nonlinear 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.Modulation

Research work on automatic demodulation of speech can be justified because of the hypothesis proposing that the humain brain has neural cells specialized in Amplitude Modulation (AM) and Frequency Modulation (FM) detection [1] [7] [8].

It is known from previous studies (see [5], for example) that fluctuations in the envelope rather than in the fine structure of the speech signal carry the information regarding speech intelligibility. The fluctuations of the envelope of a speech signal are dependent on the relative frequency, the relative energy and the relative phase of the formants (for voiced speech). This suggests again an analysis based on a model of speech "modulation".

3.Nonlinear operators

Recently Kaiser [2] 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 [3] that this operator is able to tract the envelope of an amplitude or frequency modulated signal.

Another nonlinear operator has been proposed [6] (called "Dyn") based on perceptive considerations. This operator shows surprising ability to enhance the FM-AM modulation in speech, and can be compared to the Teager energy operator. One could consider that the Teager operator is the envelope of the Dyn operator for vowels. Figure 1 illustrates this idea. Depending on the application, one can use the Dyn or the Teager operators. The top section of Fig. 1. shows four pitch periods taken from the original speech (/i/) of a male speaker (pitch frequency = 130Hz). The second section presents the band-pass-filtered speech [4] with a center frequency of 2300Hz. 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. The Dyn and the Teager operators show the modulated energy pulses characteristics of the speech signal.

Is there a link between these pulses and the formants? Are these pulses characteristic of a nonlinear coupling between the glottal source and the vocal tract? Is it possible to use these nonlinear operators to extract an information suitable for speech recognition? Most of these questions do not have a clear answer yet. And this paper proposes an analysis which can be used to answer the above questions.

4.The analysis

The actual version of the analyzer is comprised of a bank of twenty filters from 300Hz to 3300Hz. These filters simulate the frequency analysis performed by the cochlea. There are rounded exponential filters with the Equivalent Rectangular Bandwidths (ERB) proposed by Moore and Glasberg [4]. The output of each filter is then processed by either the Dyn operator or by the Teager Energy operator depending on the application, see [6] for more details.

This model is very simple and no attempt has been made to represent the exact transduction mechanism from the inner hair cells (for afferent information) or from the outer hair cells (for active feedback). The Dyn operator needs one multiplication and one addition per sample and the Teager operator, two multiplications and one addition per sample.

5.Comments

Figure 2 shows the output of the model for three pitch periods from the transition /d//i/ in the syllable "di" from a male speaker.
It can be seen that the formant information is preserved and that high-frequency components are enhanced in comparison with low-frequency components. The pulses of energy are better seen in high-frequency. From Figure 2, it can be seen that the low-frequency components are not in phase with the high-frequency components.

6.Conclusion

A new analysis has been proposed based on nonlinear transformations of the signal and on perceptive know ledge. It has been shown that this analysis preserves the formant structure [6] on the frequency dimension and has a very good time resolution. This analysis gives a three dimensionnal representation of speech: "mechanical energy", Bark scale and time scale. In comparison with a spectrogram representation, this analysis shows a more important redundancy in frequency (Bark scale) and in time (modulation) and consequently should be more resistant to noise.

7.References

[1] Gardner, R.B. and Wilson J.P. (1979). Evidence for direction-specific channels in the processing of frequency modulation. J. Acoust. soc. Amer. 66, p. 704.
[2] Kaiser, J.F. (1990). On a simple algorithm to calculate the 'energy' of a signal. Proceedings of IEEE-ICASSP'90, Albuquerque, pp. 381-384.

[3] Maragos, P. Quatieri, T. and Kaiser, J.F. (1991). Speech nonlinearities, modulations and energy operators. Proceedings of the IEEE - ICASSP'91, Toronto, pp. 421-424.
[4] Moore, B.C.J. and Glasberg, B.R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. 74, pp. 750-753.
[5] Plomp, R. (1988). Effect of amplitude compression in hearing aids in the light of the modulation-transfer function. J. Acous. Soc. Amer. 83 (6), pp. 2322-2327.
[6] Rouat, J. (1991). Dyn: a nonlinear operator for speech analysis. Université du Québec à Chicoutimi, dépt des sciences appliquées, rapport interne.
[7] Tansley, B.W. and SUFFIELD, J.B. (1983). Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. J. Acoust. Soc. Amer. '74, p. 765.
[8] Wakefield, G.H. and Viemeister, N.F. (1984). Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels ? J. Acoust. Soc. Amer. 75, p. 1588.

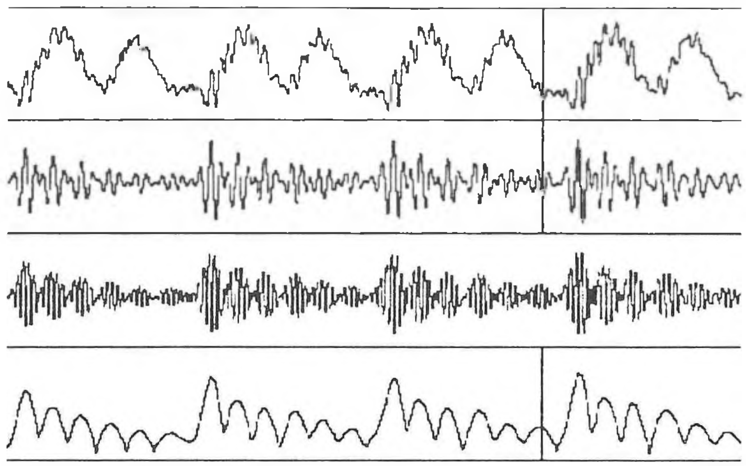

Fig. 1 : nonlinear operators on band-pass-filtered speech, C.F. 2300Hz

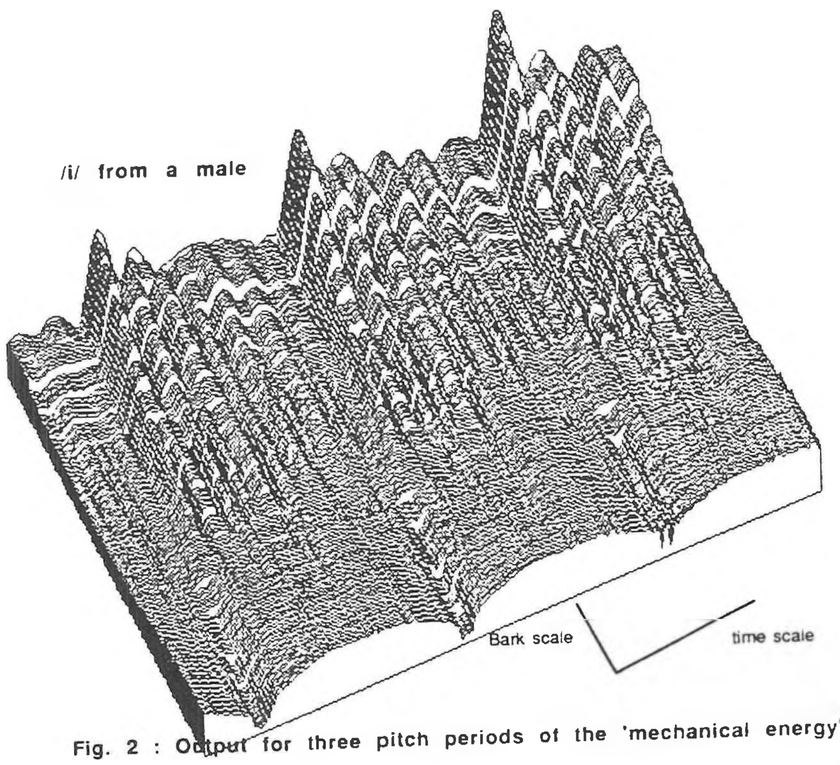

Fig. 2 : Output for three pitch periods of the 'mechanical energy'